

1. Gli oggetti del calcolo delle probabilità

1.1. Gli eventi

In termini intuitivi, gli eventi sono affermazioni, come ad esempio “dal lancio del dado uscirà il numero 6”. Di essi vorremo calcolare la probabilità. Su tali affermazioni è utile operare tramite connettivi logici, come “e”, “oppure”, ecc.

In termini rigorosi, *gli eventi sono sottoinsiemi* di un insieme “universo” Ω . Le operazioni insiemistiche di intersezione, unione, ecc. corrispondono a quelle logiche dette sopra.

Example 1. *Se ci interessa il risultato del lancio di un dado, le affermazioni che ci interessano riguarderanno i risultati possibili di tale lancio. Per usare il formalismo degli insiemi introduciamo l'insieme Ω dei numeri da 1 a 6:*

$$\Omega = \{1, \dots, 6\}.$$

I suoi sottoinsiemi rappresentano i possibili risultati del lancio. Ad esempio il sottoinsieme $\{6\}$ corrisponde all'affermazione “uscirà il numero 6”; il sottoinsieme $\{1, 3, 5\}$ corrisponde all'affermazione “uscirà un numero dispari”; e così via. All'operazione logica “uscirà il numero 6 oppure il numero 5” corrisponde l'operazione insiemistica $\{6\} \cup \{5\}$, e così via.

Dato quindi un insieme “universo” o “ambiente” Ω (la terminologia non è standardizzata), abbiamo detto che gli eventi sono sottoinsiemi di Ω . In qualche caso considereremo tutti i sottoinsiemi, in altri casi no, ma questo dettaglio per ora non è rilevante. Quindi per ragionare in modo semplice si immagini che consideriamo come eventi *tutti* i sottoinsiemi $A \subset \Omega$, anche se questo non è sempre possibile o consigliabile o utile.

Dato Ω , richiederemo sempre che la famiglia degli eventi sia chiusa per le usuali operazioni insiemistiche. Ad esempio, l'unione e l'intersezione di eventi deve essere un evento, e così il complementare di un evento. Richiederemo sempre, inoltre, che sia l'insieme Ω sia l'insieme vuoto \emptyset siano eventi. In aspetti più raffinati della teoria si deve richiedere che anche l'unione e l'intersezione *numerabile* di eventi sia un evento.

1.2. La probabilità

Degli eventi ci interessa calcolare la probabilità. In altre parole, ad ogni evento $A \subset \Omega$ vogliamo associare un numero $P(A)$, detto *probabilità* di A . Quindi, in termini rigorosi, vogliamo definire una funzione P che ha come dominio la famiglia degli eventi $A \subset \Omega$ e come codominio i numeri reali.

Ovviamente i valori specifici che assume P dipenderanno dagli esempi. Comunque richiediamo alcune proprietà generali rispetto alle operazioni insiemistiche:

1. $P(\Omega) = 1$, $P(\emptyset) = 0$, $P(A) \in [0, 1]$ per ogni $A \subset \Omega$
2. se A_1, A_2, \dots sono eventi disgiunti (ovvero $A_i \cap A_j = \emptyset$ per ogni $i \neq j$), allora

$$P\left(\bigcup A_i\right) = \sum P(A_i).$$

Exercise 1. *Mostrare che, detto A^c il complementare di A , vale*

$$P(A^c) = 1 - P(A).$$

Exercise 2. *Mostrare che $A \subset B$ implica $P(A) \leq P(B)$.*

1.3. La probabilità condizionale

Molto spesso interessa calcolare la probabilità di un evento A essendo in possesso di certe informazioni aggiuntive, riassunte dall'evento B . Ad esempio ci possiamo chiedere la probabilità che da un lancio di dado sia uscito 6, nel momento in cui ci venga data l'informazione aggiuntiva che il numero uscito è pari.

In presenza dell'informazione B , l'universo Ω si restringe a B , ovvero B diventa il nuovo universo. Ad esempio, nel lancio di un dado, senza informazioni aggiuntive prendiamo $\Omega = \{1, \dots, 6\}$, ma se sappiamo che è uscito un numero pari, ovvero sappiamo che si è verificato l'evento $B = \{2, 4, 6\}$, ora il nostro universo è proprio rappresentato dall'insieme $\{2, 4, 6\}$ (altri risultati non sono possibili).

In presenza della nuova informazione B , le varie probabilità cambiano rispetto alla situazione iniziale. La probabilità che esca 6 è $1/6$, ma se sappiamo che è uscito un numero pari essa diventa $1/3$. Dobbiamo quindi decidere una regola per calcolare la probabilità di un evento A quando sappiamo che si è verificato un evento B .

Avendo osservato sopra che il nostro universo è diventato B , la prima idea naturale è di prendere come nuova probabilità di A il numero $P(A \cap B)$, perché

$A \cap B$ è la parte di evento A che sopravvive nel restringere l'universo da Ω a B . L'unico difetto ovvio della grandezza $P(A \cap B)$ è che la probabilità dell'universo non è uno: $P(\Omega \cap B) = P(B)$, non necessariamente uguale ad 1. Basta allora dividere per $P(B)$, in modo da "rinormalizzare" tutto ad uno. Ecco la definizione rigorosa.

Definition 1. Supponiamo $P(B) > 0$. Chiamiamo *probabilità condizionale di A sapendo B* il numero

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

Si può verificare che, fissato B con $P(B) > 0$, la funzione $A \mapsto P(A|B)$ gode delle proprietà generali di una probabilità. Spesso la probabilità condizionale si usa per riscrivere l'intersezione tra due eventi:

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

A volte poi si usa per riscrivere la probabilità di un singolo evento A , dopo averlo decomposto in parti disgiunte: se B_1, B_2, \dots sono eventi disgiunti, la cui unione è tutto Ω , vale allora

$$P(A) = \sum P(A \cap B_i)$$

e quindi, usando la probabilità condizionale,

$$P(A) = \sum P(A|B_i)P(B_i).$$

Questa formula porta il nome di *formula delle probabilità totali o di ripartizione* (o di *fattorizzazione*, o di *disintegrazione*).

Exercise 3. Verificare la prima di queste uguaglianze.

Infine, come vale $P(A \cap B) = P(A|B) \cdot P(B)$, vale anche $P(A \cap B) = P(B|A) \cdot P(A)$. Quindi:

Proposition 1. Se $P(A) > 0$, allora

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

Questa è la famosa *formula di Bayes*. Si usa in problemi in cui si conoscono certe probabilità condizionali $P(A|B)$ e si vogliono calcolare quelle inverse $P(B|A)$. In quest'ottica, spesso la si riscrive sostituendo al denominatore $P(A)$ l'espressione $\sum P(A|B_i)P(B_i)$, dove B_1, B_2, \dots sono eventi disgiunti.

Exercise 4. *Un produttore produce pezzi funzionanti al 95%. Un sistema di controllo li esamina e ferma il 90% di quelli non funzionanti, mentre lascia passare tutti quelli funzionanti. Qual'è la probabilità che il sistema lasci passare un pezzo?*

Exercise 5. *Uno strumento elettronico C contiene varie componenti di cui una fondamentale A . Se A funziona sotto un certo limite di tolleranza, C ha probabilità 0.1 di funzionare male, altrimenti ha probabilità 0.05 di funzionare male. D'altra parte, la probabilità che A funzioni sotto il limite di tolleranza 0.01. Calcolare la probabilità che C funzioni male. Osservato un mal funzionamento, calcolare la probabilità che esso sia dovuto ad A .*

1.3.1. Indipendenza tra eventi

La probabilità condizionale e l'indipendenza sono i due tratti più caratteristici del calcolo delle probabilità.

Definition 2. *Due eventi A e B sono indipendenti se $P(A \cap B) = P(A) \cdot P(B)$.*

La motivazione per questa definizione sta nel seguente fatto. Avendo già a disposizione il concetto di probabilità condizionale, volendo introdurre il concetto di indipendenza tra due eventi A e B , è naturale dire che A e B sono indipendenti se la conoscenza dell'uno non influisce sulla probabilità dell'altro, ovvero se $P(A|B) = P(A)$ e $P(B|A) = P(B)$. Prendiamo ad esempio la prima, $P(A|B) = P(A)$. Per definizione essa diventa $\frac{P(A \cap B)}{P(B)} = P(A)$, ovvero $P(A \cap B) = P(A) \cdot P(B)$. Lo stesso succede partendo da $P(B|A) = P(B)$. Il ragionamento si può anche invertire (a patto di poter dividere per $P(A)$ o $P(B)$). Quindi la definizione rigorosa riflette l'idea intuitiva, col pregio di essere simmetrica rispetto ad A e B .

Exercise 6. *Consideriamo un sistema composto da due componenti c_1 e c_2 in parallelo. Supponiamo che c_1 funzioni con probabilità p_1 e c_2 funzioni con probabilità p_2 . Calcolare la probabilità che il sistema non funzioni. Supponiamo tacitamente in questo e negli esercizi successivi che l'eventuale mal funzionamento di una componente sia indipendente dall'altra.*

Exercise 7. *Due strade portano a destinazione, ma ciascuna può essere interrotta con probabilità 0.01. Calcolare la probabilità di poter raggiungere la destinazione.*

Presentiamo due soluzioni di questo problema. La prima è più veloce ma richiede l'idea di passare al complementare. La seconda si basa invece su una semplice enumerazione dei casi favorevoli, sommandone poi le probabilità.

Soluzione 1. Detto I_1 l'evento "la strada numero 1 è interrotta", detto I_2 l'analogo evento per la strada numero 2, vale per ipotesi $P(I_1) = P(I_2) = 0.01$. Detto D l'evento "è possibile raggiungere la destinazione", ovvero l'evento "almeno una delle due strade è percorribile", allora l'evento complementare D^c è "entrambe le strade sono interrotte", ovvero

$$D^c = I_1 \cap I_2.$$

Per l'indipendenza vale quindi

$$\begin{aligned} P(D^c) &= 1 - P(D) = 1 - P(I_1) \cdot P(I_2) \\ &= 1 - (0.01)^2 = 0.9999. \end{aligned}$$

Soluzione 2. Usando le stesse notazioni, osserviamo che D vale in ciascuno dei seguenti casi, mutualmente esclusivi ed esaustivi: la prima strada è percorribile mentre la seconda no, la seconda è percorribile mentre la prima no, entrambe sono percorribili. In altre parole

$$D = (I_1^c \cap I_2) \cup (I_1 \cap I_2^c) \cup (I_1 \cap I_2).$$

Essendo i tre casi mutualmente esclusivi, ovvero disgiunti, vale

$$P(D) = P(I_1^c \cap I_2) + P(I_1 \cap I_2^c) + P(I_1 \cap I_2).$$

Usando poi l'indipendenza, abbiamo

$$\begin{aligned} P(D) &= P(I_1^c) \cdot P(I_2) + \dots \\ &= 0.99 \cdot 0.01 + 0.99 \cdot 0.01 + 0.99 \cdot 0.99 = 0.9999. \end{aligned}$$

Exercise 8. *In un sistema composto da due componenti in serie, ciascuna può essere rotta con probabilità 0.1. Calcolare la probabilità che il sistema funzioni.*

Exercise 9. *Se un termostato in un impianto funziona correttamente con probabilità 0.95, qual'è la probabilità che un sistema composto da 2 termostati in serie non funzioni correttamente?*

Exercise 10. Ad un ufficio si possono rivolgere due categorie di persone, C_1 e C_2 . Ciascuna categoria C_i può richiedere o il servizio di tipo A o quello di tipo B. Da rilevamenti eseguiti in un certo periodo passato risulta che l'80% delle persone di categoria C_1 chiede il servizio B, mentre il 70% di quelle di categoria C_2 chiede il servizio A. Inoltre, il 90% delle persone che si presenta all'ufficio fa parte della categoria C_1 .

i) Calcolare la probabilità che l'ufficio riceva una richiesta di servizio di tipo A.

ii) Se si riceve una richiesta di servizio di tipo A, siamo portati a pensare che provenga da una persona di categoria C_1 o C_2 ?

iii) Si considerino le prime due persone che si rivolgono all'ufficio (supposte indipendenti). Calcolare la probabilità che presentino richieste diverse.

Soluzione. i) Indicando per semplicità gli eventi con le lettere corrispondenti a quelle del testo, abbiamo

$$\begin{aligned} P(A) &= P(A|C_1)P(C_1) + P(A|C_2)P(C_2) \\ &= 0.2 \cdot 0.9 + 0.7 \cdot 0.1 = 0.25. \end{aligned}$$

ii) Vale poi

$$P(C_1|A) = \frac{P(A|C_1)P(C_1)}{P(A)} = \frac{0.2 \cdot 0.9}{0.25} = 0.72.$$

Quindi siamo portati a credere che la richiesta arrivi da C_1 (nonostante la maggior parte delle persone di categoria C_1 chieda B!).

iii) Detto D l'evento "le richieste sono diverse", indicando con A_1 l'evento "la prima persona chiede A", con B_2 l'evento "la seconda persona chiede B", e così via, vale

$$D = (A_1 \cap B_2) \cup (A_2 \cap B_1).$$

Gli eventi $(A_1 \cap B_2)$ e $(A_2 \cap B_1)$ sono disgiunti. Gli eventi A_1 e B_2 sono indipendenti, e così gli eventi A_2 e B_1 . Quindi vale

$$\begin{aligned} P(D) &= P(A_1 \cap B_2) + P(A_2 \cap B_1) = P(A_1)P(B_2) + P(A_2)P(B_1) \\ &= 0.25 \cdot 0.75 + 0.25 \cdot 0.75 = 0.375. \end{aligned}$$

1.4. Le variabili aleatorie

In termini intuitivi, una variabile aleatoria è una grandezza che assume valori casuali, secondo certe probabilità. Ad esempio, il tempo d'attesa prima che un

servizio sia espletato è una grandezza aleatoria. In termini rigorosi, una variabile aleatoria è una funzione da un universo Ω in \mathbb{R} . Ricorreremo però a questo formalismo rigoroso solo di tanto in tanto, per cui rimandiamo una comprensione più profonda di questa definizione. Per lo più sarà per noi sufficiente usare le variabili aleatorie nella loro accezione intuitiva, a patto di seguire certe regole di calcolo che ora esporremo. A livello di notazioni, cercheremo sempre di usare lettere maiuscole per indicare variabili aleatorie, es. X, T , ecc. Come abbiamo detto, una v.a. X assume certi valori con certe probabilità. Quindi la cosa che dobbiamo specificare di una v.a. è la probabilità che assuma questo o quel valore, o che assuma valori in questo o quell'intervallo. Ad esempio, interesserà conoscere la probabilità che il tempo di servizio di un sistema sia inferiore ad una certa soglia. In termini generali, per un generico insieme A , dovremo specificare la *probabilità che X assuma valori in A* :

$$P(X \in A).$$

Qui A è un sottoinsieme di \mathbb{R} , che per evitare certi problemi tecnici un po' fastidiosi supporremo essere un intervallo (anche una semiretta o tutta la retta), aperto o chiuso o semiaperto, o l'unione finita di intervalli. Il simbolo $P(X \in A)$, nell'accezione rigorosa del concetto di v.a., non è un nuovo simbolo, ma si riconduce al concetto di probabilità P di un evento spiegato sopra. Infatti, relativamente ad una v.a. X , dovrebbe essere specificato un insieme Ω su cui è definita, per cui $(X \in A)$ è la scrittura abbreviata del sottoinsieme di Ω dato da $\{\omega \in \Omega : X(\omega) \in A\}$. Su tali sottoinsiemi dovrebbe essere definita una probabilità P , e quindi dovrebbe aver senso calcolare $P(\omega \in \Omega : X(\omega) \in A)$, ovvero con simbologia abbreviata $P(X \in A)$. Questa visione rigorosa permette di capire le cose anche meglio, ma per brevità possiamo accontentarci di usare il concetto di v.a. X ed il simbolo $P(X \in A)$ in modo un po' più intuitivo, come abbiamo fatto sopra. Dobbiamo a questo punto dare delle prescrizioni per descrivere quantitativamente le probabilità $P(X \in A)$. Evitando di studiare la situazione più generale possibile, consideriamo per ora due casi particolari.

1. Il primo è quello in cui la v.a. X assume solo un numero finito di valori (o un po' più generale, come ora diremo). Parleremo di v.a. *discreta*. Indichiamo con x_1, \dots, x_n i possibili valori di X . Ad esempio, X può essere il numero di utenti che si rivolgono ad un servizio, tra una popolazione di 1000 possibili utenti; in questo esempio i valori possibili x_1, \dots, x_n sono i numeri $0, 1, \dots, 1000$. Per una tale v.a. X , basta assegnare le probabilità

$$p_i = P(X = x_i)$$

per ogni $i = 1, \dots, n$. Infatti da esse si calcolano tutte le altre:

$$P(X \in A) = \sum_{i: x_i \in A} p_i.$$

La successione di numeri $\{p_i\}$ può essere chiamata *densità discreta* della v.a. X , oppure *massa di probabilità* di X , o *distribuzione* di X . Essa verifica

$$p_i \in [0, 1], \quad \sum p_i = 1.$$

Quanto detto fino ad ora si generalizza al caso in cui i valori possibili di X , invece che essere in numero finito, siano una quantità numerabile senza punti di accumulazione, ad esempio i numeri naturali o i numeri interi. Certe sommatorie saranno in tal caso delle serie infinite.

2. Il secondo è quello in cui c'è una funzione integrabile $f(x)$ con le proprietà

$$f(x) \geq 0, \quad \int_{-\infty}^{+\infty} f(x) dx = 1,$$

tale che

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

(qui si vede ad esempio perché abbiamo preferito restringere l'attenzione ad insiemi A che siano intervalli o un po' più generali, ma non insiemi qualsiasi). La funzione $f(x)$ si chiama *densità di probabilità* di X . Le v.a. di questo tipo si dicono *continue* (più precisamente, assolutamente continue).

1.5. La funzione di distribuzione

Data una v.a. X , si chiama *funzione di distribuzione* o *di ripartizione* di X la funzione

$$F(x) = P(X \leq x).$$

Osserviamo una sola proprietà molto importante: se X ha densità $f(x)$, vale

$$\frac{dF(x)}{dx} = f(x)$$

in ogni punto x in cui f è continua. Quindi, in molti esempi, conoscere F equivale a conoscere f .

1.6. I Valori medi

Per una v.a. discreta X con densità discreta $\{p_i\}$ e valori possibili $\{x_i\}$, si chiama *valor medio* il numero

$$E[X] = \sum x_i p_i$$

quando questa somma converge (se è finita converge sempre, ovviamente). Per una v.a. continua X con densità $f(x)$, si chiama *valor medio* il numero

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

quando questo integrale (improprio) converge. Più generalmente, data una funzione $g: \mathbb{R} \rightarrow \mathbb{R}$, poniamo

$$E[g(X)] = \sum g(x_i) p_i$$

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

rispettivamente nel caso discreto e continuo, quando queste espressioni convergono (a livello rigoroso, queste in realtà non sono nuove definizioni ma formule che si dimostrano, sotto opportune ipotesi).

Il valor medio è *lineare*:

$$E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$$

per ogni $\alpha, \beta \in \mathbb{R}$ e per ogni coppia di v.a. X ed Y . Se le v.a. X ed Y sono *indipendenti*, il valor medio del loro prodotto è uguale al prodotto dei valori medi

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

Se prendiamo la funzione $g(x) = x^n$, abbiamo i numeri

$$m_n = E[X^n]$$

detti *momenti di ordine n*. Indichiamo con μ il momento di ordine uno, ovvero il valor medio. Se prendiamo la funzione $g(x) = (x - \mu)^2$ abbiamo il numero

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

detto *varianza* di X . Usando la linearità si verifica che vale la formula

$$\text{Var}[X] = E[X^2] - \mu^2.$$

Quindi conoscendo i momenti di ordine uno e due si può calcolare la varianza. La radice quadrata della varianza si chiama *deviazione standard*, e si indica spesso con σ_X :

$$\sigma_X = \sqrt{\text{Var}(X)}.$$

Si introduce anche il *momento centrato di ordine n* , $\mu_n = E[(X - \mu)^n]$, che generalizza la varianza, e da esso l'indice di Kurtosi $\frac{\mu_4}{\mu_2^2} - 3$ e l'indice di skewness (asimmetria) $\frac{\mu_3}{\mu_2^{3/2}}$ per le cui interpretazioni rimandiamo ai testi di statistica.

Date poi due v.a. X ed Y su uno spazio probabilizzato (Ω, \mathcal{F}, P) , si introduce la *correlazione* tra X ed Y definita da

$$R_{XY} = E[XY]$$

e la *covarianza* tra X ed Y definita da

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

dove μ_X ed μ_Y sono le due medie. Per la covarianza vale una decomposizione simile a quella della varianza (di facile dimostrazione):

$$\text{Cov}(X, Y) = R_{XY} - \mu_X \mu_Y.$$

Da questo si vede subito che se X ed Y sono indipendenti, allora

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

Vale in generale la formula

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

(di facile verifica), da cui si deduce che se X ed Y sono *indipendenti*, allora vale

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

Per motivi di invarianza di scala si introduce anche una normalizzazione della covarianza, detto coefficiente di correlazione,

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Si dimostra che $-1 \leq \rho_{XY} \leq 1$.

Ricordiamo infine una regola piuttosto utile. Se X ed Y sono *indipendenti* ed hanno densità $f_X(x)$ ed $f_Y(y)$, allora la v.a. $Z = X + Y$ ha densità $f_Z(z)$ data dalla *convoluzione* delle due densità $f_X(x)$ ed $f_Y(y)$:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y) dy.$$

1.7. Variabili di Bernoulli e binomiali

Una v.a. discreta X che assume solo i valori 0 ed 1, con $P(X = 1) = p$ e $P(X = 0) = q = 1-p$, $p \in [0, 1]$, si dice di *Bernoulli di parametro p* . Nonostante la sua semplicità essa costituisce un mattone di costruzione fondamentale. Vediamo un esempio.

Example 2. *Una banca ha 1000 conti correnti aperti. Attribuisce i numeri da 1 a 1000 ai suoi correntisti. Si chiede poi qual'è la probabilità che nell'arco di una giornata si presentino più di 300 correntisti. Iniziamo la risoluzione di questo problema, che completeremo nel prossimo sottoparagrafo. La banca associa ad ogni correntista una v.a. di Bernoulli, X_1 per il primo, e così via fino ad X_{1000} per l'ultimo. La v.a. X_1 vale 1 se il correntista si presenta in banca durante il giorno in questione, 0 altrimenti. Supponendo per semplicità che i correntisti abbiano proprietà simili, indichiamo con p la probabilità che un correntista si presenti in banca in una certa giornata. Quindi $p = P(X_k = 1)$ per ogni correntista k . Finalmente, la nuova v.a. definita da $S = X_1 + \dots + X_{1000}$ rappresenta il numero di correntisti che si presentano in banca (infatti i vari addendi valgono 1 per ogni correntista che si presenta, zero altrimenti). Pertanto S descrive ciò che interessa alla banca. Dobbiamo quindi calcolare*

$$P(S > 300).$$

Dato un numero intero n ed un numero $p \in [0, 1]$, chiamiamo v.a. *binomiale di parametri n e p* (abbreviato $B(n, p)$) una v.a. X che assume i valori $0, 1, \dots, n$ con probabilità

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

per $k = 0, 1, \dots, n$.

Theorem 2. Se X_1, \dots, X_n sono v.a. di Bernoulli di parametro p , indipendenti, allora $S = X_1 + \dots + X_n$ è una v.a. $B(n, p)$.

Non diamo la dimostrazione, comunque non molto difficile. Riprendiamo invece l'esempio precedente.

