Probabilità e Statistica

Isadora Antoniano-Villalobos
isadora.antoniano@unive.it

Laurea in Informatica
(Data science/ Tecnologie e scienze dell'informazione)
Università Ca' Foscari di Venezia

Anno academico 2024/2025

Somme di variabili e teoremi limite
Testo Capitoli 5, 9, 18

Somma di v.a.

Spesso, come abbiamo già visto in alcuno esempi, siamo interessati alla somma di due o più variabili aleatorie.

- Sappiamo già come calcolare il valore atteso e la varianza di una somma di v.a..
- Tuttavia, calcolare l'intera distribuzione di una somma di v.a. può essere un'impresa davvero difficile.

Esempio: somma di Poisson indipendenti

Siano X e Y due v.a. indipendenti con distribuzione di Poisson di media rispettivamente λ e μ.

Somma di v.a.

- Siamo interessati a trovare la distribuzione di $Z = X + Y$.

$$\begin{aligned}\mathbb{P}[Z=z] &= \sum_{x=0}^{z} \mathbb{P}[X=x, Y=z-x] = \sum_{x=0}^{z} \mathbb{P}[X=x]\,\mathbb{P}[Y=z-x] \\ &= \sum_{x=0}^{z} \frac{\lambda^x}{x!} e^{-\lambda} \frac{\mu^{z-x}}{(z-x)!} e^{-\mu} = \frac{e^{-(\lambda+\mu)}}{z!} \sum_{x=0}^{z} \frac{z!}{x!(z-x)!} \lambda^x \mu^{z-x} \\ &= \frac{e^{-(\lambda+\mu)}}{z!} (\lambda+\mu)^z, \qquad z = 0, 1, 2, \ldots\end{aligned}$$

Allora $Z \sim \mathcal{P}(\lambda + \mu)$.

➔ **Esercizio:** Per le variabili del lucido precedente, si calcoli la distribuzione condizionata di $X|Z = z$.
Si tratta di una distribuzione nota?

Somma di v.a.

Esempio: somma di esponenziali indipendenti

Siano X e Y due v.a. indipendenti con distribuzione Esponenziale di parametro λ.

- Siamo interessati a trovare la distribuzione di $Z = X + Y$.
  Sia $z > 0$:

$$\begin{aligned}
F_Z(z) = \mathbb{P}[Z \leq z] &= \iint_{x+y\leq t} f(x,y)dx\,dy \\
&= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z-y} f_X(x)dx \right) f_Y(y)dy \\
&= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{z} f_X(t-y)dt \right) f_Y(y)dy \\
&= \int_{-\infty}^{z} \left(\int_{-\infty}^{+\infty} f_X(t-y)f_Y(y)dy \right) dt. \\
&= \int_{-\infty}^{z} f_Z(t)dt
\end{aligned}$$

Somma di v.a.

Allora la densità di Z è:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y) dy \\ &= \int_{-\infty}^{+\infty} \lambda^2 e^{-\lambda(z-y)} e^{-\lambda y} I_{(0,+\infty)}(z-y) I_{(0,+\infty)}(y) dy \\ &= \int_0^z \lambda^2 e^{-\lambda(z-y)} e^{-\lambda y} I_{(0,+\infty)}(z) dy \\ &= \int_0^z \lambda^2 e^{-\lambda z} I_{(0,+\infty)}(z) dy = \lambda^2 z e^{-\lambda z} I_{(0,+\infty)}(z), \end{aligned}$$

che coincide con la funzione di densità di una Gamma$(\alpha = 2, \lambda)$.

Somma di v.a.

In generale, valgono i seguenti risultati:

1. Se $\{X_i\}$, $i = 1 \ldots n$ sono i.i.d Binom $(1, p)$, allora $S_n = \sum_{i=1}^{n} \sim$ Binom (n, p).

2. Se $X_i \sim$ Pois (λ_i), $i = 1 \ldots n$ sono indipendenti, allora $S_n = \sum_{i=1}^{n} \sim$ Pois $(\sum_i \lambda_i)$.

3. Se $\{X_i\}$, $i = 1 \ldots n$ sono i.i.d Exp (λ), allora $S_n = \sum_{i=1}^{n} \sim$ Ga (n, λ).

4. Se $X_i \sim$ N (μ_i, σ_i^2) sono indipendenti, allora $S_n = \sum_{i=1}^{n} \sim$ N $(\sum_i \mu_i, \sum_i \sigma_i^2)$.

