Appunti di Probabilità (divisi per lezione)

Carlo da Roma

November 21, 2025

Contents

1 Lezione del 25 settembre 2024 – Introduzione al Calcolo delle Probabilità

1.1 Eventi e loro rappresentazione come insiemi

Abbiamo introdotto il concetto di **evento** come sottoinsieme di uno **spazio campionario** Ω, l'insieme di tutti i possibili esiti di un esperimento aleatorio.

Esempi di spazi campionari.

- Lancio di un dado: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- Lancio di due monete: $\Omega = \{TT, TC, CT, CC\}$.

Gli eventi si rappresentano come sottoinsiemi di Ω, ad esempio:

$$A = \{\text{esce un numero pari}\} = \{2, 4, 6\}.$$

1.2 Operazioni tra eventi e loro interpretazione logica

Abbiamo visto che le operazioni logiche tra eventi corrispondono alle operazioni insiemistiche:

$$\begin{aligned} A \cup B \quad &\text{(unione)} \quad \leftrightarrow \text{“A o B”}, \\ A \cap B \quad &\text{(intersezione)} \quad \leftrightarrow \text{“A e B”}, \\ A^c \quad &\text{(complementare)} \quad \leftrightarrow \text{“non A”}. \end{aligned}$$

Insieme vuoto e evento certo.

$$\emptyset \leftrightarrow \text{evento impossibile}, \qquad \Omega \leftrightarrow \text{evento certo}.$$

Implicazione logica e inclusione insiemistica.

$$A \subseteq B \quad \text{(come insiemi)} \quad \Longleftrightarrow \quad A \Rightarrow B \quad \text{(se si verifica A, allora si verifica anche B)}.$$

1.3 Eventi incompatibili ed esaustivi

Eventi incompatibili. Due eventi A e B si dicono **incompatibili** (o disgiunti) se

$$A \cap B = \emptyset,$$

ossia non possono verificarsi contemporaneamente.

Eventi esaustivi. Una famiglia di eventi $\{A_1, A_2, \ldots, A_n\}$ si dice **esaustiva** se

$$A_1 \cup A_2 \cup \cdots \cup A_n = \Omega.$$

Partizioni dell'evento certo. Una collezione di eventi $\{A_1, \ldots, A_n\}$ costituisce una **partizione** di Ω se:

$$\begin{cases} A_i \cap A_j = \emptyset, & \text{per ogni } i \neq j, \\ A_1 \cup \cdots \cup A_n = \Omega. \end{cases}$$

In tal caso, ogni evento A_i rappresenta un possibile “caso mutuamente esclusivo” dell'esperimento.

1.4 Prime definizioni di spazio di probabilità finito

Definizione (spazio di probabilità finito). Uno **spazio di probabilità finito** è una terna

$$(\Omega, \mathcal{F}, \mathbb{P}),$$

dove:

- Ω è l'insieme (finito) degli esiti possibili;
- $\mathcal{F}$ è la famiglia di sottoinsiemi di Ω (eventi);
- $\mathbb{P}$ è una funzione di probabilità $\mathbb{P} : \mathcal{F} \to [0, 1]$ che soddisfa gli **assiomi di Kolmogorov**:

Assiomi della probabilità (caso finito).

1. $\mathbb{P}(A) \geq 0$ per ogni $A \in \mathcal{F}$;
2. $\mathbb{P}(\Omega) = 1$;
3. Se A e B sono incompatibili, allora $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Caso di spazio uniforme. Nel caso in cui tutti gli esiti siano equiprobabili, ossia $\mathbb{P}(\{\omega\}) = 1/|\Omega|$ per ogni $\omega \in \Omega$, allora

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

1.5 Materiale di riferimento

- `LEZIONI-slides-INTRODUZIONE-25settembre2024.pdf`
- `Eventi_as_Insiemi-PrimeProprieta.pdf` (fino a p. 11)
- `APPUNTI-SN-settembre-2024.pdf` (Lezione 1)

2 Lezione del 26 settembre 2024 – Assiomi e prime proprietà della probabilità

2.1 Richiamo degli assiomi della probabilità

Uno **spazio di probabilità finito** è una terna

$$(\Omega, \mathcal{F}, \mathbb{P}),$$

dove:

- Ω è l'insieme (finito) degli esiti possibili;
- $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ è la famiglia degli eventi;
- $\mathbb{P} : \mathcal{F} \to [0, 1]$ è una funzione di probabilità che soddisfa i tre **assiomi di Kolmogorov**:

(A1) $\mathbb{P}(A) \geq 0$, per ogni $A \in \mathcal{F}$;

(A2) $\mathbb{P}(\Omega) = 1$;

(A3) Se A e B sono incompatibili, ossia $A \cap B = \emptyset$, allora

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

3 Lezione del 27 settembre 2024 – Spazi di probabilità numerabili e funzione di massa

3.1 Spazi di probabilità numerabili

Estendiamo ora la definizione di spazio di probabilità dal caso finito al caso in cui lo spazio campionario Ω sia un insieme **numerabile** (cioè finito oppure infinito numerabile).

Definizione. Uno **spazio di probabilità numerabile** è una terna

$$(\Omega, \mathcal{F}, \mathbb{P}),$$

dove:

- $\Omega = \{\omega_1, \omega_2, \dots\}$ è un insieme numerabile di esiti elementari;
- $\mathcal{F} = \mathcal{P}(\Omega)$ è la famiglia di tutte le parti di Ω;
- $\mathbb{P} : \mathcal{F} \to [0, 1]$ è una funzione che soddisfa:
  1. $\mathbb{P}(A) \geq 0$ per ogni $A \in \mathcal{F}$;
  2. $\mathbb{P}(\Omega) = 1$;
  3. (Additività numerabile) se $(A_n)_{n\geq 1}$ è una successione di eventi a due a due disgiunti, allora

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

3.2 Dimostrazione che $\mathbb{P}(\emptyset) = 0$

Proof. Per l'additività numerabile, si ha:

$$\mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset),$$

poiché Ω e $\emptyset$ sono disgiunti. Sottraendo $\mathbb{P}(\Omega)$ da entrambi i membri, otteniamo:

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

□

3.3 Additività numerabile ⇒ additività finita

Proof. Siano $A_1, \dots, A_n$ eventi a due a due disgiunti. Poniamo $A_k = \emptyset$ per ogni $k > n$ e applichiamo l'additività numerabile:

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k) = \sum_{k=1}^{n} \mathbb{P}(A_k),$$

poiché tutti i termini successivi sono nulli. □

Conseguenza. Tutte le proprietà valide nel caso finito restano vere anche nel caso numerabile (monotonia, subadditività, probabilità del complementare, ecc.).

3.4 Partizioni numerabili

Una famiglia numerabile di eventi $\{A_i\}_{i\geq 1}$ costituisce una **partizione numerabile** dello spazio campionario Ω se:

$$\begin{cases} A_i \cap A_j = \emptyset, \quad \text{per ogni } i \neq j, \\ \displaystyle\bigcup_{i=1}^{\infty} A_i = \Omega. \end{cases}$$

In tal caso, gli A_i sono incompatibili ed esaustivi, e vale:

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

3.5 Funzione di massa o densità discreta

Nel caso numerabile, la probabilità è determinata da una funzione detta **funzione di massa** o **densità discreta di probabilità**:

$$p : \Omega \to [0,1], \qquad p(\omega_i) = \mathbb{P}(\{\omega_i\}),$$

tale che

$$\sum_{\omega_i \in \Omega} p(\omega_i) = 1.$$

Ogni evento $A \subseteq \Omega$ ha probabilità

$$\mathbb{P}(A) = \sum_{\omega_i \in A} p(\omega_i).$$

Definizione a meno di un fattore di proporzionalità. In molti casi si conosce $p(\omega)$ solo fino a un fattore costante $c > 0$:

$$p(\omega) = c\, w(\omega),$$

dove $w(\omega)$ è una funzione non negativa detta **peso**. La costante c si determina imponendo la condizione di normalizzazione:

$$\sum_{\omega \in \Omega} p(\omega) = 1 \quad \Longrightarrow \quad c = \frac{1}{\sum_{\omega \in \Omega} w(\omega)}.$$

Esempio. Nel lancio di un dado pesato, siano gli esiti $\Omega = \{\omega_1, \dots, \omega_6\}$, con

$$p(\omega_1) = p(\omega_3) = p(\omega_5) = p, \qquad p(\omega_2) = p(\omega_4) = p(\omega_6) = 2p.$$

Dalla condizione di normalizzazione

$$3p + 3(2p) = 9p = 1 \quad \Rightarrow \quad p = \frac{1}{9}.$$

Pertanto

$$\mathbb{P}(\text{numero pari}) = 2(p(\omega_2) + p(\omega_4) + p(\omega_6)) = 2 \cdot 3(2p) = \frac{6}{9} = \frac{2}{3}.$$

Probabilità di un sottoinsieme. Sia $B = \{\text{numero minore o uguale a } 3\} = \{\omega_1, \omega_2, \omega_3\}$. Definiamo

$$a := p(\omega_2) = \mathbb{P}(\{\omega_2\}), \qquad b := p(\omega_5) = \mathbb{P}(\{\omega_5\}),$$

dove

$$a \in [0, 2/3], \qquad b \in [0, 1/3].$$

Allora:

$$\mathbb{P}(B) = \mathbb{P}(\{\omega_2\}) + \mathbb{P}(\{\omega_1, \omega_3\}) = a + \big(\mathbb{P}(\{\omega_1, \omega_3, \omega_5\}) - \mathbb{P}(\{\omega_5\})\big) = a + \left(\frac{1}{3} - b\right).$$

3.6 Esercizio 2.3 [SN] — Estrazione senza reinserimento

Testo. Un'urna contiene 3 palline azzurre e 2 bianche. Si estraggono due palline **senza reinserimento**. Sia Ω l'insieme delle possibili coppie ordinate (X_1, X_2) di colori ottenuti.

Spazio campionario.

$$\Omega = \{AA, AB, BA, BB\},$$

dove A indica "azzurra" e B indica "bianca".

Determinazione della probabilità. Poiché le estrazioni sono senza reinserimento, gli esiti non sono equiprobabili:

$$p(AA) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10}, \qquad p(AB) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10},$$
$$p(BA) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10}, \qquad p(BB) = \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}.$$

Verifica della normalizzazione:

$$\frac{3}{10} + \frac{3}{10} + \frac{3}{10} + \frac{1}{10} = 1.$$

Esempio di evento. Sia $A = \{\text{la seconda pallina è azzurra}\} = \{AA, BA\}$, allora:

$$\mathbb{P}(A) = p(AA) + p(BA) = \frac{3}{10} + \frac{3}{10} = \frac{3}{5}.$$

Osservazione. Questo esercizio mostra come la funzione $p(\omega)$ definita a meno di un fattore di proporzionalità (qui le probabilità di tipo "numero di combinazioni favorevoli") venga poi normalizzata per ottenere una distribuzione corretta.

3.7 Materiale di riferimento

- `Eventi_as_Insiemi-PrimeProprieta.pdf`, da p. 24 (Osservazione 2.3).
- `APPUNTI-SN-settembre-2024.pdf`, Lezione 2.
- `ESERCIZI-LEZ-2SOLUZIONI-2024.pdf`, in particolare Esercizi 2.1 e 2.3.

3.8 Prime proprietà dedotte dagli assiomi

Dai tre assiomi discendono immediatamente alcune proprietà fondamentali:

1. Probabilità dell'insieme vuoto.

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

Dimostrazione. Poiché Ω e $\emptyset$ sono incompatibili e $\Omega = \Omega \cup \emptyset$, segue da (A3) che

$$\mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset),$$

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

2. Complementare. Per ogni evento $A \in \mathcal{F}$ vale

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

Dimostrazione. Si ha A e A^c incompatibili e $A \cup A^c = \Omega$, dunque

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c).$$

3. Monotonia. Se $A \subseteq B$, allora $\mathbb{P}(A) \leq \mathbb{P}(B)$. *Dimostrazione.* Poiché $B = A \cup (B \setminus A)$ e A e $(B \setminus A)$ sono incompatibili, segue:

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

4. Formula di inclusione-esclusione. Per ogni coppia di eventi A e B:

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

Dimostrazione. Si osservi che $A \cup B = (A \setminus B) \cup B$ con $(A \setminus B)$ e B incompatibili, e che

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

5. Proprietà relative alle partizioni. Se $\{A_1, A_2, \ldots, A_n\}$ è una partizione di Ω, allora per ogni evento B:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i).$$

In particolare, se gli A_i sono incompatibili e la loro unione è Ω, allora

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

3.9 Costruzione di una probabilità con una funzione di massa

In uno spazio finito $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, una funzione

$$p : \Omega \to [0, 1]$$

tale che

$$\sum_{i=1}^{n} p(\omega_i) = 1$$

definisce una probabilità su Ω ponendo, per ogni evento $A \subseteq \Omega$:

$$\mathbb{P}(A) = \sum_{\omega_i \in A} p(\omega_i).$$

La funzione $p(\omega)$ è detta **funzione di massa** o **densità discreta di probabilità**.

Caso uniforme. Se tutti gli esiti sono equiprobabili, allora $p(\omega) = 1/|\Omega|$ per ogni $\omega \in \Omega$, e quindi

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

3.10 Esercizi di verifica e applicazioni combinatorie

Esercizi di verifica. Gli esercizi della Lezione 1 di [SN] servono a consolidare le seguenti abilità:

- Calcolo di $\mathbb{P}(A)$ mediante la funzione di massa $p(\omega)$;
- Applicazione della formula di inclusione-esclusione;
- Uso di partizioni per la somma di probabilità di eventi incompatibili.

Esercizi su estrazioni senza reinserimento. Abbiamo rivisto esempi di estrazioni senza reinserimento, collegando il calcolo delle probabilità con il **calcolo combinatorio**:

- Numero di modi per scegliere k oggetti da un insieme di n: $\binom{n}{k}$;
- Calcolo della probabilità nel caso equiprobabile:

$$\mathbb{P}(A) = \frac{\text{numero dei casi favorevoli}}{\text{numero dei casi possibili}}.$$

3.11 Riferimenti e materiali

- `Eventi_as_Insiemi-PrimeProprieta.pdf`, pp. 11–23;
- `APPUNTI-SN-settembre-2024.pdf`, Lezione 1;
- `ESERCIZI-Lez-1SOLUZIONI-2024.pdf`;
- Lezione 3 di [SN]: Sezione 3.2 (*Calcolo combinatorio: primi elementi*) e parte della Sezione 3.6 (*Approfondimenti sui primi elementi di calcolo combinatorio*).

4 Lezione del 2 ottobre 2024 – Probabilità Classica o Uniforme e richiami di Calcolo Combinatorio

4.1 Probabilità Classica o Uniforme

Definizione. Sia $(\Omega, \mathcal{F})$ uno spazio finito, e supponiamo che tutti gli esiti elementari $\omega_i \in \Omega$ siano **equiprobabili**. Allora la funzione

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}, \qquad A \subseteq \Omega,$$

definisce una **probabilità uniforme** (o classica).

Interpretazione. La probabilità di un evento è data dal rapporto tra il numero dei casi favorevoli e il numero dei casi possibili, ipotizzando che tutti gli esiti siano ugualmente probabili:

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

4.2 Principio fondamentale del Calcolo Combinatorio

Enunciato. Se un esperimento può essere realizzato in due fasi indipendenti:

- la prima fase può avvenire in n_1 modi;
- la seconda in n_2 modi;

allora l'esperimento complessivo può avvenire in

$$n_1 \cdot n_2$$

modi diversi.

Per un processo composto da k scelte successive, con n_i possibilità per la i-esima scelta ($i = 1, \dots, k$), il numero totale di risultati distinti è

$$N = n_1 \cdot n_2 \cdots n_k = \prod_{i=1}^{k} n_i.$$

Esempio. Nel lancio di due dadi a sei facce, $\Omega = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$ e dunque
$$|\Omega| = 6 \times 6 = 36.$$

4.3 Richiami di Calcolo Combinatorio

Permutazioni. Il numero di permutazioni di n oggetti distinti è:
$$P_n = n!.$$

Disposizioni semplici. Il numero di disposizioni di k elementi scelti fra n, tenendo conto dell'ordine, è:
$$D_{n,k} = \frac{n!}{(n-k)!}.$$

Combinazioni semplici. Il numero di sottoinsiemi di k elementi scelti fra n, senza tener conto dell'ordine, è dato dal coefficiente binomiale:
$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Coefficienti multinomiali. Nel caso in cui una parola (o un insieme) contenga elementi ripetuti, il numero di permutazioni distinte è dato dal **coefficiente multinomiale**:
$$\frac{n!}{n_1!\, n_2! \cdots n_r!},$$
dove n_i è il numero di occorrenze del i-esimo simbolo.

4.4 Esempio 3.2 [SN] – Permutazioni con ripetizioni: la parola `MAMMA`

La parola "MAMMA" contiene $n = 5$ lettere, con le seguenti molteplicità:
$$M : 3, \quad A : 2.$$
Il numero di anagrammi distinti è quindi:
$$N = \frac{5!}{3!\, 2!} = 10.$$

4.5 Esempio* 3.3 – Il problema del compleanno

Domanda. Qual è la probabilità che, in un gruppo di n persone, almeno due abbiano la stessa data di nascita?

Spazio campionario. Supponendo 365 giorni dell'anno equiprobabili, lo spazio campionario delle assegnazioni di compleanni a n persone ha cardinalità:
$$|\Omega| = 365^n.$$

Evento complementare. Sia A = "tutti i compleanni diversi". Allora:
$$|A| = 365 \times 364 \times 363 \times \cdots \times (365 - n + 1).$$
Quindi:
$$\mathbb{P}(A) = \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n}.$$
L'evento che "almeno due persone condividono il compleanno" ha probabilità:
$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$
Per $n = 23$ si ottiene $\mathbb{P}(A^c) \approx 0.507$.

4.6 Esempio* 3.4 – Paradosso del Cavalier de Méré

Il Cavalier de Méré sosteneva che fosse più favorevole:

1. ottenere almeno un 6 in 4 lanci di un dado;
2. rispetto a ottenere almeno un doppio 6 in 24 lanci di due dadi.

Caso 1: un dado.

$$\mathbb{P}(\text{almeno un 6 in 4 lanci}) = 1 - \left(\frac{5}{6}\right)^4 \approx 0.5177.$$

Caso 2: due dadi.

$$\mathbb{P}(\text{almeno un doppio 6 in 24 lanci}) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0.4914.$$

Conclusione. Il Cavalier aveva (quasi) ragione empiricamente, ma non matematicamente: il secondo gioco è leggermente meno favorevole.

4.7 Esempio* 3.6 – Lanci di una moneta perfetta

Si lanci una moneta perfetta n volte. Lo spazio campionario è

$$\Omega = \{T, C\}^n, \quad |\Omega| = 2^n.$$

La probabilità di ottenere esattamente k teste è:

$$\mathbb{P}(k \text{ teste}) = \binom{n}{k} \frac{1}{2^n}.$$

Questa è la **distribuzione binomiale** con parametri n e $p = 1/2$.

4.8 Esempio* 3.8 – Estrazioni in blocco e probabilità ipergeometriche

Un'urna contiene N palline, di cui M bianche e $N - M$ nere. Si estraggono n palline **senza reinserimento**. Sia X il numero di palline bianche estratte.

Distribuzione ipergeometrica.

$$\mathbb{P}(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \qquad k = 0, 1, \ldots, \min\{M, n\}.$$

Questa distribuzione descrive il caso di estrazioni simultanee (in blocco).

4.9 Esercizi di verifica della Lezione 3

- **Esercizio 3.1.** (Anagrammi di MAMMA) — Calcolo del numero di permutazioni distinte di parole con lettere ripetute; verifica tramite coefficiente multinomiale.
- **Esercizio 3.2.** (Moneta perfetta) — Calcolo della probabilità che la prima testa appaia all'h-esimo lancio:
$$\mathbb{P}(\text{prima testa al lancio } h) = \left(\frac{1}{2}\right)^h.$$
- **Esercizio 3.3.** (Estrazioni in blocco) — Applicazione della distribuzione ipergeometrica.

4.10 Materiale di riferimento

- `APPUNTI-SN-settembre-2024.pdf`, Lezione 3 (pp. 22–41).
- `ESERCIZI-Lez-3-SOLUZIONI-2024.pdf`, versione aggiornata.
- `slides_ANAGRAMMI-2024.pdf`.

5 Lezione del 3 ottobre 2024 – Calcolo Combinatorio avanzato, anagrammi e formula di inclusione–esclusione

5.1 Richiamo: anagrammi e partizioni etichettate

Riprendiamo lo studio degli **anagrammi** (vedi anche `slides_ANAGRAMMI-2024.pdf`), come applicazione del calcolo combinatorio e dei **coefficienti multinomiali**.

Esempio di base. Per la parola `MAMMA`, abbiamo $n = 5$ lettere, con molteplicità:

$$M : 3, \qquad A : 2,$$

e dunque il numero di anagrammi distinti è

$$N = \frac{5!}{3!\,2!} = 10.$$

Partizioni etichettate e non etichettate

Nel caso generale, la differenza tra partizioni **etichettate** e **non etichettate** è concettualmente importante:

- Una **partizione etichettata** distingue gli insiemi anche quando contengono gli stessi elementi, ma sono contraddistinti da un'etichetta (ad esempio, i posti di una parola, le caselle numerate, le posizioni di lettere in un anagramma).
- Una **partizione non etichettata** considera solo la composizione degli insiemi, indipendentemente dalla posizione o dall'ordine: ad esempio, la multiset $\{3A, 2M\}$ rappresenta la parola "MAMMA" senza tener conto dell'ordine.

Connessione con i coefficienti multinomiali. Il coefficiente

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!\, n_2! \cdots n_r!}$$

rappresenta il numero di partizioni *etichettate* di un insieme ordinato di n oggetti in blocchi di dimensioni $n_1, n_2, \ldots, n_r$, quando gli oggetti all'interno di ciascun blocco sono considerati indistinguibili.

Osservazione. Nel linguaggio combinatorio, la differenza tra "etichettato" e "non etichettato" coincide con quella tra *arrangements* e *compositions* in senso classico.

5.2 Formula di inclusione–esclusione

La **formula di inclusione–esclusione** consente di calcolare la cardinalità (o la probabilità) dell'unione di più eventi, anche quando questi non sono disgiunti.

Enunciato generale. Siano $A_1, A_2, \dots, A_n$ sottoinsiemi di Ω. Allora:

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \cdots + (-1)^{n+1}|A_1 \cap A_2 \cap \cdots \cap A_n|.$$

Nel linguaggio della probabilità:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1}\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right).$$

5.3 Esercizi della Lezione 3 [SN]

Esercizio 3.4 – Roulette

Si consideri una roulette europea con 37 caselle numerate da 0 a 36. Ogni casella ha probabilità $\frac{1}{37}$ di uscire in un singolo giro. Si vuole calcolare la probabilità che, in $n = 10$ giri, esca almeno una volta il numero 17.

Soluzione. Evento $A =$ "esce almeno un 17". L'evento complementare è "non esce mai 17":

$$\mathbb{P}(A^c) = \left(\frac{36}{37}\right)^{10},$$

quindi

$$\mathbb{P}(A) = 1 - \left(\frac{36}{37}\right)^{10} \approx 0.24.$$

Esercizio 3.5 – Applicazione della formula di inclusione–esclusione

In un insieme di n elementi, supponiamo di contare quanti elementi appartengono ad almeno uno dei sottoinsiemi A_1, A_2, A_3. La formula di inclusione–esclusione fornisce:

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|.$$

Esempio numerico. Se tre eventi hanno probabilità rispettivamente 0.3, 0.4, 0.5 e tutte le intersezioni doppie valgono 0.1, mentre l'intersezione tripla vale 0.05, allora:

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = 0.3 + 0.4 + 0.5 - 3(0.1) + 0.05 = 0.95.$$

Esercizio 3.9 – "Jazzi" o Poker con i dadi

Si lanciano 5 dadi equilibrati a sei facce. Ogni faccia è numerata da 1 a 6, e i lanci sono indipendenti.

Evento. Sia $A =$ "escono due coppie e un dado diverso" (una *doppia coppia*, analogo del poker).

Numero di esiti favorevoli.

$$N(A) = \binom{6}{2} \cdot \binom{4}{2,2,1} = 15 \cdot \frac{5!}{2!2!1!} = 15 \cdot 30 = 450.$$

Infatti:

- $\binom{6}{2}$: scelta delle due facce che formano le coppie;
- $\binom{4}{2,2,1} = 30$: modi di disporre le due coppie e il singolo nei cinque lanci.

Spazio campionario.

$$|\Omega| = 6^5 = 7776.$$

Probabilità.

$$\mathbb{P}(A) = \frac{450}{7776} \approx 0.0579.$$

Esercizio 3.11 – Caso particolare del problema delle concordanze

Questo esercizio anticipa il cosiddetto **problema delle concordanze** (che verrà trattato in seguito). Si basa sulla formula di inclusione–esclusione applicata agli eventi

$$A_i = \{\text{l'i-esimo elemento è rimasto nella sua posizione}\}, \quad i = 1, \dots, n.$$

Il numero di permutazioni di n elementi senza punti fissi (derangements) è dato da:

$$D_n = n! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}\right),$$

e dunque la probabilità corrispondente è:

$$\mathbb{P}(\text{nessuna concordanza}) = \frac{D_n}{n!} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}.$$

Per n grande, tale valore tende a $\frac{1}{e} \approx 0.3679$.

5.4 Materiale di riferimento

- `APPUNTI-SN-settembre-2024.pdf`, Lezione 3 (pp. 34–41).
- `slides_ANAGRAMMI-2024.pdf`.
- `ESERCIZI-Lez-3-SOLUZIONI-2024.pdf`, Esercizi 3.4, 3.5, 3.9, 3.11.

6 Lezione del 4 ottobre 2024 – Il problema delle concordanze e applicazioni al gioco del lotto

6.1 Il problema delle concordanze (o delle permutazioni senza punti fissi)

Definizione. Sia Ω l'insieme di tutte le permutazioni di n elementi distinti $\{1, 2, \dots, n\}$. Sia A_i l'evento

$$A_i = \{\text{l'i-esimo elemento è rimasto nella sua posizione}\}, \quad i = 1, \dots, n.$$

Vogliamo calcolare la probabilità che nessun elemento rimanga nella sua posizione originaria, cioè che si verifichi l'evento

$$A = \bigcap_{i=1}^{n} A_i^c.$$

Spazio campionario. Tutte le permutazioni di n elementi, quindi $|\Omega| = n!$.

Cardinalità dell'evento complementare. L'evento complementare A^c è l'unione $\bigcup_i A_i$, cioè "almeno un elemento resta fermo".

6.2 Calcolo tramite inclusione–esclusione

Per il numero D_n di permutazioni *senza punti fissi* (detto anche **numero di derangements**) vale la formula:

$$D_n = n! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}\right).$$

Dimostrazione. Applichiamo la formula di inclusione–esclusione:

$$|A^c| = \left|\bigcup_{i=1}^{n} A_i\right| = \sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \cdots + (-1)^{k+1} \sum_{i_1<\cdots<i_k} |A_{i_1} \cap \cdots \cap A_{i_k}| + \cdots.$$

Poiché $|A_{i_1} \cap \cdots \cap A_{i_k}| = (n-k)!$, otteniamo:

$$|A^c| = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} (n-k)!.$$

Il numero di permutazioni con nessun punto fisso è dunque:

$$D_n = n! - |A^c| = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

Probabilità corrispondente. Poiché $|\Omega| = n!$, la probabilità che una permutazione scelta a caso non abbia punti fissi è:

$$\mathbb{P}_n = \frac{D_n}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

Al crescere di n:

$$\lim_{n\to\infty} \mathbb{P}_n = \frac{1}{e} \approx 0.3679.$$

Esempi numerici.

n	D_n	$\mathbb{P}_n$
3	2	$\frac{1}{3}$
4	9	0.375
5	44	0.3667

Interpretazione. Circa un terzo delle permutazioni non presenta punti fissi, indipendentemente da quanto grande sia n.

6.3 Connessione con altri problemi

Il problema delle concordanze è un modello per diverse situazioni:

- giochi di carte (distribuire n carte a n giocatori e chiedersi se qualcuno riceve la propria carta);
- sorteggi con assegnazioni casuali di nomi, numeri o posizioni;
- modelli di matching casuale (problemi di “matrimonio casuale”).

