

Probabilità Appunti

Giovanni "Qua' Qua' dancer" Palma e Alex Basta

Contents

Chapter 1	Introduzione _____	Page _____
Chapter 2	Introduzione _____	Page _____
Chapter 3	Spazi di probabilità _____	Page _____
	3.1 Concetti introduttivi	
	3.2 Regole del calcolo probabilistico	
	Assiomi della probabilità — • Conseguenze degli assiomi —	
	3.3 Probabilità discreta	
	Probabilità uniforme —	
Chapter 4	Probabilità Condizionata _____	Page _____
	4.1 Definizione e motivazioni	
	4.2 Regola della catena	
	4.3 Indipendenza di eventi	
	Generalizzazione su n eventi — • Esercizi —	
	4.4 Bayes	
	Formula di Boyes —	
Chapter 5	Calcolo Combinatorio _____	Page _____
	5.1 Metodo (Principio) delle Scelte Successive	
	5.2 Disposizioni	
	5.3 Combinazioni	
	5.4 Tipologie di esperimenti	
Chapter 6	Variabili Aleatorie _____	Page _____
	6.1 Introduzione	
	6.2 Variabili Aleatorie Costanti	
	6.3 Variabili Aleatorie Indicatrici (o di Bernulli)	
	6.4 Eventi associati alle variabili aleatorie	
	6.5 Distribuzione (o legge) di una Variabile Aleatoria	
	6.6 Funzione di Ripartizione	

- 6.7 Variabili Aleatorie Discrete
Caratterizzazione delle VA discrete —
- 6.8 Indici di sintesi di una distribuzione
Media o valore atteso — • Varianza —
- 6.9 Distribuzioni notevoli di VA discrete
Distribuzione uniforme discreta — • Distribuzione di Bernulli — • Distribuzione binomiale di parametri $p \in [0, 1]$ e $n \in \mathbb{N}$ — • Distribuzione di Poisson —
- 6.10 Variabili Aleatorie Continue
Definizioni di densità e VA continue — • Funzione di Ripartizione Continua — • Confronto tra VA continua e discreta — • Dalla funzione di ripartizione alla densità continua — • Funzioni di variabili aleatorie continue — • Distribuzioni notevoli di VA continue —

Chapter 7

Integrale di Gauss _____ Page _____

Chapter 8

Vettori Aleatori _____ Page _____

- 8.1 Introduzione
Legge di un vettore aleatorio — • Indipendenza di variabili aleatorie —
- 8.2 Vettori aleatori discreti
Densità discreta congiunta e densità discrete marginali — • Indipendenza e Densità discreta congiunta — • Valore atteso e varianza di una funzione di (X, Y) — • Indici di sintesi della distribuzione di un vettore aleatorio discreto —

Chapter 9

Catene di Markov _____ Page _____

Chapter 10

Esercitazione _____ Page _____

- 10.1 Probabilità Condizionata
Esercizio 4 (Foglio 2) — • Esercizio 7 — • Esercizio 5 — • Esercizio 2 — • Esercizio 3 — • Esercizio Porco Rosso — • Esercizio 2.3 — • —
- 10.2 Esercitazione combinatoria
Esercizio 3 — • Esercizio 4 — • Esercizio 6 — • Esercizio 8 —
- 10.3 Probabilità condizionata e indipendenza
Es. 1 —

Chapter 1

Introduzione

Chapter 2

Introduzione

Appunti di Probabilità presi in base alle lezioni di Elly Shlein

Chapter 3

Spazi di probabilità

3.1 Concetti introduttivi

Innanzitutto andiamo a definire che cosa intendiamo per *esperimento aleatorio*, *esito*, *probabilità*

Con la dicitura *esperimento aleatorio* indicheremo qualunque fenomeno (fisico, economico, sociale, ...) il cui esito non sia determinabile con certezza a priori. Il nostro obiettivo è di fornire una descrizione matematica di un esperimento aleatorio, definendo un modello probabilistico, un *esito* invece è un ipotetico risultato di un'esperimento aleatorio sulla base di un cosiddetto *spazio campionario* un insieme che contiene tutti gli esiti possibili dell'esperimento

Example 3.1.1

- **Esperimento aleatorio:** Lancio di un dado.
- **Spazio campionario:** $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- **Esito:** 4.

Note:

In casi più complessi ci saranno vari sotto-esperimenti aleatori, come 10 lanci di un dado.

Adesso forniamo vere e proprie definizioni

Definition 3.1.1: evento

Si definisce **evento** un'affermazione riguardante l'ipotetico esito univoco dell'esperimento, di cui si può affermare con certezza se è vero o falso una volta noto l'esito

Example 3.1.2

Esper. aleatorio: Lancio del dado
 $A = \text{"esce un numero pari"}$

Definition 3.1.2: Spazio campionario

Lo **spazio campionario** è l'insieme di tutti i possibili esiti di un esperimento casuale e viene denotato con Ω

Notare che non si afferma "tutti e solo tutti", quindi **qualsiasi** insieme che contiene gli esiti possibili può essere considerato uno spazio campionario

Example 3.1.3 (Lancio dado)

Possiamo porre come spazio campionario:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

ma anche

$$\Omega = \mathbb{R}$$

Definition 3.1.3: Esiti favorevoli

Esiti per cui un evento è vero sono detti esiti favorevoli.

Definition 3.1.4: Evento in termine di insiemi

Un evento si può definire anche come il sottoinsieme dello spazio campionario Ω formato da tutti gli esiti favorevoli dell'evento.

Example 3.1.4

$\Omega = \{1, 2, 3, 4, 5, 6\} \implies A = \text{"esce un numero pari"} \implies \{2, 4, 6\}$ sono gli esiti favorevoli dell'evento A .

Note:

La definizione insiemistica di un evento dipende dallo spazio campionario Ω definito, poiché l'evento è un sottoinsieme di Ω . Tuttavia, l'insieme degli esiti favorevoli di un evento è fisso, e rappresenta l'insieme evento di cardinalità massima possibile, ovvero l'insieme degli esiti favorevoli $A \subseteq \Omega$.

Definition 3.1.5

- Ω è l'evento *certo*
- \emptyset è l'evento *impossibile*
- $\omega \in \Omega$ è un evento *elementare* ($A = \{\omega\}$)

Example 3.1.5

Lancio un dado.

$A = \text{"esce un numero tra 1 e 6"}$

$B = \text{"esce un numero maggiore di 6"}$

$C = \text{"esce il numero 3"}$

- Se $\Omega = \{1, 2, 3, 4, 5, 6\}$, allora:
 - $A = \Omega$ (evento certo),
 - $B = \emptyset$ (evento impossibile),
 - $C = \{3\}$ (evento con un solo esito favorevole).
- Se $\Omega = \mathbb{R}$, allora:
 - $A = \{1, 2, 3, 4, 5, 6\} \subset \Omega$ (evento quasi certo),
 - $B = (6, +\infty)$ (evento quasi impossibile),
 - $C = \{3\}$ (evento con un solo esito favorevole).

3.2 Regole del calcolo probabilistico

Ad ogni relazione logica possiamo associare un'operazione insiemistica:

Connettivi Logici	Connettivi Insiemistici
$A \vee B$	$A \cup B$
$A \wedge B$	$A \cap B$
$\neg A$	A^c
$A \implies B$	$A \subseteq B$
$A \iff B$	$A = B$

Note:

Nella prima colonna, A e B sono eventi come affermazioni, mentre nella colonna di destra sono degli insiemi.

3.2.1 Assiomi della probabilit 

Poniamo tre assiomi fondamentali da cui possiamo partire per derivare tutte le operazioni e propriet  che ci servono:

Note:

Per noi tutti i sottoinsiemi di Ω sono eventi (anche se non sar  sempre cos )

Definition 3.2.1: Assiomi fondamentali della probabilit 

Assioma 1. A ciascun sottoinsieme (o evento) A di Ω   assegnato un numero $\mathbb{P}(A)$ che verifica:

$$0 \leq \mathbb{P}(A) \leq 1.$$

Tale numero $\mathbb{P}(A)$ si chiama **probabilit ** dell'evento A .

Assioma 2. $\mathbb{P}(\Omega) = 1$.

Assioma 3. Vale la propriet  di **additivit  numerabile**^a: sia $A_1, A_2, \dots, A_n, \dots$ una successione di sottoinsiemi di Ω tra loro disgiunti^b e sia

$$A = \bigcup_{n=1}^{\infty} A_n.$$

Allora

$$\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

^aAnche detta **σ -additivit **.

^bIn formule: $A_i \cap A_j = \emptyset$, per ogni $i \neq j$. In altri termini, non hanno elementi in comune.

Note:

Quindi, per il primo assioma, esiste una funzione probabilit  $\mathbb{P}(A) : \mathcal{P}(\Omega) \rightarrow [0, 1]$.

Definition 3.2.2: Spazio di probabilit 

La coppia (Ω, \mathbb{P}) si dice **spazio di probabilit ** o modello matematico dell'esperimento aleatorio

3.2.2 Conseguenze degli assiomi

Theorem 3.2.1

Sia Ω spazio campionario e \mathbb{P} probabilit  su Ω ((Ω, \mathbb{P})   uno spazio di probabilit  con $\mathbb{P} : \mathcal{P}(A) \rightarrow [0, 1]$). Dagli assiomi 1, 2, 3 deduciamo le cose seguenti:

1. $\mathbb{P}(\emptyset) = 0$

2. **Additività finita:** $(A_i)_{i=1,\dots,n}. \forall i \neq j. A_i \cap A_j = \emptyset \implies \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i)$
3. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
4. **Monotonia:** $A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$

Dimostrazione: 1. Devo mostrare che $\mathbb{P}(\emptyset) = 0$. Per semplicità definiamo $p := \mathbb{P}(\emptyset)$. Uso l'assioma 3 con la successione $(A_n)_{n \in \mathbb{N}}$ dove $\forall i \in \mathbb{N}. A_i = \emptyset$, che sono tutti eventi disgiunti. Quindi:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \sum_{i=1}^{\infty} p.$$

Inoltre:

$$\bigcup_{i=1}^{\infty} A_i = \emptyset \implies p = \sum_{i=1}^{\infty} p.$$

L'equazione è soddisfatta solo per $p = 0$.

2. Supponiamo di avere una sequenza finita disgiunta A_1, \dots, A_n . Definisco $(B_i)_{i \in \mathbb{N}}$ tale che $B_i = A_i$ per $i = 1, \dots, n$ e $B_i = \emptyset$ per $i > n$. Usando l'assioma 3:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \sum_{i=1}^n \mathbb{P}(A_i).$$

3. Per definizione di complemento, $A^c \cup A = \Omega$ e $A^c \cap A = \emptyset$. Per additività:

$$\mathbb{P}(A^c) + \mathbb{P}(A) = \mathbb{P}(\Omega) = 1 \quad (\text{per l'assioma 2}).$$

4. Se $A \subseteq B$, allora $B = A \cup (B \setminus A)$, con A e $B \setminus A$ disgiunti. Per additività:

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A).$$

Q

Theorem 3.2.2 Probabilità unione non disgiunta

Siano A e B eventi:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \quad (3.1)$$

Dimostrazione: $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$. Per additività:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B).$$

Osservando che:

$$\mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B) = \mathbb{P}(A), \quad \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B) = \mathbb{P}(B),$$

si ottiene la formula.

Q

Note:

La formula si complica con un numero di eventi maggiore di 2. Per $n = 3$:

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(B \cap C) - \mathbb{P}(A \cap C) + \mathbb{P}(A \cap B \cap C).$$

3.3 Probabilità discreta

Finora sappiamo solo le "regole" che deve seguire una funzione per essere una probabilità. Passiamo ora a vedere come calcolare il valore di un certo tipo di probabilità, la *probabilità discreta*:

Definition 3.3.1: Probabilità discreta

Chiamo probabilità discreta una funzione probabilità \mathbb{P} su Ω , tale che:

$$\exists \bar{\Omega} \subseteq \Omega, \bar{\Omega} \text{ e' finito o numerabile. } \mathbb{P}(\bar{\Omega}) = 1$$

Ovvero, una probabilità è discreta se il suo spazio campionario minimo è finito o numerabile. Questa condizione è necessaria per poter poi definire un modo per effettivamente calcolare il valore della probabilità (discreta) di un qualunque evento.

Diamo prima una definizione di una tale probabilità:

Definition 3.3.2: Delta di Dirac

Sia $\Omega = \mathbb{R}$, $x_0 \in \mathbb{R}$, allora si chiama delta di Dirac centrato in x_0 la funzione:

$$\delta_{x_0} : \mathcal{P}(\mathbb{R}) \rightarrow [0, 1]$$
$$A \mapsto \delta_{x_0}(A) = \begin{cases} 1 & x_0 \in A \\ 0 & x_0 \notin A \end{cases}$$

Notare che per definizione, la funzione di Dirac è una probabilità discreta, dato che soddisfa tutti gli assiomi per essere una probabilità e il suo spazio campionario minimo è formato da un solo elemento di Ω , quindi è discreta (ma non molto utile dato che può assumere solo due valori). Però, tramite le delta di Dirac, siamo in grado di costruire qualunque altra probabilità discreta:

Sia $\Omega = \mathbb{R}$. Prendiamo un numero contabile n di eventi singoletto $x_1, x_2, \dots, x_n \in \mathbb{R}$ a cui corrispondono $p_1, p_2, \dots, p_n \in \mathbb{R}$ tale che:

$$\forall i = 1, \dots, n. p_i \in [0, 1], \quad \sum_{i=1}^n p_i = 1$$

Definiamo la funzione:

$$\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$$
$$A \mapsto \sum_{i=1}^n p_i \delta_{x_i}(A)$$

\mathbb{P} è una combinazione lineare di delta di Dirac. Essendo una combinazione convessa, $\mathbb{P} \in [0, 1]$ e si può dimostrare che soddisfa gli altri due assiomi (2 e 3), quindi è una probabilità discreta! Variando le x e le p è possibile generare qualsiasi funzione \mathbb{P} discreta.

Example 3.3.1

$\Omega = \{1, 2, 3, 4, 5, 6\}$, $\forall i = 1, \dots, 6. x_i = i, p_i = \frac{1}{6}$, la funzione \mathbb{P} associata è:

$$P(A) = \sum_{i=1}^6 \frac{1}{6} \delta_{x_i}(A)$$

$$A = \{1, 2, 3, 4, 5, 6\} \implies P(A) = 1$$

$$B = (6, +\infty) \implies P(B) = 0$$

$$C = \{1, 2, 3, 4, 5, 6, 7, 8, 9\} \implies P(C) = 1$$

Definition 3.3.3

Si chiama evento quasi certo un evento A tale che $\mathbb{P}(A) = 1$.

Definition 3.3.4

Si chiama evento quasi impossibile un evento A tale che $\mathbb{P}(A) = 0$.

Posso allargare Ω quanto voglio perché tanto fuori dall'insieme minimo che comprende tutti gli eventi possibili le probabilità che aggiungo sono quasi impossibili e quindi hanno probabilità 0 e non cambiano il valore totale della somma.

3.3.1 Probabilità uniforme

Theorem 3.3.1 Principio di probabilità uniforme

Si consideri un esperimento aleatorio con spazio campionario $\Omega = \{w_1, \dots, w_N\}$ finito e discreto, con esiti sono equiprobabili $\mathbb{P}(\{w_1\}) = \mathbb{P}(\{w_2\}) = \dots = \mathbb{P}(\{w_N\})$.

Si dice allora che \mathbb{P} è la **probabilità uniforme** su Ω e valgono le seguenti proprietà:

1. Dato un qualunque evento elementare $A = \{w_i\}$, si ha:

$$\mathbb{P}(A) = \frac{1}{N}$$

2. Dato un qualunque evento $A \subseteq \Omega$, vale la *formula di Laplace*:

$$\mathbb{P}(A) = \frac{|A|}{N} = \frac{\text{numero di esiti favorevoli}}{\text{numero di esiti possibili}}$$

Dimostrazione: 1. Dimostro il punto 1:

Per ipotesi sappiamo che $\mathbb{P}(\{w_1\}) = \mathbb{P}(\{w_2\}) = \dots = \mathbb{P}(\{w_N\})$ e per il *principio di additività* si può costruire il seguente sistema:

$$\begin{cases} \mathbb{P}(\{w_1\}) + \mathbb{P}(\{w_2\}) + \dots + \mathbb{P}(\{w_N\}) = 1 \\ \mathbb{P}(\{w_1\}) = \mathbb{P}(\{w_2\}) \\ \mathbb{P}(\{w_2\}) = \mathbb{P}(\{w_3\}) \\ \vdots \\ \mathbb{P}(\{w_{N-1}\}) = \mathbb{P}(\{w_N\}) \end{cases}$$

Da cui si ricava che:

$$\forall i \in [1, \dots, N] \quad \mathbb{P}(\{w_i\}) = \frac{1}{N}$$

2. Dimostro il punto 2: Sia $A = \{w_{i_1}, \dots, w_{i_k}\}$, con $k \leq N$. Per definizione di probabilità:

$$\mathbb{P}(A) = \mathbb{P}(\{w_{i_1}, \dots, w_{i_k}\}) = \mathbb{P}(\{w_{i_1}\}) + \dots + \mathbb{P}(\{w_{i_k}\}) = \frac{k}{N}$$



Chapter 4

Probabilità Condizionata

4.1 Definizione e motivazioni

Supponiamo di sapere che un evento di un'esperimento aleatorio si è avverato. Finora abbiamo visto solo casi in cui gli eventi non si influenzavano (*indipendenti*), ma succede spesso nella realtà che se si sa che un certo evento è avvenuto, allora questo ci dà informazioni aggiuntive che possono cambiare la probabilità di altri eventi di cui ancora non sappiamo gli esiti.

Chiamiamo B l'evento che è avvenuto e A un altro evento di cui vogliamo sapere la probabilità. Prima di avere informazioni su B , la probabilità di A era semplicemente $\mathbb{P}(A)$, ma ora ci poniamo la domanda: "se so che si è verificato B , come cambia $\mathbb{P}(A)$?". Denotiamo questa nuova probabilità con:

$$P(A|B)$$

chiamata la *probabilità condizionata di A dato B* .

Definition 4.1.1: Probabilità Condizionata

Prendo due eventi A, B su uno spazio di probabilità (Ω, \mathbb{P}) . Definisco *probabilità condizionata a B di A* la funzione:

$$\begin{aligned}\mathbb{P}(A|B) : \mathcal{P}(\Omega) &\rightarrow [0, 1] \\ A &\mapsto \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\end{aligned}$$

È possibile dimostrare che una certa funzione è anch'essa una probabilità (sempre discreta), verifichiamo gli assiomi (fissiamo $B \subseteq \Omega$ con $\mathbb{P}(B) > 0$):

1. $\mathbb{P}(A|B) \in [0, 1], \forall A \subseteq \Omega$
2. $\mathbb{P}(\Omega|B) = 1$
3. σ - addittività: $(A_i)_{i \in \mathbb{N}}$ disgiunti:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i | B\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i | B)$$

Lasciate al lettore in quanto davvero molto facili, quasi banali. Se non riesci a farle fai schifo. Vediamo ora, con un esempio, come mai è proprio questa la definizione utilizzata:

Example 4.1.1

- Lancio del dado: $\Omega = \{1, 2, 3, 4, 5, 6\}$, \mathbb{P} probabilità uniforme:

$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}, \forall \omega \in \Omega$, ovvero:

$$\mathbb{P}(A) = \frac{\text{casi favorevoli in } A}{\text{casi possibili}}$$

$A = \text{"esce un numero maggiore di 3"} = \{3, 4, 5, 6\}$ e $B = \text{"esce un numero pari"} = \{2, 4, 6\}$, domanda: quanto vale $\mathbb{P}(A|B)$?

$P(A) = \frac{4}{6}$ come abbiamo già visto.