Example 3. Supponendo indipendenti i correntisti, il numero S di correntisti che si presentano in banca è una $B(1000, p)$. Vale allora

$$P(S > 300) = \sum_{k=301}^{1000} \binom{1000}{k} p^k (1-p)^{1000-k}$$

$$1 - \sum_{k=0}^{300} \binom{1000}{k} p^k (1-p)^{1000-k}.$$

Ad esempio, se un generico correntista si reca in banca mediamente una volta alla settimana, per cui $p = \frac{1}{5}$ (cinque giorni di apertura), risulta

$$P(S > 100) = 2.2017 \times 10^{-14}.$$

Questo numero è piccolissimo. Si noti che, ragionando un po' intuitivamente, se ogni correntista si presenta mediamente una volta alla settimana, avremo mediamente 200 correntisti al giorno; ma chiaramente ci aspettiamo fluttuazioni aleatorie, per cui in teoria potremmo pensare che ogni tanto capitino un giorno con 300 o 400 presenze. Invece la probabilità di avere più di 300 presenze è 2.2017×10^{-14} , quindi un tale evento eccezionale capita mediamente due volte ogni 10^{14} giorni (ovvero mai). Un simile fatto, senza il calcolo delle probabilità, non era facilmente immaginabile. Esso ha conseguenze sulla necessità di tenere determinate somme disponibili giornalmente, di avere un certo numero di sportelli aperti e così via.

1.8. Variabili aleatorie di Poisson

Dato un numero $\lambda > 0$, si chiama v.a. di Poisson di parametro λ (abbreviato $\mathcal{P}(\lambda)$) una v.a. X che può assumere solo tutti i valori interi non negativi $0, 1, \dots$, con probabilità

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

per $k = 0, 1, \dots$

Si dimostra che il valore medio e la varianza sono entrambi pari a λ . Si può inoltre dimostrare che la somma di una $\mathcal{P}(\lambda)$ e di una $\mathcal{P}(\mu)$ *indipendenti* è una $\mathcal{P}(\lambda + \mu)$.

Theorem 3. Se $X_{n,p}$ è una v.a. $B(n, p)$ ed X_λ è una v.a. $\mathcal{P}(\lambda)$, e se $n \rightarrow \infty$, $p \rightarrow 0$, con $n \cdot p \rightarrow \lambda$, allora

$$P(X_{n,p} = k) \rightarrow P(X_\lambda = k)$$

ovvero

$$\binom{n}{k} p^k (1-p)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}$$

per ogni $k \geq 0$. In altre parole, la distribuzione $B(n, p)$ converge (sotto le suddette ipotesi) alla distribuzione $\mathcal{P}(\lambda)$.

Non eseguiamo la dimostrazione, non difficile. Questo teorema porta il nome di *teorema degli eventi rari*, nel senso che la probabilità di successo $p \rightarrow 0$, ovvero è sempre più raro che si verifichi un successo (ma questo è compensato dal fatto che il numero di prove n tende all'infinito). Grazie anche a questo teorema, la Poisson viene usata in molti problemi di conteggio, ad esempio nel conteggio del numero di guasti in un sistema con moltissimi elementi aventi ciascuno una bassa probabilità di guastarsi, o nel conteggio del numero di richieste che arrivano ad un sistema di servizio.

1.9. Variabili aleatorie esponenziali

Dato un numero $\lambda > 0$ chiamiamo esponenziale di parametro λ (abbreviato $\exp(\lambda)$) una v.a. T avente densità di probabilità

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{per } t \geq 0 \\ 0 & \text{per } t < 0 \end{cases}.$$

Exercise 11. Dimostrare che la funzione di distribuzione è, per $t \geq 0$,

$$F(t) = 1 - e^{-\lambda t}.$$

Altrettanto facilmente si dimostra che il valore medio e la varianza sono pari rispettivamente a $\frac{1}{\lambda}$ e $\frac{1}{\lambda^2}$.

Semplicissima ed utile è la formula

$$P(T > t) = e^{-\lambda t}$$

valida per una v.a. esponenziale T di parametro λ .

Le v.a. esponenziali vengono spesso utilizzate per modellare le grandezze “tempo di interarrivo tra due eventi” (chiamate telefoniche, arrivi ad una coda, ecc.) e “tempo di servizio di un utente”.

Example 4. *Un aditta di manutenzioni osserva con l’esperienza che tra una richiesta e la successiva intercorre mediamente un tempo di 6 ore. Volendo descrivere questo tempo aleatorio con una distribuzione esponenziale, prende quella di parametro $\lambda = \frac{1}{6}$ (ore⁻¹).*

1.9.1. Proprietà di assenza di memoria della legge esponenziale

Una proprietà importante della legge esponenziale è rappresentata dalla cosiddetta *assenza di memoria*. intuitivamente, se ad esempio usiamo la legge esponenziale per modellare il tempo di vita di un oggetto, la proprietà di assenza di memoria si manifesta quando qualunque sia il tempo trascorso, il tempo residuo di vita non è affetto dal passato e ha la stessa distribuzione del tempo di vita originario. In altre parole, l’oggetto non subisce logoramento, per cui la sua propensione statistica a rompersi resta invariata. Ovviamente da questo si vede che l’ipotesi di esponenzialità è piuttosto ideale nella pratica, ma la sua comodità matematica fa sì che la si supponga in molti contesti.

Example 5. *Attraverso Internet richiediamo un servizio, che può essere espletato solo quando il servente è libero. Supponiamo che la nostra richiesta non venga messa in una coda, ma che venga reiterata ogni secondo. Quando il servente si libera, prende la prima richiesta che gli arriva; se la nostra reiterazione gli arriva un istant dopo, viene scartata, e si continua ad aspettare. In questa situazione, anche se abbiamo aspettato inutilmente per 10 minuti, le nostre chances di essere accettati dall’operatore non sono aumentate: non c’è traccia nella memoria dell’operatore della nostra attesa più o meno lunga. In questo caso il tempo di attesa di connessione al servizio è molto plausibilmente esponenziale.*

Theorem 4. *Se T è esponenziale di parametro λ allora, per ogni $t, s \geq 0$, vale*

$$P(T > t + s | T > t) = P(T > s).$$

In altre parole, arrivati al tempo t ed osservato che siamo ancora in attesa ($T > t$), la probabilità che l'evento accada dopo un tempo s è uguale alla probabilità che inizialmente l'evento accadesse dopo un tempo s .

Dimostrazione. Vale

$$P(T > t + s | T > t) = \frac{P(T > t + s, T > t)}{P(T > t)}$$

ma l'evento $T > t + s, T > t$ equivale all'evento $T > t + s$, quindi

$$= \frac{P(T > t + s)}{P(T > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P(T > s).$$

La dimostrazione è completa.

1.9.2. Variabili aleatorie di Erlang

Capita spesso di considerare la somma di n v.a. esponenziali indipendenti con lo stesso parametro λ : ad esempio è il tempo che trascorre prima dell'arrivo dell' n -esima richiesta di servizio, se le richieste arrivano con intertempi esponenziali indipendenti ed ugualmente distribuiti.

Date T_1, \dots, T_n v.a. indipendenti con distribuzione esponenziale di parametro $\lambda > 0$, diciamo che $S_n = T_1 + \dots + T_n$ è una *v.a. di Erlang* di numerosità n e parametro λ .

Lemma 5. *La sua densità g_n e la sua funzione di distribuzione G_n sono date da*

$$g_n(x) = \lambda \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x} \quad \text{per } x > 0$$

$$G_n(x) = 1 - e^{-\lambda x} \left(1 + \frac{\lambda x}{1!} + \dots + \frac{(\lambda x)^{n-1}}{(n-1)!} \right) \quad \text{per } x > 0.$$

Dimostrazione. Dimostriamo il lemma per induzione. Per $n = 1$ la g_1 è la densità esponenziale, quindi l'affermazione è vera. Supponiamo che l'affermazione del lemma sia vera per n , dimostriamola per $n + 1$. Consideriamo la somma

$$S_{n+1} = T_1 + \dots + T_n + T_{n+1} = S_n + T_{n+1},$$

dove $S_n = T_1 + \dots + T_n$ ha $g_n(x) = \lambda \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x}$ come densità. Abbiamo ricordato sopra che la densità della somma di due v.a. indipendenti è la convoluzione delle densità. Quindi

$$g_{n+1}(x) = \int_{-\infty}^{+\infty} g_n(x-y) f_{T_{n+1}}(y) dy$$

dove abbiamo indicato con $f_{T_{n+1}}(y)$ la densità di T_{n+1} . Per $x > 0$ vale allora (si deve prestare un attimo di attenzione agli estremi di integrazione, motivati dal fatto che le due densità integrande sono nulle per i loro argomenti negativi)

$$\begin{aligned} g_{n+1}(x) &= \int_0^x \lambda \frac{(\lambda(x-y))^{n-1}}{(n-1)!} e^{-\lambda(x-y)} \lambda e^{-\lambda y} dy \\ &= \frac{\lambda^2 e^{-\lambda x}}{(n-1)!} \lambda^{n-1} \int_0^x (x-y)^{n-1} dy = \frac{\lambda^2 e^{-\lambda x}}{(n-1)!} \lambda^{n-1} \int_0^x t^{n-1} dt \\ &= \frac{\lambda^2 e^{-\lambda x}}{(n-1)!} \lambda^{n-1} \frac{x^n}{n} = \lambda \frac{(\lambda x)^n}{n!} e^{-\lambda x}. \end{aligned}$$

La dimostrazione per induzione che g_n è la densità è completa. Per dimostrare che G_n è la funzione di distribuzione si può eseguire l'integrale di g_n , o conoscendo già l'espressione data sopra per G_n basta far vedere che la sua derivata è g_n (derivando, i vari termini si cancellano a due a due, escluso uno) e che $G_n(0) = 0$. La dimostrazione è completa.

Remark 1. Il valore medio e la varianza di S_n sono pari rispettivamente a $\frac{n}{\lambda}$ e $\frac{n}{\lambda^2}$ (segue subito dalla definizione e dalle proprietà del valor medio).

1.9.3. Legame tra v.a. esponenziali e v.a. di Poisson

Supponiamo che, sull'asse dei tempi $[0, \infty)$ accadano degli eventi ad istanti aleatori successivi (es. gli arrivi di chiamate telefoniche ad una centrale). Indichiamo con T_1, \dots, T_n gli interarrivi tra un evento e l'altro (T_1 è l'istante in cui accade il primo evento, $T_1 + T_2$ l'istante del secondo, mentre T_2 da solo è il tempo che intercorre tra il primo evento ed il secondo, e così via). Fissato un tempo $\tau > 0$, ci chiediamo quanti eventi sono accaduti entro τ , ovvero nell'intervallo di tempo $[0, \tau]$. Indichiamo con $N(\tau)$ questo numero, aleatorio (ad esempio il numero di chiamate telefoniche arrivate entro il tempo τ).

Theorem 6. Se le T_i sono esponenziali $\exp(\lambda)$ indipendenti, allora $N(\tau)$ è una v.a. di Poisson di parametro $\lambda\tau$.

Dimostrazione. Ricordiamo che $S_n = T_1 + \dots + T_n$ è una v.a. di Erlang. $N(\tau)$ e la famiglia $(S_n)_{n \geq 0}$ sono legate da questa relazione logica:

$$N(\tau) \leq k \Leftrightarrow S_{k+1} > \tau.$$

Questa darà la chiave di tutto. Dimostriamola mostrando la validità delle due implicazioni separatamente. Se $N(\tau) \leq k$, ovvero se entro τ arrivano al più k chiamate, allora non ne possono arrivare $k+1$, quindi la $k+1$ -esima chiamata arriva dopo il tempo τ , quindi $S_{k+1} > \tau$. Viceversa, se $S_{k+1} > \tau$, cioè se la $k+1$ -esima chiamata arriva dopo il tempo τ , allora entro il tempo τ sono arrivate meno di $k+1$ chiamate, quindi al più k , quindi $N(\tau) \leq k$.

Allora

$$\begin{aligned} P(N(\tau) \leq k) &= P(S_{k+1} > \tau) = 1 - G_{k+1}(\tau) \\ &= e^{-\lambda\tau} \left(1 + \frac{\lambda\tau}{1!} + \dots + \frac{(\lambda\tau)^k}{k!} \right). \end{aligned}$$

Abbiamo così trovato la funzione di distribuzione di $N(\tau)$. Vale quindi

$$P(N(\tau) = k) = P(N(\tau) \leq k) - P(N(\tau) \leq k-1) = e^{-\lambda\tau} \frac{(\lambda\tau)^k}{k!}$$

come volevasi dimostrare.

Exercise 12. Durante le 6 ore di apertura di un negozio di bici, l'intertempo di vendita, cioè il tempo tra due vendite successive, si può assumere distribuito secondo una legge esponenziale con media pari a 2 ore. Gli intertempi tra le vendite successive della singola giornata possono considerarsi v.a. indipendenti e, inoltre, la vendita durante ogni giornata lavorativa può assumersi indipendente da quella delle altre giornate.

1. Calcolare la probabilità che nella singola giornata non si abbia nessuna vendita;
2. Valutare la probabilità che in un giorno si effettuino almeno 3 vendite;
3. Calcolare il valore medio e la varianza delle vv.aa. X ed Y , indicanti rispettivamente il numero di bici vendute in un giorno ed il numero di bici vendute in 300 giorni lavorativi.

1.9.4. Sul minimo di v.a. esponenziali

Date T_1 e T_2 v.a. esponenziali indipendenti, di parametri λ_1 e λ_2 , consideriamo la v.a. $T = \min(T_1, T_2)$. Ad esempio, se attendiamo che si liberi uno sportello ed abbiamo davanti a noi due possibili sportelli, entrambi occupati, ciascuno che si libererà dopo un tempo esponenziale, T indica l'istante in cui si libererà il primo dei due.

La v.a. T ha densità esponenziale di parametro $\lambda_1 + \lambda_2$.

Per dimostrarlo calcoliamo il complementare della funzione di distribuzione di T :

$$\begin{aligned} P(T > t) &= P(\min(T_1, T_2) > t) = P(T_1 > t, T_2 > t) \\ &= P(T_1 > t)P(T_2 > t) = e^{-\lambda_1 t} e^{-\lambda_2 t} = e^{-(\lambda_1 + \lambda_2)t}. \end{aligned}$$

Questo dimostra quanto volevamo. In generale vale:

Proposition 7. Se T_1, \dots, T_n sono v.a. esponenziali indipendenti, di parametri $\lambda_1, \dots, \lambda_n$, allora la v.a. $T = \min(T_1, \dots, T_n)$ è esponenziale di parametro $\lambda_1 + \dots + \lambda_n$.

1.10. Il processo di Poisson

Partendo da v.a. esponenziali T_1, \dots, T_n indipendenti, di parametro λ , abbiamo introdotto sopra, per ogni $\tau > 0$ la v.a. $N(\tau)$, Poisson di parametro $\lambda\tau$. Poniamo inoltre $N(0) = 0$. Se, fissato λ , facciamo variare τ , la famiglia $(N(\tau))_{\tau \geq 0}$ è un processo stocastico che porta il nome di *Processo di Poisson*.

Si immagini cosa si vede sperimentalmente quando si osserva un processo di Poisson: al tempo 0 vale $N(0) = 0$, e tale resta fino all'istante T_1 in cui accade il primo evento; in quell'istante il processo salta al valore 1: $N(T_1) = 1$. Poi si aspetta fino all'istante $T_1 + T_2$, in cui $N(\tau)$ salta al valore 2, e così via. Tracciando il grafico di $N(\tau)$ in un piano cartesiano, si vede una curva costante a tratti, col primo tratto di ordinata zero, poi quello di ordinata uno e così via.

Se $N(\tau)$ indica il numero di eventi che accadono in $[0, \tau]$, allora $N(t) - N(s)$ indica, per $t > s \geq 0$, il numero di eventi che avvengono nell'intervallo $(s, t]$.

Valgono, per gli *incrementi* del processo di Poisson ora introdotti, due proprietà molto importanti:

- 1) $N(t) - N(s)$ ha distribuzione di Poisson di parametro $\lambda(t - s)$
- 2) presi $t_n > t_{n-1} > \dots > t_1 \geq 0$, gli incrementi $N(t_n) - N(t_{n-1}), \dots, N(t_2) - N(t_1)$ sono indipendenti.

In particolare, la seconda proprietà dice che $N(t) - N(s)$ è indipendente da $N(s) - N(0)$, ovvero da $N(s)$. Quindi, arrivati al tempo s , il numero di arrivi in un successivo intervallo $(s, t]$ è indipendente dal numero di arrivi osservati fino al tempo t .

Omettiamo la dimostrazione di 1 e 2. Un'altra proprietà che enunciamo senza dimostrazione è la seguente: se si parte da un processo che soddisfa 1 e 2, allora si possono definire gli istanti in cui il processo cresce di un'unità e dimostrare che gli intertempi tra tali istanti sono v.a. esponenziali indipendenti di parametro λ .

Riprenderemo più avanti il processo di Poisson da altri punti di vista.

1.11. Esercizi

Exercise 13. Calcolare il valor medio di una v.a. $\mathcal{P}(\lambda)$.

Exercise 14. Calcolare la varianza di una v.a. $\mathcal{P}(\lambda)$, dopo aver calcolato $E[X(X-1)]$.

Exercise 15. Con la convoluzione di densità, verificare la formula di Erlang per $n = 2, 3$.

Exercise 16. Calcolare media e varianza di una $\text{esp}(\lambda)$. Interpretare graficamente la dipendenza da λ di queste grandezze.

Exercise 17. Calcolare media e varianza di $T_1 + \dots + T_n$.

Exercise 18. Ricordiamo un fatto di analisi: una funzione continua $h(t)$, definita per $t \geq 0$, con la proprietà $h(t+s) = h(t)h(s)$ per ogni $t, s \geq 0$, è per forza della forma $h(t) = e^{\alpha t}$, per un opportuno $\alpha > 0$. Inoltre $\alpha = h'(0)$. Dopo questo richiamo, risolvere il seguente problema. Per una v.a. T continua a valori positivi, con densità $f(t)$, nulla per $t < 0$, e $f(0) \in (0, \infty)$, chiamiamo assenza di memoria la proprietà

$$P(T > t + s | T > t) = P(T > s).$$

Mostrare che T ha distribuzione esponenziale, con parametro $\lambda = f(0)$. In altre parole, l'assenza di memoria è equivalente all'esponenzialità.

Exercise 19. Con le notazioni del paragrafo sul legame tra esponenziale e Poisson, esprimere l'evento $P(N(\tau) = k)$ mediante la famiglia di Erlang $(S_n)_{n \geq 0}$. Capire quindi come mai nella dimostrazione abbiamo preso $P(N(\tau) \leq k)$ al posto dell'apparentemente più semplice $P(N(\tau) = k)$.

Soluzione di alcuni esercizi.

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

$$\begin{aligned}
 E[T] &= \int_0^{\infty} t \lambda e^{-\lambda t} dt = -[te^{-\lambda t}]_0^{\infty} + \int_0^{\infty} e^{-\lambda t} dt = \frac{1}{\lambda}. \\
 Var[T] &= \int_0^{\infty} t^2 \lambda e^{-\lambda t} dt - \frac{1}{\lambda^2} = -[t^2 e^{-\lambda t}]_0^{\infty} + \int_0^{\infty} 2te^{-\lambda t} dt - \frac{1}{\lambda^2} \\
 &= \frac{2}{\lambda} \int_0^{\infty} t \lambda e^{-\lambda t} dt - \frac{1}{\lambda^2} = \frac{2}{\lambda} \frac{1}{\lambda} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.
 \end{aligned}$$

L'assenza di memoria equivale a

$$\frac{P(T > t + s, T > t)}{P(T > t)} = P(T > s)$$

ovvero a

$$P(T > t + s) = P(T > t) P(T > s).$$

La funzione $h(t) = P(T > t)$ soddisfa allora le ipotesi del teorema di analisi ricordato e quindi vale $h(t) = e^{-\alpha t}$ con

$$\begin{aligned}
 \alpha &= h'(0) = \left. \frac{dP(T > t)}{dt} \right|_{t=0} = \left. \frac{d[1 - P(T \leq t)]}{dt} \right|_{t=0} \\
 &= - \left. \frac{dP(T \leq t)}{dt} \right|_{t=0} = -f(0).
 \end{aligned}$$

1. Catene di Markov

Lo studio delle catene di Markov è piuttosto avvincente e conduce a risultati di utilità pratica, anche perché mette in relazione diverse strutture matematiche:

- i) i grafi,
- ii) le matrici,
- iii) i processi stocastici.

Come preliminare, inquadreremo il discorso parlando più in generale dei processi stocastici a tempo discreto e stati discreti.

1.1. Processi a tempo discreto e stati discreti

Chiamiamo processo stocastico a tempo discreto e stati discreti una successione di v.a.

$$X_0, X_1, \dots, X_n, \dots$$

dove ciascuna X_n è una v.a. discreta a valori in un insieme S , detto *spazio degli stati*.

L'insieme S è finito o al più numerabilmente infinito. Senza perdere in generalità, supponiamo che S sia un sottoinsieme dei numeri interi relativi \mathbb{Z} .

Penseremo all'indice n di X_n come al *tempo*; chiameremo poi *stati* i possibili valori di X_n .

Example 1. *consideriamo la coda che si forma ad uno sportello che offre un servizio. Osserviamo la coda solo ad istanti di tempo discreti, ad esempio ogni minuto. Al minuto n -esimo ci sarà in coda un certo numero di persone, aleatorio, che indichiamo con N_n . Qui la v.a. N_n può assumere tutti i valori interi non negativi. Al variare di n abbiamo quindi un esempio di processo stocastico del tipo descritto sopra.*

Al trascorrere del tempo (discreto), il processo può saltare da uno stato all'altro. Se ad un certo istante n si trova in uno stato i ed all'istante successivo $n+1$ si trova in uno stato $j \neq i$, diremo che c'è stata una *transizione*. Useremo, a volte, questo linguaggio anche quando non è necessariamente $j \neq i$, solo per unità espositiva.

Dato un processo stocastico (X_n) , ci interessa calcolare certe probabilità ad esso associate. Le più semplici sono quelle a tempo fissato, della forma

$$P(X_n = i)$$

che rappresentano la probabilità di osservare al tempo n il sistema (descritto dal processo) nello stato i . Indicheremo questi numeri col simbolo $p_i^{(n)}$ ed useremo il simbolo $p^{(n)}$ per indicare il *vettore* di componenti $p_i^{(n)}$. $p^{(n)}$ è la *distribuzione* della v.a. X_n . Per ogni istante di tempo n vale $\sum_{i \in \mathbb{Z}} p_i^{(n)} = 1$.

Oltre a queste semplici probabilità ci può interessare il calcolo di probabilità più complesse, ad esempio che coinvolgono più istanti temporali contemporaneamente. Ad esempio può essere interessante calcolare la probabilità di trovarsi nello stato j al tempo $n + 1$ sapendo di essere nello stato i al tempo n :

$$P(X_{n+1} = j | X_n = i)$$

detta *probabilità di transizione* da i a j al tempo n . Usando la definizione di probabilità condizionale, questa si riscrive nella forma

$$\frac{P(X_n = i, X_{n+1} = j)}{P(X_n = i)}.$$

Pertanto, per questo calcolo è sufficiente conoscere $p_i^{(n)}$ e le *probabilità congiunte* della forma $P(X_n = i, X_{n+1} = j)$. Si capisce che per calcolare espressioni più complesse serve la conoscenza delle generiche probabilità congiunte della forma

$$P(X_0 = i_0, \dots, X_n = i_n)$$

al variare di tutti gli n e tutti gli $i_0, \dots, i_n \in \mathbb{Z}$. In un certo senso, queste probabilità esauriscono le informazioni possibili: il processo stocastico è completamente determinato statisticamente quando sono note tutte le densità (discrete) congiunte, ovvero le densità di tutte le variabili discrete multiple (X_1, \dots, X_n) al variare di tutti gli n e tutti gli $i_0, \dots, i_n \in \mathbb{Z}$. Non dimostriamo questo fatto non banale.