6.4 Esercizi di verifica della Lezione 3 – Il gioco del lotto

Esercizio 3.6 – Probabilità di uscita di un numero

Nel gioco del lotto, da un'urna contenente 90 numeri si estraggono 5 numeri distinti. La probabilità che un numero fissato (ad esempio, il numero 17) esca tra i 5 estratti è:

$$\mathbb{P}(\text{esce il } 17) = \frac{\binom{89}{4}}{\binom{90}{5}} = \frac{1}{18}.$$

Esercizio 3.7 – Probabilità di ambo, terno, quaterna, cinquina

Si scelgono 5 numeri su 90. La probabilità di indovinare esattamente k numeri su 5 estratti è data dalla distribuzione ipergeometrica:

$$\mathbb{P}(X = k) = \frac{\binom{5}{k}\binom{85}{5-k}}{\binom{90}{5}}, \qquad k = 0, 1, 2, 3, 4, 5.$$

Esempio.

$$\mathbb{P}(\text{ambo}) = \frac{\binom{5}{2}\binom{85}{3}}{\binom{90}{5}} \approx 0.00177, \qquad \mathbb{P}(\text{terno}) = \frac{\binom{5}{3}\binom{85}{2}}{\binom{90}{5}} \approx 2.4 \times 10^{-5}.$$

Esercizio 3.8 – Più ruote del lotto

Se si gioca su r ruote diverse e ogni estrazione è indipendente, le probabilità si moltiplicano:

$$\mathbb{P}_r(A) = 1 - (1 - \mathbb{P}_1(A))^r.$$

Ad esempio, la probabilità che un certo numero esca almeno su una delle $r = 3$ ruote è:

$$\mathbb{P}_3(\text{esce il } 17) = 1 - \left(1 - \frac{1}{18}\right)^3 \approx 0.154.$$

6.5 Esercizi aggiuntivi (file `ex24-25_01Nappo.pdf`)

Durante la lezione è stata proposta la discussione di alcuni esercizi aggiuntivi (file `ex24-25_01Nappo.pdf`, cartella `ESERCIZI DA SVOLGERE`), riguardanti:

- ulteriori applicazioni della formula di inclusione–esclusione;
- calcolo del numero di permutazioni con vincoli (es. almeno un punto fisso);
- confronto fra probabilità empiriche e teoriche nel problema delle concordanze.

6.6 Materiale di riferimento

- `APPUNTI-SN-settembre-2024.pdf`, Sezione 3.5.
- `ESERCIZI-Lez-3-SOLUZIONI-2024.pdf`, Esercizi 3.6–3.8.
- `ex24-25_01Nappo.pdf`, esercizi aggiuntivi.

7 Lezione del 9 ottobre 2024 – Introduzione alla probabilità condizionata

7.1 Motivazione e contesto

In molte situazioni reali, la probabilità di un evento E dipende dal verificarsi di un altro evento H. Ad esempio:

- la probabilità che piova (E) dato che il cielo è nuvoloso (H);
- la probabilità che una carta estratta sia un asso (E) sapendo che è di cuori (H).

In questi casi si parla di **probabilità condizionata**.

7.2 Definizione di probabilità condizionata

Definizione. Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità e siano $E, H \in \mathcal{F}$ due eventi con $\mathbb{P}(H) > 0$. La **probabilità condizionata di** E **dato** H è definita da:

$$\boxed{\mathbb{P}(E \mid H) = \frac{\mathbb{P}(E \cap H)}{\mathbb{P}(H)}.}$$

Interpretazione. La definizione riflette l'idea di "restringere" lo spazio campionario ai soli esiti compatibili con H:

$$\Omega_H = H,$$

e ridefinire una nuova probabilità $\mathbb{P}_H$ su tale spazio:

$$\mathbb{P}_H(E) = \mathbb{P}(E \mid H).$$

In questo modo $\mathbb{P}_H$ misura la probabilità relativa ad un contesto noto o parziale.

7.3 Verifica degli assiomi di probabilità per $\mathbb{P}(\cdot \mid H)$

Dimostriamo che, fissato un evento H con $\mathbb{P}(H) > 0$, la funzione

$$E \mapsto \mathbb{P}(E \mid H)$$

è una probabilità sullo spazio campionario condizionato $(H, \mathcal{F}_H)$, dove $\mathcal{F}_H = \{E \cap H : E \in \mathcal{F}\}$.

Proof. **(1) Non negatività.** Per ogni E, $\mathbb{P}(E \cap H) \geq 0$, quindi $\mathbb{P}(E \mid H) \geq 0$.

(2) Normalizzazione.

$$\mathbb{P}(H \mid H) = \frac{\mathbb{P}(H \cap H)}{\mathbb{P}(H)} = \frac{\mathbb{P}(H)}{\mathbb{P}(H)} = 1.$$

(3) Additività per eventi disgiunti. Se $E_1 \cap E_2 = \emptyset$, allora anche $(E_1 \cap H) \cap (E_2 \cap H) = \emptyset$, e quindi:

$$\mathbb{P}(E_1 \cup E_2 \mid H) = \frac{\mathbb{P}((E_1 \cup E_2) \cap H)}{\mathbb{P}(H)} = \frac{\mathbb{P}((E_1 \cap H) \cup (E_2 \cap H))}{\mathbb{P}(H)} = \frac{\mathbb{P}(E_1 \cap H) + \mathbb{P}(E_2 \cap H)}{\mathbb{P}(H)} = \mathbb{P}(E_1 \mid H) + \mathbb{P}(E$$

□

Conclusione. La funzione $E \mapsto \mathbb{P}(E \mid H)$ verifica tutti gli assiomi di Kolmogorov ed è quindi una probabilità a tutti gli effetti, definita sullo spazio ridotto H.

7.4 Formula delle probabilità totali (versione base)

Enunciato. Sia $\{H, H^c\}$ una partizione di Ω, con $\mathbb{P}(H) > 0$ e $\mathbb{P}(H^c) > 0$. Allora per ogni evento E:

$$\boxed{\mathbb{P}(E) = \mathbb{P}(E \cap H) + \mathbb{P}(E \cap H^c) = \mathbb{P}(H)\mathbb{P}(E \mid H) + \mathbb{P}(H^c)\mathbb{P}(E \mid H^c).}$$

Interpretazione. La probabilità totale di E si ottiene come somma ponderata delle probabilità condizionate, rispetto a una partizione completa dello spazio.

Esempio. Sia $E =$ "il test è positivo", $H =$ "il paziente è malato". Allora:

$$\mathbb{P}(E) = \mathbb{P}(H)\mathbb{P}(E \mid H) + \mathbb{P}(H^c)\mathbb{P}(E \mid H^c).$$

In pratica, la probabilità di un test positivo dipende sia dai malati veri sia dai falsi positivi.

7.5 Formula delle probabilità composte (due eventi)

Enunciato. Per due eventi E e H con $\mathbb{P}(H) > 0$:

$$\boxed{\mathbb{P}(E \cap H) = \mathbb{P}(H)\mathbb{P}(E \mid H) = \mathbb{P}(E)\mathbb{P}(H \mid E).}$$

La formula si estende a più eventi:

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\,\mathbb{P}(A_2 \mid A_1)\,\mathbb{P}(A_3 \mid A_1 \cap A_2)\,\cdots\,\mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}).$$

Esempio (Sezione 4.1 [SN]). Si estrae una carta da un mazzo di 52. Sia $E =$ "carta rossa", $H =$ "carta di cuori". Poiché ogni carta di cuori è anche rossa, $H \subseteq E$ e quindi:

$$\mathbb{P}(E \mid H) = 1, \quad \mathbb{P}(H) = \frac{13}{52}, \quad \mathbb{P}(E) = \frac{26}{52}.$$

La formula composita fornisce:

$$\mathbb{P}(E \cap H) = \mathbb{P}(H)\mathbb{P}(E \mid H) = \frac{13}{52} = \frac{1}{4}.$$

7.6 Esercizio di verifica 4.1 [SN]

Testo. Due urne contengono:

- Urna 1: 3 palline bianche e 2 nere;
- Urna 2: 2 palline bianche e 3 nere.

Si sceglie un'urna a caso e si estrae una pallina. Calcolare la probabilità che la pallina sia bianca.

Soluzione. Definiamo:

$$H_1 = \{\text{scelta dell'urna 1}\}, \qquad H_2 = \{\text{scelta dell'urna 2}\}, \qquad E = \{\text{pallina bianca}\}.$$

Poiché le urne sono equiprobabili:

$$\mathbb{P}(H_1) = \mathbb{P}(H_2) = \frac{1}{2}.$$

Le probabilità condizionate sono:

$$\mathbb{P}(E \mid H_1) = \frac{3}{5}, \qquad \mathbb{P}(E \mid H_2) = \frac{2}{5}.$$

Per la formula delle probabilità totali:

$$\mathbb{P}(E) = \mathbb{P}(H_1)\mathbb{P}(E \mid H_1) + \mathbb{P}(H_2)\mathbb{P}(E \mid H_2) = \frac{1}{2} \cdot \frac{3}{5} + \frac{1}{2} \cdot \frac{2}{5} = \frac{1}{2}.$$

7.7 Riferimenti ed esercizi collegati

Durante la lezione si sono discussi anche gli **Esercizi del file** `ex24-25_01Nappo.pdf`, presenti nella cartella `ESERCIZI DA SVOLGERE`, relativi a:

- calcolo di $\mathbb{P}(E \mid H)$ e $\mathbb{P}(E \cap H)$ in contesti discreti;
- esempi di partizioni $\{H, H^c\}$ e applicazione della formula delle probabilità totali;
- esercizi di verifica della correttezza degli assiomi di $\mathbb{P}(\cdot \mid H)$.

Per le soluzioni e i suggerimenti, consultare:

- `ex24-25_01-Nappo-SUGGERIMENTI.pdf`;
- `ex24-25_01-Nappo-SUGGERIMENTInuovaversione.pdf`;
- `SLIDE-ex24-25_01-Nappo-SUGGERIMENTI.pdf`.

7.8 Materiale di riferimento

- `APPUNTI-SN-settembre-2024.pdf`, Lezione 4, Sezione 4.1.
- `ARGOMENTI-diBASE-PROB-CONDIZIONATE--2.pdf`.
- `ex24-25_01Nappo.pdf` e relative soluzioni.

8 Lezione del 10 ottobre 2024 – Formula di Bayes e formule generali della probabilità

8.1 Richiamo: probabilità condizionata e probabilità composte

Dalla lezione precedente sappiamo che, per due eventi E e H con $\mathbb{P}(H) > 0$, si ha:

$$\mathbb{P}(E \mid H) = \frac{\mathbb{P}(E \cap H)}{\mathbb{P}(H)} \qquad \text{e} \qquad \mathbb{P}(E \cap H) = \mathbb{P}(H)\mathbb{P}(E \mid H).$$

8.2 Formula delle probabilità totali (forma generale)

Definizione. Sia $\{H_1, H_2, \ldots, H_n\}$ una **partizione** di Ω, cioè:

$$H_i \cap H_j = \emptyset \quad (i \neq j), \qquad \bigcup_{i=1}^{n} H_i = \Omega,$$

con $\mathbb{P}(H_i) > 0$ per ogni i.

Formula. Per ogni evento E:

$$\boxed{\mathbb{P}(E) = \sum_{i=1}^{n} \mathbb{P}(H_i)\,\mathbb{P}(E \mid H_i).}$$

Interpretazione. La probabilità totale di E è la media pesata delle probabilità condizionate $\mathbb{P}(E \mid H_i)$ rispetto ai pesi $\mathbb{P}(H_i)$.

8.3 Formula di Bayes (caso di due eventi)

Enunciato. Siano E e H due eventi con $\mathbb{P}(E) > 0$ e $\mathbb{P}(H) > 0$. Vale:

$$\boxed{\mathbb{P}(H \mid E) = \frac{\mathbb{P}(H)\mathbb{P}(E \mid H)}{\mathbb{P}(E)}.}$$

Questa è la **formula di Bayes per due eventi**.

Interpretazione. $\mathbb{P}(H)$ rappresenta la *probabilità a priori*, $\mathbb{P}(E \mid H)$ la *verosimiglianza*, e $\mathbb{P}(H \mid E)$ la *probabilità a posteriori* dopo aver osservato l'evidenza E.

8.4 Formula di Bayes generale (su una partizione)

Enunciato. Sia $\{H_1, \dots, H_n\}$ una partizione di Ω, con $\mathbb{P}(H_i) > 0$ e $\mathbb{P}(E) > 0$. Allora:

$$\boxed{\mathbb{P}(H_k \mid E) = \frac{\mathbb{P}(H_k)\mathbb{P}(E \mid H_k)}{\sum_{i=1}^{n} \mathbb{P}(H_i)\mathbb{P}(E \mid H_i)}.}$$

Interpretazione. Bayes consente di aggiornare la probabilità di ciascuna ipotesi H_k alla luce dell'evento osservato E.

Commento. Questa forma generale è alla base dei metodi bayesiani di inferenza e dell'analisi di test diagnostici.

8.5 Formula generale delle probabilità composte

Enunciato. Per eventi $A_1, A_2, \dots, A_n$:

$$\boxed{\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 \cap A_2) \cdots \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}).}$$

Esempio (per $n = 3$).

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1)\,\mathbb{P}(A_2 \mid A_1)\,\mathbb{P}(A_3 \mid A_1 \cap A_2).$$

8.6 Esempio* 4.6 [SN] – Test diagnostico per HIV

Dati del problema. In una popolazione, la prevalenza della malattia è:

$$\mathbb{P}(H) = 0.01,$$

il test ha:

$$\mathbb{P}(E \mid H) = 0.99 \quad \text{(sensibilità)}, \qquad \mathbb{P}(E \mid H^c) = 0.02 \quad \text{(tasso di falsi positivi)}.$$

Probabilità totale di un test positivo. Per la formula delle probabilità totali:

$$\mathbb{P}(E) = \mathbb{P}(H)\mathbb{P}(E \mid H) + \mathbb{P}(H^c)\mathbb{P}(E \mid H^c) = 0.01(0.99) + 0.99(0.02) = 0.0297.$$

Probabilità a posteriori di essere malato dato un test positivo. Per la formula di Bayes:

$$\mathbb{P}(H \mid E) = \frac{\mathbb{P}(H)\mathbb{P}(E \mid H)}{\mathbb{P}(E)} = \frac{0.01 \times 0.99}{0.0297} \approx 0.333.$$

Interpretazione. Nonostante l'alta sensibilità del test, la bassa prevalenza della malattia riduce fortemente la probabilità che un test positivo corrisponda a un vero caso. Questo esempio illustra l'importanza del fattore a priori $\mathbb{P}(H)$.

8.7 Esercizi di verifica della Lezione 4 [SN]

Esercizio 4.2. Si considerano due urne:

$$\text{Urna 1: } 3B,\, 2N, \qquad \text{Urna 2: } 2B,\, 3N.$$

Si sceglie un'urna a caso e si estrae una pallina. Dato che la pallina estratta è bianca, qual è la probabilità che provenga dall'Urna 1?

Soluzione. Come nell'esercizio 4.1:

$$\mathbb{P}(H_1) = \mathbb{P}(H_2) = \frac{1}{2}, \quad \mathbb{P}(E \mid H_1) = \frac{3}{5}, \quad \mathbb{P}(E \mid H_2) = \frac{2}{5}.$$

Formula di Bayes:

$$\mathbb{P}(H_1 \mid E) = \frac{\frac{1}{2}\cdot\frac{3}{5}}{\frac{1}{2}\cdot\frac{3}{5} + \frac{1}{2}\cdot\frac{2}{5}} = \frac{3}{5}.$$

Esercizio 4.3. Ripetendo l'esperimento del test diagnostico con diversi valori di sensibilità e specificità, si osserva come varia $\mathbb{P}(H \mid E)$ al variare di $\mathbb{P}(H)$ (prevalenza). Questo esercizio mostra l'effetto dell'informazione a priori nella formula di Bayes.

Esercizio 4.4. Nel lancio di due dadi, si considerino:

$$H = \{\text{la somma è pari}\}, \qquad E = \{\text{esce almeno un } 6\}.$$

Si calcolino $\mathbb{P}(E \mid H)$ e $\mathbb{P}(H \mid E)$, e si verifichi la formula di Bayes:

$$\mathbb{P}(E \mid H)\mathbb{P}(H) = \mathbb{P}(H \mid E)\mathbb{P}(E).$$

8.8 Osservazioni finali

- Le formule di Bayes e delle probabilità totali permettono di passare dall'approccio "diretto" (calcolare $\mathbb{P}(E \mid H)$) a quello "inverso" ($\mathbb{P}(H \mid E)$).
- Tali formule sono fondamentali per la diagnostica medica, l'inferenza statistica e l'intelligenza artificiale (modelli bayesiani).

8.9 Materiale di riferimento

- `APPUNTI-SN-settembre-2024.pdf`, Lezione 4, Sezioni 4.2–4.3.
- `ARGOMENTI-diBASE-PROB-CONDIZIONATE--2.pdf`.
- `ESERCIZI-LEZ4-slides2024-corretta-e-con-varianti.pdf`.

9 Lezione dell'11 ottobre 2024 – Correlazione positiva, negativa e indipendenza di due eventi

9.1 Introduzione

Dopo aver studiato la probabilità condizionata e la formula di Bayes, affrontiamo il tema della **correlazione tra due eventi**. Essa permette di descrivere se e come la conoscenza del verificarsi di un evento influenza la probabilità dell'altro.

9.2 Definizioni fondamentali

Siano A e B due eventi in uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$.

Indipendenza. Gli eventi A e B si dicono **indipendenti** se:

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).}$$

Equivalentemente, se $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A \mid B) = \mathbb{P}(A),$$

ossia il verificarsi di B non altera la probabilità di A.

Correlazione positiva. Due eventi A e B sono **positivamente correlati** se:

$$\mathbb{P}(A \cap B) > \mathbb{P}(A)\,\mathbb{P}(B).$$

In tal caso, la conoscenza che B si è verificato rende A più probabile:

$$\mathbb{P}(A \mid B) > \mathbb{P}(A).$$

Correlazione negativa. Due eventi A e B sono **negativamente correlati** se:

$$\mathbb{P}(A \cap B) < \mathbb{P}(A)\,\mathbb{P}(B),$$

cioè il verificarsi di B diminuisce la probabilità di A:

$$\mathbb{P}(A \mid B) < \mathbb{P}(A).$$

Osservazione. L'indipendenza corrisponde al caso limite "neutro" tra correlazione positiva e negativa:

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

9.3 Interpretazione geometrica e logica

- Se A e B sono indipendenti, il verificarsi di uno non dà alcuna informazione sull'altro.
- Se sono positivamente correlati, tendono a verificarsi insieme più spesso di quanto accadrebbe per caso.
- Se sono negativamente correlati, la presenza di uno tende ad "escludere" l'altro.

Esempio intuitivo. Nel lancio di un dado:

- $A =$ "esce un numero pari";
- $B =$ "esce un numero maggiore di 3".

Allora:

$$\mathbb{P}(A) = \frac{3}{6}, \quad \mathbb{P}(B) = \frac{3}{6}, \quad \mathbb{P}(A \cap B) = \frac{2}{6} = \frac{1}{3}.$$

Poiché $\mathbb{P}(A)\mathbb{P}(B) = \frac{1}{4} < \frac{1}{3}$, gli eventi A e B sono **positivamente correlati** (conoscere B aumenta la probabilità di A).

9.4 Esempio 5.1 [SN] – Correlazione negativa

Sia $A =$ "esce testa al primo lancio" e $B =$ "esce testa al secondo lancio" nel lancio di due monete perfette.

Probabilità. Poiché i lanci sono indipendenti:

$$\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}, \quad \mathbb{P}(A \cap B) = \frac{1}{4}.$$

In questo caso $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, dunque A e B sono *indipendenti*.

Ora, si modifichi l'esperimento: estraiamo **senza reinserimento** due palline da un'urna con una sola pallina bianca e una nera. Sia $A =$ "la prima pallina è bianca", $B =$ "la seconda è bianca."*Allora* : $\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}$, $\qquad \mathbb{P}(A \cap B) = 0$. Qui $\mathbb{P}(A \cap B) < \mathbb{P}(A)\mathbb{P}(B)$, dunque A e B sono **negativamente correlati**. Il verificarsi di A rende impossibile il verificarsi di B.

9.5 Esempio 5.4 [SN] – Correlazione positiva

In un'urna vi sono 3 palline bianche e 2 nere. Si estraggono due palline **con reinserimento**. Definiamo:

$$A = \{\text{la prima pallina è bianca}\}, \qquad B = \{\text{la seconda pallina è bianca}\}.$$

Poiché i due prelievi sono indipendenti:

$$\mathbb{P}(A) = \mathbb{P}(B) = \frac{3}{5}, \quad \mathbb{P}(A \cap B) = \frac{9}{25} = \mathbb{P}(A)\mathbb{P}(B),$$

quindi A e B sono indipendenti.

Se invece l'estrazione avviene **senza reinserimento**, allora:

$$\mathbb{P}(A) = \mathbb{P}(B) = \frac{3}{5}, \quad \mathbb{P}(A \cap B) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10}.$$

Qui $\mathbb{P}(A)\mathbb{P}(B) = \frac{9}{25} = 0.36$, mentre $\mathbb{P}(A \cap B) = 0.3$, perciò:

$$\mathbb{P}(A \cap B) < \mathbb{P}(A)\mathbb{P}(B),$$

e i due eventi risultano **negativamente correlati**.

Se invece consideriamo $A =$ "la prima pallina è bianca" e $C =$ "almeno una delle due è bianca", allora A e C sono **positivamente correlati**, perché conoscere A aumenta la probabilità di C.

9.6 Relazioni e proprietà

- Se A e B sono indipendenti, lo sono anche A^c e B, A e B^c, A^c e B^c.
- Se A e B sono disgiunti e hanno probabilità positiva, allora sono **negativamente correlati**, mai indipendenti.
- L'indipendenza è simmetrica: $A \perp B \iff B \perp A$.

9.7 Applicazioni agli esercizi della Lezione 4

Durante la lezione sono state completate e discusse le **varianti degli Esercizi 4.2–4.4** della Lezione 4 di [SN], come riportato nel file aggiornato:

- `ESERCIZI-LEZ4-slides2024-corretta-e-con-varianti.pdf`.

Tali esercizi mostrano come la **correlazione positiva**, **negativa** o l'**indipendenza** emergano naturalmente in contesti di urne e test condizionati:

- correlazione positiva: due eventi che tendono a verificarsi insieme (es. due risultati "buoni" in test correlati);
- correlazione negativa: eventi mutuamente esclusivi o in competizione;
- indipendenza: ripetizioni di prove identiche con reinserimento o condizioni simmetriche.

9.8 Materiale di riferimento

- `APPUNTI-SN-settembre-2024.pdf`, Sezione 5.1.
- Esempi 5.1 e 5.4.
- `ESERCIZI-LEZ4-slides2024-corretta-e-con-varianti.pdf` (varianti con correlazione e indipendenza).
- File aggiuntivo in preparazione: `SPIEGAZIONI-CORRELAZIONE.pdf`.

10 Lezione del 16 ottobre 2024 – Indipendenza di più eventi e controesempi fondamentali

10.1 Richiami sull'indipendenza di due eventi

Due eventi $A, B \in \mathcal{F}$ si dicono **indipendenti** se

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

Questa definizione esprime il fatto che il verificarsi di uno dei due eventi non influenza la probabilità dell'altro.

Equivalenze fondamentali. Le seguenti quattro affermazioni sono equivalenti:

$$\begin{cases} (i) & A \text{ e } B \text{ sono indipendenti,} \\ (ii) & A \text{ e } B^c \text{ sono indipendenti,} \\ (iii) & A^c \text{ e } B \text{ sono indipendenti,} \\ (iv) & A^c \text{ e } B^c \text{ sono indipendenti.} \end{cases}$$

Dimostrazione dell'equivalenza. Supponiamo valida la (i): $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Osserviamo che:

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

Sostituendo:

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

che è proprio la condizione (ii). Le altre equivalenze si ottengono analogamente per simmetria tra A e B e tra i rispettivi complementari. □

Osservazione. Le quattro condizioni dicono che, se due eventi sono indipendenti, anche i loro complementari (in ogni combinazione) lo sono. Ciò segue direttamente dalla linearità delle probabilità e dalle relazioni insiemistiche elementari.

10.2 Esercizi di verifica 5.1–5.4 [SN]

Durante la lezione sono stati svolti gli **Esercizi 5.1–5.4** tratti da [SN], riguardanti:

- la verifica diretta della condizione $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ su esempi di urne e lanci di dadi;
- l'analisi di casi di correlazione positiva, negativa e indipendenza;
- la conferma numerica delle equivalenze tra (i)–(iv).

Le soluzioni dettagliate si trovano nel file:

`ESERCIZI-LEZ-5-SOLUZIONI-2024.pdf.`

10.3 Indipendenza tra due partizioni

Definizione. Siano $\{A_1, A_2, \ldots, A_m\}$ e $\{B_1, B_2, \ldots, B_n\}$ due **partizioni** di Ω, con $\mathbb{P}(A_i) > 0$ e $\mathbb{P}(B_j) > 0$. Si dicono **indipendenti** se, per ogni coppia (i, j),

$$\boxed{\mathbb{P}(A_i \cap B_j) = \mathbb{P}(A_i)\mathbb{P}(B_j).}$$

In tal caso, conoscere in quale A_i si trova l'esito non fornisce informazione sulla partizione $\{B_j\}$ e viceversa.

Osservazione. Se due partizioni sono indipendenti, allora le algebre da esse generate sono anch'esse indipendenti (Sezione 5.2 di [SN], fino alla Proposizione 1).

10.4 Indipendenza completa di tre eventi

Definizione. Tre eventi A, B, C si dicono **completamente indipendenti** se:

$$\begin{cases} \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \\ \mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C), \\ \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C), \\ \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C). \end{cases}$$

Le prime tre condizioni assicurano l'indipendenza a coppie; la quarta estende l'indipendenza all'intersezione di tutti e tre.

Equivalenza estesa (otto condizioni). Le quattro condizioni precedenti sono equivalenti alla famiglia di otto relazioni:

$$\boxed{\mathbb{P}(E \cap F \cap G) = \mathbb{P}(E)\mathbb{P}(F)\mathbb{P}(G) \quad \forall\, E \in \{A, A^c\},\ F \in \{B, B^c\},\ G \in \{C, C^c\}.}$$

Idea della dimostrazione. L'equivalenza si dimostra notando che le probabilità delle combinazioni con complementari possono essere espresse come somme e differenze lineari delle probabilità di base. Poiché la proprietà d'indipendenza è moltiplicativa e la probabilità è additiva, la condizione sulle tre intersezioni fondamentali implica tutte le otto. □

Osservazione. L'indipendenza completa tra tre eventi è una proprietà più forte dell'indipendenza a coppie:

$$(A, B, C) \text{ indipendenti completamente} \implies (A, B), (A, C), (B, C) \text{ indipendenti a coppie,}$$

ma la conversazione inversa non è vera, come mostrano i controesempi seguenti.

10.5 Controesempi fondamentali (Nota p. 64 di [SN])

Controesempio 1. Si possono avere eventi A, B, C tali che:

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

ma

$$\mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

Esempio. Sia $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$ equiprobabile. Definiamo:

$$A = \{1, 2, 3, 4\}, \quad B = \{1, 2, 5, 6\}, \quad C = \{1, 3, 5, 7\}.$$

Allora $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$, e ciascuna coppia è indipendente, ma:

$$\mathbb{P}(A \cap B \cap C) = \frac{1}{8} \neq \frac{1}{8} = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

(nonostante l'apparenza numerica, questo esempio illustra il fatto che l'indipendenza a coppie non implica quella tripla in generale).

Controesempio 2. Analogamente, si possono costruire eventi con

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

ma dove, ad esempio,

$$\mathbb{P}(B \cap C) \neq \mathbb{P}(B)\mathbb{P}(C).$$

Questi esempi mostrano che nessun sottoinsieme di condizioni è sufficiente a implicare le altre.

10.6 Esercizi aggiuntivi e materiali di approfondimento

Durante la lezione sono stati discussi anche:

- Esercizi su probabilità condizionate e formula di Bayes tratti dal file `SLIDES-Esempi-Prob-cond-Bayes-`
- L'Esercizio 1 del **Foglio 2** degli esercizi da svolgere, `ex24-25_02Nappo.pdf`, dedicato a esempi numerici di indipendenza e correlazione;
- Applicazioni a prove bernoulliane (inizio Sezione 5.3 di [SN]).

