Alma Mater Studiorum · Università di Bologna

FACOLTA' DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea Triennale in Matematica

PROBABILITA' DI TRANSIZIONE PER ALCUNI PROCESSI DELLA TEORIA DELLE CODE

Tesi di Laurea in Probabilità e Statistica matematica

Relatore:
Chiar.mo Prof.
Massimo Campanino

Presentata da:
Sabina Smerilli

III Sessione
Anno Accademico 2012/2013

Ai miei genitori,
Gabriele e Luciana

e a mia sorella Silvia.

Introduzione

Lo scopo di questa tesi è lo studio dei processi di Markov con tempo continuo legati alla teoria delle code. Primo passo in questa direzione sarà quello di introdurre il concetto di processo stocastico e subito dopo analizzarne un tipo particolare chiamato appunto processo di Markov, il cui nome deriva dal matematico russo A. A. Markov (1856-1922) che per primo ne sviluppò la teoria.
Si dirà che un processo stocastico con spazio degli stati discreto è markoviano se le probabilità di transizione (ossia le probabilità che regolano il passaggio da uno stato ad un altro) dipendono unicamente dallo stato di sistema immediatamente precedente (proprietà di Markov) e non dal come si è giunti in tale stato.
In questo elaborato saranno considerati solo i processi di Markov con tempo continuo, e poiché lo scopo del loro studio è la previsione delle loro probabili sequenze operative al trascorrere del tempo, verrà studiato il loro comportamento in transitorio e all'equilibrio. Quanto detto è descritto nel Capitolo 1.

Nel Capitolo 2 si passerà ad analizzare il particolare processo markoviano di nascita e morte con tempo continuo la cui caratteristica peculiare è quella di poter transitare solo negli stati adiacenti, in altre parole se si ha un sistema con spazio degli stati a valori interi non negativi che si trova in una stato k, esso potrà soltanto passare nello stato $k+1$ (nascita) o $k-1$ (morte) . Poiché anche di questo processo si studierà il comportamento in transitorio e all'equilibrio, esso potrà essere utilizzato per simulare l'evoluzione nel tempo del numero degli utenti presenti in un sistema i cui unici cambiamenti siano le nascite o le morti e dunque i processi di nascita e morte saranno degli ottimi modelli per la teoria elementare delle code.

Il terzo e ultimo capitolo si aprirà con la descrizione di quelle che sono le componenti principali di un sistema a coda (o coda). Un sistema a coda, come suggerisce la parola stessa, può essere fisicamente vista come un flusso di clienti che chiedono di ricevere un servizio erogato da un centro di servizio e nel caso tutti i servitori siano siano occupati aspetteranno il loro turno ordinatamente formando una fila d'attesa. Da un punto di vista matematico e sotto opportune condizioni la teoria delle code diventa lo studio di due processi stocastici, quali il processo d'arrivo dei clienti A e il processo di servizio B. Un processo a coda di questo tipo sarà indicato con la notazione $A/B/n$ dove n denota il numero dei servitori.
Considereremo un esempio particolare di sistema a coda e precisamente la coda $M/M/1$ dove M indica un processo markoviano (che si vedrà avere una distribuzione dei tempi di tipo esponenziale).
Si noterà in questo esempio che le proprietà studiate dei processi di nascita e morte saranno applicabili nello studio della coda $M/M/1$ per la determinazione delle probabilità che nel sistema siano presenti k utenti in un tempo t o a regime. Questo sarà possibile perché nel contesto della teoria delle code una nascita si riferirà ad un arrivo di un utente nel sistema e una morte ad una uscita dal sistema una volta espletato il servizio.

Indice

Capitolo 1

Catene di Markov

1.1 Processi stocastici

Un *processo stocastico* è semplicemente definito come una famiglia di variabili aleatorie $\{X_t\}$, dove il parametro t, che di solito ha il significato di tempo, appartiene ad un dato insieme T. In particolare, se l'insieme T è un insieme discreto, cioè la sua cardinalità è finita o numerabile, allora il processo si dice con *con tempo discreto*; se invece l'insieme T assume valori su un intervallo continuo, finito o infinito (o su insiemi di tali intervalli), allora il processo si dice con *tempo continuo.* Nel secondo caso la notazione X_t viene sostituita da $X(t)$.

I processi stocastici sono in più classificati dallo *spazio degli stati.* Questo è formato dall'insieme di tutti i possibili valori, detti stati, che le variabili aleatorie X_t o $X(t)$ possono assumere; si parlerà quindi di *processi discreti* se esse sono variabili discrete, di *processi continui* se sono continue.
Lo spazio degli stati si denota con $I(X_t)_{t \in T}$ e può essere quindi finito, infinito e numerabile o infinito e non numerabile.

Ciò che caratterizza veramente un processo stocastico è la relazione che intercorre tra le variabili casuali $X(t)$ e gli altri membri della stessa famiglia. A tale scopo specifichiamo alcune definizioni.
Si chiama *funzione di distribuzione di probabilità (DDP) di una variabile*

aleatoria X una funzione $F_X : \mathbb{R} \longrightarrow [0,1]$ data da:

$$F_X(x;t) \triangleq \mathbf{P}[X(t) \leq x]$$

per ogni t ammesso e per ogni $x \in \mathbb{R}$.
Inoltre, si chiama *funzione di distribuzione di probabilità congiunta del vettore aleatorio* $(X_1, X_2, ..., X_n)$ di dimensione n, la funzione definita da:

$$F_{X_1 X_2 \cdots X_n}(x_1, x_2, ..., x_n; t_1, t_2, ..., t_n) \triangleq \mathbf{P}[X(t_1) \leq x_1, ..., X(t_n) \leq x_n]$$

per ogni $\mathbf{x} = (x_1, x_2, ..., x_n), \mathbf{t} = (t_1, t_2, ..., t_n)$ e per tutti gli n. Usiamo la notazione vettoriale $F_{\mathbf{X}}(\mathbf{x};\mathbf{t})$ per indicare questa funzione.
Infine si definisce *funzione di densità di probabilità (ddp) congiunta*, la funzione

$$f_{\mathbf{X}}(\mathbf{x};\mathbf{t}) \triangleq \frac{\partial F_{\mathbf{X}}(\mathbf{x};\mathbf{t})}{\partial \mathbf{x}}$$

Un processo stocastico $X(t)$ è detto *stazionario* se tutte le $F_{\mathbf{X}}(\mathbf{x};\mathbf{t})$ sono invarianti rispetto a traslazioni nel tempo; cioè per ogni costante τ vale quanto segue:

$$F_{\mathbf{X}}(\mathbf{x};\mathbf{t}+\tau) \triangleq F_{\mathbf{X}}(\mathbf{x};\mathbf{t})$$

dove la notazione $\mathbf{t}+\tau$ indica il vettore $(t_1+\tau, t_2+\tau, ..., t_n+\tau)$. Queste funzioni casuali stazionarie costituiscono il maggior interesse nella teoria dei processi stocastici.

Per specificare completamente le caratteristiche probabilistiche di un processo stocastico occorre dare la $F_{\mathbf{X}}(\mathbf{x};\mathbf{t})$ (o la $f_{\mathbf{X}}(\mathbf{x};\mathbf{t})$) per tutti i possibili sottoinsiemi di $\{x_i\}, \{t_i\}$ e per tutti gli n. In generale si tratta di un compito sovrumano! Per fortuna è possibile fornire una simile descrizione in termini molto semplici per molti dei particolari processi stocastici interessanti.

1.2 Processi di Markov con tempo continuo

I processi di Markov, anche detti *catene di Markov*, devono il loro nome al matematico russo A. A. Markov che studiò questa classe di processi nei primi anni anni del secolo. Questi processi sono caratterizzati da una semplice ma molto utile dipendenza tra le variabili casuali chiamata proprietà di Markov.

Definizione 1.1. (**Proprietà di Markov**) Un processo stocastico $(X(t))_{t\in T}$ verifica la proprietà di Markov se per ogni coppia di stati $i, j \in I(X(t))_{t\in T}$, per ogni $n \in \mathbb{N}$ e per ogni sequenza $t_0 < t_1 < \cdots < t_{n-2} < t_{n-1} < t_n$, risulta:

$$\mathbf{P}[X(t_n) = j \mid X(t_0) = i_0, X(t_1) = i_1, ..., X(t_{n-2}) = i_{n-2}, X(t_{n-1}) = i]$$

$$= \mathbf{P}[X(t_n) = j \mid X(t_{n-1}) = i].$$

Intuitivamente si capisce che l'evoluzione futura del processo dall'istante t_{n-1} in poi non dipende dalla storia passata del processo stocastico; quindi lo stato corrente $X(t_{n-1}) = i$ riassume tutta l'informazione riguardante l'evoluzione passata del processo.

In questo testo considereremo unicamente processi di Markov con tempo continuo e spazio degli stati discreto e senza perdere generalità supponiamo che T sia l'insieme dei reali non negativi, cioè $T = \mathbb{R}^+ = [0, \infty)$.
Ad ogni istante di tempo $t \in T$ il sistema si troverà quindi in un determinato stato appartenente ad un insieme numerabile di stati esaustivi ed mutuamente esclusivi che da qui in poi chiameremo S. Qualora risulti $X(t) = i \in S$ si dirà che il sistema al tempo t si trova nello stato i.
Una catena di Markov è detta *catena di Markov con un numero finito di stati* nel caso in cui l'insieme S degli stati è finito.

Un'altra restrizione che assumeremo è quella di considerare solo processi di Markov stazionari, cioè il cui comportamento futuro non dipende dal tempo di osservazione. Essa può essere scritta come

$$\mathbf{P}[X(t_n) = j \mid X(t_{n-1}) = i] = \mathbf{P}[X(t_n - t_{n-1}) = j \mid X(0) = i]$$

ed una catena di Markov per la quale vale questa condizione è detta *omogenea*. Per specificare in termini probabilistici il suo comportamento futuro non è necessario dunque conoscere t_{n-1} ma esso è completamente determinato dalla conoscenza dello stato corrente i.

In particolare, la proprietà di Markov implica inoltre che non è necessario specificare per quanto tempo il processo è rimasto nello stato corrente i; quindi il tempo di permanenza in questo stato deve soddisfare la seguente proprietà di *assenza di memoria*:

$$\forall\, \tau, t \geq 0 \qquad \mathbf{P}[X_i > \tau + t \mid X_i > \tau] = \mathbf{P}[X_i > t]. \tag{1.1}$$

In termini di tempo, questa proprietà significa che la probabilità che un evento dipendente soltanto dallo stato i avvenga dopo $\tau + t$ istanti, dato che è avvenuto dopo τ, è uguale alla probabilità che avvenga dopo t istanti partendo dall'istante zero.

Teorema 1.2.1. *La distribuzione esponenziale è l'unica distribuzione che verifica la proprietà di assenza di memoria.*

Dimostrazione. La proprietà 1.1 può essere formalizzata come

$$\mathbf{P}[X_i > \tau + t \mid X_i > \tau] \triangleq h(t) \tag{1.2}$$

dove $h(t)$ è una funzione solo del tempo aggiuntivo t (e non del tempo già passato τ).
Ora la probabilità condizionata $\mathbf{P}[X_i > \tau + t \mid X_i > \tau]$ si può riscrivere come

$$\begin{aligned} \mathbf{P}[X_i > \tau + t \mid X_i > \tau] &= \frac{\mathbf{P}[X_i > \tau + t,\ X_i > \tau]}{\mathbf{P}[X_i > \tau]} \\ &= \frac{\mathbf{P}[X_i > \tau + t]}{\mathbf{P}[X_i > \tau]} \end{aligned}$$

dove l'ultima uguaglianza è data dal fatto che l'evento $X_i > \tau + t$ implica l'evento $X_i > \tau$. Dunque, da quest'ultima e da 1.2 segue che

$$\mathbf{P}[X_i > \tau + t] = \mathbf{P}[X_i > \tau] h(t). \tag{1.3}$$

Ponendo $\tau = 0$ si ottiene

$$\mathbf{P}[X_i > t] = h(t)$$

che sostituito alla 1.3 si ha

$$\mathbf{P}[X_i > \tau + t] = \mathbf{P}[X_i > \tau]\mathbf{P}[X_i > t] \quad \forall\, t, \tau \geq 0. \tag{1.4}$$

Mostriamo prima di procedere che in generale vale la seguente relazione

$$\begin{aligned} \frac{d}{dt}(\mathbf{P}[X_i > t]) &= \frac{d}{dt}(1 - \mathbf{P}[X_i \leq t]) \\ &= -f_{X_i}(t) \end{aligned} \tag{1.5}$$

dove $f_{X_i}(t)$ è la funzione di densità di probabilità (ddp) di X_i. Sfruttando la proprietà appena definita differenziamo la 1.4 rispetto a τ ottenendo

$$\frac{d\mathbf{P}[X_i > \tau + t]}{d\tau} = -f_{X_i}(\tau)\mathbf{P}[X_i > t].$$

Dividendo ambo i membri per $\mathbf{P}[X_i > t]$ e ponendo $\tau = 0$ si ottiene

$$\frac{d\mathbf{P}[X_i > t]}{\mathbf{P}[X_i > t]} = -f_{X_i}(0)d\tau.$$

Integrando quest'ultima equazione tra 0 e t si ha

$$\log_e \mathbf{P}[X_i > t] = -f_{X_i}(0)t$$

che in modo equivalente si scrive

$$\mathbf{P}[X_i > t] = e^{-f_{X_i}(0)t}.$$

Nello stesso modo di quanto si è appena fatto se applichiamo di nuovo la 1.5 per ricavare la ddp o la DDP di X_i si ottiene

$$f_{X_i}(t) = f_{X_i}(0)e^{-f_{X_i}(0)t} \quad t \geq 0$$

ed

$$F_{X_i}(t) = [1 - e^{-f_{X_i}(0)t}] \quad t \geq 0$$

che porta alla conclusione che il tempo che una catena di Markov con tempo continuo spende nello stato i è distribuito esponenzialmente con parametro $f_{X_i}(0)$, il quale può dipendere dallo stato i o più in generale che una catena di Markov con tempo continuo deve avere i tempi di permanenza negli stati distribuiti esponenzialmente. $\square$

La catena $(X(t))_{t\geq 0}$ si può in conclusione vedere come un sistema che evolve nel tempo passando da uno stato all'altro di S in maniera aleatoria.