Ora abbiamo un'informazione in più: sappiamo che B si è avverato. Questo significa che si restringe l'insieme di valori che possono essere usciti al lancio del dado. ATTENZIONE! ciò non vuol dire che cambia lo spazio campionario perché l'esperimento è lo stesso, ma cambiano i *veri* casi favorevoli e i *veri* casi possibili:

$$P(A|B) = \frac{\text{"veri casi favorevoli di A"}}{\text{veri casi possibili}} = \frac{|A \cap B|}{|B|} = \frac{2}{3}$$

- Vediamo anche cosa accade quando la probabilità non è uniforme, come con un dado a 4 facce truccato:

$$\Omega = \{1, 2, 3, 4\}, \mathbb{P}(4) = \frac{1}{15}, \mathbb{P}(3) = \frac{2}{15}, \mathbb{P}(2) = \frac{4}{15}, \mathbb{P}(1) = \frac{8}{15}$$

$$A = \{3, 4\}, B = \{2, 4\}$$

$$\mathbb{P}(A|B) = \frac{\text{"probabilità dei veri casi favorevoli di A"}}{\text{probabilità dei veri casi possibili}} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Note:

Se \mathbb{P} è la probabilità uniforme allora:

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

Note:

B è fissato nella definizione di probabilità condizionata, ovvero:

$$\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$$

Quindi il ruolo di A e B è completamente diverso

Note:

Se $B = \Omega$, allora $\mathbb{P}(A|B) = \mathbb{P}(A)$ dato che la conoscenza del fatto che si è avverato Ω è ovvio e non ci cambia.
Se $A = \Omega$, allora $\mathbb{P}(\Omega|B) = 1$ (per proprietà, dato che è sempre una probabilità)

4.2 Regola della catena

La probabilità condizionata in genere è nota e si usa per calcolare la probabilità dell'intersezione:

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

Questa formula è detta *regola della catena* e vale in generale con n eventi:

Proposition 4.2.1 Regola della catena (generalizzata)

$(A_i)_{i=1, \dots, n}$, $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0$, allora:

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2|A_1) \mathbb{P}(A_3|A_1 \cap A_2) \dots \mathbb{P}(A_n|A_1 \cap \dots \cap A_{n-1})$$

Note:

La condizione funziona grazie alla monotonia, dato che $0 < \mathbb{P}(A_1 \cap \dots \cap A_{n-1}) \leq \mathbb{P}(A_1 \cap \dots \cap A_j)$, $1 \leq j \leq n-1$ quindi siamo certi che l'intersezione degli eventi che sono avvenuti è maggiore di 0.

$$: P(A_1) \frac{P(A_1 \cap A_2)}{P(A_1)} \dots \frac{P(A_1 \cap \dots \cap A_n)}{P(A_1 \cap \dots \cap A_{n-1})} = P(A_1 \cap \dots \cap A_n)$$



TODO: migliora un po

Example 4.2.1

Un'urna contiene tre palline bianche, due palline nere e una pallina rossa. Si eseguono tre estrazioni senza reimmissione. Qual è la probabilità di estrarre nell'ordine una bianca, una rossa e una nera? Sono interessato solo ad alcuni eventi, quindi non c'è bisogno di descrivere l'intero esperimento aleatorio. Per prima cosa definisco l'evento:

A = "estrarre in ordine una bianca, una rossa e una nera"

Voglio trovare $P(A)$. Notiamo che dobbiamo determinare tre sottoesperimenti in relazione (dato che non c'è reimmissione). Quindi dopo ogni sottoesperimento cambia la composizione dell'urna, e sappiamo come calcolare la probabilità condizionata:

B_i = "estraggo una pallina bianca all' i -esimo turno"

R_i = "estraggo una pallina rossa all' i -esimo turno"

N_i = "estraggo una pallina nera all' i -esimo turno"

Esistono tre famiglie di eventi: $(B_i)_{i=1,\dots,k}$, $(R_i)_{i=1,\dots,k}$, $(N_i)_{i=1,\dots,k}$ dove i indica il turno al quale viene estratta la pallina. Quindi possiamo scrivere A come relazione fra sottoeventi:

$$A = B_1 \cap R_2 \cap N_3$$

Quindi:

$$P(A) = P(B_1 \cap R_2 \cap N_3) = P(B_1)P(R_2|B_1)P(N_3|B_1 \cap R_2)$$

Solo ora possiamo passare ai valori numerici. Dato che gli esiti sono equiprobabili e lo spazio campionario è finito, la probabilità è uniforme:

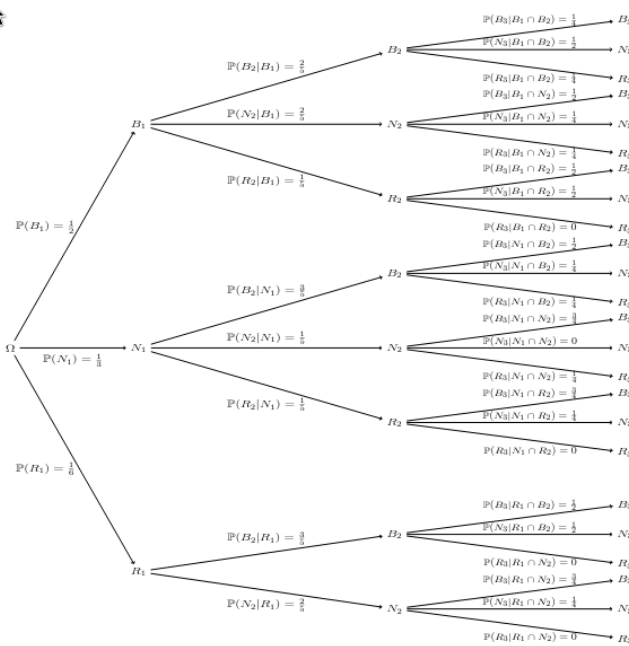
$$P(B_1) = \frac{1}{2}, P(R_2) = \frac{1}{5}, P(N_3) = \frac{1}{2}$$

$$P(A) = \frac{1}{20}$$

Per esperimenti formati da sottoesperimenti di cui conosco le probabilità condizionate, è possibile rappresentare ogni evento come un nodo:

Ω = primo nodo

e ogni probabilità come un ramo che partiziona il nodo (tanti rami quanti gli insiemi della partizione) che rappresenta poi un altro evento (condizionato dalla seconda in poi).



La regola della catena la leggo sul diagramma ad albero:

percorso: $\Omega \rightarrow B_1 \rightarrow R_2 \rightarrow N_3$ ha probabilita' $P(B_1 \cap R_2 \cap N_3)$ che si calcola facendo il prodotto delle probabilita' dei relativi rami che si usano nel percorso.

E' uno strumento utile per convincerci che stiamo usando le formule giuste, ma non le sostituisce e puo' diventare laborioso per problemi complessi.

Example 4.2.2

Ci sono due urne: la prima contiene due palline rosse e una bianca; la seconda contiene tre palline rosse e due bianche. Si lancia una moneta: se esce testa si estrae una pallina dalla prima urna, se esce croce si estrae una pallina dalla seconda urna. Qual e' la probabilita' che l'esito del lancio della moneta sia testa e la pallina estratta sia bianca?

2 sottoesperimenti:

- lancio della moneta
- estrazione da un'urna

Nota che i sottoesperimenti sono indipendenti dall'esito di altri esperimenti. Sono gli esiti, ovvero i risultati, che possono dipendere dagli esiti di altri esperimenti.

A = "esce testa ed estraggo una pallina bianca"

Devo esprimere A con eventi che

T = "esce testa"

U = estraggo una pallina bianca

Disegna lo zio pera di diagramma che non mi metto a fare, se @GiovanniPalma vuole puo' farlo

$$A = T \cap U, \quad P(A) = P(T)P(U|T)$$

4.3 Indipendenza di eventi

E' possibile che sapere che un evento B e' avvenuto non altera la probabilita' di un altro evento A . Possiamo esprimere questa relazione in modo matematico cosi':

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

Utilizzando la definizione di probabilit  condizionale, possiamo usare un'identit  equivalente che useremo come definizione:

Definition 4.3.1: Eventi indipendenti

Due eventi A, B si dicono indipendenti se:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) \quad (4.1)$$

E viene denotato $A \perp B$

Usiamo questa definizione dato che   esplicitamente simmetrica, ovvero se A   indipendente a B allora vale anche il contrario:

$$A \perp B \iff B \perp A$$

ed   definita (e banalmente vera) anche quando $\mathbb{P}(A) = 0$ o $\mathbb{P}(B) = 0$. In particolare si noti il seguente teorema:

Theorem 4.3.1 Teorema della simmetria tra eventi indipendenti

Sia $\mathbb{P}(B) > 0$ allora:

$$A \perp B \iff \mathbb{P}(A|B) = \mathbb{P}(A)$$

Dall'altro lato, sia $\mathbb{P}(A) > 0$ allora:

$$A \perp B \iff \mathbb{P}(B|A) = \mathbb{P}(B)$$

Dimostrazione: Verr  fornita solo la dimostrazione del primo punto, la seconda parte   analoga. Assumo $\mathbb{P}(B) > 0$, si ha:

- $A \perp B \implies \mathbb{P}(A|B) = \mathbb{P}(A)$:

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

- $\mathbb{P}(A|B) = \mathbb{P}(A) \implies A \perp B$:

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



Note:

Si noti che se $\mathbb{P}(A) > 0$ e $\mathbb{P}(B) > 0$ allora, le tre uguaglianze seguenti sono equivalenti:

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

Note:

- L'indipendenza   diversa dalla disgiunzione:

$$A \perp B \neq A \cap B = \emptyset$$

infatti sono relazioni ortogonali:

$$A \perp B \wedge A \cap B = \emptyset \iff \mathbb{P}(A)\mathbb{P}(B) = 0 \iff \mathbb{P}(A) = 0 \vee \mathbb{P}(B) = 0$$

- L'indipendenza   diverso dall'essere sottoinsieme non-vuoto:

$$A \perp B \neq A \subseteq B \vee B \subseteq A$$

infatti:

$$A \perp B \wedge A \subseteq B \iff \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A) \iff \mathbb{P}(B) = 1$$

Quindi in generale due eventi sono indipendenti quando la loro intersezione ha "le giuste proporzioni".

Adesso fornirò un altro teorema piuttosto importante:

Proposition 4.3.1 Sull'indipendenza di eventi complementari

Siano A, B due eventi indipendenti, allora:

$$A \perp B \iff A^c \perp B, A \perp B^c, A^c \perp B^c$$

Dimostrazione: Dimostro solo la prima parte, le altre sono analoghe.
Assumo A, B due eventi indipendenti, debbo dimostrare la seguente uguaglianza:

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(A^c) \cdot \mathbb{P}(B)$$

Dato che

$$B = \Omega \cap B = (A \cup A^c) \cap B = (A \cap B) \cup (A^c \cap B)$$

E dato che $(A \cap B)$ e $(A^c \cap B)$ sono disgiunti, per 2 (*additività finita*) si ha:

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$$

E dato che $A \perp B$ si ha:

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) = \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) + \mathbb{P}(A^c \cap B)$$

Quindi:

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(A) \cdot \mathbb{P}(B) = \mathbb{P}(B) \cdot (1 - \mathbb{P}(A)) = \mathbb{P}(A^c) \cdot \mathbb{P}(B)$$



4.3.1 Generalizzazione su n eventi

Come nel caso con solo due eventi, $n > 2$ eventi si dicono indipendenti quando, sapendo che qualsiasi numero degli altri eventi si è avverato, la probabilità dell'evento non cambia. Questo deve valere per tutti gli n eventi, ovvero:

$$\mathbb{P}\left(A_i \left| \bigcup_{j=1, j \neq i}^n A_j\right.\right) = \mathbb{P}(A_i) \quad \forall i = 1, \dots, n$$

Si può dimostrare, usando la definizione di probabilità condizionata, che questa identità equivale a dire:

Definition 4.3.2: Eventi indipendenti (per n eventi)

Sia $(A_i)_{i \in I}$ una famiglia di eventi in uno spazio di probabilità. Si dice che questi eventi sono indipendenti quando **per ogni sottoinsieme** finito $J \subseteq I$, $|J| > 2$:

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j)$$

4.3.2 Esercizi

Example 4.3.1 (Calcolo di eventi indipendenti con probabilità condizionata)

TESTO:

Si lancia un dado a 6 facce

A = "Esce un numero > 4 "

B = "Esce un numero pari"

Determinare $P(A)$ e $P(A|B)$

DETTAGLIO SVOLGIMENTO:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$A = \{5, 6\}$$

$$B = \{2, 4, 6\}$$

$$P(A) = \frac{3}{6} = \frac{1}{2}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{3}$$

Dato che $P(A) = P(A|B)$, per il teorema 4.3 si ha che $A \perp B$

Ecco un altro esercizio:

Example 4.3.2

TESTO

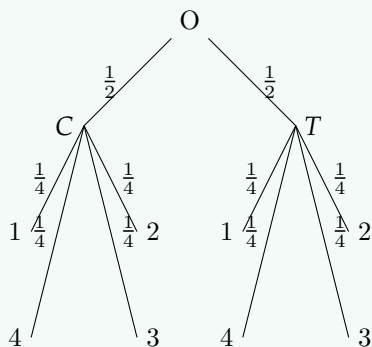
Lanciamo una moneta e un dado a 4 facce.

Determinare uno spazio di prob. che descriva l'esperimento aleatorio

SOLUZIONE

$$\Omega = \{(T, 1), (T, 2), (T, 3), (T, 4), (C, 1), (C, 2), (C, 3), (C, 4)\}$$

$$P(T, 1) = \frac{1}{8} = P(C, 4) = \frac{1}{8}$$



T = "esito del lancio moneta e testa"

C = "esito del lancio moneta è croce"

D_i = "è uscito il numero i "

$$P(C) = \frac{1}{2}, \quad P(T) = \frac{1}{2}$$

$$P(D_i|C) = \frac{1}{4}$$

$$P(D_i|T) = \frac{1}{4}$$

A = "è uscito testa e il numero i "

$$P(A) = P(T) \cdot P(D_i|T) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$$

Analogamente per $C \cap D_i$

4.4 Bayes

Ci sono delle situazioni nelle quali può esserci richiesto di calcolare la probabilità condizionata "inversa" conoscendo quella diretta. Possiamo fare ciò usando la definizione di probabilità condizionate

Theorem 4.4.1

Siano A, B due eventi t.c. $P(A), P(B) > 0$:

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

Example 4.4.1 (4.1)

Ci sono due urne: la prima urna contiene una pallina bianca e due palline rosse, mentre la seconda contiene due palline bianche e cinque palline rosse. Si lancia una moneta: se esce testa si estrae una pallina dalla prima urna, se esce croce si estrae una pallina dalla seconda. Sapendo che 'e stata estratta una pallina bianca, calcolare la probabilit'a che l'esito del lancio della moneta sia stato testa.

EVENTI

T = esce testa

B = estrazione pallina bianca

$\Omega = \{t, c\} \times \{b, r\}$

$$\mathbb{P}(T|B) = \frac{P(B|T)P(T)}{P(B)} = \frac{P(B|T)P(T)}{P(B|T)P(T) + P(B|T^c)P(T^c)} = \frac{1/3 \cdot 1/2}{1/3 \cdot 1/2 + 2/7 \cdot 1/2} \text{ probabilita' totali e bayes}$$

Example 4.4.2

Esempio 5.1. In un'urna ci sono due palline che possono essere rosse (R) o bianche (B). La composizione esatta non 'e nota, quindi le composizioni possibili sono: RR, RB, BB. (H_0, H_1, H_2) Supponiamo che, in base alle informazioni a disposizione, sia ragionevole assegnare uguale probabilit'a pari a $1/3$ alle tre composizioni (ipotesi) possibili, che denotiamo H_0, H_1 e H_2 . 1) Se si estrae una pallina dall'urna, qual 'e la probabilit'a che sia bianca? 2) Si effettuano tre estrazioni con reimmissione: sapendo che le prime due palline estratte sono bianche, qual 'e la probabilit'a che anche la terza pallina estratta sia bianca?

Il paradosso viene dal fatto che dobbiamo tenere traccia dell'asimmetria iniziale e denotare le diverse composizioni come ipotesi possibili.

EVENTI

B_i = estraggo una pallina bianca all'i-esima estrazione

1) $P(B_1) = P(B_1|H_0)P(H_0) + P(B_1|H_1)P(H_1) + P(B_1|H_2)P(H_2) = 1/2$ probabilita' totali e prob. uniforme delle H

Ce l'aspettiamo dato che c'e' completa simmetria fra i rami delle H con la distribuzione di palline

2) $P(B_3|B_1 \cap B_2) = P(B_3 \cap B_2 \cap B_1)/P(B_1 \cap B_2) =$ usare ancora il partizionamento per H per applicare probabilita' totali

4.4.1 Formula di Bayes

Quando si deve calcolare una probabilita' condizionata di eventi "nell'ordine temporale sbagliato", ad esempio se si calcolare $P(A|B)$ ma si conosce solo $P(B|A)$, si puo' utilizzare uno strumento utilissimo facilmente derivabile dalla matematica della probabilita' condizionata, ovvero la formula di Bayes:

Theorem 4.4.2 Formula di Bayes

Siano A e B due eventi tali che $\mathbb{P}(A) > 0$ e $\mathbb{P}(B) > 0$, allora:

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

Dimostrazione: Per definizione di $\mathbb{P}(A|B)$ si ha:

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

Utilizzando la regola della catena possiamo rimuovere il numeratore come segue:

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

Quindi:

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



Example 4.4.3

Ci sono due urne: la prima urna contiene una pallina bianca e due palline rosse, mentre la seconda contiene due palline bianche e cinque palline rosse. Si lancia una moneta: se esce testa si estrae una pallina dalla prima urna, se esce croce si estrae una pallina dalla seconda. Sapendo che è stata estratta una pallina bianca, calcolare la probabilità che l'esito del lancio della moneta sia stato testa.

Svolgimento:

T = "esce testa"

Chapter 5

Calcolo Combinatorio

Quando abbiamo numero finito di elementi con probabilita' uniforme, e' possibile utilizzare l'addittivita' delle probabilita' per ridurre il problema di probabilita' a un problema di conteggio (la probabilita' di un evento e' data dal numero di casi favorevoli) \implies dobbiamo imparare a contare.

Dato un certo insieme, dobbiamo calcolarne la cardinalita' utilizzando i giusti strumenti matematici:

Theorem 5.0.1 Calcolo probabilita' uniforme (Laplace)

Se Ω e' finito, $\Omega = \{w_1, \dots, w_N\}$ e $\forall i = 1, \dots, N. \mathbb{P}(\{w_i\}) = \frac{1}{N}$ (probabilita' uniforme), allora $\forall A \subseteq \Omega$ (evento):

$$\mathbb{P}(A) = \frac{|A|}{N}$$

Dobbiamo introdurre:

- **Fattoriali:**

- $0! := 1$
- $(n+1)! := (n+1) \cdot n!$

- **Coefficienti Binomiali:**

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \forall n, k \in \mathbb{N}. n \geq k$$

In generale:

$$\binom{n}{k} = \binom{n}{n-k}$$

Dal triangolo di Tartaglia, possiamo visualizzare altre proprieta' (oltre alla simmetria)

Vedremo che $\binom{n}{k}$ sono il numero di modi diversi che abbiamo per selezionare k sottoinsiemi diversi da un insieme di cardinalita' n .

Theorem 5.0.2 Formula di Newton

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

Quindi anche il coefficiente deriva da un problema di conteggio.

5.1 Metodo (Principio) delle Scelte Successive

E' un algoritmo per determinare la cardinalita' di un insieme. Vediamo un esempio:

Example 5.1.1

Alfabeto di 36 caratteri dove ognuno dei numeri corrisponde a un carattere alfanumerico.

Domanda: "Quante possibili password distinte di 8 caratteri esistono in questo alfabeto?"

$\Omega = \text{Alfanumerici} \times \dots \times \text{Alfanumerici}$ 8 volte.

- Scelte:

1. un carattere dei 36 totali

⋮

2. (e' cosi 8 volte)

Quindi $|\Omega| = 36 \times \dots \times 36 = 36^8$

E se vogliamo evitare di ripetere i caratteri? Vediamo le scelte:

1. un carattere dei 36 totali

2. un carattere dei **35 possibili**

⋮

3.

4. un carattere dei 29 possibili

Quindi $|\Omega| = 36 \times 35 \times \dots \times 29 = \frac{36!}{28!}$

Definiamo il principio generale:

Theorem 5.1.1 Non proprio un teorema

Ciascun elemetno di un insieme A puo' essere determinato tramite sola sequenza di k scelte, dove per ogni scelta ci sono n_1, \dots, n_k possibilità, allora:

$$|A| = n_1 \times \dots \times n_k$$

Note:

Puo' essere riscritto come teorema formale ma **E' TROPPO DIFFICILE PER NOI INFORMATICI** quindi non lo facciamo!!!!

Example 5.1.2

52 carte (13 tipi 4 semi)

1. $A =$ iniseme dei full (un tris e una coppia), $|A| = ?$

4 scelte:

- tipo di carta nel tris (13)
- tipo di carta nella coppia (12)
- semi nel tris (4)
- semi nella coppia (6)

$|A| = 131246 =$ casi favorevoli

2. $A =$ doppie coppie (due coppie di tipo diverso e una carta libera), $|A| = ?$

Scelte:

- tipo nella prima coppia (13)

- tipo nella seconda coppia (12)
- semi nella prima coppia (6)
- semi seconda (6)
- tipo singolo (11)
- seme singolo (4)

$|A| = 131266114$ **SBAGLIATO!!**

Perche' non ci interessa dell'ordine delle due coppie (bisogna dedurlo dalla definizione di A), quindi NON c'e' una prima e seconda coppia (anche sopra era sbagliato vederlo cosi') dato che non c'e' l'ordine.

Combinazioni dei tipi che compongono 2 coppie ($13 \times 12/2$)

5.2 Disposizioni

Definition 5.2.1: Disposizioni con ripetizione

Dato un insieme E con n elementi, indichiamo con $DR_{n,k}$ le sequenze ordinate di k elementi di E .

Sostanzialmente, $DR_{n,k} = E \times E \dots \times E$ k volte, ovvero:

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

Quindi usando il principio delle scelte successive:

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

dato che per ogni E abbiamo una scelta fra n elementi.

Note:

Indicando tale insieme con $DR_{n,k}$, l'insieme E sparisce, dato che ci interessa solo la **cardinalita'** di tale insieme e non ce ne frega un cavolo dei suoi elementi.

Example 5.2.1 (Iniziamo a calcolare le probabilita'!)

Presa un'urna con n palline numerate ($E = \{1, \dots, n\}$) e si estraggono k palline con *reinbussolamento*.

$$\Omega = E \times \dots \times E = DR_{n,k}$$

Quindi $|\Omega| = n^k$, e la probabilita' uniforme e' data da:

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|} = n^{-k}$$

Example 5.2.2

Determinare spazi campionari per i seguenti esperimenti:

- Scelta casuale di una parola di 8 lettere
- Scelta di colonne del totocalcio (risultato di 13 partite)

Quindi lo usiamo nei casi di estrazione con reinbussolamento quando ci interessa l'ordine.

Definition 5.2.2: Disposizioni

Dato un insieme E di n elementi, l'insieme delle disposizioni (senza ripetizione) di k elementi dell'insieme E e' l'insieme delle sequenze ordinate di k elementi *distinti*, ovvero:

$$D_{n,k} := \{(x_1, \dots, x_k) | x_1, \dots, x_k \in E \wedge x_i \neq x_j \text{ se } i \neq j\}$$

Note:

$D_{n,k}$ e' un sottoinsieme **stretto** di $DR_{n,k}$.

Usando le scelte successive:

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

L'esperimento aleatorio di riferimento e' l'estrazione **senze** reimmisione.