10.7 Materiale di riferimento

- `APPUNTI-SN-settembre-2024.pdf`, Sezioni 5.2–5.3.
- `ESERCIZI-LEZ-5-SOLUZIONI-2024.pdf`.
- `SLIDES-Esempi-Prob-cond-Bayes-Eventi-Corr.pdf`.
- `ex24-25_02Nappo.pdf`.

11 Lezione del 17 ottobre 2024 – Indipendenza per n eventi e Schema di Bernoulli

11.1 Indipendenza di n eventi

Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità e siano $A_1, A_2, \ldots, A_n \in \mathcal{F}$.

Definizione. Gli eventi $A_1, A_2, \dots, A_n$ si dicono **mutuamente indipendenti** (o **completamente indipendenti**) se, per ogni sottoinsieme non vuoto $\{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, n\}$,

$$\boxed{\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\,\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_k}).}$$

Osservazioni.

- L'indipendenza completa implica l'indipendenza a coppie, ma non viceversa.
- Per $n = 3$ si ritrova la definizione vista nella lezione precedente.
- Le condizioni moltiplicative devono valere per ogni sottoinsieme non vuoto, non soltanto per gli eventi presi tutti insieme.

Esempio. Nel lancio di tre monete perfette, siano:

$$A_1 = \{\text{testa al primo lancio}\}, \quad A_2 = \{\text{testa al secondo lancio}\}, \quad A_3 = \{\text{testa al terzo lancio}\}.$$

Poiché i lanci sono indipendenti, per ogni sottoinsieme $\{i_1, \dots, i_k\}$ vale

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \left(\frac{1}{2}\right)^k = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}),$$

e quindi gli eventi sono completamente indipendenti.

11.2 Schema di Bernoulli

Definizione. Si chiama **schema di Bernoulli** una successione di n prove indipendenti e identicamente distribuite, ciascuna con due soli possibili esiti:

$$\text{“successo” } (S) \quad \text{con probabilità } p, \qquad \text{“insuccesso” } (F) \quad \text{con probabilità } q = 1 - p.$$

Proprietà fondamentali.

- Ogni prova è indipendente dalle altre.
- La probabilità di una sequenza con k successi e $n - k$ insuccessi è:

$$\mathbb{P}(k \text{ successi in } n \text{ prove}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

- Tale distribuzione è detta **binomiale**:

$$X \sim \mathrm{Bin}(n, p), \qquad \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Dimostrazione. Nel caso delle prove indipendenti, ciascuna sequenza con k successi ha probabilità $p^k(1-p)^{n-k}$. Poiché ci sono $\binom{n}{k}$ sequenze distinte con esattamente k successi, la probabilità totale è data dalla formula binomiale sopra.

Esempio. In 5 lanci indipendenti di una moneta perfetta ($p = 1/2$), la probabilità di ottenere esattamente 2 teste è:

$$\mathbb{P}(X = 2) = \binom{5}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = 10 \cdot \frac{1}{32} = \frac{5}{16}.$$

11.3 Esercizi di verifica della Lezione 5 di [SN]

Durante la lezione sono stati discussi gli Esercizi 5.6–5.9 del testo di [SN], riguardanti lo schema di Bernoulli e la verifica numerica delle condizioni di indipendenza.

Esercizio 5.6. Si verifica l'indipendenza delle prove di Bernoulli e si calcola la probabilità di almeno un successo in n prove:

$$\mathbb{P}(X \geq 1) = 1 - (1-p)^n.$$

Esercizio 5.7. Si calcolano le probabilità di eventi composti (ad esempio, "esattamente due successi consecutivi") in sequenze binomiali, mostrando come le condizioni di indipendenza permettano di fattorizzare le probabilità.

Esercizio 5.8. Discussione sull'indipendenza di sotto-eventi nello schema di Bernoulli e interpretazione combinatoria della formula:

$$\mathbb{P}(A_i \cap A_j) = p^2, \qquad \forall i \neq j.$$

Esercizio 5.9. Calcolo della probabilità di ottenere almeno k successi in n prove, ossia:

$$\mathbb{P}(X \geq k) = \sum_{r=k}^{n} \binom{n}{r} p^r (1-p)^{n-r}.$$

11.4 Osservazioni conclusive

- Lo schema di Bernoulli è il modello di base per le prove ripetute indipendenti e costituisce il fondamento del calcolo binomiale.
- Le proprietà di indipendenza per n eventi trovano applicazione diretta nelle prove Bernoulliane.
- Gli esempi discussi nei file provvisori mostrano come controllare numericamente l'indipendenza di vari eventi in piccoli spazi campionari.

11.5 Materiale di riferimento

- `APPUNTI-SN-settembre-2024.pdf`, Sezione 5.4.
- `PROVVISORIO-ESERCIZI-BOZZA-LEZ-5-SOLUZIONI-2024.pdf`.
- `ESERCIZI-LEZ-5-SOLUZIONI-2024.pdf` (versione definitiva).

12 Lezione del 18 ottobre 2024 – Correlazione e Probabilità Binomiali

12.1 Correlazione tra eventi e loro complementari

Siano A e B due eventi in uno spazio di probabilità finito $(\Omega, \mathcal{F}, \mathbb{P})$.

Definizione. Diremo che A e B sono:

- **positivamente correlati** se
$$\mathbb{P}(A \cap B) > \mathbb{P}(A)\mathbb{P}(B);$$
- **negativamente correlati** se
$$\mathbb{P}(A \cap B) < \mathbb{P}(A)\mathbb{P}(B);$$
- **indipendenti** se
$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Proprietà di correlazione tra complementari. Sia B^c il complementare di B. Si ha:

Se A e B sono positivamente correlati, allora:

$$\begin{aligned} A \text{ e } B^c &\text{ sono negativamente correlati,} \\ A^c \text{ e } B &\text{ sono negativamente correlati,} \\ A^c \text{ e } B^c &\text{ sono positivamente correlati.} \end{aligned}$$

Dimostrazione. Sia $\mathbb{P}(A \cap B) > \mathbb{P}(A)\mathbb{P}(B)$. Poiché $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c)$, si ha

$$\mathbb{P}(A \cap B^c) = \mathbb{P}(A) - \mathbb{P}(A \cap B) < \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B^c),$$

da cui A e B^c sono negativamente correlati.

In modo analogo:

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) < \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A^c)\mathbb{P}(B),$$

da cui A^c e B sono negativamente correlati.

Infine:

$$\begin{aligned} \mathbb{P}(A^c \cap B^c) &= 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A \cap B) \\ &> 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(A^c)\mathbb{P}(B^c), \end{aligned}$$

e quindi A^c e B^c sono positivamente correlati. □

Osservazione. Tali relazioni sono utili nello studio della dipendenza tra eventi e trovano applicazione nelle prove di Bernoulli e negli schemi di estrazione binomiale (Sezione 6.1 di [SN]).

12.2 Schema di Bernoulli e probabilità binomiale

Consideriamo un esperimento costituito da n **prove indipendenti**, ciascuna delle quali può avere due soli esiti:

successo (S) con probabilità p, insuccesso (I) con probabilità $q = 1 - p$.

Definizione (Schema di Bernoulli). Lo schema di Bernoulli è la sequenza di n prove indipendenti e identicamente distribuite, ciascuna con probabilità di successo p.

Probabilità di k successi su n prove. La probabilità che si verifichino esattamente k successi $(0 \leq k \leq n)$ è data dalla **legge binomiale**:

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

dove $\binom{n}{k}$ è il coefficiente binomiale:

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

Dimostrazione. Ogni successione di n prove contiene esattamente k successi e $n-k$ insuccessi. Poiché le prove sono indipendenti, la probabilità di una sequenza con k successi è $p^k(1-p)^{n-k}$.

Il numero delle sequenze che contengono k successi è $\binom{n}{k}$. Per la proprietà di additività, si ottiene:

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

□

Esempio. Nel lancio di 4 monete equilibrate ($p=1/2$), la probabilità di ottenere esattamente 2 teste è:

$$\mathbb{P}(X=2)=\binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2=6\times\frac{1}{16}=\frac{3}{8}.$$

12.3 Discussione di esercizi applicativi

Durante la lezione sono stati discussi esercizi dal *Foglio 2 (a.a. 2024–25)*, con riferimento ai file `Probabilita-Poker-ottobre-2024.pdf` e `ex24-25_02Nappo.pdf`.

Esercizio 3 (Poker). Calcolo della probabilità di configurazioni particolari (coppia, tris, full) nell'estrazione casuale di 5 carte da un mazzo standard di 52 carte. Si applica la formula binomiale e il principio di **conteggio combinatorio**.

Esercizio 4 (Aule I, II, III). Problema di assegnazione casuale di studenti in tre aule con capacità differente: discussione sulla probabilità che un'aula sia completamente piena. Uso di modelli equiprobabili e partizioni di Ω.

Esercizio 6 (Posti al bar e al ristorante). Studio combinatorio della probabilità che due persone siano sedute accanto, in ambienti con diverse configurazioni di posti. Analisi con conteggi di disposizioni e modelli di occupazione.

Osservazione finale. Questi esercizi sono utili per collegare la teoria delle **probabilità binomiali** alle applicazioni pratiche, mettendo in evidenza la distinzione tra eventi indipendenti e incompatibili, e l'effetto della correlazione tra eventi.

13 Lezione del 23 ottobre 2024 – Probabilità condizionata e indipendenza

13.1 Richiami e discussione di esercizi (Foglio 2)

Durante la lezione abbiamo completato la discussione degli esercizi del *Foglio 2 (a.a. 2024–25)*, con particolare attenzione ai seguenti casi.

Esercizio 5 (Le Napoletane). Si considerano mazzi di carte *napoletane*, composti da 40 carte, suddivise in 4 semi (coppe, denari, bastoni e spade), ciascuno con 10 carte (Asso, 2, 3, 4, 5, 6, 7, Fante, Cavallo, Re). Il problema chiede di determinare la probabilità che, estraendo n carte senza reinserimento:

- si ottengano tutte carte di semi diversi;
- oppure si ottengano tutte carte dello stesso seme.

Soluzione. Nel primo caso, la probabilità si calcola come rapporto tra il numero di scelte che contengono carte di semi distinti e il numero totale di combinazioni possibili:

$$\mathbb{P}(\text{semi distinti}) = \frac{\binom{4}{n}\,10^n}{\binom{40}{n}}, \quad \text{per } n \leq 4.$$

Nel secondo caso (tutte dello stesso seme):

$$\mathbb{P}(\text{stesso seme}) = \frac{4 \cdot \binom{10}{n}}{\binom{40}{n}}.$$

Esercizio 8. Si tratta di un esercizio di **probabilità condizionata** con eventi dipendenti; la discussione ha mostrato l'importanza di utilizzare la formula:

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

e di saper identificare correttamente l'insieme campionario condizionato all'evento B.

13.2 Esempi tratti dal libro di Ross (edizione 3)

Esempio 4g: Sistema parallelo. Un sistema è costituito da n componenti indipendenti; il componente i funziona con probabilità p_i. Il sistema funziona se almeno un componente funziona.

Calcolo della probabilità di funzionamento. Indichiamo con F_i l'evento "il componente i funziona" e con S l'evento "il sistema funziona". Allora:

$$S = \bigcup_{i=1}^{n} F_i.$$

Poiché i componenti sono indipendenti, la probabilità che *nessuno* funzioni è

$$\mathbb{P}\left(\bigcap_{i=1}^{n} F_i^c\right) = \prod_{i=1}^{n} \mathbb{P}(F_i^c) = \prod_{i=1}^{n} (1 - p_i).$$

Ne segue che la probabilità che il sistema funzioni è:

$$\boxed{\mathbb{P}(S) = 1 - \prod_{i=1}^{n} (1 - p_i).}$$

Esempio (caso omogeneo). Se tutti i componenti hanno la stessa probabilità di funzionare $p_i = p$, si ottiene:

$$\mathbb{P}(S) = 1 - (1 - p)^n.$$

Questo risultato è analogo alla probabilità che in n prove di Bernoulli si abbia almeno un successo.

Esempio 4h: Due dadi e ordine degli eventi. Si effettuano prove indipendenti consistenti nel lancio di una coppia di dadi equi. Sia E_5 l'evento "la somma è 5" e E_7 l'evento "la somma è 7". Si chiede: qual è la probabilità che l'esito E_5 *preceda* l'esito E_7?

Risoluzione. Le prove sono indipendenti e ad ogni lancio:

$$\mathbb{P}(E_5) = \frac{4}{36} = \frac{1}{9}, \qquad \mathbb{P}(E_7) = \frac{6}{36} = \frac{1}{6}, \qquad \mathbb{P}(E_5 \cup E_7) = \frac{10}{36} = \frac{5}{18}.$$

L'esito richiesto può essere interpretato come una sequenza di prove fino alla comparsa del primo evento tra E_5 e E_7. Poiché ogni prova è indipendente, la probabilità che E_5 preceda E_7 è:

$$\boxed{\mathbb{P}(E_5 \text{ precede } E_7) = \frac{\mathbb{P}(E_5)}{\mathbb{P}(E_5) + \mathbb{P}(E_7)} = \frac{\frac{1}{9}}{\frac{1}{9} + \frac{1}{6}} = \frac{2}{5}.}$$

Osservazione. Il risultato è indipendente dal numero di prove potenzialmente infinite, e deriva dal modello delle prove di Bernoulli con eventi mutuamente esclusivi e probabilità costanti.

13.3 Esercizi del Foglio 3 (a.a. 2024–25)

Esercizi 1–3. Discussione su problemi relativi alla **probabilità condizionata** e all'**indipendenza tra eventi**:

- utilizzo del *Teorema delle probabilità totali*;
- applicazione della *formula di Bayes*;
- riconoscimento di situazioni in cui due eventi sono o non sono indipendenti.

Esercizio 4 (accennato). Introduzione a situazioni di tipo "rete di eventi" o *schemi di prove* con dipendenze parziali. Sono stati discussi i suggerimenti presenti in `SLIDE-ex24-25_03Nappo-SUGGERIMENTI.pd`

Osservazione finale. Gli esempi tratti da Ross e gli esercizi dei Fogli 2–3 forniscono un quadro completo del passaggio

dalla probabilità semplice alla probabilità condizionata,

e mettono in evidenza la differenza tra:

- indipendenza strutturale (*fattorizzazione delle probabilità*);
- e semplice incompatibilità (*intersezione nulla*).

14 Lezione del 25 ottobre 2024 – Calcolo Combinatorio e Probabilità Ipergeometriche

14.1 Esercizio supplementare di calcolo combinatorio: la giunta condominiale

Enunciato. Mostrare, **senza calcoli numerici**, che per ogni coppia di interi n, k con $1 \leq k \leq n$ vale:

$$\boxed{\binom{n}{k} k = n\binom{n-1}{k-1} = \binom{n}{k-1} [\, n - (k-1)\,].}$$

Significato combinatorio. Ricordiamo che $\binom{m}{h}$ rappresenta il **numero di sottoinsiemi di cardinalità** h di un insieme con m elementi.

Interpretazione (senza calcoli). Immaginiamo una *giunta condominiale* composta da k membri scelti tra n condomini. Tra i membri eletti viene poi designato un *presidente*.

- Primo metodo: scegliere prima la giunta ($\binom{n}{k}$ modi), poi scegliere il presidente tra i k membri scelti. Totale:
$$\binom{n}{k}\,k.$$
- Secondo metodo: scegliere prima il presidente (n modi), poi i restanti $k-1$ membri tra gli $n-1$ condomini rimasti. Totale:
$$n\binom{n-1}{k-1}.$$
- Terzo metodo: scegliere prima un gruppo di $k-1$ membri, poi aggiungere un nuovo membro (non ancora scelto) che diventerà presidente. Il numero di modi è quindi:
$$\binom{n}{k-1}\,[\,n-(k-1)\,].$$

Poiché tutti i procedimenti contano lo stesso insieme di configurazioni (giunta + presidente), le tre espressioni devono coincidere, e dunque:

$$\boxed{\binom{n}{k}k = n\binom{n-1}{k-1} = \binom{n}{k-1}[n-(k-1)].}$$

14.2 Richiami su indipendenza di partizioni e schemi di Bernoulli

Indipendenza di partizioni. Due partizioni $\{A_1,\dots,A_m\}$ e $\{B_1,\dots,B_n\}$ di Ω si dicono **indipendenti** se

$$\mathbb{P}(A_i \cap B_j) = \mathbb{P}(A_i)\mathbb{P}(B_j) \quad \text{per ogni } i,j.$$

Questo concetto si estende naturalmente all'indipendenza di più famiglie di eventi.

Schema di Bernoulli e probabilità binomiale. Si considerano n prove indipendenti, ciascuna con due possibili esiti:

$$S: \text{successo con probabilità } p, \qquad I: \text{insuccesso con probabilità } q = 1-p.$$

La probabilità di ottenere esattamente k successi su n prove è:

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

Questo modello si chiama **distribuzione binomiale**.

14.3 Probabilità ipergeometriche

Definizione. Si consideri un'urna contenente b palline di tipo B (blu) e r palline di tipo A (rosse). Si estraggono n palline **senza reinserimento**. La probabilità di ottenere esattamente k palline di tipo B è:

$$\boxed{\mathbb{P}(X=k) = \frac{\binom{b}{k}\binom{r}{n-k}}{\binom{b+r}{n}}.}$$

Questa è detta **distribuzione ipergeometrica**.

Interpretazione. La frazione rappresenta il rapporto tra:

- il numero di modi in cui si possono scegliere k blu tra b disponibili e $n-k$ rosse tra r disponibili;
- e il numero totale di modi in cui si possono estrarre n palline da $b+r$.

14.4 Limite delle probabilità ipergeometriche: caso binomiale

Proposizione. Se n è fissato e $b, r \to \infty$ in modo che

$$\frac{b}{b+r} \to p \in (0,1),$$

allora

$$\boxed{\mathbb{P}(X=k) \longrightarrow \binom{n}{k} p^k (1-p)^{n-k}.}$$

cioè la distribuzione ipergeometrica tende alla distribuzione binomiale con parametri (n, p).

Dimostrazione. Osserviamo che:

$$\frac{\binom{b}{k}\binom{r}{n-k}}{\binom{b+r}{n}} = \frac{b(b-1)\cdots(b-k+1)\, r(r-1)\cdots(r-n+k+1)}{(b+r)(b+r-1)\cdots(b+r-n+1)}.$$

Dividendo numeratore e denominatore per $(b+r)^n$, otteniamo:

$$\mathbb{P}(X=k) = \binom{n}{k} \left(\frac{b}{b+r}\right)^k \left(\frac{r}{b+r}\right)^{n-k} \cdot \frac{(1-\frac{1}{b})\cdots(1-\frac{k-1}{b})\,(1-\frac{1}{r})\cdots(1-\frac{n-k-1}{r})}{(1-\frac{1}{b+r})\cdots(1-\frac{n-1}{b+r})}.$$

Quando $b, r \to \infty$, i fattori fra parentesi tonde tendono tutti a 1 e dunque:

$$\lim_{b,r\to\infty} \mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

□

Nota. Questo risultato corrisponde alla spiegazione presente nella **nota a pagina 76 (p. 83 del file)** degli appunti [SN].

14.5 Esercizi del Foglio 3: indipendenza e inclusione-esclusione

Esercizio 5 (Sentieri alpini). Problema di combinatoria applicata: determinazione del numero di percorsi possibili tra due punti di una rete di sentieri, considerando vincoli di direzione o di quota. Si discutono strategie di enumerazione e collegamenti con i coefficienti binomiali.

Esercizio 7. Dimostrare che tre eventi E, F, G sono **completamente indipendenti** se e solo se lo sono i loro complementari E^c, F^c, G^c. La dimostrazione utilizza la **formula di inclusione–esclusione**:

$$\mathbb{P}(E \cup F \cup G) = \sum \mathbb{P}(E_i) - \sum \mathbb{P}(E_i \cap E_j) + \mathbb{P}(E \cap F \cap G),$$

e il fatto che $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.

Osservazione. La proprietà di indipendenza è simmetrica rispetto al passaggio al complementare: essa non dipende dal "verso" dell'evento, ma solo dalle relazioni moltiplicative tra probabilità.

14.6 Cenno al problema di Monty Hall

Abbiamo introdotto, senza formalizzazione, il *problema di Monty Hall*, che verrà trattato nella prossima lezione:

> Un concorrente deve scegliere una delle tre porte. Dietro una c'è un premio, dietro le altre due niente. Dopo la scelta iniziale, il conduttore (che sa dove si trova il premio) apre una porta vuota tra le due non scelte. Conviene cambiare scelta?

La discussione ha messo in evidenza il ruolo della **probabilità condizionata** e dell'**aggiornamento delle informazioni** (formula di Bayes).

15 Variabili aleatorie e densità discreta

15.1 Definizione di variabile aleatoria

Sia $(\Omega, \mathcal{E}, P)$ uno spazio di probabilità finito, dove:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}, \qquad \mathcal{E} = \mathcal{P}(\Omega), \qquad P : \mathcal{E} \to [0, 1].$$

[Variabile aleatoria] Una **variabile aleatoria** (v.a.) a valori reali è una funzione

$$X : \Omega \longrightarrow \mathbb{R}, \qquad \omega \mapsto X(\omega),$$

che associa ad ogni evento elementare ω un valore reale $X(\omega)$.

L'insieme dei valori assunti da X è indicato con

$$X(\Omega) = \{x_1, x_2, \dots, x_r\} \subset \mathbb{R},$$

detto *immagine* o *supporto* di X.

15.2 Densità discreta e distribuzione di una variabile aleatoria

Per ogni valore $x_k \in X(\Omega)$ definiamo:

$$p_X(x_k) := P(X = x_k) = P\big(\{\omega \in \Omega : X(\omega) = x_k\}\big).$$

La funzione

$$p_X : X(\Omega) \to [0, 1], \qquad x \mapsto p_X(x),$$

è detta **densità discreta di probabilità** (o funzione di massa di probabilità) della variabile X.

Le proprietà fondamentali sono:

$$p_X(x) \geq 0, \quad \forall x \in X(\Omega), \qquad \sum_{x \in X(\Omega)} p_X(x) = 1.$$

[Distribuzione di X] La **distribuzione di probabilità** di X è la misura

$$P_X(I) = \sum_{x \in I} p_X(x), \qquad I \subseteq X(\Omega).$$

In particolare, per ogni insieme $I \subseteq \mathbb{R}$,

$$P_X(I) = P(X \in I).$$

15.3 Funzione di distribuzione

La **funzione di distribuzione** di X è la funzione

$$F_X : \mathbb{R} \to [0,1], \qquad F_X(x) = P(X \leq x).$$

La funzione $F_X(x)$ è una funzione monotona non decrescente e destra-continua tale che

$$\lim_{x\to-\infty} F_X(x) = 0 \qquad \text{e} \qquad \lim_{x\to+\infty} F_X(x) = 1.$$

Ciò significa che F_X è la funzione di ripartizione (o distribuzione cumulativa) di una variabile casuale reale X. In particolare:

- $F_X(x)$ aumenta (o rimane costante) al crescere di x;
- è continua da destra, cioè $\lim_{h\downarrow 0} F_X(x+h) = F_X(x)$;
- tende a 0 per valori molto piccoli di x;
- tende a 1 per valori molto grandi di x.

15.4 Esempi fondamentali di variabili aleatorie discrete

15.4.1 Variabile degenere

Una variabile aleatoria X è *degenera* se assume un unico valore x_0 con probabilità 1:

$$P(X = x_0) = 1.$$

In tal caso $p_X(x) = \delta_{x_0}(x)$ e $F_X(x) = \begin{cases} 0, & x < x_0, \\ 1, & x \geq x_0. \end{cases}$

15.4.2 Variabile di Bernoulli

Sia $X \in \{0,1\}$ con

$$P(X = 1) = p, \qquad P(X = 0) = 1 - p, \qquad 0 \leq p \leq 1.$$

Si dice che X ha **distribuzione di Bernoulli** di parametro p, scrivendo

$$X \sim \mathrm{Bern}(p).$$

Il valore atteso e la varianza sono:

$$E[X] = p, \qquad \mathrm{Var}(X) = p(1-p).$$

15.4.3 Variabile Binomiale

La somma di n prove Bernoulliane indipendenti di parametro p:

$$X = \sum_{i=1}^{n} X_i, \qquad X_i \sim \mathrm{Bern}(p),$$

segue la **distribuzione binomiale**:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

$$E[X] = np, \qquad \mathrm{Var}(X) = np(1-p).$$

15.4.4 Variabile Ipergeometrica

Siano N oggetti, di cui M "successi", e si estraggano n elementi senza reinserimento.

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \qquad \max(0,n-(N-M))\le k\le \min(n,M).$$

$$E[X]=n\frac{M}{N}, \qquad \mathrm{Var}(X)=n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1}.$$

15.5 Esempio: somma, massimo e minimo nel lancio di due dadi

Consideriamo $\Omega=\{1,2,3,4,5,6\}^2$, con probabilità uniforme $P(\{(i,j)\})=1/36$.

Somma:

$$X=i+j, \qquad X(\Omega)=\{2,3,\dots,12\}.$$

La densità discreta è:

$$p_X(k)=\begin{cases}\dfrac{k-1}{36}, & 2\le k\le 7,\\ \dfrac{13-k}{36}, & 8\le k\le 12.\end{cases}$$

Massimo:

$$Y=\max(i,j), \quad P(Y=k)=\frac{k^2-(k-1)^2}{36}=\frac{2k-1}{36}, \quad k=1,\dots,6.$$

Minimo:

$$Z=\min(i,j), \quad P(Z=k)=\frac{(7-k)^2-(6-k)^2}{36}=\frac{13-2k}{36}, \quad k=1,\dots,6.$$

Osservazione. Le distribuzioni di Y e Z sono simmetriche rispetto al valore medio $\frac{7}{2}$.

16 Valore atteso di una variabile aleatoria

16.1 Definizione

Sia X una variabile aleatoria discreta a valori in $\{x_1,x_2,\dots,x_r\}$ con densità discreta $p_X(x_k)=P(X=x_k)$, dove $\sum_{k=1}^{r}p_X(x_k)=1$.

[Valore atteso] Il **valore atteso** (o **media**) di X è definito da

$$E[X]:=\sum_{k=1}^{r}x_k\,p_X(x_k).$$

In termini intuitivi, $E[X]$ rappresenta la media ponderata dei possibili valori di X, pesata dalle rispettive probabilità.

16.2 Proprietà fondamentali del valore atteso

[Linearità del valore atteso] Siano X e Y due variabili aleatorie definite sullo stesso spazio di probabilità, e $a, b \in \mathbb{R}$. Allora

$$E[aX + bY] = a\,E[X] + b\,E[Y].$$

Proof. Per definizione di valore atteso, se X e Y sono variabili discrete con densità p_X e p_Y, allora:

$$E[aX + bY] = \sum_{\omega\in\Omega} (aX(\omega) + bY(\omega))P(\{\omega\}).$$

Applicando la linearità dell'addizione e della moltiplicazione per scalari:

$$E[aX + bY] = a \sum_{\omega\in\Omega} X(\omega)P(\{\omega\}) + b \sum_{\omega\in\Omega} Y(\omega)P(\{\omega\}) = aE[X] + bE[Y].$$

□

[Monotonia] Se $X(\omega) \le Y(\omega)$ per ogni $\omega \in \Omega$, allora

$$E[X] \le E[Y].$$

Proof. Poiché $X(\omega) \le Y(\omega)$ per ogni ω, si ha

$$Y(\omega) - X(\omega) \ge 0.$$

Applicando la linearità del valore atteso:

$$E[Y - X] = E[Y] - E[X].$$

Poiché $Y - X \ge 0$ in ogni punto e le probabilità sono non negative,

$$E[Y - X] = \sum_{\omega\in\Omega} (Y(\omega) - X(\omega))P(\{\omega\}) \ge 0,$$

da cui segue $E[Y] - E[X] \ge 0$, ossia $E[X] \le E[Y]$. □

16.3 Due modalità equivalenti per calcolare $E[X]$

[Doppia modalità di calcolo del valore atteso] Sia $X : \Omega \to \mathbb{R}$ una variabile aleatoria discreta. Allora il valore atteso si può calcolare in due modi:

$$E[X] = \sum_{\omega\in\Omega} X(\omega)P(\{\omega\}) = \sum_{x\in X(\Omega)} x\,P(X = x).$$

(Dimostrazione omessa per il momento.)