La conoscenza di queste probabilità congiunte è in generale un problema molto arduo. Bisogna introdurre una qualche semplificazione e rinunciare alla totale generalità, per ottenere risultati pratici. Il concetto che viene maggiormente in aiuto è quello di catena di Markov. L'idea, dal punto di vista matematico, può partire ad esempio dal seguente calcolo. Usando il concetto di probabilità condizionale, vale

$$\begin{aligned} & P(X_0 = i_0, \dots, X_{n+1} = i_{n+1}) \\ = & P(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) \cdot P(X_0 = i_0, \dots, X_n = i_n). \end{aligned}$$

Pertanto, se fossero note le probabilità condizionali $P(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n)$, potremmo calcolare le probabilità congiunte $P(X_0 = i_0, \dots, X_{n+1} = i_{n+1})$ in modo *ricorsivo*, cioè tramite $P(X_0 = i_0, \dots, X_n = i_n)$, e così via. Ma le probabilità condizionali scritte sopra sono oggetti altrettanto complicati, salvo che per le catene di Markov, specialmente quelle omogenee.

1.2. Catene di Markov omogenee

Un processo stocastico discreto $(X_n)_{n \geq 0}$ si dice di *Markov* se

$$P(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n). \quad (1.1)$$

Questa è detta *proprietà di Markov*. Si intende che questa relazione deve valere al variare di tutti i parametri, e qualora siano ben definite le probabilità condizionate in oggetto.

Si dice poi *omogeneo* se le probabilità di transizione $P(X_{n+1} = j | X_n = i)$ non dipendono da n , ma solo da i e j . Quando questo accade, posto

$$p_{ij} = P(X_{n+1} = j | X_n = i)$$

possiamo calcolare tutte le probabilità congiunte conoscendo i soli numeri p_{ij} più la distribuzione iniziale $p_i^{(0)} = P(X_0 = i)$ (distribuzione del processo al tempo zero). Le probabilità p_{ij} si dicono *probabilità di transizione*, e precisamente p_{ij} è la probabilità di transizione da i a j in un passo temporale.

Proposition 1. *Se (X_n) è una catena di Markov omogenea con probabilità di transizione p_{ij} e distribuzione iniziale $p_i^{(0)} = P(X_0 = i)$, allora*

$$P(X_0 = i_0, \dots, X_n = i_n) = p_{i_0}^{(0)} \cdot p_{i_0 i_1} \cdots p_{i_{n-1} i_n}.$$

Dimostrazione. Usando la nota formula $P(A \cap B) = P(A|B)P(B)$ e le ipotesi, vale

$$\begin{aligned} & P(X_0 = i_0, \dots, X_n = i_n) \\ &= P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \cdot P(X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \\ &= P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \cdot P(X_n = i_n | X_{n-1} = i_{n-1}) \\ &= P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \cdot p_{i_{n-1} i_n}. \end{aligned}$$

Riapplicando in maniera ricorsiva questo calcolo al termine $P(X_0 = i_0, \dots, X_{n-1} = i_{n-1})$ e così via, otteniamo

$$\begin{aligned} &= P(X_0 = i_0, \dots, X_{n-2} = i_{n-2}) \cdot p_{i_{n-2}i_{n-1}} \cdot p_{i_{n-1}i_n} = \dots \\ &= P(X_0 = i_0) \cdot p_{i_0i_1} \cdots p_{i_{n-1}i_n}. \end{aligned} \tag{1.2}$$

La dimostrazione è completa.

Remark 1. *L'interpretazione della proprietà di Markov è la seguente: la previsione statistica dello stato $X_{n+1} = i_{n+1}$ che si può effettuare conoscendo la storia passata e presente $X_0 = i_0, \dots, X_n = i_n$, è identica alla previsione statistica effettuabile conoscendo il solo stato presente $X_n = i_n$. E' una forma di perdita di memoria del passato, una volta noto lo stato presente.*

Remark 2. *L'omogeneità temporale ha un'interpretazione immediata: la dinamica stocastica che determina i salti del processo di Markov non dipende dal tempo, ovvero agisce sempre con le stesse modalità (aleatorie ma tempo-invarianti dal punto di vista statistico).*

A livello di notazioni, esistono vari usi in letteratura per le probabilità di transizione. Ciò che sopra abbiamo indicato con p_{ij} viene a volte indicato con $p(i, j)$, con $p(j|i)$, con $p_i(j)$, con $p_{i \rightarrow j}$ e possibilmente altre.

Conoscendo le probabilità di transizione p_{ij} e la distribuzione iniziale $p^{(0)}$ è possibile anche il calcolo delle $p_i^{(n)}$.

Proposition 2. *Se (X_n) è una catena di Markov omogenea con probabilità di transizione p_{ij} e distribuzione iniziale $p^{(0)}$, allora*

$$p_i^{(n)} = \sum_{i_0, i_1, \dots, i_{n-1}} p_i^{(0)} \cdot p_{i_0i_1} \cdots p_{i_{n-1}i}.$$

Dimostrazione. Per la formula delle probabilità totali vale

$$\begin{aligned} p_i^{(n)} &= P(X_n = i) \\ &= \sum_{i_0, i_1, \dots, i_{n-1}} P(X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \end{aligned}$$

e quindi, usando la proposizione precedente,

$$p_i^{(n)} = \sum_{i_0, i_1, \dots, i_{n-1}} p_i^{(0)} \cdot p_{i_0i_1} \cdots p_{i_{n-1}i}.$$

La dimostrazione è completa.

1.3. Grafi e matrici

Lo studio delle catene di Markov omogenee diventa particolarmente semplice ed efficace utilizzando la rappresentazione tramite grafi e la descrizione con le matrici. In particolare, la formula espressa dalla proposizione precedente diventa molto più leggibile. Rappresentiamo gli stati i, j, \dots della catena con dei piccoli cerchi nel piano, detti *nodì* o *vertici* del grafo, e tracciamo una freccia dal nodo i al nodo j se $p_{ij} > 0$. Da ogni punto esce almeno una freccia. Alcuni punti possono non essere connessi da frecce, alcune coppie di punti possono essere connessi da frecce in ambo i sensi, alcuni punti possono essere connessi a se stessi, ed eventualmente solo a se stessi.

Ad ogni connessione $i \rightarrow j$ associamo il numero p_{ij} , la probabilità di transizione dallo stato i allo stato j . I numeri p_{ij} soddisfano le proprietà:

$$p_{ij} \in [0, 1], \quad \sum_j p_{ij} = 1 \text{ per ogni } i.$$

Si può ribaltare il discorso: partendo da un grafo munito di numeri p_{ij} con queste proprietà è possibile costruire una catena di Markov con quelle probabilità di transizione. Introduciamo anche la matrice quadrata

$$P = (p_{ij})$$

detta *matrice di transizione*. Essa ha elementi non negativi, con somma uno su ciascuna riga.

Usando la matrice P possiamo riscrivere una formula importante vista sopra. Osserviamo che vale

$$(P \cdot P)_{ij} = \sum_{i_1} p_{i, i_1} p_{i_1 j}$$

e più in generale

$$(P^n)_{ij} = \sum_{i_1, \dots, i_{n-1}} p_{i i_1} \cdots p_{i_{n-1} j}.$$

Pertanto vale:

Proposition 3. *Se (X_n) è una catena di Markov omogenea con probabilità di transizione p_{ij} e distribuzione iniziale $p^{(0)}$, allora*

$$p_j^{(n)} = \sum_i p_i^{(0)} \cdot (P^n)_{ij}.$$

In termini vettoriali

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

dove intendiamo i vettori $p^{(n)}$ e $p^{(0)}$ come vettori riga, ed il prodotto $p^{(0)} \cdot P^n$ come prodotto di vettore riga per matrice.

In modo analogo si può dimostrare la seguente relazione tra i vettori delle probabilità a due istanti di tempo generici $k < n$:

$$p^{(n)} = p^{(k)} P^{n-k}.$$

Questa relazione dice che se si parte dalla distribuzione $p^{(k)}$ (a cui si era arrivati in k passi, ma che ora viene vista come distribuzione iniziale) e si eseguono $n - k$ passi, si arriva alla distribuzione $p^{(n)}$. Il passato prima del tempo k non serve per determinare $p^{(n)}$, quando si conosca $p^{(k)}$.

Remark 3. A volte è utile considerare la probabilità di transizione a n passi

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

Essa è la probabilità che il sistema si trovi al tempo n nello stato j , sapendo che è partito all'istante $t = 0$ dallo stato i . Si noti che essa non è la probabilità che il sistema resti in i fino al tempo $n - 1$ e poi all'istante n avvenga la transizione $i \rightarrow j$: il passaggio da i a j in n passi può avvenire attraverso vari passi intermedi. $p_{ij}^{(n)}$ sarà la somma delle probabilità che, sapendo che il sistema è partito all'istante $t = 0$ dallo stato i , successivamente abbia seguito questo o quel cammino $i \rightarrow i_1 \rightarrow i_2 \rightarrow \dots \rightarrow i_{n-1} \rightarrow i_n = j$ arrivando in j al tempo n . Allora

$$\begin{aligned} p_{ij}^{(n)} &= P(X_n = j | X_0 = i) \\ &= \sum_{i_1, i_2, \dots, i_{n-1}} P(X_n = j, X_{n-1} = i_{n-1}, \dots, X_1 = i_1 | X_0 = i) \\ &= \sum_{i_1, i_2, \dots, i_{n-1}} P(X_n = j, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i) \frac{1}{P(X_0 = i)}. \end{aligned}$$

Usando la relazione trovata sopra per le probabilità congiunte, troviamo

$$p_{ij}^{(n)} = \sum_{i_1, i_2, \dots, i_{n-1}} p_{i, i_1} p_{i_1, i_2} \dots p_{i_{n-2}, i_{n-1}} p_{i_{n-1}, i_n}.$$

Usando il simbolismo matriciale, abbiamo dimostrato che

$$p_{ij}^{(n)} = (P^n)_{ij}$$

dando così anche un'interpretazione agli elementi della matrice P^n .

Remark 4. Un commento di natura “fisica” sulla proprietà di Markov. Essa afferma che, noto il presente, il futuro non dipende dal passato. Questa proprietà è simile a quanto avviene in fisica per il concetto di stato: esso indica (di solito) l’insieme delle informazioni ad un certo istante che permettono di determinare le analoghe informazioni ad istanti successivi (tramite le leggi del moto), senza che si debba conoscere il passato. Ad esempio, per il moto di un punto materiale, la conoscenza di posizione e velocità permette di calcolare posizione e velocità agli istanti successivi (lo stato di un punto materiale che si muove nello spazio è il vettore a 6 componenti dato dalla sua posizione e velocità). Invece, se ad un certo istante conosciamo solo la posizione, diversa sarebbe la nostra previsione sul futuro se in aggiunta ci fosse data la posizione nel passato: da quest’ultima (derivando in t) potremmo ricavare la velocità all’istante presente e poi predire lo stato futuro. Quindi la sola posizione non è lo stato di un punto materiale. Da queste analogie si capisce perché chiamiamo stati i punti i, j, \dots delle catene di Markov, e soprattutto perché è naturale in certi problemi la proprietà di Markov.

Remark 5. Senza dirlo esplicitamente, sopra abbiamo considerato matrici infinite, cioè con infinite righe e colonne. Le operazioni matriciali sono inalterate. Nei casi più complicati, facendo un prodotto di due matrici infinite si devono eseguire delle serie (si pensi al significato di $(AB)_{ij} = \sum_k a_{ik}b_{kj}$). Tuttavia, nella maggior parte degli esempi tali serie si riducono a somme finite, perché quasi tutti i termini sono nulli. Quindi anche algebricamente la situazione non è molto diversa.

Exercise 1. Data la catena a tre stati 1,2,3, con probabilità di transizione $p_{11} = \frac{1}{2}$, $p_{12} = \frac{1}{4}$, $p_{13} = \frac{1}{4}$, $p_{21} = 1$, $p_{32} = 1$, sapendo che $P(X_0 = 1) = 1$, calcolare la distribuzione di X_3 .

Soluzione n. 1. Ovviamente è corretto usare la formula $p^{(3)} = p^{(0)}P^3$. Conviene però dare una risoluzione grafica, che presentando meno calcoli è anche più sicura. A tal scopo, prendiamo ad esempio $p_1^{(3)} = P(X_3 = 1)$. Dobbiamo vedere in quali modi può il sistema, partendo da 1 (questa è la condizione data per ipotesi), trovarsi in 1 dopo tre passi. Un cammino è quello in cui passa da 1 ad 1 per tre volte di seguito, che ha probabilità $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$. Un secondo cammino va da 1 a 2, poi torna in 1, e poi va di nuovo in 1; la sua probabilità è $\frac{1}{4} \cdot 1 \cdot \frac{1}{2} = \frac{1}{8}$. Un terzo è $1 \rightarrow 1 \rightarrow 2 \rightarrow 1$, con probabilità $\frac{1}{2} \cdot \frac{1}{4} \cdot 1 = \frac{1}{8}$. Infine, un quarto è $1 \rightarrow 3 \rightarrow 2 \rightarrow 1$ che ha probabilità $\frac{1}{4} \cdot 1 \cdot 1 = \frac{1}{4}$. Complessivamente, $P(X_3 = 1) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{4} = \frac{5}{8}$. Si calcolino analogamente $P(X_3 = 2)$ e $P(X_3 = 3)$.

Soluzione n. 2. A titolo di esempio, verifichiamo il risultato precedente tramite la formula $p^{(3)} = p^{(0)}P^3$. Vale

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

e quindi

$$\begin{aligned} P \cdot P &= \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{4} \cdot 1 + \frac{1}{4} \cdot 1 & \frac{1}{2} \cdot \frac{1}{4} & \frac{1}{2} \cdot \frac{1}{4} \\ & 1 \cdot \frac{1}{4} & 1 \cdot \frac{1}{4} \\ & 1 \cdot \frac{1}{4} & 1 \cdot \frac{1}{4} \end{pmatrix} = \begin{pmatrix} \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix} \end{aligned}$$

$$P^3 = P^2 \cdot P = \begin{pmatrix} \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{5}{8} & \frac{3}{16} & \frac{3}{16} \\ \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \\ \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \end{pmatrix}.$$

Infine, essendo $p^{(0)} = (1, 0, 0)$, vale

$$p^{(3)} = p^{(0)}P^3 = (1, 0, 0) \cdot \begin{pmatrix} \frac{5}{8} & \frac{3}{16} & \frac{3}{16} \\ \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \\ \frac{3}{4} & \frac{1}{8} & \frac{1}{8} \end{pmatrix} = \left(\frac{5}{8} \quad \frac{3}{16} \quad \frac{3}{16} \right).$$

In particolare $P(X_3 = 1) = p_1^{(3)} = \frac{5}{8}$, come con l'altro metodo.

Exercise 2. Si consideri la catena di Markov a due stati 1, 2 e probabilità di transizione

$$p_{11} = p, p_{12} = 1 - p, p_{22} = 1$$

dove $p \in (0, 1)$ è un numero dato. Calcolare le probabilità di transizione ad n passi.

Soluzione. La matrice di transizione è semplicemente

$$P = \begin{pmatrix} p & q \\ 0 & 1 \end{pmatrix}.$$

Le sue potenze sono

$$P^2 = \begin{pmatrix} p & q \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p & q \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} p^2 & pq + q \\ 0 & 1 \end{pmatrix},$$

$$P^3 = \begin{pmatrix} p & q \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p^2 & pq + q \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} p^3 & p^2q + pq + q \\ 0 & 1 \end{pmatrix}$$

ecc., da cui si capisce che

$$P^n = \begin{pmatrix} p^n & p^{n-1}q + p^{n-2}q + \dots + q \\ 0 & 1 \end{pmatrix}.$$

Da essa si leggono le probabilità di transizione ad n passi. Osserviamo però che la somma sulle righe è uno, quindi accorgiamoci che

$$p^{n-1}q + p^{n-2}q + \dots + q = 1 - p^n.$$

Si verifichi algebricamente questa identità. Inoltre, si osservi che potevamo stabilire l'uguaglianza

$$p_{11}^{(n)} = p^n$$

senza fare la potenza n -esima di P , ma tramite semplici considerazioni grafiche. Quindi, in definitiva, si poteva pervenire alle probabilità di transizione ad n passi in modo grafico e con semplici considerazioni.

Exercise 3. Calcolare inoltre la probabilità che il sistema prima o poi si trovi in 2, sapendo che all'istante $t = 0$ si trova in 1. (Attenzione: la probabilità di trovarsi prima o poi in 2 (condizionata...) non è la somma su n delle probabilità di trovarsi in 2 al tempo n , sapendo che all'istante $t = 0$ si trova in 1, ovvero $\sum_{n=1}^{\infty} p_{12}^{(n)}$. Infatti questi non sono eventi disgiunti. Provate a verificare che $\sum_{n=1}^{\infty} p_{12}^{(n)} = +\infty$.)

2. Probabilità invarianti

Data una catena di Markov, il calcolo delle probabilità congiunte o di $p_i^{(n)}$ è in linea di principio possibile ed agevole usando la matrice di transizione P e la distribuzione iniziale. Con l'uso di un calcolatore non è troppo dispendioso calcolare queste probabilità anche per n alto. Se però cerchiamo risultati senza calcolatore, abbiamo visto già negli esercizi precedenti che le possibilità di calcolo delle $p_i^{(n)}$ è limitato o a pochi esempi facili oppure ad n molto basso.

Ciò che invece è spesso facile da calcolare sono le *probabilità invarianti*.

Definition 1. Un vettore di probabilità π (cioè tale che $\sum \pi_i = 1$, $\pi_i \in [0, 1]$ per ogni i), si dice invariante per la catena di Markov descritta dalla matrice di transizione P se $\pi = \pi P$.

In termini dinamici questo significa che partendo da una situazione iniziale aleatoria con distribuzione di probabilità π , dopo un passo (e quindi dopo un numero arbitrario di passi, essendo

$$\pi P^n = \pi P P^{n-1} = \pi P^{n-1} = \dots = \pi)$$

la distribuzione di probabilità è sempre π . Si può dimostrare inoltre che in questo caso il processo stocastico $\{X_n\}$ è *stazionario* in senso stretto. Per questo le componenti di π si dicono anche *probabilità stazionarie*.

Si usano vari altri nomi equivalenti, come distribuzione invariante, massa invariante.

In larga misura l'interesse per le probabilità invarianti deriva dallo studio di $p_i^{(n)}$ nel limite per $n \rightarrow \infty$. Supponiamo che questo limite esista ed indichiamolo con π_i :

$$\pi_i = \lim_{n \rightarrow \infty} p_i^{(n)}. \quad (2.1)$$

Siccome vale la relazione

$$p^{(n+1)} = p^{(n)} P$$

(come caso particolare della $p^{(n)} = p^{(k)} P^{n-k}$), passando al limite per $n \rightarrow \infty$ otteniamo

$$\pi = \pi P.$$

Quindi π è una probabilità invariante. Abbiamo dimostrato:

Lemma 1. Se le probabilità $p^{(n)}$ al passo n -esimo convergono ad un vettore di probabilità π , allora π è invariante.

Il viceversa non è vero (lo discuteremo in un'altro paragrafo): esistono esempi con una certa caratteristica di periodicità in cui ci sono soluzioni di $\pi = \pi P$ ma $p^{(n)}$ non converge. Inoltre, un po' articolato è il viceversa nel caso in cui l'equazione $\pi = \pi P$ ha più di una soluzione (il limite delle $p^{(n)}$ dipende dalle condizioni iniziali).

Nel caso di matrici infinite questo teorema richiede delle precisazioni, ma per ora restiamo ad un livello lievemente non rigoroso.

L'interpretazione che diamo dei numeri π_i è la seguente: π_i è approssimativamente la probabilità di trovare il sistema nello stato i , dopo che è trascorso un tempo sufficientemente lungo. Allora la relazione $\pi = \pi P$ rappresenta un enorme vantaggio di calcolo, in quanto il calcolo di $p^{(n)}$ era molto complicato mentre spesso non lo è la risoluzione di questo sistema lineare.

Notiamo che l'equazione $\pi = \pi P$ è equivalente all'equazione $(I - P)^T \pi^T = 0$.

Nota. Una equazione del tipo $(I - P)^T \pi^T = 0$ ha sempre la soluzione $\pi^T = 0$, che però non ci interessa (cerchiamo una distribuzione di probabilità π). In generale, tale equazione potrebbe non avere altre soluzioni (se la matrice $I - P$ ha rango massimo, sempre che stiamo studiando un sistema con un numero finito di stati, altrimenti le cose sono ancora più complicate). Se invece esistono soluzioni $\pi \neq 0$, ne esistono infinite: il nucleo della matrice $I - P$ è uno spazio vettoriale. Bisogna allora ricordarsi di imporre le altre condizioni generali che deve soddisfare una distribuzione di probabilità: $\pi_i \in [0, 1]$, $\sum_i \pi_i = 1$. In casi "buoni" (quelli che incontreremo lo sono), queste condizioni individuano una unica soluzione.

Exercise 4. *Calcolare tutte le probabilità invarianti per una generica catena a due stati.*

In pratica, per calcolare π a partire da P si ragiona nel seguente modo. Prima di tutto si effettua la cosiddetta classificazione degli stati, che descriveremo tra un attimo. Dopo essersi ricondotti, tramite essa, ad una classe irriducibile (in breve un insieme di stati in cui tutti gli stati comunicano tra di loro), si cerca una π in tale classe. Per calcolare π si può ovviamente affrontare il sistema $\pi = \pi P$, ma, almeno in casi semplici, conviene adottare un metodo grafico detto del *bilancio di flusso*. Anche questo verrà descritto tra breve.

2.1. Classificazione degli stati

2.1.1. Stati comunicanti

Si dice che lo stato i comunica con lo stato j se esiste un $n > 0$ tale che $p_{ij}^{(n)} > 0$. Questo si indica con $i \rightarrow j$. Se i non comunica con j scriveremo $i \nrightarrow j$.

2.1.2. Classe chiusa

Indicando con S l'insieme degli stati di una catena, un sottoinsieme $A \subset S$ è una classe chiusa se gli stati di A non comunicano con gli stati del complementare A^c .

2.1.3. Classe chiusa irriducibile

Una classe chiusa A si dice irriducibile se tutti i suoi stati comunicano fra loro.

Se uno stato costituisce da solo una classe irriducibile, esso si dice *assorbente*.

Una catena di Markov si dice irriducibile se tutti i suoi stati comunicano, ovvero se la classe di tutti i suoi stati è una classe irriducibile.

2.1.4. Stati transitori, stati ricorrenti

Uno stato i si dice transitorio o transiente se con probabilità strettamente positiva la catena che parte da i non ci torna più.

Al contrario, uno stato i si dice ricorrente se vale 1 la probabilità che la catena, partendo da esso, ci ritorni. Per definizione, uno stato è o ricorrente, o transitorio.

Uno stato assorbente è ricorrente.

2.1.5. Alcune proprietà generali

Enunciamo, senza dimostrazione, alcune proprietà che aiutano nella classificazione degli stati. Chiamiamo catena finita una catena il cui insieme S degli stati è finito.

Proposition 2. *Se per lo stato i vale la condizione*

$$\exists j \text{ tale che } i \rightarrow j \text{ ma } j \not\rightarrow i$$

allora i è transiente. Se la catena è finita, allora la condizione precedente è necessaria e sufficiente per la transienza di i .

Proposition 3. *Una catena finita ha almeno uno stato ricorrente.*

E' su questa proprietà che si basa il seguente fatto, che spiegheremo meglio più avanti: una catena finita ha sempre almeno una misura invariante.

La proprietà di ricorrenza è contagiosa, come esprime la seguente proposizione.

