

Appunti del corso di
Modelli Stocastici e Analisi dei Dati
(Draft)

Ilia Negri
ilia.negri@unibg.it

15 marzo 2010

Indice

1	Introduzione	5
1.1	Processi a tempo discreto	5
1.2	Processi in tempo continuo	9
2	Richiami di calcolo delle probabilità	15
2.1	La probabilità	15
2.2	Proprietà della probabilità	19
2.2.1	Principio di inclusione ed esclusione	21
2.3	Probabilità condizionata	22
2.4	Indipendenza	23
2.5	Variabili casuali discrete	25
2.5.1	Valore atteso	29
2.6	Vettori aleatori	33
2.6.1	Vettori aleatori discreti	33
3	Catene di Markov	37
3.1	Processi Stocastici	37
3.2	Notazioni di base	37
3.3	Esempi fondamentali	39
3.3.1	La rovina del giocatore	40
3.3.2	La camminata aleatoria	40
3.3.3	Modello di Eherenfest di diffusione	44
3.3.4	Modello per sequenza di successi	45
3.3.5	Fila d'attesa discreta	46
3.3.6	Catene di nascita e morte	47
3.3.7	Un modello genetico	48
3.3.8	Il problema del ballottaggio	49
3.4	Probabilità di transizione in n passi	50
3.4.1	Esempi	53

3.5	Catena di Markov con due stati	55
4	Classificazione degli stati	59
4.1	Tempi di arresto	59
4.2	Criteri per la classificazione degli stati	65
4.3	Catene con un numero finito di stati	66
5	Distribuzioni invarianti	71
5.1	Definizione e prime proprietà	71
5.2	Esistenza della distribuzione invariante	76
5.3	Calcolo della distribuzione invariante	79
5.4	Il caso in cui lo spazio degli stati è infinito	84
5.4.1	Catene di nascita e morte	85
5.4.2	File d'attesa discrete	88
6	Probabilità e tempi medi di assorbimento	91
6.1	Probabilità di assorbimento	91
6.1.1	Tempi medi di assorbimento	97
6.2	Catene con un'infinità numerabile di stati	102
6.2.1	Catene di nascita e morte	102
7	Generazione di variabili casuali	105
7.1	Introduzione	105
7.2	Il metodo della funzione inversa	105
7.3	Generazione di variabili casuali discrete	106
7.4	Generazione di variabili casuali continue	107
7.5	Simulazione di una catena di Markov	109
8	Processi di Poisson	113
8.1	Proprietà della variabile casuale esponenziale	113
8.2	Processi di Poisson omogenei	116
8.3	Probabilità condizionate	118
8.4	Processi di Poisson non omogenei	120
9	Introduzione alla teoria delle file di attesa	123
9.1	Introduzione	123
9.2	Sistemi esponenziali	124
9.2.1	Le equazioni del bilancio dettegiato	124
9.2.2	Numero medio di utenti e tempi medi di attesa	126
9.3	Esempi	127

9.3.1	Il problema di Erlang	127
9.3.2	La coda $M/G/\infty$	128
10	Esercizi	131
10.1	Catene di Markov	131
10.2	Simulazione	135
10.3	Processi di Poisson	135
10.4	Simulazione	137

Capitolo 1

Introduzione

La teoria dei processi stocastici riguarda lo studio di sistemi che evolvono nel tempo (ma anche più in generale nello spazio) secondo leggi probabilistiche. Tali sistemi o modelli descrivono fenomeni complessi del mondo reale che hanno la possibilità di essere aleatori. Tali fenomeni sono più frequenti di quanto si possa credere e si incontrano tutte quelle volte in cui le quantità alle quali siamo interessati non sono prevedibili con certezza. Quando però tali fenomeni mostrano una variabilità di possibili esiti che può essere in qualche modo spiegata o descritta, allora possiamo introdurre il modello probabilistico del fenomeno.

In questo capitolo introduttivo faremo una panoramica dei possibili campi dove i fenomeni sono modellabili con particolari processi stocastici. Questi campi spaziano dalle scienze naturali e biologiche a quelle mediche, dall'economia, all'industria e all'ingegneria. Si parlerà diffusamente di probabilità, di realizzazioni e di distribuzioni senza averle definite in modo formale (questo avverrà nel capitolo 2), ma confidiamo nel fatto che tali termini, almeno dal punto di vista intuitivo, siano famigliari al lettore. Ad esempio supporremo che frasi come *la probabilità che domani piova è 0.4* siano comprensibili così come la deduzione del fatto che quindi *la probabilità che domani non piova è 0.6*. Nei paragrafi che seguono presenteremo alcuni esempi utilizzando questi termini, per ora informali, ma che ci danno l'opportunità di introdurre una parte di terminologia che utilizzeremo anche in seguito.

1.1 Processi a tempo discreto

Il primo processo che introduciamo e che ha svariate applicazioni è la passeggiata aleatoria. Essa è la generalizzazione dell'idea di successione di variabili indipendenti, come sarà chiaro più avanti. Si consideri una particella che parte dall'origine di una retta e può muoversi a destra o a sinistra con uguale probabilità. Indichiamo

con X_n , $n = 1, 2, \dots$ la posizione della particella all'istante n . Possiamo dire che $X_n = X_{n-1} + Z_n$, dove Z_n , $n = 1, 2, \dots$ sono una successione di variabili casuali indipendenti che possono assumere solo i valori -1 o 1 con uguale probabilità. Si vede allora che la successione X_n , $n = 1, 2, \dots$, costituisce una successione di variabili aleatorie che non sono indipendenti, ma tali che il valore che assume la variabile X_n dipende solo dalla variabile X_{n-1} . Dopo un numero n pari di istanti la particella potrà trovarsi in uno qualunque dei punti di coordinate $(-n, \dots, -2, 0, 2, \dots, n)$. Dopo un numero n dispari di istanti la particella potrà trovarsi in uno qualunque dei punti di coordinate $(-n, \dots, -1, 1, \dots, n)$. Per una certa realizzazione della successione di variabili aleatorie Z_n , possiamo rappresentare la corrispondente traiettoria della particella, mettendo nell'asse delle ascisse gli istanti, $n = 1, 2, \dots$ e sull'asse delle ordinate la posizione della particella all'istante considerato. Nella terminologia della teoria dei processi stocastici i possibili valori che possono assumere ciascuna delle variabili X_n è detto spazio degli stati. Il processo è costituito dalla famiglia di variabili casuali X_n , $n = 1, 2, \dots$. Una realizzazione del processo (inteso come famiglia di variabili casuali) è detto traiettoria. Ovviamente le traiettorie possono differire molto tra loro, ma questo non sorprende trattandosi di processi aleatori. I problemi che possiamo porci sono del tipo: quale è la probabilità che la particella raggiunga un fissato punto? Quale è la probabilità che ripassi per quel punto? Quale è il tempo medio di permanenza in un punto? Vi sono poi delle generalizzazioni possibili di questo semplice modello. Ad esempio la probabilità di andare a destra o sinistra può non essere uguale. Oppure la particella può anche rimanere nel punto in cui si trova con una certa probabilità, cioè le variabili Z_n possono assumere i valori $-1, 0, 1$ rispettivamente con probabilità p_1, p_2, p_3 , $p_1 + p_2 + p_3 = 1$. Si possono poi mettere una o due barriere alla passeggiata. Cioè quando la particella arriva ad esempio nel punto k o vi rimane per sempre (barriera assorbente) o viene riflessa nel punto adiacente con probabilità 1 (barriera riflettente). Un'applicazione del modello della passeggiata aleatoria lo troviamo nelle scienze sociali o politiche. Si tratta dell'importante problema del ballottaggio. Due candidati A e B alle elezioni hanno ricevuto rispettivamente a e b voti, con $a > b$. Ci si chiede quale è la probabilità che durante lo spoglio dei voti il candidato A risulti sempre in testa.

Un'altra importante applicazione è il cosiddetto problema della rovina del giocatore. Un tizio parte con un capitale di a unità ed ad ogni istante in cui gioca può vincere o perdere un'unità del suo capitale. X_n rappresenta le unità di capitale che il giocatore possiede all'istante n . Il gioco termina quando o il tizio perde tutto il suo capitale o quando raggiunge un tetto di capitale fissato m . In questo caso gli stati 0 ed m costituiscono due barriere assorbenti. È interessante domandarsi qual è la probabilità che il giocatore finisca sul lastrico, che in termini della

passeggiata significa la probabilità che la particella sia assorbita in 0. Oppure qual è la probabilità che sia assorbita in m prima che in 0. Inoltre tale assorbimento sarà certo oppure no?

La passeggiata aleatoria costituisce un primo semplice esempio di una grande famiglia di processi che va sotto il nome di catene di Markov dal nome del matematico russo A.A. Markov, che per primo, all'inizio del secolo, le studiò e ne definì formalmente la teoria. La caratteristica fondamentale di questi processi può riassumersi dicendo che lo stato in cui la catena si troverà domani (X_{n+1}) dipende solo da dove si trova oggi (X_n) e non da dove si trovava ieri o avantieri o ancora prima ($X_{n-1}, X_{n-2}, \dots, X_1, X_0$). Come esempio in campo ingegneristico (affidabilità di un sistema) si consideri la valvola di un sistema che ad ogni controllo può trovarsi in uno dei tre stati: soddisfacente, insoddisfacente (ma funzionante), rotta. Indichiamo tali stati rispettivamente con 0, 1, 2. Si può essere interessati a scoprire se sia conveniente e quanto rimpiazzare il componente quando si trova nello stato 1 prima che si rompa. Una possibile modellazione è questa. Se il componente si trova nello stato 0, può rimanere in quello stato con probabilità p_{00} , può andare nello stato 1 con probabilità p_{01} , può andare nello stato 2 con probabilità p_{02} , con $p_{00} + p_{01} + p_{02} = 1$. Se si trova nello stato 1 può rimanere nello stato 1 con probabilità p_{11} , o può andare nello stato 2 con probabilità p_{12} , con $p_{11} + p_{12} = 1$. Infine se si trova nello stato 2 vi rimane per sempre. Le p_{ij} sono dette probabilità di transizione e sono completamente specificate dalla matrice

$$\begin{bmatrix} p_{00} & p_{01} & p_{02} \\ 0 & p_{11} & p_{12} \\ 0 & 0 & 1 \end{bmatrix}$$

i cui elementi sommano ad uno per riga. Il sistema descritto da una tale catena è possibile simularlo ed osservarne le possibili traiettorie. Si può determinare il tempo medio di rottura del pezzo e studiarne attraverso le simulazioni il comportamento.

Un esempio nel campo delle scienze sociali è costituito dal modello che descrive la mobilità degli impiegati all'interno di una azienda. Supponiamo di avere 4 livelli di impiego, indicizzati dal più basso al più alto. Siamo interessati a sapere con che probabilità un impiegato che entra al livello più basso arriverà alla fine della sua carriera al livello più alto.

È chiaro che questi esempi costituiscono un possibile modello del problema, e come tali sono basati su determinate ipotesi. In questo caso potremmo dire che l'ipotesi che le probabilità di transizione non dipendano da n è molto forte. Potrebbe essere più realistico ipotizzare che la probabilità di passare dal primo livello di carriera al successivo vari (ad esempio cali) col tempo. I processi in cui le probabilità di transizione non dipendono da n sono detti omogenei.

I fenomeni che possono essere modellati con le catene di Markov sono spesso facili da riconoscere proprio per la mancanza di memoria di quanto è avvenuto prima di un certo orizzonte temporale. Una volta che una catena di Markov è identificata la teoria che la regola limita e spiega il comportamento della stessa. Si possono determinare gli stati ricorrenti (stati per cui la catena passerà infinitamente spesso) e quelli transitori (in cui la catena vi ripasserà con una probabilità minore di 1). Studieremo dei metodi che ci permetteranno di calcolare la probabilità che la catena entri in certo stato, oppure di calcolare il tempo medio di permanenza in quello stato. Inoltre saremo in grado di determinare il comportamento della catena con il passare del tempo attraverso lo studio della distribuzione invariante.

Vediamo altri due esempi di applicazione delle catene di Markov. Le catene di Markov sono utilizzate per modellare processi biologici, nello studio della crescita delle popolazioni, nello studio delle ereditarietà genetiche ed epidemiche. A volte tali modelli vengono molto, a volte troppo, semplificati al fine di poterli trattare matematicamente ma, resta il fatto che questa semplificazione permette di capire quantitativamente il comportamento del fenomeno empirico, e l'utilizzo di tali studi ha portato notevoli contributi verso una migliore comprensione del fenomeno stesso. Ad esempio, nello studio dell'evoluzione della popolazione le catene di Markov furono utilizzate già da Galton e Watson per spiegare il fenomeno della scomparsa dei nomi di famiglia anche in una popolazione in crescita. Essi introdussero quelli che sono noti oggi come *branching processes* (la traduzione italiana non viene usata). Si consideri una popolazione (di individui, molecole, maschi con un cognome o altro ancora) i cui elementi sono capaci di generare un certo numero di discendenti. Il numero di discendenti per ogni elemento della popolazione sarà $0, 1, 2, \dots$ con probabilità rispettivamente p_0, p_1, p_2, \dots e ciascun elemento di una generazione genera indipendentemente dagli altri. Supponiamo di partire con un individuo ed esaminiamo lo sviluppo dei suoi discendenti generazione dopo generazione. Il primo individuo appartiene alla generazione 0, i suoi discendenti alla generazione 1, i discendenti dei suoi discendenti alla generazione 2 e così via. Se tutti gli individui di una generazione non si riproducono la popolazione si è estinta. Il problema è di calcolare la probabilità di estinzione di una popolazione. Le ipotesi minimali devono essere $p_0 > 0$, perché se $p_0 = 0$ tutti gli individui devono avere almeno un discendente e l'estinzione è impossibile. Se indichiamo con X_n la consistenza della popolazione nella n -esima generazione, allora

$$\mathbf{P}(X_{n+1} = k | X_n = j) = \mathbf{P}(Z_1 + \dots + Z_j = k)$$

dove le Z_i sono una successione di variabili casuali indipendenti che rappresentano il numero di discendenti che l'individuo i -esimo ha generato all'istante n . La con-

sistenza della popolazione all'istante $n + 1$ sarà k se e solo se i j elementi presenti all'istante n hanno generato k elementi. La consistenza X_n ha un numero infinito di stati: $0, 1, 2, \dots$ e lo stato 0 è uno stato "assorbente. Questo tipo di modello ha molte altre applicazioni come ad esempio lo studio delle reazioni a catena in fisica e nella fissione nucleare.

Le catene di Markov sono usate anche nella pianificazioni commerciale. Infatti le strategie e le decisioni commerciali sono sempre soggette al rischio a causa dell'incertezza degli eventi futuri. Se si fosse in grado di quantificare tale rischio, magari sulla base delle passate esperienze, allora la determinazione dell'azione migliore da compiere si ridurrebbe al calcolo della probabilità con cui particolari catene di Markov saranno in determinati stati. Vi sono vari esempi di pianificazione delle risorse che possono essere modellati con catene di Markov. Ad esempio, lo stock di un magazzino, l'acqua in un deposito o la riserva di una assicurazione. Consideriamo il problema di un magazzino con capacità pari a C unità. In ogni istante n c'è una domanda aleatoria di beni del magazzino. La consistenza X_n dice quanti beni vi sono nel magazzino alla fine del periodo n . Il manager del magazzino decide di rifornire il magazzino ogni qualvolta il numero di beni all'istante n è minore di un minimo fissato m . Ci chiediamo se è possibile, al crescere di n , trovare la distribuzione del numero di beni nel magazzino. Qual è il valore atteso del numero di richieste non soddisfatte? Di solito il manager vuole ottimizzare i costi a lungo termine causati da un troppo alto stoccaggio e da una perdita di domanda che non è in grado di soddisfare avendo sottorifornito il proprio magazzino.

1.2 Processi in tempo continuo

Tutti gli esempi fin qui visti riguardano casi in cui gli istanti in cui il processo X_n cambia di stato sono discreti ossia $n = 1, 2, \dots$. Si tratta di processi a tempo discreto. Si noti che anche i possibili stati dei sistemi considerati sono discreti. Se invece il tempo è continuo, cioè un sistema $\{X_t\}$ è descritto per un intervallo di tempo continuo, si ha un processo a tempo continuo. Uno dei più importanti processi di questo tipo è il processo di Poisson, che viene usato come modello matematico per molti fenomeni empirici, quali l'arrivo delle chiamate ad un centralino, le emissioni di particelle da una sorgente radioattiva, il numero di incidenti mortali lungo un'autostrada. Le ipotesi fondamentali sono che gli eventi possono verificarsi in un istante di tempo t qualunque di un intervallo $(0, T)$ e possono presentarsi solo uno alla volta. L'arrivo è aleatorio nel senso seguente: sia $N(t, t + \Delta t)$ il numero di eventi che si

verificano nell'intervallo $(t, t + \Delta t)$. Allora si suppone che

$$\mathbf{P}(N(t, t + \Delta t) = 0) = 1 - \rho\Delta t + o(\Delta t)$$

$$\mathbf{P}(N(t, t + \Delta t) = 1) = \rho\Delta t + o(\Delta t)$$

$$\mathbf{P}(N(t, t + \Delta t) \geq 2) = o(\Delta t).$$

Cosa significano queste formule? Esse descrivono in termini matematici proprietà e ipotesi che sono del tutto normali e realistiche nei tipi di fenomeni che il processo di Poisson cerca di modellare. Presentiamo prima di tutto alcuni oggetti non ancora definiti che compaiono nelle formule: ρ è detto tasso di occorrenza ed ha la dimensione di tempo⁻¹ mentre $o(\Delta t)$ descrive una funzione che tende a zero più rapidamente di Δt . La terza formula ci dice che in un intervallo piccolo a piacere, o meglio infinitamente piccolo, Δt , la probabilità che si verifichino 2 o più eventi è zero. La seconda e la prima ci dicono invece rispettivamente che in un intervallo piccolo a piacere Δt si possono verificare solo nessun oppure un solo evento e la probabilità con cui si verifica un solo evento è proporzionale al tasso di occorrenza per la lunghezza dell'intervallo di tempo. Se ci si pensa bene l'ipotesi è assai realistica, perché in sostanza stiamo chiedendo che in un istante infinitesimo di tempo si possa ricevere o meno una telefonata, ma sia impossibile riceverne due. Si suppone inoltre che $N(t, t + \Delta t)$ sia indipendente da $N(0, t)$. Il processo $X_t = N(0, t)$ è detto processo di Poisson con tasso di intensità ρ . Il numero di eventi in ogni intervallo $(t, t + h]$ ha una distribuzione di Poisson con media ρh

$$\mathbf{P}(N(t, t + h) = k) = \frac{(\rho h)^k e^{-\rho h}}{k!}$$

Altro aspetto importante è che il tempo di attesa per il primo evento e anche gli intervalli tra un evento e il successivo sono distribuiti come una variabile casuale esponenziale di parametro ρ : $\mathbf{P}(T > x) = e^{-\rho x}$ e sono indipendenti tra loro. Da ciò discende l'importante fatto che il processo di Poisson è markoviano, cioè la probabilità che una transizione di stato per il processo avvenga nell'intervallo $(t, t + \Delta t]$ dipende solo dallo stato occupato al tempo t e non dall'evoluzione del processo prima di t .

Le traiettorie del processo di Poisson possono essere simulate a partire dalle simulazioni dei tempi di attesa tra due eventi successivi. Naturalmente il parametro t che scandisce il tempo può essere anche interpretato come una coordinata spaziale.

Un importante campo di applicazione del processo di Poisson o di sue generalizzazioni è costituito dallo studio dei problemi di congestione in contesti quali le chiamate ai centralini, lo studio del traffico e altri problemi industriali.

Originariamente il problema che portò allo sviluppo della teoria dei processi applicata ai servizi su grande scala si sviluppò con le operazioni legate ai sistemi telefonici. Più tardi si scoprì che problemi simili per struttura si riscontravano nel commercio (pianificazione del numero di negozi, di venditori, di cassieri, di scorte) nella gestione di produzione, nel calcolo delle capacità delle strade, dei ponti, degli incroci, degli aeroporti, dei porti e così via. Lo sviluppo di tale teoria è dovuto a Erlang ingegnere impiegato alla compagnia dei telefoni di Copenhagen. Egli formulò e risolse il primo problema di questa natura.

Supponiamo che un centralino telefonico riceva chiamate dai suoi abbonati. Se all'istante in cui arriva la telefonata ci sono linee libere l'abbonato viene girato su una di quelle libere e parte la conversazione che dura un tempo non determinato a priori. Se tutte le linee sono occupate vari sistemi di servizio sono possibili. Ne consideriamo due. Un sistema di attesa e un sistema che introduce le perdite. Nel primo, la chiamata viene messa in attesa fino a quando tutte le chiamate precedenti non sono terminate. Nel secondo la chiamata è rifiutata e tutti i servizi seguenti si comportano come se non ci fosse stata. Primo fatto importante: le chiamate arrivano a caso e non è possibile dire adesso quando arriverà la prossima. Secondo fatto: la durata della conversazione non è costante e varia aleatoriamente. I due sistemi differiscono sostanzialmente sia per il tipo di strutture ingegneristiche (cioè per l'aspetto meccanico) sia per i problemi matematici. Nel sistema di attesa è essenziale determinare il tempo medio di attesa prima di accedere al servizio, detto tempo di soggiorno. Per i sistemi con perdita invece tale tempo non è di interesse. La cosa importante in questo caso è determinare la probabilità di congestione, cioè quando tutte le linee sono occupate e quindi si verifica una perdita di chiamate. Nel secondo problema la conoscenza della probabilità di congestione fornisce un quadro completo della situazione quando sono note le caratteristiche del sistema. Nel sistema di attesa invece la media del tempo di attesa, seppure fondamentale, non descrive pienamente il sistema. Altri importanti fattori sono la dispersione del tempo di attesa attorno alla media, la distribuzione della lunghezza della coda di attesa, il prolungarsi del carico dell'apparecchiatura di servizio. Per capire come un tale modello possa essere applicato ad altri contesti, si pensi ad esempio ad una biglietteria o al problema di organizzare un porto. In quest'ultimo caso ad esempio le navi non arrivano in porto secondo un programma stabilito e il tempo di scarico o carico varia. Se mancano le banchine di attracco le navi passano tempo in attesa fuori dal porto e questo costituisce una perdita economica. Un numero maggiore di banchine, d'altra parte, potrebbe accrescere il tempo di inattività dei macchinari e dei lavoratori. Questo genera l'importante problema di saper determinare il numero di banchine di attracco in modo tale da minimizzare i costi sia nei termini

di tempo di attesa delle navi, sia di inattività dei lavoratori e delle macchine. Un altro esempio di applicazione è legato al controllo dei macchinari. In genere in un sistema automatizzato un operatore controlla un certo numero di macchine ed interviene quando una di queste si guasta. Tali macchine si guastano ad istanti aleatori per varie ragioni e richiedono l'intervento dell'operatore. La durata dell'operazione di riparazione è anch'essa aleatoria. Le questioni che ci poniamo sono le seguenti. Quanto vale la probabilità che ad un certo istante un numero fissato di macchine siano inattive? Quanto è il tempo medio di inattività per la macchina in riparazione? Per una data struttura quante macchine può controllare un'operatore in modo efficiente?

Presentiamo quindi uno dei più semplici modelli di processo a coda. Supponiamo che i clienti (siano essi persone piuttosto che telefonate, navi o rotture di macchine) arrivino in un centro servizi secondo un processo di Poisson di tasso λ . I clienti sono serviti da un server uno alla volta e se un cliente che arriva trova il server occupato si mette in una lista di attesa o coda, e ci rimane fino a quando tutti i clienti prima di lui nella lista sono serviti e quindi giunge il suo turno. Supponiamo inoltre che il tempo di durata del servizio per un cliente sia distribuito secondo un variabile casuale esponenziale, cioè abbia densità data da:

$$f(x) = \nu e^{-\nu x}, \quad x > 0.$$

Con queste ipotesi, per il legame tra il processo di Poisson e la variabile casuale esponenziale si dimostra che se all'istante t un cliente è servito dal server, la probabilità che tale servizio termini nell'intervallo $(t, t + \Delta t]$ è pari a $\nu \Delta t + o(\Delta t)$. Tale probabilità è indipendente dal comportamento del processo prima del tempo t e in particolare dalla durata del servizio per quel cliente prima dell'istante t . Questa notevole proprietà vale solo per questa distribuzione del tempo di servizio. Ne vedremo alcune generalizzazioni. Si può descrivere il sistema della coda con un processo i cui stati sono rappresentati dal numero di clienti in coda, incluso quello che è servito. Un transito di stato avviene se arriva un nuovo cliente o se un cliente parte perché servito. In ogni caso si tratta di un processo con la proprietà di Markov in quanto la probabilità che una transizione di stato avvenga nell'intervallo $(t, t + \Delta t]$ dipende solo dallo stato occupato al tempo t e non dall'evoluzione del processo prima di t . Se fissiamo un'istante T il numero di clienti che arriveranno è approssimativamente λT , mentre il tempo medio di servizio è $\frac{1}{\nu}$ e se il server lavora con continuità senza interruzioni, il numero di clienti serviti nel periodo T sarà approssimativamente νT . È intuitivo quindi pensare che se $\lambda > \nu$ la coda crescerà indefinitamente fino ad esplodere, se $\lambda < \nu$ viceversa, il server sarà inattivo per un certo periodo di tempo. Noi studieremo questo tipo di modello e alcune sue generalizzazioni. Ci concen-

treremo in particolare sullo studio della distribuzione limite del processo poiché ci aspettiamo che il sistema della coda, se non esplode, mostri un certo comportamento stabile determinato dalla proporzione nel lungo periodo, della permanenza nei diversi stati. Ci chiederemo ad esempio qual è il numero medio di clienti del sistema, qual è il tempo medio che un cliente passa in coda, o qual è la distribuzione delle lunghezze della coda durante i periodi di attività del server.

Capitolo 2

Richiami di calcolo delle probabilità

2.1 La probabilità

Un esperimento il cui esito non è deterministico, ma che possiede una certa regolarità, ad esempio data da una certa frequenza dei possibili risultati, è detto aleatorio. Tali esperimenti sono molto più frequenti di quanto non si possa credere. Si pensi al numero di clienti che entrano in una filiale di una banca, il tipo di ferita riportata dal prossimo ricoverato al pronto soccorso, il prezzo del titolo della borsa di Milano tra tre settimane, la vittoria della squadra di casa la prossima partita. Questi sono solo alcuni esempi in cui l'esito dell'esperimento può essere diverso e non noto a priori. Ogni volta che ci troviamo di fronte ad un problema concreto il primo passo consiste nella costruzione di un modello adeguato in grado di descrivere tutti i possibili esiti dell'esperimento che stiamo analizzando. Assegnare una probabilità agli esiti possibili costituisce il secondo passo della modellazione del problema. In genere viene fatta basandosi su considerazioni empiriche e soggettive tenendo conto della natura del problema. La probabilità ci fornisce delle informazioni riguardo il possibile esito di un esperimento aleatorio. Poiché gli esperimenti aleatori con cui avremo a che fare sono numerosi introduciamo un formalismo che ci permetta di descriverli tutti.

Definizione 2.1.1. *L'insieme di tutti i possibili esiti di un esperimento aleatorio è detto spazio campionario ed è denotato con Ω . Gli elementi di Ω sono chiamati eventi elementari e denotati con ω .*

Esercizio 2.1.2. Provare a fornire lo spazio campionario per gli esperimenti seguenti:

1. Il lancio di una moneta regolare.
2. Il lancio di due monete regolari.
3. Il lancio di 1000 monete regolari.
4. Il numero di chiamate in arrivo ad un centralino nei prossimi 10 minuti.
5. La scelta casuale di un numero reale compreso tra 0 e 1.
6. La scelta casuale di n numeri reali compresi tra 0 e 1.
7. L'errore commesso nel misurare il peso di un certo oggetto.
8. L'errore commesso nel misurare le tre dimensioni di un parallelepipedo.

Cosa potete osservare?

Nell'esercizio precedente lo spazio degli eventi possibili per l'ultimo esempio è lo spazio tridimensionale \mathbb{R}^3 .

Esercizio 2.1.3. Provare a fornire lo spazio campionario per gli esperimenti seguenti:

1. La posizione, ad un istante t fissato, di una particella che può muoversi liberamente solo su una retta.
2. La posizione, ad un istante t fissato, di una particella che può muoversi liberamente solo su un piano.
3. La posizione, ad un istante t fissato, di una particella che può muoversi liberamente nello spazio.
4. La posizione, ad ogni istante $t \in [0, T]$, di una particella che può muoversi liberamente solo su una retta.
5. La posizione, ad ogni istante $t \in [0, T]$, di una particella che può muoversi liberamente solo su un piano.
6. La posizione, ad ogni istante $t \in [0, T]$, di una particella che può muoversi liberamente nello spazio.

Cosa potete osservare?

Gli ultimi tre esempi dell'esercizio precedente sono detti spazi di funzioni e sono, da un certo punto di vista, gli spazi degli eventi naturali per i processi stocastici in tempo continuo. Sono spazi molto complessi da studiare. Nei prossimi capitoli studieremo i processi stocastici in tempo discreto e gli spazi di possibili eventi elementari saranno elementi di spazi di funzioni più semplici da studiare.

Di fronte ad un esperimento aleatorio abbiamo almeno una certezza: che almeno uno degli esiti facenti parte dello spazio campionario si realizza. Non potendo sapere con certezza quale sarà l'esito dell'esperimento vorremmo almeno conoscere la probabilità con cui dei particolari sottoinsiemi di esiti si possono verificare. Tali sottoinsiemi sono detti *eventi* e sono di solito indicati con le lettere maiuscole A , E e altre, oppure, come nella terminologia insiemistica, enunciando la proprietà che descrive l'insieme. In generale, a seconda della forma di Ω , non tutti i sottoinsiemi di Ω hanno il diritto di chiamarsi eventi. Questo può apparire innaturale ma è dovuto ad un problema tecnico legato alla definizione di probabilità come misura in senso analitico. Ad ogni modo gli eventi a cui dovremo rinunciare saranno non importanti per i nostri scopi. Essendo sottoinsiemi di Ω gli eventi devono però soddisfare alle operazioni insiemistiche. Cioè agli eventi si applicano le operazioni insiemistiche \cup unione, \cap intersezione, il complemento, e la relazione insiemistica di inclusione \subset (\subseteq). Il risultato di tali operazioni deve essere ancora un evento. Più formalmente vediamo quali sono i sottoinsiemi che possiamo considerare eventi.

Definizione 2.1.4. *La famiglia \mathcal{A} di insiemi di sottoinsiemi di Ω tale che*

i) $\Omega \in \mathcal{A}$;

ii) se $A \in \mathcal{A}$, allora $A^C \in \mathcal{A}$;

iii) se $A_1, A_2, \dots \in \mathcal{A}$, allora $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{A}$,

è detta famiglia degli eventi di Ω ¹

Non ci soffermiamo sulle proprietà che deve avere \mathcal{A} , ma teniamo presente che nel caso lo spazio campionario Ω sia finito o numerabile, l'insieme \mathcal{A} contiene tutti i sottoinsiemi di Ω , nel caso in cui Ω sia l'insieme dei numeri reali (si pensi ad esempio all'esperimento consistente nel tempo di attesa di un cliente ad una fila in cui l'esito dell'esperimento può essere un qualunque numero $t > 0$) \mathcal{A} contiene tutti gli intervalli e le unioni di intervalli.

Da un'altro punto di vista possiamo osservare che quando fissiamo un evento A al quale siamo interessati, il realizzarsi o il non realizzarsi di A dipende da tutta

¹Una tale classe di sottoinsiemi di Ω è detta sigma algebra.

una serie di circostanze e che tutte queste circostanze possono essere viste come la struttura dell'esperimento. Il risultato dell'esperimento è detto realizzazione dell'esperimento e può oppure no fare parte di A . Noi non possiamo in generale prevedere con certezza l'esito di un esperimento ma possiamo solo fare due cose per altro importantissime: elencare le possibili realizzazioni di un esperimento (lo abbiamo fatto nei due esercizi precedenti) e dire con che probabilità si possono realizzare tali esiti (cosa a volte non semplice). Prima di vedere come viene assegnata la probabilità agli eventi possibili di un esperimento, vediamo prima quali sono le proprietà che essa deve soddisfare. La probabilità è una funzione che assegna ad ogni evento un numero compreso tra 0 e 1, in modo tale che più la probabilità dell'evento è alta più siamo indotti a credere che l'evento si realizzi. La probabilità gode di alcune proprietà che, se possono apparire bizzarre ad una prima e formale lettura, hanno invece un'immediata e naturale interpretazione pratica. Veniamo alla definizione precisa.

Definizione 2.1.5. *La probabilità è una funzione \mathbf{P} che deve soddisfare le seguenti proprietà per ogni evento $E \in \mathcal{A}$:*

$$i) \ 0 \leq \mathbf{P}(E) \leq 1,$$

$$ii) \ \mathbf{P}(\Omega) = 1;$$

iii) *per ogni successione E_1, E_2, \dots di eventi disgiunti, cioè tali che $E_i \cap E_j = \emptyset$ per ogni $i \neq j$, si ha*

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(E_i).$$

Le condizioni imposte alla probabilità sono dovute al fatto che trattando gli eventi come insiemi devono essere possibili le varie operazioni insiemistiche. Le tre condizioni sono detti *assiomi della probabilità*. Il primo assioma ci dice che la probabilità è sempre compresa tra 0 e 1. Se un evento ha probabilità 0 è detto *evento impossibile*, se ha probabilità 1 è detto *evento certo*. In ogni spazio di probabilità si ha sempre $\mathbf{P}(\emptyset) = 0$ e $\mathbf{P}(\Omega) = 1$. Il secondo assioma ci dice che è *normalizzata*, cioè il valore più grande è 1. Il terzo assioma è detto della *numerabile additività*. Si noti che la numerabile additività implica la finita additività della probabilità. Dati n eventi disgiunti vale la seguente relazione:

$$\mathbf{P}\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n \mathbf{P}(E_i),$$

che in particolare per $n = 2$ diviene $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$, se $A \cap B = \emptyset$. Questa proprietà ci dice che per calcolare la probabilità di eventi **disgiunti**, basta sommare

le probabilità dei singoli eventi. Tale assioma è abbastanza naturale da richiedere ad una probabilità. L'estensione al caso numerabile è necessaria per motivi tecnici.

La scelta dello spazio degli eventi elementari, o spazio campionario, Ω risulta quindi il primo passo nel formulare un modello probabilistico per l'esperimento che si sta considerando. In secondo luogo, accanto al concetto di spazio campionario, vi è quello di evento che può essere considerato come un particolare sottoinsieme dello spazio campionario. Questo perché in genere lo sperimentatore (cioè noi) è interessato non tanto al fatto se si è avuta una particolare realizzazione dell'esperimento, ma piuttosto se quella realizzazione appartiene ad un determinato insieme di realizzazioni. L'ulteriore passo è quello di assegnare ad ogni evento un *peso* che è chiamato probabilità. Il problema che sorge a questo punto è come assegnare la probabilità agli eventi di un esperimento.

Ricapitolando quanto descritto fin qui, quando dobbiamo descrivere o modellare un esperimento non deterministico ma casuale, tre sono le quantità da ben definire. Prima di tutto uno spazio Ω che descriva tutti i possibili esiti dell'esperimento, detto spazio degli eventi elementari o spazio campionario. In secondo luogo una classe di sottoinsiemi di Ω , indicata con \mathcal{A} , che contenga tutti gli eventi compatibili con la struttura di Ω , ai quali sia possibile associare una probabilità di realizzarsi. Infine, ma non ultima per importanza, una probabilità \mathbf{P} , in grado di quantificare con un peso, la probabilità che ogni evento di \mathcal{A} ha di realizzarsi. La tripletta $(\Omega, \mathcal{A}, \mathbf{P})$ è detta spazio di probabilità o modello probabilistico dell'esperimento.

Il problema non banale è come assegnare la probabilità agli eventi. Un esempio fondamentale è dato dal modello equiprobabile.

Se $\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$, se \mathcal{A} , la σ -algebra degli eventi, è data da tutti i sottoinsiemi di Ω e se $p_i = P(\{\omega_i\}) = \frac{1}{k}$, si ha il modello di equiprobabilità classico. Per questo modello la probabilità di un evento $A \subset \Omega$ di cardinalità k_A è $P(A) = \frac{k_A}{k}$ (regola classica nota come rapporto tra numero eventi elementari favorevoli e numero eventi elementari possibili). Questo modello, per quanto banale, è utilissimo e sta alla base delle applicazioni del calcolo combinatorio al calcolo delle probabilità.

I modelli generati dai processi stocastici non saranno semplici ma capire questo modello elementare sarà di aiuto.

2.2 Proprietà della probabilità

Dagli assiomi della probabilità si ricavano alcune importanti proprietà. Considerati due eventi $A, B \in \mathcal{A}$. Allora

1. Se $A \cap B = \emptyset$ allora $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$;

2. $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$;
3. $\mathbf{P}(B \setminus A) = \mathbf{P}(B \cap A^c) = \mathbf{P}(B) - \mathbf{P}(A)$;
4. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$;
5. Se $A \subset B$ allora $\mathbf{P}(A) \leq \mathbf{P}(B)$.

Dimostrazione. La prima relazione si dimostra osservando che $A \cup B = A \cup B \cup \emptyset \cup \emptyset \dots$. Gli insiemi sono tutti disgiunti e quindi per la terza proprietà che deve soddisfare la probabilità si ricava

$$\mathbf{P}(A \cup B) = \mathbf{P}(A \cup B \cup \emptyset \cup \emptyset \dots) = \mathbf{P}(A) + \mathbf{P}(B) + 0 + 0 + \dots,$$

da cui la relazione cercata. Per dimostrare la seconda relazione si può scrivere $\Omega = A \cup A^c$ e quindi per la prima proprietà della probabilità:

$$\mathbf{P}(A \cup A^c) = \mathbf{P}(\Omega) = 1.$$

Inoltre vale che $A \cap A^c = \emptyset$ e quindi per la proprietà 1 appena dimostrata vale che

$$\mathbf{P}(A \cup A^c) = \mathbf{P}(A) + \mathbf{P}(A^c)$$

Dalle due relazioni si ricava la tesi. Per dimostrare la terza proprietà scriviamo $B = (B \cap A) \cup (B \cap A^c)$ dove gli insiemi $B \cap A$ e $B \cap A^c$ sono disgiunti. Allora si può scrivere

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

Riordinando si ottiene la tesi. Si osservi che se $A \subset B$ allora si deduce che $\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A)$. La quarta proprietà si dimostra osservando che $A \cup B = A \cup (B \cap A^c)$. Allora, poiché A e $B \cap A^c$ sono disgiunti, si ha

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B \cap A^c) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(B \cap A),$$

dove l'ultima uguaglianza segue dalla proprietà tre appena dimostrata. Infine se $A \subset B$ allora $A \cap B = A$. Quindi per la terza proprietà si ha

$$0 \leq \mathbf{P}(B \cap A^c) = \mathbf{P}(B) - \mathbf{P}(A),$$

da cui la relazione cercata. □

2.2.1 Principio di inclusione ed esclusione

Sia $(\Omega, \mathcal{A}, \mathbf{P})$ uno spazio di probabilità. Il principio di inclusione ed esclusione permette di calcolare la probabilità dell'unione di n eventi.