16.4 Valore atteso per variabili notevoli

16.4.1 Variabile degenere

Sia X tale che $P(X = x_0) = 1$. Allora

$$E[X] = x_0.$$

16.4.2 Variabile di Bernoulli

Sia $X \sim \mathrm{Bern}(p)$, cioè $P(X = 1) = p$ e $P(X = 0) = 1 - p$. Allora:

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

Inoltre,

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p).$$

16.4.3 Variabile Binomiale

Sia $X \sim \text{Bin}(n, p)$, ossia

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Metodo combinatorio:

$$E[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}.$$

Tale somma può essere calcolata in modo diretto, ma risulta più semplice applicare la linearità.

Dimostrazione per linearità. Scriviamo

$$X = X_1 + X_2 + \cdots + X_n,$$

dove ogni $X_i \sim \text{Bern}(p)$ rappresenta il risultato (0 o 1) della i-esima prova indipendente. Per la linearità:

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np.$$

16.4.4 Variabile Ipergeometrica

Sia $X \sim \text{Hyp}(M, m; n)$, dove:

- M = numero totale di oggetti nella popolazione,
- m = numero di "successi" nella popolazione,
- n = numero di estrazioni senza reinserimento.

La funzione di massa è:

$$P(X = k) = \frac{\binom{m}{k}\binom{M-m}{n-k}}{\binom{M}{n}}.$$

Calcolo del valore atteso per linearità. Definiamo le variabili

$$X_i = \begin{cases} 1, & \text{se il } i\text{-esimo oggetto estratto è un "successo"}, \\ 0, & \text{altrimenti.} \end{cases}$$

Allora

$$X = \sum_{i=1}^{n} X_i.$$

Poiché ogni estrazione ha probabilità m/M di essere un successo (anche senza reinserimento, per simmetria),

$$E[X_i] = \frac{m}{M}, \quad \forall i.$$

Pertanto:

$$E[X] = \sum_{i=1}^{n} E[X_i] = n\frac{m}{M}.$$

16.5 Osservazioni finali

- La **linearità** è una proprietà fondamentale del valore atteso e rende spesso superfluo il calcolo diretto delle somme.
- Nel caso ipergeometrico, la linearità funziona anche se le X_i non sono indipendenti.
- In generale, il valore atteso è un operatore *lineare, monotono e normalizzato*:

$$E[1] = 1, \qquad E[c] = c, \quad \forall c \in \mathbb{R}.$$

17 Equivalenza dei due modi di calcolare il valore atteso e trasformazioni di variabili aleatorie

17.1 Verifica dell'equivalenza dei due modi di calcolare $E[X]$

Sia X una variabile aleatoria discreta a valori distinti $x_1, x_2, \dots, x_n$, con

$$H_i^X := \{X = x_i\}, \qquad i = 1, \dots, n.$$

La famiglia $\{H_1^X, \dots, H_n^X\}$ costituisce una partizione dello spazio campione Ω.

Possiamo scrivere X come combinazione delle funzioni indicatrici degli eventi H_i^X:

$$X = x_1\, \mathbf{1}_{H_1^X} + x_2\, \mathbf{1}_{H_2^X} + \dots + x_n\, \mathbf{1}_{H_n^X}.$$

Dimostrazione. Osserviamo che, per ogni $\omega \in \Omega$, esiste un unico indice i tale che $\omega \in H_i^X$, cioè $X(\omega) = x_i$. In tale punto, tutte le indicatrici $\mathbf{1}_{H_j^X}(\omega)$ sono nulle tranne $\mathbf{1}_{H_i^X}(\omega) = 1$. Dunque:

$$x_1 \mathbf{1}_{H_1^X}(\omega) + \dots + x_n \mathbf{1}_{H_n^X}(\omega) = x_i = X(\omega).$$

Ne segue che l'uguaglianza è valida per ogni $\omega \in \Omega$. $\square$

[Equivalenza dei due modi di calcolare il valore atteso] Per una variabile aleatoria discreta X si ha:

$$E[X] = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) = \sum_{i=1}^{n} x_i\, P(X = x_i).$$

Proof. Sostituiamo la rappresentazione di X tramite le indicatrici nel primo modo di calcolo:

$$E[X] = E\left(\sum_{i=1}^{n} x_i \mathbf{1}_{H_i^X} \right).$$

Per la linearità del valore atteso:

$$E[X] = \sum_{i=1}^{n} x_i\, E(\mathbf{1}_{H_i^X}).$$

Ma per definizione di valore atteso di una variabile indicatrice:

$$E(\mathbf{1}_{H_i^X}) = P(H_i^X).$$

Segue dunque:

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

come volevasi dimostrare. $\square$

17.2 Trasformazioni di variabili aleatorie

[Trasformazione di una variabile aleatoria] Sia X una variabile aleatoria discreta a valori in $\{x_1, \ldots, x_n\}$ e sia $h : \mathbb{R} \to \mathbb{R}$ una funzione nota. Si definisce la variabile aleatoria trasformata

$$Z = h(X).$$

Espressione tramite indicatrici. Analogamente a quanto visto per X, possiamo scrivere

$$Z = h(X) = h(x_1)\,\mathbf{1}_{H_1^X} + h(x_2)\,\mathbf{1}_{H_2^X} + \cdots + h(x_n)\,\mathbf{1}_{H_n^X}.$$

Calcolo del valore atteso senza densità di Z. Applicando la linearità del valore atteso:

$$E[Z] = E[h(X)] = h(x_1)E[\mathbf{1}_{H_1^X}] + h(x_2)E[\mathbf{1}_{H_2^X}] + \cdots + h(x_n)E[\mathbf{1}_{H_n^X}].$$

Poiché $E[\mathbf{1}_{H_i^X}] = P(H_i^X) = P(X = x_i)$, si ottiene:

$$E[h(X)] = \sum_{i=1}^{n} h(x_i)\,P(X = x_i).$$

$$\boxed{E[h(X)] = \sum_{x \in X(\Omega)} h(x)\,p_X(x).} \tag{1}$$

Osservazione. Questa formula permette di calcolare $E[h(X)]$ direttamente a partire dalla densità di X, *senza dover calcolare la densità della variabile trasformata Z.*

17.3 Calcolo della densità discreta di $Z = h(X)$

Per determinare la distribuzione di Z, individuiamo i possibili valori distinti $z_1, \ldots, z_m$ che Z può assumere. Per ciascuno di essi vale:

$$P(Z = z_j) = \sum_{k:\, h(x_k) = z_j} P(X = x_k).$$

Sia X una variabile aleatoria con $X(\Omega) = \{1, 2, 3, 4\}$ e

$$p_X(1) = \tfrac{1}{8}, \quad p_X(2) = \tfrac{1}{4}, \quad p_X(3) = \tfrac{3}{8}, \quad p_X(4) = \tfrac{1}{4}.$$

Definiamo $Z = h(X)$ con $h(x) = x \bmod 2$ (il resto della divisione di x per 2). Allora $Z(\Omega) = \{0, 1\}$, e:

$$P(Z = 0) = P(X \text{ pari}) = P(X = 2) + P(X = 4) = \tfrac{1}{4} + \tfrac{1}{4} = \tfrac{1}{2},$$
$$P(Z = 1) = P(X \text{ dispari}) = P(X = 1) + P(X = 3) = \tfrac{1}{8} + \tfrac{3}{8} = \tfrac{1}{2}.$$

17.4 Osservazioni sul calcolo del valore atteso

- La formula $E[h(X)] = \sum h(x)p_X(x)$ è estremamente utile per calcoli pratici: evita di determinare la distribuzione di $Z = h(X)$.
- Molti esercizi, come gli Esercizi 1 e 25 del Capitolo 4 del testo di *Ross*, mostrano che il valore atteso di una variabile trasformata può essere ottenuto in modo rapido grazie alla linearità.
- In generale, la formula

  $$E[aX + b] = aE[X] + b$$

  è un caso particolare della pre

18 Varianza e Covarianza di variabili aleatorie

18.1 Definizione di varianza

[Varianza] Sia X una variabile aleatoria discreta con valore atteso finito $E[X]$. La **varianza** di X è definita come

$$\operatorname{Var}(X) := E\left[(X - E[X])^2\right].$$

La varianza misura la *dispersione* dei valori di X attorno al suo valore medio. È sempre non negativa e si annulla solo per variabili costanti.

18.2 Proprietà fondamentali

(1) Non negatività. Poiché $(X-E[X])^2 \geq 0$ sempre, per ogni $\omega \in \Omega$, segue immediatamente che:

$$\operatorname{Var}(X) = E[(X - E[X])^2] \geq 0.$$

(2) Espressione alternativa della varianza.

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

Proof. Sviluppiamo il quadrato:

$$(X - E[X])^2 = X^2 - 2XE[X] + (E[X])^2.$$

Applicando l'operatore di valore atteso:

$$E[(X - E[X])^2] = E[X^2] - 2E[X]E[X] + (E[X])^2 = E[X^2] - (E[X])^2.$$

□

(3) Varianza di una variabile affina. Per ogni $a, b \in \mathbb{R}$ si ha:

$$\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X).$$

Proof. Si osserva che $E[aX + b] = aE[X] + b$. Allora:

$$\operatorname{Var}(aX + b) = E[(aX + b - (aE[X] + b))^2] = E[a^2(X - E[X])^2] = a^2 E[(X - E[X])^2].$$

□

18.3 Covarianza

[Covarianza] Siano X e Y due variabili aleatorie. Si definisce la loro **covarianza** come:

$$\operatorname{Cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

La covarianza misura la tendenza delle due variabili a crescere o decrescere insieme. Se $\operatorname{Cov}(X, Y) > 0$, le due variabili tendono ad aumentare insieme; se è negativa, si comportano in modo opposto.

[Formula esplicita]

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

Proof. Sviluppando:

$$E[(X - E[X])(Y - E[Y])] = E[XY - XE[Y] - YE[X] + E[X]E[Y]] = E[XY] - E[X]E[Y].$$

□

18.4 Varianza della somma

Per ogni coppia di variabili aleatorie X, Y si ha:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\ \mathrm{Cov}(X,Y).$$

Proof.

$$\mathrm{Var}(X+Y) = E[(X+Y-E[X+Y])^2] = E[(X-E[X]+Y-E[Y])^2].$$

Sviluppando il quadrato:

$$E[(X-E[X])^2] + E[(Y-E[Y])^2] + 2E[(X-E[X])(Y-E[Y])].$$

Riconoscendo i termini:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y).$$

□

Caso di indipendenza. Se X e Y sono indipendenti, allora $E[XY] = E[X]E[Y]$, per cui

$$\mathrm{Cov}(X,Y) = 0.$$

Segue che

$$\boxed{\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)}.$$

18.5 Esempi fondamentali

18.5.1 Variabile degenere

Se X è degenere, cioè $P(X=c)=1$, allora

$$E[X] = c, \quad \text{e} \quad \mathrm{Var}(X) = E[(X-c)^2] = 0.$$

18.5.2 Variabile di Bernoulli

Sia $X = \mathbf{1}_A$ la variabile indicatrice di un evento A, con $P(A) = p$.

$$E[X] = P(A) = p, \qquad E[X^2] = E[X] = p,$$

perché $X^2 = X$. Pertanto:

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1-p).$$

18.5.3 Covarianza tra indicatori

Siano A e B due eventi e $X = \mathbf{1}_A$, $Y = \mathbf{1}_B$.

$$E[XY] = E[\mathbf{1}_A \mathbf{1}_B] = P(A \cap B),$$

$$E[X]E[Y] = P(A)P(B),$$

quindi:

$$\mathrm{Cov}(X,Y) = P(A \cap B) - P(A)P(B).$$

Varianza della somma di due indicatori.

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y).$$

Sostituendo:

$$\mathrm{Var}(X+Y) = P(A)(1-P(A)) + P(B)(1-P(B)) + 2[P(A\cap B) - P(A)P(B)].$$

Caso di indipendenza. Se A e B sono indipendenti, $P(A\cap B) = P(A)P(B)$, dunque $\mathrm{Cov}(X,Y) = 0$, e

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

18.5.4 Varianza della Binomiale

Siano $A_1, A_2, \ldots, A_n$ eventi indipendenti, ciascuno con $P(A_i) = p$. Sia

$$X = \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \cdots + \mathbf{1}_{A_n}.$$

Allora $X \sim \mathrm{Bin}(n,p)$.

Poiché le $\mathbf{1}_{A_i}$ sono indipendenti:

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(\mathbf{1}_{A_i}) = \sum_{i=1}^{n} p(1-p) = np(1-p).$$

Varianza della Binomiale (spiegato meglio)

Siano $A_1, A_2, \ldots, A_n$ eventi indipendenti, ciascuno con $P(A_i) = p$. Definiamo le variabili indicatrici

$$\mathbf{1}_{A_i} = \begin{cases} 1 & \text{se } A_i \text{ si realizza,} \\ 0 & \text{altrimenti.} \end{cases}$$

e

$$X = \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \cdots + \mathbf{1}_{A_n}.$$

Allora $X \sim \mathrm{Bin}(n,p)$.

Calcoliamo la varianza di X. Usiamo la proprietà che, per variabili casuali *indipendenti*,

$$\mathrm{Var}\Big(\sum_{i=1}^{n} Y_i\Big) = \sum_{i=1}^{n} \mathrm{Var}(Y_i),$$

poiché le covarianze tra variabili indipendenti sono nulle.

Perciò

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(\mathbf{1}_{A_i}).$$

Calcoliamo ora la varianza di una singola indicatrice $\mathbf{1}_{A_i}$. Poiché $\mathbf{1}_{A_i}$ assume solo i valori 0 e 1, vale $\mathbf{1}_{A_i}^2 = \mathbf{1}_{A_i}$. Quindi

$$\mathbb{E}[\mathbf{1}_{A_i}] = P(A_i) = p, \qquad \mathbb{E}[\mathbf{1}_{A_i}^2] = \mathbb{E}[\mathbf{1}_{A_i}] = p.$$

La varianza è

$$\mathrm{Var}(\mathbf{1}_{A_i}) = \mathbb{E}[\mathbf{1}_{A_i}^2] - \big(\mathbb{E}[\mathbf{1}_{A_i}]\big)^2 = p - p^2 = p(1-p).$$

Sostituendo nella somma otteniamo

$$\operatorname{Var}(X)=\sum_{i=1}^{n} p(1-p)=np(1-p).$$

Un'alternativa (equivalente) è calcolare $\operatorname{Var}(X)=\mathbb{E}[X^2]-\mathbb{E}[X]^2$ espandendo X^2:

$$X^2=\sum_{i=1}^{n}\mathbf{1}_{A_i}^2+2\sum_{1\le i<j\le n}\mathbf{1}_{A_i}\mathbf{1}_{A_j}.$$

Prendendo l'aspettazione e usando $\mathbf{1}_{A_i}^2=\mathbf{1}_{A_i}$ e l'indipendenza ($\mathbb{E}[\mathbf{1}_{A_i}\mathbf{1}_{A_j}]=P(A_i\cap A_j)=p^2$ per $i\neq j$):

$$\mathbb{E}[X^2]=\sum_{i=1}^{n}p+2\sum_{1\le i<j\le n}p^2=np+n(n-1)p^2.$$

Poiché $\mathbb{E}[X]=np$, si ottiene

$$\operatorname{Var}(X)=\mathbb{E}[X^2]-\mathbb{E}[X]^2=\big(np+n(n-1)p^2\big)-(np)^2=np(1-p).$$

Quindi la varianza della distribuzione binomiale $\operatorname{Bin}(n,p)$ è

$$\boxed{\operatorname{Var}(X)=np(1-p).}$$

18.5.5 Varianza dell'Ipergeometrica

Sia $X\sim\operatorname{Hyp}(M,m_1;n)$, cioè M oggetti totali, m_1 successi nella popolazione, n estrazioni senza reinserimento. Sia $p=\frac{m_1}{M}$.

Allora, senza dimostrazione per ora,

$$\boxed{\operatorname{Var}(X)=n\,p(1-p)\left(1-\frac{n-1}{M-1}\right)}.$$

- Se $n=1$, la varianza si riduce a $p(1-p)$, come nella Bernoulli.
- Se $n=M$, la varianza è 0, perché si estraggono tutti gli oggetti.

18.6 Commento: significato della varianza e Legge dei Grandi Numeri

La varianza fornisce una misura quantitativa della *variabilità* o *dispersione* dei risultati attorno alla media. È uno strumento fondamentale per analizzare la stabilità delle medie campionarie.

[Legge Debole dei Grandi Numeri — enunciato] Sia $\{X_1,X_2,\dots,X_n\}$ una famiglia di variabili aleatorie indipendenti e identicamente distribuite con media μ e varianza finita σ^2. Sia

$$\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i.$$

Allora

$$\overline{X}_n\xrightarrow{P}\mu,$$

cioè per ogni $\varepsilon>0$, $P(|\overline{X}_n-\mu|>\varepsilon)\to 0$ per $n\to\infty$.

L'intuizione è che, all'aumentare del numero di prove, la media campionaria si concentra attorno alla media teorica, con varianza che tende a 0 come $\mathrm{Var}(\overline{X}_n) = \sigma^2/n$.

Esercizi consigliati: vedere gli Esercizi 19, 40 e 51 del Capitolo 4 di Ross (nei file `Dal libro di ROSS-14nov2024.pdf` e `ROSS-18-dic-2024.pdf`), in particolare l'Esercizio 51 anticipa il caso delle variabili di Poisson, che assumono un'infinità numerabile di valori.

19 Lezione del 14 novembre 2024 – Distribuzione di Poisson e varianza di somme

19.1 Richiami sulla varianza

Varianza della distribuzione uniforme discreta. Sia X una variabile aleatoria uniforme su $\{1, 2, \dots, n\}$, cioè

$$P(X = k) = \frac{1}{n}, \quad k = 1, 2, \dots, n.$$

Allora:

$$E[X] = \frac{1}{n}\sum_{k=1}^{n} k = \frac{n+1}{2},$$

$$E[X^2] = \frac{1}{n}\sum_{k=1}^{n} k^2 = \frac{(n+1)(2n+1)}{6}.$$

Segue:

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{(n^2-1)}{12}.$$

19.2 Varianza della somma di n variabili aleatorie

[Formula generale della varianza di una somma] Siano $X_1, X_2, \dots, X_n$ variabili aleatorie con varianze finite. Allora:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{1 \le i < j \le n} \mathrm{Cov}(X_i, X_j).$$

Proof. Per induzione o applicando la formula $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$ in modo iterato. □

Formula generale della varianza di una somma

Siano $X_1, X_2, \dots, X_n$ variabili aleatorie con varianze finite. Allora vale

$$\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathrm{Var}(X_i) \; + \; 2\sum_{1 \le i < j \le n} \mathrm{Cov}(X_i, X_j).$$

Proof. Ricordiamo che per due variabili aleatorie X e Y

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

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

Dimostrazione per induzione. Per $n = 1$ la formula è banale. Supponiamo vera la formula per $n-1$, cioè

$$\mathrm{Var}\Big(\sum_{i=1}^{n-1} X_i\Big) = \sum_{i=1}^{n-1} \mathrm{Var}(X_i) + 2 \sum_{1\le i<j\le n-1} \mathrm{Cov}(X_i, X_j).$$

Applicando la formula per la somma di due variabili a $\sum_{i=1}^{n-1} X_i$ e X_n otteniamo

$$\begin{aligned}
\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) &= \mathrm{Var}\Big(\sum_{i=1}^{n-1} X_i\Big) + \mathrm{Var}(X_n) + 2\,\mathrm{Cov}\left(\sum_{i=1}^{n-1} X_i,\ X_n\right) \\
&= \sum_{i=1}^{n-1} \mathrm{Var}(X_i) + 2 \sum_{1\le i<j\le n-1} \mathrm{Cov}(X_i, X_j) + \mathrm{Var}(X_n) + 2\sum_{i=1}^{n-1} \mathrm{Cov}(X_i, X_n) \\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1\le i<j\le n} \mathrm{Cov}(X_i, X_j),
\end{aligned}$$

che è esattamente la formula voluta. Questo completa il passo induttivo.

Dimostrazione iterativa (alternativa). Si applica ripetutamente la formula $\mathrm{Var}(U+V) = \mathrm{Var}(U) + \mathrm{Var}(V) + 2\,\mathrm{Cov}(U, V)$ aggiungendo le variabili una ad una; i termini di covarianza si accumulano e si ottiene la stessa espressione.

Osservazione. Se le X_i sono indipendenti allora tutte le covarianze $\mathrm{Cov}(X_i, X_j)$ per $i \neq j$ sono nulle e la formula si riduce a

$$\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

□

19.3 Applicazione: Varianza della variabile Ipergeometrica

Sia $X \sim \mathrm{Hyp}(M, m_1; n)$, dove:

- M = numero totale di oggetti nella popolazione,
- m_1 = numero di "successi" nella popolazione,
- n = numero di estrazioni senza reinserimento.

Si definiscono le variabili indicatrici:

$$X_i = \begin{cases} 1, & \text{se l'}i\text{-esima estrazione è un successo,} \\ 0, & \text{altrimenti.} \end{cases}$$

Allora $X = \sum_{i=1}^{n} X_i$, ma le X_i *non sono indipendenti.*

Usando la formula generale:

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j),$$

e calcolando le covarianze negative dovute all'assenza di reinserimento, si ottiene la nota formula:

$$\boxed{\mathrm{Var}(X) = np(1-p)\left(1 - \frac{n-1}{M-1}\right)}, \quad \text{dove } p = \frac{m_1}{M}.$$

- Se $n = 1$, si riduce a $p(1-p)$, come la Bernoulli.
- Se $n = M$, la varianza è nulla.

Varianza dell'ipergeometrica (senza reinserimento)

Sia una popolazione di dimensione M contenente m_1 oggetti di tipo "successo" e $M - m_1$ oggetti di tipo "insuccesso". Si estrae un campione di dimensione n *senza reinserimento*. Indichiamo con X il numero di successi nel campione.

Allora $X \sim \text{Ipergeo}(M, m_1, n)$ e la probabilità di successo in un singolo estratto è

$$p = \frac{m_1}{M}.$$

Calcolo della varianza (esplicitiamo i calcoli). Se le estrazioni fossero con reinserimento (caso binomiale), avremmo

$$\text{Var}(X_{\text{binomiale}}) = np(1-p).$$

Nel caso senza reinserimento, le variabili indicatrici

$$\mathbf{1}_i = \begin{cases} 1 & \text{se l'estrazione } i \text{ è un successo,} \\ 0 & \text{altrimenti,} \end{cases} \qquad i = 1, \dots, n$$

non sono più indipendenti.

Allora, applicando la formula generale della varianza di una somma:

$$\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(\mathbf{1}_i) + 2 \sum_{1 \le i < j \le n} \text{Cov}(\mathbf{1}_i, \mathbf{1}_j),$$

dove:

$$\text{Var}(\mathbf{1}_i) = p(1-p), \quad \text{Cov}(\mathbf{1}_i, \mathbf{1}_j) = -\frac{p(1-p)}{M-1} \quad (i \neq j).$$

Infatti, la covarianza negativa riflette il fatto che se un successo è già stato estratto, la probabilità di un successo successivo diminuisce:

$$\text{Cov}(\mathbf{1}_i, \mathbf{1}_j) = \mathbb{E}[\mathbf{1}_i \mathbf{1}_j] - \mathbb{E}[\mathbf{1}_i]\mathbb{E}[\mathbf{1}_j] = \frac{m_1}{M} \cdot \frac{m_1 - 1}{M-1} - \left(\frac{m_1}{M}\right)^2 = -\frac{p(1-p)}{M-1}.$$

Sostituendo nella formula generale della varianza:

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^{n} p(1-p) + 2 \sum_{1 \le i < j \le n} \left(-\frac{p(1-p)}{M-1}\right) \\ &= np(1-p) - 2\frac{p(1-p)}{M-1} \cdot \frac{n(n-1)}{2} \\ &= np(1-p)\left(1 - \frac{n-1}{M-1}\right). \end{aligned}$$

Nella formula

$$2 \sum_{1 \le i < j \le n} \text{Cov}(\mathbf{1}_i, \mathbf{1}_j),$$

la sommatoria $\sum_{1 \le i < j \le n} 1$ conta **quante coppie distinte (i, j) con $i < j$ esistono** tra le n estrazioni.

Il numero di coppie distinte è dato dal coefficiente binomiale:

$$\binom{n}{2} = \frac{n(n-1)}{2}.$$

Poiché la covarianza negativa è costante tra tutte le coppie,

$$\sum_{1 \le i < j \le n} \text{Cov}(\mathbf{1}_i, \mathbf{1}_j) = \frac{n(n-1)}{2} \left(-\frac{p(1-p)}{M-1}\right).$$

Moltiplicando per 2, come previsto dalla formula generale della varianza, otteniamo

$$2 \sum_{1 \le i < j \le n} \mathrm{Cov}(\mathbf{1}_i, \mathbf{1}_j) = -\frac{n(n-1)}{M-1}\, p(1-p),$$

che è il termine di correzione dovuto alla dipendenza tra le estrazioni senza reinserimento.

In altre parole, $\frac{n(n-1)}{2}$ rappresenta **il numero di volte che la sommatoria sulle coppie viene ripetuta**.

Conclusione. Quindi, la varianza dell'ipergeometrica è

$$\boxed{\mathrm{Var}(X) = np(1-p)\frac{M-n}{M-1}}, \qquad p = \frac{m_1}{M}.$$

19.4 Primo esempio di variabile con infiniti valori: la distribuzione di Poisson

Motivazione. Consideriamo una sequenza di prove di Bernoulli con parametro p, e sia

$$X_n \sim \mathrm{Bin}(n, p_n),$$

dove p_n dipende da n in modo che $np_n \to \lambda > 0$ per $n \to \infty$. Allora X_n rappresenta il numero di successi in n prove, ciascuna di probabilità p_n.

19.5 Approssimazione di Poisson (Legge dei piccoli numeri)

[Approssimazione di Poisson] Sia $X_n \sim \mathrm{Bin}(n, p_n)$ con $np_n \to \lambda > 0$. Allora, per ogni $k \in \mathbb{N}$,

$$\lim_{n\to\infty} P(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Proof. Abbiamo:

$$P(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}.$$

Scriviamo $p_n = \lambda/n$, e calcoliamo il limite:

$$P(X_n = k) = \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Ora, quando $n \to \infty$,

$$\frac{n!}{(n-k)!n^k} = \frac{n(n-1)\cdots(n-k+1)}{n^k} \to 1,$$

e

$$\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}, \quad \left(1 - \frac{\lambda}{n}\right)^{-k} \to 1.$$

Pertanto:

$$\lim_{n\to\infty} P(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

□

[Distribuzione di Poisson] Si dice che una variabile aleatoria X ha **distribuzione di Poisson** di parametro $\lambda > 0$ se

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

e si scrive

$$X \sim \mathrm{Poisson}(\lambda).$$

- X può assumere infiniti valori ($|X(\Omega)| = \mathbb{N}$).
- La distribuzione di Poisson è il limite di una Binomiale con $n \to \infty$ e $p \to 0$ mantenendo $np = \lambda$ costante.

19.6 Teorema di Le Cam (enunciato)

[Le Cam, enunciato] Sia $S_n = \sum_{i=1}^n X_i$, dove le X_i sono variabili di Bernoulli indipendenti con parametri p_i. Se $\lambda = \sum_{i=1}^n p_i$, allora

$$d_{TV}\left(\mathcal{L}(S_n), \text{Poisson}(\lambda)\right) \leq \sum_{i=1}^n p_i^2,$$

dove d_{TV} è la distanza di variazione totale tra distribuzioni.

Questo risultato giustifica l'uso della Poisson come approssimazione della somma di eventi rari (legge dei piccoli numeri).