Definition 5.2.3: Permutazioni

$P_n = D_{n,n}$, quindi: $|P_n| = n!$

5.3 Combinazioni

Definition 5.3.1: Combinazioni

Dato un insieme E di n elementi, indichiamo con $C_{n,k}$ la classe dei sottoinsiemi di E contenenti k elementi, ovvero:

$$C_{n,k} = \{A \subseteq E \mid |A| = k\}$$

Siamo passati da sequenze a sottoinsiemi, ovvero non ci interessa piu' dell'ordine:

sottoinsieme = sequenza non ordinata

Proposition 5.3.1 sulla combinazione come coefficiente binomiale

Sia la combinazione $C_{n,k}$, allora vale:

$$|C_{n,k}| = \binom{n}{k} \quad (5.1)$$

Dimostrazione: Mi riduco a dimostrare che:

$$\frac{n!}{(n-k)!} = |C_{n,k}| k!$$

Ovvero

$$|D_{n,k}| = |C_{n,k}| \cdot |P_k|$$

Ma $|D_{n,k}| = |C_{n,k}| \cdot |P_k|$ vale per il principio delle scelte successive, infatti:

- scelta per i k elementi selezionati: $|C_{n,k}|$
- scelta dell'ordinamento $|P_k|$



Si noti questo esempio

Example 5.3.1

Consideriamo un'urna con n palline distinte delle quali ne vengono estratte k . Determinare lo spazio di probabilità

Si ha che: $\Omega = C_{n,k}$

Mentre per il principio di probabilità uniforme si ha:

$$\mathbb{P}(\{w\}) = \frac{1}{\binom{n}{k}} = \frac{k!(n-k)!}{n!}$$

Interpretazione combinatoria della formula di Stifel

Note:

Si consideri la formula di Stifel:

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \quad (5.2)$$

Si può interpretare questa formula in termini di combinazioni come segue:

$$\underbrace{\binom{n}{k}}_{\text{n° di combinazioni di } k \text{ el-}} = \underbrace{\binom{n-1}{k-1}}_{\text{menti di } E \text{ n° di combinazioni di } k \text{ el-}} + \underbrace{\binom{n-1}{k}}_{\text{menti di } E \text{ in cui è pre-}} + \underbrace{\binom{n-1}{k}}_{\text{sente l'elemento } \bar{e} \text{ n° di combinazioni di } k \text{ elementi di } E \text{ senza l'elemento } \bar{e}}$$

dove \bar{e} è un elemento fissato

Interpretazione combinatoria del binomio di Newton

Note:

Si consideri la formula del binomio di Newton:

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k} \quad a, b \in \mathbb{R} \quad (5.3)$$

Il prodotto $(a+b)(a+b)\dots(a+b)$ di n fattori si sviluppa in una somma di monomi di grado n del tipo $a^{n-k}b^k$ dove k varia da 0 a n . Dobbiamo quindi determinare il coefficiente di ogni monomio $a^{n-k}b^k$, ossia calcolare quante volte compare facendo il prodotto $(a+b)(a+b)\dots(a+b)$ ossia calcolare quante volte compare facendo il prodotto $(a+b)(a+b)\dots(a+b)$. Questo monomio però si ottiene scegliendo il valore b da k degli n fattori e a dai rimanenti $n-k$, ovvero in $\binom{n}{k}$ modi

5.4 Tipologie di esperimenti

Riassumendo, si possono verificare 3 casi notevoli durante un esperimento aleatorio inseriti in questa tabella

Si tiene conto dell'ordine	Estrazione senza reimmissione $\Omega = D_{n,k}$ $ \Omega = \frac{n!}{(n-k)!}$	Estrazione con reimmissione $\Omega = DR_{n,k}$ $ \Omega = n^k$
Non si tiene conto dell'ordine	Estrazione simultanea $\Omega = C_{n,k}$ $ \Omega = \frac{ D_{n,k} }{k!} = \binom{n}{k}$	—

Table 5.1: Gli spazi campionari e le loro cardinalità.

Note:

La casella vuota nella tabella sopra riportata corrisponde all'insieme delle cosiddette *combinazioni con ripetizione*, un raggruppamento di k elementi estratti da n elementi distinti, nel quale uno stesso elemento può

ripetersi fino a k volte, e in cui l'ordine di estrazione non è rilevante

Si noti questo esempio:

Example 5.4.1 (Esempio sulla casella mancante)

si lancia tre volte una moneta. Qual è la probabilità che esca due volte testa?

Si hanno due approcci per scegliere lo spazio campionario:

- utilizzare le $DR_{2,3}$ con \mathbf{P} uniforme.

In questo caso:

$$\Omega = DR_{2,3} = \{(T, T, T), (T, T, C), (T, C, T), (C, T, T), (T, C, C), (C, T, C), (C, C, T), (C, C, C)\}$$

Ci sono 8 possibili sequenze perché ogni lancio ha 2 possibili risultati:

$$2^3 = 8$$

Siccome ogni sequenza ha la stessa probabilità di verificarsi, possiamo assegnare la probabilità uniforme, si ha

$$\mathbf{P}(x_1, x_2, x_3) = \frac{1}{8} \quad \forall x$$

Sarà quindi facile calcolare la probabilità di A (lascio al basta come esercizio)

- Lo spazio campionario con combinazioni con ripetizione $CR_{n,k}$.

In questo caso, dato che non ci interessa l'ordine dei lanci, dobbiamo raggruppare le sequenze "uguali":

$$CR_{2,3} = \{[T, T, T], [T, T, C], [T, C, C], [C, C, C]\}$$

Tuttavia è facile verificare che in questo caso la probabilità uniforme **non sussiste**, perché alcuni gruppi contengono più sequenze di altri

infatti:

- $\mathbf{P}([T, T, T]) = \mathbf{P}((T, T, T)) = \frac{1}{8}$
- $\mathbf{P}([T, T, C]) = \mathbf{P}((T, T, C)) + \mathbf{P}((T, C, T)) + \mathbf{P}((C, T, T)) = \frac{3}{8}$
- $\mathbf{P}([T, C, C]) = \mathbf{P}((T, C, C)) + \mathbf{P}((C, T, C)) + \mathbf{P}((C, C, T)) = \frac{3}{8}$
- $\mathbf{P}([C, C, C]) = \mathbf{P}((C, C, C)) = \frac{1}{8}$

Note:

Se scegliamo lo spazio campionario $CR_{n,k}$, dobbiamo lavorare con probabilità non uniformi, il che complica i calcoli. Invece, se lavoriamo con $DR_{n,k}$, tutte le sequenze hanno la stessa probabilità e possiamo usare il calcolo combinatorio standard per trovare i risultati.

Per questo motivo, nella maggior parte dei problemi in cui c'è ripetizione, conviene lavorare con disposizioni con ripetizione $DR_{n,k}$, perché permette di usare la probabilità uniforme.

Bho adesso si fa un altro esempio, e va benissimo così

Example 5.4.2 (Probabilità binomiale)

Si consideri un'urna che contiene b palline bianche ed r palline rosse. Si effettuano k estrazioni con reimmissione. Calcolare la probabilità dell'evento

$A_\ell = \text{"si estraggono } \ell \text{ palline bianche ed } k - \ell \text{ palline rosse"}$

soluzione: Etichettiamo le b palline bianche con b_1, b_2, \dots, b_b ; analogamente, le r palline rosse con r_1, \dots, r_r . Sia dunque:

$$E = \{b_1, \dots, b_b, r_1, \dots, r_r\}$$

E si noti che $|E| = b + r$.

Come spazio di probabilità (Ω, \mathbb{P}) è naturale considerare $\Omega = DR_{b+r, k}$ (insieme delle disposizioni con ripetizione di k elementi di E) e \mathbb{P} probabilità uniforme, quindi $\mathbb{P}(\{w\}) = \frac{1}{(b+r)^k}$

Per calcolare $\mathbb{P}(A_\ell)$ dobbiamo calcolare $|A_\ell|$ che si rivela possibile usando il principio delle scelte successive:

- scelta della sequenza di ℓ palline bianche estratte, ovvero $|DR_{b, \ell}|$ possibilità
- scelta della sequenza di $k - \ell$ palline rosse estratte, ovvero $|DR_{r, k-\ell}|$ possibilità
- scelta delle ℓ estrazioni in cui sono uscite le palline bianche: $|C_{k, \ell}|$ possibilità

In definitiva:

$$\mathbb{P}(A_\ell) = \frac{|DR_{b, \ell}| |DR_{r, k-\ell}| |C_{k, \ell}|}{|DR_{b+r, k}|} = \binom{k}{\ell} \frac{b^\ell r^{k-\ell}}{(b+r)^k}$$

O, equivalentemente, dato $p = \frac{b}{b+r}$ si ha:

$$\mathbb{P}(A_\ell) = \binom{k}{\ell} p^\ell (1-p)^{k-\ell}, \quad \ell = 0, \dots, k$$



Note:

Consideriamo lo spazio di probabilità (Ω, \mathbb{P}) con $\Omega = \{0, 1, 2, \dots, k\}$ e \mathbb{P} data da

$$\mathbb{P}(\{\ell\}) = \binom{k}{\ell} p^\ell (1-p)^{k-\ell}, \quad \ell = 0, 1, \dots, k.$$

Si noti che \mathbb{P} è effettivamente una probabilità. Infatti, per la formula del binomio di Newton vale che

$$\sum_{\ell=0}^k \mathbb{P}(\{\ell\}) = \sum_{\ell=0}^k \binom{k}{\ell} p^\ell (1-p)^{k-\ell} = (p + (1-p))^k = 1.$$

\mathbb{P} non è una probabilità uniforme. \mathbb{P} si chiama **probabilità binomiale**.

Chapter 6

Variabili Aleatorie

6.1 Introduzione

Finora, abbiamo trattato solo di eventi, per i quali ci chiediamo, fatto l'esperimento aleatorio, se sono avvenuti o meno. Possiamo pensare di domandarci non se un'affermazione si realizza o meno, ma generalizzare l'idea e chiederci quale sara' la quantita' numerica aleatoria, associata quindi ad un esperimento. Vogliamo quindi formalizzare l'idea di una quantita' che dipende dall'esito di un esperimento aleatorio:

Definition 6.1.1: Variabile Aleatoria

Diamo due affermazioni (come per gli eventi):

- **Affermazione:**

Una variabile aleatoria e' un'affermazione che riguarda il risultato dell'esperimento aleatorio. Tale affermazione identifica uno e un solo numero, una volta noto l'esito. Stiamo quindi rispondendo alla domanda "Quanto vale ... ?"

- **Funzione:**

Dato uno spazio di probabilita' (Ω, \mathbf{P}) ed un insieme $E \neq \emptyset$, si dice *variabile aleatoria* ogni funzione X del tipo:

$$X : \Omega \rightarrow E$$

Se $E = \mathbb{R}$, allora X e' una variabile aleatoria *reale*. Puo' anche essere che $E = \mathbb{R}^n$ con $n > 1$, e in questo caso si parla di variabili aleatorie *vettoriali*.

Example 6.1.1 (Lancio di due dadi)

Dato lo spazio di probabilita', definiamo una variabile aleatoria X :

$$X = \text{La somma dei due lanci}$$

Dobbiamo, quindi, prendere l'esito dell'esperimento e sommare il valore dei due dadi. Questa operazione puo' essere vista come una funzione a cui viene passata l'esito dell'esperimento e che restituisce un numero reale:

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

Questa e' la definizione di variabile aleatoria come funzione.

In questo caso, $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\} = DR_{6,2}$, quindi:

$$X((\omega_1, \omega_2)) : \omega_1 + \omega_2$$

Note:

Qualunque funzione da Ω a \mathbb{R} è una variabile aleatoria. Se Ω è più che numerabile sorgono dei problemi a causa dell'unione numerabile, quindi viene usata la σ -algebra, un insieme delle parti ristretto per evitare tali problemi.

6.2 Variabili Aleatorie Costanti

Vediamo dei casi "banali" di variabili aleatorie per capire il loro funzionamento e come vengono definite. Iniziamo con le funzioni costanti:

Definition 6.2.1: Variabile Aleatoria Costante

Una VA (variabile aleatoria) X si dice *costante* se:

$$\forall \omega \in \Omega. X(\omega) = a$$

Dove $a \in E$ è un elemento fissato. È definita $\forall \Omega$, dato che è deterministica (non dipende dall'esito dell'esperimento).

Come per gli eventi, anche le variabili aleatorie possono essere *quasi* costanti:

Definition 6.2.2: Variabili Aleatorie Quasi Costanti

Una VA X si dice *quasi costante* se:

$$\forall \omega \in \Omega. \mathbb{P}(X = a) = 1$$

Note:

La scrittura $(X = a)$ è una notazione che rappresenta l'evento $A =$ "Il valore di X sarà " a , ovvero il sottoinsieme di Ω per cui tutti gli elementi, passati a X , danno lo stesso valore a , quindi:

$$\mathbb{P}(X = a) := \mathbb{P}(\{\omega \in \Omega | X(\omega) = a\})$$

Example 6.2.1 (Lancio di un dado)

$\Omega = \mathbb{R}$, \mathbb{P} probabilità uniforme su $\{1, \dots, 6\}$ e nulla sul resto degli elementi.

Con $E = \mathbb{R}$, fisso $a \in E$ e voglio costruire $X : \Omega \rightarrow E$ tale che X è quasi costante:

$$X(\omega) = \begin{cases} a & \omega \in \{1, \dots, 6\} \\ \omega & \text{altrimenti} \end{cases}$$

Questa è effettivamente una VA quasi costante, dato che $\mathbb{P}(X = a) = 1$. Infatti abbiamo assegnato un valore costante a a tutti gli ω la cui probabilità non era nulla. Ricordiamoci che la notazione $\mathbb{P}(X = a)$ può essere riscritta meglio come $\mathbb{P}(\{\omega \in \Omega | X(\omega) = a\})$, che in questo caso corrisponde con $\mathbb{P}(\{1, \dots, 6\})$ che ovviamente è uguale a 1.

Notare che per tutti gli ω con probabilità nulla, il valore di X associato può essere qualunque cosa (non costante) e la X rimane comunque quasi costante.

6.3 Variabili Aleatorie Indicatrici (o di Bernulli)

Definition 6.3.1: Variabile Aleatoria Indicatrice

Dato un evento $A \subseteq \Omega$, definisco la variabile aleatoria indicatrice di A come:

$$X(\omega) : \mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

Quindi, dato un evento, la VA indicatrice X ci *indica* se l'esito ω appartenga o meno all'evento. Notiamo che tutta l'informazione dell'evento A e' contenuta nella VA:

$$A \rightsquigarrow \mathbb{1}_A$$

Allora le VA sono *generalizzazioni* del concetto di evento.

Example 6.3.1 (Prove ripetute e indipendenti)

Consideriamo uno schema di n prove ripetute e indipendenti con probabilità di successo p , ossia uno spazio di probabilità discreto (Ω, P) in cui sono definiti n eventi C_1, \dots, C_n indipendenti e con la stessa probabilità $p = P(C_i)$ (si ricordi il Paragrafo 1.3.4). Nel Paragrafo 1.3.4.3 abbiamo studiato gli eventi

$A_k :=$ "esattamente k prove hanno successo",

$B_\ell :=$ "il primo successo si verifica nella ℓ -esima prova",

per $0 \leq k \leq n$ e $1 \leq \ell \leq n$. Introduciamo ora due variabili aleatorie

$S :=$ "numero di successi nelle n prove",

$T :=$ "prova in cui si ha il primo successo",

definite su Ω a valori rispettivamente in $\{0, 1, \dots, n\}$ e in $\mathbb{N} \cup \{+\infty\}$, mediante

$$S(\omega) := \sum_{i=1}^n 1_{C_i}(\omega), \quad T(\omega) := \min\{i \in \{1, \dots, n\} : \omega \in C_i\}, \quad (6.1)$$

con la convenzione $\min \emptyset := +\infty$. Possiamo allora esprimere gli eventi A_k e B_ℓ nel modo seguente:

$$A_k = \{\omega \in \Omega : S(\omega) = k\}, \quad B_\ell = \{\omega \in \Omega : T(\omega) = \ell\}. \quad (6.2)$$

Questo mostra che gli eventi A_k e B_ℓ possono essere definiti in modo naturale in termini delle variabili aleatorie S e T , specificandone un sottoinsieme di valori.

6.4 Eventi associati alle variabili aleatorie

Ma se io ho X variabile aleatoria, sono capace di risalire all'evento (o eventi) associati? Le variabili aleatorie sono generalizzazioni del concetto di evento, quindi data una variabile aleatoria ci sono una moltitudine di eventi associati (o generati)

Definition 6.4.1: Evento generato da una VA

Sia $X : \Omega \rightarrow E$ una VA. $\forall A \subseteq \mathbb{R}$ indichiamo con $\{X \in A\}$ la controimmagine di A tramite X :

$$\{X \in A\} := X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$$

Quindi $\{X \in A\} \subseteq \Omega$ e' un evento, ed e' costituito da tutti e soli gli esiti ω per cui $X(\omega) \in A$.
Gli eventi di questo tipo si dicono *generati da X*.

Note:

La scrittura $\{X = a\}$ che abbiamo usato prima e' equivalente a scrivere $\{X \in \{a\}\}$. Allo stesso modo, possiamo considerare un intervallo di valori usando le disequazioni: $\{X > a\} \equiv \{X \in (a, +\infty)\}$.

Segue quindi:

Definition 6.4.2: Insieme di eventi generati da una VA

Data una VA X , l'insieme degli eventi da essa generati si indica con:

$$\sigma(X) := \{\{X \in A\} \mid A \subseteq \mathbb{R}\} \subseteq \mathcal{P}(\Omega)$$

Note:

$\forall X : \Omega \rightarrow \mathbb{R}$ v.a. possiamo scrivere:

$$\Omega = \{X \in \mathbb{R}\}$$

$$\emptyset = \{X \in \emptyset\}$$

Calcoliamo l'insieme degli eventi generati dalle VA particolari che abbiamo visto:

- **X costante:**

Fisso $a \in \mathbb{R}$, tale che $\forall \omega \in \Omega. X(\omega) = a$.

$\sigma(X) = ?$

Fissato $B \subseteq \mathbb{R}$, notiamo che ci sono solo due casi:

$$\{X \in B\} = \begin{cases} \Omega & a \in B \\ \emptyset & a \notin B \end{cases}$$

Infatti, se a appartiene a B allora $\forall \omega \in \Omega. X(\omega) \in B$ e quindi $\{X \in B\} = \Omega$. Mentre se a non appartiene a B , si ha che $\forall \omega \in \Omega. X(\omega) \notin B$ e quindi $\{X \in B\} = \emptyset$.

- **X indicativa:**

Fisso $A \subseteq \Omega$ tale che $\forall \omega \in \Omega. X(\omega) = \mathbb{1}_A(\omega)$.

$\sigma(X) = ?$

Fissato $B \subseteq \mathbb{R}$, vediamo i casi:

$$\{X \in B\} = \begin{cases} \Omega & 0 \in B \wedge 1 \in B \\ \emptyset & 0 \notin B \wedge 1 \notin B \\ A & 0 \notin B \wedge 1 \in B \\ A^c & 0 \in B \wedge 1 \notin B \end{cases}$$

Proposition 6.4.1

Sia $X : \Omega \rightarrow \mathbb{R}$ VA su (Ω, \mathcal{P}) , allora $\forall x \in \mathbb{R}$:

1. $\mathbb{P}(\{X \geq a\}) = \mathbb{P}(\{X = a\}) + \mathbb{P}(\{X > a\})$
2. $\mathbb{P}(\{X \geq a\}) = 1 - \mathbb{P}(\{X < a\})$

- : 1. $\{X \geq a\} = \{\omega \in \Omega | X(\omega) > a \vee X(\omega) = a\} = \{X > a\} \cup \{X = a\}$ dato che se $X(\omega) > a$ allora $X(\omega) \neq a$ e viceversa. Quindi l'equazione e' dimostrata per addittivita' finita.
2. Dimostriamo che $\{X \geq a\}$ e $\{X < a\}$ sono complementari. Se $X(\omega) \not\geq a$, allora $X(\omega) < a$ e viceversa, fatto. \square

Possiamo confrontare intervalli su \mathbb{R} per calcolare probailita'

6.5 Distribuzione (o legge) di una Variabile Aleatoria

Ora che sappiamo trovare l'evento generato da una variabile aleatoria e un sottoinsieme del suo codominio, possiamo calcolare la probabilita' di tale evento:

Definition 6.5.1: Distribuzione di una VA

Dati (Ω, \mathbb{P}) e $X : \Omega \rightarrow \mathbb{R}$ VA, chiamiamo legge di X la funzione:

$$\begin{aligned} \mathbb{P}_X : \mathcal{P}(\mathbb{R}) &\rightarrow [0, 1] \\ B &\mapsto \mathbb{P}(X \in B) \end{aligned}$$

Si scrive $X \sim \mathbb{P}_X$ e si legge "X ha legge \mathbb{P}_X ".

Proposition 6.5.1

La distribuzione \mathbb{P}_X di una variabile aleatoria $X : \Omega \rightarrow E$ e' una probabilita' sull'insieme E . (per noi E sara' quasi sempre \mathbb{R}).

La legge di una VA ne calcola quindi, dato un insieme di valori reali, la probabilita' che $X(\omega)$ appartenga a tale intervallo. Quindi, anche senza conoscere la funzione X , se conosciamo la sua distribuzione sappiamo quali valori puo' assumere e con quale probabilita'.

Andiamo a costruire le distribuzioni delle VA "banali" studiate precedentemente:

- Variabili Costanti:

$$a \in \mathbb{R}, X(\omega) = a. \forall \omega \in \Omega$$

$$B \subseteq \mathbb{R}. \{X \in B\} = \begin{cases} \Omega & a \in B \\ \emptyset & a \notin B \end{cases}$$

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \begin{cases} 1 & a \in B \\ 0 & a \notin B \end{cases}$$

Questa e' la delta di Dirac di a valutata su B

- Variabili Indicatrici:

$$A \subseteq \Omega, X(\omega) = \mathbb{1}_A(\omega)$$

$$B \subseteq \mathbb{R}, \{X \in B\} = \begin{cases} \Omega & 1, 0 \in B \\ \emptyset & 1, 0 \notin B \\ A & 1 \in B \wedge 0 \notin B \\ A^c & 1 \notin B \wedge 0 \in B \end{cases}$$

$$P_X(B) = P(X \in B) = \begin{cases} 1 & 1, 0 \in B \\ 0 & 1, 0 \notin B \\ P(A) & 1 \in B \wedge 0 \notin B \\ 1 - P(A) & 1 \notin B \wedge 0 \in B \end{cases}$$

Questa e' una probabilita' discreta, quindi e' possibile crearla usando una combinazione lineare di delta di Dirac:

$$= P(A)\delta_1(B) + (1 - P(A))\delta_0(B)$$

Ovvero:

$$P_X(\cdot) = P(A)\delta_1(\cdot) + (1 - P(A))\delta_0(\cdot)$$

Distribuzione (o legge) di Bernulli. Se la variabile aleatoria diventa piu' complicata, aumentano il numero di delta.