Teorema 2.2.1. *Data una successione di eventi qualsiasi A_1, A_2, \dots , per ogni $n > 1$ vale la seguente relazione*

$$\mathbf{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbf{P}(A_i) - \sum_{i < k} \mathbf{P}(A_i \cap A_k) + \dots + (-)^{n-1} \mathbf{P}(A_1 \cap \dots \cap A_n). \quad (2.1)$$

Dimostrazione. La dimostrazione viene fatta col principio di induzione. Dimostriamo la relazione per $n = 2$. Dobbiamo dimostrare che

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cap A_2).$$

Possiamo scrivere $A_1 \cup A_2 = (A_1 \cap A_2^c) \cup (A_1 \cap A_2) \cup (A_1^c \cap A_2)$, dove i tre insiemi sono disgiunti. Vale quindi

$$\begin{aligned} \mathbf{P}(A_1 \cup A_2) &= \mathbf{P}(A_1 \cap A_2^c) + \mathbf{P}(A_1 \cap A_2) + \mathbf{P}(A_1^c \cap A_2) \\ &\quad + \mathbf{P}(A_1 \cap A_2) - \mathbf{P}(A_1 \cap A_2) \\ &= \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cap A_2). \end{aligned}$$

Supponiamo vera la relazione per n , cioè valga la (2.1), dimostriamo la relazione per $n + 1$. Possiamo scrivere

$$\mathbf{P}\left(\bigcup_{i=1}^{n+1} A_i\right) = \mathbf{P}\left(\bigcup_{i=1}^n A_i \cup A_{n+1}\right).$$

Applichiamo ora il principio di inclusione ed esclusione con $n = 2$ agli insiemi $\bigcup_{i=1}^n A_i$ e A_{n+1} . Abbiamo quindi

$$\mathbf{P}\left(\bigcup_{i=1}^{n+1} A_i\right) = \mathbf{P}\left(\bigcup_{i=1}^n A_i\right) + \mathbf{P}(A_{n+1}) - \mathbf{P}\left(\bigcup_{i=1}^n (A_i \cap A_{n+1})\right) \quad (2.2)$$

Per l'ipotesi di induzione il principio vale per n . Quindi applichiamo il principio agli n insiemi $A_i \cap A_{n+1}$. Abbiamo

$$\begin{aligned} \mathbf{P}\left(\bigcup_{i=1}^n (A_i \cap A_{n+1})\right) &= \sum_{i=1}^n \mathbf{P}(A_i \cap A_{n+1}) \\ &\quad - \sum_{i < k} \mathbf{P}((A_i \cap A_{n+1}) \cap (A_k \cap A_{n+1})) + \dots \\ &\quad + \dots (-)^{n-1} \mathbf{P}((A_1 \cap A_{n+1}) \cap \dots \cap (A_n \cap A_{n+1})). \end{aligned}$$

Applicando il principio anche al primo termine del secondo membro della (2.2) e unendo i risultati abbiamo la tesi. \square

2.3 Probabilità condizionata

Dato un qualunque esperimento, supponiamo di aver definito lo spazio di probabilità $(\Omega, \mathcal{A}, \mathbf{P})$. Ancora una volta non ha importanza per il momento come è stata definita la \mathbf{P} , ma l'importante è che soddisfi i tre assiomi della probabilità. Uno dei concetti più utili in teoria delle probabilità è il concetto di probabilità condizionata. Da un lato perché spesso siamo interessati a calcolare la probabilità di certi eventi quando già disponiamo di una parziale informazione. In questo senso il calcolo della probabilità è condizionato. In secondo luogo perché a volte per calcolare la probabilità di alcuni eventi è utile condizionare rispetto ad altri. Il concetto di probabilità condizionata, quindi prescinde da come è definita la probabilità. Quando si studiaranno le catene di Markov la probabilità condizionata sarà diffusamente utilizzata. Anzi, verranno calcolate quasi solo probabilità condizionate! Si sottolinea quindi l'importanza di capire bene tale definizione.

Definizione 2.3.1. *Dati due eventi A e B tali che $\mathbf{P}(B) > 0$, la quantità*

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

si chiama probabilità condizionata di A dato B .

A volte useremo la notazione \mathbf{P}_B per indicare la probabilità condizionata rispetto ad un evento B . Quindi con $\mathbf{P}_B(A)$ si intende la probabilità di A condizionata a B , o dato B . Dalla definizione di probabilità condizionata discende uno dei primi risultati non banali legati alla probabilità. Consiste nel fatto che il nome *probabilità condizionata* non è dato per caso alla \mathbf{P}_B ma essa è a tutti gli effetti una probabilità su Ω , nel senso che soddisfa gli assiomi della definizione 2.1.5. Si provi a risolvere il seguente esercizio.

Esercizio 2.3.2. Dato un evento B tale che $\mathbf{P}(B) > 0$, dimostrare che la funzione di probabilità $\mathbf{P}(\cdot|B) = \mathbf{P}_B(\cdot)$, definita per ogni evento A come $\mathbf{P}(A|B) = \mathbf{P}_B(A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$, è una probabilità su Ω , cioè soddisfa i tre assiomi della probabilità.

Intuitivamente possiamo pensare di avere cambiato lo spazio campionario campionario grazie alla conoscenza che si è realizzato l'evento B e relativamente al suo realizzarsi andiamo a calcolare le probabilità degli altri eventi. Si noti che la probabilità \mathbf{P} e la probabilità condizionata \mathbf{P}_B sono due modi diversi di assegnare la probabilità a tutti gli eventi A di Ω . Tali probabilità possono essere minori uguali o maggiori, come vedremo negli esempi.

Esercizio 2.3.3. Dato l'evento B tale che $0 < \mathbf{P}(B) < 1$, quanto vale $\mathbf{P}_B(B)$? Se $A \subset B$ quanto vale $\mathbf{P}_B(A)$? Se $A \cap B = \emptyset$, quanto vale $\mathbf{P}_B(A)$?

Dalla definizione di probabilità condizionata si ricava un'importante regola di calcolo. Dati due eventi A e B si ha:

$$\mathbf{P}(A \cap B) = \mathbf{P}(A|B)\mathbf{P}(B). \quad (2.3)$$

Questa relazione a differenza di quella della definizione vale anche se $\mathbf{P}(B) = 0$. La (2.3) si generalizza al caso di più di due eventi:

$$\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A|B \cap C)\mathbf{P}(B|C)\mathbf{P}(C). \quad (2.4)$$

Infatti: $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A|B \cap C)\mathbf{P}(B \cap C)$. Ma $\mathbf{P}(B \cap C) = \mathbf{P}(B|C)\mathbf{P}(C)$, da cui la (2.4). In modo analogo la si dimostra per un numero maggiore di 3 eventi. Vale a dire, vale la seguente formula

$$\begin{aligned} \mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_n) &= \mathbf{P}(A_1) \cdot \mathbf{P}(A_2|A_1) \cdot \mathbf{P}(A_3|A_2 \cap A_1) \cdot \dots \cdot \\ &\quad \mathbf{P}(A_{n-1}|A_{n-2} \cap \dots \cap A_1) \cdot \mathbf{P}(A_n|A_{n-1} \cap A_{n-2} \cap \dots \cap A_1) \end{aligned}$$

Si provi a interpretarla come se gli indici avessero il significato del tempo: se quello che accade in un determinato istante non è indipendente da quello che accade l'istante successivo, per calcolare la probabilità di ciò che è accaduto in un lasso di tempo devo tener conto di tutte le relazioni.

Il seguente teorema (detto *Teorema delle probabilità totali*) illustra una proprietà che useremo spesso nel corso.

Teorema 2.3.4. *Siano H_1, H_2, \dots, H_n n eventi tali che $H_i \cap H_j = \emptyset$ per ogni $i \neq j$ e $H_1 \cup \dots \cup H_n = \Omega$. Allora per ogni evento A*

$$\mathbf{P}(A) = \sum_{i=1}^n \mathbf{P}(A|H_i)\mathbf{P}(H_i).$$

Il teorema ci dice che l'evento A può verificarsi solo in concomitanza con qualche H_i . Inoltre la formula del teorema, nota come formula delle probabilità totali, è utile perché il calcolo delle probabilità condizionate è a volte più semplice del calcolo diretto di $\mathbf{P}(A)$.

2.4 Indipendenza

In generale la probabilità condizionata $\mathbf{P}(A|B)$ non è uguale alla probabilità non condizionata $\mathbf{P}(A)$. Quando invece tali probabilità coincidono possiamo concludere che la realizzazione dell'evento B non ha conseguenze sulla realizzazione dell'evento

A . In questo caso si dice che l'evento A è indipendente dall'evento B . Dalla (2.3) si ricava quindi

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

Questa relazione è simmetrica rispetto ad A e B quindi possiamo dire che quando A è indipendente da B anche B è indipendente da A .

Definizione 2.4.1. *Due eventi A e B sono detti indipendenti se vale la relazione*

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

Si noti che tale definizione vale anche se $\mathbf{P}(B) = 0$ o $\mathbf{P}(A) = 0$.

Esempio 2.4.2. Un esperimento consiste nell'estrarre una carta da un mazzo di carte. Verificare che gli eventi "la carta estratta è di fiori" e "la carta estratta è un re" sono indipendenti.

Il seguente esempio chiarisce il concetto di eventi indipendenti.

Esempio 2.4.3. Una compagnia di assicurazione ha diviso i suoi assicurati in due categorie. Nella prima categoria vi sono i soggetti più propensi ad avere incidenti, nella seconda quelli meno propensi. Denotiamo con H_1 e H_2 rispettivamente gli eventi, "il cliente appartenga alla categoria i -esima, con $i = 1, 2$ ". La proporzione di clienti nelle due categorie rispetto alla popolazione totale sta nel rapporto di 1 : 5. Questa informazione porta a valutare $P(H_1) = \frac{1}{6}$ e $P(H_2) = \frac{5}{6}$. Supponiamo che nella prima categoria la probabilità di avere almeno un incidente in un anno sia 0.6 mentre nella seconda 0.06. Supponiamo che per un cliente di una qualunque delle due categorie, l'aver un incidente in un certo anno sia indipendente dall'averlo gli anni successivi e sia indipendente dalla categoria a cui appartiene. La probabilità che un cliente dell'assicurazione scelto a caso abbia un incidente nel corso di un anno (evento A_1) è, per il teorema delle probabilità totali

$$\mathbf{P}(A_1) = \mathbf{P}(A_1|H_1)\mathbf{P}(H_1) + \mathbf{P}(A_1|H_2)\mathbf{P}(H_2),$$

Tale probabilità vale

$$\mathbf{P}(A_1) = 0.6\frac{1}{6} + 0.06\frac{5}{6} = 0.15.$$

Denotiamo con A_2 l'evento corrispondente al fatto che un cliente abbia un'incidente il secondo anno. La probabilità che un cliente abbia incidenti per due anni di seguito (evento $A_1 \cap A_2$) è

$$\mathbf{P}(A_1 \cap A_2) = \mathbf{P}(A_1 \cap A_2|H_1)\mathbf{P}(H_1) + \mathbf{P}(A_1 \cap A_2|H_2)\mathbf{P}(H_2).$$

Tale probabilità vale

$$\mathbf{P}(A_1 \cap A_2) = (0.6)^2 \frac{1}{6} + (0.06)^2 \frac{5}{6} = 0.063.$$

Andiamo ora a calcolare la probabilità che un cliente abbia un incidente l'anno successivo dato che ne ha già avuto uno l'anno corrente. Si tratta di calcolare

$$\mathbf{P}(A_2|A_1) = \frac{\mathbf{P}(A_1 \cap A_2)}{\mathbf{P}(A_1)} = \frac{0.063}{0.15} = 0.42.$$

Si noti la differenza tra $\mathbf{P}(A_1 \cap A_2)$ e $\mathbf{P}(A_2|A_1)$. La seconda tiene conto dell'informazione che l'anno prima il cliente ha avuto un'incidente, la prima no.

Il concetto di indipendenza si estende da due a più eventi.

Definizione 2.4.4. *Gli eventi A_1, A_2, \dots, A_n sono detti indipendenti se per ogni $1 \leq i < j < k < \dots \leq n$ valgono le seguenti relazioni*

$$\begin{aligned} \mathbf{P}(A_i \cap A_j) &= \mathbf{P}(A_i)\mathbf{P}(A_j), \\ \mathbf{P}(A_i \cap A_j \cap A_k) &= \mathbf{P}(A_i)\mathbf{P}(A_j)\mathbf{P}(A_k), \\ &\vdots \\ \mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_n) &= \mathbf{P}(A_1)\mathbf{P}(A_2) \dots \mathbf{P}(A_n). \end{aligned}$$

Si badi che l'indipendenza a due a due di tre eventi non implica l'indipendenza a tre a tre.

Esempio 2.4.5. Un esperimento consiste nel lancio di un dado due volte. Consideriamo gli eventi A , “esce un numero pari al primo lancio”, B , “esce un numero pari al secondo lancio” e C , “la somma dei due lanci è un numero pari”. Gli eventi A , B e C sono indipendenti a due a due ma non a tre a tre.

2.5 Variabili casuali discrete

In generale più che all'esperimento in se si è interessati ad una qualche conseguenza della possibile realizzazione dell'esperimento. Ad esempio il giocatore che ha scommesso è interessato alla quantità della sua perdita più che al risultato dell'esperimento che ha generato tale perdita. Gli oggetti che descrivono, come nell'esempio la perdita del giocatore, sono le variabili casuali che definiremo tra breve. Dato un esperimento casuale ed uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbf{P})$ che lo descrive, la variabile casuale, è una funzione definita sullo spazio Ω degli eventi elementari, che associa

ad ogni elemento di Ω uno ed un solo valore reale. Noi presenteremo e studieremo le proprietà di un particolare tipo di variabili casuali: quelle discrete. Per questo tipo di variabile casuale definiremo le quantità che saranno importanti per lo studio che seguirà, quali ad esempio la funzione di densità (discreta), la funzione di ripartizione, e le quantità di sintesi come il valore atteso e la varianza. Deve però essere tenuto presente che con le variabili discrete non si esaurisce la casistica delle variabili aleatorie. Per quanto riguarda però le applicazioni che intendiamo affrontare in questo corso, lo studio di questo tipo di variabili sarà esauriente.

Definizione 2.5.1. *Una variabile casuale (o aleatoria) discreta X è una funzione $X : \Omega \rightarrow \mathbb{R}$ che assegna ad ogni evento elementare ω un valore reale x tale che $\mathbf{P}(\{\omega : X(\omega) = \pm\infty\}) = 0$. Al variare di $\omega \in \Omega$ la variabile X può assumere solo un numero finito di valori o al più una infinità numerabile di valori. Indichiamo con x_1, x_2, \dots , i valori assunti dalla variabile casuale.*

Vale la pena osservare che in generale una variabile casuale deve essere tale da soddisfare anche la condizione che per ogni $x \in \mathbb{R}$ l'insieme $\{\omega : X(\omega) \leq x\}$ appartenga ad \mathcal{A} . Tale condizione è sempre soddisfatta da una variabile casuale discreta.

Gli insiemi del tipo $\{\omega : X(\omega) \in A\}$, dove A è un sottoinsieme di \mathbb{R} , saranno importantissimi in tutto quello che tratteremo più avanti, per cui è bene abituarsi al loro significato: sono sottoinsiemi in Ω costituiti da tutti gli eventi elementari tali che l'immagine $X(\omega)$ appartiene al sottoinsieme A di \mathbb{R} . Di solito utilizzeremo la scrittura più compatta

$$\{\omega : X(\omega) \in A\} = \{X \in A\}.$$

Con questa notazione possiamo riscrivere le condizioni date sopra come

$$\{X \leq x\} \in \mathcal{A} \quad \text{e} \quad \mathbf{P}(\{X = \pm\infty\}) = 0.$$

La differenza sostanziale tra le variabili aleatorie e le funzioni dell'analisi matematica è che non abbiamo un valore assunto dalla funzione per un determinato elemento del dominio dove la funzione è definita. O meglio abbiamo un tale valore solo quando osserviamo l'esito dell'esperimento, poiché in questo caso conosciamo la realizzazione che si identifica con un elemento dello spazio campionario, e quindi possiamo assegnare il valore osservato della variabile casuale. In generale delle variabili aleatorie possiamo solo conoscere i possibili valori che può assumere e la probabilità con cui possono essere assunti. Chiariamo il concetto con un esempio.

Esempio 2.5.2. Consideriamo l'esperimento consistente nel lanciare n volte una moneta. Lo spazio campionario Ω è costituito dalle 2^n n -uple del tipo (x_1, \dots, x_n)

dove $x_i = 0$ se esce testa al i -esimo lancio, e $x_i = 1$ se esce croce al i -esimo lancio. Supponiamo di essere interessati al numero di croci presenti negli n lanci. Denotiamo con X tale numero. Allora X è una variabile casuale che associa ad ogni evento elementare un numero. Ad esempio ad $\omega = (x_1, \dots, x_n)$ tale che $x_i = 0$ per ogni i , associamo il valore $X(\omega) = 0$. Si tratta di una variabile casuale in quanto l'evento $\{\omega : X(\omega) = \pm\infty\} = \emptyset$ e quindi $\mathbf{P}(\{\omega : X(\omega) = \pm\infty\}) = 0$. Inoltre si osservi che per ogni $x \in \mathbb{R}$ gli insiemi $\{\omega : X(\omega \leq x)\}$ appartengono ad \mathcal{A} che in questo caso è formata da tutti i sottoinsiemi di Ω . A priori, data la struttura dell'esperimento, noi sappiamo che la variabile X può assumere un qualunque valore intero tra 0 e n , ma fino a quando non lanciamo n monete e non osserviamo l'esito di ciascun lancio, e contiamo il numero di croci, non possiamo dire che valore assumerà la variabile X . Possiamo però calcolare la probabilità che assuma uno qualunque dei valori compresi tra 0 ed n .

Un altro esempio di variabile semplice ma molto importante è la variabile indicatore.

Esempio 2.5.3. Sia $E \in \mathcal{A}$ un evento. Denotiamo la variabile indicatore con I_E definita come

$$I_E(\omega) = \begin{cases} 0 & \text{se } \omega \in E \\ 1 & \text{se } \omega \notin E \end{cases}$$

L'insieme dei valori che la variabile casuale assume prende il nome di *range* della variabile casuale. Il range della variabile casuale è un sottoinsieme di \mathbb{R} . Su tale insieme viene indotta una probabilità dalla variabile X , che indichiamo con P_X , e che viene detta *distribuzione di probabilità* della variabile X , o *legge* della variabile X . Si introduce questa nuova probabilità perché si è interessati alla probabilità con cui la variabile casuale assume un dato valore. Da questo punto di vista siamo interessati, dato un insieme di valori $A \subset \mathbb{R}$ alla probabilità che $X \in A$. La distribuzione della variabile casuale X la si calcola partendo dalla probabilità definita su Ω

$$P_X(A) = \mathbf{P}(X \in A) = \mathbf{P}(X^{-1}(A)),$$

dove $X^{-1}(A)$ è l'evento che consiste di tutti gli eventi elementari ω tali che $X(\omega) \in A$. La distribuzione P_X di una variabile casuale discreta in genere è determinata dalla densità discreta della variabile X . Diamo quindi la definizione di densità discreta.

Definizione 2.5.4. Data la variabile casuale discreta X , la funzione $p_X : \mathbb{R} \rightarrow \mathbb{R}^+$ definita da

$$p_X(x) = \mathbf{P}(X = x),$$

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

La funzione p_X gode delle seguenti proprietà

- i) $p_X(x) = 0$ tranne che per un'infinità numerabile di valori di x , cioè i valori x_i , $i = 1, 2, \dots$, che sono i valori che assume la variabile casuale X ;
- ii) $\sum_{i=1}^{+\infty} p_X(x_i) = 1$.

Una funzione $p : \mathbb{R} \rightarrow \mathbb{R}^+$ che goda delle proprietà *i)* e *ii)* è detta densità discreta. Dalla conoscenza della densità discreta possiamo risalire alla distribuzione della v.c. X . Siamo cioè in grado di calcolare $P_X(A)$ per ogni sottoinsieme A di \mathbb{R} . Infatti osserviamo che possiamo scrivere

$$\{X \in A\} = \bigcup_{x_i \in A} \{X = x_i\}$$

e gli eventi a destra sono disgiunti, per cui

$$P_X(A) = \mathbf{P}(X \in A) = \sum_{x_i \in A} \mathbf{P}(X = x_i) = \sum_{x_i \in A} p_X(x_i).$$

Definiamo uno strumento molto importante sia dal punto di vista teorico che pratico (avrà un ruolo fondamentale nella parte di simulazione) per studiare le variabili casuali. Si tratta della funzione di ripartizione.

Definizione 2.5.5. *Data la variabile casuale X , per ogni numero reale x definiamo*

$$F_X(x) = \mathbf{P}(X \leq x).$$

F_X è detta *funzione di ripartizione della variabile casuale X* .

La funzione di ripartizione ha le seguenti proprietà che si deducono direttamente dalla definizione

Proposizione 2.5.6. *La funzione di ripartizione F_X della variabile casuale X gode delle seguenti proprietà:*

- a) *Se $x_1 < x_2$ allora $F_X(x_1) \leq F_X(x_2)$;*
- b) $\lim_{x \rightarrow -\infty} F_X(x) = 0$ e $\lim_{x \rightarrow +\infty} F_X(x) = 1$;
- c) F_X è continua a destra, cioè $\lim_{h \rightarrow 0} F_X(x+h) = F_X(x)$

Una variabile casuale è completamente determinata una volta che sono noti i valori che può assumere e la probabilità con cui assume questi valori. Per identificare una variabile casuale, dobbiamo conoscere, il suo range e la sua legge di probabilità. Queste due caratteristiche possono essere determinate conoscendo l'esperimento e la probabilità associata ad ogni possibile realizzazione dell'esperimento, di cui la variabile casuale descrive la parte che interessa lo sperimentatore. In generale conoscere le realizzazioni dell'esperimento è assai complicato, mentre stabilire un modello probabilistico partendo da una variabile casuale può essere più intuitivo. Noi adotteremo questo punto di vista. È anche importante osservare che la distribuzione di una variabile casuale e la sua funzione di ripartizione sono legate univocamente², nel senso che si può dimostrare che data una funzione di ripartizione, cioè una funzione che soddisfa le tre condizioni della proposizione 2.5.6, si riesce a trovare una variabile casuale che ammette come funzione di ripartizione quella di partenza.

La funzione di ripartizione di una variabile casuale discreta è una funzione definita su tutto l'asse reale \mathbb{R} crescente ed a salti. Ogni salto avviene nei valori x_i assunti dalla variabile aleatoria X ed ha ampiezza data da $p_X(x_i)$.

Esercizio 2.5.7. Scrivere e disegnare il grafico della funzione di ripartizione della variabile casuale X che assume i valori 0 e 1 con probabilità rispettivamente p e $1 - p$. Per fare il grafico adottare $p = 0.30$.

Esercizio 2.5.8. Scrivere e disegnare il grafico della funzione di ripartizione della variabile casuale X che assume i valori $\{x_1, x_2, \dots, x_n\}$ con probabilità rispettivamente $\{p_1, p_2, \dots, p_n\}$. Per fare il grafico adottare $\{0, 1, 2, \dots, 4\}$ e $p_i = 1/5$.

Tutte le questioni probabilistiche riguardanti le v.c. possono essere risolte tramite la funzione di ripartizione o la funzione di densità discreta. Ad esempio:

$$\mathbf{P}(a < X \leq b) = F(b) - F(a) = \sum_{a < x_i \leq b} p_X(x_i).$$

Si noti che l'insieme $\{a < X \leq b\}$ è un evento (per la definizione data di v.c) e quindi ha senso calcolare la probabilità indicata.

2.5.1 Valore atteso

A volte si è interessati, non alla v.c. nella sua interezza, ma a dei particolari valori che sintetizzano le informazioni contenute nella variabile casuale e che possono essere di immediata interpretazione. Questi sono ad esempio il valore atteso e la varianza.

²in realtà vi è univocità a meno di insiemi di probabilità nulla

Definizione 2.5.9. *Supponiamo che la variabile casuale X soddisfi la condizione*

$$\sum_{i=1}^{+\infty} |x_i| p_X(x_i) < +\infty.$$

Allora si chiama valore atteso della variabile casuale X la quantità

$$\mathbf{E}(X) = \sum_{i=1}^{+\infty} x_i p_X(x_i).$$

Possiamo introdurre altre quantità legate alle variabili casuali.

Definizione 2.5.10. *Supponiamo che la variabile casuale X soddisfi la condizione*

$$\sum_{i=1}^{+\infty} |x_i^k| p_X(x_i) < +\infty.$$

Allora si chiama momento di ordine k della variabile casuale X la quantità

$$\mathbf{E}(X^k) = \sum_{i=1}^{+\infty} x_i^k p_X(x_i).$$

La varianza è un'altra quantità importante nello studio delle variabili casuali poiché misura la dispersione media della variabile casuale attorno al suo valore atteso.

Definizione 2.5.11. *Supponiamo che la variabile casuale X soddisfi la condizione*

$$\sum_{i=1}^{+\infty} |x_i^2| p_X(x_i) < +\infty.$$

Allora si chiama varianza della variabile casuale X la quantità

$$\mathbf{Var}(X) = \sum_{i=1}^{+\infty} (x_i - \mathbf{E}(X))^2 p_X(x_i).$$

Il momento di ordine k di una variabile casuale e la varianza, sono casi particolari di valori attesi di funzioni di variabili casuali.

Definizione 2.5.12. *Sia X una v.c. e $f : \mathbb{R} \rightarrow \mathbb{R}$ una funzione tale che*

$$\sum_{i=1}^{+\infty} |f(x_i)| p_X(x_i) < +\infty.$$

Allora si chiama valore atteso della variabile casuale $Y = f(X)$ la quantità

$$\mathbf{E}(f(X)) = \sum_{i=1}^{+\infty} f(x_i) p_X(x_i).$$

Allora il momento di ordine k è il valore atteso di $Y = X^k$ con $f(x) = x^k$ e $\mathbf{Var}(X)$ è il valore atteso di $Y = (X - \mathbf{E}(X))^2$ con $f(x) = (x - \mathbf{E}(X))^2$. Ricordiamo che $\mathbf{E}(X)$ è una costante, un numero.

Esercizio 2.5.13. Calcolare il valore atteso e la varianza della variabile casuale X che assume i valori 0 e 1 con probabilità rispettivamente p e $1 - p$. Calcolare il valore numerico nel caso in cui $p = 0.30$.

Esercizio 2.5.14. Calcolare il valore atteso e la varianza della variabile casuale X che assume i valori $\{0, 1, 2, \dots, 4\}$ con $p_i = 1/5$, $i = 0, 1, \dots, 4$.

Si osservi che per calcolare la varianza è consigliabile utilizzare la formula

$$\mathbf{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2$$

Abbiamo visto che formalmente una variabile casuale è una funzione a valori reali definita sullo spazio campionario Ω . Comunque nella maggior parte dei casi lo spazio Ω non è descritto e la variabile casuale è definita attraverso la sua distribuzione. Nel caso di variabili casuali discrete questa è data dalla funzione di probabilità discreta, ovvero da tutti i valori $P(X = x)$ per ogni valore x per cui $P(X = x) > 0$. Vediamo quindi alcune particolari variabili casuali discrete.

Variabile casuale di Bernoulli

Consideriamo un esperimento che abbia solo due possibili esiti classificati come successo e insuccesso. Definiamo la variabile casuale X uguale a 1 se si è verificato un successo, uguale a 0 se si è verificato un insuccesso. Sia p , $0 < p < 1$, la probabilità di ottenere un successo. Allora la densità discreta di X è data da

$$\begin{aligned} p_X(0) &= \mathbf{P}(X = 0) = 1 - p, \\ p_X(1) &= \mathbf{P}(X = 1) = p. \end{aligned}$$

Si verifica inoltre che $\mathbf{E}(X) = p$ e $\mathbf{Var}(X) = p(1 - p)$. Indicheremo tale variabile aleatoria, detta di Bernoulli, con il simbolo $B(p)$.

Variabile casuale Binomiale

Consideriamo un esperimento che consiste in n prove, ciascuna delle prove con solo due possibili esiti classificati come successo ed insuccesso. Queste n prove supponiamo che siano indipendenti l'una dalle altre. Sia p la probabilità di successo in una prova. La variabile casuale X che rappresenta il numero di successi nelle n prove è

detta variabile casuale Binomiale ed indicata con il simbolo $\text{Bin}(n, p)$. Si tratta di una variabile che assume i valori $k = 0, 1, \dots, n$ e la cui densità discreta è data da

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Si verifica che $\mathbf{E}(X) = np$ e $\mathbf{Var}(X) = np(1-p)$.

Variabile casuale Ipergeometrica

Consideriamo un'urna che contenga M palle di cui K rosse e le restanti $M - K$ blu. Supponiamo di effettuare n estrazioni da tale urna senza rimettere le palle estratte nell'urna. La variabile casuale X che rappresenta il numero di palle rosse tra le n estratte si chiama variabile casuale ipergeometrica. Essa assume tutti i valori interi tra $\max(0, n - (M - K))$ e $\min(K, n)$. La densità discreta è data da

$$P_X(x) = \frac{\binom{K}{x} \binom{M-K}{n-x}}{\binom{M}{n}}, \quad \max(0, n - (M - K)) \leq x \leq \min(K, n).$$

Si verifica che $\mathbf{E}(X) = n \frac{K}{M}$ e $\mathbf{Var}(X) = n \frac{K}{M} \frac{M-K}{M} \frac{M-n}{M-1}$.

Variabile casuale Geometrica

Consideriamo l'esperimento che consiste nel replicare prove indipendenti, ciascuna delle quali ha solo due possibili esiti (successo e insuccesso), fino al realizzarsi del primo successo. Supponiamo che in ciascuna prova la probabilità di ottenere un successo sia p . La variabile X che rappresenta il numero di prove da effettuare prima del primo successo si chiama variabile casuale Geometrica, la si indica con $\text{Geom}(p)$ e la densità discreta è data da

$$p_X(n) = (1-p)^{n-1} p, \quad n = 1, 2, \dots$$

Si tratta di una variabile casuale che può assumere un'infinità numerabile di valori. Si verifica che $\mathbf{E}(X) = \frac{1}{p}$, mentre $\mathbf{Var}(X) = \frac{1-p}{p^2}$. Un altro modo di definirla è quello di considerare la variabile casuale Y che conta il numero di insuccessi prima del primo successo. Si ha che $Y = X - 1$ e

$$p_Y(n) = (1-p)^n p, \quad n = 0, 1, 2, \dots$$

In questo caso $\mathbf{E}(Y) = \frac{1-p}{p}$, mentre $\mathbf{Var}(Y) = \frac{1-p}{p^2}$.

2.6 Vettori aleatori

L'argomento di questo testo sono i processi stocastici, in particolare le catene di Markov e i processi di Poisson. Vedremo la definizione precisa di tali processi nei prossimi capitoli. In questa fase possiamo però dire che un processo stocastico si può vedere come una famiglia di v.c. i cui elementi sono variabili casuali in generale non indipendenti. Vi sarà dipendenza di diverso tipo tra le varie variabili della famiglia, ovvero del processo stocastico. In altri corsi di statistica abbiamo già incontrato famiglie di variabili casuali, in particolare una successione di variabili v.c. indipendenti e identicamente distribuite (i.i.d.). Sia $\{X_i\}$, $i = 1, \dots, n$ una successione di v.c. i.i.d.

2.6.1 Vettori aleatori discreti

Quando si studiano le proprietà che riguardano due o più variabili aleatorie si introducono delle quantità che descrivono il comportamento delle variabili congiuntamente. Supponiamo che X e Y siano due variabili casuali discrete. Denotiamo con x_1, x_2, \dots i valori che può assumere la variabile X e analogamente con y_1, y_2, \dots i valori che può assumere la variabile Y . Abbiamo visto che tali valori possono essere al più solo un'infinità numerabile.

Definizione 2.6.1. *La densità di probabilità congiunta (discreta) delle variabili X e Y è definita come*

$$p_{XY}(x, y) = \mathbf{P}(X = x, Y = y),$$

per ogni $x, y \in \mathbb{R}$.

Abbiamo indicato semplicemente con $\{X = x, Y = y\}$ l'evento $\{X = x\} \cap \{Y = y\}$. Osserviamo che solo per un'infinità al più numerabile di punti x e y si ha $p_{XY}(x, y) > 0$. Per non appesantire le notazioni indicheremo semplicemente con $p(x, y)$ la densità congiunta nel punto (x, y) . Dalla densità congiunta si ricavano le densità marginali, cioè le densità delle variabili X e Y considerate singolarmente. Precisamente

$$p_X(x) = \sum_{y:p(x,y)>0} p(x, y)$$

e analogamente

$$p_Y(y) = \sum_{x:p(x,y)>0} p(x, y).$$

La dimostrazione è immediata. Infatti possiamo scrivere

$$\begin{aligned} p_X(x) &= \mathbf{P}(X = x) = \sum_{y:p(x,y)>0} \mathbf{P}(X = x, Y = y) \\ &= \sum_{y:p(x,y)>0} p(x, y). \end{aligned}$$

Due variabili casuali discrete X e Y sono dette indipendenti se, e solo se, si verifica che

$$\mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x)\mathbf{P}(Y = y)$$

per ogni $x, y \in \mathbb{R}$. Questo significa che se due variabili casuali sono indipendenti allora la densità congiunta è data dal prodotto delle marginali, cioè

$$p(x, y) = p_X(x)p_Y(y)$$

per ogni $x, y \in \mathbb{R}$. Date due variabili casuali discrete X e Y e una funzione $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, si chiama valore atteso congiunto della funzione aleatoria $g(X, Y)$ la quantità

$$\mathbf{E}(g(X, Y)) = \sum_{x,y} g(x, y)p(x, y).$$

Il valore atteso essendo definito come una somma di valori gode di importanti proprietà matematiche, che si riflettono poi nelle applicazioni. Una proprietà molto importante del valore atteso è la linearità.

Proposizione 2.6.2. *Siano X e Y due variabili casuali, aventi range rispettivamente R_X e R_Y e distribuzione p_X e p_Y . Si verifica che*

$$\mathbf{E}(aX + bY) = a\mathbf{E}(X) + b\mathbf{E}(Y),$$

dove a e b sono due costanti.

Provare a fare la dimostrazione per esercizio. Basta partire dalla scrittura

$$\mathbf{E}(aX + bY) = \sum_x \sum_y (ax + by)p(x, y)$$

e sviluppare la somma. Poi con le proprietà della distribuzione congiunta si arriva alla conclusione.

Date n variabili casuali discrete X_1, X_2, \dots, X_n , si definisce la funzione di densità congiunta come

$$p(x_1, x_2, \dots, x_n) = \mathbf{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n).$$

Le densità marginali si ricavano in modo ovvio. Inoltre, per il valore atteso verifica che

$$\mathbf{E} \left(\sum_{i=1}^n a_i X_i \right) = \sum_{i=1}^n a_i \mathbf{E}(X_i),$$

dove a_i sono costanti. Se X e Y sono due variabili discrete si definisce la densità condizionata di X dato $Y = y$. Se $\mathbf{P}(Y = y) > 0$, essa è data da

$$p_{X|y}(x|y) = \mathbf{P}(X = x|Y = y) = \frac{\mathbf{P}(X = x, Y = y)}{\mathbf{P}(Y = y)}.$$

La densità condizionata si utilizza per calcolare il valore atteso condizionato di una variabile rispetto al valore assunto da un'altra. Siano X e Y due variabili casuali discrete, aventi densità congiunta $p(x, y)$ e densità marginali $p_X(x)$ e $p_Y(y)$. Abbiamo la seguente

Definizione 2.6.3. Dato l'evento $\{X = x\}$ tale che $\mathbf{P}(X = x) > 0$, si definisce valore atteso condizionato della variabile casuale Y , dato l'evento $\{X = x\}$, la quantità

$$\mathbf{E}(Y|X = x) = \sum_y y p_{Y|x}(y|x)$$

Si tratta, al variare di x , della funzione di regressione. Cioè del valor medio della variabile Y , dato che la variabile X ha assunto il valore x .

Vediamo ora un esempio di una famiglia di variabili casuali discrete che ritroveremo quando daremo la definizione di catena di Markov.

Esercizio 2.6.4. Consideriamo una particella che si possa muovere sull'asse delle ascisse ad ogni istante di una unità a destra o a sinistra secondo lo schema seguente. La posizione della particella all'istante n sia descritta da una variabile casuale X_n definita ricorsivamente da $X_n = X_{n-1} + Z_n$ dove la successione di variabili Z_n , $n = 1, 2, \dots$ costituisce una successione di variabili casuali indipendenti e identicamente distribuite che assumono i valori -1 e $+1$ con probabilità rispettivamente di $\frac{1}{2}$. Si supponga che la particella all'istante 0 parta dal punto 0, valga cioè $\mathbf{P}(X_0 = 0) = 1$. Si considerino le variabili casuali X_1 e X_2 . Trovare la funzione di densità congiunta e la funzione di densità marginale. Trovare la funzione di densità condizionata di X_2 dato l'evento $X_1 = 1$. Le variabili casuali X_1 e X_2 sono indipendenti? Calcolare il valore atteso di X_2 condizionato a $X_1 = 1$.