Definizione 1.2. La probabilità condizionata $\mathbf{P}[X(t) = j \mid X(s) = i] \; \forall t > s$ che rappresenta la probabilità che il processo si trovi nello stato j al tempo t, dato che si trova nello stato i al tempo s è detta *probabilità di transizione* della catena con tempo continuo. Se la catena è omogenea si può facilmente intuire che le *probabilità di transizione sono stazionarie*, ovvero dipendono solo dalla differenza $\tau = t - s$ e si usa la seguente notazione

$$\begin{aligned} p_{ij}(\tau) =& \mathbf{P}[X(s+\tau) = j \mid X(s) = i] \\ =& \mathbf{P}[X(\tau) = j \mid X(0) = i] \qquad \forall i, j \in S, \forall \tau \geq 0, \forall s \geq 0 \end{aligned}$$

per esprimere la probabilità che il processo si trovi nello stato j dopo un intervallo di durata τ, dato che attualmente esso si trova nello stato i.

Quando $\tau = 0$ si ha:

$$p_{ij}(0) = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{altrimenti} \end{cases} \qquad \forall i, j \in S.$$

Le catene di Markov con spazio degli stati finito possono essere facilmente studiate rappresentando le probabilità di transizione stazionarie nella seguente forma matriciale.

Definizione 1.3. Si consideri una catena di Markov con un numero finito N di stati S. Si definisce *matrice di transizione* la seguente matrice $P(\tau)$:

$$P(\tau) = [p_{ij}(\tau)] \in [0,1]^{N\times N} \qquad i, j = 1, ..., N \in S$$

con P(0)= I dove I rappresenta la matrice identità (in un tempo nullo non si avrà possibilità di transizione di stato).
La matrice di transizione $P(\tau)$ è quindi una matrice quadrata di ordine N, con $0 \leq p_{ij}(\tau) = [P(\tau)]_{ij} \leq 1$ (elemento nella riga i e nella colonna j)

viene indicata la probabilità di passare dallo stato i allo stato j e gode della seguente proprietà:

$$\sum_{j=1}^{N} p_{ij}(\tau) = 1 \qquad \text{con} \quad 0 \leq p_{ij}(\tau) \leq 1 \qquad \forall i \in S, \forall \tau \geq 0 \; \forall i = 1, ..., N.$$

In modo equivalente se lo spazio degli stati non fosse finito, la matrice di transizione sarebbe una matrice con infinite righe e infinite colonne con elementi non negativi tali che la somma delle serie degli elementi di ogni riga sia uguale ad 1. Cioè

$$\sum_{j \in S} p_{ij}(\tau) = 1 \qquad \text{con} \quad 0 \leq p_{ij}(\tau) \leq 1 \qquad \forall i \in S, \forall \tau \geq 0.$$

1.2.1 Equazioni di Chapman-Kolmogorov

Si può osservare che il processo per passare dallo stato i allo stato j in un intervallo di tempo t deve passare per qualche stato intermedio k in un istante di tempo intermedio dell'intervallo t. Si ha allora la seguente formulazione dell'*equazione di Chapman-Kolmogorov* per le catene di Markov con tempo continuo:

$$p_{ij}(t) = \sum_{k \in S} p_{ik}(t_1) p_{kj}(t_2) \quad \forall i, j \in S, \forall t_1, t_2 \geq 0 \text{ con } t_1 + t_2 = t \tag{1.6}$$

Dimostrazione.

$$\begin{aligned}
p_{ij}(t_1 + t_2) &= \mathbf{P}[X(t_1 + t_2) = j \mid X(0) = i] \\
&= \sum_{k \in S} \mathbf{P}[X(t_1 + t_2) = j, X(t_1) = k \mid X(0) = i] \\
&= \sum_{k \in S} \mathbf{P}[X(t_1 + t_2) = j \mid X(t_1) = k, X(0) = i] \mathbf{P}[X(t_1) = k \mid X(0) = i] \\
&= \sum_{k \in S} \mathbf{P}[X(t_1 + t_2) = j \mid X(t_1) = k] \mathbf{P}[X(t_1) = k \mid X(0) = i] \\
&= \sum_{k \in S} p_{ik}(t_1) p_{kj}(t_2)
\end{aligned}$$

□

In modo equivalente scriviamo l'equazione di Chapman-Kolmogorov 1.6 per un qualunque intervallo $t + \Delta t$ nel modo seguente:

$$p_{ij}(t + \Delta t) = \sum_{k \in S} p_{ik}(t) p_{kj}(\Delta t) \qquad \forall i, j \in S, \forall t, \Delta t \geq 0$$

Formuliamo ora la seguente differenza

$$\begin{aligned} p_{ij}(t + \Delta t) - p_{ij}(t) &= \sum_{k \in S} p_{ik}(t) p_{kj}(\Delta t) - \sum_{k \in S} p_{ik}(t) p_{kj}(0) \\ &= \sum_{k \in S} p_{ik}(t) [p_{kj}(\Delta t) - p_{kj}(0)] \qquad \forall i, j \in S \end{aligned}$$

Dividendo ambo i membri per Δt e prendendo il limite per $\Delta t \longrightarrow 0$, troviamo

$$\begin{aligned} \frac{dp_{ij}(t)}{dt} &= \sum_{k \in S} p_{ik}(t) q_{kj} \qquad \forall i, j \in S, \forall t \geq 0 \qquad (1.7) \\ &= p_{ij}(t) q_{jj} + \sum_{k \neq j} p_{ik}(t) q_{kj} \end{aligned}$$

dove abbiamo definito i termini q_{ij}, chiamati *tassi di transizione*, con i seguenti limiti:

$$\begin{aligned} q_{ii} &= \lim_{\Delta t \to 0} \frac{p_{ii}(\Delta t) - 1}{\Delta t} \qquad \forall i \in S \\ q_{ij} &= \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t)}{\Delta t} \qquad \forall i, j \in S \quad i \neq j \end{aligned}$$

E' possibile dimostrare che questi limiti esistono sotto opportune condizioni di regolarità. Essi ci spiegano che se il sistema ad un certo istante t si trova nello stato i, allora la probabilità che avvenga una transizione (verso un qualunque altro stato diverso da i) durante l'intervallo di tempo Δt è data da $-q_{ii}\Delta t + o(\Delta t)$. Quindi possiamo dire che $-q_{ii}$ rappresenta il *tasso* col quale il processo *esce* dallo stato i quando si trova in tale stato. Analogamente la probabilità condizionata di una transizione dallo stato i allo stato j nell'intervallo di tempo Δt è data da $q_{ij}\Delta t + o(\Delta t)$. Quindi q_{ij} è il tasso (la velocità) con cui il processo *si sposta* dallo stato i allo stato j, dato che il processo sia attualmente nello stato i. Per $o(\Delta t)$ si intende una funzione tale che $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$, cioè si dice che $o(\Delta t)$ è un infinitesimo di ordine superiore a Δt. Poiché è sempre verificata la condizione $\sum_{j \in S} p_{ij}(\Delta t) = 1$, vediamo che le equazioni definite dai limiti precedenti implicano che

$$\sum_{j \in S} q_{ij} = 0 \quad \text{per ogni } i$$

Dimostrazione. Per un numero finito N di stati vale:

$$\begin{aligned} \sum_{j=1}^{N} q_{ij} &= \sum_{j=1, j \neq i}^{N} \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t)}{\Delta t} + \lim_{\Delta t \to 0} \frac{p_{ii}(\Delta t) - 1}{\Delta t} \\ &= \lim_{\Delta t \to 0} [\sum_{j=1}^{N} \frac{p_{ij}(\Delta t)}{\Delta t} - \frac{1}{\Delta t}] \\ &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} [\sum_{j=1}^{N} p_{ij}(\Delta t) - 1] = 0 \end{aligned}$$

Pertanto abbiamo interpretato i termini dell'equazione 1.7; essa è nota come l'equazione *in avanti* di Chapman-Kolmogorov per le Catene di Markov con tempo continuo e questo perché sono scritte fissando lo stato i di partenza e scorrendo poi sullo stato di arrivo.

Ora, se ad essere fissato è lo stato j di arrivo e si scorre sullo stato di partenza, è possibile ricavare la seguente equazione di Chapman-Kolmogorov *all'indietro*:

$$\begin{aligned} \frac{dp_{ij}(t)}{dt} &= \sum_{k \in S} q_{ik} p_{kj}(t) \qquad \forall i, j \in S, \forall t \geq 0 \qquad (1.8) \\ &= q_{ii} p_{ij}(t) + \sum_{k \neq i} q_{ik} p_{kj}(t). \end{aligned}$$

Diamo ora una formulazione matriciale di quanto appena detto.

Definizione 1.4. Si consideri una catena di Markov con tempo continuo e spazio degli stati numerabile. Si definisce *matrice infinitesimale* della matrice di transizione $P(t)$, o anche *matrice dei tassi di transizione*, la matrice

$$Q \triangleq [q_{ij}].$$

L'equazione di Chapman-Kolmogorov 1.6 può essere così scritta in forma matriciale come:

$$P(t_1 + t_2) = P(t_1)P(t_2).$$

Mentre le equazioni di Kolmogorov in avanti e all'inditro sono espresse rispettivamente da:

$$\frac{dP(t)}{dt} = P(t)Q, \quad \frac{dP(t)}{dt} = QP(t).$$

1.2.2 Classificazione degli stati

Lo studio della catena di Markov con tempo continuo prosegue con alcune definizioni che riguardano la struttura della catena e le proprietà che legano i suoi stati.

Definizione 1.5. Si consideri una catena di Markov e due suoi stati $i, j \in S$.

i) si dice che lo stato i *comunica* con lo stato j (o che j è accessibile da i) se esiste un intervallo di tempo $t \geq 0$ tale che $p_{ij}(t) > 0$, in tal caso useremo la seguente notazione:

$$i \prec j$$

ii) gli stati i e j si dicono *equivalenti* (o comunicanti) se lo stato i comunica con lo stato j ed a sua volta lo stato j comunica con lo stato i; cioè $\exists\, t, s \geq 0$ tali che $p_{ij}(t) > 0$ e $p_{ji}(s) > 0$.
In simboli:

$$i \prec j \quad \text{e} \quad j \prec i$$

La relazione $\prec$ di comunicazione gode delle proprietà riflessiva e transitiva mentre la simmetrica non sempre vale.
La seconda relazione che definisce due stati equivalenti è invece una relazione di equivalenza. Infatti, dati gli stati i, j, h e definita la relazione nel modo seguente

$$i \sim j \Leftrightarrow i \prec j \quad \text{e} \quad j \prec i$$

si ha:

i) la proprietà riflessiva segue dal fatto che $p_{ii}(0) = 1$;

ii) la proprietà simmetrica segue banalmente dalla definizione;

iii) dimostriamo la proprietà transitiva (se $i \sim j \quad \text{e} \quad j \sim h \Rightarrow i \sim h$):
$i \sim j \Rightarrow \exists\, t_1, t_2 \mid p_{ij}(t_1) > 0 \quad \text{e} \quad p_{ji}(t_2) > 0$;
$j \sim h \Rightarrow \exists\, t_3, t_4 \mid p_{jh}(t_3) > 0 \quad \text{e} \quad p_{hj}(t_4) > 0$;
Si può concludere che $i \sim h$ in quanto, se si sceglie $t = t_1 + t_2$ e $\tilde{t} = t_3 + t_4$, si ottiene:
$p_{ih}(t_1 + t_2) = [P(t_1 + t_2)]_{ih} = \sum_{k \in S} p_{ik}(t_1) p_{kh}(t_2) \geq \underbrace{p_{ij}(t_1)}_{>0} \underbrace{p_{jh}(t_2)}_{>0} > 0$;
$p_{hi}(t_3 + t_4) = [P(t_3 + t_4)]_{hi} = \sum_{k \in S} p_{hk}(t_3) p_{ki}(t_4) \geq \underbrace{p_{hj}(t_3)}_{>0} \underbrace{p_{ji}(t_4)}_{>0} > 0$;
da cui la tesi.

Indichiamo la classe di equivalenza con $[i] = \{j \mid j \sim i\}$. Gli stati di una catena di Markov possono quindi essere partizionati in classi di equivalenza disgiunte tali che tutti gli stati di una certa classe sono comunicanti tra loro (una classe può essere composta anche da un singolo stato).
La relazione $\prec$ tra stati si può estendere alle classi di equivalenza e diciamo che $[i]$ comunica con $[j]$, in simboli $[i] \prec [j]$, se $i \prec j$. Facilmente si vede che questa relazione è ben definita perché non dipende dalla scelta dei rappresentanti delle classi di equivalenza.