La media campionaria

Media campionaria

Siano $X_1, \ldots, X_n$ n v.a. indipendenti e tutte con la stessa distribuzione (i.i.d.). Scriviamo $\mathbb{E}[X_i] = \mu$ e $\mathrm{Var}[X_i] = \sigma^2$.
La **media campionaria** si definisce come

$$\overline{X}_n = \frac{\sum_i X_i}{n} = \frac{S_n}{n}$$

La media campionaria è una v.a. molto importante nelle applicazioni del calcolo delle probabilità in statistica.

- La sua distribuzione, come quella di S_n dipende dalla distribuzione delle X_i e, in generale, non è facile da calcolare.

La media campionaria

- Si possono però calcolare facilmente la media e la varianza di somma e media campionaria:

$$\mathbb{E}\left[S_n\right] = \mathbb{E}\left[\sum_i X_i\right] = \sum_i \mathbb{E}\left[X_i\right] = n\mu,$$

$$\mathsf{Var}\left[S_n\right] = \mathsf{Var}\left[\sum_i X_i\right] = \sum_i \mathsf{Var}\left[X_i\right] = n\sigma^2$$

$$\mathbb{E}\left[\overline{X}_n\right] = \mathbb{E}\left[\frac{S_n}{n}\right] = \frac{\mathbb{E}\left[S_n\right]}{n} = \frac{n\mu}{n} = \mu,$$

$$\mathsf{Var}\left[\overline{X}_n\right] = \mathsf{Var}\left[\frac{S_n}{n}\right] = \frac{\mathsf{Var}\left[S_n\right]}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Due importanti teoremi

I due più importanti teoremi del calcolo delle probabilità parlano proprio della media campionaria e della sua distribuzione. Sono:

- La Legge dei grandi numeri
- Il Teorema del limite centrale.

Per illustrarli, dobbiamo introdurre un risultato preliminare e due concetti di convergenza di successioni di v.a..

La disuguaglianza di Chebyshev

Disuguaglianza di Chebyshev

Sia X una v.a. con valore atteso $\mathbb{E}[X]$ e $\mathsf{Var}[X] < +\infty$. Allora, $\forall \epsilon > 0$,

$$\mathbb{P}[|X - \mathbb{E}[X]| > \epsilon] \leq \frac{\mathsf{Var}[X]}{\epsilon^2}.$$

Esempio: Il numero di richieste giornaliere di collegamento ad un server è una v.a. Y con valore atteso 130 e varianza 50. Qual è la probabilità che in un giorno si colleghino fra i 100 e i 160 clienti?

$$\mathbb{P}[100 \leq Y \leq 160] = \mathbb{P}[|Y - 130| \leq 30] \geq 1 - \frac{50}{30^2} = 0.9444.$$

Convergenza di v.a.

Convergenza di v.a.

Sia $X_1, \dots, X_n, \dots$ una sequenza di v.a. con f.r. F_n e X un'ulteriore v.a. con f.r. F.

1. Diciamo che $\{X_n\}$ **converge in probabilità** a X e scriviamo $X_n \xrightarrow{p} X$, se
$$\mathbb{P}\left[|X_n - X| > \epsilon\right] \to 0, \quad \forall \epsilon > 0.$$
2. Diciamo che $\{X_n\}$ **converge in distribuzione** a X e scriviamo $X_n \xrightarrow{d} X$, se in ogni punto x di continuità per F
$$F_n(x) \to F(x).$$
3. Diciamo che $\{X_n\}$ **converge in quasi certamente** o **con probabilità 1** a X e scriviamo $X_n \to X \quad q.c.$, se
$$\mathbb{P}\left[X_n \to X\right] = 1.$$

Le leggi dei grandi numeri (LGN)

Legge debole dei grandi numeri

Sia $X_1, \ldots, X_n, \ldots$ una sequenza di v.a. i.i.d. con $\mathbb{E}[X_i] = \mu$ e Var $[X_i] = \sigma^2 < +\infty$. Allora

$$\overline{X}_n \xrightarrow{p} \mu.$$

Infatti, per la disuguaglianza di Chebyshev applicata alla media campionaria,

$$\mathbb{P}\left[|\overline{X}_n - \mu| > \epsilon\right] \leq \frac{\text{Var}\left[\overline{X}_n\right]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0.$$

Le leggi dei grandi numeri (LGN)

Legge forte dei grandi numeri

Sia $X_1, \ldots, X_n, \ldots$ una sequenza di v.a. i.i.d. con $\mathbb{E}[X_i] = \mu$. Allora

$$\overline{X}_n \to \mu \quad q.c..$$

La dimostrazione di questo risultato è più complessa...