Proposition 4. *Se i è ricorrente e $i \rightarrow j$, allora anche j è ricorrente.*

Da questo discende:

Proposition 5. *Se una catena è irriducibile, allora gli stati sono o tutti ricorrenti, o tutti transitori.*

Da due di queste proprietà discende:

Proposition 6. *Se una catena è finita ed irriducibile, allora tutti i suoi stati sono ricorrenti.*

2.2. Probabilità invarianti (continuazione)

Indichiamo con S l'insieme degli stati di una catena di Markov omogenea $(X_n)_{n \geq 0}$. Per tutto questo capitolo supporremo che S sia un **insieme finito**. Molti risultati si generalizzano o direttamente o con alcune modifiche al caso in cui S ha un'infinità numerabile di elementi; altri risultati non si generalizzano; rimandiamo ai testi specializzati per una discussione dettagliata.

Ricordiamo che le probabilità invarianti sono importanti per almeno due ragioni. Da un lato esse sono un'approssimazione delle probabilità al tempo n , le $p_i^{(n)} = P(X_n = i)$. Infatti, come vedremo, sotto opportune ipotesi il vettore $p^{(n)}$ converge a π . Quando questo accade, il complicato calcolo delle $p^{(n)}$ si può rimpiazzare col calcolo molto più agevole della π , che richiede "solo" la risoluzione del problema algebrico $(I - P)^T \pi^T = 0$. Dall'altro, la conoscenza di π fornisce informazioni sullo stato del sistema fisico che stiamo descrivendo con la catena di Markov, se quando cominciamo ad osservare il sistema (cioè al tempo che per noi è $t = 0$) il sistema è già da tempo soggetto ad evoluzione casuale. In altre parole, quello che per noi è il tempo $t = 0$, per il sistema può essere un tempo già molto elevato, e quindi il sistema può già trovarsi nella situazione invariante (stazionaria) descritta dalla probabilità π .

Precisiamo un fatto: nel seguito per probabilità invariante intendiamo una qualsiasi soluzione dell'equazione algebrica $(I - P)^T \pi^T = 0$ (indipendentemente dal fatto che si ottenga come limite delle $p^{(n)}$).

Ci chiediamo ora:

- 1) data una catena di Markov, esiste sempre almeno una probabilità invariante π ?
- 2) è unica?
- 3) vale la *convergenza all'equilibrio*, ovvero la proprietà

$$p^{(n)} \rightarrow \pi$$

che abbiamo già discusso in modo non rigoroso? Per quali distribuzioni iniziali $p^{(0)}$ vale?

3. Esistenza ed unicità

Per rispondere ad alcune di queste domande ricordiamo i concetti sulla classificazione degli stati, precedentemente introdotti: stati comunicanti, classi chiuse, classi irriducibili, stati transitori (o transienti), stati ricorrenti. Si può dimostrare,

partendo dalle varie proprietà descritte in precedenza per questi concetti, che l'insieme S degli stati si decompone in alcuni sottoinsiemi, come segue. Indichiamo con T l'insieme degli stati transienti, e con R l'insieme degli stati ricorrenti. Vale per definizione

$$S = T \cup R.$$

A sua volta, l'insieme R si decompone nell'unione di classi irriducibili. Esistono cioè delle classi irriducibili C_1, \dots, C_N , disgiunte, con unione pari ad R :

$$R = C_1 \cup \dots \cup C_N.$$

Quindi in definitiva vale la decomposizione

$$S = T \cup C_1 \cup \dots \cup C_N.$$

Non esistono altri sottoinsiemi di stati. Invitiamo il lettore a raffigurare questa decomposizione.

Invitiamo inoltre il lettore a farsi un'idea intuitiva e dinamica di come evolve nel tempo la massa in questi insiemi. Al tempo $n = 0$ abbiamo la distribuzione di probabilità $p^{(0)}$, che dobbiamo pensare come una distribuzione di massa (discreta, cioè concentrata negli stati del sistema). Al tempo $n = 1$ questa si è trasformata nella distribuzione $p^{(1)}$, e così via. Nel limite $n \rightarrow \infty$ questa tende alla distribuzione di massa π (nota: per ora questa convergenza va presa intuitivamente, perchè non abbiamo chiarito sotto quali ipotesi valga). Ecco allora un'idea intuitiva: la massa che inizialmente si trova nell'insieme transiente T , prima o poi esce da esso, entrando nelle varie classi irriducibili. Di conseguenza, ci aspettiamo che la distribuzione limite π dia massa zero ai punti dell'insieme T . Questo si può dimostrare rigorosamente:

$$\begin{aligned} &\text{ogni probabilità invariante } \pi \text{ ha la proprietà} \\ \pi_i &= 0 \text{ per ogni stato transiente } i. \end{aligned}$$

Esprimeremo questo dicendo che T *non supporta* massa invariante.

Riprendiamo l'immagine intuitiva. La massa che si trova in una classe irriducibile C_k da questa classe non può uscire. Questo fa pensare che ci sia una massa invariante in questa classe. Questo si può dimostrare rigorosamente. Inoltre, il fatto che la classe sia irriducibile, ovvero che ogni stato comunichi con

ogni altro, fa sì che la massa invariante sia unica. Si può dimostrare rigorosamente il seguente risultato non banale:

in ogni classe irriducibile C_k esiste
una ed una sola probabilità invariante $\pi^{(k)}$.

Precisiamo il significato di questa frase. Esiste una probabilità invariante $\pi^{(k)}$ supportata da C_k , ovvero tale che

$$\pi_i^{(k)} = 0 \text{ per ogni stato } i \text{ che non appartiene a } C_k.$$

Inoltre, ne esiste una sola con questa proprietà. Si può ulteriormente dimostrare, sempre grazie alla comunicatività tra tutti gli stati della classe irriducibile, che $\pi^{(k)}$ attribuisce massa strettamente positiva ad ogni stato di C_k .

Siamo quindi arrivati a questa immagine: l'insieme transiente T non contiene massa invariante; ogni classe irriducibile C_k contiene una ed una sola massa invariante. Queste però non sono tutte le possibili probabilità invarianti della catena nel suo complesso. Detto altrimenti, se ci restringiamo ad una classe irriducibile, lì c'è una sola probabilità invariante, ma se guardiamo il sistema nel suo complesso, ce ne sono altre. Queste altre sono tutte e sole le *combinazioni convesse* delle $\pi^{(k)}$, cioè i vettori di probabilità π che si possono esprimere nella forma

$$\pi = \alpha_1 \pi^{(1)} + \dots + \alpha_N \pi^{(N)}$$

dove i coefficienti soddisfano

$$\alpha_i \in [0, 1], \quad \sum_{i=1}^N \alpha_i = 1.$$

Infatti è immediato verificare che, essendo le $\pi^{(k)}$ invarianti, lo sono anche queste combinazioni convesse (l'invarianza è una condizione lineare, quindi si conserva per le combinazioni lineari; la restrizione di considerare combinazioni convesse deriva dal fatto che π deve essere un vettore di probabilità). Si dimostra poi che non ci sono altre probabilità invarianti oltre a queste.

Resta comunque chiaro che le probabilità invarianti $\pi^{(k)}$ sono le più importanti, e che ci si deve concentrare sulla loro ricerca ed indagine.

3.1. Convergenza all'equilibrio

Quanto detto fino ad ora esaurisce le domande 1) e 2) poste sopra, su esistenza ed unicità delle distribuzioni invarianti. La domanda 3) si collega a quanto appena detto, con un'ulteriore complicazione.

Innanzitutto osserviamo cosa intuitivamente accade alla massa iniziale $p^{(0)}$ (riprendiamo l'idea intuitiva descritta sopra). La parte di questa massa che occupa T a poco a poco fluisce fuori da T ed entra, definitivamente, in qualche classe irriducibile, aggiungendosi così alla massa già presente in quella classe irriducibile (massa che non può uscire dalla classe). Il ragionamento si riconduce allora a capire cosa succede al crescere del tempo alla massa che occupa una certa classe irriducibile. Prima di far questo osserviamo che è possibile (noi non lo faremo) studiare varie proprietà di questo *assorbimento* di massa da parte delle classi irriducibili: probabilità di finire in una o un'altra classe, tempo medio di entrata in quella classe, ecc.

Restringiamoci quindi ad una classe irriducibile. In altre parole, studiamo una *catena di Markov irriducibile*, avendo capito che a questo studio ci si riconduce dal caso generale.

La difficoltà aggiuntiva a cui abbiamo accennato è che, pur restringendosi ad una catena irriducibile, la proprietà di convergenza all'equilibrio può non valere. Ecco un esempio.

Esempio. Consideriamo una catena a due stati, detti 1 e 2. Supponiamo che sia

$$p_{11} = 0, \quad p_{12} = 1, \quad p_{21} = 1, \quad p_{22} = 0.$$

In altre parole, la catena dallo stato 1 passa allo stato 2, e viceversa. Partiamo con tutta la massa iniziale concentrata nello stato 1. Dopo un passo, tutta la massa è concentrata nello stato 2, e così via alternativamente. Al variare del tempo, la massa oscilla periodicamente tra i due stati. Quindi è chiaro che questa massa non converge ad alcuna massa limite π . Quindi non si ha convergenza all'equilibrio delle $p^{(n)}$. Nonostante questo, una massa invariante esiste (ed è unica): è la massa che vale $\frac{1}{2}$ in ciascun stato.

Riflettendo su varianti più complicate di questo esempio si arriva a capire che la difficoltà è causata da qualche forma di periodicità della catena, che produce delle oscillazioni periodiche di massa, non convergenti ad un equilibrio. Una condizione

perché questo non avvenga è la seguente:

esiste n tale che $p_{ij}^{(n)} > 0$ per ogni i, j .

Questa condizione afferma che esiste un n fissato relativamente al quale tutti gli stati comunicano tra loro. Si dimostra (non è banale) che questo rompe le possibili periodicità ed implica la convergenza all'equilibrio. Il teorema afferma:

sotto questa ipotesi vale $p^{(n)} \rightarrow \pi$ per ogni distribuzione iniziale $p^{(0)}$.

Si noti che π esiste ed è unica, perché abbiamo supposto dall'iniziazione che la catena fosse irriducibile.

Osservazione. Ci si può chiedere se nei casi “periodici” non si salvi qualche forma debole di convergenza all'equilibrio. In effetti, pensando all'esempio precedente, vediamo che la massa oscilla tra i due stati senza convergere, ma in un certo senso medio (una media temporale) essa tende alla massa uguale a $\frac{1}{2}$. In assoluta generalità si può dimostrare il seguente *teorema ergodico*:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p^{(k)} = \pi.$$

In altre parole, se anche le $p^{(k)}$ non tendono a π , almeno ci tendono le medie temporali.

3.2. Calcolo delle probabilità invarianti

Mettiamoci nel caso irriducibile. Il calcolo delle probabilità invarianti, come già detto, equivale alla ricerca dell'autovettore sinistro di P con autovalore uno

$$\pi = \pi P$$

preoccupandosi di trovare, tra tutti i multipli possibili, quello che è un vettore di probabilità. Equivalentemente si cerca il vettore di probabilità che risolve l'equazione algebrica

$$(I - P)^T \pi^T = 0$$

avendo l'accortezza di imporre sempre le condizioni che sia un vettore di probabilità. Le metodologie note dell'algebra lineare, e quelle numeriche corrispondenti,

sono ovviamente di aiuto. Vogliamo però indicare alcune scorciatoie o di tipo grafico o valide in casi particolari.

1) La prima osservazione è che possiamo riscrivere la generica componente j -esima dell'equazione $\pi = \pi P$,

$$\pi_j = \sum_i \pi_i p_{ij}$$

nella forma

$$(1 - p_{jj}) \pi_j = \sum_{i \neq j} \pi_i p_{ij} \quad (3.1)$$

(basta isolare nella sommatoria il termine con $i = j$ e portarlo a sinistra). Questa equazione si presta ad una utilissima interpretazione grafica. Osservando che p_{jj} è la probabilità che la massa π_j non fluisca fuori dallo stato j . Allora $(1 - p_{jj})$ è la probabilità che la massa π_j fluisca fuori dallo stato j , e pertanto $(1 - p_{jj}) \pi_j$ è la proporzione di massa che esce dallo stato j . D'altra parte, per ogni stato $i \neq j$, $\pi_i p_{ij}$ è la massa che arriva a j dallo stato i , e la somma $\sum_{i \neq j} \pi_i p_{ij}$ rappresenta la massa che giunge allo stato j dagli altri stati. Per queste ragioni l'equazione (3.1) viene detta *equazione di bilancio del flusso*. Essa si può leggere facilmente sul grafo della catena e può essere di grande aiuto nello studio della probabilità invariante.

2) A volte, basandosi sull'intuizione, di solito dovuta a qualche particolare simmetria del grafo della catena, si riesce ad intuire a priori quanto dovrebbe valere il vettore π . In tal caso, è sufficiente *verificare* l'uguaglianza $\pi = \pi P$. Questa verifica si può svolgere algebricamente, sostituendo nell'equazione $\pi = \pi P$, oppure appoggiandosi al grafo, e verificando l'equazione di bilancio del flusso (3.1) descritta sopra.

3) In aiuto a questa semplice strategia viene anche il concetto di *equilibrio dettagliato*. Diciamo che, per un vettore π (di cui non si richiede a priori che sia invariante) vale l'equilibrio dettagliato se per ogni i, j vale

$$\pi_i p_{ij} = \pi_j p_{ji}.$$

Il significato intuitivo di questa condizione (riprendendo l'idea intuitiva usata più volte sopra) è che la percentuale p_{ij} della massa π_i che si sposta in un passo dallo

stato i allo stato j è esattamente uguale alla percentuale p_{ji} della massa π_j che si sposta in un passo dallo stato j allo stato i . Di conseguenza, in un passo, gli stati i e j si trasmettono l'un l'altro la stessa quantità di massa. E' quindi una versione molto particolare e semplice dell'equazione di bilancio del flusso.

La condizione di equilibrio dettagliato implica che π è invariante. Quest'affermazione è chiara a livello intuitivo per quanto appena detto, e si verifica algebricamente in modo semplice (si veda l'esercizio seguente).

Dal punto di vista pratico, spesso è immediato verificare la condizione di equilibrio dettagliato per un vettore π che pensiamo possa essere invariante. Oppure, a volte si riesce a scoprire una massa invariante cercandone una che verifica l'equilibrio dettagliato. In questo modo si trovano esplicitamente le probabilità invarianti di molti esempi, evitando calcoli laboriosi. Si veda il secondo esercizio dato sotto.

Esercizio. Dimostrare (algebricamente) che la condizione di equilibrio dettagliato implica la condizione di invarianza $\pi = \pi P$.

Esercizio. Si consideri una catena con 4 stati, tutti comunicanti tra loro, con tutte le probabilità di transizione uguali tra loro. Trovare la probabilità invariante.

4. Processo di nascita e morte a tempo discreto

Sviluppiamo ora un esempio che esula un poco dalla trattazione teorica precedente, in quanto il numero degli stati è infinito. Tuttavia si riesce a calcolare in modo esplicito una distribuzione invariante e si capisce che è unica. Questo esempio è fondamentale in teoria delle code.

Supponiamo che la catena (X_n) abbia come stati tutti i numeri interi non negativi, ovvero $0, 1, \dots$

Supponiamo che da ogni stato i siano possibili solo le transizioni ad $i, i + 1$ ed $i - 1$, con l'eccezione dello stato 0 da cui si può transire solo in 1 o restare in 0.

Indichiamo con λ_0 la probabilità di transire da 0 a 1, per cui $1 - \lambda_0$ è la probabilità di restare in 0.

Poniamo poi, per ogni stato $i > 0$,

$$\begin{aligned}\lambda_i &= p_{i,i+1} \\ \mu_i &= p_{i,i-1}\end{aligned}$$

per cui vale

$$p_{i,i} = 1 - \lambda_i - \mu_i.$$

La matrice di transizione, infinita, è

$$P = \begin{pmatrix} 1 - \lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \mu_1 & 1 - \lambda_1 - \mu_1 & \lambda_1 & 0 & \dots \\ 0 & \mu_2 & 1 - \lambda_2 - \mu_2 & \lambda_2 & \dots \\ 0 & 0 & \mu_3 & 1 - \lambda_3 - \mu_3 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

Supponiamo infine che tutti i numeri λ_i (incluso $i = 0$) e μ_i siano strettamente positivi.

Il significato di questa catena in teoria delle code il seguente. La catena conta il numero di persone presenti in una coda. Ad ogni istante n di tempo, essendo i il numero di utenti in coda, può non accadere nulla, ovvero restare i tale numero, con probabilità $1 - \lambda_i - \mu_i$. Oppure può venir servito un utente, che esce dalla coda, per cui il numero di utenti diventa $i - 1$, e questo accade con probabilità μ_i . Oppure infine può arrivare un nuovo utente, quindi il numero aumenta di uno, e ciò avviene con probabilità λ_i .

Ci chiediamo: nel regime stazionario, ovvero quando è passato un tempo abbastanza grande e la coda si è statisticamente assestata, qual'è la probabilità di avere i persone in coda? Dobbiamo calcolare la distribuzione invariante.

4.1. Preliminare: stati $0, \dots, N$

Come preliminarmente, consideriamo lo stesso problema con un numero finito di stati $0, \dots, N$, in modo da applicare fatti più semplici. Per le probabilità di transizione usiamo le stesse notazioni del caso infinito, salvo il fatto che dallo stato N non si può transire in $N + 1$ (quindi $\lambda_N = 0$).

Ogni stato comunica con ogni altro, avendo presupposto strettamente positivi tutti i λ_i e μ_i . Quindi $\{0, \dots, N\}$ è un'unica classe irriducibile, che supporta un'unica misura invariante π .

Il bilancio di flusso nel generico stato $i = 1, \dots, N - 1$ è

$$\pi_i (\lambda_i + \mu_i) = \pi_{i-1} \lambda_{i-1} + \pi_{i+1} \mu_{i+1}.$$

Il bilancio in $i = 0$ è invece

$$\pi_0 \lambda_0 = \pi_1 \mu_1.$$

Quindi

$$\pi_1 = \pi_0 \frac{\lambda_0}{\mu_1}.$$

Il bilancio in $i = 1$ è

$$\pi_2 \mu_2 = \pi_1 (\lambda_1 + \mu_1) - \pi_0 \lambda_0$$

e quindi, sostituendo la formula precedente,

$$\pi_2 \mu_2 = \pi_0 \frac{\lambda_0}{\mu_1} (\lambda_1 + \mu_1) - \pi_0 \lambda_0$$

ovvero

$$\pi_2 = \pi_0 \frac{\lambda_0 (\lambda_1 + \mu_1) - \mu_1 \lambda_0}{\mu_1 \mu_2} = \pi_0 \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2}.$$

Così di seguito si riconosce la validità della seguente formula (la dimostrazione rigorosa andrebbe svolta per induzione).

Theorem 1. *Posto*

$$a_i = \frac{\lambda_0 \lambda_1 \dots \lambda_{i-1}}{\mu_1 \mu_2 \dots \mu_i},$$

la distribuzione invariante per la catena ora descritta è

$$\pi_i = \frac{a_i}{\sum_{i=0}^N a_i}$$

Infatti dai calcoli precedenti si trova

$$\pi_i = \pi_0 a_i$$

e poi imponendo $\sum_{i=0}^N \pi_i = 1$ si trova

$$\pi_0 = \frac{1}{\sum_{i=0}^N a_i}.$$

4.2. Il caso infinito

Tornando alla catena di nascita e morte con infiniti stati, i calcoli sono gli stessi. L'unica differenza è che la serie $\sum_{i=0}^{\infty} a_i$ potrebbe divergere a più infinito (è a termini positivi, quindi o converge o diverge). Quando questo accade, la condizione $\sum_{i=0}^{\infty} \pi_i = 1$ non può essere soddisfatta con alcuna scelta di π_0 , quindi non si

trova, con questi calcoli, alcuna misura invariante. Può restare il dubbio teorico che ce ne sia una rintracciabile in altro modo, ma non è così, come si intuisce ovviamente sulla base del caso con stati finiti e con semplice buon senso (omettiamo però la dimostrazione). Abbiamo quindi il seguente risultato.

Theorem 2. *Posto*

$$a_i = \frac{\lambda_0 \lambda_1 \dots \lambda_{i-1}}{\mu_1 \mu_2 \dots \mu_i},$$

se risulta

$$\sum_{i=0}^{\infty} a_i < \infty$$

allora la distribuzione invariante per la catena di nascita e morte è

$$\pi_i = \frac{a_i}{\sum_{i=0}^{\infty} a_i}.$$

Se invece $\sum_{i=0}^{\infty} a_i = +\infty$, allora non ci sono distribuzioni invarianti.

Esaminiamo un caso particolare in cui possono accadere entrambe le situazioni.

Example 2. *Supponiamo che sia $\lambda_i = \lambda$ e $\mu_i = \mu$ per ogni i . Allora*

$$a_i = \left(\frac{\lambda}{\mu}\right)^i$$

e quindi

$$\sum_{i=0}^{\infty} a_i = \sum_{i=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^i$$

converge se e solo se $\frac{\lambda}{\mu} < 1$ (ricordiamo che $\frac{\lambda}{\mu} > 0$) ovvero se e solo se

$$\lambda < \mu.$$

Nelle code di servizio questo si interpreta dicendo che se il tasso di servizio è superiore al tasso di arrivo di nuovi utenti, allora c'è una distribuzione invariante, ovvero il sistema raggiunge una situazione di equilibrio. Se invece il tasso di servizio è inferiore al tasso di arrivo, allora la coda non riesce a smaltire gli arrivi, aumenta indefinitamente e non si assesta in un regime di equilibrio.

1 Processi di Markov a salti

1.1 Definizioni

Consideriamo un processo stocastico a valori in Z , o in un suo sottoinsieme di stati S , ma con parametro temporale variabile su tutto $[0, \infty)$. Il processo è quindi della forma $(X_t)_{t \geq 0}$, dove le v.a. X_t sono v.a. discrete.

Un tale processo si dice un *processo di Markov a salti* (oppure una *catena di Markov a tempo continuo*) se vale

$$P(X_{t_{n+1}} = i_{n+1} | X_{t_0} = i_0, \dots, X_{t_n} = i_n) = P(X_{t_{n+1}} = i_{n+1} | X_{t_n} = i_n)$$

al variare di tutti gli stati possibili e dei tempi $t_0 \leq \dots \leq t_{n+1}$. L'interpretazione di questa condizione è identica al caso del tempo discreto. Un tale processo si dice poi *omogeneo* se la probabilità (per ogni i, j)

$$P(X_{s+t} = j | X_s = i)$$

dipende solo dal lasso di tempo t tra le due osservazioni e non dall'istante s della prima osservazione. Poniamo allora

$$p_{ij}(t) = P(X_{s+t} = j | X_s = i)$$

che chiameremo *probabilità di transizione da i a j nel tempo t* . Ci restringeremo sempre a processi di Markov omogenei; per questi il ruolo fondamentale è svolto dalle grandezze $p_{ij}(t)$. Tramite queste probabilità di transizione si possono calcolare tutte le probabilità congiunte del tipo

$$P(X_{t_0} = i_0, \dots, X_{t_n} = i_n)$$

(in modo del tutto analogo al caso del tempo discreto) e quindi ottenere ogni informazione statistica sul processo.

Le probabilità di transizione $p_{ij}(t)$ sono l'analogo delle probabilità di transizione a n passi $p_{ij}^{(n)}$, rispetto alle catene a tempo discreto. Quello che manca, nel caso a tempo continuo, è l'analogo delle probabilità di transizione ad un passo p_{ij} , così decisive e basilari in quel contesto. Ricordiamo che nel caso a tempo discreto abbiamo basato tutta la nostra analisi sulle probabilità ad un passo p_{ij} . Nel caso a tempo continuo c'è quindi da superare la difficoltà dovuta alla mancanza di P .