Esempio 2.6.5. Facciamo riferimento all'esempio 2.6.4. Vogliamo determinare $\mathbf{P}(X_n = r)$ per ogni n e per ogni r . Innanzitutto osserviamo che fissato n , r può assumere solo i valori $(-n, \dots, -2, 0, 2, \dots, n)$ per n pari e $(-n, \dots, -1, 1, \dots, n)$

per n dispari. Quindi l'insieme dei valori che ciascuna variabile X_n può assumere è finito. Per determinare $\mathbf{P}(X_n = r)$ possiamo scrivere $X_n = Z_1 + \dots + Z_n$. Tutti i cammini possibili di n passi sono individuati da una serie di $+1$ e -1 , e sono 2^n , per cui la probabilità che si realizzi uno dei possibili cammini è $\frac{1}{2^n}$. Indichiamo con a il numero di $+1$ e con b il numero di -1 . I cammini che partono da 0 e arrivano in r in n passi sono tali che $a + b = n$ e $a - b = r$. Tali cammini sono individuati ogni volta che scelgo gli a istanti in cui abbiamo $+1$. Tali cammini sono $\binom{n}{a} = \binom{n}{b}$. Dal sistema ricaviamo $a = \frac{n+r}{2}$, se $\frac{n+r}{2}$ è intero. Ricaviamo quindi

$$\mathbf{P}(X_n = r) = \begin{cases} \binom{n}{\frac{n+r}{2}} 2^{-n} & r = 0, \pm 2 \dots \pm n, \text{ se } n \text{ è pari} \\ \binom{n}{\frac{n+r}{2}} 2^{-n} & r = \pm 1, \pm 3 \dots \pm n, \text{ se } n \text{ è dispari} \\ 0 & \text{altrimenti.} \end{cases}$$

Si osservi che ciascuna X_n è una variabile casuale discreta e che per n tendente all'infinito l'insieme dei valori che tale variabile può assumere tende all'infinito, pur rimanendo numerabile.

Capitolo 3

Catene di Markov

3.1 Processi Stocastici

Il concetto di processo stocastico è molto generale. Vale infatti la seguente definizione.

Definizione 3.1.1. *Una famiglia di variabili aleatorie*

$$X = \{X_t : t \in T\}$$

definite su uno spazio di probabilità $(\Omega, \mathcal{A}, \mathbf{P})$, è detto processo stocastico.

I processi stocastici in cui l'insieme T è discreto sono detti processi in tempo discreto. Per i processi in tempo discreto potremo quindi utilizzare la notazione $\{X_n : n \in \mathbb{N}\}$. Se invece l'insieme T è continuo (ad esempio un sottoinsieme di \mathbb{R} , o lo stesso \mathbb{R} , o \mathbb{R}^+), la famiglia di variabili aleatorie è detto processo in tempo continuo. I processi stocastici si distinguono, oltre che per il tipo di insieme in cui sono indicizzati anche per i valori che le variabili casuali possono assumere. Se possono assumere un numero di valori al più numerabile, il processo è detto discreto (se il tempo è discreto) o a salti (se il tempo è continuo). Noi studieremo in questo capitolo e nei prossimi, una famiglia di processi stocastici in cui l'insieme T è discreto e i valori che possono assumere le variabili casuali sono un numero finito o al più numerabile. Nel Capitolo 8 studieremo invece processi in cui l'insieme T è continuo e i valori che possono assumere le variabili casuali sono al più un'infinità numerabile.

3.2 Notazioni di base

Consideriamo un processo $\{X_n\}$, $n = 1, 2, \dots$, tale che ogni variabile X_n possa assumere solo un numero finito di valori o al più un'infinità numerabile di valori.

Non è detto che ogni variabile X_n debba assumere gli stessi valori. L'insieme di tutti i valori che tutte le variabili X_n possono assumere verrà denotato con \mathcal{S} ed è detto spazio degli stati. È sempre possibile rappresentare \mathcal{S} come

$$\mathcal{S} = \{0, 1, 2, \dots\}$$

o come

$$\mathcal{S} = \{0, \pm 1, \pm 2, \dots\},$$

quando \mathcal{S} è numerabile, ovvero con

$$\mathcal{S} = \{1, 2, \dots, d\}$$

se lo spazio degli stati è finito. A volte si parte dallo stato 0 invece che dallo stato 1. Se $X_n = i$ significa che il processo all'istante n si trova nello stato i . Denotiamo con $P_{ij}^n = \mathbf{P}(X_{n+1} = j | X_n = i)$. P_{ij}^n è la probabilità che il processo si trovi nello stato j all'istante $n + 1$, dato che si trovava nello stato i all'istante n .

Definizione 3.2.1. *Il processo $\{X_n\}$, $n = 1, 2, \dots$, avente spazio degli stati \mathcal{S} è detto catena di Markov se*

$$\mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbf{P}(X_{n+1} = j | X_n = i) = P_{ij}^n, \quad (3.1)$$

ogni volta che le probabilità scritte hanno senso, per ogni stato $i_0, i_1, \dots, i_{n-1}, i, j$, e per ogni $n \geq 0$. Le probabilità P_{ij}^n sono dette probabilità di transizione ad un passo.

L'interpretazione che diamo alla definizione 3.2.1 è che, per ogni istante n , la legge condizionata dello stato futuro X_{n+1} , dato il presente X_n e tutto il passato X_{n-1}, \dots, X_1, X_0 , dipende solo dal presente ed è indipendente dal passato. Le P_{ij}^n rappresentano le probabilità di tali transizioni dallo stato i allo stato j . Si noti che dipendono anche dall'istante n . Noi studieremo solo una classe particolare di catene di Markov, quelle omogenee.

Definizione 3.2.2. *Una catena di Markov per cui*

$$\mathbf{P}(X_{n+1} = j | X_n = i) = P_{ij},$$

per ogni $n \geq 0$ è detta catena di Markov omogenea.

Osserviamo che per una catena omogenea possiamo sempre esprimere le probabilità di transizione facendo riferimento all'istante iniziale e a quello immediatamente successivo

$$P_{ij} = \mathbf{P}(X_1 = j | X_0 = i).$$

In una catena di Markov omogenea le probabilità di transizione dallo stato i allo stato j in un passo non dipendono dall'istante che si sta considerando o in cui tale transizione avviene. Per le proprietà della probabilità condizionata sono soddisfatte le seguenti condizioni

$$P_{ij} \geq 0, \quad i, j \geq 0, \quad \sum_{j \in \mathcal{S}} P_{ij} = 1.$$

Le probabilità di transizione ad un passo formano la matrice P che identifica e specifica la particolare catena di Markov

$$P = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \dots & P_{0j} & \dots \\ P_{10} & P_{11} & P_{12} & \dots & P_{1j} & \dots \\ P_{20} & P_{21} & P_{22} & \dots & P_{2j} & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ P_{i0} & P_{i1} & P_{i2} & \dots & P_{ij} & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \end{bmatrix}.$$

Per ogni istante $n \geq 1$, l'evoluzione della catena è completamente determinato dalla matrice di transizione. La matrice P è detta matrice di transizione ad un passo, ed è tale che ogni elemento è maggiore o uguale a zero, e la somma degli elementi di ogni riga è uguale a uno. Le matrici con queste due proprietà sono dette matrici stocastiche. Nel caso in cui lo spazio degli stati è finito $\mathcal{S} = \{1, 2, \dots, d\}$ la matrice di transizione assume la forma di una matrice $d \times d$

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & \dots & P_{1d} \\ P_{21} & P_{22} & P_{23} & \dots & P_{2d} \\ P_{31} & P_{32} & P_{33} & \dots & P_{3d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{d1} & P_{d2} & P_{d3} & \dots & P_{dd} \end{bmatrix}$$

il cui generico elemento P_{ij} ha il significato di $\mathbf{P}(X_{n+1} = j | X_n = i)$ per ogni i, j ed n e rappresenta la probabilità che la catena ad un istante si trovi nello stato j dato che all'istante precedente si trovava nello stato i .

3.3 Esempi fondamentali

In questa sezione presentiamo alcuni importanti esempi di fenomeni che possono essere descritti da una catena di Markov. Ritroveremo questi esempi nel corso di tutto il libro dove cercheremo di rispondere alle domande naturali legate ad ogni fenomeno presentato.

3.3.1 La rovina del giocatore

Consideriamo un giocatore che ad ogni partita può vincere un euro con probabilità $p = 0.4$ e perdere un euro con probabilità $q = 1 - p = 0.6$. Supponiamo inoltre che il giocatore smetta di giocare quando la sua fortuna ha raggiunto quota N oppure quando ha perso tutto e quindi la sua fortuna ammonta a 0. Denotiamo con X_n l'ammontare della fortuna del giocatore all'istante n , cioè dopo aver giocato n partite. La famiglia di v.a. X_n è una catena di Markov. Infatti se la fortuna del giocatore all'istante n è i , $X_n = i$, con $0 < i < N$ allora per ogni possibile situazione della fortuna agli istanti precedenti, denotate rispettivamente con $i_{n-1}, i_{n-2}, \dots, i_1, i_0$ si ha

$$P(X_{n+1} = i + 1 | X_n = i, X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots, X_1 = i_1, X_0 = i_0) = 0.4.$$

Abbiamo quindi dimostrato la (3.1): la fortuna del giocatore all'istante $n + 1$ data tutta la storia passata dipende solo dalla fortuna all'istante precedente. Le probabilità di transizione risultano $P_{i,i+1} = 0.4$, $P_{i,i-1} = 0.6$ per $0 < i < N$, $P_{0,0} = 1$ e $P_{N,N} = 1$. La matrice di transizione ad un passo per $N = 5$ risulta:

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{bmatrix} & 0 & 1 & 2 & 3 & 4 & 5 \\ \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right. \end{bmatrix}$$

Si noti che lo spazio degli stati è costituito da $N + 1$ elementi: $\mathcal{S} = \{0, 1, 2, \dots, N\}$, per $N = 5$. Gli elementi $P_{00} = 1$ e $P_{NN} = 1$ ci dicono che se la catena ad un certo istante $n^* \geq 0$ raggiunge lo stato 0 o lo stato N , poi vi rimane per sempre, perché $\mathbf{P}(X_{n+1} = 0 | X_n = 0) = 1$ e $\mathbf{P}(X_{n+1} = N | X_n = N) = 1$ per ogni $n \geq n^*$.

3.3.2 La camminata aleatoria

Un importante esempio di catena di Markov che generalizza quello della rovina del giocatore è costituito dalla passeggiata aleatoria. Consideriamo una particella che può muoversi sull'asse delle ascisse secondo il seguente schema: ad ogni istante n compie un passo verso destra o verso sinistra secondo il risultato di una variabile casuale Z_n che assume valore $+1$ con probabilità $p > 0$ e valore -1 con probabilità q , con $p + q = 1$. Supponiamo che le variabili casuali $Z_n, n = 1, 2, \dots$ siano indipendenti

r . La matrice di transizione risulta (ad esempio per $r = 6$)

$$P = \begin{bmatrix} 0 & 1 & & & & & \\ q & 0 & p & & & & \\ & q & 0 & p & & & \\ & & q & 0 & p & & \\ & & & q & 0 & p & \\ & & & & q & 0 & p \\ & & & & & 1 & 0 \end{bmatrix},$$

con $p > 0$, $q > 0$ e $p + q = 1$. La passeggiata aleatoria con barriera riflettente può anche essere utilizzata come modello per il seguente gioco. Due giocatori partono con capitale rispettivamente a e b . Ad ogni istante giocano un'unità di tale capitale, ed il giocatore A vince tale unità con probabilità p , mentre il giocatore B vince tale unità con probabilità q . Quando però uno dei due giocatori perde l'ultima unità di capitale, il gioco non termina ma il giocatore che ha perduto tale ultima unità la conserva e continua a giocare. Se X_n indica il capitale del giocatore A all'istante n , la famiglia $\{X_n\}$ è una catena di Markov con spazio degli stati $\mathcal{S} = \{1, 2, \dots, a+b-1\}$ e la cui matrice di transizione, nel caso in cui ad esempio $a = 4$ e $b = 4$, è data da

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

Non necessariamente per la passeggiata aleatoria le barriere assorbenti devono essere 2. Ad esempio vi può essere una sola barriera in 0. In questo caso la matrice di transizione si modifica in modo opportuno, e avrà infinite righe e colonne.

Altre generalizzazioni

La passeggiata aleatoria ammette anche altre generalizzazioni. Ad esempio le variabili Z_n possono assumere i valori $-1, 0, +1$ rispettivamente con probabilità rispettivamente p, r, q , con $p+r+q = 1$ e $p > 0$. Un'altra possibile generalizzazione si ha se le variabili Z_n hanno una qualunque distribuzione discreta. Se la particella continua a muoversi secondo l'equazione (3.2) la camminata è detta senza restrizioni. Se invece consideriamo la camminata soggetta a qualche restrizione sul suo cammino, come ad esempio una barriera a destra e o a sinistra, la camminata è detta con barriere. Ad esempio supponiamo che una particella parta dall'origine e possa muoversi solo

fino ad una distanza a dall'origine verso destra e b verso sinistra, se il suo moto cessa quando la particella raggiunge o il punto a o il punto $-b$, i punti a e $-b$ sono detti barriere assorbenti.

Esempio 3.3.1. Consideriamo una compagnia di assicurazione che parte ad un istante 0 con un capitale X_0 . Durante i periodi $1, 2, \dots$ riceve i premi Y_1, Y_2, \dots e paga i diritti W_1, W_2, \dots . Il capitale della compagnia al tempo n sarà dunque

$$X_n = X_0 + (Y_1 - W_1) + \dots + (Y_n - W_n).$$

Se ad un qualche istante n il capitale è $X_n \leq 0$, la compagnia è rovinata e smette di operare. Se poniamo $Z_r = Y_r - W_r$ per ogni $r = 1, 2, \dots$ il capitale della compagnia si comporta come una passeggiata aleatoria che parte da X_0 ed ad ogni istante compie un salto pari a Z_r , per $r = 1, 2, \dots$. Le equazioni che definiscono il processo sono per $n = 1, 2, \dots$

$$X_n = \begin{cases} X_{n-1} + Z_n, & \text{se } X_{n-1} > 0, X_{n-1} + Z_n > 0, \\ 0 & \text{altrimenti.} \end{cases}$$

Un problema molto interessante è la determinazione della probabilità di rovina per un certo valore X_0 e per una data distribuzione dei premi e dei pagamenti dei rischi. Tale rovina corrisponde al calcolo della probabilità di assorbimento in 0 .

Il problema della rovina del giocatore può essere visto come una camminata aleatoria con due barriere assorbenti. Il prossimo esempio è un modello alternativo a quello presentato nel paragrafo 3.3.1.

Esempio 3.3.2. Consideriamo due giocatori A e B che partono rispettivamente con un capitale a e b . Il gioco consiste in una successione di prove in cui con probabilità p A vince una unità del capitale di B , mentre con probabilità q A perde una unità del suo capitale in favore di B , $p + q = 1$. Il gioco termina quando A o B terminano il loro capitale. Denotiamo con X_n il guadagno del giocatore A all'istante n . Il processo $\{X_n\}$ è una catena di Markov avente spazio degli stati $\mathcal{S} = \{-a, \dots, 0, \dots, b\}$. Se $-a < X_n < b$ possiamo scrivere

$$X_n = Z_1 + Z_2 + \dots + Z_n,$$

dove le variabili Z_i , $i = 1, 2, \dots$, sono indipendenti con distribuzione $\mathbf{P}(Z_i = +1) = p$, $\mathbf{P}(Z_i = -1) = q$. Se invece per qualche n , $X_n = -a$ o $X_n = b$ significa che il giocatore A ha perso tutto il suo capitale o il giocatore A ha vinto tutto il capitale che possedeva il giocatore B . In questo caso il gioco termina e possiamo immaginare che

il processo rimanga nello stato $-a$ o nello stato b per sempre. Si dice che la catena è assorbita e gli stati $-a$ e b sono detti stati assorbenti. La matrice di transizione risulta

$$P = \begin{bmatrix} 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ q & 0 & p & 0 & \cdot & \cdot & \cdot \\ 0 & q & 0 & p & \cdot & \cdot & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdot & \cdot & \cdot & q & 0 & p \\ 0 & \cdot & \cdot & \cdot & 0 & 0 & 1 \end{bmatrix}.$$

Dimostreremo più avanti che la probabilità che una catena di questo tipo sia assorbita in uno dei due stati assorbenti è pari a uno. Calcolare la probabilità che la catena sia assorbita in a significa calcolare la probabilità che il giocatore A sia rovinato. Possiamo ritrovare l'esempio 3.3.1 se introduciamo il seguente processo. Denotiamo con Y_n il capitale del giocatore A all'istante n . Abbiamo

$$Y_n = a + Z_1 + Z_2 + \dots + Z_n,$$

dove le variabili casuali Z_i sono definite come sopra. Anche il processo $\{Y_n\}$ è una catena di Markov. In questo caso lo spazio degli stati è $\mathcal{S} = \{0, 1, 2, \dots, N\}$, dove $N = a + b$. Gli stati assorbenti sono 0, che corrisponde all'evento in cui il giocatore A perde, e N che corrisponde all'evento in cui il giocatore A vince. La matrice di transizione è la matrice P con $N + 1$ righe e colonne.

3.3.3 Modello di Eherenfest di diffusione

Supponiamo di avere due scatole A e B in cui sono contenute d molecole numerate da 1 a d , alcune nella scatola A , le rimanenti nella scatola B . Ad ogni istante viene estratto un numero compreso tra 1 e d e la molecola contraddistinta dall'etichetta con il numero estratto, viene spostata dalla scatola in cui si trova all'altra. Questa operazione di estrazione può proseguire indefinitamente e si suppone che le estrazioni successive siano indipendenti. Denotiamo con X_n il numero di molecole nella scatola A dopo l' n -esima estrazione. La famiglia $\{X_n\}$ è una catena di Markov, con spazio degli stati $\mathcal{S} = \{0, 1, \dots, d\}$. Infatti per calcolare in che stato sarà la catena all'istante $n + 1$ la sola informazione di cui si ha bisogno, conoscendo tutta la storia $X_n, X_{n-1}, \dots, X_1, X_0$, è lo stato attuale, all'istante n . Si ha

$$P(X_{n+1} = i + 1 | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) = P_{i+1,i} = \frac{d - i}{d}.$$

poiché per incrementare di una unità le molecole nella scatola A occorre estrarre una delle $d - i$ molecole che si trova nell'urna B . Allo stesso modo si ha che $P_{i-1,i} = \frac{i}{d}$

per ogni $0 < i < d$. Se invece la catena all'istante n si trova nello stato $i = 0$, con probabilità 1 l'istante successivo si troverà nello stato 1. Se si trova nello stato d con probabilità 1 sarà nello stato $d - 1$, vale a dire $P_{0,1} = 1$ e $P_{d,d-1} = 1$. La matrice di transizione risulta pertanto

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ \frac{1}{d} & 0 & \frac{d-1}{d} & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & \frac{2}{d} & 0 & \frac{d-2}{d} & \cdot & \cdot & \cdot & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdot & \cdot & \cdot & \frac{d-1}{d} & 0 & \frac{1}{d} \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 1 & 0 \end{bmatrix}.$$

Ad esempio se $d = 6$ la matrice di transizione è

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & \frac{5}{6} & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{6} & 0 & \frac{4}{6} & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{6} & 0 & \frac{3}{6} & 0 & 0 \\ 0 & 0 & 0 & \frac{4}{6} & 0 & \frac{2}{6} & 0 \\ 0 & 0 & 0 & 0 & \frac{5}{6} & 0 & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

dove non c'è nulla si intende che la probabilità è zero.

Tale modello può essere visto come uno schema molto semplice di diffusione ad esempio del calore da un mezzo ad un'altro. Lo stesso modello può essere utilizzato per descrivere la passeggiata aleatoria di una particella soggetta ad un forza attrattiva simmetrica direttamente proporzionale alla distanza della particella dal punto di applicazione della forza. Ad esempio supponiamo che la forza attrattiva si trovi nel punto $\frac{d}{2}$. Se la particella si trova nella posizione k allora si muoverà con maggiore probabilità verso destra se $k < \frac{d}{2}$, con maggiore probabilità verso sinistra se $k > \frac{d}{2}$.

3.3.4 Modello per sequenza di successi

Consideriamo una successione di prove di Bernoulli in cui $p > 0$ è la probabilità di successo in una prova e $q = 1 - p$ la probabilità di insuccesso. Definiamo la famiglia di variabili casuali $X_n = k$ solo se l'ultimo insuccesso nella successione di prove si è registrato $n - k$ prove fa, con $k = 0, 1, 2, \dots$. Per convenzione la prova numero 0 è

un insuccesso. La matrice di transizione risulta

$$P = \begin{bmatrix} q & p & & & & \\ q & 0 & p & & & \\ q & 0 & 0 & p & & \\ q & 0 & 0 & 0 & p & \\ \vdots & \cdot & \cdot & \cdot & \cdot & \ddots \\ q & 0 & \cdot & \cdot & \cdot & p \\ \vdots & \cdot & \cdot & \cdot & \cdot & \ddots \end{bmatrix}.$$

Per tale modello possiamo calcolare facilmente la matrice di transizione di ordine 2. Precisamente abbiamo

$$P^{(2)} = \begin{bmatrix} q & qp & p^2 & & & \\ q & qp & 0 & p^2 & & \\ q & qp & 0 & 0 & p^2 & \\ q & 0 & 0 & 0 & p & \\ \vdots & \cdot & \cdot & \cdot & \cdot & \ddots \\ q & qp & \cdot & \cdot & \cdot & p^2 \\ \vdots & \cdot & \cdot & \cdot & \cdot & \ddots \end{bmatrix}.$$

In generale, gli elementi della matrice di transizione in n passi sono dati da

$$P_{jk}^{(n)} = \begin{cases} qp^k & k = 0, 1, 2, \dots, n-1 \\ p^n & k = j+n+1 \\ 0 & \text{altrimenti.} \end{cases}$$

3.3.5 Fila d'attesa discreta

Consideriamo un servizio (ad esempio una cassa di un supermercato) alla quale arrivano clienti che devono essere serviti. I clienti arrivano alla cassa in diversi istanti secondo una certa legge di probabilità. Noi non osserviamo la coda a questa cassa in ogni istante ma solo negli istanti immediatamente successivi al momento in cui un cliente è servito e lascia la coda. Durante il tempo in cui tale cliente era servito alla cassa, la coda può essere aumentata di tanti clienti quanti quelli che si sono presentati nell'intervallo di tempo necessario al suo servizio. Supponiamo che all'istante n siano presenti X_n clienti, con $X_n > 1$. Allora all'istante $n+1$, che è l'istante in cui uno di quegli X_n clienti è stato servito e lascia la coda, vi saranno in coda X_{n+1} clienti dati da:

$$\begin{aligned} X_{n+1} &= X_n - 1 + Z_n && \text{se } X_n \geq 1 \\ X_{n+1} &= Z_n && \text{se } X_n = 0, \end{aligned}$$

dove le variabili casuali Z_n sono una successione di variabili casuali discrete indipendenti aventi distribuzione

$$\mathbf{P}(Z_n = j) = p_j, \quad j = 0, 1, 2, \dots$$

e rappresentano il numero di clienti che si presentano nell'istante n . La famiglia di variabili casuali X_n costituisce una catena di Markov avente spazio degli stati $\mathcal{S} = \{0, 1, 2, \dots\}$. Per determinare la matrice di transizione osserviamo che, se $X_n = 0$, allora

$$\mathbf{P}(X_{n+1} = k | X_n = 0) = \mathbf{P}(Y_{n+1} = k) = p_k.$$

Se invece $X_n = h$, con $h \geq 1$, allora

$$\begin{aligned} \mathbf{P}(X_{n+1} = k | X_n = h) &= \frac{\mathbf{P}(X_{n+1} = k, X_n = h)}{\mathbf{P}(X_n = h)} \\ &= \frac{\mathbf{P}(X_n - 1 + Z_{n+1} = k, X_n = h)}{\mathbf{P}(X_n = h)} \\ &= \frac{\mathbf{P}(Z_{n+1} = k - h + 1, X_n = h)}{\mathbf{P}(X_n = h)} \\ &= \mathbf{P}(Z_{n+1} = k - h + 1) = p_{k-h+1}. \end{aligned}$$

La matrice di transizione risulta quindi

$$P = \begin{bmatrix} p_0 & p_1 & p_2 & p_3 & p_4 & \dots \\ p_0 & p_1 & p_2 & p_3 & p_4 & \dots \\ 0 & p_0 & p_1 & p_2 & p_3 & \dots \\ 0 & 0 & p_0 & p_1 & p_2 & \dots \\ 0 & 0 & 0 & p_0 & p_1 & \dots \\ 0 & 0 & 0 & 0 & p_0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

3.3.6 Catene di nascita e morte

Le catene di nascita e morte sono caratterizzate dal fatto se $X_n = j$, allora nell'istante successivo potrà solo verificarsi che $X_{n+1} = j + 1$, evento che corrisponde ad una nascita, o $X_{n+1} = j - 1$, evento che corrisponde ad una morte, o infine $X_{n+1} = j$ evento che corrisponde ad una situazione stabile. Lo spazio degli stati di una catene di nascita e morte è $\mathcal{S} = \{0, 1, 2, \dots\}$ oppure $\mathcal{S} = \{0, 1, 2, \dots, d\}$, le probabilità di

transizione ad un passo sono date da

$$\begin{aligned}\mathbf{P}(X_{n+1} = j - 1 | X_n = j) &= q_j \\ \mathbf{P}(X_{n+1} = j | X_n = j) &= r_j \\ \mathbf{P}(X_{n+1} = j + 1 | X_n = j) &= p_j,\end{aligned}$$

con $q_j + r_j + p_j = 1$ per ogni $j = 1, 2, \dots$. La matrice di transizione risulta pertanto:

$$P = \begin{bmatrix} r_0 & p_0 & & & & \dots \\ q_1 & r_1 & p_1 & & & \dots \\ & q_2 & r_2 & p_2 & & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ & & & \dots & q_j & r_j & p_j \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ & & & & & \dots & q_d & r_d \end{bmatrix}.$$

Se lo spazio degli stati è infinito allora anche la matrice ha infinite righe ed infinite colonne.

3.3.7 Un modello genetico

In una popolazione sono presenti tre tipi di genotipi, AA , Aa e aa , caratterizzati dai geni di tipo A e a . Indichiamo con 1, 2 e 3 rispettivamente i tre tipi di genotipo. Supponiamo che nella popolazione i genotipi siano presenti rispettivamente con frequenze p^2 , $2pq$ e q^2 , con $p + q = 1$. Coppie di genitori formano una cellula figlia cedendo uno dei due geni che possiedono, con uguale probabilità. Seguiamo generazione dopo generazione come vengono tramandati i genotipi da un capostipite (generazione zero) ai discendenti di padre in figlio. Se indichiamo con X_n il genotipo presente alla generazione n , si può ipotizzare che la famiglia $\{X_n\}$ sia una catena di Markov con tre stati, in quanto il genotipo del discendente alla generazione n dipende solo dal genotipo dell'ascendente alla generazione $n - 1$, e non da tutti i precedenti ascendenti. Indichiamo con $p_{ik} = \mathbf{P}(X_n = k | X_{n-1} = i)$ la probabilità condizionata che un discendente sia di genotipo k dato che il padre è di genotipo i . Vogliamo calcolare le probabilità di transizione p_{ik} , $i, k = 1, 2, 3$ assumendo che la probabilità che l'altro genitore sia del genotipo 1, 2 o 3 sia rispettivamente p^2 , $2pq$ e q^2 e che l'accoppiamento sia frutto del caso. I risultati sono dati dalla seguente matrice

$$P = \begin{bmatrix} p & q & 0 \\ \frac{p}{2} & \frac{1}{2} & \frac{q}{2} \\ 0 & p & q \end{bmatrix}$$

Calcoliamo ad esempio p_{11} . I genitori possono essere coppie del tipo (AA, AA) , (AA, Aa) , (AA, aa) rispettivamente con probabilità p^4 , $2p^3q$, p^2q^2 . Tali coppie possono generare un figlio di tipo AA rispettivamente con probabilità 1 , $\frac{1}{2}$, 0 . Dunque la probabilità cercata, decomponendo lo spazio campionario a seconda del genotipo del secondo genitore e applicando la definizione di probabilità condizionata, risulta

$$p_{11} = \frac{p^4 + p^3q}{p^2} = p^2 + pq = p.$$

Analogamente calcoliamo p_{21} . I genitori possono essere di tipo Aa, AA , Aa, aA , Aa, aa rispettivamente con probabilità $2pqp^2$, $2pq2pq$, $2pqq^2$. Essi possono generare un figlio di tipo AA rispettivamente con probabilità $\frac{1}{2}$, $\frac{1}{4}$, 0 . Dunque la probabilità richiesta risulta

$$p_{21} = \frac{p^3q + p^2q^2}{2pq} = \frac{1}{2}(p^2 + pq) = \frac{p}{2}.$$

Procedendo in modo analogo si possono calcolare tutte le altre probabilità.

3.3.8 Il problema del ballottaggio

Riprendiamo il problema del ballottaggio presentato nell'introduzione e riscriviamolo secondo le notazioni della passeggiata aleatoria. Il susseguirsi dello spoglio dei voti corrisponde ad una passeggiata aleatoria che parte dall'origine ed arriva nel punto di coordinate $(p+q, p-q)$ dove p e q sono rispettivamente il numero di passi verso destra e il numero di passi verso sinistra, con $p+q = n$ e $p > q$. Se vogliamo sapere con che probabilità durante lo spoglio il candidato P risulta sempre in testa, dobbiamo contare i cammini che partono dall'origine e arrivano nel punto di coordinate $p+q, p-q$ e non toccano mai l'asse delle ascisse. Tali cammini sono tanti quelli che partono dal punto di coordinate $(1, 1)$ e arrivano nel punto $(p+q, p-q)$ e non toccano l'asse delle ascisse. I cammini che partono da $(1, 1)$ e arrivano in $(p+q, p-q)$ sono $\binom{n-1}{p-1}$. A questi cammini dobbiamo togliere i cammini che toccano l'asse delle ascisse. Per ogni cammino che parte da $(1, 1)$ e arriva in $(p+q, p-q)$ toccando l'asse delle ascisse, costruiamo un cammino così fatto: tale cammino è uguale a quello che sto considerando fino all'istante in cui tocca l'asse delle ascisse, da questo punto in poi è simmetrico a quello dato rispetto all'asse delle ascisse. In questo modo un tale cammino parte da $(1, 1)$ e finisce nel punto di coordinate $(n, -x)$, dove $x = p - q$. Inoltre ognuno di questi cammini è in corrispondenza biunivoca con i cammini che partono da $(1, 1)$ terminano in (n, x) e toccano l'asse delle ascisse. Quindi i cammini che vogliamo scartare sono tanti quanti i cammini che partono dal punto $(1, 1)$ e terminano nel punto $(n, -x)$. Cioè sono tanti quanti i cammini dall'origine al punto

di coordinate $(n-1, -(x+1))$. Per contarli occorre risolvere il sistema

$$\begin{cases} p' + q' = n - 1 \\ p' - q' = -(x + 1) \end{cases}$$

che da come soluzione $p' = q - 1$ e $q' = p$. Quindi tali cammini sono $\binom{n-1}{p}$. Allora i cammini che partono da $(1, 1)$, arrivano in (n, x) e non toccano l'asse delle ascisse sono

$$\binom{n-1}{p-1} - \binom{n-1}{p} = \binom{n}{p} \frac{p-q}{p+q}.$$

La probabilità che il candidato che ha ottenuto p voti risulti sempre in testa risulta

$$\frac{\binom{n}{p} \frac{p-q}{p+q}}{\binom{n}{p}} = \frac{p-q}{p+q}.$$

3.4 Probabilità di transizione in n passi

Data una catena di Markov siamo interessati a studiarne l'evoluzione in due diverse scale temporali. La prima, che studiamo in questo paragrafo, è una scala a corto termine, la seconda è una scala a lungo termine e la vedremo fra due paragrafi. Introduciamo le probabilità di transizione ad n passi. Sia $\{X_n\}$ una catena di Markov avente spazio degli stati \mathcal{S} . Denotiamo con $P_{ij}^{(n)}$, per ogni $n \geq 1$, la probabilità che la catena si trovi nello stato j dato che si trovava nello stato i n istanti fa. In formula

$$P_{ij}^{(n)} = \mathbf{P}(X_{n+m} = j | X_m = i).$$

Si noti che $P_{ij}^{(n)}$ non dipende da m poiché la catena è omogenea, cioè possiamo sempre scrivere

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

Si osservi inoltre che $P_{ij}^{(1)} = P_{ij}$. Vale la relazione

$$P_{ij}^{(n)} = \sum_{h \in \mathcal{S}} P_{ih}^{n-1} P_{hj}. \quad (3.3)$$

Infatti possiamo scrivere

$$\begin{aligned} P_{ij}^{(n)} &= \frac{\mathbf{P}(X_n = j, X_0 = i)}{\mathbf{P}(X_0 = i)} = \sum_{h \in \mathcal{S}} \frac{\mathbf{P}(X_n = j, X_{n-1} = h, X_0 = i)}{\mathbf{P}(X_0 = i)} \\ &= \sum_{h \in \mathcal{S}} \mathbf{P}(X_n = j | X_{n-1} = h, X_0 = i) \mathbf{P}(X_{n-1} = h | X_0 = i) \\ &= \sum_{h \in \mathcal{S}} \mathbf{P}(X_n = j | X_{n-1} = h) P_{ih}^{n-1} = \sum_{h \in \mathcal{S}} \mathbf{P}(X_1 = j | X_0 = h) P_{ih}^{n-1}, \end{aligned}$$

da cui si ha la (3.3), invertendo l'ordine dei fattori. Se indichiamo con $P^{(n)}$ la matrice i cui elementi sono le probabilità $P_{ij}^{(n)}$ vale quindi la relazione $P^{(n)} = P^{(n-1)} \cdot P$ nel senso di prodotto di matrici (eventualmente infinite) righe per colonne. Da questa relazione si deduce che $P^{(2)} = P \cdot P = P^2$, e per ricorrenza $P^{(n)} = P^n$. Le equazioni di Chapman–Kolmogorov, di cui la (3.3) è un caso particolare, forniscono uno strumento per calcolare le probabilità di transizione in n passi.

Proposizione 3.4.1. *Data una catena di Markov omogenea $\{X_n\}$, $n \geq 0$ avente probabilità di transizione P_{ij} , valgono le seguenti relazioni*

$$P_{ij}^{(m+n)} = \sum_{h \in \mathcal{S}} P_{ih}^{(m)} P_{hj}^{(n)},$$

per ogni $k, l \geq 0$ e per ogni i, j .

L'interpretazione di questa legge è che la catena per andare dallo stato i allo stato j in $k + l$ passi, può andare dallo stato i ad uno stato intermedio h in k passi e quindi dallo stato h allo stato j in l passi. La probabilità di essere nello stato j dopo $k + l$ passi dato che si trovava nello stato i è calcolata sommando su tutti gli stati intermedi le probabilità degli eventi corrispondenti. Questa legge possiamo riscriverla in forma matriciale:

$$P^{(m+n)} = P^{(m)} \cdot P^{(n)},$$

La dimostrazione è la seguente. Possiamo scrivere

$$P(X_{m+n} = j | X_0 = i) = \sum_h P(X_{m+n} = j, X_m = h | X_0 = i)$$

Riscriviamo il termine della somma come

$$\begin{aligned} P(X_{m+n} = j, X_m = h | X_0 = i) &= \frac{P(X_{m+n} = j, X_m = h, X_0 = i)}{\mathbf{P}(X_0 = i)} = \\ &= \frac{P(X_{m+n} = j, X_m = h, X_0 = i)}{\mathbf{P}(X_m = h, X_0 = i)} \cdot \frac{P(X_m = h, X_0 = i)}{\mathbf{P}(X_0 = i)} = \\ &= P(X_{m+n} = j | X_m = h, X_0 = i) \cdot P(X_m = h | X_0 = i) \end{aligned}$$

Per la proprietà di Markov l'ultima espressione diviene

$$= P(X_{m+n} = j | X_m = h) \cdot P(X_m = h | X_0 = i) = P_{ih}^{(m)} P_{hj}^{(n)}$$

Accanto alle probabilità di transizione in n passi, assumono particolare rilevanza le probabilità che la catena all'istante n si trovi in un particolare stato i . Denotiamo con π^n tale probabilità. In particolare denotiamo con

$$\pi_k^n = \mathbf{P}(X_n = k), \quad \forall k \in \mathcal{S}$$

la legge della variabile casuale X_n per ogni $n \geq 1$. La legge della variabile casuale X_0 , cioè la legge iniziale della catena di Markov, è indicata analogamente con π^0 .

Proposizione 3.4.2. *Data una catena di Markov omogenea $\{X_n\}$, $n \geq 0$ avente matrice di transizione P e tale che la distribuzione di X_0 sia π^0 , la distribuzione di X_n è data da*

$$\pi_k^n = \sum_{h \in \mathcal{S}} \pi_h^0 P_{hk}^{(n)},$$

per ogni $k \in \mathcal{S}$.

Dimostrazione. Possiamo scrivere

$$\mathbf{P}(X_n = k) = \sum_{h \in \mathcal{S}} \mathbf{P}(X_n = k, X_0 = h) = \sum_{h \in \mathcal{S}} \mathbf{P}(X_n = k | X_0 = h) \mathbf{P}(X_0 = h),$$

da cui la tesi invertendo l'ordine dei fattori. □

In forma matriciale possiamo scrivere che

$$\pi^n = \pi^0 P^n.$$

Possiamo anche determinare le distribuzioni congiunte di qualunque n -upla di variabili casuali $(X_{k_1}, \dots, X_{k_n})$.

Proposizione 3.4.3. *Data una catena di Markov omogenea $\{X_n\}$, $n \geq 0$ avente matrice di transizione P e tale che la distribuzione di X_0 sia π^0 , la distribuzione di $(X_{k_1}, X_{k_2}, \dots, X_{k_n})$, dove $0 < k_1 < k_2 < \dots < k_n$, è data da*

$$\mathbf{P}(X_{k_1} = h_1, \dots, X_{k_n} = h_n) = \sum_{h \in \mathcal{S}} \pi_h^0 P_{hh_1}^{(k_1)} P_{h_1 h_2}^{(k_2 - k_1)} \dots P_{h_{n-1} h_n}^{(k_n - k_{n-1})}$$

per ogni $h_1, h_2, \dots, h_n \in \mathcal{S}$.

Dimostrazione. Poniamo

$$\nu_{h_1, \dots, h_n}^{k_1, \dots, k_n} = \mathbf{P}(X_{k_1} = h_1, \dots, X_{k_n} = h_n).$$

Possiamo scrivere

$$\begin{aligned} \nu_{h_1, \dots, h_n}^{k_1, \dots, k_n} &= \mathbf{P}(X_{k_1} = h_1, \dots, X_{k_{n-1}} = h_{n-1}) \mathbf{P}(X_{k_n} = h_n | X_{k_1} = h_1, \dots, X_{k_{n-1}} = h_{n-1}) \\ &= \mathbf{P}(X_{k_1} = h_1, \dots, X_{k_{n-1}} = h_{n-1}) P_{h_{n-1} h_n}^{(k_n - k_{n-1})}. \end{aligned}$$

Ripetendo il ragionamento si arriva alla tesi. □

3.4.1 Esempi

Consideriamo la seguente matrice di transizione che descrive il tempo in una città. Gli stati sono tre. stato 1: rovesci; stato 2: coperto; stato 3: sereno. La matrice di transizione è la seguente:

$$P = \begin{bmatrix} 0.3 & 0.7 & 0 \\ 0.2 & 0.5 & 0.3 \\ 0.1 & 0.3 & 0.6 \end{bmatrix}$$

Esercizio 3.4.4. *Dato che oggi piove, con che probabilità dopodomani ci sarà il sole?*

Dobbiamo calcolare

$$P_{1,3}^{(2)} = \mathbf{P}(X_2 = 3 | X_0 = 1).$$

Calcoliamo la matrice di transizione a due passi:

$$P^{(2)} = \begin{bmatrix} 0.23 & 0.56 & 0.21 \\ 0.19 & 0.48 & 0.33 \\ 0.15 & 0.40 & 0.45 \end{bmatrix}$$

La probabilità richiesta è $P_{1,3}^{(2)} = 0.21$.