19.7 Valore atteso e varianza della Poisson

Sia $X \sim \text{Poisson}(\lambda)$, cioè

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

Valore atteso.

$$E[X] = \sum_{k=0}^{\infty} k\, e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\lambda e^{\lambda} = \lambda.$$

Varianza. Calcoliamo $E[X^2]$:

$$E[X^2] = \sum_{k=0}^{\infty} k^2 e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} k(k-1)\frac{\lambda^k}{k!} + e^{-\lambda}\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}.$$

Perché sia $X \sim \text{Poisson}(\lambda)$, cioè

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

Per definizione, il secondo momento di X è

$$\mathbb{E}[X^2] = \sum_{k=0}^{\infty} k^2\, P(X = k).$$

Sostituendo la probabilità di massa della Poisson otteniamo

$$\mathbb{E}[X^2] = \sum_{k=0}^{\infty} k^2\, e^{-\lambda}\frac{\lambda^k}{k!}.$$

In generale, per una variabile discreta X con valori $k = 0, 1, 2, \ldots$,

$$\mathbb{E}[f(X)] = \sum_{k=0}^{\infty} f(k)\, P(X = k),$$

qui prendendo $f(k) = k^2$ otteniamo esattamente la somma sopra.

e perché: Partiamo dal secondo momento di $X \sim \text{Poisson}(\lambda)$:

$$\mathbb{E}[X^2] = \sum_{k=0}^{\infty} k^2 \, e^{-\lambda} \frac{\lambda^k}{k!}.$$

Notiamo che

$$k^2 = k(k-1) + k.$$

Quindi possiamo riscrivere la somma come due termini separati:

$$\sum_{k=0}^{\infty} k^2 \, e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} \big(k(k-1) + k\big) \, e^{-\lambda} \frac{\lambda^k}{k!} = \underbrace{\sum_{k=0}^{\infty} k(k-1) \, e^{-\lambda} \frac{\lambda^k}{k!}}_{(1)} + \underbrace{\sum_{k=0}^{\infty} k \, e^{-\lambda} \frac{\lambda^k}{k!}}_{(2)}.$$

In altre parole:

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

Nota: - Nel primo termine, $k(k-1) = 0$ per $k = 0$ e $k = 1$, quindi la somma effettiva parte da $k = 2$. - Nel secondo termine, la somma può partire da $k = 1$ perché il termine $k = 0$ è zero.

Questa scomposizione è utile perché permette di semplificare i termini usando proprietà della serie esponenziale, e ottenere facilmente

$$\mathbb{E}[X^2] = \lambda^2 + \lambda.$$

Osserviamo che:

$$\sum_{k=1}^{\infty} k(k-1) \frac{\lambda^k}{k!} = \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 e^{\lambda},$$

e

$$\sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} = \lambda e^{\lambda}.$$

Quindi:

$$E[X^2] = e^{-\lambda}(\lambda^2 e^{\lambda} + \lambda e^{\lambda}) = \lambda^2 + \lambda.$$

Segue:

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

$$\boxed{\begin{array}{c} \text{Per } X \sim \text{Poisson}(\lambda): \\[1em] E[X] = \lambda, \qquad \text{Var}(X) = \lambda. \end{array}}$$

19.8 Osservazioni e applicazioni

- La Poisson descrive il numero di eventi rari in un intervallo di tempo o spazio (arrivi, difetti, incidenti, ecc.).
- È il limite naturale della Binomiale per eventi rari e numerosi.
- È una delle principali distribuzioni discrete a supporto infinito.

Esercizi consigliati: Esercizi 52 e 53 del Capitolo 4 di *Ross*, nei file `Dal libro di ROSS-14nov2024.pdf` e `ROSS-18-dic-2024.pdf`.

20 Lezione del 15 novembre 2024 – Variabili geometriche e tempi di successo

20.1 Uniformi discrete e traslazioni

Uniforme discreta su $\{1, \dots, n\}$: momento atteso e varianza

Sia X una variabile aleatoria uniforme su $\{1, 2, \dots, n\}$, cioè

$$P(X = k) = \frac{1}{n}, \qquad k = 1, \dots, n.$$

Calcolo dell'atteso. Per definizione,

$$\mathbb{E}[X] = \sum_{k=1}^{n} k\, P(X = k) = \sum_{k=1}^{n} k \cdot \frac{1}{n} = \frac{1}{n} \sum_{k=1}^{n} k.$$

La somma dei primi n interi è

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

Quindi

$$\mathbb{E}[X] = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

Calcolo della varianza. Usiamo la definizione

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Prima calcoliamo il secondo momento:

$$\mathbb{E}[X^2] = \sum_{k=1}^{n} k^2\, P(X = k) = \frac{1}{n} \sum_{k=1}^{n} k^2.$$

La somma dei quadrati è nota:

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Dunque

$$\mathbb{E}[X^2] = \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}.$$

Ora sottraiamo il quadrato dell'atteso:

$$\mathrm{Var}(X) = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2.$$

Mettiamo in evidenza $(n+1)$:

$$\mathrm{Var}(X) = (n+1)\left(\frac{2n+1}{6} - \frac{n+1}{4}\right).$$

Troviamo il minimo comune multiplo (12):

$$\frac{2n+1}{6} = \frac{4n+2}{12}, \qquad \frac{n+1}{4} = \frac{3(n+1)}{12} = \frac{3n+3}{12}.$$

Sottraiamo:

$$\frac{4n+2}{12} - \frac{3n+3}{12} = \frac{n-1}{12}.$$

Quindi

$$\mathrm{Var}(X) = (n+1)\frac{n-1}{12} = \frac{n^2-1}{12}.$$

Conclusione. Per una uniforme discreta su $\{1, \dots, n\}$:

$$\boxed{\mathbb{E}[X] = \frac{n+1}{2}, \qquad \mathrm{Var}(X) = \frac{n^2-1}{12}.}$$

Uniforme su $\{0, 1, \dots, m\}$. Definiamo Y uniforme su $\{0, 1, \dots, m\}$. Allora Y ha la stessa distribuzione di $X - 1$, dove X è uniforme su $\{1, 2, \dots, m+1\}$.

$$\mathbb{E}[Y] = \mathbb{E}[X-1] = \mathbb{E}[X] - 1 = \frac{m+2}{2} - 1 = \frac{m}{2},$$
$$\mathrm{Var}(Y) = \mathrm{Var}(X-1) = \mathrm{Var}(X) = \frac{(m+1)^2-1}{12} = \frac{m(m+2)}{12}.$$

20.2 Formula alternativa per il valore atteso

Per una variabile aleatoria X a valori interi non negativi (cioè $X(\Omega) \subseteq \{0, 1, 2, \dots, n\}$), vale la formula

$$\mathbb{E}[X] = \sum_{k=0}^{n-1} \mathbb{P}(X > k).$$

Dimostrazione nel caso finito. Poiché X assume valori in $\{0, 1, \dots, n\}$, si ha

$$\mathbb{E}[X] = \sum_{h=0}^{n} h\, \mathbb{P}(X = h).$$

Scriviamo $h = \sum_{k=0}^{h-1} 1$ e scambiamo l'ordine delle somme:

$$\mathbb{E}[X] = \sum_{h=0}^{n} \sum_{k=0}^{h-1} \mathbb{P}(X = h) = \sum_{k=0}^{n-1} \sum_{h>k} \mathbb{P}(X = h) = \sum_{k=0}^{n-1} \mathbb{P}(X > k).$$

□

Nel caso generale, cioè quando X è una variabile intera non negativa non necessariamente limitata superiormente, la stessa identità vale come serie infinita:

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} \mathbb{P}(X > k),$$

purché la serie a destra converga (ovvero $\mathbb{E}[X] < \infty$).

Una giustificazione breve: applicando lo stesso scambio di somme su somme potenzialmente infinite e usando il teorema della convergenza monotona (o la definizione dell'aspettazione per variabili non negative), si ottiene la formula sopra. Viceversa, se la serie $\sum_{k=0}^{\infty} \mathbb{P}(X > k)$ converge, allora la sua somma è esattamente $\mathbb{E}[X]$.

20.3 Tempo di primo successo (variabile geometrica)

Sia $\{E_n\}_{n \geq 1}$ una successione di eventi indipendenti con

$$P(E_n) = p, \qquad 0 < p < 1.$$

Definiamo

$$T := \text{numero di prove necessarie per ottenere il primo successo.}$$

Allora T è detta **variabile geometrica di parametro** p.

[Distribuzione geometrica] Una variabile aleatoria X si dice **geometrica di parametro** p se

$$P(X = k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \dots$$

$$P(X > k) = (1-p)^k.$$

Si scrive

$$X \sim \mathrm{Geom}(p).$$

Interpretazione. X rappresenta il numero di prove fino al primo successo in una sequenza di prove di Bernoulli indipendenti con probabilità di successo p.

20.4 Proprietà di mancanza di memoria

[Mancanza di memoria] Per ogni $h, k \geq 0$ vale:

$$P(X - h > k \mid X > h) = P(X > k).$$

Proof. Poiché $P(X > h) = (1-p)^h$, abbiamo:

$$P(X - h > k, X > h) = P(X > h + k) = (1-p)^{h+k}.$$

Quindi:

$$P(X - h > k \mid X > h) = \frac{P(X > h+k)}{P(X > h)} = \frac{(1-p)^{h+k}}{(1-p)^h} = (1-p)^k = P(X > k).$$

□

20.5 Valore atteso e varianza della Geometrica

Valore atteso. Usando la formula $E[X] = \sum_{k \geq 0} P(X > k)$:

$$E[X] = \sum_{k=0}^{\infty} (1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

Varianza. Senza dimostrazione (vedi [SN] o [SCHEMA-VARIABILI-ALEATORIE-19dicembre2024.pdf]):

$$\boxed{\mathrm{Var}(X) = \frac{1-p}{p^2}.}$$

20.6 Numero di insuccessi prima del primo successo

Definiamo

$$T' := \text{numero di insuccessi prima del primo successo.}$$

Allora $T' = X - 1$, e quindi

$$P(T' = h) = P(X = h+1) = (1-p)^h p, \quad h \geq 0.$$

T' è detta *variabile geometrica a supporto* $\{0, 1, 2, \dots\}$. In tal caso:

$$E[T'] = \frac{1-p}{p}, \qquad \mathrm{Var}(T') = \frac{1-p}{p^2}.$$

20.7 Tempi di successi successivi

Sia T_r il **tempo del r-esimo successo**, ossia il numero di prove necessarie per ottenere r successi. Sia S_k il numero di successi dopo k prove: $S_k = \sum_{i=1}^{k} \mathbf{1}_{E_i}$. Allora:

$$T_r = \min\{k \geq 1 : S_k = r\}.$$

[Distribuzione del tempo del r-esimo successo]

$$P(T_r = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, r+2, \dots$$

Proof. Perché il r-esimo successo avvenga alla k-esima prova, è necessario che nei primi $k-1$ tentativi vi siano esattamente $r-1$ successi e che la k-esima prova sia un successo:

$$P(T_r = k) = P(S_{k-1} = r-1, E_k) = P(S_{k-1} = r-1)P(E_k),$$

dove $S_{k-1} \sim \text{Bin}(k-1, p)$. Quindi:

$$P(T_r = k) = \binom{k-1}{r-1} p^{r-1} (1-p)^{k-r} \cdot p.$$

□

Tempo di r-esimo successo in termini di insuccessi. Definiamo

$$T'_r := T_r - r,$$

ossia il numero di insuccessi prima dell'r-esimo successo. Allora:

$$P(T'_r = h) = P(T_r = h + r) = \binom{h+r-1}{r-1} p^r (1-p)^h, \quad h \geq 0.$$

Valore atteso. Poiché T_r può essere visto come somma di r variabili geometriche indipendenti (ognuna con media $1/p$):

$$E[T_r] = r\, E[\text{Geom}(p)] = \frac{r}{p}.$$

Osservazione. Le variabili T_r e T'_r sono esempi di **distribuzioni negative binomiali**. Infatti, T'_r rappresenta il numero di insuccessi prima del r-esimo successo.

20.8 Riferimenti ed esercizi

- Esercizio 71 di *Ross, Cap. 4*: introduzione alla Geometrica come tempo di primo successo.
- Esercizio 74 di *Ross, Cap. 4*: tempo del secondo successo T_2.
- Esercizio 75 di *Ross, Cap. 4*: tempo del r-esimo successo.

Per $X \sim \text{Geom}(p)$:

$$E[X] = \frac{1}{p}, \qquad \text{Var}(X) = \frac{1-p}{p^2}, \qquad P(X > k) = (1-p)^k.$$

Proprietà di mancanza di memoria:

$$P(X - h > k \mid X > h) = P(X > k).$$

21 Lezioni del 20 e 21 novembre 2024 – Densità congiunta, indipendenza e giochi aleatori

21.1 Densità discreta congiunta di due variabili aleatorie

Siano X e Y due variabili aleatorie definite sullo stesso spazio di probabilità $(\Omega, \mathcal{E}, P)$, con:

$$X(\Omega) = \{x_1, x_2, \ldots, x_n\}, \qquad Y(\Omega) = \{y_1, y_2, \ldots, y_m\}.$$

[Densità discreta congiunta] La **densità discreta congiunta** di X e Y è la funzione

$$p_{X,Y}(x_i, y_j) := P(X = x_i,\, Y = y_j), \qquad x_i \in X(\Omega),\ y_j \in Y(\Omega),$$

che soddisfa:

$$p_{X,Y}(x_i, y_j) \geq 0, \qquad \sum_{i=1}^{n} \sum_{j=1}^{m} p_{X,Y}(x_i, y_j) = 1.$$

La distribuzione congiunta può essere rappresentata in una *tabella delle probabilità congiunte*:

$X \backslash Y$	y_1	y_2	$\dots$	y_m	Somma su Y
x_1	p_{11}	p_{12}	$\dots$	p_{1m}	$p_X(x_1)$
x_2	p_{21}	p_{22}	$\dots$	p_{2m}	$p_X(x_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_n	p_{n1}	p_{n2}	$\dots$	p_{nm}	$p_X(x_n)$
Somma su X	$p_Y(y_1)$	$p_Y(y_2)$	$\dots$	$p_Y(y_m)$	1

[Densità marginali] Le **densità marginali** di X e Y si ottengono per somma rispetto all'altra variabile:

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

21.2 Indipendenza di due variabili aleatorie

[Indipendenza] Le variabili aleatorie X e Y si dicono **indipendenti** se, per ogni coppia di valori $x_i \in X(\Omega)$ e $y_j \in Y(\Omega)$, vale:

$$P(X = x_i, Y = y_j) = P(X = x_i)\, P(Y = y_j).$$

Equivalente a:

$$p_{X,Y}(x_i, y_j) = p_X(x_i)\, p_Y(y_j).$$

In termini di partizioni:

$$\mathcal{H}_X = \{H^X_{x_i} := \{X = x_i\}\}_{i=1}^{n}, \qquad \mathcal{H}_Y = \{H^Y_{y_j} := \{Y = y_j\}\}_{j=1}^{m},$$

X e Y sono indipendenti se e solo se le partizioni $\mathcal{H}_X$ e $\mathcal{H}_Y$ sono indipendenti.

Osservazione. Se esistono x_i e y_j tali che $p_X(x_i) > 0$, $p_Y(y_j) > 0$ ma $p_{X,Y}(x_i, y_j) = 0$, allora X e Y *non* sono indipendenti.

21.3 Valore atteso di una funzione di due variabili aleatorie

Sia $g : \mathbb{R}^2 \to \mathbb{R}$ una funzione nota. Il valore atteso della variabile $Z = g(X, Y)$ è definito da:

$$E[g(X, Y)] = \sum_{i=1}^{n} \sum_{j=1}^{m} g(x_i, y_j)\, p_{X,Y}(x_i, y_j).$$

Formula utile in caso di indipendenza (senza dimostrazione). Se X e Y sono indipendenti, allora:

$$E[g(X,Y)] = \sum_{i=1}^{n}\sum_{j=1}^{m} g(x_i, y_j)\, p_X(x_i)\, p_Y(y_j).$$

21.4 Proprietà fondamentali in caso di indipendenza

Se X e Y sono indipendenti e $h_1, h_2 : \mathbb{R} \to \mathbb{R}$ sono funzioni, allora:

$$E[h_1(X)h_2(Y)] = E[h_1(X)]\, E[h_2(Y)].$$

Proof. Per definizione:

$$E[h_1(X)h_2(Y)] = \sum_{i,j} h_1(x_i)h_2(y_j)p_{X,Y}(x_i, y_j).$$

Se X e Y sono indipendenti:

$$p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j).$$

Segue:

$$E[h_1(X)h_2(Y)] = \left(\sum_i h_1(x_i)p_X(x_i)\right)\left(\sum_j h_2(y_j)p_Y(y_j)\right) = E[h_1(X)]E[h_2(Y)].$$

□

Caso particolare. Prendendo $h_1(x) = x$ e $h_2(y) = y$, si ottiene:

$$E[XY] = E[X]E[Y].$$

Pertanto:

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = 0.$$

Conseguenza. Se X e Y sono indipendenti:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Più in generale, se $X_1, \ldots, X_n$ sono indipendenti:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Caso di variabili identicamente distribuite. Se $X_1, \ldots, X_n$ sono indipendenti e identicamente distribuite, con varianza comune σ^2, allora la media aritmetica

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

ha varianza:

$$\mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}.$$

Questo risultato è alla base della *Legge dei Grandi Numeri*.

21.5 Esempi ed applicazioni

21.5.1 Esempio 9.5 di [SN] – La lotteria

Sia un gioco in cui un biglietto costa 1 euro e vi sono premi distribuiti casualmente. Definiamo X come il **guadagno netto** del giocatore (premio vinto meno costo del biglietto). Allora:

$$E[X] = \text{valore medio delle vincite} - 1.$$

Il valore atteso rappresenta il prezzo "equo" del biglietto: se $E[X] = 0$, il gioco è equo; se $E[X] < 0$, è sfavorevole per il giocatore.

21.5.2 Osservazione 9.3 di [SN] – Prezzo equo e valore atteso

Il valore atteso $E[X]$ può essere interpretato come il prezzo giusto da pagare per una scommessa: è la somma che, in media, mantiene equilibrio fra rischio e rendimento nel lungo periodo.

21.5.3 Esempio 9.7 di [SN] – Gioco al raddoppio e paradosso di San Pietroburgo

Nel gioco del raddoppio (o *Martingala*), un giocatore scommette 1 euro, e ogni volta che perde raddoppia la posta finché non vince. Alla prima vittoria guadagna 1 euro, quindi "vince sempre".

Tuttavia, il numero di scommesse fino alla prima vittoria segue una distribuzione geometrica $T \sim \text{Geom}(1/2)$, e la somma totale scommessa prima della vittoria è

$$X = 1 + 2 + 4 + \cdots + 2^{T-1} = 2^T - 1.$$

Il valore atteso:

$$E[X] = \sum_{k=1}^{\infty}(2^k - 1)P(T = k) = \sum_{k=1}^{\infty}(2^k - 1)\frac{1}{2^k} = \infty.$$

$$\boxed{E[X] = \infty.}$$

Questo paradosso, noto come **Paradosso di San Pietroburgo**, mostra come un guadagno certo (1 euro) possa essere associato a una variabile con valore atteso infinito — un caso in cui il valore atteso non rappresenta più un criterio realistico di scelta.

Riferimenti: Lezioni 7–8 di [SN]; file `ESERCIZI-LEZ-7SOLUZIONI.pdf` e `SCHEMA-VARIABILI-ALEATORIE-19d`

22 Lezione del 22 novembre 2024 – Esercizi sulla distribuzione congiunta e indipendenza

22.1 Esercizi sulla distribuzione congiunta

Abbiamo svolto e discusso gli esercizi della prova del **16 luglio 2020** (`CP-INFORM-16-luglio-2020-corretto` in particolare i punti i), ii), iii), iv) e il calcolo del valore atteso e della varianza al punto v).

Nota. Nel testo originale era presente un errore di stampa nella domanda v)(a):

$$|S_{100}/100 - 1/2| \quad \text{CORRETTO IN} \quad |S_{100}/100 - 2|.$$

Tale correzione è riportata nel file aggiornato `CP-INFORM-16-luglio-2020-corretto.pdf`.

Gli esercizi avevano lo scopo di:

- esercitarsi nel calcolo della **densità discreta congiunta** $p_{X,Y}(x,y)$ e delle **marginali** $p_X(x)$ e $p_Y(y)$;

- utilizzare tabelle di probabilità per calcolare in modo rapido il valore atteso di una **trasformazione** $Z = g(X, Y)$;
- determinare la distribuzione (densità discreta) della variabile Z;
- calcolare $E[Z]$ e $\mathrm{Var}(Z)$.

22.2 Valore atteso di una trasformazione di due variabili

Come discusso nella **Proposizione 9.13** e nell'**Esempio 9.8** di [SN, pagg. 128–129], se $Z = g(X, Y)$ e (X, Y) è una coppia di variabili aleatorie discrete con densità congiunta $p_{X,Y}$, allora:

$$E[Z] = E[g(X, Y)] = \sum_x \sum_y g(x, y)\, p_{X,Y}(x, y).$$

Inoltre, se si vogliono ottenere i valori di probabilità della nuova variabile Z, si può utilizzare la formula:

$$P(Z = z) = \sum_{(x,y):\, g(x,y)=z} p_{X,Y}(x, y).$$

Esempio. Siano X, Y con valori in $\{1, 2\}$ e densità congiunta:

$$p_{X,Y} = \begin{array}{c|cc} & Y=1 & Y=2 \\ \hline X=1 & 1/4 & 1/4 \\ X=2 & 1/4 & 1/4 \end{array}$$

Sia $Z = X + Y$. Allora:

$$E[Z] = \sum_{x,y} (x + y) p_{X,Y}(x, y) = 2\left(\frac{1}{4}\right) + 3\left(\frac{1}{2}\right) + 4\left(\frac{1}{4}\right) = 3.$$

La densità discreta di Z è:

$$P(Z = 2) = \tfrac{1}{4}, \quad P(Z = 3) = \tfrac{1}{2}, \quad P(Z = 4) = \tfrac{1}{4}.$$

22.3 Indipendenza e Covarianza

Abbiamo approfondito la relazione tra indipendenza e covarianza.

Richiamo. Se X e Y sono indipendenti, allora:

$$E[XY] = E[X]E[Y] \quad \Rightarrow \quad \mathrm{Cov}(X, Y) = 0.$$

Tuttavia, l'implicazione inversa *non è vera in generale*:

$$\mathrm{Cov}(X, Y) = 0 \;\not\Rightarrow\; X, Y \text{ indipendenti}.$$

22.4 Controesempi

Controesempio 1. Siano X, Y con valori in $\{-1, 0, 1\}$ e distribuzione definita da:

$$P(X = -1, Y = -1) = P(X = -1, Y = 1) = P(X = 1, Y = -1) = P(X = 1, Y = 1) = P(X = 0, Y = 0) = \tfrac{1}{5}$$

e $P(X = i, Y = j) = 0$ per tutti gli altri (i, j).

Si verifica che:

$$E[X] = E[Y] = 0, \qquad E[XY] = 0,$$

da cui $\mathrm{Cov}(X, Y) = 0$.

Tuttavia X e Y *non sono indipendenti*, perché, ad esempio:

$$P(X = 1, Y = 0) = 0 \quad \text{mentre} \quad P(X = 1)P(Y = 0) = \tfrac{2}{5} \cdot \tfrac{1}{5} > 0.$$

Controesempio 2. Sia X una variabile aleatoria uniforme su $\{-1,0,1\}$ e sia $Y = X^2$. Allora

$$P(X=-1) = P(X=0) = P(X=1) = \frac{1}{3}.$$

Poiché $Y = X^2$, i possibili valori congiunti (X,Y) sono

$$(-1,1),\ (0,0),\ (1,1), \quad \text{tutti con probabilità } \frac{1}{3}.$$

Da ciò derivano le marginali

$$P(Y=1) = \frac{2}{3}, \qquad P(Y=0) = \frac{1}{3}.$$

Calcoliamo ora le aspettative:

$$E[X] = \frac{-1+0+1}{3} = 0, \qquad E[Y] = E[X^2] = \frac{1+0+1}{3} = \frac{2}{3}.$$

Inoltre

$$E[XY] = E[X^3] = \frac{(-1)^3 + 0^3 + 1^3}{3} = \frac{-1+0+1}{3} = 0.$$

Segue quindi

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = 0 - 0 \cdot \frac{2}{3} = 0.$$

Tuttavia, X e Y non sono indipendenti. Infatti,

$$P(X=0,\, Y=1) = 0,$$

poiché se $X = 0$ allora necessariamente $Y = 0$. D'altra parte,

$$P(X=0)\, P(Y=1) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9} > 0.$$

Pertanto,

$$P(X=0, Y=1) \neq P(X=0)P(Y=1),$$

e ne segue che X e Y non sono indipendenti nonostante $\mathrm{Cov}(X,Y) = 0$.

Osservazione (vedi anche Osservazione 10.4 in [SN, p. 138]). Questi esempi mostrano che:

$$\text{indipendenza } \Rightarrow \mathrm{Cov}(X,Y) = 0, \quad \text{ma non viceversa.}$$

In particolare, l'assenza di correlazione non implica indipendenza stocastica.

22.5 Dipendenza funzionale e indipendenza stocastica

Esempio. Nel Controesempio 2, se si restringe il dominio di X a $\{-1,+1\}$, allora $Y = X^2 = 1$ è una variabile *degenere*. Poiché Y assume un unico valore con probabilità 1, risulta indipendente da ogni altra variabile:

$$P(X=-1, Y=1) = P(X=-1) = P(X=-1)P(Y=1), \quad P(X=+1, Y=1) = P(X=+1)P(Y=1).$$

Questo mostra che una *dipendenza funzionale* non implica necessariamente una *dipendenza stocastica*, se la funzione è costante.

Osservazione generale. Un evento di probabilità 1 (o 0) è indipendente da qualunque altro evento.

Verifica. Siano A, B due eventi e supponiamo $P(B) = 1$. Allora $P(B^c) = 0$ e quindi:

$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A \cap B) + 0 = P(A \cap B).$$

Ne segue:

$$P(A \cap B) = P(A) = P(A) \cdot 1 = P(A)P(B).$$

$\square$

Conclusione. Eventi con probabilità 0 o 1 sono indipendenti da ogni altro evento. Tuttavia, questa è una forma "banale" di indipendenza: non fornisce informazioni sulla relazione stocastica effettiva tra le variabili.

Riferimenti:

- Esercizio 2 del *Compito del 16 luglio 2020*, file `CP-INFORM-16-luglio-2020-corretto.pdf`.
- *Proposizione 9.13* e *Esempio 9.8* in [SN], pagg. 128–129.
- Osservazione 10.4 di [SN], pag. 138.

Anticipazione. Nella prossima lezione (venerdì 29 novembre) verranno introdotti:

- la **disuguaglianza di Chebyshev**;
- la **Legge Debole dei Grandi Numeri**.

23 Lezione del 27 novembre 2024 – Disuguaglianza di Chebyshev e applicazioni

23.1 Proposizione 10.8 (Disuguaglianza di Chebyshev)

[Disuguaglianza di Chebyshev] Sia X una variabile aleatoria con valore atteso $\mu = E[X]$ e varianza $\sigma^2 = \mathrm{Var}(X)$. Allora, per ogni $\varepsilon > 0$,

$$P(|X - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Equivalentemente,

$$P\big(|X - E(X)| > \varepsilon\big) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

Osservazioni.

- La disuguaglianza richiede $\varepsilon > 0$, perché per $\varepsilon < 0$ si avrebbe $P(|X - \mu| > \varepsilon) = P(\Omega) = 1$, mentre per $\varepsilon = 0$ il secondo membro non avrebbe senso.
- La disuguaglianza è significativa solo se $\sigma^2/\varepsilon^2 < 1$, altrimenti si ottiene una stima banale, dato che la probabilità di ogni evento è sempre al più 1.

Proof. Partiamo dalla definizione di varianza:

$$\sigma^2 = \mathrm{Var}(X) = \sum_{\omega_i \in \Omega} p(\omega_i)\,(X(\omega_i) - \mu)^2.$$

Possiamo separare la somma in due parti, in base all'evento $\{|X - \mu| > \varepsilon\}$:

$$\sigma^2 = \sum_{\substack{\omega_i \in \Omega \\ |X(\omega_i) - \mu| > \varepsilon}} p(\omega_i)\,(X(\omega_i) - \mu)^2 + \sum_{\substack{\omega_i \in \Omega \\ |X(\omega_i) - \mu| \leq \varepsilon}} p(\omega_i)\,(X(\omega_i) - \mu)^2.$$

Poiché il secondo termine è non negativo, otteniamo

$$\sigma^2 \geq \sum_{\substack{\omega_i \in \Omega \\ |X(\omega_i)-\mu|>\varepsilon}} p(\omega_i)\,(X(\omega_i)-\mu)^2.$$

Ma su ciascun ω_i dell'evento $\{|X-\mu|>\varepsilon\}$ si ha $(X(\omega_i)-\mu)^2 > \varepsilon^2$, quindi

$$\sum_{\substack{\omega_i \in \Omega \\ |X(\omega_i)-\mu|>\varepsilon}} p(\omega_i)\,(X(\omega_i)-\mu)^2 \geq \sum_{\substack{\omega_i \in \Omega \\ |X(\omega_i)-\mu|>\varepsilon}} p(\omega_i)\,\varepsilon^2 = \varepsilon^2\, P(|X-\mu|>\varepsilon).$$

Da cui segue immediatamente

$$P(|X-\mu|>\varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

□

Chebyshev: spiegazione semplice

La disuguaglianza di Chebyshev dice quanto una variabile casuale X può allontanarsi dalla sua media μ.