Definizione 1.6. Una catena di Markov è detta *irriducibile* se tutti gli stati del processo sono equivalenti tra loro (ogni stato della catena può essere raggiunto da ogni altro stato), ossia se esiste una sola classe di equivalenza.

Vediamo adesso come è possibile differenziare tra loro gli stati di una catena.

Definizione 1.7. Sia H_j il *tempo di primo ingresso* nello stato j, cioè l'istante di tempo in cui il processo entra nello stato j per la prima volta dopo aver lasciato lo stato corrente. Inoltre, sia

$$f_{ij} = \mathbf{P}[H_j < \infty \mid X(0) = i] \quad \forall i, j \in S$$

Allora:

- uno stato j è detto *transiente* se $f_{jj} < 1$, cioè se c'è una probabilità non nulla che il processo non ritorni mai allo stato j dopo averlo lasciato;

- uno stato j è detto *ricorrente* se $f_{jj} = 1$, cioè se il processo ritorna allo stato j prima o poi con probabilità 1. In più se il tempo medio di attesa per ritornare allo stato j per la prima volta dopo averlo lasciato è finito, allora lo stato j è detto *ricorrente positivo.* Mentre è detto *ricorrente nullo* se il ritorno a j è comunque certo ma il suo tempo di attesa potrebbe anche essere infinito.

- uno stato j è detto *assorbente* (o trappola) se $q_{ji} = 0$ per tutti gli $i \neq j$, quindi se $q_{jj} = 0$.

Si osserva facilmente che per uno stato $j \in S$ *ricorrente*, la probabilità di transizione $p_{jj}(t)$ deve essere maggiore di 0 per qualche valore di $t > 0$.

E' inoltre possibile dimostrare che:

$$\text{uno stato } j \in S \text{ è } \textit{ricorrente} \text{ se e solo se } \int_1^{+\infty} p_{jj}(t)dt = +\infty; \tag{1.9}$$

$$\text{uno stato } j \in S \text{ è } \textit{transiente} \text{ se e solo se } \int_1^{+\infty} p_{jj}(t)dt < \infty. \tag{1.10}$$

Si verificano inoltre le seguenti proprietà:

1. In una stessa classe di equivalenza gli stati sono tutti ricorrenti o tutti transienti;

   Dimostrazione. Siano i e j due qualsiasi stati comunicanti della classe considerata, per definizione quindi $\exists\ t_1, t_2 \geq 0$ tali che $p_{ij}(t_1) > 0$ e $p_{ij}(t_2) > 0$. Per ogni $t \geq 0$ risulta:

   $p_{jj}(t_1+t+t_2) \geq p_{ij}(t_1)p_{jj}(t)p_{ji}(t_2)$ e $p_{jj}(t_1+t+t_2) \geq p_{ji}(t_2)p_{ii}(t)p_{ij}(t_1)$

e quindi, rispettivamente dalla prima e dalla seconda disequazione, si ha:

$$\begin{aligned}\int_0^{+\infty} p_{jj}(t)dt &\leq \frac{1}{p_{ij}(t_1)p_{ji}(t_2)} \int_0^{+\infty} p_{ii}(t_1+t+t_2)dt \\ &\leq \frac{1}{p_{ij}(t_1)p_{ji}(t_2)} \int_0^{+\infty} p_{ii}(t)dt \end{aligned} \tag{1.11}$$

$$\begin{aligned}\int_0^{+\infty} p_{jj}(t)dt &\geq \int_0^{+\infty} p_{jj}(t_1+t+t_2)dt \\ &\geq p_{ij}(t_1)p_{ji}(t_2) \int_0^{+\infty} p_{ii}(t)dt \end{aligned} \tag{1.12}$$

Se lo stato i è ricorrente allora per 1.9 $\int_1^{+\infty} p_{ii}(t)dt = +\infty$, dalla 1.2.2 segue quindi che $\int_1^{+\infty} f_{jj}(t)dt = +\infty$ e quindi anche j è ricorrente.
Se lo stato i è transiente allora per 1.10 $\int_1^{+\infty} p_{ii}(t)dt < +\infty$, dalla 1.2.2 segue quindi che $\int_1^{+\infty} f_{jj}(t)dt < \infty$ e quindi anche j è transiente. □

2. in una catena di Markov con un numero finito di stati non tutti gli stati possono essere transienti (se ciò fosse vero dopo un tempo finito il sistema potrebbe abbandonare tutti i suoi stati, il che è assurdo);

3. *tutti gli stati di una catena di Markov con un numero finito di stati irriducibile sono ricorrenti* (la catena ha una sola classe di equivalenza ed i suoi stati non possono essere tutti transienti).

1.2.3 Comportamento in transitorio e soluzione stazionaria

Si vuole ora studiare il comportamento delle catene di Markov con tempo continuo al passare del tempo.

Definizione 1.8. Sia $(X(t))_{t\geq 0}$ una catena di Markov con tempo continuo. Si denota con

$$\pi_j(t) \triangleq \mathbf{P}[X(t) = j] \quad \forall j \in S, \forall t \geq 0$$

e, in particolare con

$$\pi_j(0) \triangleq \mathbf{P}[X(0) = j] \quad \forall j \in S$$

rispettivamente la probabilità di trovare il sistema nello stato j al tempo t e la probabilità di trovare il sistema nello stato j all'inizio del processo. Quindi si ha che $\{\pi_j(0), j \in S\}$ è la distribuzione di stato iniziale.

Detto ciò, possiamo valutare $\pi_j(t)$ dalle probabilità di transizione e dalla distribuzione iniziale del processo, infatti per il teorema della probabilità totale, abbiamo:

$$\begin{aligned}\pi_j(t) = \mathbf{P}[X(t) = j] &= \sum_{i \in S} \mathbf{P}[X(t) = j \mid X(0) = i]\mathbf{P}[X(0) = i] \\ &= \sum_{i \in S} \pi_i(0) p_{ij}(t) \quad \forall j \in S, \forall t \geq 0. \end{aligned} \tag{1.13}$$

Ora, al fine di conoscere la probabilità che il processo si trovi nello stato j ad un istante arbitrario t cominciamo col derivare la 1.13. Si ha così

$$\frac{d\pi_j(t)}{dt} = \sum_{i \in S} \frac{dp_{ij}(t)}{dt} \pi_i(0) \quad \forall j \in S.$$

Sostituendo ad essa l'equazione di Chapman-Kolmogorov in avanti 1.7 segue

$$\frac{d\pi_j(t)}{dt} = \sum_{i \in S} \sum_{k \in S} p_{ik}(t) q_{kj} \pi_i(0) \quad \forall j \in S$$

ora, sommando rispetto a i, usando la 1.13, si ottiene

$$\frac{d\pi_j(t)}{dt} = \sum_{k \in S} q_{kj} \pi_k(t) \quad \forall j \in S. \tag{1.14}$$

La soluzione esplicita di questo sistema di equazioni differenziali che ci fornisce la soluzione della distribuzione del processo sullo spazio degli stati S ad un istante arbitrario t, è difficile in molti casi. Tuttavia per molte applicazioni pratiche non è necessario ottenere la distribuzione dipendente dal tempo ma è sufficiente trovare, quando esiste, una distribuzione di probabilità limite. Proseguiamo con lo studio di essa iniziando dalle seguenti definizioni.

Definizione 1.9. Una qualsiasi distribuzione di probabilità $\{z_j, j \in S\}$ definita sugli stati di una catena di Markov con tempo discreto è detta *stazionaria* se vale:

$$z_j = \sum_{i \in S} z_i p_{ij}(t) \quad \forall j \in S,\ \forall t \geq 0$$

cioè, una volta che tale distribuzione è raggiunta, essa rimane la distribuzione del processo anche negli istanti successivi. In altre parole se si sceglie come valore iniziale per la nostra distribuzione proprio z_j, cioè poniamo $\pi_j(0) = z_j$, allora la distribuzione dopo ogni intervallo di tempo t rimane $\pi_j(t) = z_j$.

Definizione 1.10. Supponendo che esistono i limiti:

$$\pi_j = \lim_{t \to \infty} \pi_j(t) \quad \forall j \in S$$

allora le probabilità $\{\pi_j, j \in S\}$ si chiamano *probabilità limite* e rappresentano le probabilità di trovare il sistema nello stato j dopo che è trascorso un tempo sufficientemente lungo.

Teorema 1.2.2. *In una catena di Markov omogenea, esistono sempre le probabilità limite*

$$\pi_j = \lim_{t \to \infty} \pi_j(t) \quad \forall j \in S. \tag{1.15}$$

Inoltre se la catena è irriducibile, i limiti sono indipendenti dalla distribuzione di probabilità di stato iniziale $\{\pi_j(0), j \in S\}$. Di più, si verifica che:

a) o tutti gli stati sono transienti, o tutti gli stati sono ricorrenti nulli, nel qual caso $\pi_j = 0$ per tutti i j e non esiste alcuna distribuzione stazionaria;

b) oppure tutti gli stati sono ricorrenti positivi (lo sono sempre se una catena è finita) e quindi $\pi_j > 0$ per ogni j, nel qual caso l'insieme delle probabilità limite $\{\pi_j, j \in S\}$ formano una distribuzione di probabilità stazionaria.

Si vede che gli stati del caso b) del teorema precedente sono quelli più interessanti; proseguiamo con il loro studio al fine di determinare univocamente le quantità π_j. Si nota che quando i limiti 1.15 esistono, per la stazionarietà deve valere

$$\lim_{t\to\infty} \frac{d\pi_j(t)}{dt} = 0 \quad \forall j \in S$$

per cui dalla 1.14 si ottiene che la distribuzione delle probabilità limite è univocamente determinata come soluzione del seguente sistema di equazioni lineari:

$$q_{jj}\pi_j + \sum_{k\neq j} q_{kj}\pi_k = 0 \quad \forall j \in S$$

il quale può essere scritto in forma matriciale come

$$\Pi Q = 0$$

dove abbiamo usato l'ovvia notazione $\Pi = [\pi_0, \pi_1, \pi_2, ...]$.
Il vettore Π si determina in maniera univoca risolvendo il sistema di equazioni lineari

$$\begin{cases} \Pi Q = 0 \\ \sum_{j\in S} \pi_j = 1 \end{cases}$$

Osservazione 1. Dalla 1.13, applicando il limite per $t \to \infty$ si vede che le probabilità di equilibrio di una catena di Markov soddisfano l'ovvia relazione

$$\pi_j = \lim_{t\to\infty} p_{ij}(t) \quad \forall j \in S.$$

Capitolo 2

Processo di nascita e morte con tempo continuo

Una sottoclasse molto importante delle catene di Markov è diventata nota come classe dei *processi di nascita e morte* e la loro caratteristica peculiare è data dal fatto che le transizioni di stato possono avvenire solo tra stati *adiacenti*. Ossia, se assumiamo che lo spazio discreto degli stati sia costituito dagli insiemi dei numeri interi non negativi, una catena di Markov è detta di nascita e morte se dallo stato i è possibile, oltre che rimanere nello stato i stesso, passare solo negli stati $i+1$, oppure $i-1$, e in nessun altro.

Questi processi prendono il nome di catena di nascita e morte perché costituiscono il modello naturale per l'evoluzione del numero di individui di una popolazione. Informalmente, dato un insieme di persone o oggetti (aventi caratteristiche comuni), si dice che si verifica una nascita ogni qualvolta un nuovo membro si aggiunge all'insieme e una morte quando un membro lascia l'insieme. Nel seguito diremo che il sistema si trova nello stato k quando la popolazione consiste di k individui e gli unici cambiamenti nella dimensione di essa possano avvenire al più uno alla volta, ovvero non sono ammesse nascite gemellari o stragi collettive.

Si parlerà di *processo di sole nascite* o in particolare del più famoso *processo di Poisson* se si verificano solamente nascite e non morti, esso può quindi solo accrescere il numero dei suoi membri.

I processi di nascita e morte sono di grande importanza perché costituiscono degli eccellenti modelli per la teoria elementare delle code che verrà studiata nel prossimo capitolo.

Formalizziamo quanto detto come segue.

Definizione 2.1. Una catena di Markov omogenea $(X(t))_{t\geq 0}$ con spazio degli stati S a valori interi non negativi si dice *processo di nascita e morte con tempo continuo* se le probabilità di transizione per intervalli di tempo Δt piccoli

$$p_{kj}(\Delta t) = \mathbf{P}[X(t+\Delta t) = j \mid X(t) = k],$$

(qui interpretata come la probabilità che l'insieme costituito da k elementi al tempo t, sia costituito da j elementi al tempo $t + \Delta t$), che dipendono solo dal valore dello stato al tempo t e non esplicitamente dal tempo t o e non dai valori assunti in istanti precedenti (proprietà di Markov) soddisfano le seguenti condizioni :

$$\begin{aligned}
&a)\ p_{k,k+1}(\Delta t) = \mathbf{P}[\text{una sola nascita in}\,(t, t+\Delta t) \mid \text{popolazione } k \text{ individui}]\\
&\qquad\qquad\quad = \lambda_k \Delta t + o(\Delta t) \quad \text{per } k \geq 0;\\
&b)\ p_{k,k-1}(\Delta t) = \mathbf{P}[\text{una sola morte in}\,(t, t+\Delta t) \mid \text{popolazione } k \text{ individui}]\\
&\qquad\qquad\quad = \mu_k \Delta t + o(\Delta t) \quad \text{per } k \geq 1;\\
&c)\quad p_{kj}(\Delta t) = o(\Delta t) \quad \text{per } \mid k - j \mid > 1.
\end{aligned}$$

dove $\lambda_k \geq 0$, $\mu_k \geq 0$ e $o(\Delta t)$ è un infinitesimo di ordine superiore a Δt .