Esercizio 3.4.5. *Dato che oggi piove, con che probabilità tra 4 giorni ci sarà il sole? Con che probabilità tra 7 giorni ci sarà il sole?*

Calcoliamo la matrice di transizione a 4 passi:

$$P^{(4)} = \begin{bmatrix} 0.19 & 0.48 & 0.33 \\ 0.18 & 0.47 & 0.35 \\ 0.18 & 0.46 & 0.37 \end{bmatrix}$$

La probabilità richiesta è $P_{1,3}^{(4)} = 0.33$. Calcoliamo la matrice di transizione a 7 passi:

$$P^{(7)} = \begin{bmatrix} 0.18 & 0.47 & 0.35 \\ 0.18 & 0.47 & 0.35 \\ 0.18 & 0.47 & 0.35 \end{bmatrix}$$

La probabilità richiesta è $P_{1,3}^{(7)} = 0.35$. Si osservi come al crescere di n la matrice tende ad avere tre righe uguali.

Consideriamo ora l'esempio della rovina del giocatore con $N = 5$. La matrice di transizione è la seguente:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Esercizio 3.4.6. *Calcolare la matrice di transizione a due passi e a venti passi.*

Le matrici sono le seguenti:

$$P^{(2)} = \begin{bmatrix} 1 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.60 & 0.24 & 0.00 & 0.16 & 0.00 & 0.00 \\ 0.36 & 0.00 & 0.48 & 0.00 & 0.16 & 0.00 \\ 0.00 & 0.36 & 0.00 & 0.48 & 0.00 & 0.16 \\ 0.00 & 0.00 & 0.36 & 0.00 & 0.24 & 0.40 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1 \end{bmatrix}$$

$$P^{(20)} = \begin{bmatrix} 1 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.92 & 0.00 & 0.00 & 0.00 & 0.00 & 0.07 \\ 0.80 & 0.00 & 0.01 & 0.00 & 0.00 & 0.19 \\ 0.63 & 0.01 & 0.00 & 0.01 & 0.00 & 0.36 \\ 0.38 & 0.00 & 0.01 & 0.00 & 0.00 & 0.61 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1 \end{bmatrix}$$

Si osservi che se il capitale è 2 o 3 abbiamo alta probabilità di essere assorbiti in 0 o 5. Si osservi poi che se il capitale iniziale è 3, con probabilità 0 sarà 2 dopo 20 giocate. Questo perchè la fortuna del giocatore si alterna tra un numero pari e uno dispari fino a quando non viene assorbito in uno dei punti 0 o 5.

Riprendiamo l'esempio della matrice che descrive il tempo in una città.

Esercizio 3.4.7. *Calcolare la probabilità che tra quattro giorni piova. Si supponga che la distribuzione iniziale sia $\pi^0 = (1, 0, 0)$ cioè lo stato iniziale è sicuramente rovescio.*

Occorre calcolare $\pi^4 = \pi^0 \cdot P^{(4)}$. Abbiamo

$$\pi^4 = (1, 0, 0) \begin{bmatrix} 0.19 & 0.48 & 0.33 \\ 0.18 & 0.47 & 0.35 \\ 0.18 & 0.46 & 0.37 \end{bmatrix} = (0.19, 0.48, 0.33)$$

La probabilità richiesta è $\pi_1^4 = 0.19$. Si osservi che a differenza dei precedenti esercizi adesso lo stato iniziale è descritto dalla sua distribuzione. Abbiamo che $\mathbf{P}(X_0 = 1) = 1$ e dobbiamo calcolare $\mathbf{P}(X_4 = 1)$. Questa probabilità non è uguale $\mathbf{P}(X_4 = 1|X_0 = 1)$ anche se i valori numerici coincidono.

Esercizio 3.4.8. *Calcolare la probabilità che tra quattro giorni piova. Si supponga che la distribuzione iniziale sia $\pi^0 = (0.3, 0.4, 0.3)$.*

Occorre calcolare $\pi^4 = \pi^0 \cdot P^{(4)}$. Abbiamo

$$\pi^4 = (0.3, 0.4, 0.3) \begin{bmatrix} 0.19 & 0.48 & 0.33 \\ 0.18 & 0.47 & 0.35 \\ 0.18 & 0.46 & 0.37 \end{bmatrix} = (0.18, 0.47, 0.35)$$

La probabilità richiesta è $\pi_1^4 = 0.18$.

3.5 Catena di Markov con due stati

Una catena di Markov a due stati può essere un buon modello per una situazione di questo tipo. Una macchina si può trovare in due stati, funzionante o rotto. Indichiamo con 0 lo stato corrispondente alla macchina rotta e con 1 lo stato corrispondente a macchina funzionante. Supponiamo che se la macchina si trova nello stato 0 il giorno n , sia α la probabilità che si trovi allo stato 1 il giorno $n + 1$, indipendentemente da dove si trovava nei giorni prima di n . Supponiamo inoltre che se si trova nello stato 1 il giorno n , sia β la probabilità che si trovi nello stato 0 il giorno $n + 1$, indipendentemente da dove si trovava nei giorni prima di n . La catena $\{X_n\}$ segue l'evolversi dello stato della macchina durante i giorni $n = 1, 2, \dots$. Si tratta di una catena di Markov per come è stata definita. La sua matrice di transizione è

$$P = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix}.$$

Per evitare casi banali supponiamo che α e β non siano contemporaneamente 0 o 1. Infatti se sono entrambe 0 se il sistema si trova nello stato 0 vi rimane per sempre e analogamente se si trova nello stato 1. Se invece sono contemporaneamente 1, il sistema salta deterministicamente ad ogni istante dallo stato 0 allo stato 1. Supporremo pertanto che $0 < \alpha + \beta < 2$. Denotiamo con π_0^0 la probabilità che la catena all'istante 0 si trovi allo stato 0. Vale a dire $\mathbf{P}(X_0 = 0) = \pi_0^0$. Poiché la catena ha solo due stati si ha che $\mathbf{P}(X_0 = 1) = 1 - \pi_0^0 = \pi_1^0$. Cerchiamo di ricavare la matrice di transizione in n passi. Abbiamo

$$\begin{aligned} P_{00}^{(n+1)} &= \mathbf{P}(X_{n+1} = 0 | X_0 = 0) = \\ &= \mathbf{P}(X_{n+1} = 0, X_n = 0 | X_0 = 0) + \mathbf{P}(X_{n+1} = 0, X_n = 1 | X_0 = 0) \\ &= \frac{\mathbf{P}(X_{n+1} = 0, X_n = 0, X_0 = 0)}{\mathbf{P}(X_0 = 0)} + \frac{\mathbf{P}(X_{n+1} = 0, X_n = 1, X_0 = 0)}{\mathbf{P}(X_0 = 0)} \\ &= \frac{\mathbf{P}(X_{n+1} = 0 | X_n = 0, X_0 = 0)}{\mathbf{P}(X_0 = 0)} \mathbf{P}(X_n = 0, X_0 = 0) + \\ &\quad + \frac{\mathbf{P}(X_{n+1} = 0 | X_n = 1, X_0 = 0)}{\mathbf{P}(X_0 = 0)} \mathbf{P}(X_n = 1, X_0 = 0) \\ &= \mathbf{P}(X_{n+1} = 0 | X_n = 0) \mathbf{P}(X_n = 0 | X_0 = 0) + \\ &\quad + \mathbf{P}(X_{n+1} = 0 | X_n = 1) \mathbf{P}(X_n = 1 | X_0 = 0). \end{aligned}$$

In termini delle probabilità di transizione possiamo riscrivere il primo e l'ultimo membro come

$$P_{00}^{(n+1)} = (1 - \alpha)P_{00}^{(n)} + \beta P_{01}^{(n)}.$$

Ma sappiamo che

$$P_{01}^{(n)} = 1 - P_{00}^{(n)},$$

da cui ricaviamo una equazione ricorrente per $P_{00}^{(n)}$

$$P_{00}^{(n+1)} = (1 - \alpha)P_{00}^{(n)} + \beta(1 - P_{00}^{(n)})$$

che possiamo infine riscrivere come

$$P_{00}^{(n+1)} = (1 - \alpha - \beta)P_{00}^{(n)} + \beta,$$

con la condizione iniziale $P_{00}^{(0)} = 1$, cioè se il sistema si trova nello stato 0 è 1 la probabilità che in zero passi sia in 0. Da questa equazione ricaviamo ad esempio che

$$P_{00}^{(2)} = (1 - \alpha)^2 + \alpha\beta,$$

che possiamo ricavare anche facendo il prodotto della matrice P per se stessa e prendendo il primo elemento in alto a destra della matrice prodotto. La soluzione generale la si trova risolvendo l'equazione ricorrente ed è data da

$$P_{00}^{(n)} = \frac{\alpha}{\alpha + \beta}(1 - \alpha - \beta)^n + \frac{\beta}{\alpha + \beta}$$

In modo analogo, cioè risolvendo altre equazioni ricorrenti si trova che la matrice di transizione in n passi è data da

$$P^{(n)} = \frac{(1 - \alpha - \beta)^n}{\alpha + \beta} \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} + \frac{1}{\alpha + \beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix}. \quad (3.4)$$

Passiamo ora al calcolo delle distribuzioni delle variabili X_n , per rispondere a domande quali, quale è la probabilità che all'istante n la macchina si trovi nello stato 0? Vogliamo determinare $\mathbf{P}(X_n = 0)$. Cerchiamo anche in questo caso un'equazione ricorrente. Posto $\pi_0^{n+1} = \mathbf{P}(X_{n+1} = 0)$ abbiamo

$$\begin{aligned} \pi_0^{n+1} &= \mathbf{P}(X_{n+1} = 0, X_n = 0) + \mathbf{P}(X_{n+1} = 0, X_n = 1) \\ &= \mathbf{P}(X_{n+1} = 0 | X_n = 0)\mathbf{P}(X_n = 0) + \mathbf{P}(X_{n+1} = 0 | X_n = 1)\mathbf{P}(X_n = 1) \\ &= (1 - \alpha)\mathbf{P}(X_n = 0) + \beta\mathbf{P}(X_n = 1) \\ &= (1 - \alpha - \beta)\mathbf{P}(X_n = 0) + \beta. \end{aligned}$$

Riscritta per $n = 1$ diventa

$$\mathbf{P}(X_1 = 0) = (1 - \alpha - \beta)\pi_0^0 + \beta.$$

Per $n = 2$ ancora abbiamo

$$\mathbf{P}(X_2 = 0) = (1 - \alpha - \beta)^2 \pi_0^0 + \beta(1 + (1 - \alpha - \beta)).$$

Si deduce quindi

$$\mathbf{P}(X_n = 0) = (1 - \alpha - \beta)^n \pi_0^0 + \beta \sum_{j=0}^{n-1} (1 - \alpha - \beta)^j.$$

Poiché

$$\sum_{j=0}^{n-1} (1 - \alpha - \beta)^j = \frac{1 - (1 - \alpha - \beta)^n}{\alpha + \beta},$$

ricaviamo la soluzione

$$\mathbf{P}(X_n = 0) = \frac{\beta}{\alpha + \beta} + (1 - \alpha - \beta)^n \left(\pi_0^0 - \frac{\beta}{\alpha + \beta} \right) \quad (3.5)$$

e per complemento

$$\mathbf{P}(X_n = 1) = \frac{\alpha}{\alpha + \beta} + (1 - \alpha - \beta)^n \left(\pi_1^0 - \frac{\alpha}{\alpha + \beta} \right) \quad (3.6)$$

Si può osservare che per le ipotesi fatte $|1 - \alpha - \beta| < 1$, per cui possiamo concludere che

$$\lim_{n \rightarrow +\infty} \mathbf{P}(X_n = 0) = \frac{\beta}{\alpha + \beta}$$

e

$$\lim_{n \rightarrow +\infty} \mathbf{P}(X_n = 1) = \frac{\alpha}{\alpha + \beta}.$$

Proviamo infine a calcolare la probabilità congiunta di $X_0 = 0$, $X_1 = 1$ e $X_2 = 0$. Questa è data da

$$\begin{aligned} \mathbf{P}(X_0 = 0, X_1 = 1, X_2 = 0) &= \mathbf{P}(X_0 = 0) \mathbf{P}(X_1 = 1 | X_0) \mathbf{P}(X_2 = 0 | X_1 = 1) \\ &= \pi_0^0 \alpha \beta \end{aligned}$$

Provare a calcolare le altre probabilità congiunte. Come ulteriore esempio di catena di Markov a due stati si consideri il seguente esempio

Esempio 3.5.1. Un virus mutante può stare in N differenti stati. Ad ogni generazione il virus rimane nello stato in cui è o muta in uno degli altri stati in modo casuale e indipendentemente dagli stati in cui si trovava precedentemente. Vogliamo trovare la probabilità che il virus dopo n generazioni sia nello stesso stato in

cui si trovava nella generazione 0. Indichiamo con $i = 1, 2, \dots, N$ gli stati del virus. Abbiamo

$$P_{ii} = 1 - \alpha, \quad P_{ij} = \frac{\alpha}{N-1}, \quad j \neq i.$$

Dobbiamo calcolare $\mathbf{P}(X_n = i | X_0 = i) = P_{ii}^{(n)}$. La catena non ha due stati ma possiamo ridurci ad una catena con due stati. Poiché non ha importanza lo stato iniziale in cui si trova il virus, possiamo considerare un'altra catena Y_n a due stati per la quale lo stato $Y_n = 0$ corrisponde a virus nello stesso stato iniziale, cioè $X_n = i$ se i è lo stato iniziale, mentre $Y_n = 1$ corrisponde a virus in uno stato diverso da quello iniziale. Abbiamo allora che la matrice di transizione è

$$P = \begin{bmatrix} 1 - \alpha & \alpha \\ \frac{\alpha}{N-1} & 1 - \frac{\alpha}{N-1} \end{bmatrix}.$$

La probabilità richiesta è dunque

$$P_{11}^{(n)} = \frac{\alpha}{\alpha + \frac{\alpha}{N-1}} \left(1 - \alpha - \frac{\alpha}{N-1} \right)^n + \frac{\frac{\alpha}{N-1}}{\alpha + \frac{\alpha}{N-1}},$$

che semplificato da

$$P_{11}^{(n)} = \frac{1}{N} + \left(1 - \frac{1}{N} \right) \left(1 - \frac{N\alpha}{N-1} \right)^n.$$

Per n che tende all'infinito tale probabilità è pari a $\frac{1}{N}$.

Capitolo 4

Classificazione degli stati

4.1 Tempi di arresto

Introduciamo ora una grandezza molto importante per lo studio delle catene di Markov. Si tratta di una ‘quasi’ variabile aleatoria che registra il tempo di primo passaggio di una catena in uno stato i .

Definizione 4.1.1. *Sia $i \in \mathcal{S}$, poniamo*

$$\tau_i = \min\{n \geq 1 : X_n = i\}.$$

τ_i è detto tempo di primo passaggio della catena nello stato i .

La quantità τ_i non è propriamente una variabile aleatoria perché τ_i può assumere anche valore infinito (quando lo stato i non è mai raggiunto) e questo può avvenire con probabilità positiva. Si osservi che la τ_i è definita senza specificare alcuno stato iniziale. Inoltre

$$\mathbf{P}(\tau_i = m) = \mathbf{P}(X_m = i, X_{m-1} \neq i, \dots, X_1 \neq i). \quad (4.1)$$

Si noti che se la catena fosse o meno nello stato i all’istante iniziale $t = 0$ è irrilevante.

Esempio 4.1.2. Consideriamo una passeggiata aleatoria e cerchiamo di capire cosa è τ_3 . L’evento $\tau_3 = 1$ si verifica solo se si verificano gli eventi $\{X_0 = 2, X_1 = 3\}$ e $\{X_0 = 4, X_1 = 3\}$. Di conseguenza

$$\mathbf{P}(\tau_3 = 1) = \mathbf{P}(X_1 = 3|X_0 = 2)\mathbf{P}(X_0 = 2) + \mathbf{P}(X_1 = 3|X_0 = 4)\mathbf{P}(X_0 = 4).$$

Se ne deduce che conoscendo la distribuzione iniziale e la matrice di transizione possiamo calcolare la distribuzione del tempo di primo passaggio. In realtà quello che a noi interessa calcolare è piuttosto la probabilità condizionata rispetto allo stato della catena all'istante 0, vale a dire, ad esempio, $\mathbf{P}(\tau_3 = 1 | X_0 = 0)$.

Sia ora

$$f_{ii} = \mathbf{P}(\tau_i < +\infty | X_0 = i)$$

Tale quantità rappresenta la probabilità che X_n ritorni allo stato i almeno una volta dato che è partita dallo stato i . La proprietà di mancanza di memoria di una catena di Markov, vale anche in un senso più generale che specifichiamo ora.

Definizione 4.1.3. *Una variabile τ si chiama tempo d'arresto (in inglese stopping time) se la realizzazione dell'evento $\{\tau = n\}$ può essere stabilita esclusivamente dalla conoscenza del processo fino all'istante n , vale a dire dalla conoscenza di X_0, X_1, \dots, X_n .*

Si noti che la variabile τ_j introdotta all'inizio della sezione è un tempo di arresto in quanto vale la relazione (4.1).

Vale il seguente importante risultato, detto proprietà di Markov forte. Questo per distinguerla dalla proprietà di Markov semplice (mancanza di memoria).

Teorema 4.1.4. *Sia τ un tempo di arresto. Data la conoscenza di $\tau = n$ e $X_\tau = i$, nessuna altra informazione su X_0, X_1, \dots, X_τ è necessaria per determinare la probabilità condizionata dell'evento $X_{\tau+1} = j$. Vale cioè la relazione*

$$\mathbf{P}(X_{\tau+1} = j | X_\tau = i, \tau = n) = P_{ij}$$

In sostanza la proprietà della mancanza di memoria vale anche per degli istanti aleatori. La catena $\{\tilde{X}_k\} = \{X_{\tau+k}\}$ si comporta come una catena di Markov partita da i .

Dimostrazione. Denotiamo con V_n l'insieme dei vettori $x = (i_0, i_1, \dots, i_n)$ tali che l'evento $\{X_0 = i_0, X_1 = i_1, X_n = i_n\}$ implichi l'evento $\tau = n$ e $X_\tau = i$ (si noti che dovrà essere $i_n = i$). Allora per il teorema delle probabilità totali vale

$$\mathbf{P}(X_{\tau+1} = j | X_\tau = i, \tau = n) = \sum_{x \in V_n} \mathbf{P}(X_{n+1} = j, X_n = x_n, \dots, X_0 = x_0)$$

Il secondo membro può essere riscritto come

$$\sum_{x \in V_n} \mathbf{P}(X_{n+1} = j | X_n = x_n, \dots, X_0 = x_0) \mathbf{P}(X_n = x_n, \dots, X_0 = x_0)$$

Ricordando che se $x \in V_n$ allora $\tau = n$ e $X_\tau = i$ ne segue che $x_n = i$ e quindi per la proprietà di Markov possiamo scrivere

$$\mathbf{P}(X_{\tau+1} = j | X_\tau = i, \tau = n) = P_{ij} \mathbf{P}(X_\tau = i, \tau = n)$$

in quanto

$$\sum_{x \in V_n} \mathbf{P}(X_n = x_n, \dots, X_0 = x_0) = \mathbf{P}(X_\tau = i, \tau = n)$$

La proprietà di Markov segue dividendo per $\mathbf{P}(X_\tau = i, \tau = n)$ la penultima equazione. \square

La proprietà di Markov forte implica che la probabilità che una catena torni nello stato i dato che ha visitato quello stato $n - 1$ volte è ancora f_{ii} . Da ciò si deduce che la probabilità di ritornare in i due volte è f_{ii}^2 , la probabilità di tornarvi 3 volte è $f_{ii}^2 \cdot f_{ii} = f_{ii}^3$ e in generale la probabilità di tornarvi n volte è f_{ii}^n .

Gli stati di una catena di Markov si dividono in due tipi a seconda dei valori assunti dalle probabilità di transitare e quindi ritornare in un determinato stato. Tale classificazione viene fatta a seconda del valore assunto dalle probabilità f_{ii} sopra definite.

Definizione 4.1.5. *Sia \mathcal{S} lo spazio degli stati di una catena di Markov. Dato $i \in \mathcal{S}$:*

lo stato i è detto transitorio se $f_{ii} < 1$.

lo stato i è detto ricorrente se $f_{ii} = 1$;

Si noti che se $f_{ii} < 1$ la probabilità che la catena torni n volte nello stato i è f_{ii}^n e questa tende a zero per $n \rightarrow \infty$, quindi da un certo n in poi la catena non ritornerà più nello stato i , da cui il nome di stato transitorio.

Se invece $f_{ii} = 1$ la probabilità che la catena torni n volte nello stato i è $f_{ii}^n = 1$ quindi la catena torna a visitare lo stato i infinitamente spesso. Lo stato è quindi chiamato ricorrente.

Esempio 4.1.6. Consideriamo la catena di Markov per la rovina del giocatore la cui matrice di transizione è la seguente

$$\begin{matrix} & \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \end{bmatrix} \\ \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

Mostriamo che la catena finirà o nello stato 0 (rovina del giocatore) o nello stato 4 (felicità per il giocatore). Quindi gli stati 1, 2, 3 sono transitori, secondo la nostra definizione. Per verificare che lo stato 0 è ricorrente basta osservare che $P_{00} = 1$ e quindi

$$\mathbf{P}(\tau_0 = 1 | X_0 = 0) = 1$$

da cui

$$f_{00} = \mathbf{P}(\tau_0 < \infty | X_0 = 0) \geq \mathbf{P}(\tau_0 = 1 | X_0 = 0) = 1$$

e quindi $f_{00} = 1$. In modo analogo si mostra che lo stato 4 è ricorrente. Per mostrare che lo stato 1 è transitorio, osserviamo che se la catena si trova nello stato 1 e al passo successivo visita lo stato 0, la catena non tornerà mai più nello stato 1. Quindi possiamo scrivere

$$\mathbf{P}(\tau_1 = \infty | X_0 = 1) \geq P_{10} = 0.6 > 0$$

quindi $f_{11} \leq 0.4 < 1$ e lo stato 1 è transitorio. Anche lo stato 2 è transitorio, perché se la catena parte dallo stato 2, la catena può andare nello stato 1 e da qui allo stato 0. Quindi

$$\mathbf{P}(\tau_2 = \infty | X_0 = 2) \geq P_{21} \cdot P_{10} = 0.6 \cdot 0.6 = 0.36$$

quindi $f_{22} \leq 0.36 < 1$ e lo stato 2 è transitorio. In modo analogo si mostra che lo stato 3 è transitorio.

Gli stati 0 e 4 della catena nell'esempio precedente sono stati ricorrenti un po' particolari: se la catena entra in uno di questi stati non ne esce più.

Definizione 4.1.7. *Uno stato i tale che $P_{ii} = 1$ è detto stato assorbente.*

Il significato è chiaro, se la catena entra nello stato assorbente non ne esce più.

Esempio 4.1.8. Consideriamo una catena di Markov avente 4 stati la cui matrice di transizione è la seguente,

$$\begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

Tutti gli stati sono ricorrenti. Per mostrare che lo stato 1 è ricorrente, osserviamo che τ_1 , il tempo di primo passaggio allo stato 1, può assumere solo i valori 1, 2, 3 e 4. In particolare $\mathbf{P}(\tau_1 = 1 | X_0 = 1) = 1/4$, in quanto l'unica traiettoria possibile è da 1 a 1. $\mathbf{P}(\tau_1 = 2 | X_0 = 1) = 1/4$ in quanto l'unica traiettoria che parte da 1 e arriva di nuovo in 1 dopo due istanti è da 1 a 4 e quindi da 4 a 1. Cioè

$$\mathbf{P}(\tau_1 = 2 | X_0 = 1) = \mathbf{P}(X_2 = 1 | X_1 = 4) \mathbf{P}(X_1 = 4 | X_0 = 1) = 1 \cdot \frac{1}{4}$$

In modo analogo si dimostra che $\mathbf{P}(\tau_1 = 3|X_0 = 1) = 1/4$ in quanto l'unica traiettoria che parte da 1 e arriva di nuovo in 1 dopo tre istanti è da 1 a 3, quindi da 3 a 4 e quindi da 4 a 1, e questa traiettoria ha probabilità $1/4$. Infine $\mathbf{P}(\tau_1 = 4|X_0 = 1) = 1/4$ in quanto l'unica traiettoria che parte da 1 e arriva di nuovo in 1 dopo quattro istanti è da 1 a 2, quindi da 2 a 3, quindi da 3 a 4 e quindi da 4 a 1, e questa traiettoria ha probabilità $1/4$. Di conseguenza

$$f_{11} = \mathbf{P}(\tau_1 < \infty) = \sum_{k=1}^4 \mathbf{P}(\tau_1 = k|X_0 = 1) = 1$$

In modo analogo si dimostra che $f_{ii} = 1$ per $i = 2, 3, 4$.

Definizione 4.1.9. Uno stato i comunica con lo stato j se esiste un $n \geq 0$ tale che $P_{ij}^n > 0$.

Se lo stato i comunica con j significa che la quantità

$$f_{ij} = \mathbf{P}(\tau_j < \infty|X_0 = i) > 0.$$

dove f_{ij} rappresenta la probabilità che la catena visiti lo stato j dato che è partita dallo stato i . Diremo anche che lo stato j è raggiungibile da i per sottolineare la direzione della comunicazione. Scriveremo $i \rightarrow j$. Se lo stato i comunica con lo stato j significa che la catena ha una probabilità positiva di trovarsi nello stato j dopo n passi da che si trovava nello stato i . Non è detto che se i è raggiungibile dallo stato j , lo stato j sia raggiungibile dallo stato i . Per convenzione poiché, $P_{ii}^0 = \mathbf{P}(X_0 = i|X_0 = i) = 1$, uno stato i è sempre raggiungibile da se stesso.

La relazione di comunicabilità gode delle proprietà transitiva (se i comunica con j e j comunica con k , anche i comunica con k). Per dimostrare quest'ultima proprietà basta osservare che esistono un $m \geq 0$ ed un $n \geq 0$ tali che $P_{ij}^m > 0$ ($i \rightarrow j$) e $P_{jk}^n > 0$ ($j \rightarrow k$). Di conseguenza per l'uguaglianza di Chapman Kolmogorov

$$P_{ik}^{m+n} = \sum_{h \in \mathcal{S}} P_{ih}^m P_{hk}^n \geq P_{ij}^m P_{jk}^n > 0,$$

da cui si deduce che i comunica con lo stato k ($i \rightarrow k$).

Esempio 4.1.10. Consideriamo la catena che descrive una passeggiata aleatoria con spazio degli stati $\mathcal{S} = \{0, 1, \dots, N-1, N\}$ avente barriere assorbenti in 0 e N (rovina del giocatore), lo stato 0 e lo stato N comunicano solo con loro stessi, e sono raggiungibili da tutti gli altri stati. Lo stato 0 non è raggiungibile dallo stato N e lo stato N non è raggiungibile dallo stato 0. Gli stati $1, \dots, N-1$ comunicano tra di loro, possono raggiungere gli stati 0 e N , ma non possono essere raggiunti da 0 e N .

Definizione 4.1.11. *Un sottoinsieme di stati $C \subset \mathcal{S}$ è detto chiuso (o classe chiusa) se ogni stato fuori da C non è raggiungibile da alcun stato in C . Vale a dire se $P_{ij} = 0$ per ogni $i \in C$ e per ogni $j \notin C$*

Osserviamo che se una classe C è chiusa, allora $P_{ij}^n = 0$ per ogni $i \in C$, per ogni $j \notin C$ e per ogni n . La dimostrazione viene fatta per induzione. Infatti per $n = 1$ $P_{ij} = 0$ per definizione. Supponiamo (ipotesi di induzione) che $P_{ij}^{n-1} = 0$ per ogni $i \in C$, per ogni $j \notin C$. Possiamo dunque scrivere

$$P_{ij}^n = \sum_{h \in \mathcal{S}} P_{ih} P_{hj}^{n-1} = \sum_{h \in C} P_{ih} P_{hj}^{n-1} + \sum_{h \in C^c} P_{ih} P_{hj}^{n-1} = 0,$$

in quanto, nella prima sommatoria, per l'ipotesi di induzione, $P_{hj}^{n-1} = 0$ per ogni $h \in C$ e per ogni $j \notin C$, mentre, nella seconda sommatoria, $P_{ih} = 0$ per ogni $i \in C$ e per ogni $h \notin C$.

Si osservi che gli stati fuori da C possono raggiungere gli stati in C ma non vale il viceversa, ovvero gli stati fuori da C possono comunicare con gli stati in C , ma gli stati di C non comunicano con gli stati non in C . In pratica se una catena entra in uno stato di una classe chiusa, poi vi rimane per sempre.

Definizione 4.1.12. *Una classe chiusa è detta irriducibile se tutti gli stati della classe chiusa comunicano tra di loro.*

Uno stato i che da solo forma una classe chiusa (irriducibile!) è uno stato assorbente. Infatti dovrà per forza essere $P_{ii} = 1$.

Definizione 4.1.13. *Una catena di Markov in cui l'intero spazio degli stati \mathcal{S} forma un'unica classe chiusa che è anche irriducibile è detta catena irriducibile.*

Per capire la differenza tra classe chiusa e classe chiusa irriducibile consideriamo una catena a tre stati. L'intero spazio degli stati è una classe chiusa, ma se c'è uno stato assorbente questa classe non è irriducibile. Una classe chiusa per essere irriducibile deve essere tale che ogni stato della classe chiusa possa essere raggiunto e possa raggiungere ogni altro stato della classe chiusa.

Esempio 4.1.14. Per la catena di Markov con la seguente matrice di transizione

$$\begin{bmatrix} 0 & p_{12} & p_{13} \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

tutti gli stati formano una classe irriducibile. La catena è detta irriducibile. In particolare per $n = 5$ abbiamo $P_{ij}^n > 0$ per ogni i e j . Il numero 5 risulta il numero

minimo di passi necessari per raggiungere qualunque stato essendo partiti da uno stato qualunque, cioè il numero di passi che sono necessari agli stati per raggiungersi vicendevolmente. Per la catena di Markov con la seguente matrice di transizione

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ 0 & 1 & 0 \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

tutti gli stati costituiscono una classe chiusa ma non irriducibile in quanto lo stato 2 non comunica con alcun stato diverso da se stesso.

Per chiarire il concetto di classe chiusa ed irriducibile consideriamo un ulteriore esempio.

Esempio 4.1.15. Consideriamo la catena di Markov avente spazio degli stati $\mathcal{S} = \{1, 2, 3, 4\}$ e matrice di transizione

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

La classe costituita dagli stati 1 e 2 è una classe irriducibile. Lo stato 3 può raggiungere gli stati 1 e 2 ma non può essere raggiunto da questi. Lo stato 4 è uno stato assorbente, quindi forma una classe irriducibile, può essere raggiunto dallo stato 3. Gli stati 1, 2 e 4 costituiscono una classe chiusa che però non è irriducibile.

4.2 Criteri per la classificazione degli stati

Abbiamo dato la definizione di stato ricorrente e stato transitorio. Adesso cerchiamo dei criteri per stabilire di quale tipo siano gli stati di una catena senza dover ricorrere per ogni stato al calcolo delle probabilità f_{ii} .

Teorema 4.2.1. *Se lo stato i comunica con j ma lo stato j non comunica con i , allora lo stato i è transitorio.*

Dimostrazione. È equivalente a dimostrare che se lo stato i è ricorrente e i comunica con j , allora lo stato j comunica con lo stato i . Sia $K = \min\{k : P_{ij}^k > 0\}$, il più piccolo numero di passi necessario per andare da i a j . Allora esistono $K - 1$ stati i_1, i_2, \dots, i_{K-1} tali che

$$P_{ii_1} P_{i_1 i_2} \cdots P_{i_{K-1} i} > 0$$

K è il minimo, per cui tutti gli stati i_k sono diversi da j altrimenti ci sarebbe un cammino più breve per andare a i a j . Se per assurdo i fosse ricorrente varrebbe la seguente

$$0 = \mathbf{P}(\tau_i = \infty | X_0 = i) \geq P_{ii_1} P_{ii_2} \cdots P_{ii_{K-1}} P_{i_{K-1}j} (1 - f_{ji})$$

e quindi $1 - f_{ji} = 0$, da cui $f_{ji} = 1$, che è in contraddizione con l'ipotesi che j non comunichi con i . Da cui l'assurdo e quindi i è transitorio. \square

Vale inoltre il seguente risultato

Teorema 4.2.2. *Se i è uno stato ricorrente e i comunica con j ($i \rightarrow j$), allora j è uno stato ricorrente.*

Dimostrazione. Se lo stato i comunica con j esiste un n tale che $P_{ij}^n > 0$. Supponiamo che questo n sia il più piccolo per il quale i comunica con j . Allora poiché i è ricorrente, la catena torna in i infinitamente spesso ripassando da j infinitamente spesso con una probabilità positiva, quindi j è ricorrente. \square

4.3 Catene con un numero finito di stati

Le catene con un numero finito di stati sono quelle che si incontrano più spesso nelle applicazioni. Denotiamo con \mathcal{S} lo spazio degli stati di tale catena.

Teorema 4.3.1. *Se \mathcal{C} è una classe con un numero finito di stati, chiusa e irriducibile, allora tutti gli stati sono ricorrenti.*

Per dimostrare questo teorema serve il seguente risultato. Indichiamo con N_j la variabile che conta il numero di visite nello stato j .

Teorema 4.3.2. *Sia j uno stato ricorrente. Allora $\mathbf{P}(N_j = +\infty | X_0 = j) = 1$. Inoltre*

$$\mathbf{P}(N_j = +\infty | X_0 = i) = \mathbf{P}(\tau_j < +\infty | X_0 = i) = f_{ij}, \quad \text{per ogni } i \in \mathcal{S}.$$

Se $f_{ij} = 0$ allora $\mathbf{E}(N_j | X_0 = i) = 0$ mentre se $f_{ij} > 0$ allora $\mathbf{E}(N_j | X_0 = i) = +\infty$. Sia j uno stato transitorio. Allora

$$\mathbf{P}(N_j < +\infty | X_0 = i) = 1 \quad \text{per ogni } i \in \mathcal{S},$$

inoltre

$$\mathbf{E}(N_j | X_0 = i) = \frac{f_{ij}}{1 - f_{jj}} \quad \text{per ogni } i \in \mathcal{S}.$$

Dimostrazione. Se lo stato i è transitorio, la probabilità che, dato che la catena è partita da i , vi faccia ritorno m volte con $m \geq 0$ è data da

$$\mathbf{P}(N_i = m | X_0 = i) = (f_{ii})^m (1 - f_{ii}), \quad m \geq 0.$$

Se ne deduce che, nel caso in cui i sia uno stato transitorio, il valore atteso condizionato a $X_0 = i$ di N_i risulta

$$\mathbf{E}(N_i | X_0 = i) = \frac{f_{ii}}{1 - f_{ii}}.$$

Quando $j = i$. Per la prima parte possiamo scrivere, se i è uno stato transitorio

$$\mathbf{P}(N_j < +\infty | X_0 = i) = \sum_{m=0}^{+\infty} \mathbf{P}(N_j = m | X_0 = i) = \sum_{m=0}^{+\infty} f_{ij}^m (1 - f_{ii}) = 1.$$

Risulta quindi

$$\mathbf{E}(N_i | X_0 = i) = \sum_{n=1}^{+\infty} \mathbf{E}(N_i^n | X_0 = i) = \sum_{n=1}^{+\infty} \mathbf{P}(X_n = i | X_0 = i) = \sum_{n=1}^{+\infty} P_{ii}^n.$$

E di conseguenza

$$\mathbf{E}(N_i | X_0 = i) = \sum_{n=1}^{+\infty} P_{ij}^n = \frac{f_{ii}}{1 - f_{ii}}.$$

Per la seconda parte osserviamo che se uno stato è ricorrente allora

$$\begin{aligned} \mathbf{P}(N_i = +\infty | X_0 = i) &= \lim_{m \rightarrow +\infty} \mathbf{P}(N_i \geq m | X_0 = i) \\ &= \lim_{m \rightarrow +\infty} \sum_{k_1=1}^{+\infty} \mathbf{P}(\tau_i = k_1 | X_0 = i) \cdots \sum_{k_m=1}^{+\infty} \mathbf{P}(\tau_i = k_m | X_0 = i) \\ &= \lim_{m \rightarrow +\infty} f_{ii}^m = 1 \end{aligned}$$

Da questo fatto si deduce che il valore atteso non può che essere infinito, in quanto la variabile casuale assume un valore infinito con probabilità 1. \square

Lemma 4.3.3. *Se una catena ha un numero finito di stati allora ci deve essere almeno uno stato ricorrente*

Dimostrazione. La dimostrazione è fatta per assurdo. Supponiamo che tutti gli stati della catena siano transitori e mostriamo che arriviamo ad una contraddizione. Indichiamo con N_j la variabile che conta il numero di visite nello stato j . Se uno stato j è transitorio ricaviamo, dal teorema 4.3.2 che

$$\mathbf{P}(N_j < +\infty | X_0 = i) = 1, \quad \forall i \in \mathcal{S}.$$

Sempre per lo stesso teorema si ha che

$$\mathbf{E}(N_j | X_0 = i) = \frac{f_{ij}}{1 - f_{jj}} < +\infty, \quad \forall i \in \mathcal{S},$$

ed inoltre si deduce che $\sum_{n=i}^{+\infty} P_{ij}^n < +\infty$ e quindi che $\lim_{n \rightarrow +\infty} P_{ij}^n = 0$. Allora possiamo scrivere

$$\begin{aligned} 0 &= \sum_{j \in \mathcal{S}} \lim_{n \rightarrow +\infty} P_{ij}^n = \lim_{n \rightarrow +\infty} \sum_{j \in \mathcal{S}} P_{ij}^n \\ &= \lim_{n \rightarrow +\infty} \mathbf{P}(X_n \in \mathcal{S} | X_0 = i) = \lim_{n \rightarrow +\infty} 1 = 1, \end{aligned}$$

da cui la contraddizione. □

Se una catena ha un numero finito di stati ed è irriducibile, allora per ogni coppia di stati i e j esiste un $n > 0$ tale che $P_{ij}^n > 0$. In particolare se la catena ha un numero finito di stati esiste un n tale che la matrice $P^{(n)}$ ha tutti gli elementi positivi.