Esempio: L'estrazione del Lotto

Immaginiamo di essere interessati all'evento che venga estratto il numero 53 nella ruota di Venezia. Indichiamo con X_i la variabile binaria che vale 1 se all'i-esima estrazione esce il numero 53 e 0 altrimenti. E' chiaro che

$$X_i \sim \mathcal{B}(1, p = 1/90).$$

Le leggi dei grandi numeri (LGN)

Per la legge dei grandi numeri, ci aspettiamo che la frequenza relativa delle volte che si estrae il 53 si avvicini sempre più a p al crescere del numero di estrazioni n. Infatti:

$$\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{p} \mathbb{E}\left[X_i\right] = p = 1/90.$$

Questo però non cambia la probabilità di estrarre 53 ad ogni estrazione, che rimane sempre uguale a $p = 1/90$, anche quando non si è osservato il 53 per molto tempo!

Il teorema del limite centrale (TLC)

Teorema del limite centrale (TLC)

Sia $X_1, \ldots, X_n, \ldots$ una sequenza di v.a. i.i.d. con $\mathbb{E}[X_i] = \mu$ e $\text{Var}[X_i] = \sigma^2 < +\infty$. Allora

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{n(S_n - \mu)}{\sigma\sqrt{n}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1),$$

Il TLC permette, nella pratica, di approssimare la distribuzione della media campionaria (o di una somma) con la distribuzione normale per valori grandi di n:

$$\overline{X}_n \mathrel{\dot\sim} \mathcal{N}(\mu, \sigma^2/n).$$

Ovvero

$$\sum_{i=1}^{n} X_i \mathrel{\dot\sim} \mathcal{N}(n\mu, n\sigma^2).$$

Il teorema del limite centrale (TLC)

Si osservi che l'approssimazione normale per la binomiale altro non è che un'applicazione del TLC alla somma di v.a. di Bernoulli (la binomiale).

Esempio: TLC per Poisson

Il numero di studenti iscritti a un corso di laurea segue una distribuzione di Poisson di media 100. Se gli iscritti superano i 120 bisogna sdoppiare i corsi. Qual è la probabilità che ciò accada?

- $X \sim \text{Pois}\,(100)$, perciò

$$\mathbb{P}\left[X \geq 120\right] = e^{-100} \sum_{i=120}^{+\infty} \frac{100^i}{i!} = \text{1-ppois(119,100)} = 0.0282.$$

Il teorema del limite centrale (TLC)

- Ma X si può vedere come una somma di 100 v.a. Pois (1) e la sua distribuzione si può dunque approssimare con quella normale, in virtù del TLC.
  Con la correzione per continuità, otteniamo:

$$\begin{aligned}\mathbb{P}\left[X \geq 120\right] = \mathbb{P}\left[X \geq 119.5\right] = \mathbb{P}\left[\sum_i X_i \geq 119.5\right] \\ \doteq \mathbb{P}\left[Z \geq \frac{119.5 - 100}{\sqrt{100}}\right] \\ = 1 - \Phi(1.95) = 0.0256.\end{aligned}$$

Catene di Markov
Testo Capitoli 24-25

Introduzione

Fino ad ora abbiamo considerato successioni di variabili aleatorie indipendenti e identicamente distribuite.

Spesso però nelle applicazioni capita di incontrare **sequenze di variabili dipendenti le une dalle altre**. Basti pensare a modelli probabilistici per fenomeni che si evolvono nel tempo, dove la variabile di interesse al tempo presente dipende da ciò che si è verificato nel passato.

- Le **catene di Markov** sono un modello in cui le variabili sono legate da un particolare tipo di dipendenza.

Definizioni

Catena di Markov (omogenea)

Sia $X_0, X_1, X_2, \ldots$ una successione di variabili casuali (discrete) a valori in un insieme finito (o numerabile) $S = \{1, 2, \ldots, M\}$, detto **spazio degli stati**.

$X = \{X_n\}_{n\geq 0}$ è una **catena di Markov** (omogenea) se

$$P(X_{n+1} = j \;\;|\;\; X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \\ = P(X_{n+1} = j | X_n = i) = p_{ij}.$$

➔ $P = (p_{ij})_{ij}$ è la **matrice di transizione** e i suoi elementi p_{ij} sono le **probabilità di transizione**.

La matrice di transizione

La matrice di transizione

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{pmatrix}$$

è tale che

- $p_{ij} \geq 0,\ i, j = 1 \ldots, M,$
- $\sum_{j=1}^{M} p_{ij} = 1,\ i = 1 \ldots, M.$

➔ Conoscere la matrice di transizione e la funzione di probabilità dello stato iniziale, $\pi^{(0)}$, permette di calcolare probabilità condizionate, congiunte e marginali della catena.