Definition 6.5.2: Distribuzioni di Bernulle

$$\mathbb{P}_x(\cdot) = \mathbb{P}(A)\delta_1(\cdot) + (1 - \mathbb{P}(A))\delta_0(\cdot)$$

6.6 Funzione di Ripartizione

Molto bella la distribuzione di una VA, ma prende come input insiemi (sottoinsiemi di \mathbb{R}) e questo ci complica un po' la vita. Vediamo una caratterizzazione piu' semplice che ci aiuterà soprattutto per le VA assolutamente continue:

Definition 6.6.1: Funzione di Ripartizione

Dati (Ω, P) SP e una VA $X : \Omega \rightarrow \mathbb{R}$, si chiama *funzione di ripartizione* o *CDF* di X la funzione:

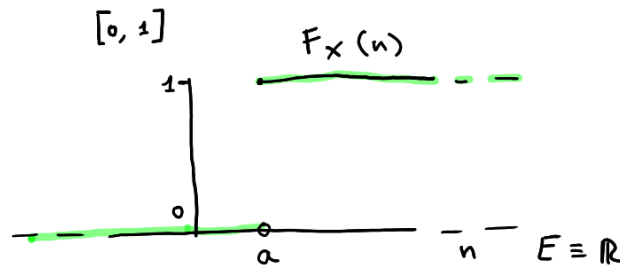
$$F_X : \mathbb{R} \rightarrow [0, 1] \\ x \mapsto P_X((-\infty, x]) = P(X \leq x)$$

Diamo un'occhiata a come sono definite le funzioni di ripartizione per i soliti due casi "banali" di VA:

- VA Costanti:

$X(\omega) = a \implies P_X(B) = \delta_a(B)$, quindi la relativa funzione di ripartizione sarà:

$$F_X(n) = P_X((-\infty, n]) = \delta_a((-\infty, n]) = \begin{cases} 1 & n \geq a \\ 0 & n < a \end{cases}$$

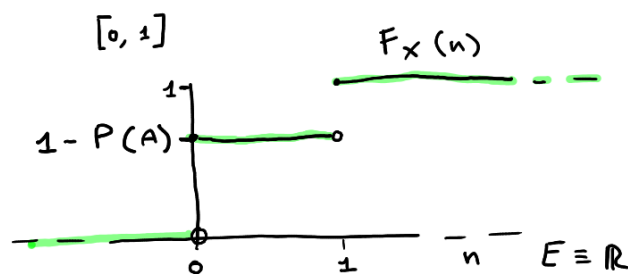


- VA Indicatrici:

$X(\omega) = \mathbb{1}_A \implies P_X(B) = P(A)\delta_1(B) + (1 - P(A))\delta_0(B)$, quindi la relativa funzione di ripartizione sarà:

$$F_X(x) = P_X((-\infty, x]) = P(A)\delta_1((-\infty, x]) + (1 - P(A))\delta_0((-\infty, x])$$

$$= \begin{cases} 1 & x \geq 1 \\ 1 - P(A) & 0 \leq x < 1 \\ 0 & x < 0 \end{cases}$$



Teorema dato da Shlein, che ci caratterizza la funzione di ripartizione:

Theorem 6.6.1 Teorema di Teodoro

(Ω, P) SP, $X : \Omega \rightarrow \mathbb{R}$, F_X fdr, allora:

1. F_X e' monotona crescente
2. F_X e' continua a destra, ovvero $\forall n \in \mathbb{R}. \lim_{y \rightarrow x^+} F_X(y) = F_X(n)$
3. $\lim_{n \rightarrow +\infty} F_X(n) = 1$
4. $\lim_{n \rightarrow -\infty} F_X(n) = 0$

Vale anche il viceversa: se $G : \mathbb{R} \rightarrow [0, 1]$ verifica 1,2,3,4, allora

$$\exists (\Omega, P), X : \Omega \rightarrow \mathbb{R}. G \equiv F_X$$

: • 1) Devo dimostrare che $\forall x, y \in \mathbb{R}, x < y. F_X(x) \leq F_X(y)$, ovvero che $P_X((-\infty, x]) \leq P_X((-\infty, y])$. Dato che $(-\infty, n] \subset (-\infty, y]$ e col fatto che P_X e' una probabilita', e' dimostrabile usando la proprieta' di monotonia delle probabilita'.

• 2) 3) 4) Per queste dimostrazioni serve dimostrare prima la *Stabilita' della Probabilita' per limiti monotoni*:



Lemma 6.6.1

Sia (Ω, P) SP

- Data " $A_x \uparrow A$ ", ovvero una successione di eventi tali che $A_x \subset A_{x+1}, \bigcup_{x=1}^{+\infty} A_x = A$, allora:

$$\lim_{x \rightarrow +\infty} P(A_x) = P(A)$$

- Data " $A_x \downarrow A$ ", ovvero una successione di eventi tali che $A_x \supset A_{x+1}, \bigcap_{x=1}^{+\infty} A_x = A$, allora:

$$\lim_{x \rightarrow +\infty} P(A_x) = P(A)$$

Stessa roba per monotona al contrario

Dato questo lemma, che non dimostriamo perche' si, possiamo concludere la dimostrazione precedente:

: 3) Dobbiamo dimostrare che $\lim_{x \rightarrow +\infty} P_X((-\infty, x]) = 1$. Usiamo il lemma appena introdotto, dato che se poniamo $A_x = (-\infty, x]$, possiamo creare una successione tale che $A_x \uparrow \mathbb{R}$, per cui $\lim_{x \rightarrow +\infty} P_X(A_x) = P_X(\mathbb{R}) = 1$. Gli altri punti sono lasciati al DarioDestroyer04 come compito per casa

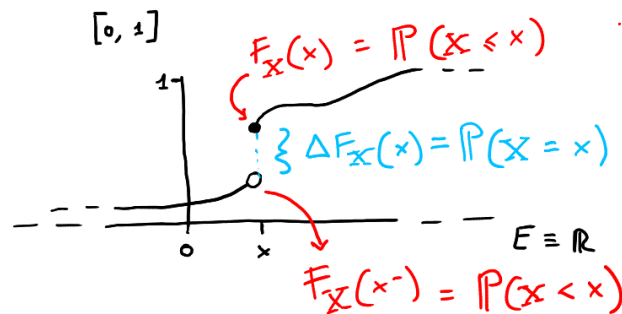


F_X e' continua a destra, quindi $\lim_{y \rightarrow x^+} F_X(y) = F_X(x) = P(X \leq x)$. Inoltre si puo' dimostrare (usando il lemma) che esiste anche il limite sinistro e vale:

$$\lim_{y \rightarrow x^-} F_X(y) = P(X < x) (F_X(x^-))$$

Inoltre, sappiamo che:

$$\underbrace{P(X \leq x)}_{F_X(x)} = \underbrace{P(X < x)}_{F_X(x^-)} + \underbrace{P(X = x)}_{\Delta F_X(x)}$$



Quindi possiamo calcolare la probabilit  di ogni intervallo in termini della F_X :

- $(a, b]$: $P_X((a, b]) = F_X(b) - F_X(a)$
- $[a, b)$: $P_X([a, b)) = F_X(b^-) - F_X(a^-)$
- (a, b) : $P_X((a, b)) = F_X(b^-) - F_X(a)$
- $[a, b]$: $P_X([a, b]) = F_X(b) - F_X(a^-)$

Note:

Conoscere $\mathbb{P}_X \forall B \subseteq \mathbb{R} \iff$ conoscere $\mathbb{P}_X \forall I$ intervallo di \mathbb{R} , dato che:

- Ogni intervallo   anche un'insieme
- Dato un sottoinsieme di \mathbb{R} , questo pu  sempre essere scritto come unione di intervalli

Quindi, F_X **determina completamente la distribuzione** di X , dato che ci da la distribuzione per tutti gli intervalli, e quindi per tutti gli insiemi.

6.7 Variabili Aleatorie Discrete

Dato che ci stiamo concentrando sulle probabilit  discrete, ci conviene applicare la definizione 3.3 anche per le VA:

Definition 6.7.1: Variabile Aleatoria Discreta

Una VA $X : \Omega \rightarrow E$ definita su uno spazio di probabilit  (Ω, P)   detta *discreta* se esiste un sottoinsieme $S_X \subset \text{Im}(X)$ finito o numerabile tale che:

$$P_X(S_X) := P(X \in S_X) = 1$$

Quindi un VA   discreta sse la sua distribuzione   una probabilit  discreta.
Il sottoinsieme S_X   chiamato supporto di X .

Note:

Se Ω o E sono finiti o numerabili, allora una VA definita su di essi sar  sicuramente discreto.

6.7.1 Caratterizzazione delle VA discrete

Che relazione esiste tra p_X e la distribuzione di X (P_X)? E tra p_X e F_X ?

Distribuzione

Essendo P_X una probabilit  discreta, possiamo definirla usando la Delta di Dirac:
Sia $S_X = \{x_1, \dots, x_n\}$ il supporto di X , allora $\exists p_1, \dots, p_n \in (0, 1], \sum_{i=1}^n p_i = 1$ tale che:

$$\forall B \subseteq E. P_X(B) = \sum_{i=1}^n p_i \delta_{x_i}(B)$$

Per definizione di P_X , sappiamo che $\forall x \in E$:

$$P_X(\{x\}) = P(X = x)$$

Per cui, mettendo insieme le due definizioni:

$$\forall x_i \in S_X. P_X(\{x_i\}) = \sum_{j=1}^n p_j \underbrace{\delta_{x_j}(\{x_i\})}_{\text{vale 1 solo quando } j=i} = p_i = P(X = x_i)$$

Quindi $\forall x_i \in S_X. p_i = P(X = x_i)$ e possiamo dire che X   una VA discreta se e solo se:

$$\forall B \subseteq E. P_X(B) = \sum_{i=1}^n P(X = x_i)$$

Dato che sara' una scrittura ricorrente, definiamo:

Definition 6.7.2: Densita' discreta di una VA

Data una VA discreta X , si definisce *densita' discreta* di X la funzione:

$$\begin{aligned} p_X : \mathbb{R} &\rightarrow [0, 1] \\ x &\mapsto P_X(\{x\}) = P(X = x) \end{aligned}$$

Possiamo quindi scrivere

$$\forall B \subseteq E. P_X(B) = \sum_{i=1}^n p_X(x_i)$$

Funzione di Ripartizione

Ricordiamo che possiamo esprimere $F_X(x)$ come $P_X((-\infty, x])$ per tutti i valori di x . Quindi, per quanto abbiamo detto sopra:

$$F_X(x) = \sum_{x_i \in (-\infty, x]} p_X(x_i) \delta_{x_i}((-\infty, x]) \quad \forall x \in \mathbb{R}$$

Ma se $x_i \notin S_X$, allora $p_X(x_i) = 0$, quindi:

- Se $\forall x_i \in S_X. x < x_i$, allora $F_X(x) = 0$
- Altrimenti, $F_X(x) = \sum_{x_i \in S_X. x_i \leq x} p_X(x_i)$

Cio' significa che il valore della fdr cambia solo nei punti del supporto, dove abbiamo dei "salti" del valore $p_X(x_i)$, mentre del resto   costante.

Conclusioni

Theorem 6.7.1 Caratterizzazione della distribuzione e fdr di una VA discreta tramite la sua densita'

Sia X una VA su (Ω, P) , le seguenti affermazioni sono equivalenti:

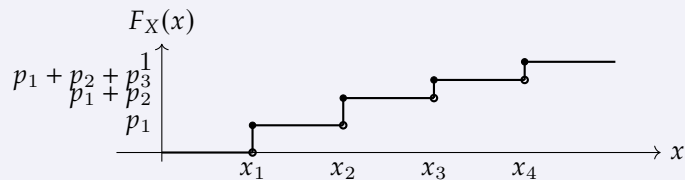
1. X   VA discreta con supporto S_X e densita' p_X

2. X ha fdr F_X **costante a tratti**: e' costante tranne nei punti di $S_X = \{x_1, \dots, x_n\}$, in cui F_X "salta" con ampiezza

$$\Delta F_X(x) = F_X(x) - F_X(x^-) = p_X(x)$$

La fdr ha quindi la seguente forma:

$$F_X(x) = \begin{cases} 0 & x < x_1, \\ p_1 & x_1 \leq x < x_2, \\ p_1 + p_2 & x_2 \leq x < x_3, \\ \vdots & \vdots \\ 1 & x \geq x_n. \end{cases}$$



Quindi F_X e' data dalla formula:

$$F_X(x) = \sum_{x_i \in S_X, x_i \leq x} p_X(x_i), \quad \forall x \in \mathbb{R}$$

3. se B è un sottoinsieme dei possibili valori di X (anche con probabilita' nulla), allora:

$$\mathbb{P}(X \in B) = \mathbb{P}_X(B) = \sum_{x_i \in B} p_X(x_i) = \sum_{x_i \in B} \Delta F_X(x_i) \quad (6.3)$$

Example 6.7.1 (Esercizio Riassuntivo (2.1 libro))

Sia $G : \mathbb{R} \rightarrow [0, 1]$ una funzione data da:

$$G(x) = \begin{cases} 0 & x < 0 \\ 1/2 & 0 \leq x < 1 \\ 2/3 & 1 \leq x < 2 \\ 11/12 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

1. Mostrare che G e' una funzione di ripartizione.
2. Mostrare che X e' discreta. Determinare supporto e densita' discreta di X .
3. Trovare P_X
4. Calcolare $P(X > 1/2), P(2 < X \leq 4), P(1 < X < 2), P(X < 3)$.
5. Mostrare che $Y = (X - 2)^2$ e' una variabile aleatoria discreta. Determinare S_{Yep_Y} .

Soluzione:

1. Dobbiamo dimostrare che G ha le quattro proprieta' delle funzioni di ripartizione:

- E' ovviamente monotona crescente
- In tutti i punti di discontinuita', e' sempre continua a dx ($\lim_{x \rightarrow k^+} G(x) = k$)

- I limiti a sx e dx sono 0 e 1

Quindi G è una funzione di ripartizione.

- Essendo $F_X(x)$ costante a tratti, X è per forza discreta con supporto $S_X = \{0, 1, 2, 3\}$.

La densità discreta $p_X(x)$ si calcola guardando il "salto" di $F_X(x)$ nel punto x :

$$\forall x \in S_X. p_X(x) = F_X(x) - F_X(x^-) = \Delta F_X(x)$$

TODO: fare graficuccio e tabellonski

- Per la caratterizzazione della distribuzione usando la densità:

$$P_X(B) = \sum_{x \in S_X} p_X(x) \delta_x(B)$$

- Va beh usa il punto prima per fare sta roba

- Y è una VA che può essere riscritta come una composizione fra una funzione $Z : \mathbb{R} \rightarrow \mathbb{R}$. $Z(x) = (x - 2)^2$ e X . Per questo motivo, se X ha lo stesso valore per degli $\omega \in \Omega$, allora anche in Y quegli ω avranno la stessa immagine. Ovvero, $|Im(Y)| \leq |Im(X)|$, che implica che $|S_Y| \leq |S_X|$. In più, possiamo dire che:

$$p_Y(y) = \sum_{x \in Z^{-1}(y)} p_X(x) \quad \forall y \in \mathbb{R}$$

Quindi $S_Y = \{y \in \mathbb{R} | \exists x \in S_X. y = Z(x)\}$ TODO: molte di ste cose non le ho dimostrate/spiegate bene, forse sarebbe meglio farlo prima?

6.8 Indici di sintesi di una distribuzione

La distribuzione di una VA discreta può essere descritta in maniera sintetica tramite due quantità numeriche:

- La *media*, che indica il valore "centrale" della distribuzione, ovvero il valore medio di X pesato con le probabilità di ogni valore
- La *varianza*, che è un indice di dispersione, ossia ci dice quanto quanto la distribuzione si concentra attorno alla media

6.8.1 Media o valore atteso

Definition 6.8.1: Valore atteso di una variabile aleatoria

Se X è una variabile aleatoria discreta con densità di probabilità p_X , allora il valore atteso di X è:

$$E[X] := \sum_{x \in S_X} x p_X(x)$$

Note:

$$\begin{aligned} \mathbb{E}[X - \mathbb{E}[X]] &= \mathbb{E}[X] - \mathbb{E}[\mathbb{E}[X]] \\ &= \mathbb{E}[X] - \mathbb{E}[X] = 0 \end{aligned}$$

Theorem 6.8.1 Stabilità delle VA con composizione

Sia X una variabile aleatoria discreta con densità di probabilità p_X e sia $g : \mathbb{R} \rightarrow \mathbb{R}$ una funzione. Allora

il valore atteso di $g(X)$ è:

$$\mathbb{E}[g(X)] = \sum_{x \in S_X} g(x)p_X(x)$$

Proposition 6.8.1 Linearita' del Valore atteso

Siano X, Y VA discrete, allora $\forall \alpha, \beta \in \mathbb{R}$:

$$\mathbb{E}[\alpha \cdot X + \beta \cdot Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]$$

Dimostrazione: Il bro fa una dimostrazione prendendo un solo caso, non ho capito TODO: boh



6.8.2 Varianza

Definition 6.8.2: Varianza di X

Se X è una variabile aleatoria discreta con densità di probabilità p_X e valore atteso $E[X]$, allora la varianza di X è:

$$\text{Var}(X) = \mathbb{E}[(X - E[X])^2] = \sum_{x \in S_X} (x - E[X])^2 p_X(x)$$

La varianza e' anche indicata con σ^2 .

Proposition 6.8.2

Si osservi che:

$$\begin{aligned} \text{var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \sum_{x \in S_X} x^2 \cdot p_X(x) - \left(\sum_{x \in S_X} x \cdot p_X(x) \right)^2 \end{aligned}$$

Definition 6.8.3: Deviazione standard

Si definisce *deviazione standard* la quantita':

$$\sigma(X) := \sqrt{\text{var}(X)}$$

Note:

Se X e' una grandezza fisica espressa in una certa unita di misura, la deviazione standard ha il vantaggio di avere la stessa unita' di misura.

6.9 Distribuzioni notevoli di VA discrete

6.9.1 Distribuzione uniforme discreta

Definition 6.9.1: Distribuzione uniforme discreta

Una variabile aleatoria X si dice avere una distribuzione uniforme discreta su un insieme di supporto $S_X = \{x_1, \dots, x_n\}$ di cardinalità n se la densità di probabilità di X è:

$$p_X(x) = \frac{1}{n}, \quad \forall x \in S_X$$

E si scrive $X \sim \text{Unif}(\{x_1, \dots, x_n\})$

Si può rappresentare questa distribuzione attraverso questa tabella:

X	x_1	x_2	\dots	x_n	
$p_X(x)$	$\frac{1}{n}$	$\frac{1}{n}$	\dots	$\frac{1}{n}$	1

In pratica il valore atteso di una variabile aleatoria con distribuzione uniforme discreta è:

$$E[X] = \sum_{x \in S_X} x p_X(x) = \sum_{x \in S_X} x \frac{1}{n} = \frac{1}{n} \sum_{x \in S_X} x = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}$$

mentre la varianza è:

$$Var(X) = \sum_{x \in S_X} (x - E[X])^2 p_X(x) = \frac{1}{n} \sum_{i=1}^n (x_i - E[X])^2$$

6.9.2 Distribuzione di Bernulli

Definition 6.9.2: Distribuzione di Bernulli di parametro $p \in [0, 1]$

Una variabile aleatoria X con un insieme di supporto $S_X = \{0, 1\}$ si dice avere una distribuzione di Bernoulli di parametro p se la sua densità di probabilità è:

$$p_X(x) = \begin{cases} p & \text{se } x = 1, \\ 1 - p & \text{se } x = 0 \end{cases}$$

Si può rappresentare questa distribuzione attraverso questa tabella:

X	0	1
$p_X(x)$	$1 - p$	p

Si noti che le variabili aleatorie di Bernoulli sono tutte Variabili aleatorie indicatrici:

$$X = \mathbb{1}_A$$

$$X(w) = \begin{cases} 1 & w \in A \\ 0 & w \notin A \end{cases} \quad X \sim B(p)$$

In pratica il valore atteso di una variabile aleatoria con distribuzione uniforme discreta è:

$$E[X] = \sum_{x \in S_X} x p_X(x) = 0 \cdot (1 - p) + 1 \cdot p = p$$

Mentre la sua varianza è:

$$Var(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p)$$

Si noti Inoltre che:

$$E[X^2] = \sum_i x_i^2 p_X(x_i) = 0 \cdot (1 - p) + 1 \cdot p = p$$

6.9.3 Distribuzione binomiale di parametri $p \in [0, 1]$ e $n \in \mathbb{N}$

Si consideri ora una generalizzazione della distribuzione di Bernoulli, detta distribuzione binomiale. Una variabile aleatoria X ha distribuzione di Bernoulli se è una variabile aleatoria indicatrice di un qualche evento A . Se invece consideriamo n prove ripetute e indipendenti di un esperimento di Bernoulli, allora la variabile aleatoria che conta il numero di successi ottenuti in n si dice che ha una distribuzione binomiale. Avremo quindi n variabili aleatorie di Bernoulli:

$$X_1 \sim B(p), X_2 \sim B(p), \dots, X_n \sim B(p)$$

Ora si consideri la seguente variabile aleatoria:

$$X = \text{"n di successi negli n esperimenti"} = X_1 + X_2 + \dots + X_n$$

Si noti questo esempio:

Example 6.9.1 (Estrazione con reimmissione da un'urna)

Consideriamo un'urna contenente b palline bianche e r palline rosse, per un totale di $b + r$ palline. Si effettuano n estrazioni con reimmissione, e definiamo la variabile aleatoria X come:

$X = \text{"numero di palline bianche estratte"}$.

Mostrare che X è una variabile aleatoria discreta e determinarne supporto e densità discreta.