Per le catene con un numero finito di stati il Teorema 4.2.1 e il Teorema 4.3.1 permettono di decidere se uno stato è transitorio o ricorrente senza dover studiare il comportamento delle probabilità di transizione. Possiamo enunciare il seguente risultato

Teorema 4.3.4. *Se lo spazio degli stati \mathcal{S} è finito, allora possiamo scrivere*

$$\mathcal{S} = T \cup R_1 \cup R_2 \cup \dots \cup R_k,$$

dove T è la classe degli stati transitori e R_1, \dots, R_k sono k classi chiuse irriducibili.

Cerchiamo ora di utilizzare questo criterio per classificare gli stati delle catene di Markov.

Esempio 4.3.5. Classifichiamo gli stati per la catena che descrive la rovina di un giocatore. Se a è il capitale del giocatore A e b il capitale del giocatore B allora la catena di Markov che descrive il capitale del giocatore A ha spazio degli stati dato da $\mathcal{S} = \{0, 1, 2, \dots, N\}$, con $N = a + b$. Uno stato i tale che $0 < i < N$ comunica con ogni altro stato. Lo stato 0 è ricorrente in quanto $P_{00}^n = 1$ per ogni n per cui la serie $\sum P_{00}^n$ diverge. Analogamente lo stato N . Poiché gli stati 0 ed N comunicano solo con loro stessi sono assorbenti. Tutti gli altri stati $0 < i < N$ sono invece transitori poiché ad esempio lo stato i può raggiungere lo stato 0 ma non vale il viceversa. Abbiamo $\mathcal{S} = T \cup R_1 \cup R_2$, dove $T = \{1, \dots, N - 1\}$, $R_1 = \{0\}$ e $R_2 = \{N\}$.

Esempio 4.3.6. Consideriamo una catena di Markov avente 9 stati la cui matrice di transizione è la seguente, dove abbiamo indicato con \star la presenza di un elemento maggiore di 0.

$$\begin{array}{c}
 [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9] \\
 \begin{array}{c}
 \left[\begin{array}{c}
 1 \\
 2 \\
 3 \\
 4 \\
 5 \\
 6 \\
 7 \\
 8 \\
 9
 \end{array} \right]
 \left[\begin{array}{cccccccc}
 & & & \star & & & & \star \\
 & \star & \star & & \star & & & \star \\
 & & & & & & & \star \\
 \star & & & & & & & & \\
 & & & & \star & & & & \\
 & \star & & & & & & & \\
 & \star & & & \star & \star & \star & & \\
 & & \star & & & & & & \\
 & & & \star & & & & & \star
 \end{array}
 \right]
 \end{array}
 \end{array}$$

Lo stato 5 è assorbente. Gli stati 3 e 8 costituiscono una classe chiusa e sono entrambi ricorrenti. Anche gli stati 1, 4 e 9 costituiscono una classe chiusa. Entrambe queste classi sono irriducibili e finite per cui tutti gli stati in esse sono ricorrenti. Gli stati 2, 6 e 7 sono invece transitori poiché riusciamo a trovare per ognuno di essi uno stato che loro riescono a raggiungere ma dal quale non sono raggiungibili. Si osservi che la classe composta dagli stati 2, 3, 5 e 8 è una classe chiusa (detta anche chiusura di 2) ma non è irriducibile. Abbiamo quindi $T = \{2, 6, 7\}$ la classe degli stati transitori, $R_1 = \{5\}$, $R_2 = \{3, 8\}$ e $R_3 = \{1, 4, 9\}$ e $\mathcal{S} = \{1, 2, \dots, 9\} = T \cup R_1 \cup R_2 \cup R_3$.

Si osservi che se dallo stato 1 si potesse andare solo nello stato 4, allora lo stato 9, sarebbe uno stato transitorio, gli stati 1 e 4 costituirebbero una classe chiusa e irriducibile. In questo caso si avrebbe $T = \{2, 6, 7, 9\}$ la classe degli stati transitori, $R_1 = \{5\}$, $R_2 = \{3, 8\}$ e $R_3 = \{1, 4\}$ e $\mathcal{S} = \{1, 2, \dots, 9\} = T \cup R_1 \cup R_2 \cup R_3$.

Esempio 4.3.7. Riprendiamo l'Esempio 4.1.8 Consideriamo una catena di Markov avente 4 stati la cui matrice di transizione è la seguente,

$$\begin{array}{c}
 \left[\begin{array}{cccc}
 1/4 & 1/4 & 1/4 & 1/4 \\
 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 1 \\
 1 & 0 & 0 & 0
 \end{array} \right]
 \end{array}$$

Mostriamo che tutti gli stati sono ricorrenti con i criteri presentati in questo paragrafo e nel precedente. Tutti gli stati comunicano tra di loro. Si tratta di una catena chiusa irriducibile con un numero finito di stati. Una catena siffatta deve avere almeno uno stato ricorrente, tutti gli stati comunicano tra di loro quindi sono tutti stati ricorrenti.

Capitolo 5

Distribuzioni invarianti

5.1 Definizione e prime proprietà

Lo studio del comportamento a lungo termine del sistema descritto da una catena di Markov è importante per due motivi. In primo luogo perché si cerca di determinare una distribuzione “di equilibrio che descriva lo stato del sistema dopo molto tempo. In questo caso l’interesse sta nel capire se il sistema ammette una tale distribuzione di equilibrio. In secondo luogo perché tale distribuzione è una buona approssimazione del comportamento del sistema anche per istanti grandi ma finiti ed è di più immediata comprensione che non la legge del sistema all’istante fissato. In questo corso troveremo la distribuzione invariante (i cui sinonimi sono distribuzione stazionaria o distribuzione di equilibrio) per catene di Markov con struttura particolare che andremo tra poco a definire. Diamo invece ora la definizione di distribuzione invariante e studiamone qualche proprietà. Supponiamo data una catena di Markov con spazio degli stati \mathcal{S} , finito o infinito e matrice di transizione P .

Definizione 5.1.1. *Una distribuzione π su \mathcal{S} , cioè una misura di probabilità su \mathcal{S} , è detta distribuzione invariante se soddisfa la relazione*

$$\pi = \pi P.$$

L’esistenza di una distribuzione invariante implica che il comportamento del sistema descritto dalla catena sia molto stabile.

Proposizione 5.1.2. *Supponiamo che una catena di Markov abbia una distribuzione invariante $\pi = (\pi_1, \pi_2, \dots)$ e che la distribuzione iniziale della catena di Markov sia questa distribuzione invariante, cioè $\pi_j = \mathbf{P}(X_0 = j)$ per ogni $j \in \mathcal{S}$. Allora la distribuzione all’istante n è sempre π per ogni n .*

Dimostrazione. Sappiamo che la distribuzione all'istante n , denotata con ν^n , è data da

$$\nu_j^n = \sum_{i \in \mathcal{S}} \pi_i P_{ij}^n,$$

dove in questo caso la distribuzione iniziale è π . Ora dimostriamo per induzione che se π è una distribuzione invariante allora

$$\sum_{i \in \mathcal{S}} \pi_i P_{ij}^n = \pi_j, \quad \forall n, \quad \forall j.$$

Questo implicherà che

$$\nu_j^n = \pi_j, \quad \forall n, \quad \forall j.$$

Infatti per definizione di distribuzione invariante è vero per $n = 1$ che $\sum_{i \in \mathcal{S}} \pi_i P_{ij} = \pi_j$. Supponiamo che per $n - 1$ valga che $\sum_{i \in \mathcal{S}} \pi_i P_{ij}^{n-1} = \pi_j$, allora la relazione vale per n qualunque, infatti

$$\sum_{i \in \mathcal{S}} \pi_i P_{ij}^n = \sum_{i \in \mathcal{S}} \pi_i \sum_{h \in \mathcal{S}} P_{ih}^{n-1} P_{hj} = \sum_{h \in \mathcal{S}} \sum_{i \in \mathcal{S}} \pi_i P_{ih}^{n-1} P_{hj} = \sum_{h \in \mathcal{S}} \pi_h P_{hj} = \pi_j.$$

□

Quindi si deduce che se la distribuzione iniziale è una distribuzione invariante allora ogni istante ha come distribuzione la stessa distribuzione invariante. Nel seguito di questo paragrafo vogliamo dare dalle condizioni per riuscire a trovare, se esiste, la distribuzione invariante di una catena di Markov. A priori non è detto che tale distribuzione esista infatti la soluzione del sistema $\pi = \pi P$ è soggetta ai vincoli $\pi_j \geq 0$ e $\sum_{j \in \mathcal{S}} \pi_j = 1$, affinché essa sia effettivamente una distribuzione sullo spazio degli stati \mathcal{S} . Inoltre se esiste una tale distribuzione, non è detto che questa sia unica. Ad esempio per la catena di Markov della rovina del giocatore le distribuzioni $\pi = (1, 0, 0, \dots, 0)$ e $\tilde{\pi} = (0, 0, \dots, 0, 1)$ sono entrambe stazionarie. Ne consegue che ogni combinazione convessa di esse, vale a dire ogni distribuzione del tipo $\pi^* = \alpha\pi + (1 - \alpha)\tilde{\pi}$ per ogni $0 < \alpha < 1$, è una distribuzione invariante per questo sistema. Ne deduciamo che se esistono due distribuzioni invarianti in realtà ne esistono infinite. È evidente che se non esistono o esistono infinite misure invarianti, la ricerca di esse perde il suo fine di descrivere un comportamento a lungo termine del sistema. Dobbiamo quindi caratterizzare le catene che possiedono una tale distribuzione e determinare delle condizioni per cui la distribuzione invariante se esiste sia anche unica. Vedremo che questa caratterizzazione si riallaccia al problema della classificazione degli stati di una catena di Markov.

Ricordiamo la variabile N_j che conta il numero di visite allo stato j . Possiamo scrivere

$$N_j = \sum_{n=1}^{+\infty} I_{\{X_n=j\}}$$

dove $I_{\{X_n=j\}}$ è la variabile indicatore dell'evento $\{X_n = j\}$, che quindi vale 1 se si verifica l'evento $\{X_n = j\}$. Vale il seguente risultato

Teorema 5.1.3. *Sia N_j la variabile che conta il numero di visite allo stato j . Allora vale la seguente relazione*

$$\mathbf{E}(N_j|X_0 = i) = \sum_{n=1}^{+\infty} P_{ij}^n$$

Dimostrazione. Abbiamo che

$$\mathbf{E}(N_j|X_0 = i) = \mathbf{E}\left(\sum_{n=1}^{+\infty} I_{\{X_n=j\}}|X_0 = i\right)$$

Ora il valore atteso condizionato della variabile indicatore è dato dalla probabilità condizionata, vale cioè la relazione $\mathbf{E}(I_A|B) = P(A|B)$. Poiché il valore atteso è lineare, il valore atteso condizionato di una somma (anche infinita), è uguale alla somma dei valori attesi condizionati. Abbiamo quindi

$$\mathbf{E}(N_j|X_0 = i) = \sum_{n=1}^{+\infty} \mathbf{E}(I_{\{X_n=j\}}|X_0 = i) = \sum_{n=1}^{+\infty} \mathbf{P}(X_n = j|X_0 = i) = \sum_{n=1}^{+\infty} P_{ij}^n.$$

□

Ricordando che per il Teorema 4.3.2 abbiamo

$$\mathbf{E}(N_j|X_0 = i) = \frac{f_{ij}}{1 - f_{jj}} \quad \text{per ogni } i \in \mathcal{S}.$$

ne deduciamo il seguente risultato

Teorema 5.1.4. *Lo stato j è ricorrente se e solo se*

$$\sum_{n=1}^{+\infty} P_{ij}^n = \mathbf{E}(N_j|X_0 = j) = +\infty$$

La dimostrazione è ovvia per i Teorema 5.1.3 precedente e per il fatto che $f_{jj} = 1$ se e solo se lo stato j è ricorrente.

Quindi se j è ricorrente $N_j = \infty$ con probabilità 1. Se invece lo stato j è transitorio $N_j < \infty$ con probabilità 1 e il valore atteso è finito. Inoltre vale la seguente proprietà

$$\lim_{n \rightarrow \infty} P_{jj}^n = 0$$

in quanto termine generale di una serie convergente.

Mostriamo che se lo stato j è ricorrente la successione P_{jj}^n converge ad un numero positivo e questo numero positivo è proprio il valore della distribuzione invariante nello stato j . Purtroppo c'è un problema che si presenta per alcune catene. Consideriamo la catena di Ehrenfest per la diffusione nel caso in cui ci siano solo tre palline. Vi sono 4 stati, il numero di palline contenute in una delle urne e la matrice di transizione è la seguente

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

La matrice di transizione in due passi è invece

$$P^2 = \begin{bmatrix} 1/3 & 0 & 2/3 & 0 \\ 0 & 7/9 & 0 & 2/9 \\ 2/9 & 0 & 7/9 & 0 \\ 0 & 2/3 & 0 & 1/3 \end{bmatrix}$$

Si noti che se inizialmente nell'urna vi è un numero pari di palline, nell'istate successivo ve ne saranno un numero dispari e viceversa. Questa alternanza implica che è impossibile tornare nello stato da cui si è partiti in un numero dispari di passi. In questo caso quindi, per tutti gli stati j , si ha $P_{jj}^n = 0$ per ogni n dispari, mentre $P_{jj}^n \neq 0$ per ogni n pari, e quindi non la successione $\{P_{jj}^n\}$ non ammette limite, anche se tutti gli stati di questa catena sono ricorrenti.

Per ovviare a questo inconveniente un primo passo è costituito dalla seguente definizione che suddivide gli stati di una catena di Markov in base alla periodicità con cui la catena ripassa per uno stato visitato in precedenza.

Definizione 5.1.5. *Uno stato j ha periodo p se p è il più grande divisore di ogni n tale per cui $P_{jj}^n > 0$. Lo stato j è aperiodico se $p = 1$.*

In pratica p è il massimo comune divisore del numero di passi che una catena deve compiere per ritornare nello stesso stato. Se indichiamo con I_j l'insieme

$$I_j = \{n \geq 1 : P_{jj}^n > 0\}$$

il periodo p è il massimo comune divisore dei numeri che stanno in I_j . Ad esempio nel caso descritto abbiamo che

$$I_j = \{n \geq 1 : P_{jj}^n > 0\} = \{2, 4, \dots, 2k, \dots\}, \quad j = 0, 1, 2, 3$$

e quindi tutti gli stati hanno periodo 2. Anche nella passeggiata aleatoria con due barriere riflettenti tutti gli stati hanno periodo $p = 2$.

Definizione 5.1.6. *Una catena è aperiodica se $p = 1$ per ogni stato della catena.*

Le catene aperiodiche avranno un ruolo fondamentale nel seguito della teoria e quindi diamo dei criteri per riuscire a capire quando una catena è aperiodica.

Teorema 5.1.7. *Sia j uno stato della catena di Markov. Se $P_{jj} > 0$, allora il periodo dello stato j è 1.*

La dimostrazione è ovvia in quanto in I_j c'è il valore $n = 1$.

Si noti che gli stati di una catena possono avere tutti periodo 1 anche se per tutti gli stati si ha $P_{jj} = 0$, vale a dire la condizione $P_{jj} > 0$ è solo sufficiente perché lo stato abbia periodo 1 ma non è necessaria.

Abbiamo il seguente risultato

Teorema 5.1.8. *Se uno stato j ha periodo 1, allora da un certo istante n_0 in poi si ha $P_{jj}^n > 0$ per ogni $n > n_0$.*

La dimostrazione non viene riportata in quanto sfrutta risultati della teoria dei numeri che esulano dalle conoscenze richieste. Il teorema ha però un'utile applicazione in quanto se una catena ha un numero finito di stati ed è aperiodica, allora esiste un n_0 tale che per ogni $n > n_0$ tutti gli elementi della matrice P^n sono positivi.

Abbiamo già osservato che se due stati comunicano allora sono dello stesso tipo (quindi se siamo in grado di stabilire se uno stato è ricorrente o transitorio, possiamo ricavare di che tipo sono gli altri stati in base alle relazioni dicomunicabilità di questi con lo stato di cui conosciamo il tipo). Il prossimo risultato fornisce un risultato analogo per la nozione di periodo.

Teorema 5.1.9. *Se due stati comunicano uno con l'altro ($i \rightarrow j$ e $j \rightarrow i$) allora hanno lo stesso periodo.*

Anche la dimostrazione di questo teorema si basa su proprietà derivanti dalla teoria dei numeri. Il risultato può però essere compreso intuitivamente, perché se per assurdo supponiamo che i e j abbiano periodo diverso allora per compiere il tragitto da $i \rightarrow i$, potremmo pensare di andare da $i \rightarrow j$, poi da $j \rightarrow j$ e quindi da

$j \rightarrow i$. Ma se j avesse periodo diverso da quello di i non sarebbe possibile questo cammino.

In base ai risultati precedenti si deduce che se una catena è irriducibile tutti gli stati hanno lo stesso periodo. Possiamo quindi dire che se una catena è irriducibile tutti i suoi stati sono o tutti ricorrenti o tutti transitori e tutti hanno lo stesso periodo o sono aperiodici. In alcuni testi le catene di Markov irriducibili e aperiodiche sono dette regolari.

5.2 Esistenza della distribuzione invariante

Enunciamo ora il teorema chiave di tutto il paragrafo e che caratterizza le catene che hanno una distribuzione invariante unica.

Teorema 5.2.1. *Se una catena è irriducibile e aperiodica ed ammette una distribuzione invariante, cioè esiste una soluzione del sistema*

$$\sum_{i \in \mathcal{S}} \pi_i P_{ij} = \pi_j, \quad \forall j \in \mathcal{S}, \quad \sum_{i \in \mathcal{S}} \pi_i = 1, \quad \pi_i \geq 0,$$

allora tutti gli stati sono ricorrenti e

$$\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j, \quad \forall i, j.$$

La distribuzione stazionaria π è inoltre unica.

Tale teorema ci dice che se una catena è irriducibile e aperiodica e ammette almeno una misura invariante allora tutti i suoi stati sono ricorrenti e inoltre la misura invariante è ottenuta come limite delle probabilità di transizione in n passi indipendentemente dallo stato iniziale. Infatti il limite che compare nel teorema non dipende dallo stato i in cui si parte. In particolare se sono soddisfatte le ipotesi del teorema la distribuzione dell'istante n , ν^n , converge alla distribuzione stazionaria indipendentemente dalla distribuzione iniziale π^0 . Nel senso che

$$\lim_{n \rightarrow \infty} \nu_j^n = \lim_{n \rightarrow \infty} \sum_{i \in \mathcal{S}} \pi_i^0 P_{ij}^n = \sum_{i \in \mathcal{S}} \pi_i^0 \lim_{n \rightarrow \infty} P_{ij}^n = \sum_{i \in \mathcal{S}} \pi_i^0 \pi_j = \pi_j.$$

Nel caso in cui lo spazio degli stati sia finito abbiamo delle notevoli semplificazioni. Innanzitutto l'ipotesi che la catena sia irriducibile implica che tutti gli stati siano ricorrenti senza bisogno di dimostrazione. Per catene con spazio degli stati finito si può inoltre dimostrare che una soluzione del sistema $\pi = \pi P$, tale che $\sum_{i \in \mathcal{S}} \pi_i = 1$ e $\pi_i \geq 0$ esiste sempre.

Cerchiamo ora di riallacciare il problema della ricerca di una distribuzione invariante con la variabile casuale che conta il numero di visite che la catena fa ad uno stato j prima dell'istante n . Nel paragrafo 4 abbiamo definito la variabile casuale N_i^n che conta il numero delle visite allo stato i prima di un istante n e la variabile casuale N_i che conta il numero delle visite allo stato i . Avevamo inoltre definito la variabile τ_i che ci da il tempo di primo passaggio della catena per stato lo stati i . Ricordiamo che il tempo medio di ritorno nello stato i è dato da

$$\mathbf{E}(\tau_i | X_0 = i) = \sum_{n=1}^{+\infty} n f_{ii}^n$$

dove f_{ii}^n è la probabilità che una catena partita dallo stato i ritorni nello stato i per la prima volta dopo n istanti. Tale valore medio ha senso solo se f_{ii}^n è una distribuzione di probabilità. E questo avviene solo se lo stato i è ricorrente poiché vale che $\sum_{n=1}^{+\infty} f_{ii}^n = 1$. Indichiamo con

$$\mu_i = \mathbf{E}(\tau_i | X_0 = i) = \sum_{n=1}^{+\infty} n f_{ii}^n$$

il tempo medio di ritorno in i , dato che la catena è partita dallo stato i . Diamo la seguente definizione

Definizione 5.2.2. *Uno stato i di una catena di Markov ricorrente per cui*

$$\mu_i = \sum_{n=1}^{+\infty} n f_{ii}^n < +\infty$$

è detto ricorrente positivo. Uno stato i di una catena di Markov ricorrente per cui

$$\mu_i = \sum_{n=1}^{+\infty} n f_{ii}^n = +\infty$$

è detto ricorrente nullo.

La necessità di tale definizione risulta dal fatto che se uno stato è ricorrente (cioè la catena ritorna in quello stato infinite volte) può accadere che il tempo di attesa per questi ritorni sia infinitamente lungo e in questo caso lo stato è detto ricorrente nullo. Se invece il tempo per tale ritorno è finito lo stato è detto ricorrente positivo.

Definizione 5.2.3. *Uno stato i di una catena di Markov che sia ricorrente positivo ed aperiodico è detto ergodico.*

Vediamo come i tempi medi di assorbimento si riallacciano alle distribuzioni invarianti poiché se solo le catene irriducibili i cui stati sono ricorrenti positivi ammettono una misura invariante e solo queste ultime catene ammettono un tempo medio di ritorno finito ci aspettiamo che un qualche legame tra distribuzione invariante e tempo medio di ritorno debba esserci.

Lemma 5.2.4. *Supponiamo che uno stato i di una catena di Markov sia ricorrente e che*

$$\lim_{n \rightarrow +\infty} P_{ii}^n = u.$$

Allora $u > 0$ se e solo se $\mu_i < +\infty$ e in questo caso

$$\lim_{n \rightarrow +\infty} P_{ii}^n = \frac{1}{\mu_i}.$$

Il risultato di questo lemma è facilmente interpretabile nel modo seguente. Se μ_i è il tempo medio tra una visita e l'altra del sistema in uno stato i , allora il numero medio di visite allo stato i prima del tempo n saranno approssimativamente

$$\mathbf{E}(N_i^n) \approx \frac{n}{\mu_i}.$$

Da questo fatto si deduce allora che

$$\mathbf{E}\left(\frac{N_i^n}{n}\right) \approx \frac{1}{\mu_i}.$$

Del resto ricordiamo che

$$\mathbf{E}(N_i^n) = \sum_{k=1}^n P_{ii}^k$$

e quindi

$$\mathbf{E}\left(\frac{N_i^n}{n}\right) = \frac{1}{n} \sum_{k=1}^n P_{ii}^k.$$

Allora ne consegue che

$$\lim_{n \rightarrow +\infty} \mathbf{E}\left(\frac{N_i^n}{n}\right) = \frac{1}{\mu_i} = \pi_i,$$

cioè la distribuzione invariante può essere ottenuta come limite, per n che tende all'infinito, della proporzione di istanti su n che la catena passa in ogni stato. In definitiva il Lemma contiene l'importante caratterizzazione delle catene che ammettono una distribuzione invariante. Infatti il teorema precedente afferma che se una catena è irriducibile ed aperiodica ed ammette una distribuzione invariante allora tale distribuzione invariante è ottenuta come limite delle probabilità di transizione

in n passi. Il Lemma ci dice che se il limite delle probabilità di transizione in n passi per uno stato ricorrente esiste, esso è finito solo se lo stato è ricorrente positivo. Abbiamo quindi il seguente risultato definitivo.

Teorema 5.2.5. *Per una catena irriducibile e aperiodica abbiamo solo tre possibilità.*

1. *La catena è transitoria. Allora per ogni i e j*

$$\lim_{n \rightarrow +\infty} P_{ij}^n = 0$$

inoltre $\sum_{n=1}^{+\infty} P_{ij}^n < +\infty$ e in questo caso non esiste misura invariante.

2. *La catena è ricorrente ma non esiste una misura invariante. Allora per ogni i e j*

$$\lim_{n \rightarrow +\infty} P_{ij}^n = 0,$$

inoltre $\sum_{n=1}^{+\infty} P_{ij}^n = +\infty$ ma $\mu_j = \sum_{n=1}^{+\infty} n f_{jj}^n = +\infty$. La catena è ricorrente nulla.

3. *La catena è ricorrente ed esiste una misura invariante. Allora per ogni i e j*

$$\lim_{n \rightarrow +\infty} P_{ij}^n = \pi_j > 0,$$

inoltre $\mu_j = \sum_{n=1}^{+\infty} n f_{jj}^n = \frac{1}{\pi_j}$. La catena è ricorrente positiva.

5.3 Calcolo della distribuzione invariante

Per le catene di Markov con un numero finito di stati il calcolo della distribuzione invariante, quando esiste ed è unica è sempre possibile, anche se, non appena il numero degli stati cresce la complessità dei calcoli da svolgere è piuttosto noiosa. Se una catena è inoltre irriducibile e aperiodica, possiamo calcolare la distribuzione invariante in tre modi.

1. Risolvendo il sistema lineare

$$\sum_{i \in \mathcal{S}} \pi_i P_{ij} = \pi_j, \quad \forall j \in \mathcal{S},$$

con i vincoli

$$\sum_{i \in \mathcal{S}} \pi_i = 1, \quad \pi_i \geq 0.$$

2. Calcolando il limite delle probabilità di transizione a più passi

$$\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j, \quad \forall i, j \in \mathcal{S}.$$

3. Calcolando il limite delle distribuzioni marginali $\nu_j^n = \mathbf{P}(X_n = j)$:

$$\lim_{n \rightarrow \infty} \nu_j^n = \pi_j \quad \forall j \in \mathcal{S}.$$

4. Calcolando $\mu_i = \sum_{n=1}^{+\infty} n f_{ii}^n$ e se finito, ponendo $\pi_i = 1/\mu_i$

La distribuzione invariante è indipendente dalla distribuzione iniziale, e questo fatto risulta di notevole importanza nelle applicazioni.

Vediamo con alcuni esempi di trovare i metodi più veloci e comodi.

Esempio 5.3.1. Catene di Markov con due stati Abbiamo studiato nel dettaglio queste catene nel paragrafo 3.5. Si tratta di una catena di Markov irriducibile. Lo spazio degli stati è $\mathcal{S} = \{0, 1\}$ e la matrice di transizione

$$P = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix}$$

Per trovare la distribuzione invariante applichiamo nell'ordine i primi tre metodi riassunti sopra.

Il primo metodo consiste nel risolvere il sistema $\pi P = \pi$, dove abbiamo denotato la distribuzione invariante con $\pi = (\pi_0, \pi_1)$, con i vincoli $\pi_0 + \pi_1 = 1$ e $\pi_0 > 0$ e $\pi_1 > 0$. Dobbiamo cioè trovare le soluzioni positive del sistema

$$\begin{cases} (1 - \alpha)\pi_0 + \beta\pi_1 = \pi_0 \\ \alpha\pi_0 + (1 - \beta)\pi_1 = \pi_1 \\ \pi_0 + \pi_1 = 1 \end{cases}$$

Dalla seconda equazione ricaviamo

$$\pi_0 = \frac{\beta}{\alpha}\pi_1,$$

che sostituita nella terza ci da

$$\frac{\beta}{\alpha}\pi_1 + \pi_1 = 1.$$

Vale a dire

$$\pi_1 = \frac{1}{1 + \frac{\beta}{\alpha}} = \frac{\alpha}{\alpha + \beta}$$

e infine

$$\pi_0 = \frac{\beta}{\alpha + \beta}$$

La distribuzione invariante risulta pertanto

$$\pi = \left(\frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta} \right).$$

Il secondo metodo consiste nel calcolare il limite della matrice P^n per n tendente all'infinito. Sappiamo dall'equazione (3.4) che

$$P^n = \frac{(1 - \alpha - \beta)^n}{\alpha + \beta} \begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix} + \frac{1}{\alpha + \beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix},$$

e quindi ricaviamo

$$\lim_{n \rightarrow \infty} P^n = \frac{1}{\alpha + \beta} \begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix}.$$

Come si vede otteniamo una matrice con le stesse righe. La distribuzione invariante risulta

$$\pi = \left(\frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta} \right).$$

Il terzo metodo consiste nel calcolo del limite della distribuzione marginale di X_n per n tendente all'infinito. Dalle equazioni (3.5) e (3.6) si ricava che

$$\lim_{n \rightarrow \infty} \nu_0^n = \lim_{n \rightarrow \infty} \mathbf{P}(X_n = 0) = \frac{\beta}{\alpha + \beta},$$

e

$$\lim_{n \rightarrow \infty} \nu_1^n = \lim_{n \rightarrow \infty} \mathbf{P}(X_n = 1) = \frac{\alpha}{\alpha + \beta},$$

e quindi ancora una volta otteniamo la distribuzione invariante

$$\pi = \left(\frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta} \right).$$

Esempio 5.3.2. Vediamo ora un esempio di applicazione in un campo, quello sociologico, dove le catene di Markov sono utilizzate per modellare numerosi fenomeni. Si tratta di un esempio sulla mobilità di classe dei soggetti sociali. Un modello matematico semplificato per questo fenomeno consiste nel considerare tre classi sociali, alta, media e bassa e nell'assumere che le transizioni da una classe all'altra per successive generazioni avvengano secondo lo schema di una catena di Markov. In pratica si fa l'ipotesi che la classe sociale cui apparterrà un figlio dipende solo dalla classe dei genitori, dove la probabilità di essere in una determinata delle tre

classi, dato che i genitori appartenevano ad una classe è riassunta nella seguente matrice, che è appunto la matrice di transizione della catena di Markov. Gli stati supponiamo che siano 1, 2, 3 rispettivamente per classe alta, media e bassa.

$$P = \begin{bmatrix} 0.45 & 0.48 & 0.07 \\ 0.05 & 0.70 & 0.25 \\ 0.01 & 0.50 & 0.49 \end{bmatrix}$$

Si tratta di una catena di Markov irriducibile, aperiodica con un numero finito di stati, quindi esiste un'unica misura invariante che si può calcolare risolvendo il sistema $\pi = \pi P$ dove $\pi = (\pi_1, \pi_2, \pi_3)$, con la condizione $\pi_1 + \pi_2 + \pi_3 = 1$, che scritto per esteso diventa

$$\begin{cases} 0.45\pi_1 + 0.05\pi_2 + 0.01\pi_3 & = \pi_1 \\ 0.48\pi_1 + 0.70\pi_2 + 0.50\pi_3 & = \pi_2 \\ 0.07\pi_1 + 0.25\pi_2 + 0.49\pi_3 & = \pi_3 \\ \pi_1 + \pi_2 + \pi_3 & = 1, \end{cases}$$

con $\pi_i > 0$. La soluzione risulta

$$\pi_1 = 0.0624; \quad \pi_2 = 0.6234; \quad \pi_3 = 0.3142.$$

In questo caso non è semplice andare a calcolare il limite per n tendente all'infinito degli elementi della matrice P^n . Il risultato ottenuto si può interpretare dicendo che in una società in cui si supponga che i movimenti tra classi sociali possano essere descritti da una catena di Markov con matrice di transizione P , avremo che in una situazione di equilibrio la popolazione sarà così suddivisa: il 6% apparterrà alla classe alta, il 62% alla classe media e il 32% alla classe bassa.

Un metodo alternativo ai quattro dati consiste nello sfruttare le proprietà della matrice di transizione. Senza entrare nei dettagli (per chi è interessato rimandiamo al testo di Cox e Miller) vale la pena di discutere questo metodo che risulta molto comodo dal punto di vista computazionale.

Supponiamo che una catena di Markov con un numero finito di stati sia irriducibile e aperiodica. Sia P la matrice di transizione di questa catena. Le proprietà della matrice P , cioè il fatto che sia irriducibile, aperiodica e la somma degli elementi di tutte le righe sia pari a uno, implicano che la matrice P ammette l'autovalore 1 con molteplicità 1 e tutti gli altri autovalori sono in modulo minori di 1. Allora l'autovettore associato all'autovalore 1, una volta normalizzato, denotato con π soddisfa la relazione $\pi P = \pi$, vale a dire è la distribuzione invariante per la catena di Markov con matrice di transizione P (π è unica poiché la catena è irriducibile e aperiodica).

Esempio 5.3.3. Supponiamo che lo stato occupazionale nel periodo n dei membri di una popolazione di lavoratori che, in termini di valore come risorsa umana, sono classificati di formazione alta, sia rappresentato da una catena di Markov $\{X_n\}$ omogenea con tre stati:

stato 1: *lavoratore con formazione di tipo alto con occupazione di livello basso,*

stato 2: *lavoratore con formazione di tipo alto con occupazione di livello alto,*

stato 3: *lavoratore con formazione di tipo alto disoccupato*

Supponendo che la distribuzione iniziale sia $\mu_0 = [0.75, 0.10, 0.15]$ e la matrice di transizione sia

$$N = \begin{bmatrix} 0.9 & 0 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.3 & 0.6 & 0.1 \end{bmatrix}$$

Gli autovalori associati a tale matrice sono

$$\lambda_1 = 1, \quad \lambda_2 = 0.7772, \quad \lambda_3 = -0.0772.$$

L'autovettore associato all'autovalore 1 è

$$v = [-1.0825, -0.4330, -0.2165],$$

che normalizzato diviene

$$\pi = [0.6250, 0.2500, 0.1250].$$

Allo stesso risultato si arriva risolvendo il sistema

$$\begin{cases} 0.9\pi_1 + 0.1\pi_2 + 0.3\pi_3 & = \pi_1 \\ +0.7\pi_2 + 0.6\pi_3 & = \pi_2 \\ 0.1\pi_1 + 0.2\pi_2 + 0.1\pi_3 & = \pi_3 \\ \pi_1 + \pi_2 + \pi_3 & = 1, \end{cases}$$

Indicata con $\pi = [\pi_1, \pi_2, \pi_3]$ la distribuzione invariante incognita e con A la matrice dei coefficienti delle ultime tre equazioni, il sistema scritto in forma matriciale diviene $A\pi = b$, dove $b = [0, 0, 1]$ è il vettore dei termini noti e la matrice A è data da

$$A = \begin{bmatrix} 0 & 0.3 & -0.6 \\ -0.1 & -0.2 & 0.9 \\ 1 & 1 & 1 \end{bmatrix}.$$

La soluzione di questo sistema risulta

$$\pi = [0.6250, 0.2500, 0.1250].$$

Anche in questo caso il significato che diamo alla distribuzione invariante è immediato. A lungo termine, in una situazione di equilibrio, ci aspettiamo che circa il 62,5 % della popolazione dei lavoratori con formazione alta, abbia un'occupazione consona al suo livello di preparazione, il 25 % abbia un'occupazione al di sotto delle sue capacità e il restante 12,5 %, sia non occupato.

Per quanto riguarda il quanto metodo, poiché non è facile calcolare il valore μ_i direttamente, viene utilizzato approssimativamente. Ad esempio un'approssimazione della distribuzione invariante per il processo simulato, potrà essere ottenuta come numero medio di visite in uno stato che la catena di Markov effettua in quello stato rispetto al tempo totale in cui osserviamo la traiettoria. Analizzeremo questa possibilità nel capitolo sulle simulazioni.

5.4 Il caso in cui lo spazio degli stati è infinito

Solo per catene con spazi degli stati infinito si possono avere tutti gli stati transitori. Il Teorema 5.2.1 ci dice che, nel caso in cui la catena abbia spazio degli stati infinito, se una misura invariante esiste, allora la catena è ricorrente e la misura invariante (trovata risolvendo il sistema $\pi = \pi P$) è unica. Cosa accade se la misura invariante non esiste? Abbiamo il seguente risultato leggermente più generale.

Teorema 5.4.1. *Se una catena è irriducibile e aperiodica e non ammette una distribuzione invariante, cioè non esiste una soluzione del sistema*

$$\sum_{i \in \mathcal{S}} \pi_i P_{ij} = \pi_j, \quad \forall j \in \mathcal{S}, \quad \sum_{i \in \mathcal{S}} \pi_i = 1, \quad \pi_i \geq 0,$$

allora

$$\lim_{n \rightarrow \infty} P_{ij}^n = 0, \quad \forall i, j.$$

Se la catena (sempre con spazio degli stati infinito) è transitoria, il fatto che $\lim_{n \rightarrow \infty} P_{ij}^n = 0$, per ogni i, j , segue dal fatto che per una catena transitoria la serie $\sum_n P_{ij}^n$ converge e quindi il suo termine generale deve tendere a zero. Quello che questo teorema ci dice di nuovo è per gli stati ricorrenti. Possiamo riassumere quanto fin qui visto nel modo seguente:

- Se una catena è irriducibile, aperiodica e transitoria allora

$$\lim_{n \rightarrow \infty} P_{ij}^n = 0, \quad \forall i, j.$$

- Se una catena è irriducibile e aperiodica e non ammette una distribuzione invariante allora

$$\lim_{n \rightarrow \infty} P_{ij}^n = 0, \quad \forall i, j.$$

- Se una catena è irriducibile e aperiodica ed ammette una distribuzione stazionaria allora

$$\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j, \quad \forall i, j.$$

Gli ultimi due casi hanno come conseguenza che la catena sia ricorrente.

In questa sezione vedremo alcuni esempi di applicazioni di catene di Markov. Il primo esempio costituisce un'esempio molto semplice e che abbiamo studiato più volte. Si tratta di una catena di nascita e morte con spazio degli stati infinito. Vogliamo studiare sotto quali condizioni ammette una distribuzione stazionaria.