Alcuni fatti, comunque, sono ancora simili nei due casi. Intanto è utile tracciare un grafo (vedremo sotto alcuni esempi). Poi, con le probabilità di transizione $p_{ij}(t)$ formiamo la *matrice di transizione*, che indichiamo con $P(t)$. Poniamo inoltre

$$p_j(t) = P(X(t) = j)$$

ed indichiamo con $p(t)$ il vettore (riga) di componenti $p_j(t)$. Vale allora la formula simile al caso discreto

$$p(t) = p(0)P(t)$$

che esprime la distribuzione di probabilità al tempo t tramite quella al tempo 0 trasformata dalla matrice $P(t)$. La dimostrazione è immediata:

$$\begin{aligned} p_j(t) = P(X(t) = j) &= \sum_i P(X(t) = j | X(0) = i) P(X(0) = i) \\ &= \sum_i p_{ij}(t) p_i(0). \end{aligned}$$

In modo analogo si ottiene anche il legame tra due istanti s e t

$$p(t) = p(s)P(t - s)$$

su cui si basiamo l'origine del concetto di probabilità invariante, ed anche la formula di composizione delle matrici

$$P(t + s) = P(t)P(s).$$

Definition 1 *Un vettore π di probabilità (cioè tale che $\pi_i \in [0, 1]$ e $\sum \pi_i = 1$) si dice invariante o stazionario per il processo a salti descritto dalla matrice $P(t)$ se $\pi = \pi P(t)$ per ogni $t \geq 0$.*

Il seguente paragrafo mostra come si può superare la difficoltà causata dalla mancanza del concetto di transizione ad un passo, transizioni che “generavano” quelle ad n passi. Anche a tempo continuo c'è un concetto di generatore.

1.2 La matrice A dei tassi di transizione (generatore infinitesimale)

E' chiaro che stiamo cercando grandezze legate a tempi piccoli (l'analogo del singolo passo temporale) da cui sia possibile calcolare tutto il resto. Ragionando allora per t piccolo viene in mente di usare lo sviluppo di Taylor al primo ordine della funzione $p_{ij}(t)$, per ogni i, j fissati. Per poter usare questo sviluppo dobbiamo supporre che $p_{ij}(t)$ sia derivabile in $t = 0$. Facciamo questa ipotesi. Nel prossimo sottoparagrafo faremo qualche considerazione più teorica su questo tema. Per ora decidiamo di imporre questa condizione di derivabilità, per poter andare avanti con la teoria.

Vale allora

$$p_{ij}(t) = p_{ij}(0) + p'_{ij}(0) \cdot t + o_{ij}(t)$$

dove con $o_{ij}(t)$ abbiamo indicato un infinitesimo, per $t \rightarrow 0$, di ordine superiore ad uno (ovvero $\lim_{t \rightarrow 0} \frac{o_{ij}(t)}{t} = 0$). Osserviamo che

$$p_{ij}(0) = P(X_0 = j | X_0 = i).$$

Quindi vale zero per $i \neq j$ ed uno altrimenti. Poniamo

$$\delta_{ij} = \begin{cases} 1 & \text{per } i = j \\ 0 & \text{per } i \neq j \end{cases}$$

Vale quindi $p_{ij}(0) = \delta_{ij}$. Indichiamo infine con a_{ij} i numeri $p'_{ij}(0)$. Con queste nuove notazioni abbiamo

$$p_{ij}(t) = \delta_{ij} + a_{ij} \cdot t + o(t).$$

Da qui in poi omettiamo gli indici i, j in $o_{ij}(t)$, visto che è uso indicare con $o(t)$ un qualsiasi infinitesimo di ordine superiore ad uno.

Queste semplici considerazioni portano a questa prima conclusione: la matrice

$$A = (a_{ij})$$

descrive $P(t)$ approssimativamente (ovvero a meno di $o(t)$) per t piccoli:

$$P(t) = I + A \cdot t + o(t)$$

dove I indica la matrice identica. La matrice A si chiama matrice dei *tassi* di transizione (o anche generatore infinitesimale). In generale, in matematica, si chiama *tasso* di qualche grandezza quella grandezza per unità di tempo, ovvero il rapporto tra quella grandezza ed il tempo corrispondente. Allora il tasso di transizione è la probabilità di transizione per unità di tempo: per $i \neq j$

$$a_{ij} = \lim_{t \rightarrow 0} \frac{p_{ij}(t)}{t}.$$

E' molto notevole il fatto che, grazie alla proprietà $P(t+s) = P(t)P(s)$, essa permetta di ricavare *esattamente* $P(t)$ per t qualsiasi. Infatti, preso un qualsiasi tempo t_0 , vale

$$\begin{aligned} \frac{dP}{dt}(t_0) &= \lim_{\varepsilon \rightarrow 0} \frac{P(t_0 + \varepsilon) - P(t_0)}{\varepsilon} = P(t_0) \lim_{\varepsilon \rightarrow 0} \frac{P(\varepsilon) - I}{\varepsilon} \\ &= P(t_0) \lim_{\varepsilon \rightarrow 0} \frac{A \cdot \varepsilon + o(\varepsilon)}{\varepsilon} = P(t_0)A \end{aligned}$$

e nello stesso modo, usando $\frac{P(t_0+\varepsilon)-P(t_0)}{\varepsilon} = \frac{P(\varepsilon)-I}{\varepsilon}P(t_0)$, si ottiene $\frac{dP}{dt}(t_0) = AP(t_0)$. Ma allora, data A , si può calcolare $P(t)$ (in linea di principio) risolvendo il sistema di equazioni lineari

$$\begin{cases} \frac{dP}{dt}(t) = P(t)A & \text{per } t \geq 0 \\ P(0) = I \end{cases}$$

oppure quello con l'equazione $\frac{dP}{dt}(t) = AP(t)$. Il calcolo ora svolto è rigoroso quando il numero di stati è finito, altrimenti richiederebbe delle precisazioni, su cui però sorvoliamo. Negli esempi specifici con infiniti stati cercheremo per lo meno ogni volta di controllare che le varie espressioni che troveremo abbiano senso (es. siano serie convergenti).

Si vedrà negli esempi che il calcolo di $P(t)$ a partire da A non è facile. E' però più semplice del calcolo di P^n a partire da P , nel caso discreto. Infatti, la risoluzione di un sistema di equazioni differenziali, per quanto sia un'operazione non banale, è più semplice da trattare analiticamente, cioè con formule esatte, dell'operazione algebrica di potenza n -esima di una matrice.

Segnaliamo che la condizione $\sum_j p_{ij}(t) = 1$ diventa $\sum_j (\delta_{ij} + a_{ij} \cdot t) = 1$ a meno di infinitesimi di ordine superiore. Essendo già $\sum_j \delta_{ij} = 1$, deve valere

$$\sum_j a_{ij} = 0.$$

1.3 Probabilità invarianti

La matrice A permette di cercare le probabilità invarianti (che è la cosa di maggior rilevanza pratica). Infatti, π è invariante se $\pi = \pi P(t)$ per ogni t . Ma allora, per t piccolo,

$$\pi = \pi (I + A \cdot t + o(t)) = \pi + \pi A \cdot t + \pi o(t)$$

ovvero $\pi A \cdot t + \pi o(t) = 0$. Dividendo per t e facendo il limite per $t \rightarrow 0$ otteniamo

$$\pi A = 0.$$

A causa di queste proprietà, A gioca il ruolo di P nel caso del tempo continuo.

Circa la ricerca di π dato A , osserviamo che vale anche qui il bilancio di flusso, nella seguente forma. Fissata l'attenzione su uno stato i , dobbiamo uguagliare la massa uscente con quella entrante. Per tempi molto piccoli la massa uscente è $\pi_i \sum_{j \neq i} p_{ij}(t)$, che approssimativamente è

$$\pi_i \sum_{j \neq i} a_{ij} \cdot t$$

($\delta_{ij} = 0$ per $j \neq i$). La massa entrante è $\sum_{j \neq i} \pi_j p_{ji}(t)$, che approssimativamente è

$$\sum_{j \neq i} \pi_j a_{ji} \cdot t.$$

L'approssimazione è tanto migliore quanto più t è piccolo. Ne discende

$$\pi_i \sum_{j \neq i} a_{ij} = \sum_{j \neq i} \pi_j a_{ji}$$

di facile lettura sul grafo (che desciveremo sotto). Nella ricerca di π , a queste equazioni andrà aggiunta come sempre la condizione $\sum_i \pi_i = 1$.

1.4 Legame tra A e le prescrizioni concrete di un problema. Grafo di una catena a tempo continuo

Le due frasi di questo titolo indicano due questioni ancora non chiarite: c'è un analogo dell'utilissimo grafo nel caso del tempo continuo? Dato un problema applicativo con le sue specifiche concrete, mentre a tempo discreto era

piuttosto naturale disegnare il grafo corrispondente, come possiamo ora determinare A ?

Il grafo nel caso continuo è simile al caso discreto, solo che in corrispondenza delle frecce bisogna ora scrivere i numeri a_{ij} al posto delle probabilità di transizione p_{ij} . Si noti che questi numeri possono essere maggiori di uno! Inoltre, non si usa scrivere i valori a_{ii} , in genere negativi, che si ottengono come complementare degli altri (vale $\sum_j a_{ij} = 0$).

Resta però il problema del significato applicativo dei numeri a_{ij} . Esse non sono probabilità di transizione ma *tassi*. La probabilità $p_{ij}(t)$ di ritrovarsi in j al tempo t essendo partiti da i all'istante iniziale, vale zero per $t = 0$, poi aumenta all'aumentare di t in quanto possono realizzarsi delle transizioni che portano il sistema in j . Il tasso di aumento è a_{ij} . Quindi a_{ij} fornisce un'informazione quantitativa sulla maggior o minor predisposizione del sistema ad effettuare transizioni da i verso j . Questa prima osservazione va tenuta presente, ma è ancora solo di tipo qualitativo.

C'è un'importantissima interpretazione dei coefficienti a_{ij} in termini di "orologi" aventi distribuzione esponenziale. Parliamone ora in modo un po' intuitivo senza dimostrare le affermazioni. Quando il sistema si trova nello stato i , attende un tempo aleatorio T_i prima di effettuare una transizione. Si può dimostrare che T_i è esponenziale, di parametro $\sum_{j \neq i} a_{ij}$. Questo fatto si interpreta nel seguente modo.

E' come se il sistema, quando è nello stato i , accendesse delle sveglie (orologi) che suonano dopo un tempo aleatorio esponenziale; una sveglia per ogni stato $j \neq i$ a cui può transire, con tempo di attesa T_{ij} esponenziale di parametro a_{ij} . Tutte le sveglie sono indipendenti. La prima che suona determina l'istante e lo stato a cui avviene la transizione (a quel punto si ricomincia da capo dal nuovo stato, ignorando le sveglie che ancora non hanno suonato). La prima che suona, suona al tempo $T_i = \min_j T_{ij}$, che per un fatto descritto nel capitolo sulle v.a. esponenziali, ha distribuzione esponenziale di parametro $\sum_{j \neq i} a_{ij}$.

Dalle cose dette, nasce il seguente modo di calcolare A a partire da prescrizioni pratiche di un problema applicativo: per ogni coppia di stati i e j dobbiamo chiederci il valore del parametro a_{ij} della v.a. esponenziale T_{ij} che descrive il tempo da attendere prima che si realizzi la transizione da i a j , in assenza della possibilità di effettuare transizioni diverse, ovvero come se il sistema avesse solo gli stati i e j isolati. Inoltre, vale $E[T_{ij}] = \frac{1}{a_{ij}}$, quindi se dal punto di vista applicativo si conosce il tempo medio di attesa prima

della transizione (questo di solito è il dato empirico), da esso si calcola a_{ij} .

Example 2 *Una macchina utensile può trovarsi in tre stati: funzionante (f), in attesa (a), rotta (r). Quando è funzionante, completa il suo lavoro dopo un tempo aleatorio $T_{f,a}$ di media 10 minuti. Però può anche rompersi, e ciò avviene (sempre quando sta funzionando) dopo un tempo aleatorio di media 10 ore. Per semplicità, supponiamo che se completa il lavoro, le vengano ripristinati certi fattori (olio ecc.) in modo che la sua propensione a rompersi riparta da zero quando verrà rimessa in funzione (senza accumulo dai funzionamenti precedenti). Allora $a_{f,a} = \frac{1}{10} \text{ min}^{-1}$, $a_{f,r} = \frac{1}{600} \text{ min}^{-1}$. Da funzionante, avviene la prima transizione (verso a oppure r), dopo un tempo aleatorio esponenziale di parametro*

$$a_{f,a} + a_{f,r} = \left(\frac{1}{10} + \frac{1}{600} \right) \text{ min}^{-1} = .10167 \text{ min}^{-1}.$$

Il tempo medio prima che avvenga una transizione dallo stato f è di

$$\frac{1}{.10167} = 9.8357$$

minuti.

1.5 Semigruppì e generatori infinitesimali

La lettura di questo paragrafo è opzionale: esso mostra fatti simili a quelli del paragrafo precedente partendo però da un punto di vista diverso.

Una famiglia di matrici $P(t)$ (non necessariamente di probabilità di transizione) si dice un semigruppò se valgono le condizioni

$$P(0) = I, \quad P(t+s) = P(t)P(s) \quad \text{per ogni } t, s \geq 0$$

dove I è la matrice identica. Il nome semigruppò è ispirato alla proprietà algebrica di queste matrici di trasformare l'operazione di somma $t+s$ in quella di prodotto $P(t+s) = P(t)P(s)$. Se invece di essere matrici fossero numeri reali, riconosceremmo la ben nota proprietà della funzione esponenziale. Ricordiamo infatti un teorema dell'analisi di funzioni di una variabile: se una funzione $f(t)$, definita per $t \geq 0$, soddisfa $f(0) = 1$ e $f(t+s) = f(t)f(s)$ per ogni $t, s \geq 0$, e se è continua, allora è la funzione esponenziale

$$f(t) = e^{\alpha t}$$

per un opportuno numero reale α . La dimostrazione non è troppo difficile se si conoscono già le funzioni esponenziale e logaritmo: basta introdurre la funzione ausiliaria $g(t) = \log f(t)$ (dopo alcune argomentazioni per assicurarsi che sia $f(t) > 0$), verificare che vale $g(0) = 0$, $g(t+s) = g(t) + g(s)$, e dimostrare (questo richiede un po' di lavoro ma è elementare) che una funzione $g(t)$ con queste proprietà, e continua, è necessariamente lineare: $g(t) = \alpha t$ per un opportuno α . Si conclude allora che vale $f(t) = e^{\alpha t}$.

Fatta questa premessa, forse non stupisce il fatto che lo stesso risultato vale per i semigruppì di matrici $P(t)$. Supponiamo per semplicità che lo spazio degli stati S sia finito. Se $P(t)$ è un semigruppò, le cui componenti $p_{ij}(t)$ sono funzioni continue, allora $P(t)$ è un esponenziale della forma

$$P(t) = e^{At}$$

dove A è una matrice. L'espressione e^{At} è la matrice definita dalla serie di Taylor

$$\sum_{k=0}^{\infty} \frac{(At)^k}{k!}.$$

Si può inoltre dimostrare che vale

$$A = \left. \frac{dP(t)}{dt} \right|_{t=0}$$

ovvero, noto $P(t)$, la matrice A si può calcolare prendendo la derivata di $P(t)$ per $t = 0$. Infine, vale l'equazione differenziale

$$\frac{dP(t)}{dt} = AP(t) = P(t)A \quad (1)$$

con la condizione iniziale

$$P(0) = I. \quad (2)$$

Questi fatti sono analoghi al caso della funzione esponenziale classica $f(t) = e^{\alpha t}$: per essa vale

$$\alpha = \left. \frac{df(t)}{dt} \right|_{t=0} \quad \text{e} \quad \frac{df(t)}{dt} = \alpha f(t) = f(t)\alpha.$$

Le dimostrazioni sono però più laboriose e le omettiamo. Nel caso in cui lo spazio degli stati non è finito valgono risultati analoghi, ma gli enunciati stessi diventano più complicati, per cui non li esprimiamo nella completa generalità.

Osserviamo, in relazione a quanto detto nel paragrafo precedente, che con questo approccio è sufficiente supporre che $P(t)$ sia continuo per $t = 0$, invece che derivabile. La derivabilità è una conseguenza della continuità e della proprietà di semigruppato.

2 Processo di nascita e morte a tempo continuo

Consideriamo la versione a tempo continuo del processo di nascita e morte visto in precedenza nell'ambito delle catene di Markov a tempo discreto.

Consideriamo quindi un processo di Markov $(X_t)_{t \geq 0}$ a valori nell'insieme S dei numeri interi non negativi. Dobbiamo specificare la matrice A . Prescriviamo una dinamica particolare, chiamata di nascita e morte. Se $j = i+1$, chiamiamo λ_i il coefficiente $a_{i,i+1}$, per ogni $i = 0, 1, \dots$ ed analogamente se $j = i-1$, chiamiamo μ_i il coefficiente $a_{i,i-1}$, per ogni $i = 1, 2, \dots$ In altre parole, scriviamo (fin qui è solo una questione di nuove notazioni)

$$\begin{aligned} p_{i,i+1}(t) &= \lambda_i t + o(t), & i = 0, 1, \dots \\ p_{i,i-1}(t) &= \mu_i t + o(t), & i = 1, 2, \dots \end{aligned}$$

Ora viene la vera ipotesi:

$$p_{ij}(t) = o(t) \text{ per ogni } j > i + 1 \text{ e per ogni } j < i - 1$$

sempre nell'ambito degli i, j ammessi come stati della catena. In altre parole, i salti di due stati hanno probabilità, per $t \rightarrow 0$, che è un infinitesimo di ordine superiore a quella dei salti di uno stato. Commenteremo più avanti questa scelta. Ne discende che

$$\begin{aligned} p_{i,i}(t) &= 1 - (\lambda_i + \mu_i)t + o(t), & i = 1, 2, \dots \\ p_{0,0}(t) &= 1 - \lambda_0 t + o(t). \end{aligned}$$

La matrice A è stata scelta. Quindi l'unica vera ipotesi particolare di questo modello è l'annullamento dei coefficienti a_{ij} per $j > i + 1$ e $j < i - 1$.

La matrice dei tassi di transizione ha la forma

$$A = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \mu_1 & -\lambda_1 - \mu_1 & \lambda_1 & 0 & \dots \\ 0 & \mu_2 & -\lambda_2 - \mu_2 & \lambda_2 & \dots \\ 0 & 0 & \mu_3 & -\lambda_3 - \mu_3 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

e l'equazione $\pi A = 0$ diventa

$$\begin{aligned} -\lambda_0 \pi_0 + \mu_1 \pi_1 &= 0 \\ \lambda_0 \pi_0 - (\lambda_1 + \mu_1) \pi_1 + \mu_2 \pi_2 &= 0 \\ \lambda_1 \pi_1 - (\lambda_2 + \mu_2) \pi_2 + \mu_3 \pi_3 &= 0 \\ &\dots \end{aligned}$$

Questo sistema è esattamente uguale a quello delle catene a tempo discreto!
Si ottiene quindi il seguente teorema.

Theorem 3 *Posto*

$$a_i = \frac{\lambda_0 \lambda_1 \dots \lambda_{i-1}}{\mu_1 \mu_2 \dots \mu_i},$$

se risulta

$$\sum_{i=0}^{\infty} a_i < \infty$$

allora la distribuzione invariante per il processo a tempo continuo di nascita e morte è

$$\pi_i = \frac{a_i}{\sum_{i=0}^{\infty} a_i}.$$

Se invece $\sum_{i=0}^{\infty} a_i = +\infty$, allora non ci sono distribuzioni invarianti.

Remark 4 *Commentiamo la condizione di annullamento dei coefficienti a_{ij} per $j > i + 1$ e $j < i - 1$. Ovviamente una prima giustificazione pragmatica per questa scelta è la semplicità, analogamente al fatto che a tempo discreto si prendevano uguali a zero le probabilità di transizione tra stati distanti due o più stati. In un certo senso, però, quella condizione aveva allora*

una maggior naturalezza anche geometrica, dovuta appunto alla scansione discreta del tempo, mentre qui, dovendo ragionare per tempi piccoli ma arbitrari, possono sorgere dei dubbi sulla naturalezza della condizione. Infatti, è ovvio che, preso un lasso di tempo $\tau > 0$ comunque piccolo ma fissato, in esso possono avvenire due transizioni, ad esempio da i ad $i + 1$ durante l'intervallo $[0, \frac{\tau}{2}]$ e da $i + 1$ a $i + 2$ durante l'intervallo $[\frac{\tau}{2}, \tau]$. Il processo, come avveniva a tempo discreto, continua a procedere attraverso salti unitari ($i \rightarrow i + 1 \rightarrow i + 2$), ma in un intervallo di tempo arbitrariamente piccolo ne può effettuare due (o un numero finito arbitrariamente grande). Quindi, chiedere ad esempio che $p_{i,i+2}(t) = o(t)$, può destare qualche perplessità. Una spiegazione convincente viene pensando che le transizioni avvengano dettate da "orologi sveglia", come spiegheremo in seguito.

Analisi della varianza (ANOVA) ad una via

Dicembre 2002

L'analisi della varianza (ANOVA) è una metodologia statistica che permette di valutare gli effetti su una variabile di interesse di *fattori di controllo* che possono assumere diversi livelli. Per esempio, gli effetti sulla durata media di un componente al variare della qualità del materiale con cui è prodotto, del tipo di metodologia di produzione adottata; del tipo di controllo di qualità sul prodotto finito, ecc. . .

Per motivi di tempo e di opportunità didattica descriveremo il metodo applicato al caso in cui si vogliono valutare gli effetti di un solo fattore di controllo. Ovvero tratteremo il caso della *analisi della varianza da una via*.

Referenze Per ulteriori informazioni o approfondimenti sull'analisi della varianza si consiglia di consultare il testo di P. Erto, *Probabilità e Statistica*, Mc Graw-Hill.

1 Analisi della varianza ad una via

Assumiamo quindi di aver misurato una certa variabile di interesse X per N valori diversi dell'unico fattore di controllo e ogni volta di aver effettuato M osservazioni. Assumiamo che le singole osservazioni siano statisticamente indipendenti e che quindi sia possibile modellarle come $N \cdot M$ variabili aleatorie indipendenti X_{ij} con $i = 1, \dots, N$ l'indice corrispondente al valore i -esimo del fattore di controllo e $j = 1, \dots, M$ l'indice associato alla j -esima osservazione (all'interno del gruppo di osservazioni corrispondenti allo stesso livello del fattore di controllo). È naturale pensare che le X_{ij} siano v.a. gaussiane di media μ_i e varianza σ^2 . La media dipende solo dal livello del fattore di controllo mentre la varianza, per semplicità, assumiamo che non dipenda dal fattore di controllo. Nelle applicazioni pratiche si dovrà, di caso in caso, valutare se sia possibile o meno assumere tale indipendenza. Si ricordi che il nostro obiettivo finale è estrarre dalle osservazioni gli effetti dei vari livelli del fattore di controllo sulla quantità misurata.

