

Riassunto CPS

Lorenzo Valentini

21 gennaio 2018

Sommario

Il seguente documento conterrà un riassunto degli appunti e dei concetti visti durante il corso CPS.

Queste appunti non vogliono sostituire il materiale fornito dal docente o in nessun modo sostituire il libro.

... alcuni parti di questo documento sono in costruzione.

Questo documento è distribuito secondo la licenza GNU GPL

Indice

1	Calcolo combinatorio	4
1.1	Permutazioni	4
1.1.1	Permutazioni semplici	4
1.1.2	Permutazioni con ripetizione	4
1.2	Disposizioni	5
1.2.1	Disposizioni semplici	5
1.2.2	Disposizioni con ripetizione	5
1.3	Combinazioni	6
1.3.1	Combinazioni semplici	6
1.3.2	Combinazioni con ripetizione	6
1.4	L'insieme delle Parti	7
1.4.1	Dimostrazione	7
1.5	Principio di inclusione - esclusione	7
1.6	Scombussolamenti	8
1.7	Parti insiemi e sottoinsiemi	8
1.7.1	Partizioni	8
2	Statistica descrittiva	10
2.1	Nozioni	10
2.2	Classi e Istogrammi	11
2.2.1	Classi	11
2.2.2	Istogrammi	11
2.3	Indici di posizione o centralità e di dispersione	12
2.3.1	La moda e la classe modale	12
2.3.2	La media campionaria	12
2.3.3	Media ponderata	14

2.3.4	La varianza e lo scarto quadratico medio	14
2.3.5	La mediana e gli altri quantili	21
3	Spazi di probabilità	24
3.0.1	Fenomeni deterministi e causali	24
3.1	Spazio di probabilità	25
3.1.1	Probabilità uniforme	26
3.1.2	Probabilità condizionale	27
3.2	Eventi indipendenti	29
4	Variabili aleatorie discrete	31
4.1	Variabili discrete	31
4.1.1	Dimostrazione	31
4.1.2	Densità	32
4.2	Densità uniforme	32
4.3	La funzione caratteristica e la densità di Bernulli	33
4.4	schema successo insuccesso	34
4.5	Densità binomiale	34
4.6	Densità geometrica	36
4.6.1	Variabile di densità geometrica modificata	36
4.6.2	Variabile di densità geometrica standard	37
4.6.3	Mancanza di memoria	38
4.7	Densità di Poisson	39
4.8	Densità congiunte e marginali	39
4.8.1	congiunte	40
4.8.2	marginali	40
4.8.3	Osservazioni sulle densità	40
4.9	Variabili aleatorie indipendenti	41
4.9.1	Lemma variabili aleatorie discrete indipendenti	42
4.10	Minimo e Massimo di variabili aleatorie	42
4.10.1	Massimo	42
4.10.2	Minimo	43
4.11	Cambi di variabile	44
4.12	Calcolo della densità	44
4.12.1	Densità binomiali	44
4.12.2	Densità poisson	45

4.13	Media o speranza matematica di variabili aleatorie	45
4.13.1	Bernulli	47
4.13.2	Binomiale	47
4.13.3	Media nei cambi di variabile(importante anche per la varianza)	47
4.13.4	Ipergeometrica	48
4.13.5	Poisson	48
4.13.6	Geometrica modificata	48
4.13.7	Geometrica standard	49
4.13.8	Uniforme	49
4.14	Varianza di variabili aleatorie	49
4.14.1	Bernulli	50
4.14.2	Binomiale	50
4.14.3	Ipergeometrica	50
4.14.4	Poisson	51
4.15	Disuguaglianza di Chebyshev	51
4.15.1	Concetto di convergenza di variabili aleatorie	51
4.16	La legge dei grandi numeri	52
5	Variabili aleatorie continue	53
6	Statistica inferenziale	54
7	Esempi & Esercizi	55

Capitolo 1

Calcolo combinatorio

Riferimenti utili:

- <http://www.webfract.it/MATJAVA/Combina1.htm>
- Questo documento è distribuito secondo la licenza GNU GPL

1.1 Permutazioni

1.1.1 Permutazioni semplici

Si dicono permutazioni semplici di n oggetti distinti le disposizioni semplici di n oggetti distinti di classe n . Il loro numero si indica con $P_n = n!$. Risulta:

$$P_n = D_{n,n} = n!$$

- Conta l'ordine;
- Non sono ammesse ripetizioni.
- Per convenzione $(n)_0 = 1$ e $(n)_1 = n$ e $(n)_n = n!$ e $0! = 1$

1.1.2 Permutazioni con ripetizione

Siano dati n oggetti di cui α identici e $n - \alpha$ distinti. Le permutazioni con ripetizione degli n oggetti sono tutti i gruppi che si possono formare con tutti gli

n oggetti sotto la condizione che sia diverso l'ordine in cui compaiono gli oggetti stessi. In questo caso il numero delle permutazioni con ripetizione è dato da:

$$P_n^\alpha = \frac{P_n}{P_\alpha} = \frac{n!}{\alpha!}$$

Analogamente, dati n oggetti di cui α identici fra loro, β identici fra loro, γ identici fra loro ($\alpha + \beta + \gamma \leq n$), il numero delle permutazioni con ripetizione è dato da:

$$P_n^{\alpha,\beta,\gamma} = \frac{P_n}{(P_\alpha P_\beta P_\gamma)} = \frac{n!}{(\alpha! \beta! \gamma!)}$$

- Ogni gruppo contiene tutti gli elementi;
- conta l'ordine;
- in ogni gruppo un elemento si ripete lo stesso numero di volte.

1.2 Disposizioni

1.2.1 Disposizioni semplici

Si dicono disposizioni semplici (fattoriale discendente) di n oggetti distinti di classe k , con $k \leq n$, tutti i gruppi che si possono formare con k degli n oggetti, considerando distinti i gruppi che differiscono per almeno un elemento o per l'ordine. Il loro numero si indica con $D_{n,k} = (n)_k$.

Risulta:

$$D_{n,k} = n(n-1) \dots (n-k+1) = \frac{n!}{(n-k)!}$$

- Conta l'ordine degli elementi in ogni disposizione.
- Non sono ammesse disposizioni ripetute.
- Per convenzione $(n)_0 = 1$ e $(n)_1 = n$ e $(n)_n = n!$

1.2.2 Disposizioni con ripetizione

Si dicono disposizioni con ripetizione di n oggetti distinti di classe k , tutti i gruppi che si possono formare con k degli n oggetti, eventualmente ripetuti, considerando distinti i gruppi che differiscono per almeno un elemento o per l'ordine. Il loro numero si indica con $DR_{n,k}$. Risulta:

$$DR_{n,k} = n^k$$

- Conta l'ordine;
- Sono ammesse disposizioni ripetute;
- In questo caso k , intero positivo, può essere anche maggiore od uguale ad n .

1.3 Combinazioni

1.3.1 Combinazioni semplici

Si dicono combinazioni semplici di n oggetti distinti di classe k , con $k \leq n$, tutti i gruppi che si possono formare con k degli n oggetti, considerando distinti i gruppi che differiscono per almeno un elemento. Il loro numero si indica con $C_{n,k} = \binom{n}{k}$. Risulta:

$$\binom{n}{k} = C_{n,k} = \frac{D_{n,k}}{P_k} = \frac{n(n-1)\dots(n-k+1)}{k!} = \frac{n!}{[k!(n-k)!]} = \frac{(n)_k}{k!}$$

- Non conta l'ordine;
- Non sono ammesse ripetizioni.
- Per convenzione e $\binom{n}{0} = 1$ e $\binom{n}{1} = \binom{n}{n} = n$

1.3.2 Combinazioni con ripetizione

Si dicono combinazioni con ripetizione di n oggetti distinti di classe k , tutti i gruppi che si possono formare con k degli n oggetti, eventualmente con ripetizione, considerando distinti i gruppi che differiscono per almeno un elemento. Il loro numero si indica con $CR_{n,k}$. Risulta:

$$CR_{n,k} = C_{n+k-1,k} = \frac{n(n+1)\dots(n+k-1)}{k!} = \frac{(n+k-1)!}{[(n-1)!k!]}$$

- Non conta l'ordine;
- Non sono ammesse ripetizioni;
- In questo caso k , intero positivo, può essere anche maggiore od uguale ad n .

1.4 L'insieme delle Parti

L'insieme delle parti: se A è un insieme, l'insieme delle parti $\mathcal{P}(A)$ è l'insieme i cui elementi sono tutti i sottoinsiemi di A incluso \emptyset e A stesso.