Le condizioni a) e b) possono essere interpretate nel seguente modo: a meno di infinitesimi di ordine superiore a Δt, la probabilità che nell'intervallo $[t, t + \Delta t]$ si verifichi una nascita non dipende da t, ma è proporzionale all'ampiezza dell'intervallo (ovvero alla durata Δt) con un coefficiente che può dipendere dallo stato attuale, ma non da quelli passati. Analogamente, la probabilità che nell'intervallo $[t, t + \Delta t]$ si verifichi una morte è proporzionale alla durata Δt con un coefficiente che dipende dallo stato attuale, ma non da quelli passati. Il coefficiente λ_k è chiamato **tasso di natalità**

e il coefficiente μ_k è chiamato **tasso di mortalità** e rappresentano quindi rispettivamente la velocità con cui si verificano le nascite e le morti quando la popolazione ha dimensione k. La proprietà c) è spiegata dalla condizione del vicino adiacente.
Dalle tre proprietà a), b), c), poiché deve valere $\sum_{k=0}^{\infty} p_{kj}(\Delta t) = 1$, si ottengono le probabilità che nell'intervallo $[t, t + \Delta t]$ non avvengono transizioni, ovvero:

$$\begin{aligned} p_{kk}(\Delta t) &= \mathbf{P}[\text{nessun cambiamento in}\, t + \Delta t \mid \text{popolazione di } k \text{ individui}] \\ &= 1 - (\lambda_k + \mu_k)\Delta t + o(\Delta t) \\ p_{00}(\Delta t) &= \mathbf{P}[\text{nessun cambiamento in}\, t + \Delta t \mid \text{popolazione di } 0 \text{ individui}] \\ &= 1 - \lambda_0 \Delta t + o(\Delta t) \end{aligned}$$

Osservazione 2. Si può facilmente notare che in termini della nostra nostra definizione precedente i tassi di natalità e mortalità non sono altro che i tassi di transizione, cioè

$$\lambda_k = q_{k,k+1}$$

e

$$\mu_k = q_{k,k-1}$$

(e rappresentano rispettivamente la velocità con cui il processo esce dallo stato k verso gli stati $k + 1$ e $k - 1$) dove la condizione del vicino adiacente richiede che $q_{kj} = 0$ per $|k - j| > 1$. Inoltre poiché abbiamo precedentemente dimostrato che $\sum_j q_{kj} = 0$ si vede che è verificata la definizione $q_{kk} = -(\mu_k + \lambda_k)$. Quindi la matrice infinitesimale per un generico processo omogeneo di nascita e morte assume la forma

$$Q = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 & \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & \cdots \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \lambda_3 & \\ \ddots & \ddots & & & & \end{bmatrix}$$

Q ha chiaramente dimensione infinita se il numero degli stati è infinito.

In sintesi,

$$q_{kj} = \begin{cases} \lambda_k & j = k+1, k \geq 0 \\ \mu_k & j = k-1, k \geq 1 \\ -(\lambda_k + \mu_k) & j = k, k \geq 1 \\ -\lambda_0 & j = k = 0 \\ 0 & \text{altrimenti} \end{cases}$$

Per facilitare la comprensione dei casi particolari che andremo ad esaminare soffermiamoci un momento a descrivere una rappresentazione grafica di un processo di nascita e morte indicata nella figura seguente. In essa gli ovali

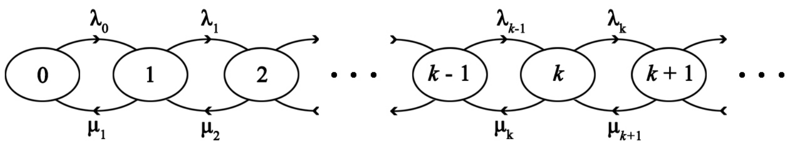

Figura 2.1: *Diagramma di un processo di nascita e morte.*

rappresentano lo stato (ovvero solo la numerosità della popolazione dato che le variabili aleatorie esponenziali sono senza memoria), mentre i coefficienti associati alle frecce esprimono i *tassi di transizione* da uno stato all'altro. Dalla figura è facile verificare che le transizioni tra stati avvengono solo se essi sono strettamente adiacenti; se il diagramma fosse stato usato per descrivere una generica catena di Markov allora gli archi avrebbero connesso tra loro anche stati non adiacenti.

2.0.4 Le equazioni di Kolmogorov

Dato un processo di nascita e morte $(X(t))_{t\geq 0}$, indichiamo con

$$P_k(t) \triangleq \mathbf{P}[X(t) = k]$$

la probabilità che al tempo t la popolazione sia costituita da k individui. Notiamo che in termini delle nostre precedenti notazioni $P_k(t) = \pi_j(t)$. Il valore di $P_k(t)$ può essere calcolato mediante l'uso dell'equazione 1.14 per $\pi_j(t)$ e i nostri particolari valori di q_{kj}. Tuttavia, pur conservando un procedimento che passa per la soluzione di un sistema di equazioni differenziali scegliamo di non utilizzare quel metodo ma bensì di passare per la derivazione di alcune equazioni che in questo caso risulterà molto diretta. Iniziamo applicando il teorema della probabilità totale attraverso il quale si può esprimere la $P_k(t + \Delta t)$ per piccoli intervalli di tempo nel seguente modo:

$$\begin{aligned} P_k(t + \Delta t) &= \mathbf{P}[X(t + \Delta t) = k] \\ &= \sum_{i=0}^{\infty} \mathbf{P}[X(t + \Delta t) = k \mid X(t) = i]\mathbf{P}[X(t) = i] \\ &= \sum_{i=0}^{\infty} P_i(t) p_{ik}(\Delta t) \end{aligned}$$

Quindi per $k \geq 1$ (in quanto non possiamo avere una dimensione di popolazione di -1 individui), dal momento che non ci dobbiamo preoccupare per le transizioni da stati non immediatamente adiacenti allo stato k in quanto abbiamo supposto che esse sono di ordine $o(\Delta t)$ su un intervallo di durata Δt, le possibili transizioni in ingresso allo stato k al tempo $t + \Delta t$ restano le tre già note (eventi esaustivi e mutuamente esclusivi) e si ha:

$$\begin{aligned} P_k(t + \Delta t) &= P_k(t)p_{kk}(\Delta t) + P_{k-1}(t)p_{k-1,k}(\Delta t) + P_{k+1}(t)p_{k+1,k}(\Delta t) + o(\Delta t) \\ &= (1 - \lambda_k \Delta t + \mu_k \Delta t)P_k(t) + \lambda_{k-1}\Delta t P_{k-1}(t) + \\ &\quad + \mu_{k+1}\Delta t P_{k+1}(t) + o(\Delta t) \end{aligned}$$

da cui portando $P_k(t)$ al primo membro e dividendo tutto per Δt, si ottiene

$$\begin{aligned} \frac{P_k(t + \Delta t) - P_k(t)}{\Delta t} &= -(\lambda_k + \mu_k)P_k(t) + \lambda_{k-1}P_{k-1}(t) + \\ &\quad + \mu_{k+1}P_{k+1}(t) + \frac{o(\Delta t)}{\Delta t}. \end{aligned} \tag{2.1}$$

Analogamente per $k = 0$, sfruttando l'ipotesi che non sia possibile avere un evento di morte quando la popolazione ha dimensione 0 ($\mu_0 = 0$), mentre si può tranquillamente avere un evento di nascita ($\lambda_0 \geq 0$), si ha:

$$\begin{aligned} P_0(t+\Delta t) &= P_0(t)p_{00}(\Delta t) + P_1(t)p_{10}(\Delta t) + o(\Delta t) \\ &= (1-\lambda_0\Delta t)P_0(t) + \mu_1 \Delta t P_1(t) + o(\Delta t) \end{aligned}$$

da cui

$$\frac{P_0(t+\Delta t) - P_0(t)}{\Delta t} = \mu_1 P_1(t) - \lambda_0 P_0(t) + \frac{o(\Delta t)}{\Delta t} \tag{2.2}$$

Passando al limite per $\Delta t \to 0$ nella 2.1 e 2.2 si ha

$$\begin{cases} \dfrac{dP_k(t)}{dt} = -(\lambda_k + \mu_k)P_k(t) + \lambda_{k-1}P_{k-1}(t) + \mu_{k+1}P_{k+1}(t) & \text{k} \geq 1 \\ \dfrac{dP_0(t)}{dt} = -\lambda_0 P_0(t) + \mu_1 P_1(t) & \text{k} = 0 \end{cases} \tag{2.3}$$

che è un insieme di equazioni differenziali. Le equazioni 2.3 prendono il nome di **equazioni di Kolmogorov** e descrivono come le probabilità dei diversi stati evolvono nel tempo per un processo di nascita e morte.
Ci resta solo da calcolarne la soluzione che ci fornisce i valori delle $P_k(t)$. Ora, notando che le equazioni presenti nel sistema sono un numero infinito si può comprendere le difficoltà nella loro soluzione. Per risolvere il sistema si potranno specificare le condizioni iniziali date dalle probabilità $P_k(0)$, $k = 0, 1, ...$, tenendo conto che, ovviamente, deve risultare $\sum_{k=1}^{\infty} P_k(0) = 1$. In particolare, se lo stato iniziale è noto con certezza ed è pari a $k > 0$, ovvero $X(0) = k$, si ha $P_k(0) = 1$ e $P_j(0) = 0$ per ogni $j \neq k$. Assegnate le condizioni iniziali è possibile trovare la soluzione; tuttavia essendo ancora alquanto laborioso, ci si limiterà a considerare solo alcuni casi particolari che ci forniranno il comportamento in transitorio di questi processi di nascita e morte.

2.0.5 Distribuzione stazionaria dei processi di nascita e morte

Anche per i processi di nascita e morte abbiamo l'interesse di studiare il comportamento all'equilibrio oltre al comportamento in transitorio; ci chiediamo quindi se le probabilità di stato $P_k(t)$ si stabilizzano quando t diventa sufficientemente grande. Per un processo di nascita e morte, si dice che esso ammette una *distribuzione stazionaria* (o di equilibrio) se esistono i limiti

$$\lim_{t\to\infty} P_k(t) \triangleq p_k \qquad k = 0, 1, ...$$

indipendentemente dai valori iniziali $P_k(0)$, $k = 0, 1, ...$, essendo naturalmente $\sum_{k=0}^{\infty} p_k = 1$. La notazione p_k indica dunque la probabilità limite che il sistema contenga k membri (oppure in modo equivalente, sia nello stato k) in un istante arbitrariamente lontano nel tempo.
Rimandiamo per il momento lo studio dell'esistenza di queste probabilità limite p_k e ammesso che esistano troviamone la soluzione generica che può essere ottenuta in due modi:

1. risolvendo il sistema di equazioni differenziali 2.3 con opportune condizioni iniziali per ottenere $P_k(t)$ e poi calcolando i limiti $\lim_{t\to\infty} P_k(t) = p_k$;

2. prendendo il limite per $t \to \infty$ ad ambo i membri di ciascuna delle equazioni differenziali 2.3 (sapendo che $\lim_{t\to\infty} P_k(t) = p_k$) e imponendo poi che, per la stazionarietà valga $\lim_{t\to\infty} \frac{dP_k(t)}{dt} = 0$, si ottiene così un sistema di equazioni algebriche sfruttando anche la nota condizione di normalizzazione $\sum_{k=0}^{\infty} p_k = 1$.