L'unica caratteristica che dipende dal livello del fattore è il valor medio μ_i delle osservazioni. Per valutare quindi un qualche effetto del cambiamento

del fattore di controllo sulla variabile misurata possiamo costruire un test statistico che tenti di rigettare l'ipotesi nulla

$$\mathcal{H}_0 = \text{“Tutti gli } \mu_i \text{ sono uguali”}$$

a favore dell'ipotesi

$$\mathcal{H}_1 = \text{“Almeno uno dei } \mu_i \text{ è diverso dagli altri”}.$$

Consideriamo quindi il seguente modello:

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

dove abbiamo espresso le v.a. X_{ij} come somme di quantità deterministiche μ e α_i e di quantità aleatorie ε_{ij} indipendenti con distribuzione gaussiana a media nulla e varianza σ^2 . Le v.a. ε_{ij} sono interpretabili come gli errori di misura o le variabilità aleatorie intrinseche alle singole osservazioni, mentre le quantità deterministiche μ, α_i sono legate alle μ_i dalla seguente semplice relazione:

$$\mu_i = \mu + \alpha_i$$

con

$$\mu = \frac{\mu_1 + \mu_2 + \dots + \mu_N}{N}$$

e di conseguenza

$$\alpha_1 + \alpha_2 + \dots + \alpha_N = 0.$$

In questo modo l'ipotesi nulla \mathcal{H}_0 è equivalente ad avere $\mu_i = \mu$ per ogni i , ovvero $\alpha_i = 0$ per ogni i mentre l'ipotesi \mathcal{H}_1 equivale al fatto che, almeno uno degli α_i è ragionevolmente diverso da zero.

Il nostro scopo è quindi quello di costruire una opportuna statistica e una regione di rigetto per \mathcal{H}_0 associata a tale statistica.

Consideriamo le seguenti quantità:

$$\bar{X} := \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M X_{ij}$$

$$\bar{X}_{i.} := \frac{1}{M} \sum_{j=1}^M X_{ij}$$

e

$$\Sigma_t := \sum_{i=1}^N \sum_{j=1}^M (X_{ij} - \bar{X})^2$$

Osserviamo che

$$\begin{aligned}
\Sigma_t &= \sum_{i=1}^N \sum_{j=1}^M (X_{ij} - \bar{X}_i + \bar{X}_i - \bar{X})^2 \\
&= \sum_{i=1}^N \sum_{j=1}^M [(\bar{X}_i - \bar{X})^2 + (X_{ij} - \bar{X}_i)^2 + 2(X_{ij} - \bar{X}_i)(\bar{X}_i - \bar{X})] \\
&= M \sum_{i=1}^N (\bar{X}_i - \bar{X})^2 + \sum_{i=1}^N \sum_{j=1}^M (X_{ij} - \bar{X}_i)^2 + \sum_{i=1}^N \sum_{j=1}^M 2(X_{ij} - \bar{X}_i)(\bar{X}_i - \bar{X}) \\
&= M \sum_{i=1}^N (\bar{X}_i - \bar{X})^2 + \sum_{i=1}^N \sum_{j=1}^M (X_{ij} - \bar{X}_i)^2
\end{aligned}$$

poichè:

$$\begin{aligned}
2 \sum_{i=1}^N \sum_{j=1}^M (X_{ij} - \bar{X}_i)(\bar{X}_i - \bar{X}) &= 2 \sum_{i=1}^N (\bar{X}_i - \bar{X}) \sum_{j=1}^M (X_{ij} - \bar{X}_i) \\
&= 2 \sum_{i=1}^N (\bar{X}_i - \bar{X}) \left[\sum_{j=1}^M X_{ij} - \sum_{j=1}^M \bar{X}_i \right] \\
&= 2 \sum_{i=1}^N (\bar{X}_i - \bar{X}) [M\bar{X}_i - M\bar{X}_i] \\
&= 0.
\end{aligned}$$

Quindi ponendo

$$\Sigma_f := \sum_{i=1}^N (\bar{X}_i - \bar{X})^2$$

ed

$$\Sigma_e := \sum_{i=1}^N \sum_{j=1}^M (X_{ij} - \bar{X}_i)^2$$

abbiamo che

$$\Sigma_t = M\Sigma_f + \Sigma_e \quad (1)$$

Un'interpretazione intuitiva di quest'ultima uguaglianza è la seguente: Σ_t può essere inteso come il momento d'inerzia di MN masse unitarie collocate nei punti X_{ij} (in una dimensione spaziale), \bar{X} è il baricentro di tutte le masse, mentre \bar{X}_i quello dei gruppi di M masse con lo stesso indice i . La formula dice allora che il momento di inerzia totale è costituito dalla somma Σ_e dei momenti di inerzia dei singoli gruppi rispetto ai loro baricentri (con i fissato) più il momento di inerzia dei gruppi rispetto al baricentro complessivo del sistema ($M\Sigma_f$).

Dal punto di vista statistico la formula (1) si interpreta invece così : la somma Σ_t degli scarti quadratici di tutte le osservazioni X_{ij} dalla loro media complessiva \bar{X} si può decomporre in due contributi: il primo $M\Sigma_f$ dipende solo dagli scarti delle medie parziali \bar{X}_i dalla media complessiva, mentre il secondo Σ_e dipende solo dagli scarti delle osservazioni dalle medie parziali del gruppo alle quali appartengono.

Se l'ipotesi \mathcal{H}_0 è vera, allora tutte le v.a. X_{ij} sono $N(\mu, \sigma^2)$ e

$$\mathbb{E} \Sigma_e = \sum_{i=1}^N \mathbb{E} \sum_{j=1}^M (X_{ij} - \bar{X}_i)^2 = N(M-1)\sigma^2$$

mentre

$$\mathbb{E} \Sigma_f = \sum_{i=1}^N \mathbb{E} (\bar{X}_i - \bar{X})^2 = \frac{(N-1)\sigma^2}{M} \quad (2)$$

(verificare!).

Queste due uguaglianze sono casi particolari del fatto che date n v.a. gaussiane Y_k $k = 1, \dots, n$ a media nulla e varianza σ_Y^2 , la v.a.

$$S_Y^2 := \sum_{k=1}^n (Y_k - \bar{Y})^2$$

(varianza empirica) con

$$\bar{Y} := \frac{1}{n} \sum_{k=1}^n Y_k$$

è tale che $\mathbb{E} S_Y^2 = (n-1)\sigma^2$.

Nella eq. (2) le v.a. gaussiane di cui calcoliamo la varianza empirica sono le \bar{X}_i che risultano a loro volta gaussiane con varianza $\text{Var}(\bar{X}_i) = \sigma^2/M$ (poiché sono le medie empiriche di M v.a. gaussiane indipendenti con varianza σ^2). Quindi definendo:

$$S_e^2 := \frac{1}{N(M-1)} \Sigma_e \quad \text{e} \quad S_f^2 := \frac{M}{N-1} \Sigma_f$$

si ha che

$$\mathbb{E} [S_e^2] = \mathbb{E} [S_f^2] = \sigma^2.$$

quindi sia S_e^2 che S_f^2 sono stimatori non distorti della varianza delle X_{ij} (questo se \mathcal{H}_0 è vera).

La v.a. Σ_e/σ^2 ha la distribuzione di un χ^2 con $N(M-1)$ gradi di libertà; possiamo infatti decomporla come segue:

$$\frac{\Sigma_e}{\sigma^2} = \sum_{i=1}^N \sum_{j=1}^M \frac{(X_{ij} - \bar{X}_i)^2}{\sigma^2} = \sum_{i=1}^N \frac{\Sigma_{e,i}}{\sigma^2}$$

con

$$\Sigma_{e,i} = \sum_{j=1}^M (X_{ij} - \bar{X}_i)^2$$

e notare che le $\Sigma_{e,i}/\sigma^2$, $i = 1, \dots, N$ hanno distribuzione χ^2 con $M - 1$ gradi di libertà e che sono indipendenti.

Invece $\Sigma_f/(\sigma^2/M) = M\Sigma_f/\sigma^2$ ha distribuzione χ^2 con $N - 1$ gradi di libertà. Un fatto notevole che non dimostreremo è che Σ_f e Σ_e sono indipendenti. Intuitivamente questo significa che le fluttuazioni intorno alla media empirica di ciascun gruppo sono indipendenti dalle fluttuazioni delle medie empiriche di ciascun gruppo rispetto a quella complessiva.

Definizione 1 Se S_n e S_m sono due v.a. χ^2 rispettivamente con n ed m gradi di libertà, allora la v.a.

$$Z_{n,m} = \frac{(S_n/n)}{(S_m/m)}$$

ha una distribuzione continua e viene chiamata Z (zeta) di Fisher con parametri n, m .

Quindi la v.a.

$$\hat{Z} = \frac{(M\Sigma_f/\sigma^2)/(N - 1)}{(\Sigma_e/\sigma^2)/[N(M - 1)]} = \frac{MN(M - 1)}{N - 1} \frac{\Sigma_f}{\Sigma_e}$$

ha distribuzione Z di Fisher con parametri $N - 1, N(M - 1)$.

Conoscendo i quantili della Z di Fisher possiamo costruire una regione di rigetto per \mathcal{H}_0 al livello di confidenza $1 - \alpha$.

Denotiamo con $z_\alpha(n, m)$ i quantili della Z di Fisher con parametri n, m al livello α , ovvero se $Z_{n,m}$ è una v.a. Z di Fisher con parametri n, m abbiamo che

$$\mathbb{P}(Z_{n,m} \leq z_\alpha(n, m)) = \alpha$$

Allora possiamo prendere come regione di rigetto per \hat{Z} il seguente intervallo (ad una coda) $R = [z_{1-\alpha}(N - 1, N(M - 1)), +\infty)$. Infatti

$$\begin{aligned} \mathbb{P}(\hat{Z} \in R) &= \mathbb{P}(\hat{Z} \geq z_{1-\alpha}(N - 1, N(M - 1))) \\ &= 1 - \mathbb{P}(\hat{Z} < z_{1-\alpha}(N - 1, N(M - 1))) = \alpha. \end{aligned}$$

La scelta di una regione di rigetto ad una coda è suggerita dalla seguente osservazione: se \mathcal{H}_0 è falsa (ovvero se alcune delle μ_i sono diverse dalle altre) allora la variabilità delle osservazioni tra i diversi gruppi tenderà a essere maggiore della variabilità delle misurazioni all'interno dei singoli gruppi e quindi la variabile \hat{Z} avrà la propensione ad assumere valori grandi. Vogliamo che il test sia sensibile a questo tipo di deviazioni dal comportamento statistico atteso per \hat{Z} .

Una volta determinata la regione di rigetto (a fissato livello di confidenza) possiamo calcolare il valore effettivo di \hat{Z} dai dati del nostro problema e confrontarlo con R . Se \hat{Z} cade all'interno di R allora possiamo rigettare \mathcal{H}_0 a favore di \mathcal{H}_1 . Abbiamo in questo caso abbastanza evidenze sperimentali per sostenere che i diversi valori del fattore di controllo influenzano in maniera deterministica la variabile che stiamo osservando. Occorreranno però analisi di natura diversa (e.g. una regressione) per quantificare l'entità di tali effetti o anche la dipendenza dettagliata dai vari livelli del fattore di controllo.

Esercizio 1 *Verificare che nel caso generale (ovvero senza porre $\alpha_i = 0$ per ogni i) si ha che*

$$\mathbb{E} [S_e^2] = \sigma^2$$

ed

$$\mathbb{E} [S_f^2] = \sigma^2 + \frac{M}{N-1} \sum_{i=1}^N \alpha_i^2$$

Questo implica che il valore atteso di S_f^2 è minimo quando \mathcal{H}_0 è vera.

2 Esempio di ANOVA ad una via

Per migliorare le caratteristiche di durata (a sollecitazioni dinamiche) di un componente aeronautico si è deciso di aggiungere delle fibre di carbonio al materiale composito con cui è prodotto. Si è tuttavia incerti su quale sia la migliore lunghezza delle fibre tra le quattro possibili per questa particolare applicazione. Si decide pertanto di effettuare una breve sperimentazione per appurare innanzitutto se la lunghezza delle fibre abbia o meno un effetto significativo sulla durata del componente. Nel caso che l'effetto esista, si vuole conoscere quale sia la lunghezza ottima da adottare.

Si sono pertanto approntati 4 gruppi, di 5 prototipi ciascuno, per ognuna della 4 alternative di lunghezza disponibili. Sottoponendoli a prove di durata si sono ottenuti i risultati riportati nella tabella sottostante.

livello	osservazioni	media e varianza	interv. di confidenza della media
1	8.99, 9.89, 9.61, 9.31, 9.35	9.43, 0.1146	[9.01, 9.85]
2	11.62, 11.40, 11.11, 10.80, 11.72	11.33, 0.1426	[10.86, 11.80]
3	11.85, 11.33, 11.84, 13.83, 11.05	11.998, 1.1866	[10.63, 13.33]
4	7.57, 11.27, 8.94, 7.89, 9.28	8.99, 2.12785	[7.18, 10.80]

Dai dati si ricava facilmente che $\Sigma_t = 45.717$ e che $\Sigma_f = 6.286$, poi abbiamo che $N = 4$ mentre $M = 5$. Dunque $\Sigma_e = \Sigma_t - M\Sigma_f = 45.716 - 31.430 = 14.287$, quindi

$$\hat{Z}^* = \frac{MN(M-1)}{N-1} \frac{\Sigma_f}{\Sigma_e} = \frac{5 \cdot 4 \cdot 4}{3} \frac{6.286}{14.287} = 11.733$$

Se ora calcoliamo la probabilità che una variabile Z di Fisher con parametri $3, 16$ ($N-1, N(M-1)$) abbia un valore più grande di \hat{Z}^* si ottiene

$$\mathbb{P}(Z_{3,16} > \hat{Z}^*) = 0.0003$$

da cui si può concludere che l'ipotesi \mathcal{H}_0 può essere rigettata fino al livello di significatività $1 - 0.0003 = 0.9997$. Viceversa se fissiamo un livello di significatività $1 - \alpha = 0.95$ standard, abbiamo che la regione di rigetto è $R = [3.29, +\infty)$ che porta lo stesso a rigettare \mathcal{H}_0 . Abbiamo quindi una forte evidenza che le μ_i sono diverse al variare del fattore di controllo. Stimando in maniera standard un intervallo di confidenza al 95% sulla media (riportato anch'esso sulla tabella) si può osservare che il terzo gruppo di osservazioni presentano una media sensibilmente maggiore degli altri gruppi. Sarà quindi preferibile utilizzare fibre corrispondenti al livello 3.

Osservazione Di seguito riportiamo la tabella della funzione di ripartizione della Z di Fisher, da cui si possono anche stimare i quantili. Si noti che nella tabella l'indice m corrisponder al primo parametro della Z (il numero di gradi di libertà della variabile χ^2 al numeratore) mentre l'indice n corrisponde al secondo parametro (il numero di gradi di libertà della variabile χ^2 al denominatore).

Sui metodi Monte Carlo

Novembre 2002

1 Somme di variabili indipendenti

Alcuni dei fenomeni più interessanti che si incontrano nello studio della teoria della probabilità sono legati alle proprietà delle somme di molti addendi indipendenti. Supponiamo che X_1, \dots, X_n, \dots sia una sequenza di v.a. indipendenti e consideriamo le somme parziali

$$S_n := X_1 + \dots + X_n$$

come mostreremo, specificando un pò meglio le condizioni necessarie, accade che nel limite $n \rightarrow \infty$ si ha

$$\frac{S_n}{n} \rightarrow \mathbb{E} X_1$$

cioè le somme parziali, opportunamente normalizzate, tendono al valore medio probabilistico dei singoli addendi. Il significato di questa frase sarà chiarito più sotto. Si noti come S_n/n non sia altro che la *media empirica* delle v.a. X_1, \dots, X_n . Si immagina infatti la seguente situazione: in un esperimento una certa misurazione viene ripetuta n volte ottenendo n valori: x_1, \dots, x_n (useremo le lettere minuscole per denotare il risultato di un particolare esperimento). Questi x_1, \dots, x_n sono dei numeri e non delle variabili aleatorie, però è possibile costruire un *modello probabilistico* della situazione reale di fronte alla quale ci troviamo. Nel modello esistono variabili aleatorie X_1, \dots, X_n che sono delle funzioni dallo spazio degli eventi Ω a valori in \mathbb{R} . Il nostro esperimento è un particolare evento $\omega \in \Omega$ e i valori da noi ottenuti x_1, \dots, x_n sono i valori che le v.a. X_1, \dots, X_n prendono in corrispondenza di questo determinato evento ω :

$$X_i(\omega) = x_i$$

per ogni $i = 1, \dots, n$.

La media empirica $\bar{X}_N := S_n/n$ è quella particolare v.a. che sull'evento corrispondente al nostro esperimento vale esattamente $(x_1 + \dots + x_n)/n$, appunto la media delle nostre osservazioni sperimentali.

Se eseguo molte misurazioni "indipendenti" e medio i risultati di queste osservazioni, il risultato che ottengo (e cioè \bar{X}_N) non sarà più un valore

casuale (o fluttuante, o comunque soggetto alla aleatorietà intrinseca delle singole misurazioni) ma sarà un valore ben determinato che nel modello probabilistico corrisponde al valor medio del singolo addendo X_i .

Teorema 1 (Legge (debole) dei grandi numeri) *Siano X_1, \dots, X_N variabili aleatorie indipendenti ed identicamente distribuite (i.i.d) di media $\mathbb{E} X_1 = \mu$ e varianza $\text{Var}(X_1) = \sigma^2$. Definendo la media empirica come*

$$\bar{X}_N := \frac{X_1 + \dots + X_N}{N} \quad (1)$$

si ha che al tendere di N all'infinito \bar{X}_N converge in probabilità a μ . Ovvero, fissati a piacere ε, δ positivi esiste $N_0(\varepsilon, \delta)$ tale che

$$\mathbb{P}(|\bar{X}_N - \mu| \geq \delta) \leq \varepsilon$$

per ogni $N > N_0(\varepsilon, \delta)$.

DIM. Calcoliamo la media e la varianza di \bar{X}_N :

$$\mathbb{E} \bar{X}_N = \frac{\mathbb{E} X_1 + \dots + \mathbb{E} X_N}{N} = \mu$$

mentre

$$\begin{aligned} \text{Var}(\bar{X}_N) &= \mathbb{E} (\bar{X}_N - \mu)^2 = N^{-2} \mathbb{E} \left[\sum_{i=1}^N (X_i - \mu) \right]^2 \\ &= N^{-2} \sum_{i=1}^N \mathbb{E} [(X_i - \mu)]^2 = \frac{\sigma^2}{N} \end{aligned}$$

Quindi la media delle \bar{X}_N non varia con N mentre la varianza diventa sempre più piccola. Quindi intuitivamente è chiaro che la probabilità che \bar{X}_N prenda valori lontani dalla media deve diventare piccola per N grandi (si ricordi che la varianza misura la “dispersione” di una variabile aleatoria attorno alla sua media). Per formalizzare questa intuizione abbiamo bisogno della seguente disuguaglianza (di *Chebichev*): per ogni v.a. Z con varianza finita vale che

$$\mathbb{P}(|Z - \mathbb{E} Z| \geq a) \leq \frac{\text{Var}(Z)}{a^2} \quad (2)$$

che dimosteremo nel seguito. Applicando la (2) alla v.a. \bar{X}_N si ottiene:

$$\mathbb{P}(|\bar{X}_N - \mu| \geq a) \leq \frac{\sigma^2}{Na^2}$$

e scegliendo $a = \delta$ e $N_0(\varepsilon, \delta)$ in modo tale che $\sigma^2 \leq \varepsilon \delta^2 N_0(\varepsilon, \delta)$ si ottiene

$$\mathbb{P}(|\bar{X}_N - \mu| \geq \delta) \leq \frac{\sigma^2}{N\delta^2} \leq \frac{\sigma^2}{N_0(\varepsilon, \delta)\delta^2} \leq \varepsilon$$

per ogni $N \geq N_0(\varepsilon, \delta)$ che prova la tesi. \square

Lemma 1 (Disuguaglianza di Chebishev) Per ogni v.a. Z con varianza finita, vale che

$$\mathbb{P}(|Z - \mathbb{E} Z| \geq a) \leq \frac{\text{Var}(Z)}{a^2}. \quad (3)$$

DIM. Sia $F(z)$ la seguente funzione

$$F(z) = \begin{cases} 1 & \text{se } |z - \mathbb{E} Z| \geq a; \\ 0 & \text{se } |z - \mathbb{E} Z| < a. \end{cases}$$

allora sarà vero che, per ogni possibile valore di Z ,

$$(Z - \mathbb{E} Z)^2 \geq (Z - \mathbb{E} Z)^2 F(Z)$$

dato che il membro destro si ottiene dal sinistro moltiplicando per 0 od 1 a seconda del valore di Z . Poi, possiamo anche scrivere che

$$(Z - \mathbb{E} Z)^2 F(Z) \geq a^2 F(Z)$$

poichè quando l'espressione $(Z - \mathbb{E} Z)^2$ al primo membro è minore di a^2 la funzione $F(Z)$ vale zero e quindi entrambe i membri si annullano, mentre se $(Z - \mathbb{E} Z)^2$ è maggiore di a^2 allora la disuguaglianza è banalmente verificata. Si ottiene quindi

$$(Z - \mathbb{E} Z)^2 \geq a^2 F(Z)$$

e prendendo il valore medio di entrambe i membri

$$\text{Var}(Z) = \mathbb{E}(Z - \mathbb{E} Z)^2 \geq a^2 \mathbb{E} F(Z) = a^2 \mathbb{P}(|z - \mathbb{E} Z| \geq a)$$

che è la tesi che volevamo dimostrare. \square

Nel seguito avremo bisogno di un risultato più forte di quello ottenuto nel teorema precedente (legge debole dei grandi numeri). In particolare vorremo che, fissato l'esperimento ω , la sequenza di medie empiriche $\overline{X}_n(\omega)$ (ora sono numeri e non v.a.) converga (nel senso della convergenza delle serie numeriche) ad $\mathbb{E} X_1$. Questo risultato è garantito (nelle ipotesi che abbiamo già formulato sopra) dal seguente teorema che non dimostriamo:

Teorema 2 (Legge forte dei grandi numeri) Nelle ipotesi del teorema precedente vale che, esiste $A \subset \Omega$ tale che $\mathbb{P}(A) = 1$ e che

$$\lim_{n \rightarrow \infty} \overline{X}_n(\omega) = \mathbb{E} X_1$$

per ogni $\omega \in A$.

Si noti che il teorema non può escludere che esistano casi particolari in cui la convergenza non ha luogo. Ma assicura se l'insieme $\Omega \setminus A$ di questi casi particolari ha probabilità nulla e quindi in un certo senso è ininfluenza.

Il prossimo risultato caratterizza il modo in cui le medie empiriche si avvicinano asintoticamente al valore limite.

Teorema 3 (Teorema del Limite Centrale (TLC)) Sia X_1, \dots, X_n, \dots una sequenza i.i.d. di v.a. con media μ e varianza σ^2 , la v.a.

$$Z_n = \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$$

nel limite $n \rightarrow \infty$ converge in legge a una v.a. Gaussiana standard.

DIM. Per comodità, introduciamo le v.a.,

$$Y_i := \frac{X_i - \mu}{\sigma}$$

aventi media nulla e varianza unitaria (verificare), si noti che la sequenza Y_1, \dots, Y_n, \dots è ancora i.i.d.. In questo modo $Z_n = \sqrt{n} \bar{Y}_n$. Calcoliamo la funzione generatrice dei momenti di Z_n :

$$\Phi_{Z_n}(\theta) := \mathbb{E} \left[e^{\theta Z_n} \right] = \prod_{i=1}^n \mathbb{E} \left[e^{\theta Y_i / \sqrt{n}} \right] = \mathbb{E} \left[e^{\theta Y_1 / \sqrt{n}} \right]^n = [\Phi_{Y_1}(\theta / \sqrt{n})]^n$$

dove $\Phi_{Y_1}(\theta) := \mathbb{E} e^{\theta Y_1}$ è la funzione generatrice dei momenti di Y_1 e dove si è utilizzata, in maniera cruciale, l'indipendenza e l'uguaglianza delle distribuzioni delle Y_i .