1.4.1 Dimostrazione

Sia A tale che $A = \{a_1, a_2, \dots, a_n\}$ è un insieme finito $|A| = n$. Esiste una corrispondenza biunivoca tra $\mathcal{P}(A)$ e le liste in $\{0, 1\}$ di lunghezza n .

$$F : \mathcal{P}(A) \rightarrow \{0, 1\}^n$$

Dove le liste di lunghezza n formate da $\{0, 1\}$ chiamate s_i con i da $i = 1, \dots, 2^n$ aventi 1 in corrispondenza di valori a_j tale che appartengono ad A con $j = 1, \dots, n$

$$s_i = \begin{cases} 1 & \text{se } a_j \in s_i \text{ elemento in pos } k \text{ per } k = 1, \dots, n \\ 0 & \text{altrimenti} \end{cases} \quad (1.1)$$

1.5 Principio di inclusione - esclusione

Si vuole determinare la cardinalità dell'unione di due o più insiemi nel caso siano note le cardinalità tra alcuni di esse.

Introduciamo la funzione caratteristica $\mathcal{X}_A : U \rightarrow \{0, 1\}$ di un sottoinsieme $A \subset U$ dove U generico insieme di elementi.

$$\mathcal{X}_A = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases} \quad (1.2)$$

per ogni $x \in U$.

$$|A| = \sum_{x \in U} \mathcal{X}_A(x)$$

Sussiste la relazione elementare

$$\mathcal{X}_{A \cup B}(x) = \mathcal{X}_A(x) + \mathcal{X}_B(x) - \mathcal{X}_{A \cap B}(x)$$

per $x \in U$.

Ne segue il principio di inclusione-esclusione

$$|A \cup B| = |A| + |B| - |A \cap B|$$

1.6 Scombussolamenti

Sono permutazioni in cui nessun elemento rimane al suo posto. Sia A un insieme di elementi di cardinalità $|A| = n$ il numero di scombussolamenti sarà dato da

$$\begin{aligned} |\bigcap_{i \in I} A_i| &= (n - |I|)! \\ |S| &= \sum_{i=0}^n (-1)^i (n)_{n-i} \end{aligned} \tag{1.3}$$

1.7 Parti insiemi e sottoinsiemi

Definizione formale di partizione. Sia A un insieme di sottoinsiemi $A = \{A_1, A_2, \dots\}$ non vuoti (finito o infinito) si dice partizione di $|A|$ se

- $A = A_1 \cup A_2 \cup \dots$
- $A_i \cap A_j = \emptyset$ per ogni $i \neq j$

Ogni elemento di A sarà contenuto esattamente in uno dei sottoinsiemi A_1, A_2, \dots

1.7.1 Partizioni

Numeri di Bell

Il numero di partizioni in un insieme è detto numero di Bell e si indica con B_n .

Fissato $B_0 = 1$ per convenzione e preso $n > 0$ di un insieme $\{1, \dots, n\}$ dove n compare in un sottoinsieme di cardinalità k .

$$B_n = \sum_{k=1}^n \binom{n-1}{k-1} B_k$$

Numeri di Stirling

Possiamo raffinare il calcolo del numero di partizioni di un insieme finito se vogliamo tener conto anche del **numero di sottoinsiemi coinvolti**.

Chiameremo numeri di Stirling di seconda specie il numero $S_{n,k}$ di partizioni di $\{1, \dots, n\}$ costituiti da k sottoinsiemi.

Per convenzione si pone $S_{0,0} = 1$ e $s_{0,k} = 0$ per ogni $k > 0$.

$$s_{n,k} = \sum_{i=0}^k (-1)^i \frac{(k-i)^n}{i!(k-i)!}$$

Capitolo 2

Statistica descrittiva

2.1 Nozioni

Definizione dei concetti di base. Alcune informazioni (come immagini e definizioni) contenute in questo capitolo sono state prese da Wikiedia e Google.

1. Popolazione

Si definisce **popolazione** l'insieme degli elementi di un insieme (finito o infinito) preso in esame.

Ogni elemento della popolazione è definito come **unità statistica**.

2. Si definisce **carattere** di una popolazione la proprietà che si vuole analizzare: un carattere è una funzione che ha come dominio la popolazione.

I caratteri possono essere quantitativi ovvero che assumono valori discreti o continui che rappresentano una "misura del carattere" o qualitativi.

I **valori** (assunti che possono essere assunti) dal carattere si dicono **modalità**.

3. Si definisce **campione** il sottoinsieme della popolazione preso in esame di cui vogliamo conoscere o conosciamo il valore assunto dal carattere in esame.
Il campione è l'oggetto dell'indagine statistica.
4. Si definisce **frequenza assoluta** di una modalità il numero di volte che si presenta all'interno del campione in analisi.

2.2 Classi e Istogrammi

2.2.1 Classi

Si definiscono classi raggruppamenti di modalità del carattere. Se risultano molto numerose o addirittura infinite è utile raggrupparle.

Le proprietà delle classi sono:

1. Le modalità (valore di un carattere) di un carattere (proprietà degli elementi della popolazione) una volta raggruppati in **classi** vengono considerate nuove modalità.
2. Pertanto anche le classi avranno **frequenze assolute e relative**.
3. L'**ampiezza** è definita come l'ampiezza positiva tra i confini.
4. Il **punto medio** è il valore centrale da i confini(range) della classe

2.2.2 Istogrammi

L'istogramma è un grafico che permette di visualizzare alcune informazioni relative a un carattere. Ad ogni modalità viene associato un rettangolo avente superficie proporzionale alla frequenza della modalità.

- **Base:** Rappresentata dal range dei valori. Se le modalità sono state raggruppate in classi si richiede che la base del rettangolo sia proporzionale all'ampiezza della classe in tutti gli altri casi le basi possono essere prese uguali.
- **Altezza:** Rappresentata dalla frequenza assoluta di quella classe o modalità.

2.3 Indici di posizione o centralità e di dispersione

Gli **indici di posizione** sono singoli valori che sintetizzano la distribuzione di un certo carattere.

2.3.1 La moda e la classe modale

La moda è data dalla modalità con **frequenza assoluta** massima.

Nel caso di classi si parla di classe modale. Questo carattere è di tipo qualitativo. Il carattere in esame definito come (x_1, \dots, x_n) assume i seguenti valori nelle modalità (m_1, \dots, m_k)

$$\begin{aligned} \text{eq:moda } x_i &= \max(m_j) \\ &\text{per } i = 1, \dots, n \\ &\text{per } j = 1, \dots, k \end{aligned} \tag{2.1}$$

2.3.2 La media campionaria

Consideriamo un carattere quantitativo x su un campione di n elementi, dove x_1, \dots, x_n sono i valori assunti dal carattere.

La media campionaria:

$$\bar{x} = \frac{1}{n} \sum_{i=0}^{n-1} x_i \quad (2.2)$$

Osservazione sulla media campionaria

In generale si può affermare che dati due campioni considerati sullo stesso caratteri fare la media del unione dei caratteri non è uguale che fare la media delle medie dei caratteri.

Equivalenze:

Definiamo quanto segue

$$m_1, \dots, m_k$$

modalità assunte dal carattere analizzato.

Con a_i e p_i frequenze assolute e relative. valori assunti dal carattere sono x_1, \dots, x_n dove la modalità 1 m_1 è presente a_1 volte e così via m_k a_k volte.

Le **frequenze assolute** rappresentate come

$$a_1, \dots, a_k$$

il numero di volte con cui la modalità di un carattere si presenta all'interno del campione.

Le **frequenze relative** sono rappresentate

$$p_1, \dots, p_k$$

Inoltre possiamo notare che le frequenze assolute e relative sono legate tra loro dalla seguente relazione

$$p_i = \frac{a_i}{n}$$

Per i quali valgono le seguenti equivalenze:

•

$$\sum_{i=1}^n x_i = \sum_{j=1}^k a_j m_j$$

•

$$\bar{x} = \frac{1}{n} \left(\sum_{j=1}^k a_j m_j \right)$$

•

$$\bar{x} = \sum_{j=1}^k p_j m_j$$

2.3.3 Media ponderata

La **media ponderata** si utilizza se è necessario assegnare un peso (w) ai vari valori ottenuti nel campionamento.

$$\text{media ponderata} = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i}$$

Quindi si può definire la media ponderata come la somma tra i vari prodotti dei valori assunti dal carattere x_1, \dots, x_n e i corrispondenti pesi w_1, \dots, w_n divisi per la somma dei pesi.