5.4.1 Catene di nascita e morte

Consideriamo una catena di Markov con spazio degli stati $\mathcal{S} = \{0, 1, \dots\}$ e matrice di transizione

$$P = \begin{bmatrix} r_0 & p_0 & & & & & & \dots \\ q_1 & r_1 & p_1 & & & & & \dots \\ & q_2 & r_2 & p_2 & & & & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ & & & \dots & q_j & r_j & p_j & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Abbiamo visto che se $q_i = q > 0$ per ogni $i \geq 1$ e $p_i = p > 0$ per ogni $i \geq 0$ allora la catena è transitoria se $q < p$, mentre la catena è ricorrente se $q \geq p$. Proviamo a cercare una distribuzione invariante. Se la troviamo allora la catena, essendo irriducibile e aperiodica, sarà ricorrente e dovremo ritrovare la condizione $q \geq p$. Cerchiamo quindi una soluzione $\pi = (\pi_0, \pi_1, \dots)$ del sistema avente infinite equazioni e infinite incognite $\pi = P\pi$. Scritto per esteso diventa

$$\begin{cases} \pi_0 &= \pi_0 r_0 + \pi_1 q_1 \\ \pi_1 &= \pi_0 p_0 + \pi_1 r_1 + \pi_2 q_2 \\ \vdots & \\ \pi_j &= \pi_{j-1} p_{j-1} + \pi_j r_j + \pi_{j+1} q_{j+1} \\ \vdots & \end{cases}$$

Osserviamo che $p_j + r_j + q_j = 1$ per ogni $j \geq 1$ e $p_0 + r_0 = 1$, dunque possiamo riscrivere il sistema come

$$\begin{cases} \pi_1 q_1 - \pi_0 p_0 & = 0 \\ \pi_2 q_2 - \pi_1 p_1 & = \pi_1 q_1 - \pi_0 p_0 \\ \vdots & \\ \pi_{j+1} q_{j+1} - \pi_j p_j & = \pi_j q_j - \pi_{j-1} p_{j-1} \\ \vdots & \end{cases}$$

Sostituendo la prima equazione nella seconda, la seconda nella terza e così via otteniamo

$$\begin{cases} \pi_1 q_1 - \pi_0 p_0 & = 0 \\ \pi_2 q_2 - \pi_1 p_1 & = 0 \\ \vdots & \\ \pi_{j+1} q_{j+1} - \pi_j p_j & = 0 \\ \vdots & \end{cases}$$

e quindi

$$\pi_{j+1} = \pi_j \frac{p_j}{q_{j+1}}, \quad j = 0, 1, \dots$$

Sostituendo a catena da π_1 in poi, otteniamo

$$\begin{cases} \pi_1 & = \pi_0 \frac{p_0}{q_1} \\ \pi_2 & = \pi_1 \frac{p_1}{q_2} = \pi_0 \frac{p_0 p_1}{q_1 q_2} \\ \vdots & \\ \pi_{j+1} & = \pi_j \frac{p_j}{q_{j+1}} = \pi_0 \frac{p_0 \cdots p_{j-1}}{q_1 \cdots q_j} \\ \vdots & \end{cases}$$

Poniamo ora

$$\delta_j = \frac{p_0 \cdots p_{j-1}}{q_1 \cdots q_j} \quad j = 1, 2, \dots, \quad \delta_0 = 1$$

possiamo riscrivere le equazioni del sistema come

$$\pi_j = \pi_0 \delta_j, \quad j = 1, 2, \dots$$

con la condizione

$$\sum_{j=0}^{+\infty} \pi_j = \pi_0 + \sum_{j=1}^{+\infty} \pi_0 \delta_j = \sum_{j=0}^{+\infty} \pi_0 \delta_j = 1.$$

Se si verifica che

$$\sum_{j=0}^{+\infty} \delta_j < +\infty,$$

cioè la serie converge, possiamo ricavare

$$\pi_0 = \frac{1}{\sum_{j=0}^{+\infty} \delta_j}$$

e quindi

$$\pi_j = \frac{1}{\sum_{j=0}^{+\infty} \delta_j} \delta_j, \quad j = 1, 2, \dots$$

Possiamo concludere che se la serie $\sum \delta_j$ converge, troviamo una distribuzione invariante, ed inoltre tale distribuzione è unica. Supponiamo che $q_i = q$ e $p_i = p$ per ogni $i \geq 1$. Allora abbiamo

$$\delta_i = \frac{p_0}{q} \left(\frac{p}{q}\right)^{i-1}, \quad i = 1, 2, \dots$$

e quindi la serie

$$\sum_{j=0}^{+\infty} \delta_j = 1 + \sum_{j=1}^{+\infty} \frac{p_0}{q} \left(\frac{p}{q}\right)^{i-1},$$

converge se e solo se $q > p$. Quindi solo se $q > p$ la serie converge e posso trovare la soluzione esplicitamente

$$\begin{aligned} \sum_{j=0}^{+\infty} \delta_j &= 1 + \frac{p_0}{q} \sum_{j=1}^{+\infty} \left(\frac{p}{q}\right)^{i-1} = 1 + \frac{p_0}{q} \frac{1}{1 - \frac{p}{q}} \\ &= \frac{q - p + p_0}{q - p}. \end{aligned}$$

Se $q = p$ abbiamo $\delta_i = 2p - 0$, quindi la serie $\sum \delta_i$ diverge e non esiste la distribuzione stazionaria. Possiamo quindi riassumere i risultati ottenuti: per la catena di nascita e morte con matrice di transizione P avente $p_i = p$, $q_1 = q$ per ogni $i \geq 1$ vale la seguente classificazione

- se $\frac{p}{q} > 1$, la catena è transitoria;
- se $\frac{p}{q} = 1$, la catena è ricorrente nulla;
- se $\frac{p}{q} < 1$, la catena è ricorrente positiva.

5.4.2 File d'attesa discrete

Le catene di Markov si prestano a modellare semplici sistemi (ad esempio centralini telefonici) in cui ad istanti prefissati un numero aleatorio di eventi (ad esempio le chiamate ad un centralino) si verificano e un numero aleatorio di eventi è evaso (ad esempio le chiamate evase dal centralino). Anche se una tale modellazione del fenomeno si presta ad altri contesti (diversi dal centralino telefonico) per ragioni storiche chiameremo chiamate gli eventi in arrivo e centralini gli organi atti ad evadere gli arrivi. Un esempio molto semplice consiste nel considerare le chiamate in arrivo e le chiamate evase ad ogni istante come variabili di Bernoulli con parametri rispettivamente α e β . Indichiamo quindi con A_n e con B_n due successioni di variabili casuali di Bernoulli indipendenti e identicamente distribuite rispettivamente di parametri α e β . Supponiamo inoltre che per ogni n , A_n e B_n siano indipendenti. Il numero di chiamate in attesa al centralino all'istante n , è dato da

$$\begin{cases} X_n = X_{n-1} + A_{n-1} - B_{n-1} & \text{se } X_{n-1} + A_{n-1} - B_{n-1} \geq 0, \\ X_n = 0 & \text{altrimenti.} \end{cases}$$

Il numero di chiamate in attesa al centralino all'istante n risulta pertanto una catena di Markov con spazio degli stati infinito. Calcoliamo la matrice di transizione. Per $i = 1, 2, \dots$ abbiamo

$$P_{i,j} = \begin{cases} \beta(1 - \alpha) & \text{se } j = i - 1, \\ 1 - \beta(1 - \alpha) - \alpha(1 - \beta) & \text{se } j = i, \\ \alpha(1 - \beta) & \text{se } j = i + 1. \end{cases}$$

Se invece $i = 0$ abbiamo

$$P_{0,j} = \begin{cases} 1 - \alpha(1 - \beta) & \text{se } j = 0, \\ \alpha(1 - \beta) & \text{se } j = 1. \end{cases}$$

Si tratta di una catena di tipo nascita e morte e dunque sappiamo che il comportamento della catena è caratterizzato completamente dai valori delle probabilità di transizione verso uno stato inferiore e verso uno stato superiore. In particolare in questo caso abbiamo che

- se $\alpha > \beta$ la catena è transitoria e dunque la coda al centralino risulta divergere;
- se $\alpha = \beta$ la catena è ricorrente nulla e anche in questo caso la coda al centralino risulta divergere;
- se $\alpha < \beta$ la catena è ricorrente positiva e la coda al centralino diverge con probabilità nulla.

Consideriamo ora un esempio in cui abbiamo un centralino con un buffer in grado di tenere in attesa solo un numero finito di telefonate, m . Supponiamo che le modalità di arrivo e di evasione delle telefonate siano le stesse descritte poco sopra. In questo caso il numero di telefonate in attesa è una catena di Markov con spazio degli stati finito, $\mathcal{S} = \{0, 1, \dots, m\}$ e matrice di transizione i cui elementi sono

$$\begin{cases} P_{i,i-1} = (1 - \alpha)\beta & \text{se } i = 1, 2, \dots, m, \\ P_{i,i} = 1 - (1 - \alpha)\beta - \alpha(1 - \beta) & \text{se } i = 1, 2, \dots, m - 1, \\ P_{i,i+1} = \alpha(1 - \beta) & \text{se } i = 0, 1, \dots, m - 1, \\ P_{m,m} = 1 - (1 - \alpha)\beta, \\ P_{0,0} = 1 - \alpha(1 - \beta). \end{cases}$$

Per valutare le prestazioni del centralino dobbiamo calcolare la probabilità con cui la catena si trova nello stato m , che corrisponde allo stato di saturazione. Se α e β sono entrambi maggiori di zero si tratta di una catena di Markov irriducibile e dunque ammette una distribuzione stazionaria che per quanto visto rappresenta il limite, per n che tende all'infinito, della proporzione di tempo che la catena passa nello stato m . Con conti analoghi a quelli fatti nel caso di catena di nascita e morte generale, si può mostrare che in questo caso la distribuzione invariante ha la forma

$$\pi_j = \frac{1}{\sum_{j=0}^m \delta_j} \delta_j, \quad j = 0, 1, 2, \dots, m$$

dove in questo caso

$$\delta_j = \left(\frac{\alpha(1 - \beta)}{(1 - \alpha)\beta} \right)^j, \quad j = 1, 2, \dots, m, \quad \delta_0 = 1.$$

Posto dunque

$$\delta = \frac{\alpha(1 - \beta)}{(1 - \alpha)\beta},$$

possiamo dedurre che

$$\pi_j = \frac{\delta^j}{\sum_{j=0}^m \delta^j} = \frac{\delta^j(1 - \delta)}{1 - \delta^{m+1}}, \quad j = 1, 2, \dots, m.$$

In particolare

$$\pi_m = \frac{\delta^m(1 - \delta)}{1 - \delta^{m+1}}$$

Se $\alpha = 0.75$, $\beta = 0.80$ e il buffer è pari a $m = 8$ abbiamo $\pi_m = 0.0271$, che sta ad indicare che il 2,7% del tempo il centralino è occupato.

Capitolo 6

Probabilità e tempi medi di assorbimento

6.1 Probabilità di assorbimento

Consideriamo una catena con un numero finito di stati che indichiamo con $\mathcal{S} = \{1, 2, \dots, r\}$. Sia C una classe chiusa di \mathcal{S} . Se la catena entra in C , cioè se per qualche k $X_k \in C$, allora la catena vi rimane per sempre, cioè $X_n \in C$ per ogni $n \geq k$. Per le applicazioni risulta importante andare a calcolare la probabilità che una catena partita da uno stato i qualunque, entri prima o poi nella classe C . Indichiamo tale probabilità con

$$\lambda_i = \mathbf{P}(X_n \in C, \text{ per qualche } n \mid X_0 = i).$$

Se $i \in C$ allora chiaramente $\lambda_i = 1$. Se invece $i \in D$, dove D è un'altra classe chiusa disgiunta dalla precedente, allora $\lambda_i = 0$. Le probabilità λ_i sono dunque di un qualche interesse quando i è uno stato transitorio che non si trova in C . Abbiamo quindi il seguente risultato.

Teorema 6.1.1. *Sia $\{X_n\}$ una catena di Markov avente spazio degli stati finito. Supponiamo che C sia una classe chiusa e indichiamo con T la classe degli stati transitori non in C . Allora la probabilità di assorbimento della catena nella classe C , partita dallo stato $i \in T$, è soluzione del sistema lineare*

$$\lambda_i = \sum_{h \in C} P_{ih} + \sum_{j \in T} P_{ij} \lambda_j. \quad (6.1)$$

Tale soluzione è anche l'unica.

Dimostrazione. Per la dimostrazione si veda ad esempio il testo di Baldi o il classico testo di Feller. Qui cerchiamo di giustificarla da un punto di vista intuitivo. L'idea

è che se la catena si trova nello stato i , essa può finire nella classe assorbente o in un passo solo andando direttamente dallo stato i allo stato $h \in C$. Oppure può andare in un passo in un altro stato transitorio j e poi da questo andare nella classe chiusa C . \square

Prima di passare al calcolo delle probabilità di assorbimento osserviamo che se lo spazio degli stati può essere suddiviso in un sottoinsieme T di tutti gli stati transitori e in un'unica classe C contenente tutti gli stati ricorrenti allora possiamo scrivere $T \cup C = \mathcal{S}$ e $T \cap C = \emptyset$. In questo caso il sistema

$$\lambda_i = \sum_{h \in C} P_{ih} + \sum_{j \in T} P_{ij} \lambda_j$$

ammette la soluzione $\lambda_i = 1$ per ogni $i \in T$. Infatti abbiamo verificata l'uguaglianza

$$1 = \sum_{h \in C} P_{ih} + \sum_{j \in T} P_{ij} = \sum_{k \in \mathcal{S}} P_{ik} = 1.$$

Questo dimostra che una catena di Markov con spazio degli stati finito lascia gli stati transitori per non tornarvi mai più con probabilità uno. Nelle applicazioni risulta interessante calcolare la probabilità di assorbimento in una classe C quando questa non esaurisce l'insieme degli stati ricorrenti della catena di Markov.

Esempio 6.1.2. Consideriamo un impianto per il confezionamento di uova prodotte da galline da allevamento. L'impianto può essere schematizzato con cinque fasi in sequenza: ingresso, spazzolatura, lavaggio, trattamento, confezionamento. Il 99% delle uova che entrano nell'impianto passano alla fase di spazzolatura, mentre le rimanenti vengono scartate. Il 97% delle uova che hanno superato la fase di spazzolatura passano alla successiva (lavaggio), mentre le rimanenti sono scartate. Il 96% delle uova che hanno subito il processo di lavaggio vengono passate alla fase di trattamento, mentre le restanti sono scartate. Infine, il 99% delle uova che sono state trattate vengono confezionate, mentre il restante 1% viene scartato. Possiamo modellare la dinamica di un uovo con una catena di Markov omogenea con sei stati, dati rispettivamente da:

stato 1: ingresso

stato 2: spazzolatura

stato 3: lavaggio

stato 4: trattamento

stato 5: confezionamento

stato 6: scarto

e matrice di transizione ad un passo data da

$$M = \begin{bmatrix} 0 & 0.99 & & & 0.01 \\ 0 & & 0.97 & & 0.03 \\ 0 & & & 0.96 & 0.04 \\ 0 & & & & 0.99 & 0.01 \\ 0 & & & & & 1 \\ 0 & & & & & & 1 \end{bmatrix}.$$

Per convenzione dove non compare nulla si intende che la probabilità di transizione è zero. Dalla forma della matrice M si deduce che gli stati 5 e 6 sono assorbenti, mentre gli altri stati sono transitori. La probabilità di assorbimento nella classe 5 rappresenta la probabilità che un uovo finisca confezionato e quindi passi alla distribuzione, la probabilità di assorbimento in 6 rappresenta la probabilità che un uovo venga scartato. Calcoliamo tali probabilità nel caso in cui lo stato iniziale dell'uovo sia quello in ingresso, cioè lo stato 1. Consideriamo la classe chiusa costituita dallo stato 5. Il sistema 6.1 in questo caso diviene, posto $T = \{1, 2, 3, 4\}$,

$$\lambda_i = \sum_{h=5} P_{ih} + \sum_{j \in T} P_{ij} \lambda_j, \quad \forall i \in T.$$

Si tratta quindi di 4 equazioni, che, considerata la forma della matrice M assumono la forma

$$\begin{aligned} \lambda_1 &= P_{12} \lambda_2 \\ \lambda_2 &= P_{23} \lambda_3 \\ \lambda_3 &= P_{34} \lambda_4 \\ \lambda_4 &= P_{45}. \end{aligned}$$

Se siamo interessati alla probabilità che, dato che lo stato iniziale è lo stato 1, la catena sia assorbita nello stato 5, basta risolvere il sistema e trovare il valore λ_1 . Risulta $\lambda_1 = P_{12} \cdot P_{23} \cdot P_{34} \cdot P_{45}$, e rappresenta la probabilità che un uovo entrato nell'impianto esca confezionato. Chiaramente $\tilde{\lambda}_1 = 1 - \lambda_1$, rappresenta la probabilità che un uovo entrato nell'impianto venga scartato. Il valore $\lambda_2 = P_{23} \cdot P_{34} \cdot P_{45}$ rappresenta la probabilità che un uovo che si trova nello 2 (cioè ha superato la prima selezione ed è stato spazzolato), sia assorbito nello stato 5, venga cioè confezionato.

Esempio 6.1.3. Riprendiamo il caso della catena che descrive una passeggiata aleatoria con barriere assorbenti in 0 e in N . Possiamo dire che la probabilità che la catena che parte da uno stato k transitorio, sia assorbita nella classe costituita dai due stati assorbenti 0 e N è 1. Abbiamo già calcolato la probabilità di assorbimento in 0 ricorrendo alla soluzione di un sistema di equazioni ricorrenti. Ora calcoliamo la probabilità di assorbimento nello stato N risolvendo il sistema (6.1). L'insieme degli stati transitori è $T = \{1, 2, \dots, N-1\}$. La classe chiusa C è costituita dal solo stato N . Il sistema risulta

$$\lambda_i = p\lambda_{i+1} + q\lambda_{i-1}, \quad i = 1, 2, \dots, N-1.$$

Osserviamo che per $i = 1$ l'equazione risulta $\lambda_1 = p\lambda_2$, mentre per $i = N-1$ $\lambda_{N-1} = p + q\lambda_N$. Possiamo inoltre porre $\lambda_0 = 0$ e $\lambda_N = 1$, in quanto se la catena si trova nello stato 0 è nulla la probabilità che sia assorbita in N , mentre se si trova nello stato N vi rimane. Poiché $p + q = 1$ possiamo scrivere

$$(p+q)\lambda_i = p\lambda_{i+1} + q\lambda_{i-1}, \quad i = 1, 2, \dots, N-1,$$

che diviene

$$p(\lambda_i - \lambda_{i+1}) = q(\lambda_{i-1} - \lambda_i) \quad i = 1, 2, \dots, N-1,$$

e quindi

$$\lambda_{i+1} - \lambda_i = \frac{q}{p}(\lambda_i - \lambda_{i-1}) \quad i = 1, 2, \dots, N-1.$$

Riscrivendo per esteso il sistema e sostituendo da un'equazione all'altra otteniamo, tenendo presente che $\lambda_0 = 0$,

$$\begin{aligned} \lambda_2 - \lambda_1 &= \frac{q}{p}\lambda_1 \\ \lambda_3 - \lambda_2 &= \frac{q}{p}(\lambda_2 - \lambda_1) = \left(\frac{q}{p}\right)^2 \lambda_1 \\ &\vdots \\ \lambda_i - \lambda_{i-1} &= \left(\frac{q}{p}\right)^{i-1} \lambda_1 \\ &\vdots \\ \lambda_N - \lambda_{N-1} &= \left(\frac{q}{p}\right)^{N-1} \lambda_1. \end{aligned}$$

Sommando le prime $i-1$ equazioni, per ogni $i = 2, 3, \dots, N$, otteniamo

$$(\lambda_2 - \lambda_1) + (\lambda_3 - \lambda_2) + \dots + (\lambda_i - \lambda_{i-1}) = \sum_{j=1}^{i-1} \left(\frac{q}{p}\right)^j \lambda_1.$$

Otteniamo quindi

$$\lambda_i - \lambda_1 = \sum_{j=1}^{i-1} \left(\frac{q}{p}\right)^j \lambda_1,$$

da cui

$$\lambda_i = \sum_{j=0}^{i-1} \left(\frac{q}{p}\right)^j \lambda_1.$$

Ricordando che, se $b \neq 1$, $(1 - b^n) = (1 - b)(1 + b + b^2 + \dots + b^{n-1})$, otteniamo per $q \neq p$ e per ogni $i = 2, \dots, N$

$$\lambda_i = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \frac{q}{p}} \lambda_1.$$

Ricordando però che $\lambda_N = 1$ ricaviamo

$$\lambda_1 = \frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N},$$

da cui

$$\lambda_i = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N}, \quad i = 1, \dots, N - 1.$$

Se invece $p = q = \frac{1}{2}$, otteniamo $\lambda_i = i\lambda_1$ e osservando che anche in questo caso $\lambda_N = 1$, otteniamo $\lambda_1 = \frac{1}{N}$ e $\lambda_i = \frac{i}{N}$, per ogni $i = 2, \dots, N - 1$. Si osservi che la probabilità che la catena sia assorbita in 0 la ricaviamo come complemento a 1 della probabilità che la catena sia assorbita in N , in virtù del fatto che la catena è sicuramente assorbita in uno dei due stati ricorrenti. In particolare ritroviamo il risultato già ottenuto in precedenza, cioè indicata con p_i la probabilità che la catena partita da i sia assorbita in 0 abbiamo

$$p_i = 1 - \lambda_i = 1 - \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N} = \frac{\left(\frac{q}{p}\right)^i - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}. \quad (6.2)$$

Osserviamo inoltre che se $q \geq p$ allora $\lim_{N \rightarrow +\infty} \lambda_i = 0$, mentre se $q < p$ allora $\lim_{N \rightarrow +\infty} \lambda_i = 1$. Vale a dire, ritornando all'esempio del giocatore che la rovina contro un giocatore con capitale infinito è certa per chi gioca alla pari o svantaggiato.

Consideriamo la passeggiata aleatoria data dall'equazione (3.2), ma supponiamo che la camminata parta dal punto $k > 0$ e vi siano due barriere assorbenti, in 0 e in N , $N > k$. Denotiamo con E l'evento

$$A = \{X_n = 0 \text{ per qualche } n\}$$

e con

$$p_k = \mathbf{P}(A|X_0 = k).$$

Sia inoltre F l'evento corrispondente a $Z_1 = +1$. Possiamo scrivere

$$\mathbf{P}(A|X_0 = k) = \mathbf{P}(A|F, X_0 = k)\mathbf{P}(F) + \mathbf{P}(A|F^c, X_0 = k)\mathbf{P}(F^c). \quad (6.3)$$

Infatti

$$\mathbf{P}(A|X_0 = k) = \mathbf{P}(A \cap F|X_0 = k) + \mathbf{P}(A \cap F^c|X_0 = k).$$

Possiamo quindi riscrivere il primo termine a secondo membro come

$$\begin{aligned} \mathbf{P}(A \cap F|X_0 = k) &= \frac{\mathbf{P}(A \cap F, X_0 = k)}{\mathbf{P}(X_0 = k)} = \mathbf{P}(A|F, X_0 = k) \frac{\mathbf{P}(F, X_0 = k)}{\mathbf{P}(X_0 = k)} \\ &= \mathbf{P}(A|F, X_0 = k)\mathbf{P}(F). \end{aligned}$$

Agendo in modo analogo per il secondo termine otteniamo la (6.3). Osserviamo che se si verifica l'evento F posso considerare la mia particella nel punto $k + 1$ e cominciare la camminata da lì. Abbiamo quindi che

$$\mathbf{P}(A|F, X_0 = k) = \mathbf{P}(A|X_0 = k + 1) = p_{k+1}.$$

In modo analogo

$$\mathbf{P}(A|F^c, X_0 = k) = \mathbf{P}(A|X_0 = k - 1) = p_{k-1}.$$

Possiamo quindi riscrivere la (6.3) come

$$p_k = p_{k+1}p + p_{k-1}q, \quad 0 < k < N,$$

con le condizioni $p_0 = 1$ e $p_N = 0$. Se $p \neq q$, la soluzione è data da

$$p_k = \frac{(q/p)^k - (q/p)^N}{1 - (q/p)^N}.$$

Se invece $p = q$, la soluzione è

$$p_k = 1 - \frac{k}{N}.$$

Esempio 6.1.4. Riprendiamo l'esempio 3.3.2. Il processo $\{Y_n\}$ è una catena di Markov. Se la catena viene assorbita in 0 il giocatore A è rovinato. In questo caso $k = a$, $N = a + b$ e $p = q = \frac{1}{2}$. La probabilità che A sia rovinato è

$$p_a = 1 - \frac{a}{a+b} = \frac{b}{a+b}.$$

Esempio 6.1.5. Un giocatore entra in una bisca con 1000 dollari e punta ogni volta un dollaro in un gioco dove egli vince con probabilità $p = 18/37$ e perde con probabilità $q = 19/37$. Ha deciso di andarsene non appena abbia vinto 1001 dollari o abbia perso tutto. Calcolare la probabilità che il giocatore se ne vada avendo vinto e la probabilità che se ne vada avendo perso. Calcolare il valore atteso di tale vincita. Possiamo rivedere il problema in questi termini. Il capitale del giocatore che entra nella bisca è una catena di Markov che parte nello stato 1000 ed evolve nel tempo come una passeggiata aleatoria con barriere assorbenti in 0 e in 1001. La probabilità che lo stato aumenti di una unità è pari a $18/37$, mentre la probabilità che diminuisca di una unità è pari a $19/37$. La probabilità di assorbimento in 0 è dunque

$$p_{1000} = \frac{(19/18)^{1000} - (19/18)^{1001}}{1 - (19/18)^{1001}} = 0.0526.$$

La probabilità di assorbimento in 1001 la possiamo calcolare come probabilità di assorbimento in 0 della catena che descrive il capitale del banco della bisca all'istante n . Tale catena parte dallo stato 1 ed ha barriere assorbenti in 0 (che corrisponde all'evento che il giocatore abbia raggiunto 1001 come capitale) e in 1001 (che corrisponde al fatto che il giocatore abbia perso tutto). La probabilità che la catena incrementi di un'unità il proprio stato è $19/37$. Quindi

$$p_1 = \frac{(18/19)^1 - (18/19)^{1001}}{1 - (18/19)^{1001}} = 0.9474.$$

Si osservi che la somma delle due probabilità è pari a 1. Per il calcolo del valore atteso della vincita del giocatore, osserviamo che tale giocatore alla fine della partita può trovarsi o con una vincita di 1001 o avere perso tutto. Se indichiamo con V tale variabile casuale, essa assume i valori 0 e 1001 con probabilità rispettivamente 0.0526 e 0.9474. Il valore atteso è dunque

$$\mathbf{E}(V) = 0 \cdot 0.0526 + 1001 \cdot 0.9474 = 948.3158.$$

6.1.1 Tempi medi di assorbimento

Se una catena di Markov ha spazio degli stati finiti la probabilità di assorbimento nella classe chiusa C composta da tutti gli stati assorbenti è certa, cioè vale 1.

Questo significa che la variabile casuale

$$\tau_C = \inf\{n > 0 : X_n \in C\}$$

che descrive il primo istante in cui la catena entra nella classe C assume solo valori finiti. Ha quindi senso calcolarne il valore atteso. Tale valore atteso, molto importante nelle applicazioni prende il nome di tempo medio di assorbimento. Indichiamo con η_i il tempo medio di assorbimento della catena nella classe C dato che la catena è partita nello stato i , vale a dire

$$\eta_i = \mathbf{E}(\tau_C | X_0 = i) = \sum_{n=1}^{+\infty} n \mathbf{P}(\tau_C | X_0 = i).$$

Per ricavare il valore dei tempi medi di assorbimento η_i , ci serviamo del seguente risultato che si ricava da un procedimento analogo a quello usato nel teorema 6.1.1.

Teorema 6.1.6. *Sia $\{X_n\}$ una catena di Markov avente spazio degli stati finito. Supponiamo che C sia una classe chiusa e indichiamo con T la classe degli stati transitori non in C . Allora il tempo medio di assorbimento della catena nella classe C , partita dallo stato $i \in T$, è soluzione del sistema lineare*

$$\eta_i = 1 + \sum_{j \in T} P_{ij} \eta_j. \quad (6.4)$$

Accanto al tempo medio di assorbimento nelle applicazioni a volte si è interessati al calcolo della probabilità che il tempo di assorbimento nella classe chiusa abbia avuto luogo prima di un istante n fissato. Si è cioè interessati al calcolo delle seguenti probabilità

$$\rho_i(n) = \mathbf{P}(\tau_C \leq n | X_0 = i).$$

Si dimostra che tali probabilità soddisfano la seguente equazione ricorrente

$$\rho_i(n+1) = \sum_{j \in C} P_{ij} + \sum_{k \in T} P_{ik} \rho_k(n).$$

Poiché possiamo calcolare

$$\rho_i(1) = \mathbf{P}(\tau_C = 1 | X_0 = i) = \sum_{j \in C} P_{ij},$$

per ricorrenza siamo in grado di calcolare il valore di $\rho_i(n)$ fino al valore n che ci interessa.

Esempio 6.1.7. Si riprenda l'esempio presentato nel capitolo introduttivo riguardante lo studio sull'affidabilità di un sistema. Si consideri un componente di un sistema che ad ogni controllo può trovarsi in uno dei tre stati: soddisfacente, insoddisfacente (ma funzionante), rotta. Indichiamo tali stati rispettivamente con 0, 1, 2. La matrice delle probabilità di transizione sia

$$\begin{bmatrix} 0.7 & 0.28 & 0.02 \\ 0 & 0.75 & 0.25 \\ 0 & 0 & 1 \end{bmatrix}$$

Si classifichino gli stati. Si calcoli il tempo medio di assorbimento nella classe C degli stati assorbenti (se esistono). Si calcoli infine la probabilità che il tempo di assorbimento nella classe C sia inferiore ai 15 giorni.

Esempio 6.1.8. Consideriamo la catena di Markov con spazio degli stati $\mathcal{S} = 1, 2, 3, 4, 5$ e matrice di transizione data da

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

Le classi $C_1 = \{1\}$ e $C_2 = \{5\}$ sono classi chiuse la classe $T = \{2, 3, 4\}$ è costituita da tutti gli stati transitori.

Calcoliamo la probabilità di assorbimento nella classe chiusa C_1 . Dobbiamo risolvere il sistema:

$$\lambda_i = \sum_{h \in C_1} P_{ih} + \sum_{j \in T} P_{ij} \lambda_j, \quad i \in T$$

dove $T = \{2, 3, 4\}$. Si tratta di 3 equazioni in tre incognite, la prima sommatoria è un termine solo. Scriviamolo per esteso: Per $i = 2$:

$$\lambda_2 = P_{21} + (P_{22}\lambda_2 + P_{23}\lambda_3 + P_{24}\lambda_4)$$

Per $i = 3$:

$$\lambda_3 = P_{31} + (P_{32}\lambda_2 + P_{33}\lambda_3 + P_{34}\lambda_4)$$

Per $i = 4$:

$$\lambda_4 = P_{41} + (P_{42}\lambda_2 + P_{43}\lambda_3 + P_{44}\lambda_4)$$

Che riscritto diviene:

$$\begin{array}{llll} (1 - P_{22})\lambda_2 & -P_{23}\lambda_3 & -P_{24}\lambda_4 & = P_{21} \\ -P_{32}\lambda_2 & +(1 - P_{33})\lambda_3 & -P_{34}\lambda_4 & = P_{31} \\ -P_{42}\lambda_2 & -P_{43}\lambda_3 & +(1 - P_{44})\lambda_4 & = P_{41} \end{array}$$

Si noti che alcuni coefficienti P_{ij} sono nulli, e che sostituendo i valori P_{ij} della matrice P otteniamo il sistema:

$$\begin{aligned} 1 \cdot \lambda_2 - \frac{1}{2} \lambda_3 - 0 \cdot \lambda_4 &= \frac{1}{2} \\ -\frac{1}{2} \lambda_2 + 1 \cdot \lambda_3 - \frac{1}{2} \lambda_4 &= 0 \\ 0 \cdot \lambda_2 - \frac{1}{2} \lambda_3 + 1 \cdot \lambda_4 &= 0 \end{aligned} \tag{6.5}$$

Si tratta di un sistema di tre equazioni in tre incognite $[\lambda_2, \lambda_3, \lambda_4]$ che risolto ci da

$$\lambda_2 = 0.75, \quad \lambda_3 = 0.50, \quad \lambda_4 = 0.25.$$

Il significato di queste probabilità è il seguente. Se la catena parte dallo stato 2, ha una probabilità dello 0.75 di essere assorbita nello stato 1, se parte dallo stato 3, viene assorbita nello stato 1 con probabilità 0.50, infine se parte dallo stato 4 ha una probabilità di 0.25 di essere assorbita nello stato 1.

Consideriamo ora la classe $C = \{1, 5\}$ che contiene tutti gli stati assorbenti. Sappiamo che, da qualunque stato transitorio della classe $T = \{2, 3, 4\}$ la catena parte, la probabilità che sia assorbita in C è 1. Ha senso allora andare a calcolare il tempo medio di assorbimento in C , condizionato dallo stato di partenza. Si tratta di risolvere il sistema

$$\eta_i = 1 + \sum_{j \in T} P_{ij} \eta_j, \quad i \in T,$$

che diviene, scritto per esteso,

$$\begin{aligned} 1 + P_{22} \eta_2 + P_{23} \eta_3 + P_{24} \eta_4 &= \eta_2 \\ 1 + P_{32} \eta_2 + P_{33} \eta_3 + P_{34} \eta_4 &= \eta_3 \\ 1 + P_{42} \eta_2 + P_{43} \eta_3 + P_{44} \eta_4 &= \eta_4 \end{aligned}$$

Il sistema riordinato diviene

$$\begin{aligned} (1 - P_{22}) \eta_2 - P_{23} \eta_3 - P_{24} \eta_4 &= 1 \\ -P_{32} \eta_2 + (1 - P_{33}) \eta_3 - P_{34} \eta_4 &= 1 \\ -P_{42} \eta_2 - P_{43} \eta_3 + (1 - P_{44}) \eta_4 &= 1 \end{aligned} \tag{6.6}$$

Si tratta di un sistema di tre equazioni in tre incognite $[\eta_1, \eta_2, \eta_3]$, che sostituendo i valori della matrice P diviene

$$\begin{aligned}\eta_2 - \frac{1}{2}\eta_3 &= 1 \\ -\frac{1}{2}\eta_2 + \eta_3 - \frac{1}{2}\eta_4 &= 1 \\ -\frac{1}{2}\eta_3 + \eta_4 &= 1\end{aligned}$$

La soluzione risulta

$$\eta_2 = 3, \quad \eta_3 = 4, \quad \eta_4 = 3.$$

Vale a dire che, se la catena parte dallo stato 2, verrà assorbito in media in tre istanti, se parte dallo stato 3, verrà assorbito in media in 4 istanti, se parte dallo stato 4, verrà assorbito in media in 3 istanti. Si noti che l'assorbimento in questo caso è sia nello stato 1 che nello stato 5.

Esempio 6.1.9. Riprendiamo la catena di nascita e morte presentata nel paragrafo 3.3.6, nel caso in cui $r_0 = 1$ e $r_d = 1$. Si tratta di un caso generale, di cui l'esempio precedente è un caso particolare. Gli stati 0 e d sono ricorrenti mentre gli altri stati sono transitori. Se vogliamo calcolare le probabilità di assorbimento della catena nello stato 0, scrivendo il sistema

$$\lambda_i = \sum_{h \in C} P_{ih} + \sum_{j \in T} P_{ij} \lambda_j,$$

ricaviamo come soluzione

$$\lambda_i = \frac{\gamma_i + \dots + \gamma_{d-1}}{1 + \gamma_1 + \dots + \gamma_{d-1}},$$

dove

$$\gamma_i = \frac{q_1 \cdots q_i}{p_1 \cdots p_i}.$$

Si osservi quindi che le probabilità di assorbimento non dipendono da r_i . Nel caso particolare in cui $p_i = p$ e $q_i = q$ per ogni i si ritrovano le formule (6.2). Poiché gli stati assorbenti sono solo 2 possiamo interpretare le probabilità di assorbimento in 0 o in d rispettivamente come la probabilità che la catena raggiunga prima lo stato 0 dello stato d e come la probabilità che la catena raggiunga prima lo stato d dello stato 0. In formula si dimostra quindi che

$$\mathbf{P}(\tau_0 < \tau_d | X_0 = i) = \frac{\sum_{k=i}^{d-1} \gamma_k}{\sum_{k=0}^{d-1} \gamma_k}$$

e analogamente

$$\mathbf{P}(\tau_d < \tau_0 | X_0 = i) = \frac{\sum_{k=0}^{i-1} \gamma_k}{\sum_{k=0}^{d-1} \gamma_k}.$$

6.2 Catene con un'infinità numerabile di stati

In questo paragrafo vogliamo dare qualche indicazione anche sugli stati di una catena con un'infinità numerabile di stati. Supponiamo che l'insieme degli stati sia $\mathcal{S} = \{0, 1, 2, \dots\}$. Supponiamo che la catena sia irriducibile, cioè tutti gli stati comunichino tra di loro. Allora la catena può essere o ricorrente o transitoria. Quindi dovremo individuare se uno stato è ricorrente e transitorio e quindi avremo deciso se la catena lo è. Avremo quindi bisogno di un criterio che ci faccia decidere se uno stato è ricorrente. Il criterio è una conseguenza della seguente proposizione che non dimostriamo.

Proposizione 6.2.1. *Per ogni stato i la probabilità di visitare prima o poi lo stato j soddisfa la seguente relazione*

$$f_{ij} = P_{ij} + \sum_{h \neq j} P_{ih} f_{hj}. \quad (6.7)$$

Anche senza dimostrazione il significato della relazione è chiaro: se la catena parte dallo stato i può visitare j in un passo oppure andare in un passo ad uno stato h diverso da j e quindi da questo andare prima o poi in j . Da questa proposizione deduciamo il seguente criterio assai utile. Uno stato è ricorrente se e solo se $f_{hi} = 1$ per ogni stato h tale che $P_{ih} > 0$. Infatti se $f_{hi} = 1$ per ogni stato h tale che $P_{ih} > 0$ allora dalla relazione (6.7) ricaviamo

$$f_{ii} = P_{ii} + \sum_{h \neq i} P_{ih} f_{hi} = 1,$$

dunque lo stato i è ricorrente. Viceversa se lo stato i è ricorrente allora $f_{ii} = 1$ e la relazione (6.7) è soddisfatta se $f_{hi} = 1$ per ogni stato h tale che $P_{ih} > 0$.

6.2.1 Catene di nascita e morte

Riprendiamo la catena di tipo nascita e morte presentata nella sezione 3.3.6. Supponiamo però che lo spazio degli stati sia $\mathcal{S} = \{0, 1, 2, \dots\}$ e la matrice di transizione

$$P = \begin{bmatrix} r_0 & p_0 & & & \dots \\ q_1 & r_1 & p_1 & & \dots \\ & q_2 & r_2 & p_2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ & & & \dots & q_j & r_j & p_j \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

sia tale che $p_j > 0$ per ogni $j \geq 0$ e $q_j > 0$ per ogni $j \geq 1$. Con queste condizioni la catena risulta irriducibile, e quindi o tutti gli stati sono ricorrenti o tutti gli stati sono transitori. Il comportamento della catena risulta determinato solo dal comportamento della serie i cui elementi sono i rapporti $\gamma_i = \frac{q_1 q_2 \cdots q_i}{p_1 p_2 \cdots p_i}$. Precisamente vale il seguente risultato.