Quando n è grande possiamo espandere $\Phi_{Y_1}(\theta / \sqrt{n})$ in serie di Taylor:

$$\Phi_{Y_1}(\theta / \sqrt{n}) = \Phi_{Y_1}(0) + \Phi'_{Y_1}(0) \frac{\theta}{\sqrt{n}} + \frac{1}{2} \Phi''_{Y_1}(0) \left(\frac{\theta}{\sqrt{n}} \right)^2 + O \left(\left(\frac{\theta}{\sqrt{n}} \right)^3 \right)$$

e ricordando che $\Phi_{Y_1}(0) = 1$ (sempre), $\Phi'_{Y_1}(0) = \mathbb{E} Y_1 = 0$, $\Phi''_{Y_1}(0) = \mathbb{E} Y_1^2 = \text{Var}(Y_1) = 1$, si ha

$$\begin{aligned} \Phi_{Z_n}(\theta) &= \left[1 + \frac{\theta^2}{2n} + O \left(\theta^2 / n^{3/2} \right) \right]^n = \exp \left\{ n \log \left[1 + \frac{\theta^2}{2n} + O \left(\theta^3 / n^{3/2} \right) \right] \right\} \\ &= \exp \left[\frac{\theta^2}{2} + O \left(\theta^3 / n^{1/2} \right) \right] \end{aligned}$$

da cui si ottiene che

$$\lim_{n \rightarrow \infty} \Phi_{Z_n}(\theta) = \exp \left(\frac{\theta^2}{2} \right)$$

che è proprio la funzione generatrice della Gaussiana standard. Quindi poichè le f.g. di Z_n convergono alla f.g. di una Gaussiana standard possiamo affermare che la distribuzione di Z_n converge, al tendere di n all'infinito, alla distribuzione di una Gaussiana standard. \square

Una conseguenza importante del TLC è la possibilità di approssimare la distribuzione (opportunamente normalizzata) di una somma di v.a. indipendenti con quella gaussiana, la cui forma analitica è ben nota. In particolare,

se X_1, \dots, X_n sono v.a. i.i.d e $Z \sim N(0, 1)$ allora per n sufficientemente grande (di solito $10 \div 50$ è abbastanza) si ha che

$$\begin{aligned} \mathbb{P}(a \leq \bar{X}_n \leq b) &= \mathbb{P}\left(\sqrt{n}\frac{a-\mu}{\sigma} \leq \sqrt{n}\frac{\bar{X}_n-\mu}{\sigma} \leq \sqrt{n}\frac{b-\mu}{\sigma}\right) \\ &\approx \mathbb{P}\left(\sqrt{n}\frac{a-\mu}{\sigma} \leq Z \leq \sqrt{n}\frac{b-\mu}{\sigma}\right) \\ &= F\left(\sqrt{n}\frac{b-\mu}{\sigma}\right) - F\left(\sqrt{n}\frac{a-\mu}{\sigma}\right) \end{aligned}$$

dove $F(x)$ è la funzione di ripartizione di una Gaussiana standard

$$F(x) := \mathbb{P}(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du$$

Intuitivamente, il teorema del limite centrale, asserisce che, per n sufficientemente grande

$$\bar{X}_n \approx \mathbb{E} X_1 + n^{-1/2} Z$$

cioè che \bar{X}_n differisce dal suo valore limite $\mathbb{E} X_1$ per uno scarto che va a zero come $n^{-1/2}$ e che ha una distribuzione gaussiana.

2 Stima del valore medio

I risultati della sezione precedente suggeriscono un metodo per stimare il valore medio $\mu = \mathbb{E} X$ di una variabile aleatoria X (si suppone sempre che $\mathbb{E}|X| < \infty$, cioè che abbia senso parlare del valore medio di X). Disponendo di un campione x_1, \dots, x_n di rango n estratto da una popolazione di v.a. X_1, \dots, X_n tutte distribuite come X , possiamo stimare μ come

$$\mu_{\text{stima}} = \frac{x_1 + \dots + x_n}{n}.$$

La legge forte dei grandi numeri ci assicura che $\mu_{\text{stima}} \rightarrow \mu$ nel limite in cui il campione ha rango arbitrariamente grande ($n \rightarrow \infty$). Nelle applicazioni pratiche il rango del campione è finito e sorge il problema di caratterizzare l'errore che si commette approssimando μ con μ_{stima} . Formalizzando il problema consideriamo la v.a. corrispondente a μ_{stima} , ovvero la media empirica:

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}.$$

Abbiamo il seguente risultato:

Teorema 4 Sia $\sigma^2 = \text{Var}(X)$ e $\mu = \mathbb{E} X$, allora per ogni livello di confidenza $1 - \alpha$ fissato, con $0 < \alpha < 1$ si ha che, per $n \rightarrow \infty$,

$$\mathbb{P}\left(|\bar{X}_n - \mu| \leq \frac{\sigma}{\sqrt{n}} \phi_{1-\alpha/2}\right) \approx 1 - \alpha. \quad (4)$$

dove ϕ_α denota il quantile della legge gaussiana standard al livello α .

DIM. Osserviamo che se G è una v.a. gaussiana standard ($N(0,1)$) si ha che

$$\mathbb{P}(|G| \leq \phi_{1-\alpha/2}) = 1 - \alpha \quad (5)$$

dove ricordiamo che ϕ_α è definito come l'unico numero tale che

$$\mathbb{P}(G \leq \phi_\alpha) = \alpha.$$

Il teorema del limite centrale garantisce che

$$Z = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$$

tende ad avere distribuzione gaussiana standard nel limite $n \rightarrow \infty$, applicando quindi la (5) a Z (in maniera approssimata) si ottiene

$$1 - \alpha \approx \mathbb{P}(|Z| \leq \phi_{1-\alpha/2}) = \mathbb{P}\left(\frac{\sqrt{n}}{\sigma} |\bar{X}_n - \mu| \leq \phi_{1-\alpha/2}\right) \quad (6)$$

che è ciò che volevamo dimostrare. \square

Fissando un certo livello di confidenza $1 - \alpha$ (per es. 95% oppure 99%) possiamo affermare che la media “vera” μ è approssimata dalla stima empirica \bar{X}_n con un errore che al più è pari a $\phi_{1-\alpha/2}\sigma/\sqrt{n}$. Usualmente $\phi_{1-\alpha/2} \approx 2$. Se n aumenta l'errore che riesco a garantire, a parità di livello di confidenza, è minore. Viceversa, se ammetto un livello di confidenza minore, l'errore parimenti diminuisce (si noti però che in questo caso pago la maggior *precisione* del risultato con una sua minore *affidabilità*).

Una osservazione: Il risultato che diamo non è, ne può essere, certo. Utilizzando un metodo statistico non è possibile ottenere risultati certi, se non nel limite in cui il rango del campione considerato diventa infinito. Con un numero finito di osservazioni possiamo fare delle affermazioni che sono plausibili e il cui grado di plausibilità è misurato dal livello di confidenza.

Per stimare l'errore nel modo appena esposto è necessario conoscere σ ovvero la varianza della v.a. X . Usualmente questo non è possibile nelle applicazioni e sorge il problema di stimare la varianza utilizzando il campione a nostra disposizione.

Uno stimatore ben noto per la varianza è il seguente:

$$S_n^2 := \frac{1}{n-1} \left[\sum_{i=1}^n (X_i - \bar{X}_n)^2 \right] \quad (7)$$

che gode della proprietà (dimostrarlo):

$$\mathbb{E} S_n^2 = \text{Var}(X) = \sigma^2.$$

e il cui valore dipende solamente dal campione X_1, \dots, X_n .

Volendo sostituire questo stimatore al valore esatto di σ^2 il risultato enunciato nel teorema precedente viene sostituito dal seguente:

Teorema 5 Per ogni livello di confidenza $1 - \alpha$ fissato, con $0 < \alpha < 1$ si ha che, per $n \rightarrow \infty$,

$$\mathbb{P} \left(|\bar{X}_n - \mu| \leq \frac{S_n}{\sqrt{n}} t_{1-\alpha/2}(n-1) \right) \approx 1 - \alpha. \quad (8)$$

dove $t_\alpha(m)$ denota il quantile della legge T di Student con m gradi di libertà al livello α .

DIM. Consideriamo

$$Z = \sqrt{n} \frac{\bar{X}_n - \mu}{S}.$$

Un risultato che non dimostriamo suggerisce che Z è approssimativamente distribuita come una v.a. T di Student con $n-1$ gradi di libertà. Procedendo come nella dimostrazione del teorema precedente si ottiene l'enunciato. \square

Riassumendo: per ottenere una stima di μ dato un campione x_1, \dots, x_n si devono calcolare la media empirica \bar{x}_n e la *varianza empirica* s_n^2 del campione utilizzando le formule:

$$\bar{x}_n = \frac{x_1 + \dots + x_n}{n}$$

e

$$s_n^2 = \frac{1}{n-1} \left[\sum_{i=1}^n (x_i - \bar{x}_n)^2 \right].$$

Fissando un livello di confidenza $1 - \alpha$ si può affermare che la media vera μ di X giace nel seguente intervallo

$$\bar{x}_n - \frac{s_n}{\sqrt{n}} t_{1-\alpha/2}(n-1) \leq \mu \leq \bar{x}_n + \frac{s_n}{\sqrt{n}} t_{1-\alpha/2}(n-1)$$

scrittura che si può abbreviare in

$$\mu = \bar{x}_n \pm \frac{s_n}{\sqrt{n}} t_{1-\alpha/2}(n-1).$$

3 Generazione di numeri casuali

I metodi MC si basano sulla possibilità di disporre di un campione di una v.a. la cui distribuzione sia nota. Il modo più semplice ed efficiente di ottenere questi campioni è quello di utilizzare un *generatore di numeri casuali* che non è altro che un particolare programma di calcolatore (o una funzione all'interno di un linguaggio di programmazione o di un pacchetto applicativo come ad esempio MathCad, MathLab, Excel, ecc. . .). In realtà il calcolatore non è in grado di generare una sequenza di numeri "veramente" casuale. Ciò che ogni generatore di numeri casuali (o più propriamente *pseudo-casuali*) fa è applicare iterativamente una trasformazione deterministica R partendo da un valore iniziale x_0 (*seme*) per ottenere una sequenza $\{x_n\}$ tale che

$x_n = R(x_{n-1})$. Dal punto di vista statistico questa sequenza si comporta come una successione di campioni di v.a. indipendenti, tutti con la stessa distribuzione che usualmente è quella di una v.a. uniforme sull'intervallo $[0, 1]$. Un generatore è più o meno buono a seconda del fatto che le sequenze che è in grado di generare si comportano effettivamente come ci si aspetta da un campione casuale, oppure mostrano delle regolarità che denunciano la natura deterministica della procedura che le ha generate. Usualmente i generatori presenti nei pacchetti applicativi hanno una qualità media.

Assumiamo che si disponga di un generatore di numeri casuali (GNC) che fornisce un campione casuale di rango arbitrario di una v.a. uniforme in $[0, 1]$. Nelle applicazioni avremmo bisogno di campioni estratti da popolazioni avente distribuzioni diverse da quella uniforme in $[0, 1]$. Sorge quindi la necessità di ottenere delle v.a. con distribuzione prefissata date delle v.a. indipendenti e uniformi in $[0, 1]$.

Ci viene in aiuto il seguente risultato

Lemma 2 *Sia X una v.a. continua che ha densità $p_X(x)$ e funzione di ripartizione $F_X(x)$, allora, se U è una v.a. uniforme in $[0, 1]$ si ha che, la v.a. Y definita come*

$$Y = F_X^{-1}(U)$$

ha la stessa distribuzione di X .

DIM. Ricordiamo che $F_X(t) = \mathbb{P}(X \leq t)$ e che $F_X'(t) = p_X(t) \geq 0$. Se $A = \{t : p_X(t) > 0\}$ è l'insieme dei valori che X può prendere, allora F_X è strettamente crescente sull'insieme A e quindi invertibile. Consideriamo quindi $F_X : A \rightarrow [0, 1]$ e sia $F_X^{-1} : [0, 1] \rightarrow A$ la sua inversa, cioè la funzione tale che

$$F_X^{-1}(F_X(t)) = t, \quad \text{per ogni } t \in A$$

e

$$F_X(F_X^{-1}(u)) = u, \quad \text{per ogni } u \in [0, 1].$$

Si noti che se $a, b \in A$, $a \leq b$ se e solo se $F_X(a) \leq F_X(b)$. Abbiamo quindi che, se $t \in A$,

$$\begin{aligned} F_Y(t) &= \mathbb{P}(Y \leq t) = \mathbb{P}(F_X^{-1}(U) \leq t) = \mathbb{P}(F_X(F_X^{-1}(U)) \leq F_X(t)) \\ &= \mathbb{P}(U \leq F_X(t)) = F_X(t) \end{aligned}$$

dove $F_U(t)$ è la funzione di ripartizione di una variabile uniforme in $[0, 1]$, cioè:

$$F_U(t) = \begin{cases} 0 & \text{se } t < 0 \\ t & \text{se } 0 \leq t \leq 1 \\ 1 & \text{se } t > 1 \end{cases}$$

Essendo quindi $F_Y(t) = F_X(t)$ per ogni $t \in A$ ed entrambe le funzioni essendo crescenti e continue, si ha che $F_Y(t) = F_X(t)$ per ogni $t \in \mathbb{R}$. \square

Questo lemma ci permette di costruire una v.a. Y avente distribuzione data (specificata da $F(t)$) ogni qualvolta fossimo in grado di ottenere in maniera esplicita l'inversa della funzione di ripartizione $F(t)$.

Esempio 1 Sia F la funzione di ripartizione di una v.a. esponenziale di parametro λ :

$$F(t) = \begin{cases} 0 & \text{se } t < 0 \\ 1 - e^{-\lambda t} & \text{se } t \geq 0. \end{cases}$$

Abbiamo che $A = \mathbb{R}_+$ (i numeri reali positivi). Calcoliamo l'inversa di $F(t) : A \rightarrow [0, 1]$. Dato $t \geq 0$ sia $h = F(t) \in [0, 1]$, allora

$$h = 1 - e^{-\lambda t}$$

che implica

$$t = -\lambda^{-1} \log(1 - h)$$

quindi $F^{-1}(h) = -\lambda^{-1} \log(1 - h)$. Applicando il lemma precedente abbiamo che, se U è una v.a. uniforme in $[0, 1]$, allora

$$Y = F^{-1}(U) = -\lambda^{-1} \log(1 - U)$$

è una v.a. esponenziale di parametro λ .

Il problema della generazione di v.a. discrete è invece risolto con l'aiuto del seguente lemma:

Lemma 3 Sia X una v.a. discreta che prende valori x_i , $i = 1, \dots, n$ con probabilità $p_i = \mathbb{P}(X = x_i) > 0$ tali che $\sum_{i=1}^n p_i = 1$. Definiamo le quantità q_i , $i = 0, \dots, n$ tali che

$$q_i = \sum_{j=1}^i p_j, \quad \text{per } i = 1, \dots, n$$

ponendo $q_0 = 0$. Sia U una v.a. uniforme in $[0, 1]$. Sia $F : [0, 1] \rightarrow \mathbb{R}$ la funzione

$$F(u) := \begin{cases} x_1 & \text{se } u \in [q_0, q_1) = [0, q_1), \\ x_2 & \text{se } u \in [q_1, q_2), \\ \vdots & \\ x_n & \text{se } u \in [q_{n-1}, q_n) = [q_{n-1}, 1]; \end{cases}$$

allora la v.a. $Y = F(U)$ ha la stessa distribuzione di X .

DIM. Calcoliamo $\mathbb{P}(Y = x_i)$:

$$\mathbb{P}(Y = x_i) = \mathbb{P}(F(U) = x_i) = \mathbb{P}(U \in [q_{i-1}, q_i)) = (q_i - q_{i-1}) = p_i$$

mentre $\mathbb{P}(Y = x) = 0$ se $x \neq x_i$ per ogni $i = 1, \dots, n$. Quindi Y ha la stessa distribuzione di X . \square

Esempio 2 Supponiamo di voler generare una v.a. di Bernoulli con parametro $1/3$, i.e. una v.a. X tale che

$$\mathbb{P}(X = 0) = 2/3, \quad \mathbb{P}(X = 1) = 1/3, \quad \mathbb{P}(X \notin \{0, 1\}) = 0.$$

Consideriamo la funzione $F : [0, 1] \rightarrow \mathbb{R}$ tale che

$$F(u) = \begin{cases} 0 & \text{se } u \in [0, 2/3) \\ 1 & \text{se } u \in [2/3, 1] \end{cases}$$

allora, per il lemma precedente, $X = F(U)$ ha la distribuzione cercata.

Il lemma 2 non è praticamente applicabile quando non si conosce una forma analitica della funzione di ripartizione della v.a. di interesse, oppure quando la funzione di ripartizione non può essere invertita. In questi casi occorrono vie alternative per generare dei numeri casuali con la distribuzione voluta. Un caso importante per il quale il lemma 2 risulta inapplicabile è quello della legge Gaussiana. In questo caso la funzione di ripartizione Φ è data dall'integrale

$$\Phi(t) = \int_{-\infty}^t \frac{e^{-u^2/2}}{\sqrt{2\pi}} du$$

di cui non si ha a disposizione una forma analitica e quindi non si conosce neppure la funzione inversa Φ^{-1} . Una tecnica per generare v.a. Gaussiane è descritta nella seguente proposizione:

Lemma 4 Siano date due v.a. R, Θ indipendenti e tali che R sia una v.a. Esponenziale di parametro 1 e Θ una v.a. uniforme in $[0, 2\pi]$. Allora la v.a.

$$X := \sqrt{2R} \cos(\Theta)$$

ha una distribuzione Gaussiana standard.

DIM. È sufficiente dimostrare che

$$\mathbb{P}(X \in A) \stackrel{?}{=} \int_A \frac{e^{-u^2/2}}{\sqrt{2\pi}} du$$

per ogni intervallo $A \subset \mathbb{R}$. Sia $\mathbb{1}_A : \mathbb{R} \rightarrow \mathbb{R}$ la funzione caratteristica dell'insieme A , ovvero

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{se } x \in A, \\ 0 & \text{altrimenti.} \end{cases}$$

Allora

$$\begin{aligned}
\mathbb{P}(X \in A) &= \mathbb{E}[\mathbb{1}_A(X)] = \mathbb{E}\left[\mathbb{1}_A\left(\sqrt{2R}\cos(\Theta)\right)\right] \\
&= \int_0^\infty dr \int_0^{2\pi} d\theta \mathbb{1}_A\left(\sqrt{2r}\cos(\theta)\right) \frac{e^{-r}}{2\pi} \\
&= \int_0^\infty \rho d\rho \int_0^{2\pi} d\theta \mathbb{1}_A(\rho \cos(\theta)) \frac{e^{-\rho^2/2}}{2\pi} \quad (\text{ponendo } \rho^2 = 2r) \\
&= \iint_{\mathbb{R}^2} dx dy \mathbb{1}_A(x) \frac{e^{-(x^2+y^2)/2}}{2\pi} \quad (\text{coord. polari} \rightarrow \text{coord. cartesiane}) \\
&= \int_{\mathbb{R}} dx \mathbb{1}_A(x) \frac{e^{-x^2/2}}{\sqrt{2\pi}} \int_{\mathbb{R}} dy \frac{e^{-y^2/2}}{\sqrt{2\pi}} \\
&= \int_{\mathbb{R}} dx \mathbb{1}_A(x) \frac{e^{-x^2/2}}{\sqrt{2\pi}} = \int_A \frac{e^{-x^2/2}}{\sqrt{2\pi}}
\end{aligned}$$

dove abbiamo fatto uso della trasformazione tra coordinate cartesiane e polari e dell'identità:

$$\int_{\mathbb{R}} dy \frac{e^{-y^2/2}}{\sqrt{2\pi}} = 1.$$

Abbiamo quindi verificato che la v.a. che abbiamo costruito ha la distribuzione gaussiana standard. \square

4 Stima MC di una funzione di distribuzione

Alcuni problemi applicativi in cui la metodologia MC è largamente applicabile sono quelli in cui è necessario valutare la distribuzione di probabilità di una v.a. per la quale non si dispone di forme analitiche chiuse o che è il risultato di un calcolo complesso. Per esempio: supponiamo di avere una v.a. X di cui conosciamo la distribuzione (e per la quale sappiamo generare dei campioni casuali), e una funzione $F : \mathbb{R} \rightarrow \mathbb{R}$ che non conosciamo nei dettagli o che risulta essere molto complessa. Il nostro scopo è valutare la distribuzione di probabilità della v.a. $Y = F(X)$. Un caso concreto potrebbe essere quello in cui X rappresenta la velocità del vento in un particolare giorno e F la funzione che associa alla velocità del vento la concentrazione di inquinanti nell'atmosfera nelle vicinanze di un'industria. In questo caso un complesso programma di simulazioni potrebbe calcolare F dato X senza che noi possiamo essere in grado di esprimere in forma semplice la dipendenza di $Y = F(X)$ da X .

Valutare la distribuzione di Y è equivalente a calcolare la funzione di ripartizione:

$$F_Y(t) := \mathbb{P}(Y \leq t) = \mathbb{P}(F(X) \leq t)$$

Consideriamo la seguente funzione $H_t : \mathbb{R} \rightarrow [0, 1]$:

$$H_t(x) := \begin{cases} 1 & \text{se } x \leq t, \\ 0 & \text{altrimenti.} \end{cases}$$

Utilizzando H_t riscriviamo la funzione di ripartizione di Y come il seguente valor medio

$$F_t(t) = \mathbb{P}(F(X) \leq t) = \mathbb{E}[H_t(F(X))]$$

abbiamo quindi ricondotto il problema di valutare la funzione di ripartizione a quello di stimare dei valori medi. Possiamo quindi adottare la metodologia MC che abbiamo esposto in precedenza. Dato un campione di rango N di X : X_1, \dots, X_N introduciamo la stima MC di F_Y :

$$\hat{F}_Y(t) := \frac{1}{N} \sum_{i=1}^N H_t(F(X_i)) = \frac{1}{N} \sum_{i=1}^N H_t(Y_i)$$

dove $Y_i = F(X_i)$. Consideriamo la lista di valori Y_1, Y_2, \dots, Y_N e riscriviamoli chiamando \hat{Y}_1 il più piccolo, \hat{Y}_2 il secondo in ordine di grandezza, e così via fino a \hat{Y}_N che quindi corrisponderà al massimo degli Y_i . Abbiamo quindi che

$$\hat{Y}_1 \leq \hat{Y}_2 \leq \dots \leq \hat{Y}_{N-1} \leq \hat{Y}_N.$$

Assumendo che non succeda mai che $\hat{Y}_i = \hat{Y}_{i+1}$ (cioè che due tra i vari valori Y_j siano uguali) si può dare una semplice descrizione della funzione $\hat{F}_Y(t)$ (vista come funzione di t una volta estratto il campione): questa funzione vale 0 per $t \rightarrow -\infty$ e ha un salto di ampiezza $1/N$ ogni volta che $t = \hat{Y}_i$ per qualche valore di i . Più in particolare possiamo dire che

$$\hat{F}_Y(t) = \begin{cases} 0 & \text{se } t < \hat{Y}_1; \\ k/N & \text{se } \hat{Y}_k \leq t < \hat{Y}_{k+1}; \\ 1 & \text{se } \hat{Y}_N \leq t; \end{cases}$$

Per i risultati che abbiamo esposto in precedenza, quando $N \rightarrow \infty$ si ha che $\hat{F}_Y(t) \rightarrow F_Y(t)$, quindi più N è grande più la nostra stima della funzione di ripartizione di Y sarà buona.