Gli **indici di dispersione** servono ad esprimere la compattezza o la rarefazione dei dati. Maggiore è l'indice di dispersione più i dati (valori assunti dal carattere) saranno rarefatti.

2.3.4 La varianza e lo scarto quadratico medio

Lo **scarto quadratico medio** (o **deviazione standard** deviazione standard o scarto tipo) è un indice di dispersione statistico, vale a dire una stima della variabilità di una popolazione di dati o di una variabile casuale.

Lo scarto quadratico medio è uno dei modi per esprimere la dispersione dei dati intorno ad un indice di posizione, quale può essere, ad esempio, la media aritmetica o una sua stima. Ha pertanto la stessa unità di misura dei valori osservati (al contrario della varianza che ha come unità di misura il quadrato dell'unità di misura dei valori di riferimento). In statistica la precisione si può esprimere come lo scarto quadratico medio. Il termine "standard deviation" è stato introdotto in statistica da Pearson nel 1894 assieme alla lettera greca σ (sigma) che lo rappresenta. Il termine italiano "deviazione standard" ne è la traduzione più utilizzata nel linguaggio comune; il termine dell'Ente Nazionale Italiano di Unificazione è tuttavia "scarto tipo", definito come la radice quadrata positiva della varianza per lo meno fin dal 1984.

Se non indicato diversamente, lo scarto quadratico medio è la radice quadrata della varianza,[4] la quale viene coerentemente rappresentata con il quadrato di sigma (σ^2).

Varianza campionaria

La varianza campionaria è definita come segue dalla seguente dimostrazione: A questo punto scegliamo un punto t sulla retta reale e calcoliamo la media dei quadrati delle distanze di t dai valori x_i .

In generale otteniamo la funzione

$$t \rightarrow \sum_{i=1}^n (x_i - t)^2 \mid x_i, t \in \mathbb{R}, i = 1, \dots, n \quad (2.3)$$

Ricordandosi ora che il grafico della funzione $y = ax^2 + bx + c$, con $a > 0$, rappresenta una parabola con concavità verso l'alto e vertice in corrispondenza del punto $-b/2a$. A questo punto l'equazione di

cui (punto 2.3 pagina 15) risolvendo diventa

$$\begin{aligned}
 & \sum_{i=1}^n (x_i - t)^2 = \\
 &= \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2tx_i + t^2) = \\
 &= \frac{1}{n} \sum_{i=1}^n (x_i^2) - \frac{1}{n} \sum_{i=1}^n (2tx_i) + \frac{1}{n} \sum_{i=1}^n (t^2)
 \end{aligned} \tag{2.4}$$

questa equazione ammette un *min* per $t = \bar{x}$, dove \bar{x} corrisponde (2.2 pagina 13) sostituendo all'equazione(2.4) otteniamo

$$\begin{aligned}
 & \sum_{i=1}^n (x_i - \bar{x})^2 = \\
 &= \frac{1}{n} \sum_{i=1}^n (x_i^2) - \frac{1}{n} \sum_{i=1}^n (2\bar{x}x_i) + \frac{1}{n} \sum_{i=1}^n (\bar{x})^2 = \\
 &= \frac{1}{n} \sum_{i=1}^n (x_i)^2 - 2\frac{1}{n}n\bar{x}\bar{x} + \frac{1}{n}n\bar{x}^2 =
 \end{aligned} \tag{2.5}$$

$$\text{semplificando e sostituendo } \frac{1}{n} \sum_{i=1}^n (x_i)^2$$

con la definizione di media(2.2 pagina 13)

$$\begin{aligned}
 &= \bar{x}^2 - 2\bar{x}^2 + \bar{x}^2 = \\
 &= \bar{x}^2 - \bar{x}^2
 \end{aligned} \tag{2.6}$$

Quindi in definitiva abbiamo ottenuto che

$$(\text{varianza}) \sigma_x^2 = \bar{x}^2 - \bar{x}^2$$

La varianza quindi rappresenta un indice della dispersione dei dati. In generale ci dobbiamo aspettare di trovare quasi la piena totalità dei valori registrati nell'intorno $[\pm 3\sigma_x]$ per il quale il centro è individuato dalla media \bar{x} .

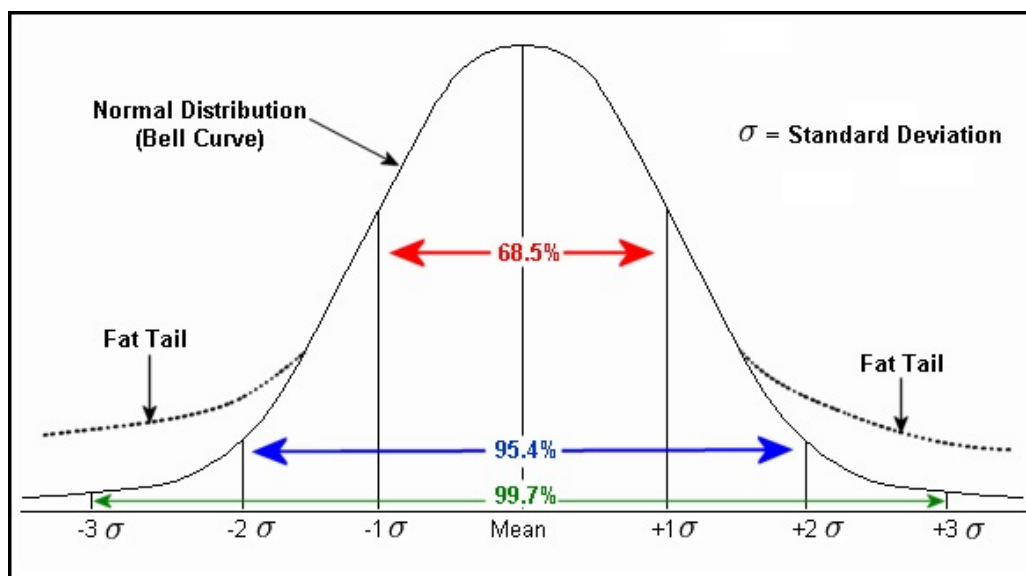


Figura 2.1: L'immagine mostra la Gaussiana o "normal distribution", dove si può notare l'andamento a campana dei valori.(Image google)

Scarto quadratico medio (o deviazione standard) campionario

Lo scarto quadratico medio è uno dei modi per esprimere la dispersione dei dati intorno ad un indice di posizione, quale può essere, ad esempio, la media aritmetica o una sua stima. Ha pertanto la stessa unità di misura dei valori osservati (al contrario della varianza che ha come unità di misura il quadrato dell'unità di misura dei valori di riferimento). In statistica la precisione si può esprimere come lo scarto quadratico medio.

$$\sigma_x = \sqrt{\bar{x^2} - \bar{x}^2} \quad (2.7)$$

Coefficiente di variazione

Il **coefficiente di variazione** o deviazione standard relativa, indicato con σ^* , è un indice di dispersione che permette di confrontare

misure di fenomeni riferite a unità di misura differenti, in quanto si tratta di un numero adimensionale (ovvero non riferito ad alcuna unità di misura). È un indice della precisione di una misura. Come si calcola:

facendo il rapporto tra lo **scarto quadratico medio (o deviazione standard) campionario** e il valore assoluto della **media campionaria**.

$$\sigma_x^* = \frac{\sigma_x}{|\bar{x}|} \quad (2.8)$$

Quello che abbiamo ottenuto è un numero puro senza unità di misura.

Covarianza

In statistica e in teoria della probabilità, la covarianza di due variabili statistiche o variabili aleatorie è un numero che fornisce una misura di quanto le due varino assieme, ovvero della loro dipendenza. Quindi è un indice di variabilità congiunta.

$$\begin{aligned}\sigma_{x,y} &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= \bar{xy} - \bar{x}\bar{y}\end{aligned}\tag{2.9}$$

e quindi abbiamo che la **covarianza è la media del prodotto meno il prodotto delle medie**. Alcuni esempi di covarianza con $(x_i - \bar{x})(y_i - \bar{y})$ teoricamente ci possiamo aspettare che varino così ma non è detto che questo accada necessariamente.

- Positiva

In questo caso diremo che i **caratteri sono direttamente o positivamente correlati** ovvero

x_i (> (maggiore) oppure minore <) \bar{x} e allo stesso modo avremo che y_i (> oppure <) \bar{y} . Direttamente correlati alle proprie medie.