Soluzione: Si ha che $S_X = \{0, 1, 2, \dots, n\}$. Occorre calcolare $p_X(k)$ per $k = 0, 1, 2, \dots, n$. Sia:

$$A_k = \{X = k\}, \quad \forall k \in S_X$$

l'evento si può definire:

$$A_k = \{\text{"esattamente } k \text{ palline bianche estratte"}\}$$

Ovviamete la probabilità di A_k è:

$$\mathbb{P}(A_k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Dove $p = \frac{b}{b+r}$ è la probabilità di estrarre una pallina bianca in una singola estrazione. Quindi

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

Come seguirà dalla definizione, X ha distribuzione binomiale di parametri n e p , ovvero $X \sim B(n, p)$



Si può arrivare così alla definizione di distribuzione binomiale:

Definition 6.9.3

Una variabile aleatoria X con un insieme di supporto $S_X = \{0, 1, \dots, n\}$ si dice avere una distribuzione binomiale di parametri p e n se la sua densità di probabilità è:

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad \forall x \in S_X$$

riassunta nella seguente tabella:

X	0	1	2	...	n
$p_X(x)$	$\binom{n}{0} p^0 (1-p)^n$	$\binom{n}{1} p^1 (1-p)^{n-1}$	$\binom{n}{2} p^2 (1-p)^{n-2}$...	$\binom{n}{n} p^n (1-p)^0$

In tal caso scriviamo

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

Note:

Quando $n = 1$ si ha la distribuzione di Bernoulli, ovvero $B(1, p)$

Proposition 6.9.1

Siano $0 \leq p \leq 1$, $n \in \mathbb{N}$ e $X \sim B(n, p)$. Allora

$$\begin{aligned} E[X] &= np \\ \text{Var}(X) &= np(1-p) \end{aligned}$$

Dimostrazione: 1. Calcoliamo prima l'attesa $\mathbb{E}[X]$:

$$\begin{aligned}
 \mathbb{E}[X] &= \sum_{k=0}^n k p_X(k) = \sum_{k=1}^n k p_X(k) \\
 &= \sum_{k=1}^n k \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \sum_{k=1}^n k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
 &= \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
 &= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\
 &\stackrel{h=k-1}{=} np \sum_{h=0}^{n-1} \frac{(n-1)!}{h!(n-1-h)!} p^h (1-p)^{n-1-h} \\
 &\stackrel{\text{form. binom., Newton}}{=} np.
 \end{aligned}$$

Qui abbiamo usato la formula del binomio e il teorema di Newton per concludere che:

$$\sum_{h=0}^{n-1} \binom{n-1}{h} p^h (1-p)^{n-1-h} = 1.$$

2. Calcoliamo ora la varianza $\text{Var}(X)$: dato che $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$, resta da calcolare $\mathbb{E}[X^2]$. Inoltre, $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \mathbb{E}[X(X-1)] + np$, quindi dobbiamo calcolare $\mathbb{E}[X(X-1)]$. Si ha che:

$$\begin{aligned}
 \mathbb{E}[X(X-1)] &= \sum_{k=0}^n k(k-1) p_X(k) = \sum_{k=2}^n k(k-1) p_X(k) \\
 &= \sum_{k=2}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \sum_{k=2}^n k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
 &= \sum_{k=2}^n \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} \\
 &= n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} (1-p)^{n-k} \\
 &\stackrel{h=k-2}{=} n(n-1)p^2 \sum_{h=0}^{n-2} \frac{(n-2)!}{h!(n-2-h)!} p^h (1-p)^{n-2-h} \\
 &\stackrel{\text{form. binom., Newton}}{=} n(n-1)p^2.
 \end{aligned}$$

Qui abbiamo nuovamente usato la formula del binomio e il teorema di Newton per concludere che:

$$\sum_{h=0}^{n-2} \binom{n-2}{h} p^h (1-p)^{n-2-h} = 1.$$

6.9.4 Distribuzione di Poisson

Caso limite della distribuzione binomiale, serve per eventi con probabilità molto piccola. Viene passato un solo parametro λ

$$n \rightarrow +\infty, p \rightarrow 0, np \rightarrow \lambda > 0$$

$X \sim Poi(\lambda), \lambda > 0, S_X = \{0, 1, \dots\} = \mathbb{N}_0$ (il supporto è infinito numerabile):

$$\forall k \in \mathbb{N}_0. p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Lo sviluppo in serie di Taylor dell'esponenziale $\forall x \in \mathbb{R}$:

$$\sum_{k=0}^{+\infty} \frac{x^k}{k!} = e^x$$

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=0}^{+\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda \end{aligned}$$

$$\begin{aligned} var(X) &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X(X-1)] &= \sum_{k=0}^{+\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda^2 \sum_{k=2}^{+\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} \\ &= \lambda^2 \cdot 1 = \lambda^2 \end{aligned}$$

6.10 Variabili Aleatorie Continue

È possibile trovarsi in una situazione in cui una variabile aleatoria può assumere un'infinita *continua* (e quindi non numerabile) di valori. Per questo motivo è utile studiare le variabili aleatorie continue, che si differenziano rispetto a quelle discrete in diversi aspetti.

Example 6.10.1

$X = \text{tempo di vita di un componente}$

Quindi X può assumere qualunque valore positivo o 0, da cui si ha

$$Im(X) = [0, +\infty)$$

E' molto piu' naturale modellare l'intervallo di tempi come il continuo, al posto di discretizzarlo (e' una forzatura)

Ma, per le proprieta' del continuo, questa scelta ci impone che

$$p_X(x) = P(X = x) = 0 \quad \forall x$$

Ovvero, le densita' discrete e' nulla per tutti i possibili valori singoli di una VA continua.

Domanda cruciale: "Come costruiamo la distribuzione di X ?"

Possiamo provare a calcolare la probabilita' di un intervallo di valori. Tornando all' esempio 6.10, ci aspettiamo che:

$$\forall a, b \in [0, +\infty), a < b. \quad P(a \leq X \leq b) > 0$$

Non ci interessa quindi il valore in un punto della nostra funzione di densita', ma il valore sugli intervalli. Data una funzione $f_X : \mathbb{R} \rightarrow \mathbb{R}$ (che chiameremo *densita' continua*), possiamo provare a usare l'integrale per calcolare la probabilita' agli eventi $\{X \in [a, b]\} = P(a \leq X \leq b)$

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

Notare che se $a = b$, allora il valore dell'integrale e' nullo, proprio come ci aspettavamo. Per assicurarci che l'integrale di tale funzione sia effettivamente una probabilita', la f_X deve verificare due proprieta':

- Deve essere ≥ 0 per ogni elemento del dominio, perche' l'integrale di una funzione negativa su un intervallo sara' sicuramente negativa
- $P(-\infty \leq X \leq +\infty)$ deve essere 1, dato che e' l'evento certo, quindi:

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1$$

6.10.1 Definizioni di densita' e VA continue

Diamo quindi la definizione formale di *densita' continua*:

Definition 6.10.1: Densita' continua

Una funzione $f_X : \mathbb{R} \rightarrow \mathbb{R}$ e' una *densita' continua* se:

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

Example 6.10.2

Un possibile valore per la densita' continua dell'esempio 6.10:

$$f_X(x) = \begin{cases} 0 & x < 0 \\ e^{-x} & x \geq 0 \end{cases}$$

todo: graficuz

calcola se valgono le proprieta' sopra

La probabilita' dell'intervallo graficamente e' l'area del sottografo di f_X nello stesso intervallo. Il vincolo e' che l'area di tutto il sottografo sia 1.

Note:

Notare che, a differenza della densita' discreta che ha valore massimo 1, la densita' continua non ha valore limite superiore. Cio' e' dovuto al fatto che ci interessa il suo integrale su un intervallo che non e' determinato dal valore della funzione in un solo punto, quindi puo' benissimo valere $f_X(x) > 1$ (puo' anche essere illimitata).

Example 6.10.3

Nell'esempio di prima, se mettiamo $\lambda e^{-\lambda x}$ per qualunque l'integrale e' sempre 1!

Dalla definizione di densita' continua possiamo definire anche le VA continue:

Definition 6.10.2: Variabile Aleatoria Continua

Sia (Ω, P) SP e X VA. Si dice che X e' *continua* se \exists una densita' di probabilita', indicata con f_X , tale che:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx \quad \forall [a, b] \subset \mathbb{R} \quad (6.4)$$

Example 6.10.4

Esempi di utilizzo della formula:

- $P(X \leq a) = \int_{-\infty}^a f_X(x) dx$
- $P(X < a) = \int_{-\infty}^a f_X(x) dx$
- $P(X > a) = \int_a^{+\infty} f_X(x) dx$
- $P(X \in B) = \int_B f_X(x) dx$
- $P(X \in [a, b]) = \int_a^b f_X(x) dx$
- $P(X \in (a, b)) = \int_a^b f_X(x) dx$
- $P(X \in [a, b)) = \int_a^b f_X(x) dx$
- $P(X \in (a, b]) = \int_a^b f_X(x) dx$
- $P(X \in \{a\}) = \int_a^a f_X(x) dx = 0$

Note:

Non esiste una sola funzione di densita' continua per una variabile. Infatti, data f_X , e' possibile modificare il suo valore in un numero limitato o infinito numerabile di punti senza cambiare il valore dell'integrale:

$$\forall [a, b] \subset \mathbb{R}. \quad \int_a^b f_X dx = \int_a^b g_X dx$$

In questo caso, anche g_X e' una densita' di X . Vedremo che cio' non causa problemi, e che in molti casi esiste una versione canonica.

A causa di tale non unicita', non e' possibile definire un supporto, che sarebbe dato dall'insieme

$$S_X = \{x \in \mathbb{R} | f_X(x) > 0\}$$

dato che per diverse densita' della stessa VA gli insiemi sono diversi. Anche in questo caso possiamo considerare la densita' canonica (se c'e'), da cui deriva il supporto canonico.

6.10.2 Funzione di Ripartizione Continua

Una delle funzioni che ci permettono di caratterizzare le VA discrete ora diventa continua: la funzione di ripartizione (Bonzo non ci crede, e' da investigare)

Infatti, $P(a \leq X \leq a) = P(X = a) = \int_a^a f_X(x)dx = 0$, quindi per avere una probabilita' positiva dobbiamo per forza considerare un intervallo. Quindi, tornando alle proprieta' della funzione di ripartizione, vediamo che:

- F_X e' sempre continua a destra
- $F_X(x) - F_X(x^-) = P(X = x)$

Ma per quello che abbiamo detto, F_X e' continua anche a sx, quindi no more scalini.

Ebbene buon bastianini, sai come lo vedremo? GRAZIE A QUESTO EFFICACISSIMO TEOREMA:

Theorem 6.10.1 Teorma della della densità e della continuità di f_X

Sia X una variabile aleatoria continua con densità di probabilità f_X . Allora:

- La densità discreta discreta di X è identivamente uguale a zero, ovvero:

$$p_X(x) = 0 \quad \forall x \in \mathbb{R}$$

- La funzione di ripartizione F_X è continua e si può calcolare come:

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

Dimostrazione: • Dalla definizione di $p_X(x)$, abbiamo che:

$$p_X(x) = P(X = x) = \int_x^x f_X(y)dy = 0$$

Dimostrato

- Dalla definizione di F_X si ha che

$$F_X(x) = P(X \leq x) = P(-\infty < X \leq x) = \int_{-\infty}^x f_X(y)dy$$

Devo dimostrare che F_X è continua a destra e a sinistra:

- Devo dimostrare che F_X è continua a destra, ovvero che

$$F_X(x) = \lim_{a \rightarrow x^+} F_X(a)$$

Si ha che:

$$\lim_{a \rightarrow x^+} F_X(a) = \lim_{a \rightarrow x^+} \mathbb{P}(X \leq a) = \lim_{a \rightarrow x^+} \int_{-\infty}^a f_X(y)dy = \int_{-\infty}^x f_X(y)dy = F_X(x)$$

Dimostrato

- Devo dimostrare che F_X è continua a sinistra, ovvero che

$$F_X(x) = \lim_{a \rightarrow x^-} F_X(a)$$

E si ha che $\lim_{a \rightarrow x^-} F_X(a) = \mathbb{P}(X < x)$, allora

$$\lim_{a \rightarrow x^-} F_X(a) = \mathbb{P}(X < x) = \lim_{a \rightarrow x^-} \int_{-\infty}^a f_X(y)dy = \int_{-\infty}^x f_X(y)dy = F_X(x)$$



Quindi, la funzione di ripartizione di una VA continua e' una funzione integrale, ed e' percio' *assolutamente continua* (quindi e' continua con altre proprieta' in piu')

Note:

Si può notare quindi che:

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \int_a^b f_X(x) dx$$

sono in particolare date da

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

A differenza delle VA discrete, quindi, l'inclusione o meno degli estremi non cambia il valore della probabilità su un intervallo.

6.10.3 Confronto tra VA continua e discreta

Variabili aleatorie discrete	Variabili aleatorie continue
densità discreta p_X	densità continua f_X
$\mathbb{P}(X \in B) = \sum_{x_i \in B} p_X(x_i)$	$\mathbb{P}(X \in B) = \int_B f_X(x) dx$
F_X è costante a tratti: $F_X(x) = \sum_{x_i \leq x} p_X(x_i)$	F_X è una funzione integrale o, equivalentemente, è una funzione assolutamente continua: $F_X(x) = \int_{-\infty}^x f_X(y) dy$

6.10.4 Dalla funzione di ripartizione alla densità continua

Dal teorema precedente sappiamo che

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

Viene quindi da chiedersi come calcolare il contrario, ci dobbiamo porre quindi due domande:

- Che proprietà deve avere F_X per essere sicuri che X è continua?
- Come si trova f_X partendo da F_X ?

Per rispondere alla prima domanda, sappiamo che X è continua sse F_X è assolutamente continua. Ma dimostrare ciò è abbastanza difficile, quindi solitamente se non ci è detto che X è continua si usa tale proprietà:

Proposition 6.10.1 Condizione sufficiente per continuità assoluta

Se una funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ è C^1 (differenziabile almeno una volta) a tratti, allora è assolutamente continua

Dato per vero che X è continua, dobbiamo calcolare la sua densità:

Proposition 6.10.2

Sia X una variabile aleatoria continua con F_X la sua funzione di ripartizione. Allora una densità di probabilità è data da:

$$f_X(x) := \frac{d}{dx} F_X(x) \quad \forall x \text{ in cui } F_X \text{ è derivabile}$$

Dimostrazione: Sia $F_X(x)$ la funzione di ripartizione di una variabile aleatoria continua X . Per definizione:

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

Per il teorema fondamentale del calcolo integrale, se f_X è continua, allora F_X è derivabile e la derivata di F_X è uguale alla funzione integranda f_X . Quindi:

$$\frac{d}{dx} F_X(x) = f_X(x).$$

Dimostrato.



6.10.5 Funzioni di variabili aleatorie continue

Siano $h : \mathbb{R} \rightarrow \mathbb{R}$ una qualunque funzione e X una variabile aleatoria continua. Poniamo:

$$Y = h(X)$$

Ricordiamo che quando X è discreta, Y è necessariamente anch'essa una variabile aleatoria discreta. Al contrario, quando X è continua, non possiamo dire nulla su Y . In particolare, Y potrebbe essere discreta, continua, mista. Adesso viene mostrato come calcolare la densità di probabilità di Y in funzione della densità di probabilità di X .

Metodo generale per calcolare la densità di Y

Per calcolare la densità di Y in funzione della densità di X calcolare le CDF, in pratica:

- Calcolare la funzione di ripartizione $F_Y(y)$:

$$F_Y(y) = P(Y \leq y) = P(h(X) \leq y)$$

- Calcolare la densità di Y :

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} P(h(X) \leq y)$$

Example 6.10.5

Sia $h(x) = x^2$, quindi $Y = h(X) = X^2$. Si trovi la densità di probabilità di Y in funzione della densità di probabilità di X .

Soluzione: Iniziamo col determinare la funzione di ripartizione di Y :

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y)$$

Dobbiamo risolvere la disuguaglianza $X^2 \leq y$, distinguendo i due casi:

- Se $y < 0$, allora $X^2 \leq y$ non ha soluzioni, quindi $F_Y(y) = 0$.
- Se $y \geq 0$, allora $X^2 \leq y$ è equivalente a $-\sqrt{y} \leq X \leq \sqrt{y}$, quindi:

$$F_Y(y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = P(X \leq \sqrt{y}) - P(X < -\sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

Quindi, la funzione di ripartizione di Y è:

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ F_X(\sqrt{y}) - F_X(-\sqrt{y}) & y \geq 0 \end{cases}$$

Ora calcoliamo la densità di probabilità di Y :

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(\sqrt{y}) - \frac{d}{dy} F_X(-\sqrt{y})$$

Applicando la regola della catena, otteniamo:

$$f_Y(y) = \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) - \frac{1}{2\sqrt{y}} f_X(-\sqrt{y})$$

Quindi, la densità di probabilità di Y è:

$$f_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) - \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) & y \geq 0 \end{cases}$$

6.10.6 Distribuzioni notevoli di VA continue

Distribuzione uniforme continua

Definition 6.10.3: Distribuzione uniforme continua

Una variabile aleatoria X si dice avere una distribuzione uniforme continua su un intervallo $[a, b]$ se la sua densità di probabilità è:

$$f_X(x) = \frac{1}{b-a} \cdot \mathbb{1}_{(a,b)}(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{altrimenti} \end{cases} \quad (6.5)$$

Si ha, quindi, che

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1 = \int_a^b \frac{1}{b-a} dx = \frac{1}{b-a} (b-a)$$

Pertanto la funzione di densità di probabilità è costante in $[a, b]$ e nulla al di fuori di questo intervallo. Si può rappresentare questa distribuzione attraverso questa tabella:

X	a	b	
$f_X(x)$	$\frac{1}{b-a}$	$\frac{1}{b-a}$	1

Il grafico della distribuzione è il seguente: TODO: non compila

Si possono fare ulteriori considerazioni:

Proposition 6.10.3 Funzione di ripartizione di una variabile aleatoria continua con distribuzione costante

Sia X una variabile aleatoria continua con distribuzione uniforme continua su $[a, b]$. Allora la funzione di ripartizione di X è:

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Note:

Si noti che la funzione di ripartizione è continua e crescente in $[a, b]$.

Dimostrazione: La funzione di ripartizione è definita come:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(y) dy$$

Quindi, per calcolare la funzione di ripartizione, dobbiamo considerare i tre casi:

- Se $x < a$, allora $F_X(x) = 0$.
- Se $a \leq x \leq b$, allora:

$$F_X(x) = \int_{-\infty}^x f_X(y) dy = \int_a^x f_X(y) dy = \int_a^x \frac{1}{b-a} dy = \frac{x-a}{b-a}$$

- Se $x > b$, allora $F_X(x) = 1$



Ora passiamo a calcolare il valore atteso e la varianza di una variabile aleatoria continua con distribuzione uniforme continua:

Proposition 6.10.4

Sia X una variabile aleatoria continua con distribuzione uniforme continua su $[a, b]$. Allora:

$$E[X] = \frac{a+b}{2}$$
$$Var(X) = \frac{(b-a)^2}{12}$$

Dimostrazione: Ricordando che

$$f_X(x) = \frac{1}{b-a}$$

$$\begin{aligned} E[X] &= \int_{-\infty}^{+\infty} x f_X(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \left(\frac{b^2}{2} - \frac{a^2}{2} \right) \\ &= \frac{1}{b-a} \cdot \frac{(b-a)(b+a)}{2} = \frac{a+b}{2} \end{aligned}$$

Quindi, il valore atteso di una variabile aleatoria continua con distribuzione uniforme continua su $[a, b]$ è:

$$E[X] = \frac{a+b}{2}$$

Per calcolare la varianza, utilizziamo la formula:

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

Calcoliamo prima $E[X^2]$:

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_a^b x^2 f_X(x) dx = \int_a^b x^2 \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{1}{b-a} \left(\frac{b^3}{3} - \frac{a^3}{3} \right) \\ &= \frac{1}{3(b-a)} (b^3 - a^3) = \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} \\ &= \frac{(b^2 + ab + a^2)}{3} \end{aligned}$$

Quindi, la varianza è:

$$\begin{aligned} Var(X) &= E[X^2] - (E[X])^2 = \frac{(b^2 + ab + a^2)}{3} - \left(\frac{a+b}{2} \right)^2 \\ &= \frac{(b^2 + ab + a^2)}{3} - \frac{(a^2 + 2ab + b^2)}{4} \\ &= \frac{4(b^2 + ab + a^2) - 3(a^2 + 2ab + b^2)}{12} \\ &= \frac{(b^2 - 2ab + a^2)}{12} = \frac{(b-a)^2}{12} \end{aligned}$$



Distribuzione esponenziale

La distribuzione esponenziale si ha quando si ha un processo di Poisson, ovvero quando si ha un evento che accade in un certo intervallo di tempo, ad esempio si usa per descrivere il tempo di vita di un macchinario oppure di un componente elettronico

Definition 6.10.4: Distribuzione esponenziale

Una variabile aleatoria X si dice avere una distribuzione esponenziale con parametro $\lambda > 0$ se la sua densità di probabilità è: TODO: non compila plot

$$f_X(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}_{(0,+\infty)}(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{altrimenti} \end{cases}$$

il cui grafico è: TODO

Si noti che

Proposition 6.10.5 Funzione di ripartizione di una variabile aleatoria continua con distribuzione esponenziale

Sia X una variabile aleatoria continua con distribuzione esponenziale con parametro $\lambda > 0$. Allora la funzione di ripartizione di X è:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases} \quad (6.6)$$

In tal caso scriviamo

$$X \sim \text{Exp}(\lambda)$$

Dimostrazione: La funzione di ripartizione è definita come:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(y) dy$$

Quindi, per calcolare la funzione di ripartizione, dobbiamo considerare i due casi:

- Se $x < 0$, allora $F_X(x) = 0$.
- Se $x \geq 0$, allora:

$$F_X(x) = \int_{-\infty}^x f_X(y) dy = \int_0^x f_X(y) dy = \int_0^x \lambda e^{-\lambda y} dy = -e^{-\lambda y} \Big|_0^x = -e^{-\lambda x} + 1$$

Quindi, la funzione di ripartizione di X è:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

**Note:**

Si noti che la funzione di ripartizione è continua e crescente in $[0, +\infty)$.

Proposition 6.10.6

Sia X una variabile aleatoria continua con distribuzione esponenziale con parametro $\lambda > 0$. Allora:

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

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

Dimostrazione: Ricordando che

$$f_X(x) = \lambda e^{-\lambda x}$$

$$\begin{aligned}
 E[X] &= \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^{+\infty} x \lambda e^{-\lambda x} dx \\
 &= -\frac{1}{\lambda^2} e^{-\lambda x} \left(x + \frac{1}{\lambda} \right) \Big|_0^{+\infty} \\
 &= -\frac{1}{\lambda^2} (0 + 0) + \frac{1}{\lambda^2} = \frac{1}{\lambda}
 \end{aligned}$$

Quindi, il valore atteso di una variabile aleatoria continua con distribuzione esponenziale è:

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

Per calcolare la varianza, utilizziamo la formula:

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

Calcoliamo prima $E[X^2]$:

$$\begin{aligned}
 E[X^2] &= \int_{-\infty}^{+\infty} x^2 f_X(x) dx = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx \\
 &= -\frac{1}{\lambda^3} e^{-\lambda x} \left(x^2 + 2\frac{x}{\lambda} + \frac{1}{\lambda^2} \right) \Big|_0^{+\infty} \\
 &= -\frac{1}{\lambda^3} (0 + 0 + 0) + \frac{1}{\lambda^3} = \frac{2}{\lambda^2}
 \end{aligned}$$

Quindi, la varianza è:

$$\begin{aligned}
 Var(X) &= E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda} \right)^2 \\
 &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}
 \end{aligned}$$



Note:

Si noti che la varianza è sempre positiva e tende a zero quando λ tende a infinito.