Una volta stimata la funzione di ripartizione possiamo anche ottenere stime dei quantili di Y dalla seguente considerazione. Supponiamo di voler conoscere il quantile a livello α di Y , ovvero il numero y_α tale che:

$$\mathbb{P}(Y \leq y_\alpha) = F_Y(y_\alpha) = \alpha.$$

Possedendo una stima \hat{F}_Y di F_Y , una stima \hat{y}_α del quantile y_α è data da un qualunque valore di t (in generale non unico) tale che:

$$\hat{F}_Y(\hat{y}_\alpha) \approx \alpha.$$

Esempio 3 Abbiamo estratto 6 campioni di una v.a. X uniforme in $[0, 1]$:

$$X_1 = 0.128251, X_2 = 0.0164056, X_3 = 0.142219,$$

$$X_4 = 0.0797794, X_5 = 0.0164213, X_6 = 0.78388.$$

Sia F la funzione $F(x) = x^2$: stimare il quantile al livello 0.8 della v.a. $Y = F(X)$.

Procediamo come illustrato sopra: calcoliamo $Y_i = F(X_i)$ ottenendo

$$Y_1 = 0.0164484, Y_2 = 0.000269142, Y_3 = 0.0202264,$$

$$Y_4 = 0.00636475, Y_5 = 0.00026966, Y_6 = 0.614467.$$

Gli \hat{Y}_i si ottengono dagli Y_i con un semplice ordinamento crescente:

$$\hat{Y}_1 = 0.000269142, \hat{Y}_2 = 0.00026966, \hat{Y}_3 = 0.00636475,$$

$$\hat{Y}_4 = 0.0164484, \hat{Y}_5 = 0.0202264, \hat{Y}_6 = 0.614467.$$

Dobbiamo ora trovare $\hat{y}_{0.8}$ tale che $\hat{F}_Y(\hat{y}_{0.8}) \approx 0.8$, poichè $N = 6$ la \hat{F}_Y è una funzione a gradini tutti di ampiezza $1/6 \approx 0.17$ quindi $F_Y(t) = 4/6 \approx 0.666$ quando $\hat{Y}_4 \leq t < \hat{Y}_5$ e $F_Y(t) = 5/6 \approx 0.8333$ quando $\hat{Y}_5 \leq t < \hat{Y}_6$. Si può scegliere come stima del quantile $\hat{y}_{0.8} = \hat{Y}_5 = 0.0202264$. Come esercizio si calcoli il vero quantile di Y e lo si confronti con il valore ottenuto dalla stima MC. Si vedrà che l'accordo non è molto buono. Ciò è dovuto alla piccola dimensione ($N = 6$) del campione estratto.

1 Esercizi sul legame tra esponenziale e Poisson, su catene di Markov e su processi di Markov a salti

Exercise 1 *Un sistema può trovarsi in uno qualsiasi di n stati, denotati con i numeri $1, \dots, n$. Quando il sistema si trova nello stato 1, resta lì per un tempo aleatorio esponenziale di media $5m$ (minuti), poi salta nello stato 2. Quando si trova in 2, lì resta per un tempo aleatorio esponenziale di media $5m$, poi salta nello stato 3. Lo stesso accade se si trova negli stati $i = 3, \dots, n - 1$. Invece, se si trova nello stato n , resta sempre lì. Trovare la distribuzione di probabilità in regime stazionario.*

Soluzione. Usiamo come modello un processo di Markov a tempo continuo e stati $1, 2, \dots, n$. Ricordiamo che i tassi di transizione a_{ij} (con $i \neq j$) possono essere interpretati mediante orologi esponenziali: pertanto vale $a_{12} = \frac{1}{5}$, \dots , $a_{n-1,n} = \frac{1}{5}$. Infatti (come abbiamo visto nel capitolo sul processo di nascita e morte), se T_{12} indica l'istante aleatorio, distribuito in modo esponenziale, in cui il sistema salta da 1 a 2, e λ_{12} è il parametro di T_{12} , allora $a_{12} = \lambda_{12}$. D'altra parte, il valor medio di T_{12} vale $\frac{1}{\lambda_{12}}$. Per ipotesi è $E[T_{12}] = 5$, quindi $a_{12} = \frac{1}{5}$, e così via. Invece vale $a_{21} = 0$, \dots , $a_{n,n-1} = 0$ così come sono uguali a zero i tassi tra stati che distano più di uno. Pertanto la matrice dei tassi di transizione è

$$A = \begin{pmatrix} -\frac{1}{5} & \frac{1}{5} & 0 & \dots & \\ 0 & -\frac{1}{5} & \frac{1}{5} & 0 & \\ & \dots & \dots & \dots & \\ & & 0 & -\frac{1}{5} & \frac{1}{5} \\ & & \dots & 0 & 0 \end{pmatrix}.$$

Tutti gli stati sono transienti, escluso n che assorbente. Pertanto l'unica misura invariante è la massa unitaria concentrata in n :

$$\pi = (0, \dots, 0, 1).$$

Verifichiamolo anche algebricamente: se $\pi = (\pi_1, \dots, \pi_n)$ è una misura invari-

ante, allora l'equazione $\pi A = 0$ si scrive

$$\left\{ \begin{array}{l} -\frac{1}{5}\pi_1 = 0 \\ \frac{1}{5}\pi_1 - \frac{1}{5}\pi_2 = 0 \\ \dots \\ \frac{1}{5}\pi_{n-2} - \frac{1}{5}\pi_{n-1} = 0 \\ \frac{1}{5}\pi_{n-1} = 0 \end{array} \right.$$

da cui si deduce $\pi_1 = 0, \dots, \pi_{n-1} = 0$. Dalla condizione di normalizzazione $\pi_1 + \dots + \pi_n = 1$ si trova infine $\pi_n = 1$.

Exercise 2 *Un imprenditore ha una ricchezza aleatoria X , che può essere pari a $0, 1, 2, 4, 8, 16, \dots$. Egli agisce nel seguente modo: quando ha la ricchezza X la investe tutta. Con probabilità $\frac{1}{3}$ questa, nell'arco di un mese, si raddoppia, mentre con probabilità $\frac{2}{3}$ si dimezza, salvo quando è pari ad 1, nel qual caso con probabilità $\frac{2}{3}$ si estingue invece che dimezzarsi. Quando poi la sua ricchezza è pari a zero, con probabilità $\frac{2}{3}$ resta a zero, mentre con probabilità $\frac{1}{3}$ essa torna ad 1 grazie al dono di parenti o amici. Supponiamo che parta da una ricchezza pari a 2. Calcolare la probabilità che dopo 4 mesi egli abbia una ricchezza pari a 4. Col trascorrere del tempo, la sua ricchezza raggiunge un regime stazionario? Se sì, qual'è in tale regime la probabilità di avere una ricchezza pari a 8?*

Soluzione. Usiamo una catena di Markov (X_n) a tempo discreto e stati $0, 1, 2, 4, 8, 16, \dots$ (tutte le potenze di due). In realtà, essendo le potenze di due in corrispondenza biunivoca con gli interi non negativi, usiamo come stati i numeri $0, 1, 2, 3, 4, 5, \dots$ e nell'interpretazione applicativa dei dati e dei risultati pensiamo che questi stati $k \in \mathbb{N}$ siano gli esponenti delle potenze di due, 2^k , ovvero le possibili ricchezze dell'imprenditore. C'è poi un ulteriore dettaglio: alla ricchezza pari a 0 facciamo corrispondere il numero -1 . Quindi, in definitiva, tutti gli stati della nostra catena sono i numeri interi maggiori o uguali a -1 .

La scadenza temporale è quella mensile. Vale (per ipotesi)

$$p_{i,i+1} = \frac{1}{3}, \quad p_{i,i-1} = \frac{2}{3},$$

per tutti i valori di i possibili, ed inoltre $p_{-1,-1} = \frac{2}{3}$.

Prima domanda: dobbiamo calcolare $P(X_4 = 2)$ sapendo che al tempo zero la ricchezza è pari a 1 (si ricordi che gli stati sono gli esponenti delle potenze di due). Raffigurando il grafico della catena si vede che vale

$$P(X_4 = 2) = 0$$

in quanto non ci sono cammini lunghi 4 che portano da 1 a 2 (si noti che la probabilità di restare in uno stato è nulla, eccetto che per -1).

La catena raggiunge un regime stazionario? Si tratta di una catena di nascita e morte, che rientra inoltre nel caso particolare in cui le probabilità λ_i e μ_i (usando le notazioni della teoria di tali catene) non dipendono da i , e valgono $\lambda = \frac{1}{3}$, $\mu = \frac{2}{3}$. Siccome $\lambda < \mu$, vale la condizione per l'esistenza di un regime stazionario.

In tale regime, la distribuzione invariante è

$$\pi_i = \frac{a_i}{\sum_{j=0}^{\infty} a_j}.$$

Dobbiamo però prestare attenzione a dove variano gli indici. Nelle notazioni della teoria delle catene di nascita e morte $i \geq 0$, mentre nel nostro esempio è ≥ -1 . Indichiamo allora con $(\pi_i)_{i \geq 0}$ la misura invariante della teoria e con $(\pi'_i)_{i \geq -1}$ la misura invariante del nostro esempio. Vale

$$\pi'_i = \pi_{i+1}.$$

Pertanto

$$\pi'_i = \frac{a_{i+1}}{\sum_{j=0}^{\infty} a_j} = \frac{\mu - \lambda}{\mu} \left(\frac{\lambda}{\mu}\right)^{i+1}.$$

ricordando che

$$\sum_{j=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^j = \frac{1}{1 - \frac{\lambda}{\mu}} = \frac{\mu}{\mu - \lambda}.$$

In conclusione, la probabilità richiesta dall'esercizio che la ricchezza sia pari a 8, ovvero che lo stato sia 3, è

$$\pi'_3 = \frac{1}{\frac{3}{2}} \left(\frac{1}{2}\right)^4 = \frac{1}{32} = 0.03125.$$

Exercise 3 Una grossa azienda ha un parco automezzi da trasporto molto ampio, in cui c'è un continuo andirivieni dovuto a meccanismi complessi, per cui non è facilmente prevedibile il numero di automezzi disponibili ad un generico istante. Supponiamo che un adetto incaricato di osservare l'andirivieni noti che quando ci sono n automezzi, il tempo da attendere prima di vedere un rientro ha media 20 minuti, mentre il tempo da attendere prima di vedere una partenza ha media 16,66 minuti. Trovare la probabilità di avere almeno 10 automezzi nel regime stazionario. Si supponga che i tempi siano esponenziali.

Soluzione. Usiamo un processo di Markov a tempo continuo e stati uguali a $0, 1, 2, \dots$ (i numeri interi non negativi). Detto $T_{i,i+1}$ il tempo da attendere prima di un nuovo rientro e $T_{i,i-1}$ quello prima di una nuova partenza, vale

$$a_{i,i+1} = \frac{1}{E[T_{i,i+1}]} = 0.05 \text{ min}^{-1}, \quad a_{i,i-1} = \frac{1}{E[T_{i,i-1}]} = 0.06 \text{ min}^{-1}.$$

Si tratta quindi di un processo di nascita e morte a tempo continuo, del caso particolare con tassi costanti. Usando le notazioni generali della teoria, essendo $\lambda < \mu$ si raggiunge un regime stazionario. In esso la probabilità invariante è

$$\pi_i = \frac{\left(\frac{\lambda}{\mu}\right)^i}{\sum_{j=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^j} = \frac{\mu - \lambda}{\mu} \left(\frac{\lambda}{\mu}\right)^i = \frac{1}{6} \left(\frac{5}{6}\right)^i.$$

Pertanto la probabilità di avere almeno 10 automezzi è

$$\sum_{i=10}^{\infty} \frac{1}{6} \left(\frac{5}{6}\right)^i = 1 - \sum_{i=0}^9 \frac{1}{6} \left(\frac{5}{6}\right)^i = 0.161.$$

Exercise 4 L'azienda delle autostrade capisce che deve potenziare un certo casello, troppo spesso intasato da code. Mette allora a monte del casello (un chilometro prima, ad esempio), un osservatore, che nota quanto segue: tra un passaggio di una macchina e della successiva trascorre un tempo aleatorio distribuito esponenzialmente, con tempo medio di interarrivo pari a 5 secondi. Qual'è la probabilità che passino 5 macchine in trenta secondi? Qual'è la probabilità che ne passino più di 10?

Soluzione. Indichiamo con T_1 l'istante del primo passaggio di autovettura (passaggio al punto di osservazione), con T_2 il tempo che trascorre prima del secondo passaggio (quindi T_2 è un tempo di interarrivo), e così via per T_3 , ecc. Ciascun T_n è distribuito esponenzialmente con parametro $\lambda = \frac{1}{E[T_1]} = \frac{1}{5} s^{-1}$. Inoltre li supponiamo indipendenti.

Dobbiamo calcolare la probabilità che passino 5 macchine in 30s. Ci sono due interpretazioni di questa frase. L'interpretazione meno stretta è che si voglia la probabilità che passino *almeno* 5 macchine, che equivale a chiedere che l'istante di passaggio della quinta macchina, $T_1 + \dots + T_5$, avvenga prima dei trenta secondi. In questo caso dobbiamo calcolare

$$P(T_1 + \dots + T_5 \leq 30s).$$

Invece l'interpretazione più stretta è che passino *esattamente* 5 macchine.

Risolviamo il primo caso, a titolo di esempio (il caso veramente richiesto è il secondo). Ricordando il legame tra esponenziale ed Erlang, abbiamo

$$\begin{aligned} P(T_1 + \dots + T_5 \leq 30s) &= 1 - P(T_1 + \dots + T_5 > 30s) \\ &= 1 - \left(1 + 6 + \dots + \frac{(6)^4}{4!}\right) e^{-6} = 0.71494. \end{aligned}$$

La risoluzione del secondo caso è immediata se si ricorda il teorema che dice che il numero di arrivi in un lasso di tempo t è distribuito secondo Poisson, di parametro λt . Quindi la probabilità che tale numero sia 5, in $t = 30$ secondi, è

$$e^{-\lambda t} \frac{(\lambda t)^5}{5!} = e^{-6} \frac{(6)^5}{5!} = 0.16062.$$

Infine l'esercizio chiedeva la probabilità che in trenta secondi passino più di dieci macchine, ovvero almeno 11 macchine, ovvero

$$P(T_1 + \dots + T_{11} \leq 30s).$$

Come sopra, questa è

$$\begin{aligned} &1 - P(T_1 + \dots + T_{11} > 30s) \\ &= 1 - \left(1 + 6 + \dots + \frac{(6)^{10}}{10!}\right) e^{-6} = 4.2621 \times 10^{-2}. \end{aligned}$$

Exercise 5 Si consideri una catena di Markov (X_n) con 6 stati e con le seguenti probabilità di transizione:

$$\begin{aligned} p_{12} &= \frac{1}{2}, & p_{13} &= \frac{1}{2}, & p_{21} &= 1, & p_{31} &= 1, \\ p_{43} &= \frac{1}{3}, & p_{44} &= \frac{1}{3}, & p_{45} &= \frac{1}{3}, \\ p_{55} &= \frac{1}{2}, & p_{56} &= \frac{1}{2}, & p_{65} &= \frac{1}{2}, & p_{66} &= \frac{1}{2}. \end{aligned}$$

- 1) Disegnare il grafo della catena e scrivere la matrice di transizione.
- 2) Calcolare $P(X_0 = 4, X_1 = 4, X_2 = 3, X_3 = 1)$, sapendo che la catena parte dallo stato 4.
- 3) Calcolare $p^{(2)}$, la distribuzione al tempo 2, sapendo che la catena parte dallo stato 4.
- 4) Classificare gli stati.
- 5) Trovare tutte le distribuzioni invarianti.

Soluzione. Tralasciamo di scrivere la soluzione del punto 1). Per il 2) vale

$$\begin{aligned} P(X_0 = 4, X_1 = 4, X_2 = 3, X_3 = 1) &= p_{31}p_{43}p_{44}P(X_0 = 4) \\ &= 1 \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot 1 = \frac{1}{9}. \end{aligned}$$

Per il 3), vale (osservando il grafo)

$$\begin{aligned} P(X_2 = 1) &= \frac{1}{3} \cdot 1 = \frac{1}{3}, & P(X_2 = 2) &= 0, & P(X_2 = 3) &= p_{43}p_{44}P(X_0 = 4) = \frac{1}{9}, \\ P(X_2 = 4) &= \frac{1}{9}, & P(X_2 = 5) &= \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} = \frac{5}{18}, & P(X_2 = 6) &= \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}. \end{aligned}$$

A titolo di verifica, sommando questi numeri si ottiene

$$\frac{1}{3} + 0 + \frac{1}{9} + \frac{1}{9} + \frac{5}{18} + \frac{1}{6} = 1.$$

Per il 4), osserviamo che lo stato 4 è transitorio, gli stati 1,2,3 sono ricorrenti e formano una classe chiusa irriducibile, gli stati 5,6 sono ricorrenti e formano una classe chiusa irriducibile.

Per il punto 5), in base a teoremi generali ed alla classificazione, sappiamo che nella classe $\{123\}$ c'è un'unica distribuzione invariante, anche nella classe

{56} c'è un'unica distribuzione invariante, ogni distribuzione invariante ha massa zero nello stato 4, ed ogni distribuzione invariante è combinazione convessa delle due dette sopra. Troviamo allora queste due.

Scriviamo come $\pi = (\pi_1, \pi_2, \pi_3)$ la distribuzione invariante nella classe {123}. Per il bilancio di flusso in 2 e 3 vale

$$\frac{1}{2}\pi_1 = \pi_2, \quad \frac{1}{2}\pi_1 = \pi_3.$$

Sostituendo allora nella condizione di normalizzazione $\pi_1 + \pi_2 + \pi_3 = 1$ troviamo

$$\pi_1 + \frac{1}{2}\pi_1 + \frac{1}{2}\pi_1 = 1$$

ovvero $\pi_1 = \frac{1}{2}$. Riutilizzando le due espressioni trovate sopra abbiamo poi $\pi_2 = \frac{1}{4}$ e $\pi_3 = \frac{1}{4}$. In definitiva la distribuzione invariante nella classe {123} è

$$\pi = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4} \right).$$

Si noti che a questo risultato si poteva arrivare anche on un ragionamento di simmetria: pensando gli stati 2,3 come un'unico stato, il sistema porta {1} in {23} con probabilità uno e {23} in {1} con probabilità uno, quindi la distribuzione invariante deve dare massa $\frac{1}{2}$ sia a {1} sia a {23}. Poi, siccome i pesi relativi a {2} e {3} sono uguali, la massa $\frac{1}{2}$ in {23} deve essere suddivisa in $\frac{1}{4}$ in {2} e $\frac{1}{4}$ in {3}.

Troviamo la distribuzione invariante $\pi = (\pi_5, \pi_6)$ nella classe {56}. Qui si può applicare come sempre il bilancio di flusso, ma è evidente che la situazione è perfettamente simmetrica tra {5} e {6}. Quindi la distribuzione invariante è

$$\pi = \left(\frac{1}{2}, \frac{1}{2} \right).$$

Completiamo questi calcoli con alcune riscritture. Guardando le cose dal punto di vista dell'intera catena, la prima distribuzione invariante trovata è

$$\pi = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0 \right)$$

mentre la seconda è

$$\pi = \left(0, 0, 0, 0, \frac{1}{2}, \frac{1}{2}\right).$$

Infine, tutte le altre distribuzioni invarianti sono combinazioni convesse di queste due, ovvero sono della forma

$$\begin{aligned}\pi &= \alpha \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0, 0, 0\right) + \beta \left(0, 0, 0, 0, \frac{1}{2}, \frac{1}{2}\right) \\ &= \left(\alpha\frac{1}{2}, \alpha\frac{1}{4}, \alpha\frac{1}{4}, 0, \beta\frac{1}{2}, \beta\frac{1}{2}\right)\end{aligned}$$

con $\alpha + \beta = 1$, $\alpha, \beta \geq 0$.

Exercise 6 *Si consideri un processo di Markov (X_t) a tempo continuo e stati discreti 1,2,3. Supponiamo che il tasso di transizione da 1 a 2 sia pari a 10, anche quello da 2 a 1 sia 10, e così via siano pari a 10 i tassi di transizione da 2 a 3, da 3 a 2, da 1 a 3, da 3 a 1. Tracciare il grafo associato al processo, scrivere la matrice dei tassi di transizione e trovare le distribuzioni invarianti.*

Soluzione. La matrice dei tassi di transizione è

$$A = \begin{pmatrix} -20 & 10 & 10 \\ 10 & -20 & 10 \\ 10 & 10 & -20 \end{pmatrix}.$$

Il processo è irriducibile, quindi ha un'unica distribuzione invariante $\pi = (\pi_1, \pi_2, \pi_3)$. Ovviamente per simmetria vale

$$\pi = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right).$$

A titolo di esempio, verifichiamolo anche in altri due modi.

Vale $\pi A = 0$, quindi

$$\begin{cases} \pi_1(-20) + \pi_2 10 + \pi_3 10 = 0 \\ \pi_1 10 + \pi_2(-20) + \pi_3 10 = 0 \\ \pi_1 10 + \pi_2 10 + \pi_3(-20) = 0 \end{cases}$$

più la condizione $\pi_1 + \pi_2 + \pi_3 = 1$ (infatti il sistema precedente è sottodeterminato). Il sistema si può riscrivere nella forma

$$\begin{cases} 2\pi_1 = \pi_2 + \pi_3 \\ 2\pi_2 = \pi_1 + \pi_3 \\ 2\pi_3 = \pi_1 + \pi_2 \end{cases}$$

per cui usando la prima nell'equazione $\pi_1 + \pi_2 + \pi_3 = 1$ troviamo $\pi_1 + 2\pi_1 = 1$, ovvero $\pi_1 = \frac{1}{3}$. Con lo stesso metodo si trovano gli altri.

Il secondo metodo è quello del bilancio di flusso. Il bilancio in 1 è

$$(10 + 10)\pi_1 = 10\pi_2 + 10\pi_3.$$

Questa è la prima delle tre equazioni scritte sopra. Scrivendo il bilancio negli altri due stati si trovano le altre due equazioni e quindi ci si riconduce a risolvere lo stesso sistema scritto sopra. Si vede che non c'è un sostanziale vantaggio ad usare il bilancio di flusso, se non semplicemente di tipo grafico piuttosto che eseguire un prodotto vettore-matrice (invece nelle catene a tempo discreto può esserci un lieve vantaggio perchè il bilancio di flusso ingloba l'operazione di sottrazione di π insita nell'equazione $\pi = \pi P$).

Exercise 7 Si consideri una catena di Markov (X_n) con 3 stati e con le seguenti probabilità di transizione:

$$\begin{aligned} p_{12} &= \frac{1}{2}, & p_{13} &= \frac{1}{2}, & p_{21} &= 1, \\ p_{31} &= \frac{1}{2}, & p_{32} &= \frac{1}{2}. \end{aligned}$$

- 1) Disegnare il grafo della catena e scrivere la matrice di transizione.
- 2) Calcolare $P(X_0 = 1, X_1 = 3, X_2 = 1)$, sapendo che la catena parte dallo stato 1 con probabilità 0.7.
- 3) Calcolare $p^{(2)}$, la distribuzione al tempo 2, sapendo che la catena parte dallo stato 2.
- 4) Classificare gli stati.
- 5) Trovare tutte le distribuzioni invarianti.

Exercise 8 Si consideri una catena di Markov con stati 1,2,3, 4, 5,6,7. Supponiamo che gli stati 1,2,3 comunichino con probabilità di transizione uniforme (ovvero $p_{11} = p_{12} = p_{13} = \frac{1}{3}$ e così via per gli altri). Ugualmente,

supponiamo che gli stati 4,5,6 comunichino con probabilità di transizione uniforme. Supponiamo infine che lo stato 4 comunichi con probabilità uniforme con se stesso e con 1 e 4. Raffigurare la catena e scrivere la matrice di transizione. Supponendo che la catena parta dallo stato 4 con probabilità 1, calcolare la distribuzione al tempo 3. Classificare gli stati e trovare tutte le probabilità invarianti.