- Negativa

In questo caso diremo che i **caratteri sono inversamente o negativamente** y_i (< oppure >) \bar{y} . inversamente correlati alle proprie medie.

- Uguale 0

In questo caso la covarianza è nulla e i due caratteri si dicono **incorrelati**. Questo non significa che non esista alcun tipo di dipendenza tra le due ma semplicemente che i valori non sono disposti lungo una retta.

Retta ai minimi quadrati

Il metodo dei minimi quadrati (in inglese OLS: Ordinary Least Squares) è una tecnica di ottimizzazione (o regressione) che permette di trovare una funzione, rappresentata da una curva ottima (o curva di regressione), che si avvicini il più possibile ad un insieme di dati (tipicamente punti del piano). In particolare, la funzione trovata deve essere quella che minimizza la somma dei quadrati delle distanze tra i dati osservati e quelli della curva che rappresenta la funzione stessa.

Questa retta di equazione $y = ax + b$ che minimizza la media dei quadrati degli errori commessi stimando i valori y_i del secondo carattere $ax_i + b$.

$$y_i = ax_i + b$$

Considerando una retta generica $y = ax + b$ che dipende dai parametri a e b si vuole ciò è minimizzare la quantità

$$S(a, b) = \frac{1}{n} \sum_{i=1}^n (y_i - ax_i - b)^2$$

che corrisponde:

$$S(a, b) = \sigma_y^2 + a^2 \sigma_x^2 - 2a\sigma_{x,y} + (\bar{y} - b - a\bar{x})^2$$

dove

$$a = \frac{\sigma_{x,y}}{\sigma_x^2}$$

$$b = (\bar{y} - a\bar{x})$$

(2.10)

Coefficiente di correlazione

La covarianza è sensibile ai cambiamenti di scala e per questo si tende ad utilizzare altri tipi di coefficienti per misurare la correlazione tra 2 variabili.

Il **coefficiente di correlazione** è dato dalla formula:

$$\rho_{x,y} = \frac{\sigma_{x,y}}{\sigma_x \sigma_y} \quad (2.11)$$

si nota velocemente che il coefficiente è invariante per cambi di variabile di primo grado riprendiamo ora l'errore medio $Err = S(a, b)$ con i valori ottenuti dalla sostituzione di a e b con

$$a = \frac{\sigma_{x,y}}{\sigma_x^2},$$

$$b = (\bar{y} - a\bar{x}).$$

otteniamo che

$$\begin{aligned} Err &= \sigma_y^2 + \frac{\sigma_{x,y}^2}{\sigma_x^2} - 2 \frac{\sigma_{x,y}^2}{\sigma_x^2} \\ &= \sigma_y^2 - \frac{\sigma_{x,y}^2}{\sigma_x^2} \\ &= \sigma_y^2 (1 - \rho_{x,y}^2) \end{aligned} \quad (2.12)$$

Infine si conclude deducendo che il coefficiente di correlazione è un numero puro sempre compreso tra ± 1 (l'errore è necessariamente non negativo) ciò significa che i dati si dispongono lungo una retta. Maggiore è il suo valore assoluto migliore sarà l'approssimazione ei dati ottenuti tramite la retta ai minimi quadrati.

2.3.5 La mediana e gli altri quantili

Quantili

I quantili di ordine n suddividono i valori assunti dal carattere in n gruppi equinumerosi questo solo dopo aver ordinato in maniere crescente l'insieme dei valori. Casi più noti

- Con $n = 2$ un'unico quantile chiamato **mediana**.
Che può essere individuata come segue a seconda della cardinalità del campione:
 - Nel caso pari si effettua la media dei due valori centrali.
 - Nel caso dispari è individuato dall'elemento in posizione centrale.
- Con $n = 4$, in questo caso i quantili di ordine 4 sono chiamati **quartili**, chiaramente in questo caso il secondo quantile è la mediana. Proprio in questa tipologia di quantili è possibile ricavarsi un **indice di dispersione** chiamato **scarto interquartile** spesso utilizzato in sostituzione al campo di variazione molto più oneroso. Lo scarto interquartile corrisponde alla differenza tra il 3 e il primo quartile.
- Con $n = 100$ i quantili sono chiamati **percentili** spesso utilizzati per indicare valori che gli corrispondono in percentuale ovvero se prendiamo il 90(*novantesimo*) percentile vorrà dire che il 90% dei valori del campione assumono un valore minore uguale a quello.

Altre tipologie di medie

Talvolta è necessario trasformare opportunamente i dati in nostro possesso per ottenere medie più significative. Tali per cui dati A e B intervalli limitato o non dove $f : A \rightarrow B$ necessariamente invertibile.

$$\bar{x}_f = f'\left(\frac{1}{n} \sum_{i=1}^n f(x_i)\right) \quad (2.13)$$

- **Media geometrica**

$f(x) = \log x$ dove la media \bar{x} corrisponde a $\bar{x}_{log} = (x_1 \cdot \dots \cdot x_n)^{\frac{1}{n}}$

$$\star \bar{x}_{log} = \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}} \quad (2.14)$$

- **Media quadratica**

$f(x) = x^2$

$$\bar{x}_{x^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \quad (2.15)$$

- **Media armonica**

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

$$\bar{x}_{\frac{1}{x}} = \left(\frac{1}{n} \sum_{i=1}^n x_i^{-1} \right)^{-1} \quad (2.16)$$

Capitolo 3

Spazi di probabilità

3.0.1 Fenomeni deterministi e causali

Un **fenomeno** è un qualunque accadimento che porta ad un certo risultato appartenente ad un insieme.

Tipologia di fenomeno:

- **Deterministico** che il risultato del fenomeno può essere predetto con esattezza prima della sua effettuazione.
- **Aleatorio** o casuale che non è possibile prevederne il risultato.

L'**insieme di tutti i possibili risultati di un fenomeno** si indica con Ω .

Un **evento** si definisce un sotto insieme di Ω . Un evento è un insieme di possibili risultati di un fenomeno aleatorio.

Una **valutazione di probabilità** è una funzione che associa ad ogni evento un numero tanto più grande quanto più riteniamo che tale evento possa accadere.

Vorremmo certe proprietà sugli eventi ad esempio se A e B sono eventi disgiunti allora verificheremo che $P(A \cup B) = P(A) + P(B)$.

3.1 Spazio di probabilità

Se e_1, e_2, \dots sono eventi diciamo che formano una **successione** di eventi.

def. Famiglia coerente di eventi

Definizione. Sia \mathcal{A} una collezione di sottoinsiemi di **omega** Ω . diciamo che \mathcal{A} è una famiglia coerente di eventi se:

1. $\emptyset, \Omega \in \mathcal{A}$
2. $E^c \in \mathcal{A}$ se $E \in \mathcal{A}$
3. Unioni ed intersezioni e successioni di elementi di \mathcal{A} stanno ancora in \mathcal{A}

$$\bigcap_{i=1}^{\infty} A_i, \bigcup_{i=1}^{\infty} A_i \text{ sono eventi}$$

Una famiglia coerente di eventi \mathcal{A} viene solitamente chiamata **σ -algebra**. Data una **σ -algebra** una probabilità su \mathcal{A} è una funzione

$$P : \mathcal{A} \rightarrow \mathbb{R}_{\geq 0}$$

che soddisfa le seguenti proprietà:

- $P(\Omega) = 1$
- data una **successione** di eventi $\{\mathcal{A}_1, \mathcal{A}_2, \dots\}$ a 2 a 2 disgiunti avrò

$$P\left(\bigcup_{i=1}^{\infty} \mathcal{A}_i\right) = \sum_{i=1}^{\infty} P(\mathcal{A}_i)$$

def. Spazio di probabilità

Definizione. uno spazio di probabilità è una terna (Ω, \mathcal{A}, P) , dove Ω è un insieme, \mathcal{A} è una famiglia coerente di eventi e P una probabilità su \mathcal{A} .

Fasi per lo studio del fenomeno aleatorio:

1. Determinare lo spazio di probabilità che permette di studiare il fenomeno. Questa fase è detta modellizzazione.
2. La seconda fase invece puramente di calcolo matematico consiste nello studio del modello realizzato nella fase uno, ovvero dello spazio di probabilità.