Distribuzione Normale (o Gaussiana)

Definition 6.10.5: Distribuzione Normale

Una variabile aleatoria X si dice avere una distribuzione normale con media $\mu \in \mathbb{R}$ e varianza $\sigma^2 > 0$ se la sua densità di probabilità è:

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

in tal caso scriviamo

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Qui la sua funzione di ripartizione è

Proposition 6.10.7 Funzione di ripartizione di una variabile aleatoria continua con distribuzione normale

Sia X una variabile aleatoria continua con distribuzione normale con media $\mu \in \mathbb{R}$ e varianza $\sigma^2 > 0$. Allora la funzione di ripartizione di X è:

$$F_X(x) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{x - \mu}{\sigma\sqrt{2}} \right) \right)$$

Dimostrazione: La funzione di ripartizione è definita come:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(y) dy$$

Quindi, per calcolare la funzione di ripartizione, dobbiamo considerare i due casi:

- Se $x < -\infty$, allora $F_X(x) = 0$.
- Se $x > +\infty$, allora $F_X(x) = 1$.
- Se $-\infty < x < +\infty$, allora:

$$F_X(x) = \int_{-\infty}^x f_X(y) dy = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

Questo integrale non ha una soluzione analitica, ma può essere calcolato numericamente utilizzando la funzione errore erf.



Distribuzione Normale standard

La distribuzione normale standard è un caso particolare della distribuzione normale in cui:

- La media $\mu = 0$
- La varianza $\sigma^2 = 1$

La notazione è

$$Z \sim \mathcal{N}(0, 1)$$

Per dovere di cronaca riporto la definizione

Definition 6.10.6: Distribuzione Normale standard

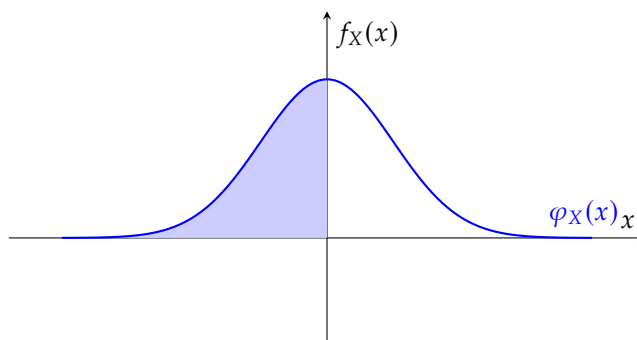
Una variabile aleatoria Z si dice avere una distribuzione normale standard se la sua densità di probabilità è:

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

in tal caso scriviamo

$$Z \sim \mathcal{N}(0, 1)$$

Si noti che la distribuzione normale standard è simmetrica rispetto all'asse y e ha una forma a campana



Si noti la funzione di ripartizione

Proposition 6.10.8 funzione di ripartizione della distribuzione normale standard

Sia Z una variabile aleatoria continua con distribuzione normale standard. Allora la funzione di ripartizione

di Z è:

$$F_Z(z) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{z}{\sqrt{2}} \right) \right)$$

Dimostrazione: La funzione di ripartizione è definita come:

$$\Phi_Z(x) = P(Z \leq z) = \int_{-\infty}^z f_Z(y) dy$$

Quindi, per calcolare la funzione di ripartizione, dobbiamo considerare i due casi:

- Se $z < -\infty$, allora $F_Z(z) = 0$.
- Se $z > +\infty$, allora $F_Z(z) = 1$.
- Se $-\infty < z < +\infty$, allora:

$$F_Z(z) = \int_{-\infty}^z f_Z(y) dy = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

Questo integrale non ha una soluzione analitica, ma può essere calcolato numericamente utilizzando la funzione errore erf.



Questo è il grafico della funzione di ripartizione della distribuzione normale standard: TODO (non c'ho voglia di farlo ora)

Proposizione utilina

Proposition 6.10.9

Sia X una variabile aleatoria continua con distribuzione normale standard $Z \sim \mathcal{N}(0, 1)$. Allora:

$$E[Z] = 0$$

$$\operatorname{Var}(Z) = 1$$

Dimostrazione: Ricordando che

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Si ha:

$$\begin{aligned} E[Z] &= \int_{-\infty}^{+\infty} z f_Z(z) dz = \int_{-\infty}^{+\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= 0 \end{aligned}$$

Quindi, il valore atteso di una variabile aleatoria continua con distribuzione normale standard è:

$$E[Z] = 0$$

Per calcolare la varianza, utilizziamo la formula:

$$\operatorname{Var}(Z) = E[Z^2] - (E[Z])^2$$

Calcoliamo prima $E[Z^2]$:

$$\begin{aligned} E[Z^2] &= \int_{-\infty}^{+\infty} z^2 f_Z(z) dz = \int_{-\infty}^{+\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\ &= 1 \end{aligned}$$

Quindi, la varianza è:

$$\operatorname{Var}(Z) = E[Z^2] - (E[Z])^2 = 1 - 0 = 1$$



Adesso vengono qui riportate alcune proprietà della distribuzione normale

Proposition 6.10.10

Se $X \sim \mathcal{N}(\mu, \sigma^2)$, allora la trasformazione

$$Z = \frac{X - \mu}{\sigma}$$

è una variabile aleatoria standardizzata, ovvero $Z \sim \mathcal{N}(0, 1)$

Dimostrazione: Sia $X \sim \mathcal{N}(\mu, \sigma^2)$, quindi la funzione di densità di probabilità di X è:

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

Devo dimostrare che $Z = \frac{X-\mu}{\sigma}$ ha una distribuzione normale standard, quindi:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Si ha che la funzione di ripartizione di Z è:

$$F_Z(z) = P(Z \leq z) = P\left(\frac{X - \mu}{\sigma} \leq z\right) = P(X \leq z\sigma + \mu) = F_X(z\sigma + \mu)$$

Dove $F_X(x)$ è la funzione di ripartizione di X .

Per trovare la funzione di densità di probabilità di Z , deriviamo $F_Z(z)$ rispetto a z :

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} F_X(z\sigma + \mu) = f_X(z\sigma + \mu) \cdot \sigma$$

Sostituendo $f_X(x)$, otteniamo:

$$f_Z(z) = f_X(z\sigma + \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z\sigma + \mu - \mu)^2}{2\sigma^2}} \cdot \sigma = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Questo dimostra che Z ha una distribuzione normale standard



Chapter 7

Integrale di Gauss

perché sì

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

Chapter 8

Vettori Aleatori

8.1 Introduzione

Si introduce tale concetto di *vettori aleatori* per trattare variabili aleatorie che non sono più scalari, ma vettori. Infatti ciò che è stato presentato fino ad ora, dato uno spazio di probabilità (Ω, \mathbf{P}) , sono esempi con una sola variabile aleatoria, che assume valori in \mathbb{R} , tuttavia, è possibile definire più variabili aleatorie in un qualsiasi spazio di probabilità, questo ci induce, così, a considerare la tupla (X, X_1, \dots, X_n) definita su Ω e che assume valori in \mathbb{R}^n (in base alla quantità di variabili aleatorie). Questa tupla è detta *vettore aleatorio*

Definition 8.1.1: Vettore aleatorio

Sia (Ω, \mathbf{P}) uno spazio di probabilità, una qualunque funzione

$$(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n \quad (8.1)$$

è definita **vettore aleatorio**

8.1.1 Legge di un vettore aleatorio

Come per le variabili aleatorie, possiamo associare ad ogni vettore aleatorio la sua distribuzione o legge. Diamo la definizione solo per il caso bidimensionale.

Definition 8.1.2: Legge di un vettore aleatorio

Sia (Ω, \mathbf{P}) uno spazio di probabilità $(X, Y) : \Omega \rightarrow \mathbb{R}^2$. Si definisce **legge di un vettore aleatorio** la probabilità

$$\mathbf{P}_{(X,Y)} = \mathcal{P}(\mathbb{R}^2) \rightarrow [0, 1] \quad (8.2)$$

definita da

$$\mathbf{P}_{(X,Y)}(B) = \mathbf{P}((X, Y) \in B) \quad \forall B \subset \mathbb{R}^2 \quad (8.3)$$

Per dire che (X, Y) ha distribuzione o legge $\mathbf{P}_{(X,Y)}$ si scrive

$$(X, Y) \sim \mathbf{P}_{(X,Y)}$$

Note:

Se $B \subset \mathbb{R}^2$ è il prodotto cartesiano di due insiemi B_1 e B_2 in \mathbb{R} , allora

$$\mathbf{P}((X, Y) \in B_1 \times B_2) = \mathbf{P}(\{X \in B_1\} \cap \{Y \in B_2\})$$

Infatti l'evento $\{(X, Y) \in B_1 \times B_2\}$ è dato da

$$\{(X, Y) \in B_1 \times B_2\} = \{X \in B_1\} \cap \{Y \in B_2\}$$

$\mathbf{P}(\{X \in B_1\} \cap \{Y \in B_2\})$ verrà scritto nella forma nella forma $\mathbf{P}(X \in B_1, Y \in B_2)$.

8.1.2 Indipendenza di variabili aleatorie

Si può estendere il concetto di indipendenza di eventi a quello di indipendenza di variabili aleatorie. Intuitivamente due variabili aleatorie X e Y sono indipendenti se la conoscenza del valore di X non fornisce informazioni sul valore di Y e viceversa. Matematicamente n variabili aleatorie X_1, \dots, X_n sono indipendenti se gli eventi da esse generati, ovvero

$$\{X_1 \in B_1\} \dots \{X_n \in B_n\} \text{ al variare di tutti i sottoinsiemi } B_1, \dots, B_n \text{ in } \mathbb{R}$$

sono indipendenti

Definition 8.1.3: Variabili aleatorie indipendenti

Sia (Ω, \mathbf{P}) uno spazio di probabilità e X_1, \dots, X_n variabili aleatorie definite su Ω . Si dice che X_1, \dots, X_n sono **indipendenti** se

$$\mathbf{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^n \mathbf{P}(X_i \in B_i) \quad \forall B_1, \dots, B_n \subset \mathbb{R} \quad (8.4)$$

Inoltre la notazione per indicare due indipendenti X e Y è

$$X \perp\!\!\!\perp Y$$

Note:

Si dice anche che due variabili aleatorie sono indipendenti se la distribuzione congiunta si fattorizza nel prodotto delle marginali

Concludiamo questa sezione con il seguente risultato, in cui si afferma che funzioni di variabili aleatorie indipendenti sono indipendenti

Proposition 8.1.1 indipendenza di funzioni di variabili aleatorie

Siano X e Y due variabili aleatorie indipendenti e siano $f : \mathbb{R} \rightarrow \mathbb{R}$ e $g : \mathbb{R} \rightarrow \mathbb{R}$ due funzioni arbitrarie. Allora le variabili aleatorie $f(X)$ e $g(Y)$ sono indipendenti

Dimostrazione: Siano $B_1, B_2 \subset \mathbb{R}$ due sottoinsiemi arbitrari. Si deve dimostrare che

$$\mathbf{P}(f(X) \in B_1, g(Y) \in B_2) = \mathbf{P}(f(X) \in B_1) \cdot \mathbf{P}(g(Y) \in B_2)$$

Siano

$$\begin{aligned} f^{-1}(B_1) &= \{x \in \mathbb{R} : f(x) \in B_1\} \\ g^{-1}(B_2) &= \{y \in \mathbb{R} : g(y) \in B_2\} \end{aligned}$$

Le controimmagini di B_1 e B_2 tramite f e g , rispettivamente. Allora

$$\{f(X) \in B_1\} = \{X \in f^{-1}(B_1)\} \quad \text{e} \quad \{g(Y) \in B_2\} = \{Y \in g^{-1}(B_2)\}$$

Quindi

$$\begin{aligned} \mathbf{P}(f(X) \in B_1, g(Y) \in B_2) &= \mathbf{P}(\{X \in f^{-1}(B_1)\} \cap \{Y \in g^{-1}(B_2)\}) \\ &= \mathbf{P}(X \in f^{-1}(B_1)) \cdot \mathbf{P}(Y \in g^{-1}(B_2)) \\ &= \mathbf{P}(f(X) \in B_1) \cdot \mathbf{P}(g(Y) \in B_2) \end{aligned}$$



Note:

Per 8.1.2 si deduce che se X e Y sono indipendenti allora non può esistere alcuna dipendenza funzionale tra X e Y , tranne nel caso in cui almeno una tra X e Y sia una costante.

In altre parole, non può esistere alcuna funzione $f : \mathbb{R} \rightarrow \mathbb{R}$ tale che

$$Y = f(X)$$

8.2 Vettori aleatori discreti

In questa sezione studiamo una particolare classe di vettori aleatori, i vettori aleatori discreti (bidimensionali)

Definition 8.2.1: vettore aleatorio discreto

Sia (Ω, \mathbb{P}) uno spazio di probabilità e (X, Y) un vettore aleatorio. Si dice che (X, Y) è un **vettore aleatorio discreto** se sia X che Y sono variabili aleatorie discrete

Si noti che il vettore (X, Y) assume un numero finito o al più numerabile di valori, dati al più da tutte le coppie $\mathcal{S}_X \times \mathcal{S}_Y$. Quindi il supporto di (X, Y) è un sottoinsieme di $\mathcal{S}_X \times \mathcal{S}_Y$. Adesso presento la densità discreta

Definition 8.2.2: Densità discreta congiunta

Sia (Ω, \mathbb{P}) uno spazio di probabilità e (X, Y) un vettore aleatorio discreto. La funzione $p_{(X,Y)} : \mathbb{R}^2 \rightarrow [0, 1]$ definita da

$$p_{(X,Y)}(x, y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}((X, Y) = (x, y)) \quad \forall (x, y) \in \mathcal{S}_X \times \mathcal{S}_Y$$

è detta **densità discreta congiunta** di X e Y

Si noti che $p_{(X,Y)}$ è la probabilità che il vettore aleatorio (X, Y) assuma il valore (x, y) , quindi $p_{(X,Y)}(x, y)$ verifica le seguenti disuguaglianze

$$0 \leq p_{(X,Y)}(x, y) \leq 1 \quad \forall (x, y) \in \mathcal{S}_X \times \mathcal{S}_Y$$

Theorem 8.2.1 sulla legge congiunta di un vettore aleatorio discreto

Sia (Ω, \mathbb{P}) uno spazio di probabilità e (X, Y) un vettore aleatorio discreto. Sia \mathcal{S}_X e \mathcal{S}_Y i supporti di X e Y , rispettivamente. Allora:

- $p_{(X,Y)}(x, y) = 0 \quad \forall (x, y) \notin \mathcal{S}_X \times \mathcal{S}_Y$
- $\sum_{x \in \mathcal{S}_X} \sum_{y \in \mathcal{S}_Y} p_{(X,Y)}(x, y) = 1$
- Vale la seguente formula

$$\mathbb{P}((X, Y) \in B) = \sum_{(x,y) \in B} p_{(X,Y)}(x, y) \quad \forall B \subset \mathcal{S}_X \times \mathcal{S}_Y \quad (8.5)$$

8.2.1 Densità discreta congiunta e densità discrete marginali

In questa sezione si indaga la relazione tra la densità congiunta $p_{(X,Y)}$ e le marginali p_X e p_Y . Ci si ricordi che

$$\begin{aligned} p_X(x) &= \mathbb{P}(X = x) \\ p_Y(y) &= \mathbb{P}(Y = y) \\ p_{(X,Y)}(x, y) &= \mathbb{P}(X = x, Y = y) \end{aligned}$$

Dalla formula delle probabilità totali otteniamo il seguente risultato

Theorem 8.2.2

Sia (X, Y) un vettore aleatorio discreto. Allora

$$p_X(x) = \sum_{y \in \mathcal{S}_Y} p_{(X,Y)}(x, y) \quad \forall x \in \mathcal{S}_X$$

$$p_Y(y) = \sum_{x \in \mathcal{S}_X} p_{(X,Y)}(x, y) \quad \forall y \in \mathcal{S}_Y$$

Dimostrazione: Dimostriamo solo la prima formula nel caso in cui \mathcal{S}_Y è un insieme finito: $\mathcal{S}_Y = \{y_1, \dots, y_m\}$. Quindi dobbiamo dimostrare che, per ogni $x_i \in \mathcal{S}_X$ fissato, vale

$$p_X(x_i) = \sum_{j=1}^m p_{(X,Y)}(x_i, y_j).$$

Riscritta in termini di \mathbf{P} diventa

$$\mathbf{P}(X = x_i) = \sum_{j=1}^m \mathbf{P}(X = x_i, Y = y_j)$$

Poniamo

$$A = \{X = x_i\}, \quad B_j = \{Y = y_j\}, \quad \forall j = 1, \dots, m.$$

Gli eventi B_1, \dots, B_m sono una partizione di Ω . Quindi, dalla formula delle probabilità totali abbiamo che

$$\mathbf{P}(A) = \sum_{j=1}^m \mathbf{P}(A \cap B_j),$$

che corrisponde all'uguaglianza.



Tabella della densità discreta congiunta

Nel caso in cui sia \mathcal{S}_X che \mathcal{S}_Y siano finiti, quindi

$$\mathcal{S}_X = \{x_1, \dots, x_n\} \quad \text{e} \quad \mathcal{S}_Y = \{y_1, \dots, y_m\}$$

possiamo riportare i valori di $p_{(X,Y)}$ in una tabella:

$X \backslash Y$	y_1	y_2	\dots	y_m	p_X
x_1	$p_{(X,Y)}(x_1, y_1)$	$p_{(X,Y)}(x_1, y_2)$	\dots	$p_{(X,Y)}(x_1, y_m)$	$p_X(x_1)$
x_2	$p_{(X,Y)}(x_2, y_1)$	$p_{(X,Y)}(x_2, y_2)$	\dots	$p_{(X,Y)}(x_2, y_m)$	$p_X(x_2)$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
x_n	$p_{(X,Y)}(x_n, y_1)$	$p_{(X,Y)}(x_n, y_2)$	\dots	$p_{(X,Y)}(x_n, y_m)$	$p_X(x_n)$
p_Y	$p_Y(y_1)$	$p_Y(y_2)$	\dots	$p_Y(y_m)$	1

Per il 8.2.1 si ha che i valori di p_X si ottengono sommando i valori di $p_{(X,Y)}$ lungo le righe, mentre i valori di p_Y si ottengono sommando i valori di $p_{(X,Y)}$ lungo le colonne. Infine, sommando i valori dell'ultima colonna (quindi i valori di p_X) si ottiene 1. Analogamente, sommando i valori dell'ultima riga (quindi i valori di p_Y) si ottiene ancora 1. Questo spiega la presenza del numero 1 nell'angolo in basso a destra della tabella.

8.2.2 Indipendenza e Densità discreta congiunta

Conoscendo la densità discreta congiunta $p_{(X,Y)}$ di un vettore aleatorio discreto è possibile ricostruire le marginali p_X e p_Y . Non è possibile viceversa

Example 8.2.1 (Esercizio sulla densità discreta congiunta)

Si lanciano due monete e si considerano le seguenti variabili aleatorie

X = “vale 1 se l’esito del lancio della prima moneta è testa, vale 0 altrimenti”

Y = “vale 1 se l’esito del lancio della seconda moneta è testa, vale 0 altrimenti”

Se si considera il vettore aleatorio (X, Y) , allora

$X \backslash Y$	0	1	p_X
0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
p_Y	$\frac{1}{2}$	$\frac{1}{2}$	1

Se si considera il vettore aleatorio (X, X) , allora

$X \backslash X$	0	1	p_X
0	$\frac{1}{2}$	0	$\frac{1}{2}$
1	0	$\frac{1}{2}$	$\frac{1}{2}$
p_X	$\frac{1}{2}$	$\frac{1}{2}$	1

Adesso si osservi il seguente teorema

Theorem 8.2.3 Indipendenza di variabili aleatorie discrete

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

$$p_{(X,Y)}(x, y) = p_X(x)p_Y(y) \quad \forall (x, y) \in \mathcal{S}_X \times \mathcal{S}_Y \quad (8.6)$$

Dimostrazione: • (\Leftarrow) si deve dimostrare che se vale 8.2.2 allora X e Y sono indipendenti.