Teorema 6.2.2. *Si consideri la catena di Markov di tipo nascita e morte con spazio degli stati \mathcal{S} e matrice di transizione P . La catena risulta ricorrente se e solo se*

$$\sum_{i=1}^{+\infty} \gamma_i = +\infty,$$

ogni stato della catena risulta invece transitorio se

$$\sum_{i=1}^{+\infty} \gamma_i < +\infty.$$

Dimostrazione. Basta studiare la ricorrenza o transitorietà dello stato 0. Dallo stato 0 si può andare solo nello stato 1. Per il criterio basta stabilire quanto vale f_{10} . Se $f_{10} = 1$ lo stato 0 risulta ricorrente, se $f_{10} < 1$ lo stato 0 risulta transitorio. Per definizione abbiamo

$$f_{10} = \mathbf{P}(\tau_0 < +\infty | X_0 = 1).$$

Abbiamo inoltre che per ogni $m > 1$ vale la seguente relazione

$$\mathbf{P}(\tau_m \geq m - 1 | X_0 = 1) = 1,$$

da cui risulta

$$\mathbf{P}(\tau_0 < m - 1 | X_0 = 1) = \mathbf{P}(\tau_0 < \tau_m | X_0 = 1).$$

Per ogni $m > 1$ possiamo ripensare alla catena di nascita e morte con stati da 0 a m e con 0 e m stati assorbenti. Allora denotando con λ_1 la probabilità di assorbimento in 0 della catena partita da 1 abbiamo che

$$\lambda_1 = \mathbf{P}(\tau_0 < \tau_m | X_0 = 1) = \frac{\sum_{k=1}^{m-1} \gamma_k}{\sum_{k=0}^{m-1} \gamma_k}.$$

Ricaviamo quindi

$$\begin{aligned} \mathbf{P}(\tau_0 < +\infty | X_0 = 1) &= \lim_{m \rightarrow +\infty} \mathbf{P}(\tau_0 < m - 1 | X_0 = 1) = \lim_{m \rightarrow +\infty} \mathbf{P}(\tau_0 < \tau_m | X_0 = 1) \\ &= \lim_{m \rightarrow +\infty} \frac{\sum_{k=1}^{m-1} \gamma_k}{\sum_{k=0}^{m-1} \gamma_k} = 1 - \lim_{m \rightarrow +\infty} \frac{1}{\sum_{k=0}^{m-1} \gamma_k} = 1 - \frac{1}{\sum_{k=0}^{+\infty} \gamma_k} \end{aligned}$$

Quindi ricaviamo che se la serie $\sum_{k=0}^{+\infty} \gamma_k$ diverge allora $\mathbf{P}(\tau_0 < +\infty | X_0 = 1) = 1$ e quindi la catena è ricorrente. Se invece la serie converge la catena è transitoria. In particolare se la serie è transitoria per ogni sottoinsieme finito di stati $R = \{0, 1, 2, \dots, k\}$ la catena esce da R per non tornarvi più. Quindi la catena tende ad andare sempre oltre k per ogni $k > 0$. Tale fatto è confermato anche dal senso comune nel caso seguente semplificato. Supponiamo che ad esempio $p_j = p$ per ogni $j \geq 0$ e $q_j = q$ per ogni $j \geq 1$. Allora $\frac{q_j}{p_j} = \frac{q}{p}$ e $\gamma_j = \left(\frac{q}{p}\right)^j$. Abbiamo quindi che la serie

$$\sum_{i=1}^{+\infty} \gamma_i = +\infty$$

se e solo se $\frac{q}{p} \geq 1$. Abbiamo quindi che se $q \geq p$ la catena è ricorrente. Se $q < p$ la catena è transitoria. Questo fatto è chiaro anche intuitivamente. Se q è più grande di p la catena tenderà a tornare nello stato 0, viceversa se q è più piccolo di p la catena tenderà ad allontanarsi sempre più dallo stato 0. \square

Capitolo 7

Generazione di variabili casuali

7.1 Introduzione

In questo capitolo cercheremo di fornire i rudimenti delle tecniche per generare variabili aleatorie da una distribuzione nota. Quasi tutti i linguaggi di basso livello dei calcolatori hanno la possibilità di generare campioni aleatori da una distribuzione uniforme su $[0, 1]$. Intorno a tali numeri, detti più precisamente pseudoaleatori, si può aprire un'ampia discussione, dai risvolti filosofici matematici e statistici, su come un computer, che in genere agisce in modo rigorosamente deterministico, possa generare effettivamente dei numeri da considerare aleatori e uniformemente distribuiti. Noi diamo per acquisita tale discussione e partiremo dal fatto che per noi la nostra macchina genera numeri aleatori uniformemente distribuiti su $[0, 1]$ e indipendenti. Da questo fatto acquisito cercheremo di capire come è possibile generare da questi numeri campioni indipendenti di variabili casuali aventi una distribuzione nota, e, ma questo lo vedremo più avanti, traiettorie di processi più complessi. Precisamente consideriamo il problema di generare i valori di una v.c. discreta X avente funzione di probabilità

$$\mathbf{P}(X = x_i) = p_i, \quad i = 1, 2, \dots, \quad \sum_i p_i = 1,$$

e il problema di generare i valori di una v.c. X continua avente funzione di ripartizione

$$F_X(x) = \mathbf{P}(X \leq x), \quad x \in \mathbb{R}.$$

7.2 Il metodo della funzione inversa

Il metodo della funzione inversa della funzione di ripartizione è uno dei metodi più semplici e più naturali per generare variabili aleatorie di qualunque (o quasi) natura.

Il metodo è giustificato dalla seguente proposizione.

Proposizione 7.2.1. *Sia X una v.c. avente funzione di ripartizione $F_X(x)$, $x \in \mathbb{R}$.*

- a) *Se X è continua allora $U = F_X(X)$ è distribuita uniformemente su $[0, 1]$.*
- b) *Se X non è continua allora la disuguaglianza $\mathbf{P}(F_X(X) \leq t) \leq t$ è ancora valida per ogni $t \in [0, 1]$.*
- c) *Se $F^{[-1]}(y) = \inf\{x : F(x) \geq y\}$ per ogni $y \in [0, 1]$, e se U è uniformemente distribuita su $[0, 1]$, allora $F^{[-1]}(U)$ ha come funzione di ripartizione la funzione F_X , cioè $\mathbf{P}(F^{[-1]}(U) \leq x) = F_X(x)$, per ogni $x \in \mathbb{R}$.*

Questa proposizione ci consente di generare il valore di una variabile casuale discreta o continua partendo dalla generazione di una variabile casuale uniforme su $[0, 1]$.

7.3 Generazione di variabili casuali discrete

Supponiamo di voler generare un valore della variabile casuale discreta X che assume i valori x_1, x_2, \dots, x_r avente distribuzione di probabilità

$$\mathbf{P}(X = x_i) = p_i, \quad i = 1, \dots, r, \quad \sum_{i=1}^r p_i = 1.$$

Denotiamo con F_i i valori della funzione di ripartizione nei punti di salto,

$$F_i = \sum_{j=1}^i p_j, \quad i = 1, \dots, r, \quad F_0 = 0$$

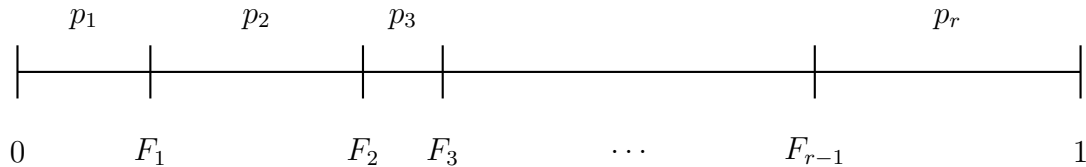
Data la variabile U uniforme su $[0, 1]$, definiamo una variabile X come

$$X = \begin{cases} x_1 & \text{se } U \leq p_1 \\ x_2 & \text{se } p_1 < U \leq F_2 \\ \vdots & \\ x_j & \text{se } F_{j-1} < U \leq F_j \\ \vdots & \\ x_r & \text{se } F_{r-1} < U \leq F_r \end{cases}$$

La variabile X così definita ha come distribuzione quella voluta, infatti

$$\mathbf{P}(X = x_j) = \mathbf{P}(F_{j-1} < U \leq F_j) = F_j - F_{j-1} = p_j, \quad j = 1, \dots, r.$$

Quindi per generare un valore della variabile discreta X avente un numero finito di valori basta generare un numero aleatorio da una variabile casuale uniforme e quindi fare il controllo con i valori della funzione di ripartizione della variabile casuale X . L'idea che sta alla base di questo metodo lo possiamo visualizzare. Immaginiamo di segnare i valori cumulati sull'intervallo $[0, 1]$:



Supponiamo di aver generato un numero u compreso tra 0 e 1. Il numero u casca all'interno di un intervallo delimitato per qualche i da F_{i-1} e F_i . Definiamo il valore simulato \tilde{x} come

$$\tilde{x} = \begin{cases} x_1 & \text{se } 0 < u \leq F_1 \\ x_2 & \text{se } F_1 < u \leq F_2 \\ \vdots & \\ x_j & \text{se } F_{j-1} < u \leq F_j \\ \vdots & \\ x_r & \text{se } F_{r-1} < u \leq 1 \end{cases}$$

Il valore \tilde{x} ottenuto lo pensiamo come una realizzazione della variabile casuale X .

7.4 Generazione di variabili casuali continue

Supponiamo di voler generare i valori assunti da una variabile casuale X avente densità $f(x)$ e funzione di ripartizione $F(x)$, $x \in \mathbb{R}$. Il punto *a*) della proposizione 7.2.1 ci fornisce la chiave per generare tali valori. Essendo la variabile casuale continua la funzione di ripartizione è invertibile. Sia F^{-1} tale funzione inversa e denotiamo con U la variabile casuale Uniforme su $[0, 1]$. Definiamo la variabile casuale $X = F^{-1}(U)$. Tale variabile casuale ha come funzione di ripartizione la funzione F . Infatti

$$\mathbf{P}(X \leq x) = \mathbf{P}(F^{-1}(U) \leq x) = \mathbf{P}(U \leq F(x)) = F(x).$$

Quindi per generare il valore della variabile aleatoria X generiamo un valore della variabile Uniforme U , diciamo il valore u . Calcoliamo $x = F^{-1}(u)$ e questo è il valore generato per la variabile X .

Esempio 7.4.1. Sia X la variabile casuale esponenziale. La funzione di ripartizione è

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

dove $\lambda > 0$ è il parametro della distribuzione che rappresenta l'inverso del valor medio della variabile. La funzione inversa di tale variabile è

$$F^{-1}(u) = -\frac{1}{\lambda} \log(1 - u), \quad 0 \leq u \leq 1.$$

Poiché la variabile casuale $1 - U$ è ancora uniformemente distribuita su $[0, 1]$, per generare il valore x della variabile esponenziale basta generare il valore u di una variabile casuale uniformemente distribuita e quindi porre

$$x = -\frac{1}{\lambda} \log(u), \quad 0 \leq u \leq 1.$$

Se una variabile casuale discreta X assume un numero infinito di valori possiamo generare i valori assunti da tale variabile con qualche artificio, tenendo presente le proprietà di tale variabile e gli eventuali legami che tale variabile ha con altre variabili note e di cui sappiamo generare i valori.

Esempio 7.4.2. In un campionamento bernulliano in cui la probabilità di un successo è p , il numero di prove N necessarie per ottenere il primo successo segue una distribuzione geometrica, $N = 1, 2, \dots$,

$$\mathbf{P}(N = k) = (1 - p)^{k-1} p, \quad k = 1, 2, \dots$$

Sia λ tale che $1 - p = e^{-\lambda}$. Allora la variabile N può essere rappresentata come $N = \lfloor X \rfloor + 1$ dove X segue la legge $Exp(\lambda)$ e abbiamo denotato con $\lfloor X \rfloor$ il più grande valore intero minore di X . In questo modo $\mathbf{P}(N = k + 1) = \mathbf{P}(\lfloor X \rfloor = k)$ e quindi

$$\mathbf{P}(N = k + 1) = \mathbf{P}(k \leq X < k + 1) = e^{-\lambda k} - e^{-\lambda(k+1)} = (1 - p)kp.$$

Quindi per generare il valore n della variabile casuale N , si genera un valore della variabile casuale U , diciamo u , e quindi si pone

$$n = \left\lfloor -\frac{1}{\lambda} \log u \right\rfloor + 1.$$

7.5 Simulazione di una catena di Markov

Una catena di Markov è una famiglia di variabili casuali X_n $n = 1, 2, \dots$ che possono assumere i valori di uno spazio degli stati \mathcal{S} e la cui dinamica è completamente determinata dalla matrice di transizione M e dal vettore delle probabilità π_0 . Quello che noi possiamo simulare è una traiettoria della catena di Markov, cioè i valori assunti dalle variabili casuali X_1, X_2, \dots, X_r , che costituiscono la catena di Markov. Chiaramente possiamo simulare solo traiettorie con un numero finito di passi. Vediamo che succede con un esempio semplice. Supponiamo che gli stati della catena siano i valori $s = [1, 2]$, la matrice di transizione sia

$$M = \begin{bmatrix} 0.8 & 0.2 \\ 0.9 & 0.1 \end{bmatrix}$$

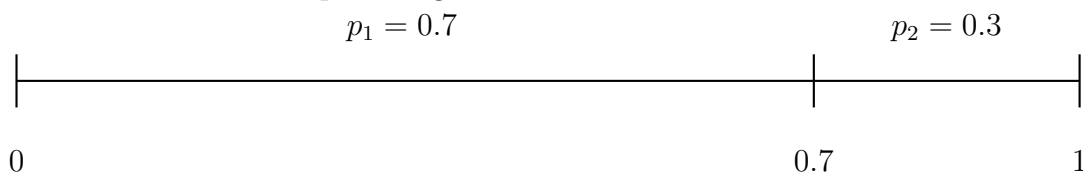
e la distribuzione iniziale sia $p = [1, 0]$, vale a dire, la catena parte dallo stato 1.

L'algoritmo per simulare una traiettoria lunga 10 passi può essere schematizzato nel seguente modo:

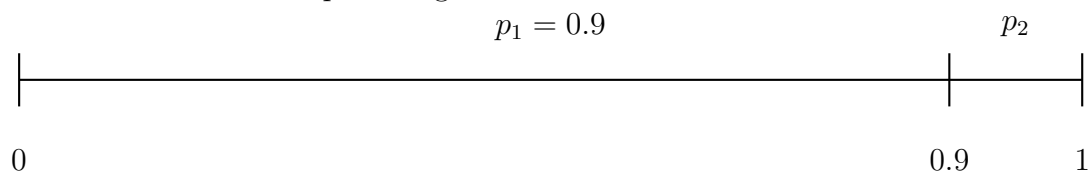
- $x(1)=0$ (stato iniziale = 0)
- ciclo **for** per k da 1 a 10
- se al passo precedente la catena era nello stato 0, cioè $x(k)=0$, simula una variabile casuale avente come distribuzione quella nella prima riga e genera il valore $x(k+1)$
- se al passo precedente la catena era nello stato 1, cioè $x(k)=1$, simula una variabile casuale avente come distribuzione quella nella seconda riga e genera il valore $x(k+1)$
- fine del ciclo **for**

Quando la catena ha due stati l'algoritmo di simulazione di ogni passo può essere fatto con un ciclo **if**

- se al passo k sono nello stato 0 devo generare una variabile casuale che ha come distribuzione la prima riga della matrice di transizione:



- se il numero aleatorio è minore o uguale a 0.7 il passo successivo sarà 0, altrimenti sarà 1
- se al passo k sono nello stato 0 devo generare una variabile casuale che ha come distribuzione la prima riga della matrice di transizione:



- se il numero aleatorio è minore o uguale a 0.9 il passo successivo sarà 1, altrimenti sarà 0

Se la catena ha k stati: $s = [s(1), \dots, s(k)]$ e la dinamica è determinata da una matrice $k \times k$

$$M = \begin{bmatrix} p(1,1) & p(1,2) & \dots & p(1,h) & \dots & p(1,k) \\ p(2,1) & p(2,2) & \dots & p(2,h) & \dots & p(2,k) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p(i,1) & p(i,2) & \dots & p(i,h) & \dots & p(i,k) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ p(k,1) & p(k,2) & \dots & p(k,h) & \dots & p(k,k) \end{bmatrix}$$

L'algoritmo di simulazione diventa molto contorto se utilizziamo un ciclo **if** per ogni stato della catena. In generale possiamo dire che se ci troviamo nello stato $s(i)$ dobbiamo simulare una variabile casuale che ha distribuzione data dalla riga i -esima della matrice M :

$$[p(i,1), p(i,2), \dots, p(i,h), \dots, p(i,k)].$$

Determinato lo stato successivo (ad esempio h) si deve simulare, al passo seguente, una variabile casuale avente distribuzione data dalla riga h -esima. I punti dell'algoritmo o programma di simulazione sono, nell'ordine:

1. occorre uno stato iniziale
2. una distribuzione iniziale
3. la matrice M della dinamica
4. il numero dei passi da simulare
5. un vettore dove mettere i valori simulati ad ogni passo (la traiettoria)

6. il numero di traiettorie da simulare

Supponiamo che lo stato iniziale sia dato e sia memorizzato nel vettore della traiettoria al posto 1: dobbiamo simulare una variabile casuale avente distribuzione data dalla riga corrispondente allo stato iniziale. Questo valore simulato corrisponde allo stato della catena all'istante 2, lo memorizziamo nel vettore della traiettoria al posto 2. Individuiamo quindi la riga corrispondente allo stato simulato al passo precedente e dobbiamo simulare una variabile casuale avente distribuzione data dalla riga corrispondente allo stato simulato al passo precedente. Si ripete il ciclo fino alla lunghezza della traiettoria richiesta. Se la catena ha k stati dati da: $s = [1, 2, \dots, k]$, l'algoritmo avrà questo aspetto:

```

M = [...] (definiamo la matrice di transizione)

n = 10 (definiamo la lunghezza della traiettoria)

F=cumsum(M')' (cumulate per riga della matrice di transizione)

x(1) = j (stato iniziale: per ora j fissato)

for i = 1 : n
    u = rand(1)
    h = 1
    while u > F(x(i), h)
        h = h + 1
    end
    x(i + 1) = h
end

```

Se lo stato iniziale non è noto, prima di $x(1) = j$ occorre simulare il valore dello stato iniziale secondo la distribuzione iniziale, e l'algoritmo avrà questo aspetto:

```

p = [...] (definiamo la distribuzione iniziale)

f=cumsum(p) (vettore cumulato della distribuzione iniziale)

j = 1

while u > f(j)
    j = j + 1
end

x(1) = j (valore iniziale simulato)

```


Capitolo 8

Processi di Poisson

8.1 Proprietà della variabile casuale esponenziale

Per definire il processo di Poisson risulta fondamentale conoscere la variabile casuale esponenziale e le sue proprietà. In questa sezione presenteremo tale variabile casuale e le variabili casuali strettamente collegate ad essa come la v.c. Gamma e la variabile casuale di Poisson.

Si dice che una v.c. T ha la distribuzione esponenziale con tasso λ se

$$P(T > t) = e^{-\lambda t}, \quad t \geq 0 \quad (8.1)$$

Si noti che dalla (8.1) si ricava che la funzione di ripartizione è

$$F_T(t) = P(T \leq t) = 1 - e^{-\lambda t}, \quad t \geq 0$$

e la densità

$$f_t(t) = \frac{dF_T(t)}{dt} = \lambda e^{-\lambda t}, \quad t \geq 0.$$

Il calcolo per il valore atteso e la varianza ci da

$$E(T) = \frac{1}{\lambda}, \quad \text{Var}(T) = \frac{1}{\lambda^2}$$

Se interpretiamo T come il tempo di attesa prima che un evento si verifichi, e se λ descrive il tasso con cui gli eventi di cui T descrive il tempo di attesa per il verificarsi del primo di questi eventi, si verificano nell'unità di tempo, allora è chiaro che il tempo di attesa risulta il reciproco di λ . Tanto più il tasso è elevato, tanto meno ci aspettiamo di aspettare l'arrivo del primo evento.

Indicheremo con $T \sim \text{Exp}(\lambda)$ una variabile casuale che segue la distribuzione esponenziale con tasso λ .

Esercizio 8.1.1. Sia $S \sim \text{Exp}(1)$. Dimostrare che $\frac{S}{\lambda} \sim \text{Exp}(\lambda)$.

Suggerimento: mostrare che $P(S/\lambda > t)$ coincide con la (8.1).

Una delle proprietà più importanti della distribuzione esponenziale è la **man-canza di memoria**. Si può descrivere in questo modo. Se il tempo di attesa per un autobus è esponenziale e voi state aspettando l'autobus da t minuti, allora la probabilità che lo dobbiate ancora aspettare per ulteriori s minuti è uguale a quella di doverlo aspettare per s minuti in tutto.

Teorema 8.1.2. Sia $T \sim \text{Exp}(\lambda)$ e $t, s \geq 0$. Allora vale la seguente relazione

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

Dimostrazione. Osserviamo che gli eventi $A = \{T > t + s\}$ e $B = \{T > t\}$ sono tali che $A \subset B$. Allora dalla definizione di probabilità condizionata abbiamo che

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

□

Un'altra proprietà importante della variabile casuale esponenziale è descritta dal seguente esempio. Se vi sono più variabili casuali esponenziali che descrivono il tempo di attesa per il realizzarsi di un certo evento e queste sono tutte indipendenti, allora il minimo di questi tempi è ancora una variabile casuale esponenziale.

Teorema 8.1.3. Siano T_1, \dots, T_n v.c. indipendenti esponenziali con parametro rispettivamente $\lambda_1, \dots, \lambda_n$. Allora la v.c. $V = \min(T_1, \dots, T_n)$ è una v.c. esponenziale con parametro $\sum_{i=1}^n \lambda_i$.

Dimostrazione. Osserviamo che se $\min(T_1, \dots, T_n) > t$ per $t > 0$, allora $T_i > t$ per ogni $i = 1, \dots, n$. Allora

$$\begin{aligned} \mathbf{P}(\min(T_1, \dots, T_n) > t) &= \mathbf{P}(T_1 > t, \dots, T_n > t) \\ (\text{per l'indipendenza}) &= \prod_{i=1}^n \mathbf{P}(T_i > t) = \prod_{i=1}^n e^{-\lambda_i t} = e^{-(\sum_{i=1}^n \lambda_i)t} \end{aligned}$$

Per la (8.1) si ha la tesi. □

Prima di enunciare le altre proprietà ricordiamo una definizione importante.

Definizione 8.1.4. Date due variabili casuali T e S con densità congiunta $f_{T,S}(s, t)$ in (s, t) e densità marginali $f_t(t)$ e $f_S(s)$, la densità condizionata della v.c T , dato il valore $S = s$ è definita come

$$f_{T|S=s}(t|s) = \frac{f_{T,S}(t, s)}{f_S(s)}.$$

Se le variabili casuali T e S sono indipendenti la densità congiunta è uguale al prodotto delle marginali e quindi $f_{T,S}(t, s) = f_t(t)f_S(s)$. Se invece si è interessati alla probabilità che un tempo di attesa sia minore di un altro abbiamo il seguente risultato.

Teorema 8.1.5. Siano S e T due v.c. indipendenti esponenziali con parametro rispettivamente λ e μ . Allora

$$\mathbf{P}(S < T) = \frac{\lambda}{\lambda + \mu}.$$

Dimostrazione. Osserviamo che integrando su tutti i possibili valori di $s > 0$ possiamo scrivere

$$\mathbf{P}(S < T) = \int_0^{+\infty} P(T > s|S = s)f_S(s)ds$$

Poiché T e S sono indipendenti la probabilità condizionata di T non dipende dal valore di S e quindi

$$\mathbf{P}(T > s|S = s) = \int_s^{+\infty} f_{T|S=s}(t|s)dt = \int_s^{+\infty} f_T(t)dt = \mathbf{P}(T > s).$$

Possiamo quindi scrivere

$$\begin{aligned} \mathbf{P}(S < T) &= \int_0^{+\infty} P(T > s)f_S(s)ds = \int_0^{+\infty} e^{-\mu s}\lambda e^{-\lambda s} ds \\ &= \frac{\lambda}{\lambda + \mu} \int_0^{+\infty} (\lambda + \mu)e^{-(\lambda + \mu)s} ds = \frac{\lambda}{\lambda + \mu}. \end{aligned}$$

L'integrale nell'ultimo passaggio è posto uguale a uno perchè l'integranda è la densità di una v.c. esponenziale con parametro $\lambda + \mu$. \square

Il prossimo risultato riguarda la distribuzione di somme di variabili casuali. Ricordiamo che una v.c X è detta distribuita come una v.c. Gamma con parametri α e λ e si indica con $\text{Gamma}(\alpha, \lambda)$, se la funzione di densità ha la forma

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}, \quad t > 0$$

dove la funzione Γ è definita, per ogni $\alpha \in \mathbb{R}$ da

$$\Gamma(\alpha) = \int_0^{+\infty} \lambda^\alpha t^{\alpha-1} e^{-\lambda t} dt = \int_0^{+\infty} z^{\alpha-1} e^{-z} dz$$

Teorema 8.1.6. *Siano S_1, \dots, S_n v.c. esponenziali indipendenti e tutte aventi parametro λ . Allora la v.c. $T = S_1 + \dots + S_n$ è una variabile casuale Gamma(n, λ). La densità di T può essere scritta come*

$$f_T(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \quad t > 0.$$

La dimostrazione si basa sulla caratterizzazione delle v.c. attraverso la loro funzione caratteristica ma esula dal corso. Osserviamo che per $n = 1$ la variabile casuale Gamma è una v.c. esponenziale. Inoltre si può dimostrare che per $\alpha = n$ intero la funzione Gamma(n) = $(n-1)!$.

8.2 Processi di Poisson omogenei

In questa sezione daremo due definizioni equivalenti di processo di Poisson con tasso costante λ . La prima ci permette di costruire il processo di Poisson direttamente dalle realizzazioni di una successione di v.c. esponenziali indipendenti. La seconda definizione identifica il processo di Poisson in base ad alcune proprietà che lo caratterizzano ed è importante perchè può essere generalizzata per definire processi di Poisson a tasso non costante nel tempo o processi di Poisson nello spazio.

Definizione 8.2.1. *Siano S_1, S_2, \dots , v.c. esponenziali indipendenti con tasso λ . Si ponga $T_0 = 0$ e per ogni $n > 0$, $T_n = S_1 + S_2 + \dots + S_n$. Si chiama processo di Poisson a tasso λ il processo $\{N(s) : s \geq 0\}$ definito come*

$$N(s) = \max\{n : T_n \leq s\}.$$

Se interpretiamo i valori di S_n come i tempi di arrivo tra l' $(n-1)$ -esimo e l' n -esimo evento di interesse (ad esempio le chiamate ad un centralino) allora T_n è il tempo di arrivo dell' n -esimo evento e $N(s)$ conta il numero di eventi fino all'istante s incluso.

Per chiarire l'idea supponiamo che $S_1 = t_1, S_2 = t_2, S_3 = t_3, S_4 = t_4$ e $S_5 = t_5$, allora $T_1 = t_1, T_2 = t_1 + t_2, T_3 = t_1 + t_2 + t_3, T_4 = t_1 + t_2 + t_3 + t_4$ e $T_5 = t_1 + t_2 + t_3 + t_4 + t_5$. Allora $N(s) = 0$ per $0 \leq s < T_1$, $N(s) = 1$ per $T_1 \leq s < T_2$, $N(s) = 2$ per $T_2 \leq s < T_3$, $N(s) = 3$ per $T_3 \leq s < T_4$ e $N(s) = 5$ per $T_4 \leq s < T_5$.

Il prossimo risultato chiarisce perché $\{N(s) : s \geq 0\}$ è chiamato processo di Poisson. Ricordiamo che X è una v.c. di Poisson con parametro μ se

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

Il valore atteso di una v.c. di Poisson è $\mathbf{E}(X) = \mu$ e la sua varianza è $\text{Var}(X) = \mu$.

Proposizione 8.2.2. *Sia $\{N(s) : s \geq 0\}$ un processo di Poisson a tasso λ . Allora il numero di eventi che si verificano prima dell'istante s è una v.c. di Poisson con parametro λs . Vale a dire*

$$\mathbf{P}(N(s) = k) = \frac{(\lambda s)^k e^{-\lambda s}}{k!}, \quad k = 0, 1, 2, \dots$$

Dimostrazione. Osserviamo che $N(s) = k$ se e solo se $T_k \leq s < T_{k+1}$. Inoltre se $T_k = t$ allora $0 < t \leq s$ e $T_{k+1} > s$ vale a dire $T_k + S_{k+1} > s$ da cui, poichè $T_k = t$, $S_{k+1} > s - t$. Teniamo presente inoltre che T_k è una v.c. Gamma(λ, k) e S_{k+1} è una v.c. Exp(λ). Allora integrando su tutti i possibili valori di t abbiamo

$$\begin{aligned} \mathbf{P}(N(s) = k) &= \mathbf{P}(T_k \leq s < T_{k+1}) = \int_0^s f_{T_k}(t) P(S_{k+1} > s - t | T_k = t) dt \\ (S_{k+1} \text{ ind } T_k) &= \int_0^s f_{T_k}(t) P(S_{k+1} > s - t) dt = \int_0^s \frac{\lambda^k t^{k-1} e^{-\lambda t}}{(k-1)!} e^{-\lambda(s-t)} dt \\ &= \frac{\lambda^k e^{-\lambda s}}{(k-1)!} \int_0^s t^{k-1} dt = \frac{\lambda^k e^{-\lambda s}}{(k-1)!} \frac{s^k}{k} = \frac{(\lambda s)^k e^{-\lambda s}}{k!} \end{aligned}$$

Abbiamo dimostrato la tesi. □

La proposizione ci dice che il numero medio di arrivi fino all'istante s incluso è λs . Ovvero il numero medio di arrivi nell'unità di tempo è λ .

Una proprietà importante del processo di Poisson è che possiede incrementi indipendenti. Vale il seguente risultato.

Proposizione 8.2.3. *Sia $\{N(s) : s \geq 0\}$ un processo di Poisson a tasso λ . Allora per ogni $s > 0$ fissato, il processo $N(t+s) - N(s)$, per $t \geq 0$ è un processo di Poisson a tasso λ ed è indipendente da $N(r)$ per ogni $0 \leq r < s$.*

Dimostrazione. La dimostrazione segue dal fatto che gli istanti degli arrivi prima dell'istante s e dopo l'istante s fino all'istante $t+s$ sono v.c. indipendenti e esponenziali di parametro λ . □

Si osservi che per questa proprietà la v.c. $N(t+s) - N(s)$, cioè il numero di arrivi nell'intervallo $(s, s+t]$ è una v.c. di Poisson con media λt . Da questa proposizione segue per induzione il seguente importante risultato.

Proposizione 8.2.4. *Sia $\{N(s) : s \geq 0\}$ un processo di Poisson a tasso λ . Allora se $0 \leq t_1 < t_2 < t_3 < \dots < t_{n-1} < t_n$ le v.c.*

$$N(t_2) - N(t_1), N(t_3) - N(t_2), \dots, N(t_n) - N(t_{n-1})$$

sono indipendenti.

La proposizione 8.2.4 è nota come la proprietà degli incrementi indipendenti. Cioè gli incrementi di un processo di Poisson sono v.c. di Poisson indipendenti.

Le proposizioni 8.2.2 e 8.2.4 caratterizzano il processo di Poisson, nel senso che possiamo dare come definizione equivalente la seguente.

Definizione 8.2.5. *Un processo $\{N(s) : s \geq 0\}$ che soddisfa le seguenti proprietà*

i) $N(0) = 0$;

ii) fissato $s > 0$, $N(t + s) - N(s)$ è una v.c. di Poisson con media λt per ogni $t \geq 0$;

iii) ha incrementi indipendenti;

è detto processo di Poisson con tasso λ .

Si osservi che se un processo di Poisson è definito tramite la prima definizione data, allora per le proposizioni 8.2.2 e 8.2.4 soddisfa le condizioni date dalla seconda definizione. Per mostrare che le due definizioni sono equivalenti occorre mostrare che se un processo è definito tramite la seconda definizione, allora i tempi di arrivo tra due eventi successivi sono v.c. esponenziali indipendenti di tasso λ . Siano S_1, S_2, \dots i tempi di arrivo degli eventi che il processo di Poisson conta. Allora l'istante del primo arrivo sarà maggiore di t se e solo se non si sono realizzati eventi prima dell'istante t , t incluso.

$$\mathbf{P}(S_1 > t) = \mathbf{P}(N(t) = 0) = e^{-\lambda t}.$$

L'ultima uguaglianza segue dalla proprietà *ii)* della seconda definizione. Calcoliamo allora

$$\mathbf{P}(S_2 > t | S_1 = s) = \mathbf{P}(N(t + s) - N(s) = 0) = e^{-\lambda t}$$

Questo perché il processo di Poisson ha incrementi indipendenti per la proprietà *iii)* della seconda definizione. Abbiamo dimostrato che la distribuzione di S_2 è indipendente da quella di S_1 ed è esponenziale. Il risultato segue per ogni S_n per induzione.

8.3 Probabilità condizionate

In questa sezione studiamo importanti proprietà di eventi condizionati da eventi legati al processo di Poisson. La prima proprietà riguarda la distribuzione dei tempi di arrivo degli eventi di Poisson dato che si sa che prima di un certo istante si sono verificati esattamente n eventi. Precisamente supponiamo di sapere che $N(t) = n$ allora

i tempi di arrivo degli n eventi, vale a dire T_1, T_2, \dots, T_n sono v.c. uniformemente distribuite e indipendenti.

Proposizione 8.3.1. *Sia $\{N(s) : s \geq 0\}$ un processo di Poisson a tasso costante λ . Siano T_1, T_2, \dots , i tempi di arrivo di eventi successivi. Dati $0 < t_1 < t_2 \dots < t_n < t$ allora*

$$f_{T_1, \dots, T_n | N(t)=n}(t_1, \dots, t_n | n) = \frac{n!}{t^n}.$$

per ogni $t > 0$ fissato e per ogni n .

La dimostrazione si basa sul fatto che n v.c. U_1, \dots, U_n indipendenti e uniformemente distribuite sull'intervallo $[0, t]$ hanno la densità data da

$$f_{U_1, \dots, U_n}(u_1, \dots, u_n) = \frac{1}{t^n}, \quad 0 \leq u_i \leq t, \quad i = 1, \dots, n$$

Se indichiamo con V_1, V_2, \dots, V_n le statistiche d'ordine di U_1, U_2, \dots, U_n , rispettivamente, vale a dire il valore più piccolo, poi il secondo e così via fino al più grande, abbiamo che

$$f_{V_1, \dots, V_n}(t_1, \dots, t_n) = \frac{n!}{t^n}, \quad 0 < t_1 < t_2 \dots < t_n < t.$$

Quest'ultimo risultato si deriva facilmente ma è un po' laborioso di calcolo per cui non lo riportiamo. Il lettore provi però a dimostrarlo per $n = 2$. La dimostrazione è conclusa osservando che

$$\begin{aligned} f_{T_1, \dots, T_n | N(t)=n}(t_1, \dots, t_n | n) &= \frac{f_{S_1, \dots, S_n, S_{n+1}}(t_1, t_2 - t_1, \dots, t_n - t_{n-1}, t - t_n)}{\mathbf{P}(N(t) = n)} \\ &= \frac{\lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2 - t_1)} \dots \lambda e^{-\lambda(t_n - t_{n-1})} \lambda e^{-\lambda(t - t_n)}}{\frac{e^{-\lambda t} (\lambda t)^n}{n!}} \\ &= \frac{n!}{t^n} \end{aligned}$$

La prima uguaglianza segue dal fatto che se voglio conoscere la distribuzione dei tempi di arrivo degli eventi dato che so che prima di t se ne sono verificati esattamente n , e indico con t_1, \dots, t_n gli istanti in cui questi si verificano, allora posso esprimere la stessa densità in funzione degli istanti di arrivo degli eventi successivi. Il fatto che $N(t) = n$ mi dice che l' $(n+1)$ -esimo arrivo sarà in un istante tra t_n e t .

La proposizione precedente implica che condizionando al fatto che si sono verificati n eventi prima dell'istante t , i tempi di arrivo si trovano nell'intervallo $[0, t]$ distribuiti uniformemente.

Dalla proposizione 8.3.1 segue la prossima importante proprietà che descrive la distribuzione del numero di eventi prima di s condizionato sempre all'evento $N(t) = n$.

Proposizione 8.3.2. *Sia $\{N(s) : s \geq 0\}$ un processo di Poisson a tasso costante λ . Se $s < t$ e $0 \leq m \leq n$, allora*

$$\mathbf{P}(N(s) = m | N(t) = n) = \binom{n}{m} \left(\frac{s}{t}\right)^m \left(1 - \frac{s}{t}\right)^{n-m}$$

Dimostrazione. Per la definizione di probabilità condizionata e sfruttando la proprietà degli incrementi indipendenti possiamo scrivere:

$$\begin{aligned} \mathbf{P}(N(s) = m | N(t) = n) &= \frac{\mathbf{P}(N(s) = m, N(t) = n)}{\mathbf{P}(N(t) = n)} \\ &= \frac{\mathbf{P}(N(s) = m, N(t) - N(s) = n - m)}{\mathbf{P}(N(t) = n)} \\ &= \frac{\frac{e^{-\lambda s} (\lambda s)^m}{m!} \frac{e^{-\lambda(t-s)} (\lambda(t-s))^{n-m}}{(n-m)!}}{\frac{e^{-\lambda t} (\lambda t)^n}{n!}} \\ &= \frac{n!}{(n-m)! m!} \left(\frac{s}{t}\right)^m \left(1 - \frac{s}{t}\right)^{n-m}, \end{aligned}$$

da cui la tesi. □

8.4 Processi di Poisson non omogenei

A volte non è molto realistico pensare che il tasso degli arrivi degli eventi che il processo di Poisson conta sia costante nel tempo. Si introduce allora il processo di Poisson con tasso non costante. Questa generalizzazione, se da un lato permette di modellare più realisticamente un numero maggiore di fenomeni, fa però perdere molte delle proprietà che aveva il processo di Poisson a tasso costante.