- Non serve sapere com'è fatta la distribuzione di X.
- Basta che la varianza σ^2 esista.
- L'idea è: **quasi tutti i valori di** X si trovano vicino alla media.

In forma semplice:

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Questo significa che **almeno**

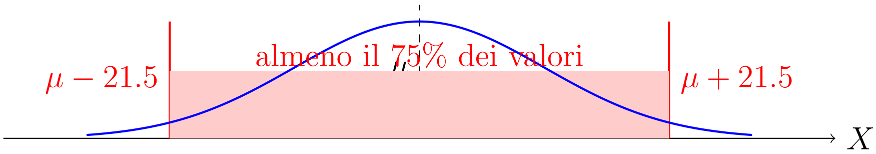
$$1 - \frac{1}{k^2}$$

dei valori è dentro l'intervallo $\mu - k\sigma$ e $\mu + k\sigma$.

Esempi semplici

- **Con $k = 2$:**
  Almeno il 75% dei valori si trova tra $\mu - 2\sigma$ e $\mu + 2\sigma$.
- **Con $k = 3$:**
  Almeno l'89% dei valori è tra $\mu - 3\sigma$ e $\mu + 3\sigma$.

Idea visiva



In breve

Chebyshev garantisce che **una grande parte dei valori** non si allontana molto dalla media, anche senza conoscere la forma della distribuzione.

23.2 Verso la Legge Debole dei Grandi Numeri (caso finito)

Siano $X_1, X_2, \ldots, X_n$ variabili aleatorie *indipendenti* e *identicamente distribuite (i.i.d.)*, con

$$E[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2 < \infty.$$

Definiamo la *media aritmetica campionaria*:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

—

Calcolo della varianza della media. La varianza di una somma di variabili indipendenti è la somma delle varianze. Inoltre, moltiplicare una variabile per una costante $1/n$ moltiplica la varianza per $(1/n)^2$. Quindi:

$$\mathrm{Var}(\overline{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Interpretazione intuitiva: la varianza della media diminuisce all'aumentare di n, cioè la media campionaria è più stabile man mano che consideriamo più osservazioni.

—

Applicazione della disuguaglianza di Chebyshev. Per ogni $\varepsilon > 0$:

$$P\big(|\overline{X}_n - \mu| \geq \varepsilon\big) \leq \frac{\mathrm{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

All'aumentare di n, il lato destro tende a 0. In parole semplici: più osservazioni consideriamo, più la media campionaria si avvicina alla media vera μ, con probabilità arbitrariamente alta.

—

Conclusione. Questo risultato mostra la **Legge Debole dei Grandi Numeri** nel caso finito: la media di un campione i.i.d. tende in probabilità al valore atteso della popolazione.

$$\boxed{X_1, \ldots, X_n \text{ i.i.d.} \iff \begin{cases} \text{indipendenti} \\ \text{identicamente distribuite} \end{cases}}$$

Nota. Non è necessario che le variabili siano identiche: basta che abbiano *stesso valore atteso e varianza finita* e siano *non correlate* ($\mathrm{Cov}(X_i, X_j) = 0$ per $i \neq j$). L'indipendenza è una condizione sufficiente ma non necessaria: la disuguaglianza di Chebyshev funziona anche con variabili non indipendenti purché siano non correlate e abbiano varianza finita. *Interpretazione intuitiva:* la varianza della media diminuisce all'aumentare di n, cioè la media campionaria è più stabile man mano che consideriamo più osservazioni.

—

Applicazione della disuguaglianza di Chebyshev. Per ogni $\varepsilon > 0$:

$$P\big(|\overline{X}_n - \mu| \geq \varepsilon\big) \leq \frac{\mathrm{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

All'aumentare di n, il lato destro tende a 0. In parole semplici: più osservazioni consideriamo, più la media campionaria si avvicina alla media vera μ, con probabilità arbitrariamente alta.

—

Conclusione. Questo risultato mostra la **Legge Debole dei Grandi Numeri** nel caso finito: la media di un campione i.i.d. tende in probabilità al valore atteso della popolazione.

Nota. Non è necessario che le variabili siano identiche: basta che abbiano *stesso valore atteso e varianza finita* e siano *non correlate* ($\mathrm{Cov}(X_i, X_j) = 0$ per $i \neq j$). L'indipendenza è una condizione sufficiente ma non necessaria: la disuguaglianza di Chebyshev funziona anche con variabili non indipendenti purché siano non correlate e abbiano varianza finita.

23.3 Intervalli di confidenza con Chebyshev (spiegazione semplice)

Idea generale. Un intervallo di confidenza è un intervallo che, con alta probabilità, contiene il vero valore della media μ. In simboli:

$$P(\mu \in I_n) \geq 1 - \delta,$$

dove $1 - \delta$ è il livello di confidenza (ad esempio 95%).

Come usare Chebyshev. La disuguaglianza di Chebyshev garantisce che la media campionaria $\overline{X}_n$ non si allontana troppo dalla media vera μ:

$$P(|\overline{X}_n - \mu| < k\,\sigma/\sqrt{n}) \geq 1 - \frac{1}{k^2}.$$

Se scegliamo k in modo che

$$\frac{1}{k^2} = \delta,$$

allora il livello di confidenza è proprio $1 - \delta$.

Ne ricaviamo un intervallo di confidenza molto semplice:

$$I_n = \left[\overline{X}_n - \frac{k\sigma}{\sqrt{n}},\ \overline{X}_n + \frac{k\sigma}{\sqrt{n}}\right], \qquad P(\mu \in I_n) \geq 1 - \delta.$$

Interpretazione intuitiva. Più osservazioni n raccogliamo:

- più la media campionaria $\overline{X}_n$ è precisa;
- più l'intervallo diventa stretto (perché c'è un fattore $\frac{1}{\sqrt{n}}$).

Caso delle variabili di Bernoulli

Situazione. Ogni X_i vale:

$$X_i = \begin{cases} 1 & \text{con probabilità } p, \\ 0 & \text{con probabilità } 1 - p, \end{cases}$$

e vogliamo stimare p.

Media e varianza. Per le Bernoulli:

$$E[X_i] = p, \qquad \mathrm{Var}(X_i) = p(1 - p).$$

Ma non conosciamo p. Tuttavia, sappiamo che:

$$p(1 - p) \leq \frac{1}{4} \quad \text{perché il massimo è in } p = \tfrac{1}{2}.$$

Uso di Chebyshev. Sostituendo questo limite:

$$P\left(|\overline{X}_n - p| < \frac{k}{2\sqrt{n}}\right) \geq 1 - \frac{1}{k^2}.$$

Perché è utile? Anche senza conoscere p:

- otteniamo comunque un intervallo di confidenza valido,
- l'intervallo dipende solo da n,
- è garantito per qualunque distribuzione Bernoulli.

In breve

Chebyshev permette di costruire intervalli di confidenza **senza conoscere la distribuzione precisa** (o senza conoscere p nel caso Bernoulli). Gli intervalli non sono molto stretti, ma funzionano sempre.

Commento finale

L'approccio di Chebyshev è completamente generale: funziona per qualunque variabile con varianza finita. Per questo gli intervalli ottenuti sono più "larghi", ma hanno il vantaggio di richiedere pochissime ipotesi. Quando però si conosce qualcosa in più sulla distribuzione (ad esempio normalità), si ottengono intervalli molto più precisi.

Perché compare $1/k^2$ e non $\frac{1}{(k\sigma/\sqrt{n})^2}$? La forma generale della disuguaglianza di Chebyshev è:

$$P(|Y - E[Y]| \geq a) \leq \frac{\mathrm{Var}(Y)}{a^2}.$$

Nel nostro caso $Y = \overline{X}_n$ e sappiamo che:

$$\mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}.$$

Scegliamo come soglia:

$$a = k\frac{\sigma}{\sqrt{n}}.$$

Inseriamo questi valori nella disuguaglianza:

$$P\left(|\overline{X}_n - \mu| \geq k\frac{\sigma}{\sqrt{n}}\right) \leq \frac{\frac{\sigma^2}{n}}{\left(k\frac{\sigma}{\sqrt{n}}\right)^2}.$$

Semplifichiamo il rapporto:

$$\frac{\sigma^2/n}{k^2\sigma^2/n} = \frac{1}{k^2}.$$

Osserviamo che:

- il fattore σ^2 si semplifica,
- anche il fattore $1/n$ si semplifica,
- l'unica cosa che rimane è $1/k^2$.

Conclusione. Il motivo per cui il risultato finale contiene solamente $1/k^2$ e *non* $\frac{1}{(k\sigma/\sqrt{n})^2}$ è che tutti i termini σ^2 e $1/n$ si cancellano automaticamente quando si applica Chebyshev alla media campionaria. Rimane soltanto $1/k^2$, esattamente come nella forma standard della disuguaglianza.

23.4 Esercizi svolti

23.4.1 Esercizio: estrazione di carte (da [SN] e ROSS)

Si estraggono due carte senza reinserimento da un mazzo di 40, di cui:

- 10 di bastoni (B),
- 10 di denari (D),
- 20 di coppe/spade (CS).

Definiamo:

$$X = X_B = \text{numero di carte di bastoni estratte}, \quad Y = X_D = \text{numero di carte di denari estratte}.$$

I possibili valori della coppia (X, Y) sono:

$$\{(h,k) : h,k \geq 0,\ h+k \leq 2\} = \{(0,0),(0,1),(0,2),(1,0),(1,1),(2,0)\}.$$

Per simmetria, X e Y hanno distribuzioni marginali identiche:

$$X \sim Y \sim \text{Hyp}(40, 10; 2).$$

La densità congiunta vale:

$$p_{X,Y}(h,k) = \frac{\binom{10}{h}\binom{10}{k}\binom{20}{2-h-k}}{\binom{40}{2}}, \quad h,k \geq 0,\ h+k \leq 2.$$

Ad esempio:

$$p_{X,Y}(0,1) = \frac{10 \cdot 20}{\frac{40 \cdot 39}{2}} = \frac{10}{39}, \qquad p_{X,Y}(0,2) = \frac{\frac{10 \cdot 9}{2}}{\frac{40 \cdot 39}{2}} = \frac{9}{4 \cdot 39}.$$

23.4.2 Esercizio: Transistor difettosi (ROSS, cap. 6, es. 6)

Una scatola contiene 5 transistor, di cui esattamente 2 difettosi. Si testano uno alla volta finché entrambi i difettosi sono trovati.

Sia:

$$N_1 = \text{numero di test fino al primo difettoso}, \quad N_2 = \text{numero di ulteriori test fino al secondo difettoso}.$$

I possibili valori sono:

$$1 \leq h \leq 4, \quad 1 \leq k \leq 4, \quad h+k \leq 5,$$

cioè 10 coppie (h, k):

$$(1,1),(1,2),(1,3),(1,4),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1).$$

Si trova che:

$$P(N_1 = h, N_2 = k) = \frac{1}{10}, \quad \text{per ciascuna di queste 10 coppie.}$$

23.5 Variabili standard e standardizzate

[Variabile standard] Una variabile aleatoria Z è detta **standard** se:

$$E[Z] = 0, \qquad \text{Var}(Z) = 1.$$

[Standardizzazione] Data X con $E[X] = \mu$ e $\text{Var}(X) = \sigma^2 > 0$, si definisce la **standardizzata** di X come:

$$X^* = \frac{X - \mu}{\sigma}.$$

Essa soddisfa $E[X^*] = 0$ e $\text{Var}(X^*) = 1$.

23.6 La media come migliore “stima centrale”

Idea. Se vogliamo scegliere un numero t che sia il più vicino possibile ai valori di X, la scelta migliore è sempre la media $E[X]$. In particolare, la media è il numero che rende minimo l'errore quadratico medio:

$$E[(X-t)^2].$$

Teorema (versione semplice)

Tra tutti i numeri reali t,

$$E[(X-t)^2]$$

è minimo quando $t = E[X]$. Il valore minimo ottenuto è la varianza:

$$\min_t E[(X-t)^2] = \mathrm{Var}(X).$$

Dimostrazione intuitiva

Sviluppiamo l'espressione:

$$E[(X-t)^2] = E[X^2] - 2t\,E[X] + t^2.$$

Questa è una parabola in t. Le parabole hanno il minimo nel punto in cui la derivata si annulla. Calcoliamo:

$$f'(t) = -2E[X] + 2t = 0 \quad \Rightarrow \quad t = E[X].$$

Sostituendo otteniamo:

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

Significato semplice.

- Scegliere t troppo grande o troppo piccolo aumenta gli errori $(X-t)^2$.
- La media è il punto che bilancia meglio tutti gli scarti.
- La “distanza media minima” che possiamo ottenere è proprio la varianza.

Immagine mentale. Immagina tutti i possibili valori che X può assumere su una retta. La media è il punto che sta, nel complesso, più vicino a tutti questi valori. Quanto oscillano complessivamente rispetto alla media? Questa oscillazione è la varianza.

23.7 Cenno alla retta di regressione di Y rispetto a X

La retta di regressione standardizzata è:

$$\frac{y-\mu_Y}{\sigma_Y} = \rho_{X,Y}\frac{x-\mu_X}{\sigma_X},$$

dove:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}.$$

Equivale alla forma affine:

$$y = a^*x + b^*, \qquad a^* = \rho_{X,Y}\frac{\sigma_Y}{\sigma_X}, \quad b^* = \mu_Y - a^*\mu_X.$$

Proprietà fondamentale (senza dimostrazione). La retta di regressione $y = a^*x + b^*$ è quella che minimizza:

$$E[(Y - aX - b)^2],$$

ossia:

$$E[(Y - a^*X - b^*)^2] \leq E[(Y - aX - b)^2], \quad \forall\, a, b \in \mathbb{R}.$$

24 Lezione del 28 novembre 2024 – Modelli di occupazione, disuguaglianza di Cauchy e paradosso di San Pietroburgo

24.1 Esercizio 6 di ROSS e variabili X_1, X_2, X_3

Riprendiamo l'esercizio:

> *Una scatola contiene 5 transistor, dei quali esattamente 2 sono difettosi. I transistor vengono testati uno alla volta fino a trovare entrambi i difettosi. Sia N_1 il numero di test per individuare il primo transistor difettoso e N_2 il numero di ulteriori test per trovare il secondo. Determinare la densità discreta congiunta di (N_1, N_2).*

Nuova impostazione. Consideriamo ora di estrarre tutti e 5 i transistor e definiamo:

$$\begin{cases} X_1 := \text{numero di transistor buoni prima del primo difettoso,} \\ X_2 := \text{numero di transistor buoni tra il primo e il secondo difettoso,} \\ X_3 := \text{numero di transistor buoni dopo il secondo difettoso.} \end{cases}$$

Poiché vi sono in totale 3 transistor buoni e 2 difettosi, si ha:

$$X_1 + X_2 + X_3 = 3, \qquad X_i \geq 0,\ X_i \in \mathbb{N}.$$

Il legame con (N_1, N_2) è:

$$P(N_1 = h_1, N_2 = h_2) = P(X_1 = h_1 - 1,\ X_2 = h_2 - 1,\ X_3 = 3 - h_1 - h_2).$$

Osservazione combinatoria. Ogni disposizione possibile dei due difettosi fra i 5 posti ha probabilità uguale. Ci sono $\binom{5}{2} = 10$ casi equiprobabili, corrispondenti alle diverse scelte delle posizioni dei due transistor difettosi.

L'evento $\{X_1 = k_1, X_2 = k_2, X_3 = k_3\}$ (con $k_1 + k_2 + k_3 = 3$) corrisponde a specificare in quali posizioni appaiono i due difettosi. Ad esempio:

$$(X_1, X_2, X_3) = (0, 2, 1) \iff (D_1, B_2, B_3, D_4, B_5).$$

Essendo tutti gli eventi equiprobabili, abbiamo:

$$P(X_1 = k_1, X_2 = k_2, X_3 = k_3) = \begin{cases} \frac{1}{10}, & \text{se } k_1 + k_2 + k_3 = 3, \\ 0, & \text{altrimenti.} \end{cases}$$

Segue quindi la distribuzione congiunta uniforme per (N_1, N_2) su 10 coppie ammissibili.

24.2 Modelli di occupazione

Abbiamo introdotto il concetto di **vettori di occupazione** $(k_1, \ldots, k_n)$ definiti da:

$$k_i \in \mathbb{N}_0, \qquad k_1 + \cdots + k_n = r.$$

Tali vettori rappresentano schemi di ripartizione di r oggetti indistinguibili in n "celle" (che possono essere distinte o meno).

Numero di vettori di occupazione. Il numero di soluzioni intere non negative di $k_1+\cdots+k_n = r$ è:

$$\binom{n+r-1}{r}.$$

Perché vale questa formula? Vogliamo contare quante soluzioni intere non negative esistono per

$$k_1 + \cdots + k_n = r.$$

L'idea è usare il metodo delle **stelle e sbarre**. Rappresentiamo gli r oggetti con r stelle:

$$\star \ \star \ \cdots \ \star$$

e separiamo le n celle con $n-1$ sbarre verticali:

$$|$$

Una distribuzione come $(k_1, k_2, \ldots, k_n)$ corrisponde a una sequenza contenente:

$$r \text{ stelle} \quad \text{e} \quad n-1 \text{ sbarre.}$$

Ad esempio, la soluzione $(2, 1, 2)$ quando $r = 5$ e $n = 3$ si rappresenta come:

$$\star \ \star \ | \ \star \ | \ \star \ \star$$

Ogni sequenza diversa di r stelle e $n-1$ sbarre rappresenta una diversa soluzione. Il numero totale di tali sequenze è il numero di modi di scegliere le posizioni delle r stelle (o equivalentemente delle $n-1$ sbarre) fra i $r+n-1$ simboli totali:

$$\binom{r+n-1}{r}.$$

Questa formula funziona anche quando alcune celle sono vuote: basta che due sbarre stiano vicine, oppure che una sbarra sia a inizio o fine della sequenza.

Esempio grafico Consideriamo $r = 5$ oggetti e $n = 3$ celle. Una possibile soluzione è $(k_1, k_2, k_3) = (5, 0, 0)$, rappresentata graficamente qui sotto:

$k_1 = 5$	$k_2 = 0$	$k_3 = 0$
$\star \ \star \ \star \ \star \ \star$	$\vert$	$\vert$
Cella 1	Cella 2	Cella 3

Modelli principali.

- **Bose–Einstein:** oggetti indistinguibili, celle indistinguibili o equivalenti, con occupazioni multiple possibili.
- **Fermi–Dirac:** oggetti indistinguibili, ma ogni cella può contenere al più un elemento (no doppie occupazioni).
- **Maxwell–Boltzmann:** oggetti e celle distinguibili, ogni oggetto può andare in qualunque cella indipendentemente.

24.3 Esempi di modelli di occupazione

24.3.1 Esercizio 14.2 (Estrazioni di assi) – Modello Bose–Einstein

Caso analogo a quello dei transistor: le variabili $(X_1, X_2, \dots, X_n)$ rappresentano il numero di elementi di ciascun tipo ottenuti da estrazioni ripetute.

24.3.2 Esempio 14.3 e 14.4 (modello Fermi–Dirac e Maxwell–Boltzmann)

Situazione. Cinque persone entrano in un ascensore che serve 4 piani. Per $i = 1, 2, 3, 4$, si definisce:

$$X_i := \text{numero di persone che scendono al piano } i.$$

Assunzione. Ogni persona sceglie in modo indipendente e uniforme uno dei 4 piani:

$$P(Y_j = k) = \frac{1}{4}, \quad j = 1, \dots, 5, \; k = 1, \dots, 4.$$

dove Y_j è il piano scelto dalla persona j.

Calcolo di probabilità di un vettore di occupazione. L'evento $\{X_1 = 1, X_2 = 0, X_3 = 2, X_4 = 2\}$ corrisponde all'unione di tutti gli anagrammi delle combinazioni $(1, 3, 3, 4, 4)$.

Il numero di tali anagrammi è:

$$\frac{5!}{1!\,0!\,2!\,2!} = 30.$$

Poiché ciascun vettore $(Y_1, \dots, Y_5)$ ha probabilità $1/4^5 = 1/1024$, si ottiene:

$$P(\{X_1 = 1, X_2 = 0, X_3 = 2, X_4 = 2\}) = 30 \cdot \frac{1}{4^5} = \frac{30}{1024}.$$

Analogamente:

$$P(\{X_1 = 1, X_2 = 0, X_3 = 0, X_4 = 4\}) = \frac{5!}{1!0!0!4!}\frac{1}{4^5} = 5 \cdot \frac{1}{1024} = \frac{5}{1024}.$$

Questo è un tipico esempio di **modello di Maxwell–Boltzmann**.

24.4 Disuguaglianza di Cauchy–Schwarz

[Disuguaglianza di Cauchy–Schwarz] Per due variabili aleatorie X e Y con secondo momento finito vale:

$$|E[XY]| \leq \sqrt{E[X^2]\,E[Y^2]}.$$

Proof. Definiamo, per $t \in \mathbb{R}$, la funzione

$$g(t) = E[(X - tY)^2].$$

Essa è non negativa per ogni t in quanto è l'attesa di un quadrato. Espandendo:

$$g(t) = E[X^2] - 2tE[XY] + t^2E[Y^2].$$

Si tratta di un polinomio di secondo grado con coefficiente direttore $E[Y^2] \geq 0$. Poiché $g(t) \geq 0$ per ogni t, il discriminante deve essere non positivo:

$$(-2E[XY])^2 - 4E[Y^2]E[X^2] \leq 0.$$

Da cui:

$$(E[XY])^2 \leq E[X^2]E[Y^2],$$

che completa la dimostrazione. □

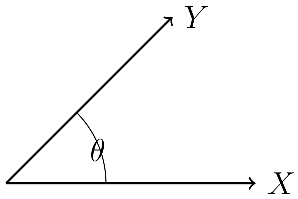
Interpretazione geometrica. Lo spazio delle variabili aleatorie a secondo momento finito è uno spazio vettoriale reale dove è possibile definire il prodotto scalare:

$$\langle X, Y \rangle := E[XY].$$

La quantità $\|X\| := \sqrt{E[X^2]}$ è la norma indotta. La disuguaglianza di Cauchy–Schwarz assume quindi la forma classica:

$$|\langle X, Y \rangle| \leq \|X\| \, \|Y\|.$$

Nel caso centrato ($X - E[X]$, $Y - E[Y]$), il coefficiente di correlazione è il *coseno dell'angolo* tra i due "vettori". Questo spiega perché $\rho_{X,Y} \in [-1, 1]$.



Interpretazione geometrica: $\rho_{X,Y} = \cos\theta$.

Corollario (Covarianza e coefficiente di correlazione). Applicando la disuguaglianza di Cauchy–Schwarz alle variabili centrali $X - E[X]$ e $Y - E[Y]$ si ottiene

$$|\operatorname{Cov}(X, Y)| = |E[(X - E[X])(Y - E[Y])]| \leq \sqrt{E[(X - E[X])^2] \, E[(Y - E[Y])^2]} = \sigma_X \sigma_Y,$$

dove

- $\operatorname{Cov}(X, Y)$ è la *covarianza*, che misura quanto le due variabili tendono a variare insieme;
- $\sigma_X = \sqrt{\operatorname{Var}(X)}$ e $\sigma_Y = \sqrt{\operatorname{Var}(Y)}$ sono le *deviazioni standard*, che misurano la dispersione di ciascuna variabile intorno alla media.

Da questa disuguaglianza segue che il *coefficiente di correlazione*

$$\rho_{X,Y} = \frac{\operatorname{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

soddisfa sempre

$$-1 \leq \rho_{X,Y} \leq 1,$$

fornendo una misura adimensionale della correlazione lineare tra X e Y.

Esempio numerico 1 (correlazione perfetta). Sia X uniforme su $\{-1, 0, 1\}$ e $Y = 3X$. Allora:

$$E[XY] = 3E[X^2], \qquad E[X^2] = \tfrac{2}{3}, \qquad E[Y^2] = 9E[X^2] = 6.$$

Quindi:

$$|E[XY]| = 2\sqrt{1.5} \qquad \text{e} \qquad \sqrt{E[X^2]E[Y^2]} = 2\sqrt{1.5}.$$

La disuguaglianza è verificata con uguaglianza (dipendenza lineare).

Esempio numerico 2 (correlazione nulla). Sia X uniforme su $\{-1, 1\}$ e Y uniforme e indipendente su $\{-2, 2\}$. Allora:

$$E[X] = E[Y] = 0, \qquad E[XY] = E[X]E[Y] = 0,$$

mentre

$$E[X^2] = 1, \qquad E[Y^2] = 4.$$

La disuguaglianza dà:

$$0 = |E[XY]| \leq 2 = \sqrt{1 \cdot 4}.$$

In questo caso $\rho_{X,Y} = 0$, ma X e Y non sono costanti né nulli.

24.5 Paradosso di San Pietroburgo

Il paradosso di San Pietroburgo è un classico esempio in teoria della probabilità che mette in evidenza i limiti del valore atteso come misura di utilità.

Definizione del gioco. Il gioco funziona così:

- Si lancia una moneta equa fino a ottenere testa.
- Se la testa appare al k-esimo lancio, il giocatore riceve un guadagno pari a 2^k.

—

Interpretazione tramite variabile geometrica. Sia

$$T \sim \text{Geom}(p), \quad P(T = k) = (1-p)^{k-1}p, \quad k \geq 1,$$

la variabile aleatoria che rappresenta il numero di lanci necessari per ottenere la prima testa.

Il guadagno G del gioco è proporzionale a 2^T:

$$G = 2^T.$$

—

Calcolo dell'attesa. Il valore atteso del guadagno è

$$E[G] = E[2^T] = \sum_{k=1}^{\infty} 2^k (1-p)^{k-1} p.$$

Per $p = 1/2$ (moneta equa), la serie diventa

$$E[2^T] = \sum_{k=1}^{\infty} 2^k \left(\frac{1}{2}\right)^{k-1} \frac{1}{2} = \sum_{k=1}^{\infty} 1 = +\infty.$$

Quindi, sebbene ogni singolo guadagno sia finito, la **media teorica del guadagno è infinita**.

—

Nota importante. La variabile T può teoricamente assumere anche il valore infinito, ma:

$$P(T = \infty) = 1 - P(T < \infty) = 1 - \sum_{k=1}^{\infty} (1-p)^{k-1} p = 1 - p \sum_{h=0}^{\infty} (1-p)^h = 1 - p \cdot \frac{1}{1-(1-p)} = 0.$$

Quindi T è finita quasi certamente: $P(T < \infty) = 1$.

—

Esempio numerico. Supponiamo di giocare con una moneta equa ($p = 1/2$):

- Probabilità di ottenere il guadagno 2 al primo lancio: $P(T = 1) = 1/2$
- Probabilità di ottenere 4 al secondo lancio: $P(T = 2) = 1/4$
- Probabilità di ottenere 8 al terzo lancio: $P(T = 3) = 1/8$

La media teorica del guadagno è

$$E[G] = 2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + 8 \cdot \frac{1}{8} + \cdots = 1 + 1 + 1 + \cdots = \infty.$$

—

Conclusione. Il paradosso mostra che, pur avendo probabilità 1 di vincere un importo finito ad ogni partita, il valore atteso del guadagno è infinito. Questo mette in evidenza un limite dell'uso del valore atteso come unico criterio di decisione e ha portato allo sviluppo di concetti come *utilità attesa* e *funzioni di utilità* in economia e teoria dei giochi.

Riferimenti:

- Esercizio 6, Capitolo 6 di ROSS.
- Lezione 13, *Approfondimenti di Calcolo Combinatorio* in [SN].
- Esempi 14.2–14.4 di [SN].

25 Lezione del 29 novembre 2024 – Somma di variabili geometriche e tempo del secondo successo

25.1 Premessa

La lezione (solo online a causa dello sciopero) ha riguardato le domande Q1–Q4 del compito del *16 luglio 2020* (`CP-INFORM-16-luglio-2020-corretto.pdf`). In particolare, per rispondere ai punti della domanda Q4 si sono ripresi i risultati principali sulle variabili geometriche.