3.1.1 Probabilità uniforme

Se Ω insieme finito dei risultati tutti equiprobabili, dove \mathcal{A} insieme di tutti i possibili sotto insiemi Ω sono a loro volta eventi. Il calcolo della probabilità quindi corrisponde ad un problema di combinatoria che è individuato dal rapporto tra *casi favorevoli* su *casi possibili*.

$$P(\mathcal{A}) = \frac{|\mathcal{A}|}{|\Omega|} \quad (3.1)$$

Dati due eventi qualunque A e B sia $B = (B \cap A) \cup (B \cap A^C)$ dove l'unione è disgiunta ne segue che:

$$P(B) = P(B \cap A) + P(B \cap A^C) \quad (3.2)$$

Unione disgiunta

Definizione.

L'unione di due o più insiemi è detta **disgiunta** se gli insiemi, presi

a due a due, la loro intersezione è l'insieme vuoto \emptyset .

Diremo che A_1, A_2, \dots sono **partizioni di Ω** se:

$$\begin{aligned} A_i \cap A_j &= \emptyset, \forall i \neq j \\ \bigcup_{i=1}^{\infty} A_i &= \Omega \end{aligned} \quad (3.3)$$

Se gli A_i sono anche eventi diremo che A_1, A_2, \dots è una successione di partizione di eventi di Ω . Sia quindi A_1, A_2, \dots **partizione di eventi di Ω** : Avremo che $B = \sum_{i=1}^{\infty} (B \cap A_i)$

$$P(B) = \sum_{i=1}^{\infty} P(B \cap A_i) \quad (3.4)$$

Dalla proprietà (3.2)

- $P(A^C) = 1 - P(A)$ ponendo $B = \Omega$
- $P(A) \leq P(B)$ se $A \leq B$

Ricordiamo anche che

$$\begin{aligned} P(A \cup B) &= P(A) + P(B \setminus A) \\ \text{dove } \setminus \text{ indica l'insieme complemento } B - A \\ &= P(A) + P(B \cap A^C) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned} \quad (3.5)$$

Si ricorda anche che il complementare del unione è uguale all'intersezione dei complementari e quindi

$$P(\cup A_i) = 1 - P(\cap A_i^C) \quad (3.6)$$

3.1.2 Probabilità condizionale

In teoria della probabilità la probabilità condizionata di un evento A rispetto a un evento B è la probabilità che si verifichi A , sapendo che B è verificato. Questa probabilità, indicata $P(A|B)$ probabilità

A dato B o $P_B(A) = P'(A)$, esprime una "correzione" delle aspettative per A , dettata dall'osservazione di B .

Dato uno spazio di probabilità (Ω, A, P) dove A è un evento tale che $P(A) > 0$, formiamo un nuovo spazio di probabilità (Ω, A, P') dove P' è una valutazione di probabilità.

$$P(B | A) = P'(B) = \frac{P(A \cap B)}{P(A)} \quad (3.7)$$

dimostrazione

- $P'(\Omega) = 1$
- *siano B_i disgiunti.*

$$P'(B) = \frac{P(A \cap (B_1 \cup B_2 \cup \dots))}{P(A)} = \frac{P((A \cap B_1) \cup (A \cap B_2) \cup \dots)}{P(A)} = P'(B_1) + P'(B_2) + \dots \quad (3.8)$$

La probabilità $P'(B)$ si dice **probabilità condizionale** di B rispetto ad A o anche

("probabilità di B dato A ") e si indicherà con $P(B | A)$. Due eventi si dicono indipendenti se il verificarsi di uno non modifica in alcun modo il verificarsi dell'altro.

Formula delle probabilità totali

Se A_1, A_2, \dots è una partizione di eventi di Ω si ha per ogni evento B

$$P(B) = \sum_{i=1}^n P(B | A_i)P(A_i)$$

Formula di Bayes

$$\text{Osservando che } P(A | B) = \frac{P(A \cap B)}{P(B)} \text{ e } P(B | A) = \frac{P(A \cap B)}{P(A)} \quad (3.9)$$

deduciamo la **formula di Bayes**

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)} \quad (3.10)$$

3.2 Eventi indipendenti

Due eventi A e B si dicono indipendenti se la probabilità che essi si verifichino **simultaneamente** sia **uguale** alla probabilità che si verifichino **separatamente** con $P(A) \neq 0$ e $P(B) \neq 0$ in modo che le probabilità risultino significative.

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

Similmente si può dimostrare come se gli eventi

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

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

dove

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

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

A questo punto $P'(A)$ e $P'(B)$ sono le probabilità condizionali di ottenere l'evento A o B sapendo che si sia già verificato l'evento B per $P'(A)$ o A per $P'(B)$ questo implica che se gli eventi sono indipendenti $P(A) = P'(A) = P(A|B)$ lo stesso vale per $P(B)$. Ora lo dimostreremo solo per il caso dell'evento A ma vale anche per B

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

come da ipotesi se A e B sono indipendenti

$$\begin{aligned} P(A) &= P'(A) \\ P(A) \cdot P(B) &= \frac{P(A \cap B)}{P(B)} P(B) \\ P(A) \cdot P(B) &= P(A \cap B) \end{aligned}$$

Capitolo 4

Variabili aleatorie discrete

video sulle variabili aleatorie **Definizione:**

Sia (Ω, \mathcal{A}, P) uno spazio di probabilità. Una funzione $X : \Omega \rightarrow \mathbb{R}$ è una variabile aleatoria se

$$\{\omega \in \Omega : X(\omega) \leq a\}$$

è un evento per ogni $a \in \mathbb{R}$.

4.1 Variabili discrete

Una variabile X si dice discreta se i valori che assume sono finiti o numerabili.

4.1.1 Dimostrazione

Data la funzione $X : \Omega \rightarrow \mathbb{R}$ che assume una quantità al più numerabile di valori è una variabile aleatoria se e solo se

$$\{X = t\} \forall t \in \mathbb{R}$$

.

4.1.2 Densità

Def. La funzione $p_X : \mathbb{R} \rightarrow \mathbb{R}$ data da $p(t) = P(X = t)$ si dice **densità** della variabile aleatoria discreta X .

La funzione p_X soddisfa le seguenti proprietà.

1. $p_X(t) \neq 0$ per una quantità al più numerabile di valori di t
2. $p_X(t) \geq 0 : \forall t \in \mathbb{R}$
3. $\sum_{t \in \mathbb{R}} p_X(t) = 1$

Per ogni $A \subset \mathbb{R}$ si ha $\{X \in A\}$ è un evento. Infatti a_1, a_2, \dots i valori assunti da X che stanno anche in A : avremo che $\{X \in A\} = \cup_i \{X = a_i\}$ è unione numerabile di eventi.

4.2 Densità uniforme

Se $A = \{x_1, x_2, \dots, x_n\}$ una variabile X che assume i valori in a tutti con la stessa probabilità $\frac{1}{n}$ si dice uniforme su A . Scriveremo in questo caso

$$X \sim U(A)$$

oppure $X \sim U(x_1, \dots, x_n)$.

La densità è quindi casi favorevoli su casi possibili

$$P(A) = \frac{|A|}{|\Omega|}$$
$$p(t) = \begin{cases} \frac{1}{n}, & \text{se } t \in \{x_1, \dots, x_n\} \\ 0, & \text{altrimenti} \end{cases} \quad (4.1)$$

4.3 La funzione caratteristica e la densità di Bernulli

Se A è un evento dove la sua funzione caratteristica \mathcal{X}_A è data da

$$\mathcal{X}_A(\omega) = \begin{cases} 1, & \text{se } \omega \in A \\ 0, & \text{se } \omega \notin A \end{cases} \quad (4.2)$$

è una variabile aleatoria. Verifichiamo che $\{\mathcal{X}_A \leq t\}$ è un evento per ogni $t \in \mathbb{R}$.

$$\{\mathcal{X}_A \leq t\} = \begin{cases} \emptyset & \text{se } t < 0 \\ A^C & \text{se } 0 \leq t < 1 \\ \Omega & \text{se } t > 1 \end{cases} \quad (4.3)$$

La sua densità è (densità di Bernulli)

$$p(t) = \begin{cases} P(A) & \text{se } t = 1 \\ P(1 - A^C) & \text{se } t = 0 \\ 0 & \text{se } t \neq 0, 1 \end{cases} \quad (4.4)$$

Una variabile aleatoria X che assume solo valori 0 e 1 si dice **variabile di Bernulli**. Se la probabilità di assumere il valore 1 è p quella per 0 è $1 - p$, indicheremo questa variabile con $X \sim B(1, p)$. Possiamo quindi concludere che per ogni evento A la variabile \mathcal{X}_A è una variabile di Bernulli.

$$X \sim B(1, P(A)) \quad (4.5)$$

Requisiti di utilizzo:

- schema successo-insuccesso
- prove indipendenti
- con rimpiazzo
- caso con t finito

4.4 schema successo insuccesso

Una sequenza di fenomeni aleatori ognuno dei quali può dare "SUCCESSO" ovvero che si verifica un certo evento oppure "IN-SUCCESSO" che nn si verifica. X = numero di successi

Un esempio importante è quello a prove indipendenti, cioè:

1. Ogni tentativo ha la stessa probabilità di dare successo.
2. Ogni prova non è influenzata dalle altre.

4.5 Densità binomiale

Questa tipologia di variabile densità si suppone che le prove **siano indipendenti l'una dall'altra** cioè che vengano effettuate tutte nelle medesime condizione, **indipendentemente** dal risultato delle prove precedenti: **schema successo insuccesso con rimpiazzo a prove indipendenti**.