Definizione 8.4.1. *Sia $\lambda(s)$ per $s \geq 0$ una funzione positiva. Allora il processo $\{N(s) : s \geq 0\}$ è un processo di Poisson con tasso dato da $\lambda(s)$, per $s \geq 0$ se soddisfa le seguenti proprietà:*

- i) $N(0) = 0$;*
- ii) $N(t)$ ha incrementi indipendenti;*
- iii) $N(t+s) - N(s)$ è una v.c. di Poisson con valore atteso $\int_s^{t+s} \lambda(r) dr$.*

Se la funzione di tasso è costante $\lambda(t) = \lambda$ allora la *iii)* diventa

$$\mathbf{E}(N(t+s) - N(s)) = \int_s^{t+s} \lambda(r) dr = \lambda(t+s-s) = \lambda t$$

e ritroviamo un processo di Poisson omogeneo.

Vediamo di capire quale è la distribuzione dei tempi di arrivo per un processo di Poisson non omogeneo. Se denotiamo con T_1 la v.c. che descrive il tempo del primo arrivo abbiamo

$$\mathbf{P}(T_1 > t) = \mathbf{P}(N(t) = 0) = e^{-\int_0^t \lambda(r) dr}.$$

Derivando rispetto a t otteniamo la densità

$$f_{T_1}(t) = -\frac{d}{dt} e^{-\int_0^t \lambda(r) dr} = \lambda(t) e^{-\int_0^t \lambda(r) dr}.$$

Ponendo

$$\mu(t) = \int_0^t \lambda(r) dr$$

possiamo scrivere

$$f_{T_1}(t) = \lambda(t) e^{-\mu(t)}$$

Quindi i tempi di attesa non sono v.c. esponenziali. Inoltre non sono neppure indipendenti. Si può dimostrare infatti che

$$f_{T_1, \dots, T_n}(t_1, \dots, t_n) = \lambda(t_1) \cdots \lambda(t_n) e^{-\mu(t_n)}.$$

Esempio 8.4.2. *Si supponga che un negozio sia aperto dalle 10:00 alle 18:00. Si supponga che il tasso di arrivo dei clienti parta da 0 alle 10:00 e cresca linearmente fino a 4 alle 12:00 e poi sempre linearmente fino a 6 alle 14:00. Si supponga poi che decresca fino a 2 alle 16:00 e quindi fino a zero alle 18:00.*

1. Quale è la distribuzione del numero di arrivi in un giorno?
2. Quale è la probabilità che non ci siano arrivi prima delle 12:00?
3. Se il negozio chiude alle 17:30 invece che alle 18:00 quale è il numero atteso di clienti persi?
4. Quale è la probabilità che almeno un cliente trovi il negozio chiuso?

Il processo di Poisson non omogeneo può essere visto anche come una scelta opportunamente pesata di un processo di Poisson omogeneo (dall'inglese thinning, assottigliamento). Questo modo di vedere il processo di Poisson porta anche ad un algoritmo per generare traiettorie di processi di Poisson non omogenei.

Proposizione 8.4.3. *Sia $\{N(t) : t \geq 0\}$ un processo di Poisson a tasso costante λ . Sia $p(s)$ una funzione che per ogni s è positiva e minore di 1. Se un evento del processo di Poisson si verifica al tempo s allora introduciamo un processo $Q(t) : t \geq 0$ che conta l'evento precedente con probabilità $p(s)$. Allora il processo $\{Q(t) : t \geq 0\}$ è un processo di Poisson a tasso non costante e pari a $\lambda(s) = \lambda p(s)$.*

La dimostrazione può essere fatta in più modi ma tutte le dimostrazioni necessitano di introdurre nuovi concetti (quali il processo di Poisson nello spazio) o un'altra caratterizzazione del processo di Poisson tramite concetti dell'analisi infinitesimale. Per questi motivi preferiamo ometterla. Questa proposizione però è il punto di partenza per simulare un valore di un processo di Poisson con tasso non costante. Supponiamo che $\lambda(r) < \lambda$ per ogni $r \geq 0$. Allora si generano gli istanti in cui si verificano gli eventi di un processo di Poisson a tasso costante λ . Se s è uno di questi istanti lo si considera un istante in cui si è verificato anche un evento per il processo con tasso non costante se $u < \frac{\lambda(s)}{\lambda}$ dove u è un numero casuale tra 0 e 1.

Capitolo 9

Introduzione alla teoria delle file di attesa

9.1 Introduzione

In questo capitolo studiamo una classe generale di modelli per descrivere gli arrivi degli utenti ad un centro in grado di erogare un servizio. Si suppone che un cliente acceda al servizio immediatamente se questo è vuoto, altrimenti si mette in fondo alla coda ad aspettare il suo turno per essere servito. Non appena un utente è servito lascia immediatamente la coda e il servizio passa ad un successivo utente. Denotiamo con sistema l'insieme costituito dal servizio, o dall'ente che eroga il servizio e dall'eventuale utenza in coda. Per questi tipi di modelli, le quantità notevoli che andremo a studiare e che interessa determinare ai fini di una gestione ottimale e del controllo delle file d'attesa, saranno il numero medio di clienti in coda, o il numero medio di clienti nel sistema, il tempo medio che un utente passa in coda o il tempo medio che un cliente passa nel sistema (tempo di attesa più tempo di servizio). Il modello che utilizziamo per descrivere il numero di utenti presenti nel sistema, cioè gli utenti che stanno usufruendo del servizio e gli utenti in coda, è un processo stocastico con parametro continuo. Sia cioè $\{X_t : t \geq 0\}$ un processo stocastico dove la variabile casuale X_t , per ogni $t \geq 0$, descrive il numero di utenti nel sistema all'istante t . Il numero degli utenti all'istante t può essere un numero intero qualunque. Se $X_t = n$ diciamo che il sistema all'istante t si trova nello stato n . Le seguenti quantità sono importanti per studiare il comportamento del processo $\{X_t : t \geq 0\}$ in condizione di equilibrio. Denotiamo con

$$\pi_n = \lim_{t \rightarrow \infty} \mathbf{P}(X_t = n),$$

la proporzione, a lungo termine, del tempo in cui nel sistema sono presenti n utenti. Tale probabilità è detta anche probabilità all'equilibrio e ha lo stesso significato che

per le catene di Markov ha la distribuzione invariante.

9.2 Sistemi esponenziali

Supponiamo che gli utenti arrivino ad un servizio dotato di una sola stazione di servizio (che possiamo chiamare sportello) in accordo ad un processo di Poisson di tasso λ . Questo significa che i tempi di attesa tra due successivi arrivi sono variabili casuali esponenziali di media $1/\lambda$. Ogni utente al suo arrivo va direttamente allo sportello, se questo è libero, altrimenti si mette in fondo alla coda in attesa di raggiungere il proprio turno ad accedere al servizio. Quando l'utente presente allo sportello è servito, lascia immediatamente il sistema e un successivo cliente, se presente in coda, accede al servizio. Si suppone che i tempi impiegati dalla stazione di servizio per espletare il servizio su un utente siano delle variabili casuali esponenziali di media $1/\mu$. Questo tipo di sistema di fila d'attesa è denotato con $M/M/1$ dove la prima M si riferisce al fatto che gli arrivi sono senza memoria (dall'inglese *memoryless*), la seconda al fatto che anche i tempi di servizio sono senza memoria, il numero 1 si riferisce al fatto che c'è solo una stazione che eroga il servizio.

9.2.1 Le equazioni del bilancio dettagliato

Per il calcolo delle probabilità limite, che vanno interpretate come la proporzione del tempo che il sistema passa in un certo stato n , ci si basa su una considerazione, che seppure la enunciamo ad un livello euristico, può essere dimostrata rigorosamente a partire dalle equazioni differenziali che le probabilità di transizione del sistema devono soddisfare. Ad ogni modo al principio che porta alla definizione delle equazioni del bilancio dettagliato (le cui soluzioni saranno le probabilità limite) possiamo crederci anche partendo da considerazioni più banali. L'idea è di accettare il fatto che il tasso col quale il sistema entra in un certo stato è uguale al tasso con cui ne esce. O meglio il numero di volte che il sistema entra in uno stato in un intervallo di tempo fissato deve essere uguale al numero di volte con cui ne esce. La morale è in sostanza che per ritornare in uno stato il sistema deve prima essere uscito da quello stato! Inoltre, se in un intervallo di tempo piccolo la probabilità che si verifichino due arrivi degli eventi di interesse è trascurabile, significa che il sistema può solo andare da uno stato agli stati adiacenti. Le equazioni che si ricavano uguagliando il flusso in entrata con il flusso in uscita da ogni stato sono dette equazioni del bilancio dettagliato e le soluzioni sono le probabilità limite del nostro sistema. Cominciamo a dedurre l'equazione per lo stato $k = 0$. Il sistema può lasciare lo stato 0 solo se

arriva un utente, e quindi va nello stato 1. La proporzione di tempo che il sistema passa nello stato 0 è π_0 mentre la proporzione di utenti che arrivano in un'unità di tempo è λ . Ne consegue che il tasso di uscita dallo stato 0 è dato da $\lambda\pi_0$. Per calcolare il tasso di entrata osserviamo che nello stato 0 ci possiamo arrivare solo dallo stato 1 in seguito ad un utente presente nel sistema che lascia il sistema. La proporzione di tempo che il sistema passa nello stato 1 è π_1 , il numero di utenti che lascia il sistema nell'unità di tempo è μ . Se ne deduce che il tasso di entrata nello stato 0 è dato da $\mu\pi_1$. Uguagliando il tasso di entrata nello stato 0 con il tasso di uscita dallo stato 0 abbiamo la prima equazione del bilancio dettagliato:

$$\lambda\pi_0 = \mu\pi_1.$$

Ripetiamo lo stesso ragionamento per lo stato 1. Dobbiamo tenere presente che il sistema esce dallo stato 1 o in seguito ad un arrivo (e quindi passa nello stato 2) o in seguito ad una partenza (e quindi passa nello stato 0), mentre si entra nello stato 1 dallo stato 0, in seguito ad un arrivo, o dallo stato 2 in seguito ad una partenza. Si deduce quindi che il tasso di uscita è pari a $(\lambda + \mu)\pi_1$, mentre il tasso di entrata è $\lambda\pi_0 + \mu\pi_2$. Anche qui uguagliando il tasso di entrata con il tasso di uscita dallo stato 1 abbiamo la seconda equazione del bilancio dettagliato:

$$(\lambda + \mu)\pi_1 = \lambda\pi_0 + \mu\pi_2.$$

In modo analogo si ragiona per gli altri stati e deduciamo il sistema delle equazioni del bilancio dettagliato:

$$\begin{aligned} \lambda\pi_0 &= \mu\pi_1 \\ (\lambda + \mu)\pi_n &= \lambda\pi_{n-1} + \mu\pi_{n+1} \quad n = 1, 2, \dots, \end{aligned}$$

al quale va aggiunta la condizione $\sum_{i=0}^{+\infty} \pi_i = 1$. Ricavando dalla prima equazione π_1 , dalla seconda π_2 e così via, ed andando a sostituire ricorsivamente i termini ricavati da un'equazione nella successiva, il sistema può essere riscritto come

$$\pi_n = \left(\frac{\lambda}{\mu}\right)^n \pi_0, \quad n = 1, 2, \dots, \quad \sum_{i=0}^{+\infty} \pi_i = 1.$$

Dalla condizione $\sum_{i=0}^{+\infty} \pi_i = 1$ ricaviamo

$$\sum_{i=0}^{+\infty} \left(\frac{\lambda}{\mu}\right)^i \pi_0 = 1,$$

da cui, se $\frac{\lambda}{\mu} < 1$

$$\pi_0 = 1 - \frac{\lambda}{\mu}$$

e quindi

$$\pi_n = \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right), \quad n \geq 1.$$

La condizione $\lambda < \mu$ è equivalente al fatto che il tempo medio di servizio è minore del tempo medio tra due arrivi successivi. Questa condizione risulta del tutto naturale se vogliamo che esista una distribuzione limite. Infatti se $\lambda > \mu$, il numero atteso di utenti prima dell'istante t è λ in quanto gli arrivi si ipotizzano seguire la legge di un processo di Poisson. D'altra parte se vi sono utenti presenti nel sistema, poiché il tempo di servizio di un utente segue una variabile casuale esponenziale di media $1/\mu$, ne consegue che il numero atteso di utenti serviti prima dell'istante t è μt . Quindi il numero medio di utenti presenti nel sistema prima dell'istante t è pari a

$$\lambda t - \mu t = (\lambda - \mu)t,$$

che se $\lambda - \mu > 0$ tende all'infinito se t tende all'infinito. Le probabilità appena calcolate si interpretano come proporzione del tempo che ci attendiamo che il sistema stia nello stato n . Ad esempio π_0 è la proporzione di tempo che ci attendiamo di trovare lo sportello vuoto, cioè privo di clienti in coda.

9.2.2 Numero medio di utenti e tempi medi di attesa

Andiamo ora a introdurre la variabile N che indica il numero di utenti presenti nel sistema. Il valore atteso di questa variabile rappresenta il numero medio di utenti che ci aspettiamo di trovare nel sistema. Per il calcolo abbiamo

$$\mathbf{E}(N) = \sum_{n=0}^{+\infty} n\pi_n = \sum_{n=0}^{+\infty} n \left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right) = \frac{\lambda}{\mu - \lambda},$$

poiché ricordiamo che $\sum_{n=0}^{+\infty} nx^n = \frac{x}{(1-x)^2}$, se $x < 1$. Possiamo concludere che più è piccola la differenza tra λ e μ più è grande il numero atteso di utenti che ci aspettiamo di trovare presenti nel sistema. Indichiamo con W la variabile che indica il tempo di permanenza di un utente nel sistema. Per calcolare il tempo medio di permanenza nel sistema dobbiamo ipotizzare un legame tra le variabili N e W . Tale legame è noto come principio di Little ed afferma che il numero di utenti nel sistema è pari al tasso di arrivo moltiplicato per il tempo di attesa nel sistema. In formula diventa

$$N = \lambda W,$$

da cui ricaviamo che $\mathbf{E}(N) = \lambda \mathbf{E}(W)$ e quindi il tempo medio di permanenza nel sistema risulta

$$\mathbf{E}(W) = \frac{\lambda}{\mu - \lambda} \frac{1}{\lambda} = \frac{1}{\mu - \lambda}.$$

Indichiamo con N_Q e W_Q le variabili che indicano il numero di utenti in coda e il tempo di permanenza in coda. Ricaviamo che il tempo medio di permanenza in coda è dato dal tempo medio di permanenza nel sistema meno il tempo di servizio. Abbiamo quindi

$$\mathbf{E}(W_Q) = \mathbf{E}(W) - \mathbf{E}(T) = \frac{1}{\mu - \lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)}.$$

Ipotizzando che anche tra le variabili N_Q e W_Q vi sia il legame $N_Q = \lambda W_Q$ abbiamo

$$\mathbf{E}(N_Q) = \frac{\lambda^2}{\mu(\mu - \lambda)}.$$

9.3 Esempi

9.3.1 Il problema di Erlang

Si consideri un centro telefonico dotato di un centralino con s apparecchi in grado di rispondere alle chiamate. Se un utente arriva nel sistema e trova un apparecchio libero accede al servizio, se invece l'utente che arriva trova tutte le macchine occupate, non entra nel sistema. Si tratta di un sistema di coda a perdere, nel senso che non si forma nessuna coda. Sarà allora importante calcolare la probabilità di congestione di un tale sistema. Supponiamo che gli arrivi seguano un processo di Poisson e i tempi di servizio di ogni macchina siano delle variabili casuali Esponenziali di parametro μ indipendenti. Tale coda è un tipo particolare di coda del tipo $M/M/s$. Per determinare la distribuzione limite dobbiamo dedurre le equazioni del bilancio dettagliato per questo sistema. Supponiamo di descrivere lo stato del sistema con la variabile N_t che indica quante macchine sono occupate all'istante t . Se il sistema si trova nello stato i , passerà nello stato $i - 1$ quando la prima delle macchine occupate si libera. Il tempo d'attesa affinché la prima macchina si liberi è una variabile casuale esponenziale di parametro $i\mu$. (Si ricordi il risultato sulla variabile casuale $\min(T_1, \dots, T_n)$ che abbiamo dimostrato essere uguale a una v.c. esponenziale con parametro uguale alla somma dei parametri delle variabili casuali T_1, \dots, T_n). In questo caso $\min(T_1, \dots, T_i) \sim \text{Exp}(i\mu)$. Le equazioni del bilancio dettagliato per

questo sistema sono

$$\begin{aligned}
 \lambda\pi_0 &= \mu\pi_1 \\
 (\lambda + \mu)\pi_1 &= \lambda\pi_0 + 2\mu\pi_2 \\
 (\lambda + 2\mu)\pi_2 &= \lambda\pi_1 + 3\mu\pi_3 \\
 &\vdots \\
 (\lambda + n\mu)\pi_n &= \lambda\pi_{n-1} + (n-1)\mu\pi_{n+1} \\
 &\vdots \\
 s\mu\pi_s &= \lambda\pi_{s-1}.
 \end{aligned}$$

Tenendo conto della condizione $\sum_{i=0}^s \pi_i = 1$ ricaviamo la soluzione

$$\pi_i = \frac{\left(\frac{\lambda}{\mu}\right)^i}{i!} \bigg/ \sum_{i=0}^s \frac{\left(\frac{\lambda}{\mu}\right)^i}{i!} \quad i = 0, 1, \dots, s.$$

La probabilità di congestione è data da π_s e rappresenta la proporzione di volte che il centralino si trova saturo.

9.3.2 La coda $M/G/\infty$

Nel modellare il traffico delle linee telefoniche si può pensare in prima approssimazione che ogni utente che chiami trovi la linea libera, ci siano cioè infinite linee telefoniche. Questo non può ovviamente essere vero nella realtà ma il supporlo possibile ci può aiutare nel progettare un sistema nel quale siamo in grado di offrire il servizio nel 99.99% dei casi. Si supponga che gli arrivi delle chiamate seguano un processo di Poisson a tasso λ , mentre le durate delle telefonate seguano una generica v.c. avente funzione di ripartizione G tale che $G(0) = 0$ e $\mathbf{E}(G) = \mu$. Se il sistema parte dallo stato 0 all'istante $t = 0$, la probabilità che una chiamata iniziata all'istante s sia finita prima dell'istante t è $G(t - s)$. Quindi il numero di chiamate che sono ancora in corso all'istante t è una v.c. di Poisson con media data da

$$\int_s^t \lambda(1 - G(t - u))du = \lambda \int_0^{t-s} (1 - G(r))dr$$

Cioè se indichiamo con $N(t)$ il numero di chiamate arrivate all'istante t , si tratta di un processo di Poisson non omogeneo con funzione di tasso data da $\lambda(r) = \lambda(1 - G(t - r))$. Ponendo $s = 0$ e facendo tendere t all'infinito si vede che nel lungo periodo il numero di chiamate in arrivo nel sistema è dato da una v.c. di Poisson con media

$$\lambda \int_0^\infty (1 - G(r))dr = \lambda\mu$$

in quanto per le v.c. continue X a valori positivi si può calcolare il valore atteso come $\mathbf{E}(X) = \int_0^{\infty} (1 - F(x)) dx$ dove abbiamo denotato con F la funzione di ripartizione di X . In conclusione il numero medio di utenti in linea è pari al tasso con il quale arrivano le chiamate moltiplicato per il tempo medio di durata.

Capitolo 10

Esercizi

10.1 Catene di Markov

1. Nel Mondo di Oz accadono strane cose e tra le altre il tempo è veramente un disastro! In particolare, non si verificano mai due giorni di tempo sereno S di seguito se oggi è bel tempo, domani può nevicare N o piovere P con uguale probabilità se nevicava o piove e il tempo cambia, solo la metà delle volte il tempo volge al bello se nevicava o piove, nel 50% il tempo rimane invariato, altrimenti nel restante 50% dei casi il tempo cambia in uno dei rimanenti stati.
 - (a) Costruire la matrice di transizione.
 - (b) Dato che oggi nevicava con che probabilità tra 3 giorni ci sarà il sole?
 - (c) Il mago di Oz tira a caso un dado con tre facce, una per ogni tipo di tempo, per determinare il tempo di oggi. Calcolare la probabilità che tra 4 giorni sia sereno.
2. Un taxi si può muovere dall'aeroporto A a due hotel B e C secondo le seguenti regole. Se si trova all'aeroporto, sarà in uno dei due hotel con uguale probabilità. Se si trova in uno dei due hotel, andrà all'aeroporto con probabilità $3/4$ e andrà nell'altro hotel con probabilità $1/4$.
 - (a) Scrivere la matrice di transizione.
 - (b) Calcolare il periodo di ogni stato.
 - (c) Classificare gli stati della catena.
 - (d) Si supponga che il taxi sia all'aeroporto all'istante 0. Trovare la probabilità con cui sarà in uno dei possibili luoghi (aeroporto A, hotel B, hotel C) all'istante 2.

- (e) Si supponga che il taxi sia all'aeroporto all'istante 0. Calcolare la probabilità che il taxi sia all'hotel C dopo tre istanti.
- (f) Dato che il taxi si trova all'aeroporto all'istante 0, trovare la probabilità con cui sarà in uno dei possibili luoghi (aeroporto A, hotel B, hotel C) all'istante 2.
- (g) Dato che il taxi si trova all'aeroporto all'istante 0, calcolare la probabilità che il taxi sia all'hotel C dopo tre istanti.
- (h) Trovare la percentuale di volte in cui il taxi sarà in uno dei possibili luoghi in un periodo di tempo sufficientemente grande.
3. Alla fine del mese una grossa azienda di vendita al dettaglio classifica ogni suo cliente in base al ritardo nei pagamenti in questo modo: stato 1 = non in ritardo, stato 2 = ritardo di 30-60 giorni, stato 3 = ritardo di 60-90 giorni, stato 4 ritardo maggiore di 90 giorni. L'esperienza dell'azienda induce a credere che i clienti si muovano da uno stato all'altro secondo la matrice di transizione seguente:

$$\begin{bmatrix} 0.90 & 0.10 & 0.00 & 0.00 \\ 0.80 & 0.00 & 0.20 & 0.00 \\ 0.50 & 0.00 & 0.00 & 0.50 \\ 0.10 & 0.00 & 0.00 & 0.90 \end{bmatrix}$$

- (a) La traiettoria $1 \rightarrow 2 \rightarrow 3 \rightarrow 2$ è compatibile con la matrice di transizione? Spiegare.
- (b) Quali stati comunicano tra di loro?
- (c) Calcolare il periodo di tutti gli stati.
- (d) Classificare gli stati della catena.
- (e) A lungo andare quanti clienti in ogni categoria ci saranno in percentuale?
4. Lo stato di un prodotto durante le sue fasi di lavorazione e di utilizzo può essere schematizzato come uno dei seguenti
- (1) prima fase di tornitura
 - (2) seconda fase di tornitura
 - (3) rispetta le specifiche e funziona
 - (4) rispetta le specifiche e non funziona
 - (5) non rispetta le specifiche

Un pezzo che entra nel ciclo di fabbricazione e si trova nello stato 1, può nell'istante di tempo successivo, non rispettare le specifiche con probabilità 0.1, essere sottoposto ad una seconda fase di tornitura con probabilità 0.3, può rispettare le specifiche e funzionare con probabilità 0.45. Un pezzo che si trova nello stato 2, può, nell'istante di tempo successivo, non rispettare le specifiche con probabilità 0.1, può rispettare le specifiche e funzionare con probabilità 0.30. Un pezzo che rispetta le specifiche e funziona, nell'istante successivo può rimanere nello stesso stato con probabilità 0.60, oppure andare nello stato 4 con probabilità 0.40. Un pezzo che rispetta le specifiche e non funziona, nell'istante successivo può rimanere nello stesso stato con probabilità 0.30, oppure andare nello stato 3 con probabilità 0.70. Lo stato 5 è uno stato assorbente.

- (a) Sapendo che se un pezzo si trova nello stato 1, all'istante successivo non può essere nello stato 1 e che se un pezzo si trova nello stato 2 all'istante successivo non può trovarsi nello stesso stato e neppure nello stato 1, costruire la matrice di transizione P . (Verificare che la somma per riga sia 1)
 - (b) La traiettoria $1 \rightarrow 2 \rightarrow 4 \rightarrow 3 \rightarrow 3 \rightarrow 4 \rightarrow 4$ è compatibile con tale matrice di transizione? (motivare la risposta passo per passo)
 - (c) Classificare gli stati della catena.
 - (d) Determinare tutte le classi chiuse e le classi chiuse ed irriducibili.
 - (e) È possibile per un pezzo che parte dallo stato 1, raggiungere lo stato 3 in 3 passi? Scrivere una possibile traiettoria.
 - (f) Calcolare la probabilità che un pezzo sia dopo tre istanti nello stato 3 dato che all'istante iniziale si trovava nello stato 1.
 - (g) Si consideri solo la sottomatrice 2×2 della matrice P costituita dagli stati 3 e 4. Tale matrice ammette una distribuzione invariante? Se sì calcolarla, altrimenti dire perché non esiste.
5. Una delle due principali aziende produttrici di detersivi ha iniziato uno studio sulle abitudini dei consumatori di detersivi. Da un'indagine preliminare risulta che il 50% di consumatori utilizza prodotti diversi dai due marchi principali (di seguito denominati genericamente con $P1$), il 30% di consumatori utilizza il prodotto dell'azienda concorrente, denominato $P2$, mentre il restante 20% preferisce il prodotto dell'azienda che ha avviato lo studio, denominato $P3$.

In seguito a fluttuazioni del mercato dovute a svariati motivi (campagne pubblicitarie, disponibilità del prodotto, qualità percepita dai consumatori) si è portati a modellare il cambiamento di abitudini nel consumo di quei tre prodotti con una catena di Markov. Lo studio dell'evoluzione delle abitudini dei consumatori prosegue per 12 mesi e si riscontra che dopo ogni mese il 60% dei consumatori di $P1$ resta fedele a $P1$, il 70% di consumatori di $P2$ resta fedele a $P2$ mentre il 90% di consumatori di $P3$ resta fedele a $P3$. Gli altri consumatori si ripartiscono in modo uguale tra i due prodotti restanti.

- (a) Classificare gli stati della catena.
 - (b) Che percentuale di consumatori del terzo prodotto ci saranno il terzo mese dopo l'inizio dell'indagine?
 - (c) Come saranno ripartite le percentuali di consumatori dei tre prodotti il sesto mese dall'inizio dell'indagine e alla fine dell'indagine?
 - (d) Dire se esiste e in caso affermativo calcolare la distribuzione invariante.
6. Le hostess di un supermercato, rispetto alla posizione sul lavoro, possono essere in uno dei seguenti stati
- (a) a tempo indeterminato attive
 - (b) a tempo indeterminato in malattia
 - (c) a tempo determinato in formazione
 - (d) a tempo determinato in prova
 - (e) a tempo determinato in malattia
 - (f) senza contratto

Si supponga che il passaggio da uno stato all'altro avvenga secondo un modello del tipo catena di Markov con matrice di transizione data da

$$\begin{bmatrix} 0.95 & 0.05 & 0 & 0 & 0 & 0 \\ 0.25 & 0.75 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & 0.45 & 0.2 & 0.05 \\ 0.4 & 0 & 0 & 0.5 & 0.05 & 0.05 \\ 0 & 0 & 0.05 & 0.9 & 0.05 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- (a) La traiettoria $3 \rightarrow 3 \rightarrow 4 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 6$ è compatibile con tale matrice di transizione? (motivare la risposta).

- (b) Classificare gli stati della catena.
- (c) Determinare tutte le classi chiuse ed irriducibili della catena.
- (d) Dato che all'istante iniziale una lavoratrice si trova nello stato 5, con che probabilità sarà nello stato 2 tra 3 istanti?
- (e) Con riferimento alla matrice che descrive la dinamica delle lavoratrici a tempo indeterminato, si calcoli la distribuzione invariante e se ne spieghi il significato.

10.2 Simulazione

10.3 Processi di Poisson

1. Si supponga che il tempo in ore necessario a riparare una macchina sia una v.c. esponenziale con media 2.
 - (a) Quanto vale la probabilità che la riparazione duri più di due ore?
 - (b) Quanto vale la probabilità che la riparazione duri più di 5 ore dato che sono già trascorse 3 ore dall'inizio della riparazione?
2. La durata X in ore di una macchina, prima che si verifichi un guasto, segue la legge Esponenziale di valore atteso $\mathbf{E}(X) = 2$ ore.
 - (a) Calcolare la probabilità che il primo guasto si verifichi prima di un'ora.
 - (b) Calcolare la probabilità che il terzo guasto si verifichi dopo 3.45 ore, nell'ipotesi che la realizzazione di due guasti successivi siano eventi indipendenti.
3. I clienti giungono ad uno sportello di una banca secondo un processo di Poisson $N(t)$ con tasso $\lambda = \frac{3}{4}$. (È stato calcolato che giungono in media 3 clienti ogni 4 minuti).
 - (a) Qual è la probabilità che non arrivino clienti nel prossimo minuto?
 - (b) Qual è la probabilità che ne arrivino almeno 2?
 - (c) Qual è la probabilità che arrivino almeno 8 clienti in 3 minuti?
 - (d) Qual è la probabilità che in tre minuti arrivino al più sei clienti?
 - (e) Qual è la probabilità che in 30 secondi arrivino almeno 2 clienti?

- (f) Dato che in tre minuti non sono arrivati clienti, qual'è la probabilità che nei prossimi tre minuti arrivino due clienti?
4. I clienti arrivano ad un negozio secondo un processo di Poisson ad un tasso di tre clienti in un ora.
- (a) Se il negozio si prevede apra alle 10:00 ma il commesso per un ritardo dovuto al traffico arriva alle 12:00, quanto vale la probabilità che nessun cliente arrivi nel negozio in quelle due ore?
- (b) Di che tipo è la distribuzione che descrive il tempo d'attesa per l'arrivo del primo cliente, dopo che il commesso è arrivato?
5. Dato il processo di Poisson $N(t)$, $t \geq 0$ con tasso $\lambda = 2$, calcolare

$$P(N(3) = 4 | N(1) = 1) \quad \text{e} \quad P(N(1) = 1 | N(3) = 4).$$

6. Le telefonate ad un centralino arrivano secondo un processo di Poisson con un tasso di 10 telefonate in un ora. Dato che due chiamate sono arrivate nei primi 5 minuti, quanto vale la probabilità che:
- (a) entrambe siano arrivate nei primi due minuti,
- (b) almeno una sia arrivata nei primi due minuti.
7. Si supponga di avere un processo di Poisson $\{N(t) : t \geq 0\}$ a tasso costante $\lambda = \frac{1}{10}$ al secondo.
- (a) Si calcoli la probabilità che in 20 secondi si verifichino esattamente 5 eventi. Giustificare la risposta.
- (b) Si calcoli la probabilità che in 10 secondi non si verifichino eventi. Giustificare la risposta.
- (c) Si calcoli la probabilità che in 5 secondi si verifichi almeno un evento. Giustificare la risposta.
- (d) Quale è la probabilità che il primo arrivo avvenga dopo 5 secondi? Giustificare la risposta.
- (e) Dato che nei primi 5 secondi non si è verificato nessun evento, calcolare la probabilità che nel prossimo secondo si verifichino due eventi. Giustificare la risposta.

- (f) Calcolare la probabilità che nei primi due secondi si verifichi un evento e nell'intervallo da due a cinque secondi un altro evento. Giustificare la risposta.
- (g) Dato che in 15 secondi si sono verificati 11 eventi, qual è la probabilità che nei primi 10 secondi se ne siano verificati 7?
8. La funzione di tasso istantaneo di un processo di Poisson non omogeneo, che conta il numero di eventi che si verificano nell'intervallo $0 \leq t \leq 10$ misurato in secondi, è data da:

$$\lambda(s) = \frac{1}{5}s, \quad 0 \leq s \leq 10$$

- (a) Calcolare il numero atteso di eventi alla fine del periodo di osservazione.
- (b) Calcolare la probabilità che non si siano verificati eventi nell'intervallo $[0, 5]$.
- (c) Si supponga che nell'intervallo $10 \leq t \leq 20$ il tasso del processo di Poisson sia dato da

$$\lambda(s) = 4 - \frac{1}{5}s, \quad 10 \leq s \leq 20$$

Calcolare il numero atteso di eventi nell'intero intervallo $[0, 20]$

9. Un sistema è composto da un solo server che espleta un servizio secondo una v.c. esponenziale con media 20 minuti. I clienti arrivano al servizio secondo un processo di Poisson con tasso $\lambda = 2$ clienti all'ora e se ne vanno se le due sedie della sala d'aspetto sono occupate.
- (a) Qual è lo spazio degli stati per questo sistema?
- (b) Scrivere e risolvere le equazioni del bilancio dettagliato.
- (c) Che percentuale di tempo le sedie saranno entrambe occupate?
- (d) Che percentuale di tempo il sistema sarà vuoto?
- (e) Qual è il numero medio di clienti nel sistema?
- (f) Qual è il tempo medio di permanenza nel sistema?

10.4 Simulazione

In questa sezione sono raccolti alcuni esercizi, sia sulle catene di Markov che sui processi di Poisson, che necessitano delle tecniche di simulazione presentate nel Capitolo 7

1. Una delle due principali aziende produttrici di detersivi ha iniziato uno studio sulle abitudini dei consumatori di detersivi. Da un'indagine preliminare risulta che il 50% di consumatori utilizza prodotti diversi dai due marchi principali (di seguito denominati genericamente con $P1$), il 30% di consumatori utilizza il prodotto dell'azienda concorrente, denominato $P2$, mentre il restante 20% preferisce il prodotto dell'azienda che ha avviato lo studio, denominato $P3$. In seguito a fluttuazioni del mercato dovute a svariati motivi (campagne pubblicitarie, disponibilità del prodotto, qualità percepita dai consumatori) si è portati a modellare il cambiamento di abitudini nel consumo di quei tre prodotti con una catena di Markov. Lo studio dell'evoluzione delle abitudini dei consumatori prosegue per 12 mesi e si riscontra che dopo ogni mese il 60% dei consumatori di $P1$ resta fedele a $P1$, il 70% di consumatori di $P2$ resta fedele a $P2$ mentre il 90% di consumatori di $P3$ resta fedele a $P3$. Gli altri consumatori si ripartiscono in modo uguale tra i due prodotti restanti.
 - (a) Classificare gli stati della catena.
 - (b) Che percentuale di consumatori del terzo prodotto ci saranno il terzo mese dopo l'inizio dell'indagine?
 - (c) Come saranno ripartite le percentuali di consumatori dei tre prodotti il sesto mese dall'inizio dell'indagine e alla fine dell'indagine?
 - (d) Rappresentare in un grafico l'andamento della percentuale dei consumatori di $P3$ durante i 12 mesi di indagine.
 - (e) Fare lo stesso per il prodotto $P1$ e per il prodotto $P2$.
 - (f) Rappresentare in un unico grafico i tre andamenti.
 - (g) Simulare il comportamento di un consumatore per 12 mesi.
 - (h) Simulare il comportamento di 1000 consumatori nei 12 mesi.
 - (i) Riportare in un grafico la ripartizione percentuale dei 1000 clienti tra i 3 prodotti alla fine sesto mese e alla fine del periodo di studio. Confrontare queste ripartizioni simulate con le percentuali calcolate nel punto 3.
 - (j) Calcolare la distribuzione invariante.
 - (k) Stimare la distribuzione invariante come inverso del tempo medio di ritorno, simulando una traiettoria lunga 10000 passi. Confrontare i valori stimati con quelli veri.
2. Un sistema per l'elaborazione delle informazioni si può trovare in 3 stati. Stato 1 corrispondente a funzionante, stato 2 corrispondente a funzionante ma non

soddisfacente, stato 3, guasto. Dall'analisi del sistema effettuata per un lungo periodo alla fine di ogni turno di lavoro si è giunti alla seguente conclusione. Si può ipotizzare che lo stato del sistema ad ogni istante sia descritto dallo stato di una catena di Markov omogenea in cui la rilevazione dello stato venga effettuata alla fine di ogni turno di lavoro. La matrice di transizione, dedotta dalle informazioni riportate dai tecnici che rilevano lo stato del sistema alla fine di ogni turno di lavoro, è data da:

$$\begin{pmatrix} 0.7 & 0.28 & 0.02 \\ 0 & 0.75 & 0.25 \\ 0 & 0 & 1 \end{pmatrix}$$

- (a) Classificare gli stati della catena.
- (b) Calcolare la matrice di transizione P^n , per $n = 2, \dots, 20$ e commentarne l'evoluzione.
- (c) Dato che oggi la catena si trova nello stato 1, quale è la probabilità che tra 9 giorni sia nello stato 3?
- (d) Dato che oggi la catena si trova nello stato 1, calcolare la probabilità che il giorno n si trovi nello stato 3. Rappresentare in un grafico l'andamento di tale probabilità per $n = 1, 2, \dots, 20$.
- (e) Calcolare il tempo medio di assorbimento nello stato 3, sia nel caso in cui la catena è partita dallo stato 1 sia quando è partita dallo stato 2.
- (f) Simulare il comportamento del sistema supponendo che si parta dallo stato 1 e seguirne l'evoluzione. Rappresentare graficamente l'evoluzione del sistema simulato. In particolare osservare il tempo impiegato a raggiungere lo stato 3 (tempo di assorbimento).
- (g) Simulare il comportamento di 1000 sistemi che partono dallo stato 1, e rappresentare graficamente la distribuzione dei tempi di assorbimento.
- (h) Confrontare la media dei tempi di assorbimento ottenuti con le 1000 simulazioni con il tempo medio calcolato al punto 5. quando la catena parte dallo stato 1.

Bibliografia

- [1] Baldi, P., (1992), *Calcolo delle Probabilità e Statistica*, McGraw-Hill, Milano.
- [2] Billingsley, P., (1995), *Probability and Measure*, 3rd ed., Wiley & Sons, New York.
- [3] Cox, D.R., Miller, H.D., (1965) *The theory of stochastic Processes*, Chapman and Hall, London.
- [4] Feller, W., (1971), *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed., Vol. II, 2nd ed., Wiley & Sons, New York, 1968.
- [5] S.M Iacus, G. Masarotto (2007) *Laboratorio di Statistica con R*, 2nd Ed. McGraw-Hill, Milano.
- [6] Hoel, P.G., Port, S.C., Stone, C.J., (1972), *Introduction to Stochastic Processes*, Houghton Mifflin Co, Boston.
- [7] Resnick, S. I., (1992), *Adventures in Stochastic Processes. The Random World of happy Hayry*, Birkh auser, Boston.
- [8] Ross, S., (1997), *Simulation*, Academic Press, San Diego.
- [9] Ross, S., (1997), *Introduction to Probability Models*, Academic Press, San Diego.