25.2 Richiami sulle variabili geometriche

[Variabile geometrica] Una variabile aleatoria W si dice **geometrica di parametro** p, e si scrive $W \sim \mathrm{Geom}(p)$, se:

$$P(W = k) = (1-p)^{k-1}p, \qquad k = 1, 2, 3, \dots$$

Essa rappresenta il numero di prove necessarie fino al *primo successo* in una successione di eventi indipendenti di probabilità p.

Valore atteso e varianza.

$$E[W] = \frac{1}{p}, \qquad \mathrm{Var}(W) = \frac{1-p}{p^2}.$$

Pertanto, se $E[W] = \beta$, allora $p = \frac{1}{\beta}$. Nel caso della Q4, con $\beta = 4$, si ottiene $p = \frac{1}{4}$.

25.3 Somma di due variabili geometriche indipendenti

Consideriamo due variabili aleatorie indipendenti X e Y distribuite come $\mathrm{Geom}(p)$. Definiamo:

$$Z = X + Y.$$

Densità discreta di Z. Per $n \geq 2$,

$$P(Z = n) = \sum_{h=1}^{n-1} P(X = h, Y = n-h) = \sum_{h=1}^{n-1} (1-p)^{h-1}p\,(1-p)^{n-h-1}p = (n-1)(1-p)^{n-2}p^2.$$

$$\boxed{P(Z = n) = (n-1)p^2(1-p)^{n-2}, \qquad n \geq 2.}$$

Questa è esattamente la legge del **tempo del secondo successo** in una successione di prove indipendenti.

25.4 Interpretazione: il tempo del secondo successo

Sia $(A_n)_{n\geq 1}$ una successione di eventi indipendenti tali che $P(A_n) = p$. Definiamo:

$$T_1 := \text{tempo del primo successo}, \qquad T_2 := \text{tempo del secondo successo}.$$

Per ipotesi, le prove proseguono finché si ottiene il secondo successo.

Definizione delle variabili incrementali. Poniamo:

$$\Delta_1 = T_1, \qquad \Delta_2 = T_2 - T_1.$$

- Δ_1 rappresenta il numero di prove fino al primo successo;
- Δ_2 rappresenta il numero di prove dopo il primo successo fino al secondo.

Distribuzione congiunta di Δ_1 e Δ_2. Per $h, k \geq 1$:

$$P(\Delta_1 = h, \Delta_2 = k) = P\big((A_1)^c \cap \cdots \cap (A_{h-1})^c \cap A_h \cap (A_{h+1})^c \cap \cdots \cap (A_{h+k-1})^c \cap A_{h+k}\big).$$

Poiché gli eventi sono indipendenti:

$$P(\Delta_1 = h, \Delta_2 = k) = (1-p)^{h-1} p\, (1-p)^{k-1} p = P(X = h)\, P(Y = k).$$

$$\boxed{P(\Delta_1 = h, \Delta_2 = k) = P(X = h, Y = k).}$$

Conseguenza. Le variabili Δ_1 e Δ_2 sono indipendenti e geometriche di parametro p, quindi:

$$T_2 = \Delta_1 + \Delta_2 \quad \text{ha la stessa distribuzione di } Z = X + Y.$$

$$\boxed{P(T_2 = n) = (n-1)p^2(1-p)^{n-2}, \qquad n \geq 2.}$$

25.5 Collegamento con la famiglia dei tempi di successo

Più in generale, il tempo del r-esimo successo T_r segue la distribuzione:

$$P(T_r = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \qquad n \geq r,$$

che coincide con una **Binomiale negativa** $\mathrm{NegBin}(r, p)$.

Valore atteso e varianza.

$$E[T_r] = \frac{r}{p}, \qquad \mathrm{Var}(T_r) = \frac{r(1-p)}{p^2}.$$

25.6 Osservazione finale

Il legame tra $Z = X + Y$ e T_2 mostra come, nello schema di Bernoulli, la somma di variabili geometriche indipendenti corrisponda ai *tempi successivi di successo*. Tale struttura è utile anche per la simulazione di processi di Poisson discreti e per l'analisi di modelli di attesa.

Riferimenti:

- Domande Q1–Q4 del compito del 16 luglio 2020 (`CP-INFORM-16-luglio-2020-corretto.pdf`);
- Lezione su variabili geometriche e tempi di successo in [SN, Capitolo 4];
- File `Dal libro di ROSS-18-dic-2024.pdf` e `ROSS-14nov2024.pdf`.

26 Lezione del 5 dicembre 2024 – Valore atteso totale e somma di variabili indipendenti

26.1 Formula del valore atteso totale

[Formula del valore atteso totale] Sia X una variabile aleatoria e sia $\{B_i\}_{i=1}^n$ una partizione finita di Ω tale che $P(B_i) > 0$ per ogni i. Allora:

$$E[X] = \sum_{i=1}^{n} E[X \mid B_i]\, P(B_i).$$

La formula si estende al caso di partizioni numerabili. Essa permette di esprimere il valore atteso di una variabile come media pesata dei valori attesi condizionati.

26.2 Applicazione: calcolo di $E[X]$ per $X \sim \mathrm{Geom}(p)$

Usando la proprietà di mancanza di memoria. Una variabile $X \sim \mathrm{Geom}(p)$ gode della proprietà:

$$P(X > h + k \mid X > h) = P(X > k),$$

ossia la distribuzione "riparte da zero" dopo ogni insuccesso.

Poniamo $E[X] = m$. Condizionando rispetto al risultato della prima prova:

$$E[X] = E[X \mid A_1]P(A_1) + E[X \mid (A_1)^c]P((A_1)^c),$$

dove A_1 è l'evento "successo alla prima prova". Segue:

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

da cui:

$$E[X] = p + (1 - p)(1 + E[X]) \quad \Rightarrow \quad E[X] = \frac{1}{p}.$$

Usando le serie di potenze. Poiché $P(X = k) = (1 - p)^{k-1}p$, otteniamo:

$$E[X] = \sum_{k=1}^{\infty} k(1 - p)^{k-1}p = p \sum_{k=1}^{\infty} k(1 - p)^{k-1}.$$

Richiamando la serie geometrica derivata:

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1 - x)^2}, \quad |x| < 1,$$

e ponendo $x = 1 - p$, segue:

$$E[X] = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

26.3 Somma di variabili aleatorie indipendenti

26.3.1 Caso binomiale

Se $X \sim \mathrm{Bin}(r, \theta)$ e $Y \sim \mathrm{Bin}(s, \theta)$ sono indipendenti, allora

$$X + Y \sim \mathrm{Bin}(r + s, \theta).$$

Proof. Per ogni n intero:

$$P(X+Y=n)=\sum_{k=0}^{n}P(X=k)P(Y=n-k).$$

Sostituendo le densità binomiali:

$$P(X+Y=n)=\sum_{k=0}^{n}\binom{r}{k}\theta^k(1-\theta)^{r-k}\binom{s}{n-k}\theta^{n-k}(1-\theta)^{s-n+k}.$$

Raccogliendo i termini:

$$P(X+Y=n)=\theta^n(1-\theta)^{r+s-n}\sum_{k=0}^{n}\binom{r}{k}\binom{s}{n-k}.$$

Usando l'identità di Vandermonde:

$$\sum_{k=0}^{n}\binom{r}{k}\binom{s}{n-k}=\binom{r+s}{n},$$

otteniamo:

$$P(X+Y=n)=\binom{r+s}{n}\theta^n(1-\theta)^{r+s-n},$$

ossia $X+Y\sim\mathrm{Bin}(r+s,\theta)$. □

26.3.2 Caso di due variabili Poisson indipendenti

Se $X\sim\mathrm{Poisson}(\lambda)$ e $Y\sim\mathrm{Poisson}(\mu)$ sono indipendenti, allora:

$$X+Y\sim\mathrm{Poisson}(\lambda+\mu).$$

Proof. Per $n\geq 0$:

$$P(X+Y=n)=\sum_{k=0}^{n}e^{-\lambda}\frac{\lambda^k}{k!}e^{-\mu}\frac{\mu^{n-k}}{(n-k)!}=e^{-(\lambda+\mu)}\sum_{k=0}^{n}\frac{\lambda^k\mu^{n-k}}{k!(n-k)!}.$$

Osservando che $\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}\lambda^k\mu^{n-k}=\frac{(\lambda+\mu)^n}{n!}$, segue:

$$P(X+Y=n)=e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^n}{n!}.$$

□

26.4 Distribuzioni condizionate di $X\mid X+Y=n$

26.4.1 Caso binomiale

Se $X\sim\mathrm{Bin}(r,\theta)$ e $Y\sim\mathrm{Bin}(s,\theta)$ indipendenti, allora:

$$X\mid X+Y=n\sim\mathrm{Hyp}(r+s,\,r;\,n),$$

ossia:

$$P(X=k\mid X+Y=n)=\frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}}.$$

26.4.2 Caso Poisson

Se $X \sim \text{Poisson}(\lambda)$ e $Y \sim \text{Poisson}(\mu)$ indipendenti, allora:

$$X \mid X+Y=n \sim \text{Bin}\left(n, \frac{\lambda}{\lambda+\mu}\right),$$

poiché:

$$P(X=k \mid X+Y=n) = \frac{\dfrac{e^{-\lambda}\lambda^k}{k!}\dfrac{e^{-\mu}\mu^{n-k}}{(n-k)!}}{e^{-(\lambda+\mu)}\dfrac{(\lambda+\mu)^n}{n!}} = \binom{n}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^k \left(\frac{\mu}{\lambda+\mu}\right)^{n-k}.$$

26.4.3 Caso geometrico

Se $X, Y \sim \text{Geom}(p)$ indipendenti, allora:

$$X \mid X+Y=n \sim \text{Unif}\{1,2,\ldots,n-1\},$$

in quanto:

$$P(X=k \mid X+Y=n) = \frac{P(X=k, Y=n-k)}{P(X+Y=n)} = \frac{(1-p)^{k-1}p\,(1-p)^{n-k-1}p}{(n-1)p^2(1-p)^{n-2}} = \frac{1}{n-1}.$$

26.5 Conclusione e collegamenti

Le tre famiglie precedenti condividono una *stabilità rispetto alla somma*:

- Binomiali $\Rightarrow$ Binomiale con parametri sommati;
- Poisson $\Rightarrow$ Poisson con parametri sommati;
- Geometriche $\Rightarrow$ Binomiale negativa.

Inoltre, il comportamento condizionato rispetto a $X+Y=n$ evidenzia una forte simmetria probabilistica:

Distribuzione originale	Condizionata a $X+Y=n$
$\text{Bin}(r,\theta), \text{Bin}(s,\theta)$	$\text{Hyp}(r+s, r; n)$
$\text{Poisson}(\lambda), \text{Poisson}(\mu)$	$\text{Bin}(n, \frac{\lambda}{\lambda+\mu})$
$\text{Geom}(p), \text{Geom}(p)$	$\text{Unif}\{1,\ldots,n-1\}$

Riferimenti:

- Sezione 9.4 di [SN] — Valore atteso condizionato e formula del valore atteso totale;
- Sezione 8.3.1 di [SN] — Distribuzione della somma e distribuzione condizionata di $X \mid X+Y$;
- File `SCHEMA-VARIABILI-ALEATORIE-19dicembre2024.pdf`;
- File `domande-4dicembre2024.pdf` e `domande-4dicembre2024SUGGERIMENTI.pdf`.

27 Lezione del 6 dicembre 2024 – Problema del collezionista e distribuzioni multi-ipergeometriche e multinomiali

27.1 Il problema del collezionista di figurine

Il cosiddetto *problema del collezionista di figurine* (in inglese *coupon collector problem*) è uno dei più classici esempi di variabili aleatorie discrete.

Descrizione del problema. Si supponga di avere un totale di n figurine diverse, numerate $1, 2, \dots, n$. Ad ogni acquisto si ottiene una figurina estratta in modo equiprobabile tra le n disponibili (e indipendentemente dalle precedenti estrazioni). Indichiamo con T_n il numero di acquisti necessari per completare la collezione, ossia per ottenere almeno una volta ciascuna delle n figurine.

Strategia di calcolo. Si scompone T_n come somma di variabili aleatorie indipendenti:

$$T_n = X_1 + X_2 + \cdots + X_n,$$

dove:

$$X_k = \text{numero di acquisti necessari per ottenere una nuova figurina quando ne abbiamo già } k-1.$$

Distribuzione di X_k. Dopo aver ottenuto $k-1$ figurine distinte, la probabilità di trovarne una nuova è:

$$p_k = \frac{n-(k-1)}{n} = \frac{n-k+1}{n}.$$

Pertanto, X_k è una variabile aleatoria geometrica con parametro p_k, cioè:

$$P(X_k = h) = (1-p_k)^{h-1} p_k, \quad h = 1, 2, \dots$$

e dunque:

$$E[X_k] = \frac{1}{p_k} = \frac{n}{n-k+1}.$$

Valore atteso del tempo totale. Per linearità del valore atteso:

$$E[T_n] = \sum_{k=1}^{n} E[X_k] = n \sum_{k=1}^{n} \frac{1}{n-k+1} = n \sum_{j=1}^{n} \frac{1}{j}.$$

Indicando con $H_n = \sum_{j=1}^{n} \frac{1}{j}$ l'n-esimo numero armonico, si ottiene:

$$E[T_n] = nH_n \approx n(\log n + \gamma),$$

dove $\gamma \approx 0.5772$ è la costante di Eulero–Mascheroni.

Esempio numerico. Per $n = 5$ tipi di figurine, il valore atteso del numero di acquisti necessari è:

$$E[T_5] = 5\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}\right) = 5 \cdot \frac{137}{60} \approx 11.42.$$

27.2 Distribuzioni multi-ipergeometriche e multinomiali

Distribuzione ipergeometrica classica. Si ricordi che una variabile X ha distribuzione $\mathrm{Hyp}(M, m; n)$ se:

$$P(X = k) = \frac{\binom{m}{k}\binom{M-m}{n-k}}{\binom{M}{n}}, \quad k = 0, 1, \dots, \min\{m, n\}.$$

Distribuzione multi-ipergeometrica. Consideriamo un'urna contenente M oggetti suddivisi in r categorie, di cui:

$$m_1, m_2, \dots, m_r, \quad \text{con} \quad \sum_{i=1}^{r} m_i = M.$$

Si estraggono n oggetti senza reinserimento e si definiscono:

$$X_i = \text{numero di oggetti di categoria } i \text{ estratti.}$$

La distribuzione congiunta è detta *multi-ipergeometrica*:

$$P(X_1 = k_1, \dots, X_r = k_r) = \frac{\prod_{i=1}^{r} \binom{m_i}{k_i}}{\binom{M}{n}},$$

dove $k_1 + \cdots + k_r = n$ e ciascun $k_i \geq 0$.

Distribuzione multinomiale. Se invece le estrazioni avvengono *con reinserimento*, e ogni estrazione ha probabilità p_i di appartenere alla categoria i, allora $(X_1, \dots, X_r)$ ha distribuzione multinomiale:

$$P(X_1 = k_1, \dots, X_r = k_r) = \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r},$$

con $k_1 + \cdots + k_r = n$ e $k_i \geq 0$.

Osservazione. Entrambe le distribuzioni possono essere viste come casi particolari di un *modello di occupazione*, in cui si distribuiscono n elementi in r categorie, distinguendo se la scelta è con o senza reinserimento.

27.3 Distribuzioni condizionate: casi geometrici e poissoniani

Caso geometrico. Abbiamo già visto che se X e Y sono indipendenti e $X, Y \sim \text{Geom}(p)$ (definite su $\{1, 2, 3, \dots\}$), allora:

$$X \mid (X + Y = n) \sim \text{Unif}\{1, 2, \dots, n - 1\}.$$

Nel caso alternativo, in cui le variabili sono definite su $\{0, 1, 2, \dots\}$, ossia con densità $P(X = k) = (1 - p)^k p$, allora:

$$X \mid (X + Y = k) \sim \text{Unif}\{0, 1, \dots, k\}.$$

La differenza deriva solo dal punto di partenza dell'indice della distribuzione.

Caso poissoniano. Se $X \sim \text{Poisson}(\lambda)$ e $Y \sim \text{Poisson}(\mu)$ sono indipendenti, allora:

$$X + Y \sim \text{Poisson}(\lambda + \mu),$$

e condizionatamente a $X + Y = n$:

$$X \mid (X + Y = n) \sim \text{Bin}\left(n, \frac{\lambda}{\lambda + \mu}\right).$$

Questo risultato si può interpretare come un'analogia con la formula del valore atteso totale e con la struttura della somma di conteggi indipendenti.

27.4 Osservazioni finali

- Il problema del collezionista di figurine mostra come il valore atteso possa essere calcolato tramite una decomposizione in variabili geometriche indipendenti.
- Le distribuzioni multi-ipergeometriche e multinomiali estendono i modelli discreti a più categorie.
- Le leggi condizionate per coppie di variabili indipendenti (Geometriche o Poissoniane) mostrano una profonda simmetria: la condizionata mantiene una forma "naturale" (Uniforme o Binomiale).

Riferimenti:

- File `domande-6dicembre2024-solo testo.pdf`;
- File `domande-6dicembre2024SOLUZIONI.pdf` (attenzione: contiene refusi);
- File `SCHEMA-VARIABILI-ALEATORIE-19dicembre2024.pdf`.

28 Lezione dell'11 dicembre 2024 – Generalizzazioni per variabili i.i.d. e introduzione alle variabili continue

$$\boxed{X_1, \dots, X_n \text{ i.i.d.} \iff \begin{cases} \text{indipendenti} \\ \text{identicamente distribuite} \end{cases}}$$

28.1 Generalizzazione al caso di n variabili aleatorie indipendenti

28.1.1 Caso di variabili geometriche i.i.d.

Consideriamo $X_1, X_2, \dots, X_n$ variabili aleatorie indipendenti e identicamente distribuite, ciascuna con distribuzione geometrica di parametro p, ossia:

$$P(X_i = k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \dots$$

Denotiamo con:

$$S_n := X_1 + X_2 + \cdots + X_n.$$

Distribuzione della somma. È noto che S_n ha distribuzione *Binomiale negativa*, denotata $\mathrm{NegBin}(n, p)$:

$$P(S_n = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}, \quad k = n, n+1, n+2, \dots$$

Siano $X_1, X_2, \dots, X_n$ variabili aleatorie i.i.d. con distribuzione geometrica di parametro p, cioè

$$\mathbb{P}(X_i = k) = (1-p)^{k-1}p, \qquad k = 1, 2, 3, \dots$$

e definiamo

$$S_n := X_1 + X_2 + \cdots + X_n.$$

La variabile X_i rappresenta il numero di prove necessarie per ottenere il primo successo in una sequenza di prove Bernoulliane con probabilità di successo p. Di conseguenza, S_n rappresenta il numero totale di prove necessarie per ottenere n successi.

Vogliamo calcolare

$$\mathbb{P}(S_n = k), \qquad k = n, n+1, n+2, \dots$$

Dire che $S_n = k$ significa che il n-esimo successo avviene esattamente alla prova k. Quindi:

- nelle prime $k-1$ prove devono essere avvenuti esattamente $n-1$ successi;
- la prova k deve essere un successo.

Il numero di modi per scegliere le posizioni dei $n-1$ successi nelle prime $k-1$ prove è

$$\binom{k-1}{n-1}.$$

In ogni tale sequenza ci sono n successi complessivi e $k-n$ insuccessi complessivi; poiché le prove sono indipendenti, la probabilità di una qualunque di queste sequenze è:

$$p^n(1-p)^{k-n}.$$

Dunque, moltiplicando il numero di sequenze per la probabilità di ciascuna sequenza, otteniamo:

$$\mathbb{P}(S_n = k) = \binom{k-1}{n-1} p^n(1-p)^{k-n}, \qquad k = n, n+1, n+2, \dots$$

Questo coincide con la distribuzione Binomiale Negativa $\mathrm{NegBin}(n, p)$, che descrive il numero totale di prove necessarie per ottenere n successi.

Si può interpretare S_n come il *tempo di n-esimo successo* in una sequenza di prove indipendenti, ciascuna con probabilità di successo p.

Valore atteso e varianza.

$$E[S_n] = \frac{n}{p}, \qquad \mathrm{Var}(S_n) = \frac{n(1-p)}{p^2}.$$

Distribuzione condizionata. Vogliamo descrivere la distribuzione del vettore $(X_1, X_2, \dots, X_n)$ condizionatamente al valore della somma

$$S_n = X_1 + X_2 + \dots + X_n = k.$$

Ricordiamo che le variabili X_i sono geometriche con parametro p, quindi

$$\mathbb{P}(X_i = x_i) = (1-p)^{x_i - 1}p, \qquad x_i = 1, 2, 3, \dots$$

e sono indipendenti.

Dunque, per qualunque n-upla $(x_1, \dots, x_n)$ di interi positivi che soddisfa

$$x_1 + \dots + x_n = k,$$

la probabilità congiunta è:

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = p^n(1-p)^{(x_1+\dots+x_n)-n} = p^n(1-p)^{k-n}.$$

Questa probabilità è la *stessa* per tutte le n-uple che sommano a k. Perciò, condizionando all'evento $\{S_n = k\}$, tutte queste n-uple rimangono equiprobabili.

$$(X_1, X_2, \dots, X_n) \,\big|\, (S_n = k) \sim \mathrm{Unif}\{(x_1, \dots, x_n) \in \mathbb{N}^n : x_1 + \dots + x_n = k\}.$$

Rimane da determinare quanti sono gli n-upleti di interi positivi che sommano a k. È un classico problema di partizioni con vincoli di positività.

Il numero di soluzioni intere positive dell'equazione

$$x_1 + x_2 + \dots + x_n = k$$

è dato da
$$\binom{k-1}{n-1}.$$

Infatti, ponendo $y_i = x_i - 1 \geq 0$, l'equazione diventa
$$y_1 + y_2 + \cdots + y_n = k - n,$$
che ha $\binom{k-1}{n-1}$ soluzioni non negative secondo la formula delle "stars and bars".

In conclusione: condizionatamente a $S_n = k$, il vettore $(X_1, \ldots, X_n)$ è distribuito *uniformemente* su tutti gli $\binom{k-1}{n-1}$ vettori di interi positivi che sommano a k.

28.1.2 Caso di variabili Poisson indipendenti (non necessariamente identiche)

Perché siano $X_1, X_2, \ldots, X_n$ variabili aleatorie indipendenti, con:
$$X_i \sim \text{Poisson}(\lambda_i), \qquad \lambda_i > 0.$$

Allora la loro somma:
$$S_n := X_1 + X_2 + \cdots + X_n$$
ha distribuzione:
$$S_n \sim \text{Poisson}(\lambda_1 + \lambda_2 + \cdots + \lambda_n).$$

Distribuzione condizionata. Condizionatamente a $S_n = k$, il vettore $(X_1, X_2, \ldots, X_n)$ ha distribuzione multinomiale:
$$(X_1, X_2, \ldots, X_n) \mid (S_n = k) \sim \text{Multin}\left(k; \frac{\lambda_1}{\Lambda}, \frac{\lambda_2}{\Lambda}, \ldots, \frac{\lambda_n}{\Lambda}\right),$$
dove $\Lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

In altre parole:
$$P(X_1 = k_1, \ldots, X_n = k_n \mid S_n = k) = \frac{k!}{k_1!k_2!\ldots k_n!} \left(\frac{\lambda_1}{\Lambda}\right)^{k_1} \left(\frac{\lambda_2}{\Lambda}\right)^{k_2} \cdots \left(\frac{\lambda_n}{\Lambda}\right)^{k_n},$$
per $k_1 + \cdots + k_n = k$.

Osservazione. Questa proprietà è perfettamente analoga al caso binomiale:
$$X_i \sim \text{Bin}(r_i, \theta), \quad Y_i \sim \text{Bin}(s_i, \theta) \Rightarrow X + Y \sim \text{Bin}(r + s, \theta),$$
con distribuzione condizionata ipergeometrica.

28.2 Introduzione alle variabili aleatorie non necessariamente discrete

Fino a questo punto ci siamo concentrati esclusivamente su variabili aleatorie *discrete*, cioè variabili che possono assumere soltanto un insieme finito o numerabile di valori, ciascuno con una probabilità ben definita. In questi casi la descrizione del comportamento della variabile era affidata alla funzione di massa di probabilità, che elencava i valori possibili e le rispettive probabilità. Tuttavia, molte grandezze di interesse pratico non possono essere modellate in questo modo: pensiamo, ad esempio, al tempo necessario per completare un'attività, alla temperatura di un ambiente o alla posizione di un punto lungo una retta. Per trattare tali situazioni è necessario introdurre variabili aleatorie non necessariamente discrete, le quali richiedono strumenti più generali e flessibili per essere descritte.

28.2.1 Funzione di distribuzione (o di ripartizione)

La *funzione di distribuzione* di una variabile aleatoria X è definita come:

$$F_X(x) := P(X \leq x), \qquad x \in \mathbb{R}.$$

- F_X è monotona crescente;
- $\lim_{x\to-\infty} F_X(x) = 0$ e $\lim_{x\to+\infty} F_X(x) = 1$;
- F_X è continua da destra;
- Se X è continua, allora F_X è continua su tutto $\mathbb{R}$.

Una funzione si dice *monotona crescente* se, per ogni coppia di punti $x_1 < x_2$, vale $F(x_1) \leq F(x_2)$; in altre parole, la funzione non diminuisce mai al crescere della variabile. Una funzione è invece *continua* in un punto x_0 se il valore del limite $\lim_{x\to x_0} F(x)$ coincide con $F(x_0)$: questo significa che la funzione non presenta salti o interruzioni in quel punto. La *continuità da destra* in x_0 richiede soltanto che $\lim_{x\downarrow x_0} F(x) = F(x_0)$.

28.2.2 Variabile aleatoria uniforme su $(0, 1)$

Si dice che X ha distribuzione uniforme su $(0, 1)$, e si scrive $X \sim \mathrm{Unif}(0, 1)$, se la sua densità è:

$$f_X(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{altrove.} \end{cases}$$

Funzione di distribuzione.

$$F_X(x) = \begin{cases} 0, & x \leq 0, \\ x, & 0 < x < 1, \\ 1, & x \geq 1. \end{cases}$$

Valore atteso e varianza.

$$E[X] = \int_0^1 x\,dx = \frac{1}{2}, \qquad \mathrm{Var}(X) = \int_0^1 (x - 1/2)^2\,dx = \frac{1}{12}.$$

28.2.3 Variabile aleatoria esponenziale di parametro $\lambda > 0$

Si dice che X ha distribuzione esponenziale di parametro $\lambda > 0$, e si scrive $X \sim \mathrm{Exp}(\lambda)$, se la sua densità è:

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

Funzione di distribuzione.

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

La funzione di distribuzione e la densità della variabile aleatoria esponenziale sono collegate dalla relazione generale tra densità e funzione di ripartizione. In particolare, poiché X è una variabile aleatoria continua, la sua funzione di distribuzione si ottiene integrando la densità:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt.$$

Nel caso della distribuzione esponenziale, la densità è nulla per $t < 0$, mentre per $t \geq 0$ vale $f_X(t) = \lambda e^{-\lambda t}$. Di conseguenza, per $x < 0$ l'integrale è nullo e si ottiene $F_X(x) = 0$, mentre per $x \geq 0$ si ha:

$$F_X(x) = \int_0^x \lambda e^{-\lambda t}\, dt = 1 - e^{-\lambda x}.$$

La funzione di distribuzione è dunque l'integrale cumulativo della densità, e la densità si può recuperare derivando la funzione di distribuzione:

$$f_X(x) = F'_X(x) \quad \text{per ogni } x \geq 0.$$

Valore atteso e varianza.

$$E[X] = \frac{1}{\lambda}, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Proprietà di mancanza di memoria.

$$P(X > s + t \mid X > s) = P(X > t), \qquad \forall s, t \geq 0.$$

È l'analogo continuo della proprietà vista per le variabili geometriche.

28.3 Esercizi di riferimento

Gli esempi discussi e i calcoli di questa lezione si trovano nei seguenti materiali:

- Esercizi 4 e 5 del file `domande-6dicembre2024-solo testo.pdf`;
- File `domande-6dicembre2024SOLUZIONI.pdf` e `domande-6dicembre2024-BOZZA-soluzioni.pdf` (con refusi);
- File `SCHEMA-VARIABILI-ALEATORIE-19dicembre2024.pdf`, aggiornato e corretto rispetto alla versione del 16 novembre;
- Esercizi proposti 8.1, 8.2 e 8.4 del materiale [SN];
- File `ESERCIZI-MAX-SOMMA-8-1_8-2_verif8_4.pdf`.