La probabilità di ottenere una sequenza fissata in cui abbiamo t successi e $n - t$ insuccessi è data da $p^t(1 - p)^{n-t}$, dove:

- t è il numero di successi.
- $n - t$ è il numero di insuccessi.
- p è la probabilità di ottenere un successo, $p^t = \prod_{i=1}^t p$ per t successi.
- $(1 - p)$ la probabilità di non ottenere un successo, $(1 - p)^{n-t}$ per $n - t$ insuccessi.
- $\binom{n}{t}$ sono le combinazioni il cui possiamo trovare t successi su n tentativi.

La densità X è definita come

$$X \sim B(n, p)$$

e la chiameremo

variabile binomiale, si calcola:

$$p_x(t) = \begin{cases} \binom{n}{t} p^t (1-p)^{n-t}, & \text{se } t = 0, 1, \dots, n \\ 0, & \text{Altrimenti} \end{cases} \quad (4.6)$$

Requisiti di utilizzo:

- schema successo-insuccesso
- prove indipendenti
- con rimpiazzo
- caso con t finito

Schema successo insuccesso senza rimpiazzo. Fenomeno: estrazione di n palline senza rimpiazzare la pallina estratta. Il numero totale di palline è dato da b bianche e r rosse.

Diremo che la variabile X = numero di successi ottenuti numero di palline bianche ottenute.

Lo spazio Ω è dato da tutte le possibili combinazioni di n oggetti scelti dall'insieme $b + r$, con probabilità uniforme.

L'evento " $X=t$ " è dato da tutte le combinazioni costituite da t bianche e $n - t$ rosse.

- Le t bianche le posso scegliere in $\binom{b}{t}$ modi.
- le $n - t$ rosse sono scelte in $\binom{r}{n-t}$ modi.

La densità di X è definita come

$$X \sim B(n, b, r)$$

e la chiameremo **variabile ipergeometrica**, si calcola:

$$p_x(t) = \begin{cases} \frac{\binom{b}{t} \binom{r}{n-t}}{\binom{b+r}{n}}, & \text{se } t \in \{0, 1, \dots, \min(n, b)\} \\ 0, & \text{Altrimenti} \end{cases} \quad (4.7)$$

Requisiti di utilizzo:

- schema successo-insuccesso
- prove dipendenti
- senza rimpiazzo
- caso con t finito

4.6 Densità geometrica

4.6.1 Variabile di densità geometrica modificata

Si consideri uno schema **successo-insuccesso senza rimpiazzo** caso numerabile. Sia la variabile aleatoria T data dal tempo di primo successo, a questo punto chiameremo

$$T \sim \tilde{G}(p)$$

la variabile aleatoria geometrica modificata di parametro p , dove p è la probabilità di avere successo.

$$p_x(t) = \begin{cases} p(1-p)^{t-1} & \text{se } t = 1, 2, \dots \\ 0 & \text{Altrimenti} \end{cases} \quad (4.8)$$

È la probabilità che il primo successo (o evento in generale) richieda l'esecuzione di k prove indipendenti, ognuna con probabilità di successo p . Se la probabilità di successo in ogni prova è p , allora la probabilità che alla k -esima prova si ottenga il primo successo è $P(T = t) = p(1-p)^{t-1}$ Requisiti di utilizzo:

- schema successo-insuccesso
- prove indipendenti
- con rimpiazzo

- caso con t numerabile
- senza l'elemento 0 perchè $T \sim \tilde{G}(p)$ geometrica modificata è equivalente $T - 1 \sim G(p)$ geometrica standard
- calcola la probabilità di fare un certo numero k di tentativi fino ad ottenere il primo successo (al $k - \text{esimo}$ tentativo).

La formula qui sopra è usata, dunque, per **calcolare la probabilità di fare un certo numero k di tentativi fino ad ottenere il primo successo** (al $k - \text{esimo}$ tentativo).

- Qui sotto invece, la seguente distribuzione esprime la probabilità di avere k **fallimenti prima di ottenere il primo successo**.

4.6.2 Variabile di densità geometrica standard

Si consideri uno schema **successo-insuccesso con rimpiazzo** caso numerabile. Sia la variabile aleatoria T data dal tempo di primo successo, a questo punto chiameremo

$$T \sim G(p)$$

la variabile aleatoria geometrica standard di parametro p , dove p è la probabilità di avere successo.

$$p_x(t) = \begin{cases} p(1-p)^t & \text{se } t = 0, 1, 2, \dots \\ 0 & \text{Altrimenti} \end{cases} \quad (4.9)$$

Requisiti di utilizzo:

- schema successo-insuccesso
- prove indipendenti
- con rimpiazzo
- caso con t numerabile

- calcola la probabilità di avere k fallimenti prima di ottenere il primo successo.

4.6.3 Mancanza di memoria

La densità geometrica gode della proprietà di mancanza di memoria ed è l'unica densità che gode di questa proprietà. Sia X una variabile aleatoria che assume solo valori naturali. Diremo che X gode della proprietà di mancanza di memoria se per ogni $k, m \in \mathbb{N}$

$$P(X \geq k + m \mid X \geq k) = P(X \geq m)$$

Nel caso in cui X rappresenti il tempo di attesa affinché un certo evento si verifichi la proprietà di mancanza di memoria dice che aver già aspettato un certo tempo ($T = k$), non influisce in nessun modo sulla probabilità che il tempo che dobbiamo ancora aspettare sia almeno m . Si può dimostrare che dato che

$$\begin{aligned} P(X \geq k + m \mid X \geq k) &= \frac{P(X \geq k + m)}{P(X \geq k)} \\ &= \frac{(1 - p)^{k+m}}{(1 - p)^k} \quad (4.10) \\ &= (1 - p)^m = P(X \geq m) \end{aligned}$$

La proprietà di mancanza di memoria caratterizza la densità geometrica nel senso che se X è una variabile aleatoria discreta che soddisfa la proprietà di mancanza di memoria $P(X \geq k + m \mid X \geq k) = P(X \geq m)$ allora è una variabile geometrica.

Link utili

- Wikipedia:Densità geometrica
- Esempi appunti del prof.

4.7 Densità di Poisson

Questa è una densità ricavata dallo sviluppo del polinomio interpolatore di Taylor.

$$e^x = \sum_{k \geq 0} \frac{x^k}{k!} \quad (4.11)$$

per $x = \lambda$. L'altro limite notevole

$$\left(1 + \frac{x}{n}\right)^n \rightarrow e^x \quad (4.12)$$

Si può verificare che una variabile di Poisson approssimata sotto certe ipotesi sia come una variabile binomiale $X \sim B(n, \frac{\lambda}{n})$.

La densità di Poisson è data da

$$p_x(t) = \begin{cases} e^{-\lambda} \frac{\lambda^t}{t!} & \text{se } t = 0, 1, \dots \\ 0 & \text{Altrimenti} \end{cases} \quad (4.13)$$

dove λ è un parametro strettamente positivo.

Quindi potremo identificare questa variabile come una variabile $X \sim B(n, p)$ dove n è molto grande $n > \{20 \text{ o } 100 \dots\}$ e p molto piccolo $p < \frac{1}{20}$ $np < 10$.