Il secondo modo di procedere è più conveniente perché permette di ottenere la distribuzione stazionaria direttamente senza dover prima determinare le probabilità $P_k(t)$ dipendenti dal tempo.
Si ha quindi il sistema di equazioni algebriche

$$0 = -(\lambda_k + \mu_k)p_k + \lambda_{k-1}p_{k-1} + \mu_{k+1}p_{k+1} \quad k \geq 1 \tag{2.4}$$

$$0 = -\lambda_0 p_0 + \mu_1 p_1 \qquad k = 0 \tag{2.5}$$

ottenuto dalle 2.3 passando al limite per $t \to \infty$ in entrambi i membri di ciascuna equazione.
Si noti che queste stesse equazioni possono essere ottenute direttamente dal

diagramma dei tassi di transizione uguagliando il flusso entrante ed uscente in ogni singolo stato in quanto ad interessarci è il caso di equilibrio.
Per esempio, se riguardiamo la figura 2.1 e consideriamo lo stato k in equilibrio, si vede che il flusso entrante nello stato k viene dallo stato $k-1$ con tasso λ_{k-1} e dallo stato $k+1$ con tasso μ_{k+1} ottenendo che la quantità $\lambda_{k-1}p_{k-1}+\mu_{k+1}p_{k+1}$ può essere vista come il "tasso del flusso entrante" allo stato k ed analogamente la quantità $(\lambda_k+\mu_k)p_k$ come "tasso di flusso uscente".
Da questa interpretazione, ciascuna equazione del sistema rappresenta, per ogni stato, una *equazione di bilancio* che per $k \geq 1$ ci consente di riscrivere la 2.4 come:

$$\lambda_{k-1}p_{k-1}+\mu_{k+1}p_{k+1}=(\lambda_k+\mu_k)p_k.$$

In particolare considerando il caso $k=0$, dall'equazione 2.5 si ha:

$$p_1=\frac{\lambda_0}{\mu_1}p_0.$$

Nel caso $k=1$, essendo $-\lambda_0 p_0+\mu_1 p_1=0$, la 2.4 diventa :

$$\mu_2 p_2-\lambda_1 p_1=0 \qquad \Rightarrow \qquad p_2=\frac{\lambda_1}{\mu_2}p_1=\frac{\lambda_1\lambda_0}{\mu_2\mu_1}p_0.$$

In generale, procedendo iterativamente per $k \geq 1$ abbiamo

$$-\lambda_{k-1}p_{k-1}+\mu_k p_k=0$$

da cui

$$p_k=\frac{\lambda_{k-1}}{\mu_k}p_{k-1}=\frac{\lambda_{k-1}}{\mu_k}\left(\frac{\lambda_{k-2}}{\mu_{k-1}}p_{k-2}\right)=\frac{\lambda_{k-1}\lambda_{k-2}\cdots\lambda_1\lambda_0}{\mu_k\mu_{k-1}\cdots\mu_2\mu_1}p_0$$

ovvero, per un generico stato $k \geq 1$ si ha

$$p_k=\frac{\prod_{i=0}^{k-1}\lambda_i}{\prod_{j=1}^{k}\mu_j}p_0 \tag{2.6}$$

che è l'espressione che fornisce p_k in funzione di p_0. Inoltre poiché deve risultare $\sum_{k=1}^{\infty}p_k+p_0=1$, si può ricavare p_0 come segue

$$p_0=1-\sum_{k=1}^{\infty}p_k=1-\sum_{k=1}^{\infty}\frac{\prod_{i=0}^{k-1}\lambda_i}{\prod_{j=1}^{k}\mu_j}p_0$$

da cui

$$p_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \frac{\prod_{i=0}^{k-1} \lambda_i}{\prod_{j=1}^{k} \mu_j}}. \tag{2.7}$$

La 2.6 e la 2.7 ci forniscono la soluzione cercata.
Queste equazioni ci mostrano anche la condizione per l'esistenza delle probabilità limite, ovvero esse esistono se la serie al denominatore della 2.7 è convergente, cioè

$$\sum_{k=1}^{\infty} \frac{\prod_{i=0}^{k-1} \lambda_i}{\prod_{j=1}^{k} \mu_j} < \infty. \tag{2.8}$$

In questo caso si può anche dimostrare che gli stati sono ricorrenti positivi e pertanto esiste una distribuzione di probabilità stazionaria p_k per tutti i $k \in S$. Si noti che la condizione di convergenza è verificata per tutti i casi in cui si restringe lo spazio ad un numero finito di stati, perché da quel punto in poi la serie sarà una somma di zeri e quindi converge.

2.0.6 Il processo di Poisson

Il più semplice processo di nascita e morte $((X(t))_{t\geq 0}$ è il *processo di sole nascite* in cui si assume $\mu_k = 0$ per ogni k.
Se oltre a questa condizione per semplificare il problema si assume che $\lambda_k = \lambda$ per ogni $k = 0, 1, 2, ...$, allora questo processo è noto come *processo di Poisson* che denoteremo con $N(t)$.

In questo caso è possibile esplicitare l'insieme di equazioni differenziali 2.3 particolarizzate alle restrizioni appena fatte, che si riduce a:

$$\begin{cases} \dfrac{dP_k(t)}{dt} = -\lambda P_k(t) + \lambda P_{k-1}(t) & \text{k} \geq 1 \\ \dfrac{dP_0(t)}{dt} = -\lambda P_0(t) & \text{k} = 0 \end{cases} \tag{2.9}$$

dove siano $P_k(0)$ per $k = 0, 1, 2, ...$ sono le condizioni iniziali.
Passiamo alla sua risoluzione utilizzando la funzione generatrice di $P_k(t)$ definita come

$$P(z,t) = \sum_{k=0}^{\infty} P_k(t) z^k.$$

Ora, applicando la proprietà $P(z,t+\Delta t) = P(z,t)P(z,\Delta t)$ si ha:

$$\begin{aligned} P(z,t+\Delta t) &= P(z,t)\{\mathbf{P}[N(\Delta t)=0] + \mathbf{P}[N(\Delta t)=1]z + \mathbf{P}[N(\Delta t)>1]\} = \\ &= P(z,t)[(1-\lambda\Delta t + o(\Delta t)) + (\lambda\Delta t + o(\Delta t))z + o(\Delta t)] \\ &= P(z,t)[1-\lambda\Delta t + \lambda\Delta t z + o(\Delta t)]. \end{aligned}$$

Procediamo derivando rispetto a t nel modo che segue:

$$\begin{aligned} \frac{P(z,t+\Delta t)-P(z,t)}{\Delta t} &= \frac{P(z,t)[1-\lambda\Delta t+\lambda\Delta t z+o(\Delta t)]-P(z,t)}{\Delta t} \\ &= P(z,t)[-\frac{\lambda\Delta t}{\Delta t}(1-z)] + \frac{o(\Delta t)}{\Delta t} \end{aligned}$$

e mandando Δt a zero ho la derivata

$$\frac{\partial P(z,t)}{\partial t} = -\lambda(1-z)P(z,t).$$

Sapendo inoltre che $\frac{P'(z,t)}{P(z,t)} = \frac{d}{dt}log[P(z,t)]$ si ha

$$\frac{\partial \log P(z,t)}{\partial t} = -\lambda(1-z).$$

Proseguiamo la risoluzione assumendo per semplicità che il sistema parta al tempo zero dallo stato con popolazione nulla, cioè

$$P_k(0) = \begin{cases} 1 & \text{k} = 0 \\ 0 & k \neq 0 \end{cases}$$

e per la proprietà

$$\log[P(z,t)] = \log[P(z,0)] + \left[\frac{P'(z,t)}{P(z,t)}\right]t$$

si ha così $\log P(z,t) = \log P(z,0) - \lambda(1-z)t$
ed essendo

$$P(z,0) = \sum_{k=0}^{\infty} \mathbf{P}[N(0)=k]z^k = \mathbf{P}[N(0)=0]z^0 = 1 \Rightarrow \log P(z,0) = 0$$

si ottiene $\log P(z,t) = -\lambda(1-z)t$,
ossia

$$\begin{aligned} P(z,t) &= e^{-\lambda(1-z)t} \\ &= e^{-\lambda t} e^{\lambda z t} \\ &= \sum_{k=0}^{\infty} \left(\frac{(\lambda t)^k}{k!} e^{-\lambda t} \right) z^k \\ &= \sum_{k=0}^{\infty} \mathbf{P}[N(t)=k] z^k \end{aligned}$$

Allora

$$P_k(t) = \mathbf{P}[N(t)=k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{2.10}$$

che altro non è che una distribuzione di Poisson con il valore atteso del numero di arrivi nell'intervallo $(0,t)$ pari a λt, da cui il nome processo di Poisson.

Osservazione 3. Abbiamo visto per quali valori le $P_k(t)$ soddisfano il sistema 2.3 e sappiamo che utilizzando il teorema della probabilità totale vale la relazione $P_k(t) = \sum_{i \in S} P_i(0) p_{ik}(t) \quad \forall k \in S,\ \forall t \geq 0$.
Ora, in particolare, se lo stato iniziale è noto con certezza ed è pari ad i, ovvero $P_i(0) = 1$ e $P_k(0) = 0\ \forall k \neq i$ si possono conoscere le probabilità di transizione a partire da i stesso per un processo di Poisson.
Infatti:

a. se $i = 0$ allora risolvendo il sistema otteniamo $p_{0k}(t) = \dfrac{(\lambda t)^k}{k!} e^{-\lambda t}$

b. se lo stato i di partenza del sistema è arbitrario, allora le probabilità di transizione sono:

$$\begin{cases} p_{ik}(t) = 0 & \text{per } k < i \\ p_{ik}(t) = \dfrac{(\lambda t)^{k-i}}{(k-i)!} e^{-\lambda t} & \text{per } k \geq i \end{cases}$$

Osservazione 4. Ovviamente dato che le morti non sono possibili, il processo di Poisson può solo accrescere il numero dei suoi membri; è quindi impossibile ritornare in uno stato dopo averlo lasciato (tutti gli stati sono transitori e non esiste una distribuzione stazionaria).

In generale in un processo di sole nascite il rischio è quello dell'*esplosione*: se i λ_k crescono velocemente allora il sistema in un tempo finito si trova a fare un'infinità di passaggi fino a non essere più in nessuno stato, cioè dopo un numero finito di istanti si può avere $N(t) = \infty$. Se accadesse ciò potrebbe succedere che $\sum_{k \in S} P_k(t) < 1$, ovvero ci sarebbe una probabilità positiva di uscire dal sistema. Quindi si può dimostrare che affinché non si verifichi un'esplosione è necessario e sufficiente che la serie $\sum_k \frac{1}{\lambda_k}$ sia divergente e in particolare si ha che un processo di Poisson non esplode se e solo se $\frac{1}{\lambda} = \infty$.

Capitolo 3

Teoria delle code

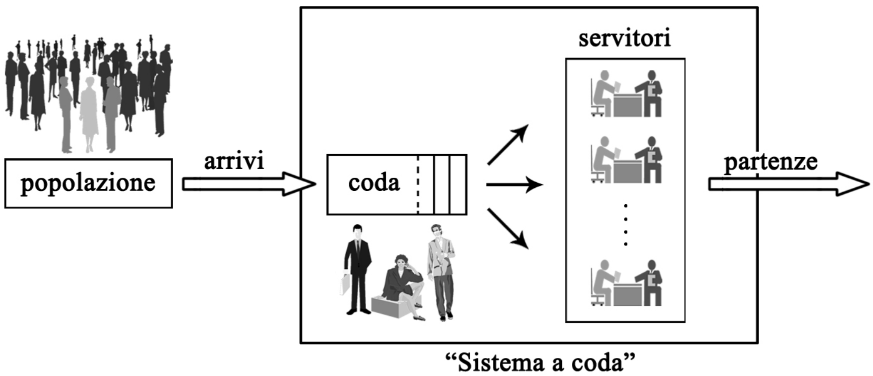

Figura 3.1: *Rappresentazione schematica di un sistema a coda.*

3.1 Sistemi a coda e loro componenti

Un *sistema a coda*, o semplicemente coda, da un punto di vista fisico può essere descritto da un insieme non vuoto di stazioni di servizio e da un flusso di clienti, provenienti da un insieme chiamato popolazione, che arrivano alle stazioni per essere serviti. La popolazione, come il numero dei servitori, può essere finita o infinita. Si assume che tutti i servitori hanno caratteristiche identiche (possono quindi lavorare in parallelo) e si suppone inoltre che se all'arrivo dei clienti ci sono servitori inattivi questi ultimi non possono rifiutarsi di prestare servizio mentre se tutte le stazioni di servizio sono occupate allora i clienti aspettano il loro turno disponendosi in modo ordinato e formando così una coda unica. Questi ultimi verranno poi chiamati per usufruire del servizio secondo certe regole precise conosciute come discipline di servizio. Le discipline di servizio usualmente considerate, poiché sono molto comuni nella realtà e matematicamente trattabili sono:

- servizio in *ordine di arrivo*
  FCFS (first-come first-served) o FIFO (first-in first-out)
- servizio in *ordine inverso di arrivo*
  LCFS (last-come first-served) o LIFO (last-in first-out)
- servizio in *ordine casuale*
  SIRO (service in random order)
- servizio basato su *classi di priorità*
  (vedi centri di emergenza quali pronto soccorso)

Un altro elemento che caratterizza un sistema a coda è la capacità di sistema, cioè il numero dei clienti in attesa di essere serviti più quelli che contemporaneamente sono serviti.

Dal punto di vista dinamico, invece, una coda è costituita essenzialmente da due processi stocastici: il *processo d'arrivo* dei clienti e il *processo di servizio.*

Il processo d'arrivo descrive il modo secondo cui i clienti si presentano. In genarale è un processo stocastico ed è definito in termini della distribuzione del tempo di interarrivo, cioè dell'intervallo di tempo che intercorre tra l'arrivo di due clienti successivi.

Il processo dei servizi, invece, descrive il modo secondo cui i servitori distribuiscono il servizio. Anche questo di solito è un processo stocastico ed è definito in relazione alla distribuzione di probabilità dei tempi di servizio dei diversi servitori, cioè la quantità di tempo spesa da un cliente all'interno della stazione di servizio.

Tutti gli elementi che definiscono una coda sono evidenziati nella notazione A/B/n/K/m/Z detta di Kendall, dove le lettere rispettivamente indicano:

- A: la distribuzione degli intertempi d'arrivo;
- B: la distribuzione dei tempi di servizio;
- n: il numero dei servitori;
- K: la capacità del sistema (default: infinita);
- m: la dimensione della popolazione (default: infinita);
- Z: la disciplina di servizio (default: FCFS);

In particolare A e B possono essere sostituite dalle seguenti sigle:

M : Markovian (distribuzione esponenziale)

D : Deterministic (distribuzione costante)

N : Normal (distribuzione normale, Gaussiana)

$\mathbf{E_k}$: Erlangian (distribuzione di Erlang di ordine k)

G : General (distribuzione generica)

Si userà la notazione troncata A/B/n per descrivere un processo in cui i valori K e m si suppongono uguali ad infinito e la disciplina di servizio sia del tipo FCFS.
Al di là di questo, i sistemi a coda possono essere soggetti a comportamenti anomali da parte degli utenti nella forma di: abbandoni della coda, passaggio

a code diverse, rinunce prima di accodarsi, corruzioni per ottenere posizioni migliori, inganni vari e sorpassi, e altri interessanti comportamenti umani.