La matrice di transizione

Esempio: il mondo di Oz

Nel mondo di Oz ci sono tre possibili situazioni meteorologiche:
1=pioggia, 2=sole, 3=neve.

Inoltre:

- non ci sono mai due giorni consecutivi di sole;
- se oggi c'è sole, domani nevica o piove con la stessa probabilità;
- se nevica o piove, con probabilità 0.5 domani rimane invariato e con probabilità 0.5 domani cambia a caso.

➔ Se oggi c'è il sole, come sarà il tempo fra due giorni?

La matrice di transizione

- Il modello che descrive bene la situazione è una catena di Markov con spazio degli stati $S = \{1, 2, 3\}$ e matrice di transizione

$$P = \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}$$

  Lo stato iniziale è $X_0 = 2$ (sole) e ci domandiamo quanto vale $P(X_2 = j | X_0 = 2)$, $j = 1, 2, 3$.

Transizione a due passi

In generale,

$$
\begin{aligned}
& P(X_2 = j | X_0 = i) \\
& \quad = \sum_{k \in S} P(X_2 = j | X_1 = k) P(X_1 = k | X_0 = i) \\
& \quad = \sum_{k \in S} p_{kj} p_{ik} = p_{ij}^{(2)},
\end{aligned}
$$

dove $p_{ij}^{(2)}$ è l'elemento ij della matrice $P^2 = P \cdot P$.

➔ La matrice P^2 è chiamata la **matrice di transizione a due passi** della catena di Markov.

Transizione a due passi

Esempio: il mondo di Oz fra due giorni

Nel nostro esempio siamo interessati alla seconda righa di P^2:

$$p_{21}^{(2)} = P(X_2 = 1 | X_0 = 2) = \begin{pmatrix} 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1/2 \\ 1/2 \\ 1/4 \end{pmatrix} = \frac{3}{8},$$

$$p_{22}^{(2)} = P(X_2 = 2 | X_0 = 2) = \begin{pmatrix} 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1/4 \\ 0 \\ 1/4 \end{pmatrix} = \frac{1}{4},$$

$$p_{23}^{(2)} = P(X_2 = 3 | X_0 = 2) = \begin{pmatrix} 1/2 & 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1/4 \\ 1/2 \\ 1/2 \end{pmatrix} = \frac{3}{8}.$$

Transizione a n passi

Reiterando l'operazione a due passi, si ottiene la **matrice di transizione a n passi**:

$$P^n = P \cdot P \cdot \ldots \cdot P,$$

i cui elementi sono le probabilità condizionate a n passi

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i).$$

Distribuzioni marginali

Le **distribuzioni marginali** della catena si ottengono dalla funzione di probabilità iniziale della catena, $\pi^{(0)}$, e dalla matrice di transizione a n passi:

$$\begin{aligned} \pi_i^{(n)} &= P(X_n = i) = \sum_{k \in S} P(X_n = i | X_0 = k) P(X_0 = k) \\ &= \sum_{k \in S} p_{ki}^{(n)} \pi_k^{(0)}. \end{aligned}$$

Perciò,

$$\pi^{(n)} = \pi^{(0)} \cdot P^n.$$

Distribuzioni marginali

Esempio: il mondo di Oz fra due giorni

Qual è la distribuzione del tempo fra due giorni?

☞ Oggi c'è il sole! e $\Rightarrow \; \pi_2^{(0)} = 1$

$$
\begin{aligned}
\pi^{(2)} &= \pi^{(0)} \cdot P^2 \\
&= \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \begin{pmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 0 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 7/16 & 3/16 & 3/8 \\ 3/8 & 2/8 & 3/8 \\ 3/8 & 3/16 & 7/16 \end{pmatrix} \\
&= \begin{pmatrix} 3/8 & 2/8 & 3/8 \end{pmatrix}
\end{aligned}
$$

Catene regolari

Catene regolari

Una catena di Markov si dice **regolare** se esiste un indice n per cui P^n ha tutti elementi strettamente positivi.

- **Esempio (mondo di Oz):** P ha uno zero, ma P^2 no! Perciò la catena è regolare.
- **Esempio:** $P = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix}$ non è regolare (provatelo!).