Siano B_1 e B_2 due sottoinsiemi arbitrari di \mathbb{R} . Si deve dimostrare che

$$\mathbb{P}(X \in B_1, Y \in B_2) = \mathbb{P}(X \in B_1) \cdot \mathbb{P}(Y \in B_2)$$

Dato che $\mathbb{P}(X \in B_1, Y \in B_2) = \mathbb{P}((X, Y) \in B_1 \times B_2)$ per la 8.5 si ha che

$$\mathbb{P}(X \in B_1, Y \in B_2) = \sum_{(x,y) \in B_1 \times B_2} p_{(X,Y)}(x, y) = \sum_{x \in B_1} \sum_{y \in B_2} p_{(X,Y)}(x, y)$$

Per si ha che

$$\sum_{x \in B_1} \sum_{y \in B_2} p_{(X,Y)}(x, y) = \sum_{x \in B_1} \sum_{y \in B_2} p_X(x)p_Y(y)$$

Eseguendo ulteriori passaggi si ha che

$$\sum_{x \in B_1} \sum_{y \in B_2} p_X(x)p_Y(y) = \sum_{x \in B_1} p_X(x) \cdot \sum_{y \in B_2} p_Y(y) = \mathbb{P}(X \in B_1) \cdot \mathbb{P}(Y \in B_2)$$

Quindi

$$\mathbb{P}(X \in B_1, Y \in B_2) = \mathbb{P}(X \in B_1) \cdot \mathbb{P}(Y \in B_2)$$

Dimostrato

- (\Rightarrow) si deve dimostrare che se X e Y sono indipendenti allora vale 8.2.2. Siano $B_1 = \{x\}$ e $B_2 = \{y\}$ due sottoinsiemi arbitrari di \mathbb{R} per la definizione di indipendenza si ottiene l’uguaglianza 8.2.2

Dimostrato



Example 8.2.2

Siano X e Y due variabili aleatorie discrete con densità discreta congiunta

X \ Y	-1	5	10	p_X
0		0.12		0.4
5				
p_Y	0.3			1

1. Completare la tabella in modo che X e Y siano indipendenti.
2. Calcolare $P(X < Y)$.
3. Calcolare $P(|XY| \geq 5)$ e $P(X + Y > 5)$.
4. Siano $U = |XY|$ e $V = X + Y$. Trovare la densità discreta congiunta di U e V e le densità marginali

Soluzione: Sia data al bonzo



8.2.3 Valore atteso e varianza di una funzione di (X, Y)

Nel seguito capiterà spesso di dover calcolare valore atteso e varianza di una funzione di (X, Y) :

$$h(X, Y)$$

Risulta dunque particolarmente utile il seguente risultato

Theorem 8.2.4 sul valore atteso e varianza di una funzione di un vettore aleatorio discreto

Siano (X, Y) un vettore aleatorio discreto e $h : \mathbb{R}^2 \rightarrow \mathbb{R}$. Allora

$$\mathbb{E}[h(X, Y)] = \sum_{(x,y) \in \mathcal{S}_X \times \mathcal{S}_Y} h(x, y) p_{(X,Y)}(x, y)$$

e

$$\text{Var}(h(X, Y)) = \mathbb{E}[(h(X, Y) - \mathbb{E}[h(X, Y)])^2] = \sum_{(x,y) \in \mathcal{S}_X \times \mathcal{S}_Y} (h(x, y) - \mathbb{E}[h(X, Y)])^2 p_{(X,Y)}(x, y)$$

Vale inoltre la formula

$$\text{Var}(h(X, Y)) = \mathbb{E}[h(X, Y)^2] - (\mathbb{E}[h(X, Y)])^2 = \sum_{(x,y) \in \mathcal{S}_X \times \mathcal{S}_Y} h(x, y)^2 p_{(X,Y)}(x, y) - (\mathbb{E}[h(X, Y)])^2$$

Corollary 8.2.1 somma di X e Y

Siano X e Y due variabili aleatorie discrete. Siano inoltre a e b due costanti. Allora

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

Dimostrazione: Per il teorema 8.2.3 si ha che con $h(x, y) = ax + by$ si ottiene

$$\begin{aligned}
 \mathbb{E}[aX + bY] &= \sum_{(x,y) \in \mathcal{S}_X \times \mathcal{S}_Y} (ax + by)p_{(X,Y)}(x, y) \\
 &= \sum_{(x,y) \in \mathcal{S}_X \times \mathcal{S}_Y} axp_{(X,Y)}(x, y) + \sum_{(x,y) \in \mathcal{S}_X \times \mathcal{S}_Y} byp_{(X,Y)}(x, y) \\
 &= a \sum_{(x,y) \in \mathcal{S}_X \times \mathcal{S}_Y} xp_{(X,Y)}(x, y) + b \sum_{(x,y) \in \mathcal{S}_X \times \mathcal{S}_Y} yp_{(X,Y)}(x, y) \\
 &= a \sum_{x \in \mathcal{S}_X} x \sum_{y \in \mathcal{S}_Y} p_{(X,Y)}(x, y) + b \sum_{y \in \mathcal{S}_Y} y \sum_{x \in \mathcal{S}_X} p_{(X,Y)}(x, y) \\
 &= a \sum_{x \in \mathcal{S}_X} xp_X(x) + b \sum_{y \in \mathcal{S}_Y} yp_Y(y) \\
 &= a\mathbb{E}[X] + b\mathbb{E}[Y]
 \end{aligned}$$



Corollary 8.2.2 Prodotto X e Y

Siano X e Y due variabili aleatorie discrete. Se X e Y sono indipendenti, allora

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

Note:

NON VALE IL CONTRARIO: se $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ non è necessariamente vero che X e Y sono indipendenti

Dimostrazione: Per il teorema 8.2.3 si ha che con $h(x, y) = xy$ si ottiene

$$\begin{aligned}
 \mathbb{E}[XY] &= \sum_{(x,y) \in \mathcal{S}_X \times \mathcal{S}_Y} xyp_{(X,Y)}(x, y) \\
 &\stackrel{\uparrow}{=} \sum_{(x,y) \in \mathcal{S}_X \times \mathcal{S}_Y} xyp_X(x)p_Y(y) \\
 &= \sum_{x \in \mathcal{S}_X} xp_X(x) \sum_{y \in \mathcal{S}_Y} yp_Y(y) \\
 &= \mathbb{E}[X] \cdot \mathbb{E}[Y]
 \end{aligned}$$



8.2.4 Indici di sintesi della distribuzione di un vettore aleatorio discreto

In termini intuitivi, la covarianza è una misura di come due variabili aleatorie (chiamiamole X e Y) tendono a variare insieme. Ti dice se, in media, quando una variabile assume valori sopra la sua media, l'altra tende a fare lo stesso, oppure se tende a fare il contrario.

In altre parole, la covarianza misura la direzione della relazione lineare tra due variabili aleatorie. Se la covarianza è positiva, significa che quando una variabile aumenta, l'altra tende ad aumentare. Se è negativa, significa che quando una variabile aumenta, l'altra tende a diminuire. Se è zero, non c'è una relazione lineare tra le due variabili.

Definition 8.2.3: Covarianza

Siano X e Y due variabili aleatorie discrete. La **covarianza** di X e Y è definita come

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \sum_{(x,y) \in \mathcal{S}_X \times \mathcal{S}_Y} (x - \mathbb{E}[X])(y - \mathbb{E}[Y])p_{(X,Y)}(x, y)$$

Se $Cov(X, Y) = 0$ si dice che X e Y sono **scorrelate**

Note:

Si noti che la covarianza di X e Y è definita come il valore atteso della variabile aleatoria $h(X, Y)$, dove

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

Quindi la seconda uguaglianza in 8.2.4 è una conseguenza del teorema 8.2.3W

Note:

Si noti che

$$\text{Cov}(X, X) = \text{Var}(X)$$

Note:

La covarianza è simmetrica, ovvero

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)$$

Per calcolare la covarianza di X e Y è utile la seguente formula

Theorem 8.2.5 Formula per la covarianza

Siano X e Y due variabili aleatorie discrete. Allora

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \sum_{(x,y) \in \mathcal{S}_X \times \mathcal{S}_Y} xyp_{(X,Y)}(x, y) - \mathbb{E}[X]\mathbb{E}[Y]$$

Note:

Se le variabili aleatorie X e Y sono indipendenti, allora sono scorrelate

Dimostrazione: Per definizione, si ha che

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY - X\mathbb{E}[Y] - \mathbb{E}[X]Y + \mathbb{E}[X]\mathbb{E}[Y]] \end{aligned}$$

Dalla linearità del valore atteso, si ottiene (si noti che $\mathbb{E}[X]$ e $\mathbb{E}[Y]$ sono costanti)

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \end{aligned}$$

Infine se le variabili aleatorie X e Y sono indipendenti allora dal Corollario 2.2 si ha che

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

quindi $\text{Cov}(X, Y) = 0$, ovvero X e Y sono scorrelate



La covarianza interviene nella formula della varianza di $X + Y$

Theorem 8.2.6 Varianza della somma di due variabili aleatorie discrete

Siano X e Y due variabili aleatorie discrete. Allora

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

Se X e Y sono scorrelate, allora

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

Dimostrazione: Applicando il teorema 8.2.3 con $h(x, y) = x + y$ si ha che

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E}[(X + Y - \mathbb{E}[X + Y])^2] = \mathbb{E}[(X + Y - \mathbb{E}[X] - \mathbb{E}[Y])^2] \\ &= \mathbb{E}[(X - \mathbb{E}[X])^2 + (Y - \mathbb{E}[Y])^2 + 2(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y). \end{aligned}$$



Note:

La covarianza è un indicatore di dipendenza tra due variabili aleatorie X e Y . Più precisamente, supponiamo³ che $\text{Var}(X) > 0$ e $\text{Var}(Y) > 0$. In tal caso, ha senso definire il **coefficiente di correlazione**

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

Si può dimostrare che

$$-1 \leq \rho_{X,Y} \leq 1.$$

Ricordiamo che se $\text{Cov}(X,Y) = 0$ allora X e Y si dicono scorrelate. Si noti che $\text{Cov}(X,Y) = 0$ equivale a $\rho_{X,Y} = 0$. Al contrario, quando la correlazione è massima in valore assoluto (quindi $\rho_{X,Y} = -1$ oppure $\rho_{X,Y} = 1$), si ha che

$$\rho_{X,Y} = \pm 1 \iff Y = aX + b.$$

Più precisamente, $\rho_{X,Y} = \pm 1$ se e solo se esistono due costanti $a \neq 0$ e $b \in \mathbb{R}$ tali che $Y = aX + b$. La correlazione misura dunque se esiste tra X e Y una dipendenza di tipo lineare. Quindi quando X e Y sono scorrelate ($\text{Cov}(X,Y) = 0$) significa solamente che non esiste una dipendenza lineare tra X e Y . Ricordiamo invece che se X e Y sono indipendenti allora non esiste alcuna dipendenza funzionale tra X e Y (non solo di tipo lineare). Perciò se sappiamo solamente che $\text{Cov}(X,Y) = 0$ non possiamo dire che X e Y sono indipendenti. Riassumendo:

$$\begin{aligned} X \perp\!\!\!\perp Y &\implies \text{Cov}(X,Y) = 0, \\ \text{invece } X \not\perp\!\!\!\perp Y &\not\implies \text{Cov}(X,Y) = 0. \end{aligned}$$

Chapter 9

Catene di Markov

Lavoriamo ancora con successioni di VA - ovvero un processo stocastico (che puo' essere a tempo discreto o continuo, dato che l'interpretazione tipica e' che gli esperimenti vengono fatti in successione nel tempo) $(X_n)_{n \in \mathbb{N}}$ sono VA discrete:

- $X_1 \rightarrow S_{X_1}$
- $X_1 \rightarrow S_{X_1}$
- $X_1 \rightarrow S_{X_1}$
- $X_1 \rightarrow S_{X_1}$

Dove tutti i supporti sono finiti o infiniti numerabili tali per cui $\exists S$ supporto finito o infinito numerabile che:

$$S_{X_1} \subseteq S, \dots, S_{X_i} \subseteq S$$

Definition 9.0.1

$(X_n)_{n \in \mathbb{N}}$ di VA discrete e' detta *catena di Markov* (a tempo discreto) se:

- $\exists S$ finito o inf. numerabile che contiene tutti i supporti
- Vale la proprieta' di Markov:

$$\forall i, j, i$$

la i e' il valore presente mentre la j e' il valore futuro, in pratica il valore del prossimo stato e' condizionato solo dall'ultimo valore e non da quelli prima

in realta noi lavoreremo nel caso di S finito, questa e' un'ipotesi innocua perche' e' facile passare alla versione infinita

la seconda non e' innocua, ci dice che ci sara' solo una matrice di transizione e il tempo non gioca nessun ruolo, quindi sapendo il valore attuale sappiamo le probabilita' di trovarci negli altri stati possibili il prossimo tempo la somma delle righe della matrice deve fare 1 dato che sono tutte le probabilita'

ci possiamo porre la domanda: ora mi trovo in un punto, qual'e' la probabilita' che fra x passi mi trovi in un altro specifico punto

abbiamo visto poi un paio di esempi con grafi

che cosa si aspetta di vedere negli esercizi d'esame: formule! non numeri a caso. Sulle catene di Markov possiamo lavorare sul grafo (anche per determinare le classi di comunicazione) ma bisogna che sul grafo ci sia un segno che fa vedere il ragionamento (e' cosi in generale per tutto il compito)

classifichiamo gli stati in base al livello di accessibilita'.

Legge delle VA data la matrice delle probabilita' di transizione

queste dipendono dal tempo e dalla distribuzione dello stato di partenza, anche se in realta' non deve essere per forza il primo stato, basta uno qualunque perche' tanto ci interessa solo dello stato finale. Quindi se viene dato il penultimo stato ci basta guardare un solo passaggio (e' meglio!)

, anche se in realta' non deve essere per forza il primo stato, basta uno qualunque perche' tanto ci interessa solo dello stato finale. Quindi se viene dato il penultimo stato ci basta guardare un solo passaggio (e' meglio!)
, anche se in realta' non deve essere per forza il primo stato, basta uno qualunque perche' tanto ci interessa solo dello stato finale. Quindi se viene dato il penultimo stato ci basta guardare un solo passaggio (e' meglio!)

Chapter 10

Esercitazione

10.1 Probabilità Condizionata

10.1.1 Esercizio 4 (Foglio 2)

Testo

Un'urna contiene r palline rosse e b palline bianche. Si eseguono due estrazioni senza reimmissione.

- (a) Determinare uno spazio di probabilità che descriva l'esperimento aleatorio.
- (b) Calcolare la probabilità che la prima pallina estratta sia rossa.
- (c) Calcolare la probabilità che la prima pallina sia rossa e la seconda bianca.
- (d) Calcolare la probabilità che le due palline abbiano colori diversi.
- (e) Calcolare la probabilità che la seconda pallina estratta sia rossa.

Soluzione

Siano:

R_i = "ho estratto una pallina rossa all' i -esima iterazione" B_i = "ho estratto una pallina bianca all' i -esima iterazione"

Definiamo lo spazio degli esiti. Poiché le estrazioni avvengono senza reimmissione, ogni esito è una coppia ordinata di palline diverse. Denotiamo:

$$\Omega = \{(p_1, p_2) : p_1, p_2 \text{ sono palline dell'urna e } p_1 \neq p_2\}$$

La cardinalità totale è:

$$|\Omega| = (r + b)(r + b - 1)$$

Infatti appena estraiamo una pallina l'urna conterrà $(r + b - 1)$ palline, la totalità di palline -1. Sfruttiamo il principio di probabilità uniforme, secondo cui ogni esito ha probabilità $1/|\Omega|$

(a) **Spazio di probabilità:**

Lo spazio di probabilità è (Ω, P) con

$$P(\{\omega\}) = \frac{1}{(r + b)(r + b - 1)} \quad \forall \omega \in \Omega.$$

(b) **Probabilità che la prima pallina sia rossa:**

Poiché ci sono r palline rosse su un totale di $r + b$, si ha:

$$P(R_1) = \frac{r}{r + b}.$$

(c) **Probabilità che la prima pallina sia rossa e la seconda bianca:**

Dato che la prima è rossa, nell'urna rimangono $r + b - 1$ palline, di cui b sono bianche:

$$P(R_1 \cap B_2) = P(B_2 | R_1)P(R_1) = \frac{r}{r+b} \cdot \frac{b}{r+b-1}.$$

(d) **Probabilità che le due palline abbiano colori diversi:**

Questa condizione si verifica in due modi: rossa poi bianca oppure bianca poi rossa. Quindi:

$$\begin{aligned} P((R_1 \cap B_2) \cup (B_1 \cap R_2)) &= P(B_2 | R_1)P(R_1) + P(R_2 | B_1)P(B_1) \\ &= \frac{r}{r+b} \cdot \frac{b}{r+b-1} + \frac{b}{r+b} \cdot \frac{r}{r+b-1} \\ &= \frac{2rb}{(r+b)(r+b-1)}. \end{aligned}$$

(e) **Probabilità che la seconda pallina sia rossa:**

Usiamo la formula della probabilità totale, considerando le possibili estrazioni del primo turno:

$$\begin{aligned} P(R_2) &= P(B_2 | R_1) \cdot P(R_1) + P(B_2 | B_1) \cdot P(B_1) \\ &= \frac{r-1}{r+b-1} \cdot \frac{r}{r+b} + \frac{r}{r+b-1} \cdot \frac{b}{r+b} \end{aligned}$$

Semplificando:

$$P(R_2) = \frac{r(r-1) + rb}{(r+b)(r+b-1)} = \frac{r^2 - r + rb}{(r+b)(r+b-1)} = \frac{r(r+b-1)}{(r+b)(r+b-1)} = \frac{r}{r+b}$$

Esercizio 6

Testo

Supponiamo che un'urna contenga 1 pallina rossa e 1 pallina bianca. Una pallina viene estratta e se ne osserva il colore. La pallina estratta viene poi rimessa nell'urna insieme a un'altra pallina dello stesso colore (estrazione con rinforzo). Siano

R_i = evento che all' i -esima estrazione venga estratta una pallina rossa,

B_i = evento che all' i -esima estrazione venga estratta una pallina bianca.

Si calcolino:

(1) $P(R_2)$

- (2) Sapendo che la seconda pallina estratta è rossa, quale è l'evento più probabile per la prima estrazione: che la pallina estratta sia stata rossa oppure bianca?

Soluzione

(1) **Calcolo di $P(R_2)$:**

Caso 1: Se alla prima estrazione esce una pallina rossa (evento R_1):

- La probabilità di estrarre una rossa al primo turno è $P(R_1) = \frac{1}{2}$.
- Dopo l'estrazione, la pallina rossa viene rimessa insieme a un'altra rossa, dunque l'urna contiene 2 rosse e 1 bianca. Quindi:

$$P(R_2 | R_1) = \frac{2}{3}.$$

Caso 2: Se alla prima estrazione esce una pallina bianca (evento B_1):

- $P(B_1) = \frac{1}{2}$.

- Dopo il rinforzo, l'urna contiene 1 rossa e 2 bianche, dunque:

$$P(R_2 | B_1) = \frac{1}{3}.$$

Applicando la formula della probabilità totale:

$$\begin{aligned} P(R_2) &= P(R_2 | R_1) \cdot P(R_1) + P(R_2 | B_1) \cdot P(B_1) \\ &= \frac{2}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} \\ &= \frac{1}{3} + \frac{1}{6} = \frac{1}{2}. \end{aligned}$$

(2) **Confronto tra $P(R_1 | R_2)$ e $P(B_1 | R_2)$:**

Innanzitutto si tenga conto il teorema di Bayes:

Siano A, B due eventi, t.c. $P(A), P(B) > 0$, allora $P(A | B) = \frac{P(B|A)P(A)}{P(B)}$

Dimostrazione: Abbiamo $P(A \cap B) = P(B \cap A)$, quindi $P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$

Usiamo il teorema di Bayes per calcolare $P(R_1 | R_2)$:

$$P(R_1 | R_2) = \frac{P(R_2 | R_1) P(R_1)}{P(R_2)} = \frac{\frac{2}{3} \cdot \frac{1}{2}}{\frac{1}{2}} = \frac{2}{3}.$$

Poiché $P(B_1 | R_2) = 1 - P(B_1^c | R_2) = 1 - P(R_1 | R_2) = 1 - \frac{2}{3} = \frac{1}{3}$, risulta che,

$$P(R_1 | R_2) > P(B_1 | R_2).$$

Quindi, sapendo che la seconda pallina è rossa, è più probabile che la prima pallina estratta fosse rossa.

10.1.2 Esercizio 7

Testo

Si consideri una popolazione in cui una persona su 100 abbia una certa malattia. Un test è disponibile per diagnosticare tale malattia. Si supponga che il test non sia perfetto, in quanto esso risulta positivo (ovvero indica la presenza della malattia) nel 5% dei casi quando è effettuato su persone sane, mentre risulta negativo (indicando l'assenza della malattia) nel 2% dei casi quando è effettuato su persone malate. Si calcolino le probabilità che:

- il test risulti positivo quando effettuato su una persona malata,
- il test risulti positivo,
- una persona sia malata se il test risulta positivo.

Soluzione

I dati del problema sono:

- $P(M) = 0.01$: probabilità che una persona sia malata.
- $P(S) = 0.99$: probabilità che una persona sia sana.
- $P(T^+ | M) = 0.98$: probabilità che il test risulti positivo se la persona è malata.
- $P(T^- | M) = 0.02$: probabilità che il test risulti negativo se la persona è malata.
- $P(T^+ | S) = 0.05$: probabilità che il test risulti positivo se la persona è sana.
- $P(T^- | S) = 0.95$: probabilità che il test risulti negativo se la persona è sana.

Costruiamo un diagramma di verità:

	Malato	Sano
Positivo	0.98	0.05
Negativo	0.02	0.95

1. **Probabilità che il test risulti positivo su una persona malata:** Questa probabilità è data direttamente dai dati:

$$P(T^+ | M) = 0.98.$$

2. **Probabilità che il test risulti positivo:** Utilizziamo la formula della probabilità totale:

$$\begin{aligned} P(T^+) &= P(T^+ | M) P(M) + P(T^+ | S) P(S) \\ &= 0.98 \cdot 0.01 + 0.05 \cdot 0.99 \\ &= 0.0098 + 0.0495 \\ &\approx 0.0593. \end{aligned}$$

3. **Probabilità che una persona sia malata, dato un test positivo:**

Applichiamo il teorema di Bayes:

$$\begin{aligned} P(M | T^+) &= \frac{P(T^+ | M) P(M)}{P(T^+)} \\ &= \frac{0.98 \cdot 0.01}{0.0593} \\ &\approx \frac{0.0098}{0.0593} \\ &\approx 0.165. \end{aligned}$$

Quindi, circa il 16,5% delle persone con test positivo sono effettivamente malate.

10.1.3 Esercizio 5

Testo

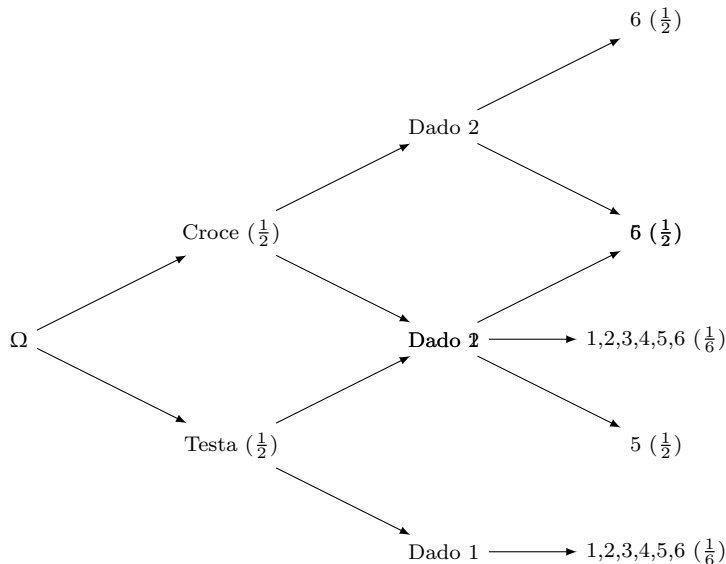
Si consideri il seguente esperimento:

ci sono quattro dadi: due non truccati e due truccati. I dadi truccati hanno tre facce con il numero 6 e tre facce con il numero 5. Si lancia una moneta (non truccata). Se viene *testa* si lanciano i due dadi non truccati, mentre se viene *croce* si lanciano i due dadi truccati.

Si richiede di calcolare:

- (a) La probabilità che la somma dei due dadi sia 11.
- (b) Sapendo di aver ottenuto 11 dalla somma dei due dadi, calcolare la probabilità che il lancio della moneta sia stato *croce*.

Soluzione



Sia:

$$T = \{\text{moneta: testa}\}, \quad C = \{\text{moneta: croce}\},$$

con $P(T) = P(C) = \frac{1}{2}$.