3.2 Code markoviane M/M/n

Siamo adesso pronti per affrontare lo studio completo dei particolari sistemi a coda M/M/n, particolarmente semplici ed analiticamente trattabili. Questi sono anche chiamati code Markoviane e come è noto dalla simbologia si tratta di sistemi in cui il processo degli arrivi è un processo di Poisson e la distribuzione dei tempi di servizio è esponenziale ed infine n indica il numero delle stazioni di servizio che può variare da uno a infinito.
Si suppone che i tempi di interarrivo e i tempi di servizio siano variabili casuali indipendenti ed identicamente distribuite e che i tempi di servizio siano indipendenti dal processo degli arrivi Poissoniano.
Tali assunzioni in certi ambiti possono essere molto limitative, infatti ad esempio il processo di arrivo dei clienti ad una banca varia durante le ore della giornata ed un cliente può essere servito solo se già arrivato. In altre parole il processo d'arrivo è indipendente ma condiziona il processo dei servizi; un servitore non può servire in anticipo clienti non ancora arrivati (non può esistere una coda negativa).

Si nota inoltre che questo tipo di sistema a coda, in quanto descritto da due processi con distribuzione dei tempi di tipo esponenziale e quindi privi di memoria, corrisponde ad una catena di Markov con tempo continuo con spazio degli stati a valori interi non negativi. Le catene di Markov saranno così usate come modelli per descrivere l'evoluzione dei sistemi M/M/n e anche (potenzialmente) la distribuzione di probabilità a regime dello stato (cioè la distribuzione di probabilità del numero di clienti all'equilibrio) o le probabilità di transizione tra stati.

3.2.1 M/M/1

Il sistema a coda M/M/1, composto da un solo servitore, è il più semplice caso particolare dei precedenti sistemi M/M/n. Si suppone sempre, come già detto, che esso è una coda in cui si sceglie che gli arrivi dei clienti seguono un processo di Poisson di parametro λ e che la distribuzione dei tempi di servizio sia esponenziale di parametro μ ed indipendente dal processo degli arrivi.
Vogliamo utilizzare i forti strumenti delle catene di Markov per affrontare lo studio di questo sistema a coda. Prendendo come definizione di stato il numero dei clienti presenti nel sistema, il modello risultante si vede facilmente che è una catena di nascita e morte con tempo continuo e spazio degli stati $S = \mathbb{Z}^+ \cup \{0\}$ con i seguenti coefficienti di nascita e morte costanti

$$\lambda_k = \lambda \qquad\qquad k = 0, 1, 2, ...$$
$$\mu_k = \mu \qquad\qquad k = 1, 2, 3, ...$$

e con la nota interpretazione di considerare l'arrivo di un cliente in coda come una nascita e l'uscita di un cliente dal sistema come una morte.
Mostriamo per una maggiore chiarezza una rappresentazione grafica di questa coda attraverso il diagramma dei tassi di transizione tra gli stati nella seguente figura.

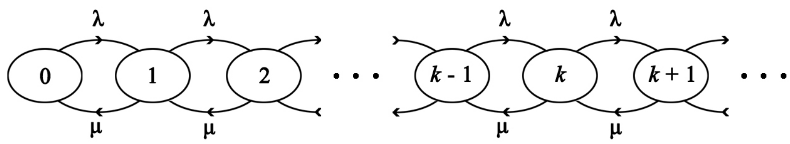

Figura 3.2: *Diagramma delle frequenze di transizione di stato per la coda M/M/1.*

Gli elementi del generatore infinitesimale per una catena di nascita e morte di questo tipo possono quindi essere scritti come

$$q_{ij} = \begin{cases} \lambda & j = i+1, i \geq 0 \\ \mu & j = i-1, i \geq 1 \\ -(\lambda+\mu) & j = i, i \geq 1 \\ -\lambda & j = i = 0 \\ 0 & \text{altrimenti} \end{cases}$$

Sia ora $(X(t))_{t\geq 0}$ il processo di nascita e morte appena descritto che rappresenta il numero di clienti presenti nel sistema M/M/1. Vogliamo, usufruendo dei risultati del capitolo 2, descrivere il comportamento in transitorio e all'equilibrio di questo sistema.
Procediamo a tal fine il nostro studio cercando la soluzione delle equazioni di Kolmogorov 2.3 che ci forniscono i valori delle $P_k(t)$, cioè delle probabilità che al tempo t ci siano k clienti nel sistema.
In questo caso le equazioni differenziali si riducono al seguente sistema

$$\begin{cases} \dfrac{dP_k(t)}{dt} = -(\lambda+\mu)P_k(t) + \lambda P_{k-1}(t) + \mu P_{k+1}(t) & \text{k} \geq 1 \\ \dfrac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t) & \text{k} = 0 \end{cases} \tag{3.1}$$

dove $P_k(0)$ per $k = 0, 1, 2, ...$ sono le condizioni iniziali.
Per trovare la soluzione facciamo uso della funzione generatrice per la variabile discreta k e della trasformata di Laplace per la variabile continua t.

Definiamo la funzione generatrice di $P_k(t)$ dipendente dal tempo come

$$P(z,t) \triangleq \sum_{k=0}^{\infty} P_k(t) z^k \tag{3.2}$$

la quale, poiché $\sum_{k=0}^{\infty} P_k(t) = 1$, è sicuramente analitica nel disco unitario $|z| \leq 1$.
Il metodo scelto procede moltiplicando la k-esima equazione differenziale di 3.1 per z^k e sommando poi su tutti i k consentiti ($k \geq 1$) per ottenere la

seguente equazione differenziale

$$\sum_{k=1}^{\infty} \frac{dP_k(t)}{dt} z^k = -(\lambda+\mu) \sum_{k=1}^{\infty} P_k(t) z^k + \lambda \sum_{k=1}^{\infty} P_{k-1}(t) z^k + \mu \sum_{k=1}^{\infty} P_{k+1}(t) z^k.$$

Ora, se $F(z)$ è la funzione generatrice di f_n ed a è un parametro di f_n, applicando la proprietà

$$\frac{\partial}{\partial a} f_n \Longleftrightarrow \frac{\partial}{\partial a} F(z)$$

alla nostra funzione generatrice $P(z,t)$, possiamo, nel primo membro dell'ultima equazione scritta, spostare l'operatore di derivazione al di fuori della sommatoria come segue:

$$\frac{\partial}{\partial t}\left(\sum_{k=1}^{\infty} P_k(t) z^k\right) = -(\lambda+\mu) \sum_{k=1}^{\infty} P_k(t) z^k + \lambda \sum_{k=1}^{\infty} P_{k-1}(t) z^k + \mu \sum_{k=1}^{\infty} P_{k+1}(t) z^k.$$

Ora, con lo scopo di riscrivere questa equazione in termini della funzione generatrice $P(z,t)$ occorre togliere alla stessa $P(z,t)$ il termine $k=0$ nella sommatoria del membro di sinistra e nella prima sommatoria del membro di destra. La seconda sommatoria del membro di destra non ha termini in più mentre nell'ultimo si sottrae a $P(z,t)$ sia i primi due termini che un fattore moltiplicativo z. Si ottiene così:

$$\begin{aligned}
&\frac{\partial}{\partial t}[P(z,t) - P_0(t)] \\
&\quad = -(\lambda+\mu)[P(z,t) - P_0(t)] + \lambda z P(z,t) + \frac{\mu}{z}[P(z,t) - P_0(t) - P_1(t) z].
\end{aligned}$$

Possiamo applicare l'equazione al contorno ($k=0$) del sistema 3.1, mai utilizzata fino ad ora, all'equazione precedente per eliminare alcuni suoi termini come segue:

$$\frac{\partial}{\partial t} P(z,t) = -\lambda P(z,t) - \mu[P(z,t) - P_0(t)] + \lambda z P(z,t) + \frac{\mu}{z}[P(z,t) - P_0(t)].$$

Moltiplicando ambo i membri per z e riordinando i termini si ha

$$\begin{aligned}
&z\frac{\partial}{\partial t} P(z,t) \\
&\quad = -\lambda z P(z,t) - \mu z P(z,t) + \mu z P_0(t) + \lambda z^2 P(z,t) + \mu P(z,t) - \mu P_0(t) \\
&\quad = (1-z)[\mu P(z,t) - \lambda z P(z,t) - \mu P_0(t)] \\
&\quad = (1-z)[(\mu - \lambda z) P(z,t) - \mu P_0(t)] \qquad (3.3)
\end{aligned}$$

che è un'equazione differenziale (alle derivate parziali) lineare di primo ordine in $P(z,t)$. Questa può essere ancora trasformata, definiamo pertanto la trasformata di Laplace della funzione $P(z,t)$ come

$$P^*(z,s) \triangleq \int_0^\infty e^{-st} P(z,t)\, dt$$

e la trasformata di Laplace di $P_0(t)$ come

$$P_0^*(s) \triangleq \int_0^\infty e^{-st} P_0(t)\, dt.$$

Se $F^*(s)$ è la trasformata di Laplace di $f(t)$, sfruttando le proprietà

$$\frac{df(t)}{dt} \Longleftrightarrow sF^*(s) - f(0)$$

e

$$af(t) \Longleftrightarrow aF^*(s)$$

e quest'ultime definizioni, la 3.3 diventa

$$z[sP^*(z,s) - P(z,0)] = (1-z)[(\mu - \lambda z)P^*(z,s) - \mu P_0^*(s)]$$

che è un'equazione algebrica della trasformata di Laplace doppia $P^*(z,s)$ (se consideriamo la funzione generatrice come una funzione di Laplace discreta). Esplicitiamo mediante alcuni semplici calcoli quest'ultima equazione rispetto a $P^*(z,s)$ come segue

$$\begin{aligned} zsP^*(z,s) - zP(z,0) =& \mu P^*(z,s) - \lambda z P^*(z,s) - \mu P_0^*(s) \\ & - \mu z P^*(z,s) + \lambda z^2 P^*(z,s) + \mu z P_0^*(s) \end{aligned}$$

$$\begin{aligned} P^*(z,s)[zs - \mu + \lambda z + \mu z - \lambda z^2] &= zP(z,0) - \mu(1-z)P_0^*(s) \\ P^*(z,s) &= \frac{zP(z,0) - \mu(1-z)P_0^*(s)}{zs - (1-z)(\mu - \lambda z)} \end{aligned} \tag{3.4}$$

Dalla definizione della funzione generatrice $P(z,t)$ si ha che

$$P(z,0) = \sum_{k=0}^{\infty} P_k(0) z^k. \tag{3.5}$$

dove sappiamo che $P_k(0)$ indica la nostra condizione iniziale. Assumiamo ora che il sistema non sia più vuoto al tempo 0 ma ci siano i utenti, ossia

$$P_k(0) = \begin{cases} 1 & k = i \\ 0 & k \neq i \end{cases}$$

dove, se $i = 0$ si ritrova la condizione iniziale semplificata. Sostituendo la nuova condizione iniziale nella 3.5 e sapendo che $\sum_{k=0}^{\infty} P_k(0) = 1$ si ha la relazione:

$$P(z, 0) = z^i$$

che sostituita nella 3.4 permette di scrivere

$$P^*(z, s) = \frac{z^{i+1} - \mu(1-z)P_0^*(s)}{sz - (1-z)(\mu - \lambda z)} \triangleq \frac{N_s(z)}{D_s(z)}. \tag{3.6}$$

Rimane da determinare la funzione incognita $P_0^*(s)$ per avere un'espressione esplicita della nostra trasformata doppia $P^*(z, s)$. A tal fine enunciamo ed applichiamo il teorema di Rouché.

Teorema 3.2.1. ***Teorema di Rouché***
Siano f(z) e g(z) due funzioni di una variabile complessa z, analitiche dentro e lungo una linea chiusa C; inoltre per ogni punto di C sia soddisfatta la condizione

$$|g(z)| < |f(z)|.$$

Allora le funzioni h(z)=g(z)+f(z) e f(z) hanno lo stesso numero di zeri all'interno di C, dove ogni zero deve essere contato con la sua molteplicità.

Dimostrazione. Ci serviremo del seguente *teorema dell'argomento* per dimostrare il teorema di Rouché.