In questo caso la variabile binomiale $X \sim B(n, \frac{\lambda}{n})$ la potremmo approssimare con una variabile aleatoria delle densità di Poisson.

$$X \sim P(n \cdot p)$$

infine avremo che

$$X \sim P(\lambda)$$

4.8 Densità congiunte e marginali

Date X_1, \dots, X_m che sono m variabili aleatorie discrete aventi stessa tipologia di densità e definite su uno spazio di probabilità la

X

$$\underline{X} = (X_1, \dots, X_m) : \Omega \rightarrow \mathbb{R}^m \quad (4.14)$$

4.8.1 congiunte

\underline{X} viene indicata come variabile m -dimensionale. La variabile \underline{X} assume al più una quantità numerabile di valori (se le X_i sono discrete).

Infatti se $\underline{x} = (x_1, \dots, x_m)$, allora

$$\{\underline{X} = \underline{x}\} = \{X_1 = x_1\}, \dots, \{X_m = x_m\}. \quad (4.15)$$

Deduciamo per ogni $\{\underline{X} = \underline{x}\}$ è un evento per ogni $\underline{x} \in \mathbb{R}$.

La densità di una variabile m -dimensionale è data dalla funzione

$$p : \mathbb{R}^m \rightarrow \mathbb{R}_{\geq 0}$$

ottenuta da

$$p(\underline{x}) = P(\underline{X} = \underline{x})$$

Questa densità si dice **congiunta delle variabili** X_1, \dots, X_m .

$$\sum_{\underline{x} \in \mathbb{R}^m} p(\underline{x}) = 1$$

4.8.2 marginali

Quando abbiamo a che fare con m variabili aleatorie X_1, \dots, X_m le **singole densità di queste ultime si dicono densità marginali**, per distinguerle dalla densità congiunta.

Solitamente verranno indicate con p_1, \dots, p_m

4.8.3 Osservazioni sulle densità

Date due variabili aleatorie discrete aventi la stessa densità possiamo calcolarci la **densità congiunta** di queste due sfruttando la tabella a doppia entrata è sempre vero che la somma delle varie righe e delle varie colonne individua le densità marginali delle due variabili. Se X e Y sono due variabili di Bernoulli le loro densità

marginali saranno $X \sim B(1, P(X))$ e $Y \sim B(1, P(Y))$ la tabella a doppia entrata che ci mostra la densità congiunta è riportata qui sotto $p_X(t) \forall t(0, 1)$ e $p_Y(t) \forall k(0, 1)$. quindi è sempre vero sia che

$X \setminus Y$	$k = 0$	$k = 1$	—
$t = 0$	$(0, 0)$	$(0, 1)$	$p_X(0)$
$t = 1$	$(1, 0)$	$(1, 1)$	$p_X(1)$
—	$p_Y(0)$	$p_Y(1)$	—

Tabella 4.1: La figura mostra una tabella a doppia entrata con le densità congiunte e marginali di due variabili aleatorie X e Y

gli eventi siano dipendenti o indipendenti che la somma delle densità congiunte identifica unicamente le densità marginali $p_{(X,Y)}(t, k)$ valori assunti dalle densità congiunte es. $\{(0, 0), (0, 1); (1, 0), (1, 1)\}$.

$$\begin{aligned} p_X(t) &= \sum_{k \in \mathbb{R}} p_{(X,Y)}(t, k) \\ p_Y(k) &= \sum_{t \in \mathbb{R}} p_{(X,Y)}(t, k) \end{aligned} \quad (4.16)$$

4.9 Variabili aleatorie indipendenti

Date n variabili aleatorie X_1, \dots, X_n diremo che esse sono indipendenti se gli eventi $\{X_1 \in A_1\} \dots, \{X_n \in A_n\}$ sono **indipendenti** per ogni scelta degli insiemi $A_i \subseteq \mathbb{R}$.

$$P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n)$$

Dati n punti di $\mathbb{R} = \{x_1, \dots, x_n\}$. Se le variabili X_1, \dots, X_n sono indipendenti se sussiste la relazione:

$$p(\underline{x}) = p_1(x_1) \cdots p_n(x_n)$$

ma è vero anche il contrario se le variabili sono discrete. :)
 Nel caso gli eventi fossero dipendenti la densità congiunta sarebbe
 differente dal prodotto delle marginali.

$$p(\underline{x}) \neq p_1(x_1) \cdots p_n(x_n)$$

4.9.1 Lemma variabili aleatorie discrete indipendenti

Siano X e Y due variabili aleatorie discrete. Allora sono indipendenti se e solo se

$$p_{X,Y}(x, y) = p_X(x)p_Y(y) \quad | \quad \forall \quad x, y \in \mathbb{R}$$

4.10 Minimo e Massimo di variabili aleatorie

Data una qualunque variabile aleatoria X (non necessariamente discreta) la sua funzione di ripartizione è data da

$$F_x : \mathbb{R} \rightarrow \mathbb{R}$$

$$t \mapsto P(X \leq t)$$

4.10.1 Massimo

La domanda in sto caso è qual'è la probabilità che occorran
 almeno k lanci?

Definiamo $Z = \max(S, T)$ due variabili geometriche modificate
indipendenti di parametri p e q .

$$S \sim \tilde{G}(p) \quad T \sim \tilde{G}(q)$$

La funzione di ripartizione di Z se assume solo valori interi non
 negativi è

$$F_Z(k) = P(Z \leq k) = P(\max(S, T) \leq k) = P(S \leq k, T \leq k) = P(S \leq k)P(T \leq k)$$

dove si sfrutta che S e T siano indipendenti.

$$F_Z(k) = (1 - (1 - p)^k)(1 - (1 - q)^k) \quad (4.17)$$

Sostituendo all'equazione otterremo la probabilità di $P(Z = k)$ con 4.17 con i rispettivi $F_Z(k)$ e $F_Z(k - 1)$

$$P(Z = k) = F_Z(k) - F_Z(k - 1) \quad (4.18)$$

sostituendo e semplificando

$$\begin{aligned} &= p(1 - p)^{k-1} + q(1 - q)^{k-1} + (pq - p - q)(1 - p)^{k-1}(1 - q)^{k-1} \\ &= (1 - p)^{k-1}(1 - q)^{k-1} \left(\frac{1}{p(1 - p)^{k-1}} + \frac{1}{q(1 - q)^{k-1}} + (pq - p - q) \right) \end{aligned}$$

4.10.2 Minimo

Se invece fossimo interessati al minimo ovvero lanciamo finchè non otteniamo un successo.

Si consideri $W = \min(S, T)$ per la quale la sua funzione di ripartizione è calcola in $1 - F_W(k) = P(Z > k)$

$$\begin{aligned} 1 - F_W(k) &= P(W > k) = P(\min(S, T) > k) \\ &= P(S > k, T > k) \\ &= P(S > k)P(T > k) \end{aligned} \quad (4.19)$$

Anche in questo caso S e T sono due variabili geometriche modificate indipendenti di parametri p e q .

$$\begin{aligned} 1 - F_W(k) &= (1 - p)^k(1 - q)^k \\ &= (1 - p - q + pq)^k \end{aligned} \quad (4.20)$$

dove otteniamo che

$$F_W(k) = 1 - (1 - p - q + pq)^k \quad (4.21)$$

$$W \sim \tilde{G}(p + q - pq)$$

4.11 Cambi di variabile

Se $\Phi : \mathbb{R}^m \rightarrow \mathbb{R}$ allora $\Phi(\underline{X})$ è una variabile aleatoria discreta se \underline{X} è una variabile aleatoria m -dimensionale discreta. Se \underline{X} assume solo valori $\underline{x}_1, \underline{x}_2, \dots$ allora $\Phi(\underline{X})$ assume i valori $\Phi \underline{x}_1, \Phi \underline{x}_2, \dots$ e quindi $\Phi(\underline{X})$ assume una quantità al più numerabile di valori (al massimo una quantità numerabile di valori).

Dato $y \in \mathbb{R}$ si ha

$$\Phi(\underline{X}) = y = \underline{X} \in \Phi^{-1}(y) = \bigcup_{\underline{x} \in \Phi^{-1}(y)} \{\underline{X} = \underline{x}\} \quad (4.22)$$

questo è un evento perchè unione numerabile di eventi.

quindi i cambi di variabile sono necessari per effettuare delle operazioni su somme di densità di variabili. Ne segue che $\Phi(\underline{X})$ è una variabile aleatoria. Abbiamo inoltre la seguente interpretazione per la densità della variabile $\Phi(\underline{X})$

$$p_{\Phi(\underline{X})}(y) = \sum_{\underline{x} \in \Phi^{-1}(y)} p_{\underline{X}}(\underline{x}) \quad (4.23)$$

4.12 Calcolo della densità

Date due variabili, anche se definite su due spazi completamente diversi, ma hanno la stessa densità, vanno considerate "come se fossero la stessa cosa".