Caso 1: Dadi non truccati (se esce testa)

I dadi non truccati hanno facce 1, 2, 3, 4, 5, 6 con probabilità uniforme. La somma 11 si ottiene con le coppie (5, 6) e (6, 5). Quindi:

$$P(11 | T) = \frac{2}{36} = \frac{1}{18}.$$

Caso 2: Dadi truccati (se esce croce)

I dadi truccati assumono solo i valori 5 e 6 con probabilità $\frac{3}{6} = \frac{1}{2}$ ciascuno. La somma 11 si ottiene con le coppie (5, 6) e (6, 5), dunque:

$$P(11 | C) = 2 \cdot \left(\frac{1}{2} \cdot \frac{1}{2} \right) = \frac{1}{2}.$$

Calcolo della probabilità totale di ottenere 11:

$$\begin{aligned} P(11) &= P(11 | T) P(T) + P(11 | C) P(C) \\ &= \frac{1}{18} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \\ &= \frac{1}{36} + \frac{1}{4} = \frac{1}{36} + \frac{9}{36} = \frac{10}{36} = \frac{5}{18}. \end{aligned}$$

Calcolo della probabilità condizionata $P(C | 11)$: Utilizzando il teorema di Bayes:

$$P(C | 11) = \frac{P(11 | C) P(C)}{P(11)} = \frac{\frac{1}{2} \cdot \frac{1}{2}}{\frac{5}{18}} = \frac{\frac{1}{4}}{\frac{5}{18}} = \frac{1}{4} \cdot \frac{18}{5} = \frac{9}{10}.$$

Quindi, se la somma è 11, la probabilità che la moneta abbia dato *croce* è $\frac{9}{10}$.

10.1.4 Esercizio 2

Testo

Si consideri l'esperimento di lanciare due volte un dado.

- Determinare uno spazio di probabilità che descriva l'esperimento aleatorio.
- Si considerino i seguenti eventi:

- $A =$ “numero dispari sul primo dado”
- $B =$ “numero dispari sul secondo dado”
- $C =$ “la somma dei due risultati ‘e dispari”

Gli eventi A , B e C sono indipendenti?

(c) Si considerino ora gli eventi:

- $E =$ “il risultato del secondo lancio ‘e 1, 2 o 5”
- $F =$ “il risultato del secondo lancio ‘e 4, 5 o 6”
- $G =$ “la somma dei due risultati ‘e 9”

Gli eventi E , F e G sono indipendenti?

Soluzione

(a) $(\Omega, \mathbb{P}) = ?$:

- $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$
- I due sottoesperimenti sono indipendenti e hanno probabilita’ uniforme, quindi:

$$\begin{aligned}\mathbb{P}((x_1, x_2)) &= \mathbb{P}(x_1)\mathbb{P}(x_2) \\ &= \frac{1}{6} \frac{1}{6} \\ &= \frac{1}{36} \\ &= \frac{1}{|\Omega|}\end{aligned}$$

Quindi anche l’esperimento principale ha probabilita’ uniforme.

(b) Dimostriamo per controprova che A, B e C non sono indipendenti:

$$\begin{aligned}\mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A)\mathbb{P}(B|A)\mathbb{P}(C|A \cap B) \\ &= \frac{1}{2} \frac{1}{2} 0 \\ &= 0 \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \text{ (dato che e' sicuramente } > 0)\end{aligned}$$

(c) Dimostriamo per controprova che D, E e F non sono indipendenti;

$$\begin{aligned}\mathbb{P}(E|F) &= \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} \\ &= \frac{1}{3} \neq \mathbb{P}(E) (= \frac{1}{2})\end{aligned}$$

10.1.5 Esercizio 3

Testo

I componenti prodotti da una ditta possono avere due tipi di difetti con percentuali del 3% e del 7% rispettivamente e in modo indipendente l’uno dall’altro. Qual ‘e la probabilita’ che un componente scelto a caso

- presenti entrambi i difetti?
- sia difettoso?
- presenti il primo difetto, sapendo che ‘e difettoso?
- presenti uno solo dei difetti, sapendo che ‘e difettoso?

Soluzione

D_1 = "prodotto presenta il difetto 1"

D_2 = "prodotto presenta il difetto 2"

- (a) $\mathbb{P}(D_1 \cap D_2) = ?$ dato che non ci viene detto niente, possiamo ipotizzare che gli eventi sono indipendenti, quindi:

$$\begin{aligned}\mathbb{P}(D_1 \cap D_2) &= \mathbb{P}(D_1)\mathbb{P}(D_2) \\ &= \mathbb{P}(D_1)\mathbb{P}(D_2) \\ &= 0.03 \cdot 0.07\end{aligned}$$

- (b) $\mathbb{P}(D_1 \cup D_2) = ?$ proviamo a usare DeMorgan, tenendo in mente che i complementari di eventi indipendenti sono anch'essi indipendenti:

$$\begin{aligned}\mathbb{P}(D_1 \cup D_2) &= 1 - \mathbb{P}(D_1^c \cap D_2^c) \\ &= 1 - \mathbb{P}(D_1^c)\mathbb{P}(D_2^c) \\ &= 1 - 0.97 \cdot 0.93\end{aligned}$$

- (c) $\mathbb{P}(D_1 | D_1 \cup D_2) = ?$ possiamo usare Bayes e il risultato trovato nel punto prima:

$$\begin{aligned}\mathbb{P}(D_1 | D_1 \cup D_2) &= \frac{\mathbb{P}(D_1 \cup D_2 | D_1)\mathbb{P}(D_1)}{\mathbb{P}(D_1 \cup D_2)} \\ &= \frac{1 \cdot 0.03}{1 - 0.97 \cdot 0.93}\end{aligned}$$

- (d) $\mathbb{P}((D_1 \cap D_2^c) \cup (D_1^c \cap D_2) | D_1 \cup D_2) = ?$

$$\begin{aligned}\mathbb{P}((D_1 \cap D_2^c) \cup (D_1^c \cap D_2) | D_1 \cup D_2) &= \frac{\mathbb{P}(D_1 \cup D_2 | ((D_1 \cap D_2^c) \cup (D_1^c \cap D_2)))\mathbb{P}((D_1 \cap D_2^c) \cup (D_1^c \cap D_2))}{\mathbb{P}(D_1 \cup D_2)} \\ &= \frac{1 \cdot \mathbb{P}((D_1 \cap D_2^c) \cup (D_1^c \cap D_2))}{\mathbb{P}(D_1 \cup D_2)}\end{aligned}$$

Ci rimane quindi da calcolare la probabilità al numeratore. Notare che:

$$\begin{aligned}(D_1 \cap D_2^c) \cap (D_1^c \cap D_2) &= D_1 \cap D_2^c \cap D_1^c \cap D_2 \\ &= D_1 \cap D_1^c \cap D_2 \cap D_2^c \\ &= \emptyset\end{aligned}$$

Quindi, essendo eventi disgiunti possiamo applicare l'addittività finita:

$$\begin{aligned}\mathbb{P}((D_1 \cap D_2^c) \cup (D_1^c \cap D_2)) &= \mathbb{P}(D_1 \cap D_2^c) + \mathbb{P}(D_1^c \cap D_2) \\ &= \mathbb{P}(D_1)\mathbb{P}(D_2^c) + \mathbb{P}(D_1^c)\mathbb{P}(D_2)\end{aligned}$$

10.1.6 Esercizio Porco Rosso

Testo

L'aereo dei pirati del cielo, appena riparato, è stato dato alle fiamme. Porco Rosso vuole scoprire chi è stato. Durante le indagini si è scoperto che la settimana prima del delitto i pirati del cielo hanno detto al meccanico della ditta Piccolo che non gli avrebbero pagato la riparazione dell'idrovolante. Interrogato da Porco Rosso, Piccolo cerca di scagionarsi dicendo che a seguito di insolvenza solo 1 meccanico su 1000 si vendica. Porco Rosso però si accorge che questa stima non è più significativa: bisogna valutare la probabilità di vendetta sapendo che l'aereo è stato effettivamente dato alle fiamme. Porco Rosso allora considera questi eventi:

- A : dei clienti risultano insolventi contro il proprio meccanico,
- B : l'aereo di un cliente viene distrutto dal meccanico,

- C: l'aereo di un cliente viene distrutto ma non dal proprio meccanico.

A questo punto è necessario il vostro aiuto!

- Esprimere in funzione di A , B e C la probabilità fornita da Piccolo e calcolarla.
- Quali eventi tra A , B e C possono essere ritenuti disgiunti?
- Quali eventi tra A , B e C possono essere ritenuti indipendenti?
- Esprimere in funzione di A , B e C l'evento D : l'aereo di un cliente viene distrutto.
- Esprimere la probabilità condizionata $P(B|A \cap D)$ in funzione solo di $P(B|A)$ e $P(C)$. Porco Rosso non riesce a trovare $P(C)$, tuttavia trova che 1 aereo ogni 10000 viene distrutto.
- Limitare (dal basso o dall'alto) il valore di $P(B|A \cap D)$.
- È il caso che Porco Rosso continui ad indagare su Piccolo?

Soluzione

- $P(B|A) = \frac{1}{1000} = \frac{P(A \cap B)}{P(A)}$,
- $B \cap C = \emptyset$
- Quali tra A,B,C sono indipendenti? $P(A \cap B) = P(A) \cdot P(B)$, $P(A \cap C) = P(A) \cdot P(C)$
- $D = \text{"Aereo distrutto"} = B \cup C$
- $P(B|A \cap D)$ sono in funzione di $P(B|A)$ e $P(C)$, quindi $= \frac{B \cap A \cap D}{P(A \cap D)} = \frac{B \cap A \cap (B \cup C)}{P(A \cap (B \cup C))} = \frac{((B \cap A))}{P(A \cap (B \cup C))}$

10.1.7 Esercizio 2.3

Testo

Nel gioco del lotto si estraggono senza reimmissione cinque numeri da un'urna che contiene 90 palline da 1 a 9

- Determinare uno spazio di probabilità che descriva l'esperimento aleatorio
- Come cambia la risposta al punto precedente se le estrazioni avvengono con reimmissione?

Svolgimento

- $\Omega = \{(x_1, x_2, x_3, x_4, x_5) \mid x_i \in \{1, \dots, 90\}, i = 1, \dots, 5, x_i \text{ distinti}\}$
 - Definisco:
 $\forall i \in \{1, \dots, 5\} \forall j \in \{1, \dots, 90\}, E_{ij} : \text{"straggo il numero } j \text{ all'i-esima estrazione"}$
 $\mathbb{P}(E_{1,x_1} \cap E_{2,x_2} \cap E_{3,x_3} \cap E_{4,x_4} \cap E_{5,x_5}) = \mathbb{P}(E_{1,x_1})\mathbb{P}(E_{2,x_2}|E_{1,x_1}) \dots \mathbb{P}(E_{5,x_5}|E_{1,x_1} \cap E_{2,x_2} \cap E_{3,x_3} \cap E_{4,x_4} \cap E_{5,x_5})$
 assunto che vi è probabilità uniforme $= \frac{1}{90} \cdot \frac{1}{89} \cdot \frac{1}{88} \cdot \frac{1}{87} \cdot \frac{1}{86} = \frac{1}{|\Omega|}$
- $\Omega = \{(x_1, x_2, x_3, x_4, x_5) \mid x_i \in \{1, \dots, 90\}, i = 1, \dots, 5\}$
 sotto-esperimento: estrazione dall'urna, dato che avviene la reimmissione si ha che ogni sotto-esperimento è indipendente, quindi si ha:

$\mathbb{P}(\cdot)$ uniforme per $\Omega : \mathbb{P}(\omega)$

10.1.8

Testo

Un'urna contiene 10 palline di cui 6 bianche e 4 rosse. Si estraggono due palline senza reimmissione. Calcolare la probabilità dell'evento

$B_2 =$ la seconda è bianca

Svolgimento

Sia B_i = "estraggo una pallina bianca all'i-esima estrazione"

$P(B_2)$?

TODO: giaga alberello

Formula delle prob totali:

$$P(B_2) = P(B_2 \cap B_1) + P(B_2 \cap B_1^c) = P(B_2|B_1)P(B_1)P(B_2|B_1^c)P(B_1^c) = \frac{5}{9} \cdot \frac{6}{10} + \frac{6}{9} \cdot \frac{4}{10}$$

10.2 Esercitazione combinatoria

10.2.1 Esercizio 3

Testo

Tre amici si danno appuntamento nel bar della piazza centrale della città senza sapere che ci sono quattro bar. Qual è la probabilità che scelgano lo stesso bar? Tre bar differenti?

Svolgimento

Ognuno di loro è come se estraesse un bar da un'urna. Quindi i casi totali sono, ovviamente, 4^3 . Il primo brodo che sceglie il bar è fortunello, infatti può scegliere qualunque bar, quindi i casi favorevoli saranno 4, mentre gli altri potranno scegliere solo e soltanto il bar che ha scelto il brodez iniziale. Quindi

$$\frac{4 \cdot 1 \cdot 1}{4^3} = \frac{1}{16}$$

Nel caso dei bar differenti? Il primo può prendere un bar a caso, l'altro potrà scegliere solo sugli altri tre bar ed infine l'ultimo solo 2, matematicamente:

$$\frac{4 \cdot 3 \cdot 2}{4^3}$$

10.2.2 Esercizio 4

Testo

Si calcoli la probabilità di ottenere 2 palline rosse estraendone 4 da un'urna che contiene 3 palline rosse e 7 bianche. Si confrontino i casi di estrazioni con reimmissione, senza reimmissione e simultanea.

Svolgimento

- Estrazione con reimmissione:

$$\frac{\binom{2}{4} \cdot 3 \cdot 3 \cdot 7 \cdot 7}{10^3}$$

Quando estraggo la prima pallina, ho 10 possibilità, ed è così per tutte le estrazioni. Dobbiamo considerare anche l'ordine (dato che siamo nel caso con reimmissione), quindi devo vedere in quanti modi posso scegliere le "posizioni" rosse in 4 estrazioni (es. rrb, rrb, ...). Una volta scelte le posizioni (usando il binomiale $\binom{4}{2}$), possiamo ora semplicemente moltiplicarlo per il numero di casi favorevoli per un ordine (es. rrb = $3 \cdot 3 \cdot 7 \cdot 7$)

- Estrazione senza reimmissione:

$$\frac{\binom{2}{4} \cdot 3 \cdot 2 \cdot 7 \cdot 6}{10 \cdot 9 \cdot 8}$$

Stessa roba di prima, ma la probabilità considerando un solo ordine cambia (es. rrb = $3 \cdot 2 \cdot 7 \cdot 6$) e cambiano anche i casi possibili

- Estrazione simultanea:

Non c'è più l'ordine, quindi usiamo i binomi per rappresentare questa cosa

$$\frac{\binom{3}{2} \cdot \binom{7}{2}}{\binom{10}{4}}$$

10.2.3 Esercizio 6

Testo

Giocate 6 numeri al Superenalotto. Verranno estratti 6 numeri senza ripetizione dai primi 90 naturali, seguiti dall'estrazione di un 7 numero jolly, diverso quindi dai precedenti. Qual è la probabilità di fare 5+1 (indovinare 5 dei primi 6 numeri estratti e in più il numero jolly)?

Svolgimento

Intanto vengono estratti i primi 6 numeri senza emissione e si ha, ovviamente, come casi possibili:

$$90 \cdot 89 \cdot 88 \cdot 87 \cdot 86 \cdot 85 \cdot 84$$

Dato che devo indovinare solo 5 numeri, debbo sbagliare un numero, pertanto, per indicare che ho 6 modi per sbagliare il numerino (infatti che numero sbaglio tra i miei 6? il primo? secondo? terzo? devo tenere conto di tutte le possibilità) utilizzo il coefficiente binomiale $\binom{6}{1}$ moltiplicato per 84 (ovvero la probabilità di "sbagliare" tra i novanta numeri), infine moltiplico per 1 (una possibilità su tutti casi possibili) tante volte quanto devo indovinare i numeri. Pertanto si ha:

$$\frac{\binom{6}{1} \cdot 84 \cdot 1 \cdot 1 \cdot \dots}{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86 \cdot 85 \cdot 84} = \frac{6}{90 \cdot 89 \cdot 88 \cdot 87 \cdot 86 \cdot 85}$$

Esercizio 7

Testo

Supponiamo di giocare a poker con un mazzo da 52 carte, identificate dal seme (cuori, quadri, fiori, picche) e dal tipo (un numero da 2 a 10 oppure J, Q, K, A). Si calcolino:

- (a) la probabilità di avere una scala reale massima, ovvero le 5 carte 10, J, Q, K, A dello stesso seme;
- (b) la probabilità di avere una scala reale, ovvero una qualsiasi scala di 5 carte dello stesso seme (ricordiamo che con l'asso possiamo fare la scala A, 2, 3, 4, 5 ma anche 10, J, Q, K, A);
- (c) la probabilità di avere una scala, ovvero una qualsiasi scala di 5 carte non necessariamente dello stesso seme;
- (d) la probabilità di avere colore, ovvero un sottoinsieme di 5 carte dello stesso seme;
- (e) la probabilità di avere poker, ovvero un sottoinsieme di 5 carte in cui ci sono 4 carte dello stesso tipo;
- (f) la probabilità di avere un tris, ovvero un sottoinsieme di 5 carte in cui ci sono 3 carte dello stesso tipo e le altre due di tipo diverso tra loro e dalle prime tre;
- (g) la probabilità di avere una coppia, ovvero un sottoinsieme di 5 carte in cui ci sono 2 carte dello stesso tipo e le altre tre di tipo diverso tra loro e dalle prime due.

Svolgimento

- (a) L'ordine non importa \implies estrazione simultanea. Ci sono $\binom{52}{5}$ modi per estrarre una mano (casi possibili), i casi favorevoli sono 4 per il seme, poi le carte in mano sono fissate (abbiamo una sola possibilità una volta scelto il seme)

$$\frac{4 \cdot 1}{\binom{52}{5}}$$

- (b) I casi favorevoli sono ora 10 per ogni seme (10 pioli più bassi diversi da cui iniziare la scala), quindi i casi favorevoli sono di più:

$$\frac{4 \cdot 10}{\binom{52}{5}}$$

- (c) Ora per ogni carta della scala abbiamo 4 possibilita' per il seme, quindi una volta scelto il piolo piu' basso moltiplico per 4^5 :

$$\frac{10 \cdot 4^5}{\binom{52}{5}}$$

- (d) Una volta scelto il seme (4) ci sono $\binom{13}{5}$ insiemi distinti di tipi di carte che possiamo avere:

$$\frac{4 \cdot \binom{13}{5}}{\frac{52}{5}}$$

- (e) In un mazzo e' impossibile avere 5 carte dello stesso tipo, il max e' proprio 4. Scegliamo quindi il tipo con cui vogliamo fare poker (13) e l'insieme dei semi (1 dato che sono tutti i semi per forza). Poi scegliamone una a caso fra tutte le carte rimaste (48):

$$\frac{13 \cdot 1 \cdot 48}{()}$$

E se volessimo trovare solo tre dello stesso tipo? Scegliamo sempre il tipo (13), poi per ogni tipo possiamo avere $\binom{4}{3}$ insiemi per scegliere i semi. Poi fra le restanti carte non possiamo scegliere l'ultima carta con lo stesso tipo, quindi $\binom{48}{2}$ (fra le 48 carte che possiamo prendere ne scelgo 2)

$$\frac{13 \cdot \binom{4}{3} \cdot \binom{48}{2}}{\binom{52}{5}}$$

- (f) Il ragionamento e' simile a prima, ma per le ultime 2 carte scegliamo prima il tipo (che deve essere diverso) poi abbiamo 4 semi per ogni tipo che va bene:

$$\frac{13 \cdot \binom{4}{3} \cdot \binom{12}{2} \cdot 4^2}{\binom{52}{5}}$$

- (g) Caso di prima ma per una coppia, stesso ragionamento:

$$\frac{13 \cdot \binom{4}{2} \cdot \binom{12}{3} \cdot 4^3}{\binom{52}{5}}$$

10.2.4 Esercizio 8

Testo

Un mazzo di carte da scopone scientifico è costituito da 40 carte, identificate dal seme (spade, coppe, bastoni, denari) e dal tipo (asso, 2, 3, 4, 5, 6, 7, fante, cavallo, re). Supponendo di pescare 10 carte dal mazzo, calcolare la probabilità di estrarre:

- (a) l'asso di bastoni,
 (b) l'asso di spade e l'asso di coppe.

Svolgimento

- (a) : Si ha che , dato che l'ordine di pescaggio non è importante, appartiene ad un genere di problemi dove l'estrazione è simultanea. Pertanto si ha che i casi possibili sono $\binom{40}{10}$. Dopodiche per avere un asso di bastoni ho una possibilità su tutti casi possibili, inoltre le altre carte possono essere qualsiasi, quindi ho nove modi per pescare le 39 carte rimanenti, ovvero $\binom{39}{9}$

$$\frac{1 \cdot \binom{39}{9}}{\binom{40}{10}}$$

- (b) : Ugualmente, qui occorre tenere presente il numero di possibilità che si ha per pescare 8 carte in un insieme di 38 carte, per la possibilità di beccare l'asso di spade e l'asso di coppe, ovviamente ne ho 1 su tutti i casi possibili, quindi:

$$\frac{1 \cdot \binom{38}{8}}{\binom{40}{10}}$$

10.3 Probabilità condizionata e indipendenza

10.3.1 Es. 1

testo

Una roulette semplificata è formata da 12 numeri che sono “rosso” (R) e “nero” (N) in base allo schema seguente:

1	2	3	4	5	6	7	8	9	10	11	12
R	R	N	N	R	N	N	R	N	N	R	R

(a) Determinare uno spazio di probabilità che descriva l'esperimento aleatorio.

Siano

- A = “esce un numero pari”,
- B = “esce un numero rosso”,
- C = “esce un numero ≤ 3 ”,
- D = “esce un numero ≤ 6 ”,
- E = “esce un numero ≤ 8 ”,
- F = “esce un numero dispari ≤ 3 ”.

Calcolare le seguenti probabilità condizionate: (b) $P(A|C), P(C|A), P(B|C), P(A|F), P(D|F), P(A|B), P(B|A)$.

Stabilire quindi se: (c) gli eventi A, B e D sono a 2 a 2 indipendenti; (d) A, B, D costituiscono una famiglia di eventi indipendenti; (e) A, B, E costituiscono una famiglia di eventi indipendenti; (f) A, C, E costituiscono una famiglia di eventi indipendenti; (g) F è indipendente da A ; F è indipendente da D .

Svolgimento

(a) : $\Omega = \{1, 2, \dots, 12\}$

(b) :

- $\mathbb{P}(A|C) = \frac{1}{3} = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} =$
- $P(C|A) = \frac{1}{6}$
- $P(B|C) = \frac{2}{3}$
- $P(A|F) = 0$
- $P(D|F) = \frac{2}{2} = 1 = \frac{P(F \cap D)}{P(F)} = \frac{P(F)}{P(F)}$
- $P(A|B) = \frac{3}{6}$
- $\frac{3}{6}$

(c) : A, B, D sono indipendenti? Bisogna calcolare se $P(\text{cap4})$