Teorema 3.2.2. *Sia $f(z)$ una funzione della variabile complessa z, analitica dentro e lungo una curva semplice chiusa C, fatta eccezione per un numero finito di punti singolari z_k giacenti dentro C. Supponiamo che tutti gli z_k siano dei poli e che la funzione $f(z)$ sul contorno C non ha né zeri né punti singolari.*

Allora la differenza tra il numero totale degli zero N_f ed il numero totale dei poli P_f della funzione $f(z)$ interni a C, contati con le loro molteplicità, è definita dall'espressione:

$$N_f - P_f = \frac{1}{2\pi i}\int_C \frac{f'(z)}{f(z)}dz. \tag{3.7}$$

Questo teorema si dimostra calcolando l'integrale al secondo membro tramite il teorema dei residui ed osservando che se $f(z)$ ha uno zero di ordine m in un punto z_0, allora $f'(z)/f(z)$ ha un polo semplice in z_0 e

$$\mathrm{Res}\Big(\frac{f'}{f}, z_0\Big) = m;$$

mentre se $f(z)$ ha un polo di ordine m in un punto z_0, allora $f'(z)/f(z)$ ha ancora un polo semplice in z_0 e

$$\mathrm{Res}\Big(\frac{f'}{f}, z_0\Big) = -m.$$

Dimostriamo il primo caso: se $f(z)$ ha uno zero di ordine m in z_0, scriviamo

$$f(z) = (z - z_0)^m g(z)$$

dove $g(z_0) \neq 0$ e $g(z)$ è analitica dentro e lungo C. Allora, per $z \neq z_0$ vale

$$\frac{f'(z)}{f(z)} = \frac{m}{z - z_0} + \frac{g'(z)}{g(z)}.$$

Poiché $g(z_0) \neq 0$, allora $1/g(z)$ è analitica in un intorno di z_0 e per definizione si ha che il residuo di f'/f in z_0 è

$$\mathrm{Res}\Big(\frac{f'}{f}, z_0\Big) = m.$$

La dimostrazione del secondo caso è analoga.
Questo teorema prende il nome di teorema dell'argomento perché, se cambiamo la variabile di integrazione,

$$\frac{1}{2\pi i}\int_C \frac{f'(z)}{f(z)}dz = \frac{1}{2\pi i}\int_{C_f} \frac{1}{\zeta}d\zeta \tag{3.8}$$

dove C_f è l'immagine di C mediante f,

$$C_f : \quad \zeta = f(z(t)), \quad t \in [a,b]$$

con $[a,b]$ intervallo di parametrizzazione; l'ultimo integrale per definizione si scrive come

$$\frac{1}{2\pi i}\int_{C_f} \frac{1}{\zeta} d\zeta = \mathrm{Ind}_{C_f}(0)$$

dove $\mathrm{Ind}_{C_f}(0)$ è l'indice di avvolgimento della curva C_f rispetto all'origine, ossia, il numero di giri che un punto mobile sulla curva C_f compie intorno all'origine. In modo intuitivo si ha così che la scrittura 3.7 indica la variazione dell'argomento di ζ quando ζ percorre C_f.

Passiamo alla dimostrazione del teorema di Rouché. Innanzitutto $h = f + g$ è analitica perché è la somma di due funzioni analitiche ed inoltre siccome h non può avere zeri sul contorno C è quindi applicabile ad essa il teorema dell'argomento. Di più, essendo essa analitica su C e al suo interno, si ha che P_h, cioè il numero dei poli di h all'interno di C, è uguale a zero. Applicando il teorema dell'argomento alla funzione h, si ha così

$$N_h = \frac{1}{2\pi i}\int_{C} \frac{h'(z)}{h(z)} dz = \mathrm{Ind}_{C_h}(0).$$

Ora, sostituendo all'interno di quest'ultimo integrale l'espressione

$$\frac{h'(z)}{h(z)} = D(\log(h(z))$$

dove con il simbolo D si intende la derivata complessa; e poiché $h = f + g =$

$f[1+\frac{g}{f}]$, si scrive:

$$
\begin{aligned}
N_h &= \frac{1}{2\pi i}\int_C D(\log(h(z))dz \\
&= \frac{1}{2\pi i}\int_C D(\log(f(z)+g(z)))dz \\
&= \frac{1}{2\pi i}\int_C D\Bigg(\log\bigg(f(z)\Big(1+\frac{g(z)}{f(z)}\Big)\bigg)\Bigg)dz \\
&= \frac{1}{2\pi i}\int_C D\Bigg(\log f(z)+\log\bigg(1+\frac{g(z)}{f(z)}\bigg)\Bigg)dz \\
&= \frac{1}{2\pi i}\int_C D(\log f(z))+D\Bigg(\log\bigg(1+\frac{g(z)}{f(z)}\bigg)\Bigg)dz \\
&= \frac{1}{2\pi i}\int_C D(\log f(z))dz+\frac{1}{2\pi i}\int_C D\Bigg(\log\bigg(1+\frac{g(z)}{f(z)}\bigg)\Bigg)dz \\
&= \frac{1}{2\pi i}\int_C \frac{f'(z)}{f(z)}dz+\frac{1}{2\pi i}\int_C \frac{D(1+\frac{g(z)}{f(z))}}{1+\frac{g(z)}{f(z)}}dz \\
&= \mathrm{Ind}_{C_f}(0)+\mathrm{Ind}_{C_{1+\frac{g(z)}{f(z)}}}(0)
\end{aligned}
$$

Per completare la dimostrazione occorre mostrare che $\mathrm{Ind}_{C_{1+\frac{g(z)}{f(z)}}}(0) = 0$. Infatti, la condizione del teorema $|g| < |f|$ mostra che

$$\left|\frac{g(z)}{f(z)}\right| < 1,$$

ossia che i punti $w = 1+\frac{g(z)}{f(z)}$ hanno parte reale positiva.
Si vede anche che $|w-1| < 1$ e da ciò segue che la curva parametrizzata dalla funzione $1+\frac{g(z)}{f(z)}$ non gira attorno all'origine ma è contenuta nel cerchio di centro A(1,0) e raggio unitario, e quindi, percorrendola, l'argomento è certamente contenuto tra $-\frac{\pi}{2}$ e $+\frac{\pi}{2}$ e la sua variazione è uguale a zero.
Si dimostra così che

$$N_h = \mathrm{Ind}_{C_h}(0) = \mathrm{Ind}_{C_f}(0) = N_f;$$

cioè le due funzioni $h = f+g$ ed f hanno il medesimo numero di zeri racchiusi da C. $\square$

Useremo questo teorema per dimostrare che $D_s(z)$ ha un unico zero nel disco unitario $|z| \leq 1$; troviamo però dapprima i valori di z che sono degli zeri di $D_s(z)$.
Scriviamo $D_s(z)$ come

$$D_s(z) = -[\lambda z^2 - (s + \mu + \lambda)z + \mu],$$

i suoi zeri sono così

$$\alpha_1(s) = \frac{s + \mu + \lambda + \sqrt{(s + \mu + \lambda)^2 - 4\lambda\mu}}{2\lambda}$$

e

$$\alpha_2(s) = \frac{s + \mu + \lambda - \sqrt{(s + \mu + \lambda)^2 - 4\lambda\mu}}{2\lambda}$$

dove $\alpha_k(s)$ ha $\mathrm{Re}(s) > 0$ per $k = 1, 2$; quindi

$$D_s(z) = -\lambda[z - \alpha_1(s)][z - \alpha_2(s)].$$

Decomponiamo ora $D_s(z)$ come somma di $f_s(z)$ e $g_s(z)$ dove

$$f_s(z) = (s + \mu + \lambda)z$$

e

$$g_s(z) = -(\lambda z^2 + \mu)$$

sono funzioni a variabile complessa z analitiche in $|z| \leq 1$.
Ora, per $|z| = 1$ ($\mathrm{Re}(s) > 0$) vale

$$|f_s(z)| = |\lambda + \mu + s| \geq \lambda + \mu + \mathrm{Re}(s) > \lambda + \mu$$

e

$$|g_s(z)| \leq \lambda + \mu$$

pertanto

$$|g_s(z)| < |f_s(z)|.$$

Dunque per il teorema di Rouché, come $f_s(z)$ ha un unico zero all'interno del disco unitario $|z| \leq 1$ ($\mathrm{Re}(s) > 0$), così anche $D_s(z) = f_s(z) + g_s(z)$ ha

un solo zero lì.
Poiché come dimostreremo $|\alpha_2(s)| < |\alpha_1(s)|$ e non vi è uno zero in $|z| = 1$, questo zero deve essere $\alpha_2(s)$. Dimostriamo quest'ultima disequazione, cioè $|\alpha_2(s)| < |\alpha_1(s)|$.

Dimostrazione. Per semplicità denotiamo $h \triangleq \sqrt{(s+\mu+\lambda)^2 - 4\lambda\mu}$ e affinché la tesi sia verificata si deve dimostrare la condizione equivalente $|\alpha_2(s)|^2 - |\alpha_1(s)|^2 > 0$; cioè deve essere soddisfatta la condizione

$$\mathrm{Re}(s+\mu+\lambda)\mathrm{Re}(h) + \mathrm{Im}(s+\mu+\lambda)\mathrm{Im}(h) > 0 \tag{3.9}$$

Analizziamo per $\mathrm{Re}(s) > 0$ i tre casi possibili dipendenti dal segno di $\mathrm{Im}(s)$.

1. $\mathrm{Im}(s) = 0$
   Si può esprimere h come

   $$h = \sqrt{s^2 + 2s(\mu+\lambda) + (\mu+\lambda)^2 - 4\lambda\mu} = \sqrt{s^2 + 2s(\mu+\lambda) + (\mu+\lambda)^2}$$

   In questo caso siccome s è un numero reale positivo anche h lo sarà e quindi vale 3.9.

2. $\mathrm{Im}(s) > 0$
   Si ha quindi che $s+\mu+\lambda$ è un punto che si trova nel primo quadrante del piano cartesiano in quanto $\mathrm{Re}(s+\mu+\lambda) > 0$ e $\mathrm{Im}(s+\mu+\lambda) > 0$. Il numero complesso $(s+\mu+\lambda)^2$ può pertanto essere rappresentato nel primo o nel secondo quadrante. Se sottraiamo ad esso il numero reale $4\lambda\mu$ si ottiene $(s+\mu+\lambda)^2 - 4\lambda\mu = h^2$ che rimane nel primo o nel secondo quadrante ed in più se prendiamo la sua radice quadrata si trova il punto h nel primo quadrante.
   Quindi anche in questo caso è soddisfatta la condizione 3.9.

3. $\mathrm{Im}(s) < 0$
   In questo caso $\mathrm{Re}(s+\mu+\lambda) > 0$ e $\mathrm{Im}(s+\mu+\lambda) < 0$. Procedendo come nel caso 2) si verifica che h è un punto che può essere rappresentato nel

quarto quadrante; quindi $\mathrm{Re}(h) > 0$ e $\mathrm{Im}(h) < 0$.

Quindi anche in questo caso vale 3.9.

$\square$

Inoltre poiché $P^*(z,s)$ per la sua analiticità deve convergere nella regione $|z| \leq 1$ per $\mathrm{Re}(s) > 0$, allora in questa regione gli zeri di $D_s(z)$ devono essere zeri anche per $N_s(z)$; dunque $\alpha_2(s)$ è uno zero anche di $N_s(z)$.
Così, si scrive

$$[\alpha_2(s)]^{i+1} - \mu[1-\alpha_2(s)]P_0^*(s) = 0$$

da cui

$$P_0^*(s) = \frac{[\alpha_2(s)]^{i+1}}{\mu[1-\alpha_2(s)]}.$$

Prendendo $\alpha_2(s) = \alpha_2$ e $\alpha_1(s) = \alpha_1$ per semplificare la scrittura, si può riscrivere $P^*(z,s)$ come:

$$\begin{aligned} P^*(z,s) &= \frac{z^{i+1} - \mu(1-z)\frac{\alpha_2^{i+1}}{\mu(1-\alpha_2)}}{-\lambda(z-\alpha_1)(z-\alpha_2)} \\ &= \frac{(1-\alpha_2)z^{i+1} - (1-z)\alpha_2^{i+1}}{-\lambda(z-\alpha_1)(z-\alpha_2)(1-\alpha_2)}. \end{aligned}$$

Il numeratore può essere scritto come

$$\begin{aligned} & z^{i+1} - \alpha_2^{i+1} - \alpha_2 z(z^i - \alpha_2^i) \\ &= (z-\alpha_2)(z^i + z^{i-1}\alpha_2 + \cdots + z\alpha_2^{i-1} + \alpha_2^i) \\ &\quad - \alpha_2 z(z-\alpha_2)(z^{i-1} + z^{i-2}\alpha_2 + \cdots + z\alpha_2^{i-2} + \alpha_2^{i-1}) \\ &= (z-\alpha_2)[(1-\alpha_2)(z^i + z^{i-1}\alpha_2 + \cdots + z\alpha_2^{i-1} + \alpha_2^i) + \alpha_2^{i+1}]. \end{aligned}$$

Quindi,

$$\begin{aligned} P^*(z,s) &= \frac{(z-\alpha_2)[(1-\alpha_2)(z^i + z^{i-1}\alpha_2 + \cdots + z\alpha_2^{i-1} + \alpha_2^i) + \alpha_2^{i+1}]}{-\lambda(z-\alpha_1)(z-\alpha_2)(1-\alpha_2)} \\ &= \frac{(1-\alpha_2)(z^i + z^{i-1}\alpha_2 + \cdots + z\alpha_2^{i-1} + \alpha_2^i) + \alpha_2^{i+1}}{\lambda\alpha_1(1-z/\alpha_1)(1-\alpha_2)} \\ &= \frac{(1-\frac{z}{\alpha_1})^{-1}}{\lambda\alpha_1}\frac{[(1-\alpha_2)(z^i + z^{i-1}\alpha_2 + \cdots + z\alpha_2^{i-1} + \alpha_2^i) + \alpha_2^{i+1}]}{(1-\alpha_2)}. \end{aligned}$$

Ora, facendo uso dell'equazione $\left(1-\frac{z}{\alpha_1}\right)^{-1} = \sum_{k=0}^{\infty}\left(\frac{z}{\alpha_1}\right)^k$ e notando che $|z/\alpha_1| < 1$ si ha

$$P^*(z,s) = \frac{1}{\lambda\alpha_1}\sum_{k=0}^{\infty}\left(\frac{z}{\alpha_1}\right)^k (z^i + z^{i-1}\alpha_2 + \cdots + z\alpha_2^{i-1} + \alpha_2^i) + \frac{1}{\lambda\alpha_1}\sum_{k=0}^{\infty}\left(\frac{z}{\alpha_1}\right)^k \frac{\alpha_2^{i+1}}{(1-\alpha_2)}. \tag{3.10}$$

Prima di procedere, riportiamo alcune semplici notazioni

$$\alpha_1 + \alpha_2 = \frac{\lambda+\mu+s}{\lambda}, \quad \alpha_1\alpha_2 = \frac{\mu}{\lambda}, \quad s = -\lambda(1-\alpha_1)(1-\alpha_2).$$

Dobbiamo a questo punto cercare di invertire la soluzione rispetto ad entrambe le due variabili trasformate z ed s per ottenere una soluzione più esplicita in termini di k, cioè $P_k(t)$. Iniziamo calcolando la trasformata inversa rispetto a z, cioè $P_k^*(s)$ che è il coefficiente di z^k della trasformata 3.2. E' chiaro che il contributo di questo coefficiente è la somma dei coefficienti di z^k dell'equazione 3.10. Dunque, il contributo del coefficiente del primo termine a destra di $P^*(z,s)$ ottenuto moltiplicando z^{i-m} per z^{k-i+m} in generale è:

$$\frac{1}{\lambda\alpha_1}\frac{\alpha_2^m}{\alpha_1^{k-i+m}} = \frac{1}{\lambda\alpha_1}\frac{\alpha_2^m\alpha_1^m}{\alpha_1^{k-i+2m}} = \frac{(\mu/\lambda)^m}{\lambda\alpha_1^{k-i+2m+1}}. \tag{3.11}$$

dove abbiamo considerato $k \geq i$ e utilizzato $\alpha_1\alpha_2 = \mu/\lambda$.