4.12.1 Densità binomiali

Date n variabili U_1, U_2, \dots, U_n indipendenti, tutte di tipo $U_i \sim B(1, p)$ per ogni $i=1, \dots, n$. Ma queste hanno la stessa densità di una variabile binomiale $X \sim B(n, p)$ in uno schema successo-insuccesso senza rimpiazzo. Si ha quindi necessariamente che $U_1 + U_2 + \dots + U_n$. Quindi possiamo dire più in generale date X e Y indipendenti e

aventi stessa densità

$$X \sim B(n_1, p) \quad Y \sim B(n_2, p)$$

$$X + Y \sim B(n_1 + n_2, p)$$

4.12.2 Densità poisson

Supponiamo ora si avere sempre X e Y due variabili aleatorie discrete **indipendenti** rappresentanti la densità di poisson.

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

A questo punto sappiamo che possiamo approssimare una variabile di Poisson con una variabile binomiale di parametri:

$$X \sim B(n, \frac{\lambda}{n}) \quad Y \sim B(m, \frac{\mu}{m})$$

con n e m strettamente maggiori di 0

$$n, m \gg 0$$

ovvero molto grandi dell'ordine della $\geq 10^6$

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

4.13 Media o speranza matematica di variabili aleatorie

$E[X]$ = è uguale alla media ponderata ottenuta moltiplicando i valori assunti dalla variabile X per ogni k per le corrispettive densità $p_X(k)$

$$E[X] = \sum_{k \in \mathbb{R}} k p_X(k)$$

Somma di medie

È un'applicazione lineare $\Phi \mathbb{R}^2 \rightarrow \mathbb{R}$

$$E[X + Y] = \sum_{x,y \in \mathbb{R}} (x + y) \cdot p_{X,Y}(x, y)$$

dalla quale si ricava

$$E[X + Y] = E[X] + E[Y]$$

Prodotto di medie

È un'applicazione lineare $\Phi \mathbb{R}^2 \rightarrow \mathbb{R}$

$$(t, k) \rightarrow t \cdot k$$

$$E[X \cdot Y] = \sum_{t,k \in \mathbb{R}} tk \cdot p_{X,Y}(t, k) = \sum_{t,k \in \mathbb{R}} tk \cdot p_X(t)p_Y(k)$$

dalla quale si ricava

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

Prodotto per uno scalare

$$\begin{aligned} E[cX] &= \\ \sum_{k \in \mathbb{R}} ck \cdot p_X(k) &= \\ c \sum_{k \in \mathbb{R}} k \cdot p_X(k) &= \\ cE[X] \end{aligned}$$

4.13.1 Bernulli

$$X \sim B(1, p)$$

il suo valore atteso sarà uguale

$$E[X] := 0 \cdot (1 - p) + 1 \cdot p$$

di una Bernulli corrisponde

$$E[X] := p$$

4.13.2 Binomiale

$$X \sim B(n, p)$$

il suo valore atteso di una Binomiale

$$E[X] := n \cdot p$$

quindi

4.13.3 Media nei cambi di variabile(importante anche per la varianza)

Definiamo $X \sim B(n, p)$ con densità $p_X(t)$ con $t = 0, \dots, n$ avremo che il valore atteso di $E[X] = np$. Se invece volessimo $E[X^2]$ potremmo fare in due modi il primo prevede di applicare una trasformazione alla densità $X \rightarrow X^2$

$$E[X^2] = \sum_{x \in \mathbb{R}} x \cdot p_{X^2}(x)$$

$$E[X^2] = \sum_{x \in \mathbb{R}} x^2 \cdot p_X(x)$$

4.13.4 Ipergeometrica

Data una variabile ipergeometrica

$X =$ " numero di volte che è uscito (a) nelle (n) estrazioni".

$$X \sim H(n, a, b)$$

il suo valore atteso quindi corrisponderà a quante volte ottengo

$$X \sim B(n, \frac{a}{a+b})$$

$$X_1 \sim B(1, \frac{a}{a+b}), \dots, X_n \sim B(1, \frac{a}{a+b})$$

$$E[X] = E[X_1] + E[X_2] + \dots + E[X_n]$$

$$E[X] = n \cdot \frac{a}{a+b}$$

4.13.5 Poisson

Valore atteso di una variabile di Poisson

$$X \sim P(\lambda)$$

se la approssimiamo ad una variabile Binomiale il calcolo si riconduce a quello di una binomiale ma allora possiamo dire che:

$$X \sim B(n, \frac{\lambda}{n})$$

$$E[X] = n \cdot \frac{\lambda}{n}$$

$$E[x] = \lambda$$

4.13.6 Geometrica modificata

$$X \sim \tilde{G}(p)$$

allora la sua media sarà

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

4.13.7 Geometrica standard

$$X \sim G(p) = X + 1 \sim \tilde{G}(p)$$

allora la sua media sarà

$$E[X] = \frac{1-p}{p}$$

4.13.8 Uniforme

$$X \sim U(\{X_1, \dots, X_n\})$$

dove la speranza matematica è data da

$$E[X] = \frac{1}{n}X_1 + \dots, \frac{1}{n}X_n = \bar{X}$$

equivale ad effettuare la media campionaria delle singole variabili.

4.14 Varianza di variabili aleatorie

Indicheremo la varianza di variabili aleatorie discrete con il simbolo $\text{Var}(X)$ questa come in statistica descrittiva rappresenta la misura della dispersione. Condizione necessaria per calcolare la varianza di variabili aleatorie discrete è che esse siano indipendenti.

$$\text{Var}(X) = E[X^2] - E[X]^2$$

Somma di varianze

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$$

dove la covarianza gode della seguente proprietà:

Se X e Y sono indipendenti allora la $Cov(X, Y) = 0$.

$$Cov(X, Y) = E[X \cdot Y] - (E[X] \cdot E[Y])$$

Questo dalle proprietà viste in precedenza sulla media $E[X \cdot Y] = E[X] \cdot E[Y]$

Prodotto per uno scalare

$$Var(cX) = c^2 Var(X)$$

4.14.1 Bernulli

$$X \sim B(1, p)$$

la sua varianza sarà uguale a

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

4.14.2 Binomiale

$$X \sim B(n, p)$$

la sua varianza sarà uguale a

$$Var(X) = n \cdot p(1 - p)$$

4.14.3 Ipergeometrica

Per questa tipologia di variabile è impossibile poter calcolare la varianza perchè dipendenti.

4.14.4 Poisson

$$X \sim P(\lambda)$$

In questo caso approssimiamo e semplifichiamo

$$X \sim B(n, \frac{\lambda}{n})$$

$$Var(X) = \lambda$$

Il parametro λ in una variabile di Poisson rappresenta sia la **media** che la **varianza** della variabile stessa.

4.15 Disuguaglianza di Chebyshev

La disuguaglianza di Chebishev, verrà utilizzata per calcolare la probabilità che X si discosti dalla sua media di una certa quantità, tale probabilità è proporzionale alla sua varianza. Sia data X variabile aleatoria discreta, $\varepsilon > 0$

$$P(|X - E[X]| > \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

4.15.1 Concetto di convergenza di variabili aleatorie

Definizione

Se X_1, X_2, \dots è una successione di variabili aleatorie diciamo che convergono in probabilità alla variabile aleatoria X se per ogni ε fissato ha

$$\lim_{n \rightarrow \infty} (P(|X_n - E[X]| > \varepsilon)) = 0$$

4.16 La legge dei grandi numeri

Teorema (media)

Sia $X_n \in \mathbb{N}$ una successione di variabili aleatorie discrete indipendenti ed aventi tutte la stessa densità. Con media e varianza espresse come

μ media di X_i

σ^2 varianza di X_i

$$E[\bar{X}_n] = \mu$$
$$Var(\bar{X}_n) = \frac{\sigma_X^2}{n}$$

Allora posto

$$\bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$$

La successione \bar{X}_n converge in probabilità alla variabile aleatoria μ .

Teorema (varianza)

Sia $X_n \in \mathbb{N}$ una successione di variabili aleatorie discrete indipendenti ed aventi tutte la stessa densità.

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

Allora la successione $\bar{\sigma}_n^2$ converge in probabilità alla variabile costante σ^2 .

Capitolo 5

Variabili aleatorie continue

Capitolo 6

Statistica inferenziale

Capitolo 7

Esempi & Esercizi