Osservazione finale. Con questa lezione si conclude la parte dedicata alle *variabili aleatorie discrete* e si introduce formalmente la teoria delle *variabili aleatorie continue*, che sarà approfondita nelle lezioni successive (in particolare con le distribuzioni Uniforme, Esponenziale e Normale).

29 Lezione del 12 dicembre 2024 – Variabili aleatorie continue e spazi di probabilità generali

29.1 Spazi di probabilità generali

Quando lo spazio campionario Ω non è numerabile (finito o infinito), una probabilità non può essere definita su *tutti* i sottoinsiemi di Ω. Si introduce allora la seguente definizione.

[Spazio di probabilità] Uno *spazio di probabilità* è una terna:

$$(\Omega, \mathcal{F}, \mathbb{P})$$

dove:

- Ω è lo spazio campionario;
- $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ è una σ-*algebra* di sottoinsiemi di Ω (cioè una famiglia di insiemi, detti *eventi*, tale che):

1. $\Omega \in \mathcal{F}$;
2. se $A \in \mathcal{F}$ allora anche $A^c \in \mathcal{F}$;
3. se $A_n \in \mathcal{F}$ per ogni $n \geq 1$, allora $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

- $\mathbb{P} : \mathcal{F} \to [0, 1]$ è una funzione che assegna probabilità agli eventi, *numerabilmente additiva*:

$$\text{se } A_i \cap A_j = \varnothing \text{ per } i \neq j, \quad \Rightarrow \quad \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Solo gli elementi di $\mathcal{F}$ si chiamano **eventi**, e la probabilità $\mathbb{P}$ è definita solo su di essi.

29.1.1 Proprietà di continuità della probabilità

[Continuità della probabilità] Siano (A_n) e (B_n) due successioni di eventi tali che:

- $A_n \subseteq A_{n+1}$ per ogni $n \geq 1$, allora:

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

- $B_n \supseteq B_{n+1}$ per ogni $n \geq 1$, allora:

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

Questa proprietà è detta *continuità della probabilità* ed è discussa in dettaglio nella Proposizione 15.1 di [SN] e nella Sezione 2.6 del ROSS.

29.2 Variabile aleatoria in spazi generali

[Variabile aleatoria] Una *variabile aleatoria* è una funzione $X : \Omega \to \mathbb{R}$ tale che, per ogni $x \in \mathbb{R}$,

$$\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}.$$

Questa condizione garantisce che la probabilità $\mathbb{P}(X \leq x)$ sia ben definita per ogni x, e permette di introdurre la:

[Funzione di distribuzione o di ripartizione] La funzione di distribuzione (o di ripartizione) di X è la funzione:

$$F_X(x) := \mathbb{P}(X \leq x), \quad x \in \mathbb{R}.$$

29.2.1 Proprietà della funzione di distribuzione

1. $0 \leq F_X(x) \leq 1$ per ogni $x \in \mathbb{R}$;
2. F_X è monotona crescente: se $x \leq y$ allora $F_X(x) \leq F_X(y)$;
3. F_X è continua da destra e ammette limiti a sinistra, con:

$$\mathbb{P}(X = x) = F_X(x) - F_X(x^-);$$

4. Normalizzazione:

$$\lim_{x\to-\infty} F_X(x) = 0, \quad \lim_{x\to+\infty} F_X(x) = 1.$$

29.3 Variabili aleatorie assolutamente continue

Una variabile aleatoria X si dice *assolutamente continua* quando la sua probabilità non è concentrata su punti isolati, ma è distribuita in modo continuo lungo la retta reale. Ciò significa che esiste una funzione non negativa

$$f_X : \mathbb{R} \to [0, +\infty),$$

detta *densità di probabilità*, tale che:

1. l'area totale sotto il grafico di f_X è uguale a 1, cioè

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

2. la probabilità che X cada in un intervallo $[a, b]$ si ottiene integrando la densità su tale intervallo:

$$P(a \leq X \leq b) = \int_a^b f_X(t)\, dt.$$

Da questa proprietà segue che la funzione di distribuzione

$$F_X(x) = P(X \leq x)$$

è ottenibile come integrale cumulato della densità:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

In particolare, la densità rappresenta la "derivata" della funzione di distribuzione (quando questa è derivabile), mentre la probabilità di un singolo punto è sempre nulla:

$$P(X = x) = 0, \qquad \forall x \in \mathbb{R}.$$

29.3.1 Esempio: Variabile uniforme su un intervallo

Se $X \sim \mathrm{Unif}(a, b)$, la densità è:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{altrove.} \end{cases}$$

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

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

29.3.2 Esempio: Variabile esponenziale di parametro $\lambda > 0$

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

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

$$E[X] = \frac{1}{\lambda}, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

29.3.3 Esempio: Variabile di Cauchy

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \qquad x \in \mathbb{R}.$$

Questa variabile è continua ma non ammette valore atteso né varianza.

29.4 Esercizi svolti

Esercizio 1. Sia X una variabile aleatoria uniforme sull'intervallo $[-2, 10]$. Calcolare:

$$P(-5 < X < 4), \quad P(|X| \geq 2), \quad P(X = 5).$$

Soluzione. La densità è:

$$f_X(x) = \begin{cases} \frac{1}{12}, & -2 \leq x \leq 10, \\ 0, & \text{altrove.} \end{cases}$$

Allora:

$$P(-5 < X < 4) = \int_{-2}^{4} \frac{1}{12}\, dx = \frac{6}{12} = \frac{1}{2}.$$

$$P(|X| \geq 2) = P(X \leq -2) + P(X \geq 2) = 0 + \int_{2}^{10} \frac{1}{12}\, dx = \frac{8}{12} = \frac{2}{3}.$$

Poiché X è continua, $P(X = 5) = 0$.

Esercizio 2. Sia X una v.a. con densità:

$$f(x) = \begin{cases} cx^2, & 1 \leq x \leq 5, \\ 0, & \text{altrove,} \end{cases}$$

dove c è una costante.

a) Determinare c.

$$1 = \int_{-\infty}^{+\infty} f(x)\, dx = c \int_{1}^{5} x^2\, dx = c \cdot \frac{5^3 - 1^3}{3} = c \cdot \frac{124}{3} \Rightarrow c = \frac{3}{124}.$$

b) Calcolare $P(X > 3)$.

$$P(X > 3) = \int_{3}^{5} f(x)\, dx = \frac{3}{124} \int_{3}^{5} x^2\, dx = \frac{3}{124} \cdot \frac{5^3 - 3^3}{3} = \frac{98}{124} = \frac{49}{62}.$$

29.5 Osservazioni finali

- Il passaggio da Ω discreto a Ω non numerabile richiede l'uso della σ-algebra per rendere rigoroso il concetto di evento.
- Le proprietà di continuità della probabilità giustificano la definizione di funzione di distribuzione come limite di probabilità di insiemi crescenti o decrescenti.
- Le variabili assolutamente continue ammettono una densità f_X, che consente di calcolare le probabilità mediante integrali:

$$P(a \leq X \leq b) = \int_{a}^{b} f_X(x)\, dx.$$

Riferimenti:

- Proposizione 15.1 di [SN];
- Sezione 2.6 del ROSS;
- File `SCHEMA-VARIABILI-ALEATORIE-19dicembre2024.pdf` (versione corretta);
- Esercizi del file `domande-12dicembre2024.pdf`.

30 Lezione del 13 dicembre 2024 – Variabili Normali e analogie tra discrete e continue

30.1 Richiami e caratterizzazione delle funzioni di distribuzione

Abbiamo ripreso la definizione generale di variabile aleatoria e di funzione di distribuzione

$$F_X(x) := \mathbb{P}(X \leq x),$$

e ricordato le quattro proprietà fondamentali:

[label=**(0)**]$0 \leq F_X(x) \leq 1$ per ogni $x \in \mathbb{R}$; F_X è monotona crescente in senso lato; F_X è continua da destra e ammette limiti a sinistra, con

$$\mathbb{P}(X = x) = F_X(x) - F_X(x^-);$$

$F_X(x)$ è normalizzata:

$$\lim_{x\to-\infty} F_X(x) = 0, \qquad \lim_{x\to+\infty} F_X(x) = 1.$$

[Caratterizzazione delle funzioni di distribuzione] Se $F : \mathbb{R} \to [0, 1]$ è una funzione che soddisfa le proprietà (1)–(4), allora esiste uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ e una variabile aleatoria X tale che:

$$F(x) = \mathbb{P}(X \leq x), \quad \forall x \in \mathbb{R}.$$

(Dimostrazione omessa.)

30.2 Analogie tra variabili discrete e continue

1.

Variabili discrete	**Variabili continue**
Densità discreta $p_X(x) = \mathbb{P}(X = x)$	Densità continua $f_X(x)$ con $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$
$\mathbb{P}(X \in A) = \sum_{x\in A} p_X(x)$	$\mathbb{P}(X \in A) = \int_A f_X(x)\,dx$
$\mathbb{E}[X] = \sum_x x\,p_X(x)$	$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x\,f_X(x)\,dx$
$\mathrm{Var}(X) = \sum_x (x - \mathbb{E}[X])^2 p_X(x)$	$\mathrm{Var}(X) = \int_{-\infty}^{+\infty} (x - \mathbb{E}[X])^2 f_X(x)\,dx$

Abbiamo inoltre ricordato che per alcune variabili, come la *Cauchy* con densità

$$f(x) = \frac{1}{\pi(1 + x^2)}, \quad x \in \mathbb{R},$$

il valore atteso non esiste (e quindi nemmeno la varianza).

30.3 Variabili aleatorie Normali (o Gaussiane)

30.3.1 Definizione

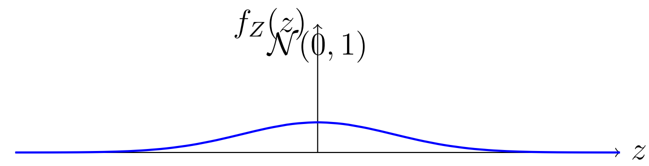
30.4 Distribuzioni Normali

[Normale standard] Una variabile aleatoria Z si dice *normale standard* se ha media 0 e varianza 1, e la sua densità è data da

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad z \in \mathbb{R}.$$

Questa densità è simmetrica rispetto a 0, ha forma di "campana" ed è il modello fondamentale da cui derivano tutte le altre distribuzioni normali.

$$E[Z] = 0, \qquad (Z) = 1.$$



30.4.1 Variabile normale generale

Siano $\mu \in \mathbb{R}$ e $\sigma > 0$. Una variabile aleatoria Y si dice *normale con media* μ *e varianza* σ^2 se

$$Y = \mu + \sigma Z, \qquad Z \sim \mathcal{N}(0, 1).$$

La trasformazione $y = \mu + \sigma z$ "sposta" la campana (di μ) e la rende più larga o più stretta (di fattore σ).

Densità. La densità di Y è

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \qquad y \in \mathbb{R}.$$

Valore atteso e varianza. Usando $Y = \mu + \sigma Z$,

$$E[Y] = \mu, \qquad (Y) = \sigma^2.$$

Funzione di distribuzione. Sia $Y \sim \mathcal{N}(\mu, \sigma^2)$. Allora, per ogni $y \in \mathbb{R}$,

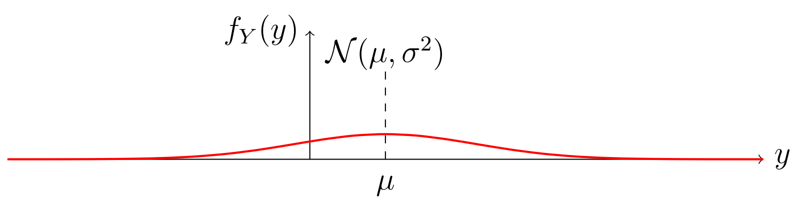
$$Y \le y \quad \Longleftrightarrow \quad Z \le \frac{y - \mu}{\sigma},$$

dove $Z \sim \mathcal{N}(0, 1)$. Da cui segue che la funzione di distribuzione di Y è

$$F_Y(y) = \Pr(Y \le y) = \Phi\left(\frac{y - \mu}{\sigma}\right),$$

dove Φ è la funzione di distribuzione della normale standard:

$$\Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt.$$



Interpretazione geometrica.

- μ determina il *centro della campana*;
- σ controlla la *larghezza*: più σ è grande, più la curva è "schiacciata" e larga;
- la trasformazione $Y = \mu + \sigma Z$ permette di ridurre ogni problema sulla normale generale allo studio della normale standard.

Osservazioni.

- La trasformazione $Y = \mu + \sigma Z$ mostra che la normale è stabile per trasformazioni lineari.
- Gli stessi risultati (sul valore atteso e sulla varianza) valgono per qualunque variabile Z nota, non solo per quella normale.

30.5 Esercizi svolti

Esercizio 1. Sia X una variabile aleatoria esponenziale con $E[X] = \frac{1}{2}$.

[label=()]Determinare il parametro λ. Calcolare la mediana, ossia il valore x_m tale che $P(X \leq x_m) = P(X > x_m)$.

Soluzione. Per la distribuzione esponenziale vale $E[X] = 1/\lambda$, dunque:

$$\frac{1}{\lambda} = \frac{1}{2} \quad \Rightarrow \quad \lambda = 2.$$

La funzione di distribuzione è:

$$F_X(x) = 1 - e^{-2x}, \quad x \geq 0.$$

La mediana x_m soddisfa $F_X(x_m) = \frac{1}{2}$, quindi:

$$1 - e^{-2x_m} = \frac{1}{2} \quad \Rightarrow \quad e^{-2x_m} = \frac{1}{2} \quad \Rightarrow \quad x_m = \frac{\ln 2}{2}.$$

Esercizio 2. Sia X una variabile aleatoria continua con densità

$$f(x) = k\, e^{-3|x|}, \quad x \in \mathbb{R}.$$

Determinare:

[label=()]il valore della costante k; la probabilità $P(|X| \leq 1)$.

Soluzione. Essendo f simmetrica rispetto a 0, calcoliamo:

$$1 = \int_{-\infty}^{+\infty} f(x)\, dx = 2k \int_0^{+\infty} e^{-3x}\, dx = 2k \cdot \frac{1}{3} = \frac{2k}{3}.$$

Quindi $k = \frac{3}{2}$.

Per la seconda domanda:

$$P(|X| \leq 1) = \int_{-1}^{1} f(x)\, dx = 2 \int_0^1 \frac{3}{2} e^{-3x}\, dx = 3 \left[-\frac{1}{3} e^{-3x} \right]_0^1 = 1 - e^{-3}.$$

30.6 Osservazioni conclusive

- Le proprietà (0)–(3) non solo sono necessarie, ma anche sufficienti per definire una funzione di distribuzione.
- Le variabili Normali costituiscono un modello fondamentale per la probabilità continua: sono stabili rispetto a somme (teorema del limite centrale).
- La normale standard Z e le sue trasformazioni lineari $Y = \mu + \sigma Z$ vengono usate come riferimento universale nelle applicazioni statistiche.

Riferimenti:

- File `SCHEMA-VARIABILI-ALEATORIE-19dicembre2024.pdf`;
- Tavole della funzione $\Phi(x)$ per la distribuzione normale standard;
- Esercizi 1 e 2 di questa lezione (svolti in classe).

31 Lezione del 18 dicembre 2024 – Teorema del Limite Centrale e Approssimazione Normale

31.1 Richiami: spazi di probabilità e funzioni di distribuzione

Negli spazi di probabilità generali, un esperimento aleatorio è descritto da una terna

$$(\Omega, \mathcal{F}, \mathbb{P}),$$

dove $\mathcal{F}$ è una *sigma-algebra* di sottoinsiemi di Ω e rappresenta la famiglia degli *eventi osservabili.*

Interpretazione intuitiva. Possiamo pensare a $\mathcal{F}$ come all'insieme delle situazioni in cui possiamo effettivamente dire se un evento si è verificato oppure no, dato il nostro stato di informazione. Si consideri il lancio di un dado, ma con le facce pari coperte di rosso e quelle dispari di blu. In tale contesto, non è possibile sapere se sia uscito precisamente il numero 2, ma solo se è uscito un numero pari oppure dispari. Gli eventi "pari" e "dispari" appartengono quindi alla sigma-algebra, mentre l'evento "è uscito 2" no.

Trasformazioni di variabili. Se $Y = a + bX$ con $b > 0$, allora

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X \leq (y-a)/b) = F_X\left(\frac{y-a}{b}\right).$$

Se X è assolutamente continua, allora $f_X(x) = F'_X(x)$ (derivata in senso debole).

31.2 Distribuzione Normale e calcolo di probabilità

Per una variabile $W \sim \mathcal{N}(\mu, \sigma^2)$, la funzione di distribuzione si ottiene tramite quella della normale standard:

$$\mathbb{P}(a < W \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right),$$

dove

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt.$$

Tali valori si leggono nelle *tabelle della Gaussiana standard.*

31.3 Approssimazione normale e teorema del limite centrale

Gli argomenti principali sono tratti dal file `APPROSS-NORMALE-PROB1-INFORM-2024-25.pdf`.

[Teorema del Limite Centrale – versione classica] Siano $X_1, X_2, \dots, X_n$ variabili aleatorie indipendenti e identicamente distribuite, con

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2 < +\infty.$$

Sia

$$S_n = X_1 + X_2 + \cdots + X_n, \qquad Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Allora, per ogni $x \in \mathbb{R}$,

$$\lim_{n\to+\infty} \mathbb{P}(Z_n \leq x) = \Phi(x),$$

cioè Z_n converge in distribuzione a una variabile normale standard $\mathcal{N}(0,1)$.

[Versione di De Moivre–Laplace (caso binomiale)] Sia $S_n \sim \mathrm{Bin}(n,p)$ e siano

$$\mu = np, \qquad \sigma^2 = np(1-p),$$

e la variabile standardizzata

$$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}.$$

Allora, per ogni $x \in \mathbb{R}$,

$$\lim_{n\to+\infty} \mathbb{P}(S_n^* \le x) = \Phi(x).$$

Interpretazione. Per n grandi, la densità discreta della distribuzione binomiale $\mathrm{Bin}(n,p)$ si avvicina alla densità continua della distribuzione normale $\mathcal{N}(np, np(1-p))$.

Osservazioni pratiche.

- Si può usare la normale per approssimare probabilità binomiali anche per valori moderati di n, a patto che $np \ge 5$ e $n(1-p) \ge 5$.
- Nel calcolo di probabilità approssimate si introduce talvolta la *correzione di continuità*:

$$\mathbb{P}(S_n = k) \approx \Phi\left(\frac{k + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - 0.5 - np}{\sqrt{np(1-p)}}\right).$$

31.4 Esercizi e riferimenti

- **Dal libro di Ross** (`Dal libro di ROSS-18-dic-2024.pdf`), Capitolo 5:
  - Esercizio 4 (tranne il punto (c));
  - Esercizi 7, 8 e 11;
  - Esempi 4b e 4c.
- **Dal file** `Esercizi-su-variabili-continue.pdf`: Esercizi 1, 2, 3, 5 (punti 1) e 2)), 6, 7. È stato suggerito, solo a voce, di svolgere anche gli Esercizi 8 e 9. Gli Esercizi 10 e 11 sono stati invece svolti a lezione. Versione aggiornata con commenti: `Esercizi-su-variabili-continue-20dic-202`

31.5 Osservazioni finali

- Il Teorema del Limite Centrale giustifica l'uso della normale come modello universale per la somma di molte variabili indipendenti, anche non gaussiane.
- La versione di De Moivre–Laplace è il caso particolare per variabili binomiali.
- In pratica, ogni volta che una somma o media coinvolge molti contributi indipendenti e non troppo asimmetrici, la distribuzione risultante si approssima a una normale.

32 Lezione del 19 dicembre 2024 – Legge dei Grandi Numeri, Teorema Centrale del Limite e Intervalli di Confidenza

32.1 Collegamento tra Legge dei Grandi Numeri (LGN) e Teorema Centrale del Limite (TCL)

Legge dei Grandi Numeri (forma debole). Siano $X_1, X_2, \dots, X_n$ variabili aleatorie indipendenti e identicamente distribuite (i.i.d.) con valore atteso μ e varianza $\sigma^2 < +\infty$. Posto

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

si ha:

$$\forall \varepsilon > 0, \quad \lim_{n\to+\infty} \mathbb{P}\big(|\overline{X}_n - \mu| > \varepsilon\big) = 0.$$

In altre parole, $\overline{X}_n$ converge in probabilità a μ.

Disuguaglianza di Chebyshev. La LGN può essere dimostrata o motivata tramite la disuguaglianza di Chebyshev:

$$\mathbb{P}\big(|\overline{X}_n - \mu| \geq \varepsilon\big) \leq \frac{\mathrm{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

Il termine a destra tende a zero per $n \to +\infty$, mostrando la concentrazione della media intorno a μ.

Teorema Centrale del Limite (TCL). Il TCL fornisce una descrizione più fine del comportamento di $\overline{X}_n$:

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1),$$

ossia la variabile standardizzata converge in distribuzione alla normale standard. Pertanto, per grandi n,

$$\overline{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Legame tra LGN e TCL.

- La LGN assicura che la media campionaria tende a μ con probabilità prossima a 1 (stabilità a lungo termine).
- Il TCL quantifica *quanto velocemente* la media campionaria si concentra intorno a μ, fornendo un'approssimazione distribuzionale utile per stimare probabilità e intervalli.

32.2 Intervalli di Confidenza

Con la disuguaglianza di Chebyshev. Poiché

$$\mathbb{P}\left(|\overline{X}_n - \mu| < k\frac{\sigma}{\sqrt{n}}\right) \geq 1 - \frac{1}{k^2},$$

possiamo dire che, con confidenza almeno $1 - \frac{1}{k^2}$,

$$\mu \in \left[\overline{X}_n - k\frac{\sigma}{\sqrt{n}},\ \overline{X}_n + k\frac{\sigma}{\sqrt{n}}\right].$$

Esempio: per $k = 2$, confidenza almeno 75%; per $k = 3$, almeno 88.9%.

Con l'approssimazione del TCL. Se n è sufficientemente grande, allora

$$\mathbb{P}\left(|\overline{X}_n - \mu| \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \approx 1 - \alpha,$$

dove $z_{\alpha/2}$ è il quantile della normale standard tale che $\Phi(z_{\alpha/2}) = 1 - \alpha/2$.

Esempi:

$$z_{0.025} = 1.96 \quad \text{(intervallo di confidenza al 95\%)}, \qquad z_{0.005} = 2.58 \quad \text{(intervallo al 99\%)}.$$

Quindi, per grandi n,

$$\mu \in \left[\overline{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

è un intervallo di confidenza approssimato al livello $1 - \alpha$.

(Esempi numerici e figure tratti dal file `APPROSS-NORMALE-PROB1-INFORM-2024-25.pdf`, Esercizi 16.1 e 16.2.)

32.3 Esercizi e applicazioni

Esercizi di simulazione (file `Esercizi-su-variabili-continue-20dic-2024-con-commenti.pdf`).

- **Esercizio 7:** simulazione numerica della media di campioni casuali e verifica empirica della Legge dei Grandi Numeri.
- **Esercizi 8 e 9:** simulazioni per osservare la convergenza empirica alla distribuzione normale (verifica sperimentale del TCL).

Trasformazioni di variabili aleatorie.

- **Esercizio 4:** se X è una variabile aleatoria con funzione di distribuzione $F_X(x)$ e $Y = h(X)$, si calcola
  $$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(h(X) \leq y).$$
  Caso particolare:
  $$h(x) = x^3 \quad \text{oppure} \quad h(x) = x^2.$$
  Nel caso di h monotona crescente (dimostrazione svolta a lezione), si ha:
  $$F_Y(y) = F_X(h^{-1}(y)).$$
  Per la generalizzazione ai casi in cui h è invertibile a tratti, vedere la Sezione 12 del file `SCHEMA-VARIABILI-ALEATORIE-19dicembre2024.pdf`.

Esercizio teorico 12: variabili aleatorie simmetriche. Una variabile X si dice *simmetrica* rispetto a $a \in \mathbb{R}$ se la sua densità o funzione di distribuzione soddisfa

$$f_X(a + x) = f_X(a - x), \quad \forall x.$$

In tal caso, $E(X) = a$ (se esiste). Sono simmetriche, ad esempio:

- la variabile normale $N(\mu, \sigma^2)$, simmetrica rispetto a μ;
- la variabile uniforme su $[a, b]$, simmetrica rispetto a $(a + b)/2$;
- la variabile di Cauchy centrata, simmetrica rispetto a 0 (anche se senza valore atteso).

Schema Finale: Valori Attesi, Varianze, Covarianze e Formule Condizionate

Valore atteso

Per una variabile aleatoria discreta X con densità p_X:

$$E[X] = \sum_x x\, p_X(x).$$

Lineare:

$$E[aX + bY + c] = aE[X] + bE[Y] + c.$$

Per funzione:

$$E[g(X)] = \sum_x g(x)\, p_X(x).$$

Valori attesi notevoli

$$E[X_{\text{Bernoulli}(p)}] = p$$
$$E[X_{\text{Bin}(n,p)}] = np$$
$$E[X_{\text{Iperg}(N,M,n)}] = n\frac{M}{N}$$
$$E[X_{\text{Pois}(\lambda)}] = \lambda$$
$$E[X_{\text{Geom}(p)}] = \frac{1}{p}$$
$$E[X_{\text{Unif}(a,b)}] = \frac{a+b}{2}$$
$$E[X_{\text{Exp}(\lambda)}] = \frac{1}{\lambda}$$
$$E[X_{\mathcal{N}(\mu,\sigma^2)}] = \mu$$

Varianza

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

Scalatura:

$$\text{Var}(aX + b) = a^2\text{Var}(X).$$

Varianze notevoli

$$\text{Var}(\text{Bernoulli}(p)) = p(1-p)$$
$$\text{Var}(\text{Bin}(n,p)) = np(1-p)$$
$$\text{Var}(\text{Iperg}(N,M,n)) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$$
$$\text{Var}(\text{Pois}(\lambda)) = \lambda$$
$$\text{Var}(\text{Geom}(p)) = \frac{1-p}{p^2}$$
$$\text{Var}(\text{Unif}(a,b) = \frac{(b-a)^2}{12}$$
$$\text{Var}(\text{Exp}(\lambda)) = \frac{1}{\lambda^2}$$
$$\text{Var}(\mathcal{N}(\mu,\sigma^2)) = \sigma^2$$

Covarianza

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

Proprietà:

$$\text{Cov}(aX + b, cY + d) = ac\,\text{Cov}(X,Y).$$

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

Indipendenza:

$$X \perp Y \quad \Rightarrow \quad \text{Cov}(X,Y) = 0.$$

Correlazione

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$

Valore atteso condizionato

$$E[X \mid Y = y] = \sum_x x\, p_{X|Y}(x \mid y).$$

Legge del valore atteso totale (tower property):

$$E[X] = E[\,E[X \mid Y]\,].$$

Varianza condizionata

$$\mathrm{Var}(X) = E[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(E[X \mid Y]).$$

Covarianza condizionata

$$\mathrm{Cov}(X,Y) = E[\mathrm{Cov}(X,Y \mid Z)] + \mathrm{Cov}(E[X \mid Z], E[Y \mid Z]).$$

Indipendenza

Variabili indipendenti:

$$p_{X,Y}(x,y) = p_X(x)\, p_Y(y).$$

$$E[XY] = E[X]E[Y], \quad \mathrm{Cov}(X,Y) = 0.$$

Somma di indipendenti:

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i).$$

Distribuzioni Condizionate Fondamentali

Binomiale

$$X \mid (X+Y=n) \sim \mathrm{Bin}\left(n, \frac{p_X}{p_X + p_Y}\right).$$

Poisson Se $X \sim \mathrm{Pois}(\lambda_1)$, $Y \sim \mathrm{Pois}(\lambda_2)$ indipendenti:

$$X + Y \sim \mathrm{Pois}(\lambda_1 + \lambda_2),$$

$$X \mid (X+Y=n) \sim \mathrm{Bin}\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right).$$

Geometrica

$$X + Y \text{ (geom. indipendenti)} \implies \text{neg-bin}.$$

Densità congiunta e valore atteso di funzioni

$$E[g(X,Y)] = \sum_{x,y} g(x,y)\, p_{X,Y}(x,y).$$

Se X, Y indipendenti:

$$p_{X,Y}(x,y) = p_X(x)p_Y(y).$$

Varianza della somma (caso generale)

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

Contatti e Risorse

daromacarlo@gmail.com