Distribuzione stazionaria

Se P è regolare allora esiste $\pi = (\pi_1, \ldots, \pi_M)$ tale che

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \pi_1 & \pi_2 & \ldots & \pi_M \\ \pi_1 & \pi_2 & \ldots & \pi_M \\ \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \ldots & \pi_M \end{pmatrix}$$

Distribuzione stazionaria

Si dimostra che π è l'unica **distribuzione stazionaria** della catena, ovvero tale che

$$\pi P = \pi, \qquad \sum_{i=1}^{M} \pi_i = 1.$$

Catena stazionaria

Catena stazionaria

Se la distribuzione iniziale di X_0 è la distribuzione stazionaria π, allora

$$\pi P^n = \pi, \qquad \forall n,$$

e tutte le distribuzioni marginali della catena sono uguali a π.
Si dice allora che la catena di Markov è una **catena stazionaria**.

Esempio: distribuzione stazionaria nel mondo di Oz

☞ Provate a dimostrare che

$$\lim_n P^n = \begin{pmatrix} 0.4 & 0.2 & 0.4 \\ 0.4 & 0.2 & 0.4 \\ 0.4 & 0.2 & 0.4 \end{pmatrix}.$$

Allora $\pi = (0.4, 0.2, 0.4)$ è la distribuzione stazionaria della catena.

Catena stazionaria

Infatti, la condizione $\pi P = \pi$, con $\sum_{i=1}^{3} \pi_i = 1$, porta al sistema

$$\begin{cases} 4\pi_1 = 2\pi_1 + 2\pi_2 + \pi_3 \\ 4\pi_2 = \pi_1 + \pi_3 \\ 4\pi_3 = \pi_1 + 2\pi_2 + 2\pi_3 \\ \pi_1 + \pi_2 + \pi_3 = 1 \end{cases} \dots \begin{cases} \pi_3 = 2\pi_2 \\ \pi_1 = 2\pi_2 \\ \pi_1 + \pi_2 + \pi_3 = 1, \end{cases}$$

che dà come soluzione proprio $\pi = (2/5, 1/5, 2/5)$.

La passeggiata aleatoria

La passeggiata aleatoria

Una **passeggiata aleatoria** (*random walk* in inglese) è una catena di Markov con spazio degli stati $S = \mathbb{Z}$ che ad ogni istante si muove di un passo a destra o a sinistra con probabilità rispettivamente p e $1-p$:

$$p_{ii+1} = p, \qquad p_{ii-1} = 1-p, \qquad p_{ij} = 0 \text{ altrimenti.}$$

In queste condizioni,

- se $p > 1/2$ il sistema andrà verso $+\infty$;
- se $p < 1/2$ il sistema andrà verso $-\infty$;
- se $p = 1/2$ il suo andamento è meno prevedibile (ma per la legge dei grandi numeri, se parte da 0 tende a 0... pensateci!).

Simulazione della passeggiata aleatoria

Per simulare una **traiettoria** della passeggiata aleatoria:

1. si simulano n valori $x_1, \ldots, x_n$ da una variabile aleatoria di Bernoulli, $X \sim \mathrm{Be}(p)$;
2. si trasformano i valori simulati in $y_i = 2x_i - 1$ (tutti valori ± 1);
3. si calcolano le somme cumulate parziali del vettore $y_1, \ldots, y_n$.

La passeggiata aleatoria con barriere

Si possono aggiungere alla passeggiata aleatoria due **barriere non assorbenti**, in modo che, una volta raggiunte, il sistema rimbalzi allo stato precedente.

➔ In questo caso $S = \{-L, \ldots, -2, -1, 0, 1, 2, \ldots, L\}$, con $p_{-L,\,-L+1} = p_{L,\,L-1} = 1$:

$$P = \begin{pmatrix} 0 & 1 & \ldots & \ldots & \ldots & 0 \\ 1-p & 0 & p & 0 & \ldots & 0 \\ 0 & 1-p & 0 & p & 0 & \ldots \\ \vdots & \ldots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & 1-p & 0 & p \\ 0 & \ldots & \ldots & \ldots & 1 & 0 \end{pmatrix}$$

La passeggiata aleatoria con barriere

Alternativamente, si possono aggiungere alla passeggiata aleatoria due **barriere assorbenti**, in modo che, una volta raggiunte, il sistema non si muova più da lì.

- In questo caso $S = \{-L, \ldots, -2, -1, 0, 1, 2, \ldots, L\}$, con $p_{-L,-L} = p_{L,L} = 1$:

$$P = \begin{pmatrix} 1 & 0 & \ldots & \ldots & \ldots & 0 \\ 1-p & 0 & p & 0 & \ldots & 0 \\ 0 & 1-p & 0 & p & 0 & \ldots \\ \vdots & \ldots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & 1-p & 0 & p \\ 0 & \ldots & \ldots & \ldots & 0 & 1 \end{pmatrix}$$