Il secondo termine di 3.10 ci fornisce i rimanenti coefficienti di z^k. Il loro contributo sempre per $k \geq i$ e dato che $|\alpha_2| < 1$ è

$$\begin{aligned}\frac{\alpha_2^{i+1}}{\lambda\alpha_1^{k+1}(1-\alpha_2)} &= \frac{\alpha_2^{i+1}}{\lambda\alpha_1^{k+1}}(1+\alpha_2+\alpha_2^2+\cdots) \\ &= \frac{1}{\lambda}\frac{\alpha_2^{k+i+2}}{(\alpha_1\alpha_2)^{k+1}}(1+\alpha_2+\alpha_2^2+\cdots) \\ &= \frac{1}{\lambda}\left(\frac{\lambda}{\mu}\right)^{k+1}\sum_{j=k+i+2}^{\infty}\left(\frac{\mu}{\lambda}\right)^j\frac{1}{\alpha_1^j}\end{aligned} \tag{3.12}$$

Quindi mettendo ora insieme i due contributi, per $k \geq i$ si ha:

$$P_k^*(s) = \frac{1}{\lambda}\left[\frac{1}{\alpha_1^{k-i+1}} + \frac{(\mu/\lambda)}{\alpha_1^{k-i+3}} + \frac{(\mu/\lambda)^2}{\alpha_1^{k-i+5}} + \cdots + \frac{(\mu/\lambda)^i}{\alpha_1^{k-i+1}} + \left(\frac{\lambda}{\mu}\right)^{k+1}\sum_{j=k+i+2}^{\infty}\frac{(\mu/\lambda)^j}{\alpha_1^j}\right] \tag{3.13}$$

che è la trasformata di Laplace della probabilità in transitorio $P_k(t)$ cercata. Per conoscere $P_k(t)$ occorre trovare la trasformata inversa di tutti i termini del membro destro di 3.13. Iniziamo cercando la trasformata inversa del termine generico

$$\alpha \triangleq \left[\frac{s+\sqrt{s^2-4\lambda\mu}}{2\lambda}\right]^{-\nu}.$$

Esso è la trasformata di Laplace di

$$(2\lambda)^{-\nu}\nu(2\sqrt{\lambda\mu})^{-\nu}t^{-1}I_\nu(2\sqrt{\lambda\mu}t) = \nu\left(\sqrt{\frac{\lambda}{\mu}}\right)^{\nu} t^{-1}I_\nu(2\sqrt{\lambda\mu}\,t) \tag{3.14}$$

dove $I_\nu(z) \triangleq \sum_{m=0}^{\infty}\dfrac{(z/2)^{\nu+2m}}{m!(\nu+m)!}$ è la funzione di Bessel di primo tipo di ordine ν.

Ora, se $F^*(s)$ è la trasformata di Laplace di $F(t)$, allora per la proprietà delle trasformate

$$F^*(s+a) \Longleftrightarrow e^{-at}F(t);$$

siccome $\alpha_1 = \dfrac{s+\mu+\lambda+\sqrt{(s+\mu+\lambda)^2-4\lambda\mu}}{2\lambda}$ è uguale ad α tranne che per il fatto che ad s è stata sommata la costante $\lambda+\mu$, si ha:

$$(\alpha_1)^{-\nu} \Longleftrightarrow e^{-(\lambda+\mu)t}\left(\sqrt{\frac{\lambda}{\mu}}\right)^{\nu}\nu t^{-1}I_\nu(2\sqrt{\lambda\mu}\,t) \tag{3.15}$$

Si è arrivati a poter scrivere:

$$\begin{aligned}
P_k(t) = &\frac{e^{-(\lambda+\mu)t}}{\lambda}\left[\left(\sqrt{\frac{\lambda}{\mu}}\right)^{k-i+1}(k-i+1)t^{-1}I_{k-i+1}(2\sqrt{\lambda\mu}\,t)\right.\\
&+\frac{\mu}{\lambda}\left(\sqrt{\frac{\lambda}{\mu}}\right)^{k-i+3}(k-i+3)t^{-1}I_{k-i+3}(2\sqrt{\lambda\mu}\,t)+\cdots\\
&+\left(\frac{\mu}{\lambda}\right)^{i}\left(\sqrt{\frac{\lambda}{\mu}}\right)^{k+i+1}(k+i+1)t^{-1}I_{k+i+1}(2\sqrt{\lambda\mu}\,t)\\
&\left.+\left(\frac{\lambda}{\mu}\right)^{k+1}\sum_{j=k+i+2}^{\infty}\left(\sqrt{\frac{\mu}{\lambda}}\right)^{j} jt^{-1}I_j(2\sqrt{\lambda\mu}\,t)\right].
\end{aligned} \tag{3.16}$$

Sostituendo la nota relazione

$$\frac{2\nu}{z} I_\nu(z) = I_{\nu-1}(z) - I_{\nu+1}(z),$$

si ha:

$$\begin{aligned}
P_k(t) = & \frac{e^{-(\lambda+\mu)t}}{\lambda} \Bigg[\left(\sqrt{\frac{\lambda}{\mu}} \right)^{k-i+1} \sqrt{\lambda\mu}\, [I_{k-i}(2\sqrt{\lambda\mu}\, t) - I_{k-i+2}(2\sqrt{\lambda\mu}\, t)] \\
& + \left(\sqrt{\frac{\lambda}{\mu}} \right)^{k-i+1} \sqrt{\lambda\mu}\, [I_{k-i+2}(2\sqrt{\lambda\mu}\, t) - I_{k-i+4}(2\sqrt{\lambda\mu}\, t)] + \cdots + \\
& + \left(\sqrt{\frac{\lambda}{\mu}} \right)^{k-i+1} \sqrt{\lambda\mu}\, [I_{k+i}(2\sqrt{\lambda\mu}\, t) - I_{k+i+2}(2\sqrt{\lambda\mu}\, t)] \\
& + \left(\frac{\lambda}{\mu} \right)^{k+1} \sum_{j=k+i+2}^{\infty} \left(\sqrt{\frac{\mu}{\lambda}} \right)^{j} \sqrt{\lambda\mu}\, [I_{j-1}(2\sqrt{\lambda\mu}\, t) - I_{j+1}(2\sqrt{\lambda\mu}\, t)] \Bigg].
\end{aligned} \tag{3.17}$$

dove la seconda espressione del primo termine si cancella con la prima espressione del secondo termine.
L'ultimo termine si può scrivere come

$$\begin{aligned}
& \left(\frac{\lambda}{\mu} \right)^{k+1} \sum_{j=k+i+2}^{\infty} \left(\sqrt{\frac{\mu}{\lambda}} \right)^{j} \sqrt{\lambda\mu}\, [I_{j-1}(2\sqrt{\lambda\mu}\, t) - I_{j+1}(2\sqrt{\lambda\mu}\, t)] \\
& = \left(\frac{\lambda}{\mu} \right)^{k+1} \Bigg[\left(\sqrt{\frac{\mu}{\lambda}} \right)^{k+i+2} \sqrt{\lambda\mu}\, I_{k+i+1}(2\sqrt{\lambda\mu}\, t) \\
& + \sqrt{\frac{\mu}{\lambda}} \sum_{j=k+i+2}^{\infty} \left(\sqrt{\frac{\mu}{\lambda}} \right)^{j} \sqrt{\lambda\mu}\, I_j(2\sqrt{\lambda\mu}\, t) + \left(\sqrt{\frac{\mu}{\lambda}} \right)^{k+i+1} \sqrt{\lambda\mu}\, I_{k+i+2}(2\sqrt{\lambda\mu}\, t) \\
& \qquad - \sqrt{\frac{\lambda}{\mu}} \sum_{j=k+i+2}^{\infty} \left(\sqrt{\frac{\mu}{\lambda}} \right)^{j} \sqrt{\lambda\mu}\, I_j(2\sqrt{\lambda\mu}\, t) \Bigg] \\
& = \left(\sqrt{\frac{\lambda}{\mu}} \right)^{k-i} \sqrt{\lambda\mu}\, I_{k+i+1}(2\sqrt{\lambda\mu}\, t) + \left(\sqrt{\frac{\lambda}{\mu}} \right)^{k+i+1} \sqrt{\lambda\mu}\, I_{k+i+2}(2\sqrt{\lambda\mu}\, t) \\
& \qquad + \left(1 - \frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^{k} \sqrt{\frac{\lambda}{\mu}} \sum_{j=k+i+2}^{\infty} \left(\sqrt{\frac{\mu}{\lambda}} \right)^{j} \sqrt{\lambda\mu}\, I_j(2\sqrt{\lambda\mu}\, t)
\end{aligned} \tag{3.18}$$

da cui segue che la seconda espressione del terzo termine della 3.17 (senza considerare i termini segnati dai tre puntini) si cancella con il secondo termine della 3.18. Quindi si ha

$$
\begin{aligned}
P_k(t) = & \frac{e^{-(\lambda+\mu)t}}{\lambda} \Bigg[\left(\sqrt{\frac{\lambda}{\mu}}\right)^{k-i+1} \sqrt{\lambda\mu}\, [I_{k-i}(2\sqrt{\lambda\mu}\, t) \\
& - \left(\sqrt{\frac{\lambda}{\mu}}\right)^{k-i+1} \sqrt{\lambda\mu}\, [I_{k-i+4}(2\sqrt{\lambda\mu}\, t) \\
& + \left(\sqrt{\frac{\lambda}{\mu}}\right)^{k-i+1} \sqrt{\lambda\mu}\, [I_{k+i}(2\sqrt{\lambda\mu}\, t) \\
& + \left(\sqrt{\frac{\lambda}{\mu}}\right)^{k-i} \sqrt{\lambda\mu}\, I_{k+i+1}(2\sqrt{\lambda\mu}\, t) \\
& + \left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^{k} \sqrt{\frac{\lambda}{\mu}} \sum_{j=k+i+2}^{\infty} \left(\sqrt{\frac{\mu}{\lambda}}\right)^{j} \sqrt{\lambda\mu}\, I_j(2\sqrt{\lambda\mu}\, t) \Bigg]
\end{aligned}
\tag{3.19}
$$

Pertanto, si può concludere dicendo che per $k \geq i$ si ha

$$
\begin{aligned}
P_k(t) = & e^{-(\lambda+\mu)t} \Bigg[\left(\sqrt{\frac{\mu}{\lambda}}\right)^{i-k} I_{k-i}(2\sqrt{\lambda\mu}\, t) \\
& + \left(\sqrt{\frac{\mu}{\lambda}}\right)^{i-k+1} I_{k+i+1}(2\sqrt{\lambda\mu}\, t) \\
& + \left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^{k} \sum_{j=k+i+2}^{\infty} \left(\sqrt{\frac{\mu}{\lambda}}\right)^{j} I_j(2\sqrt{\lambda\mu}\, t) \Bigg].
\end{aligned}
\tag{3.20}
$$

che è la soluzione finale dell'analisi in transitorio.
Ora, per $k < i$ se si ripetono i vari passaggi appena visti si trova che il valore di $P_k(t)$ rimane invariato.

Anche per le code M/M/1 si ha che $P_k(t) = \sum_{i \in S} P_i(0) p_{ik}(t) \quad \forall k \in S, \forall t \geq 0$ con $\sum_{i \in S} P_i(0) = 1$. Così, se con certezza lo stato di partenza è pari ad i, ovvero $P_i(0) = 1$ e $P_k(0) = 0 \ \forall k \neq i$ si possono conoscere *le probabilità di transizione $p_{ik}(t)$ per un sistema a coda M/M/1.* Infatti:

a. se $i = 0$ allora si ha $p_{0k}(t) = P_k(t)$;

b. se lo stato i di partenza della coda è arbitrario, allora le probabilità di transizione sono:

$$\begin{cases} p_{ik}(t) = 0 & \text{per } k < i \\ p_{ik}(t) = P_{k-i}(t) & \text{per } k \geq i \end{cases}.$$

Oltre a questi risultati dipendenti dal tempo possiamo ora passare allo studio del comportamento all'equilibrio del sistema a coda M/M/1.
La condizione 2.8 del precedente capitolo particolarizzata con i nostri coefficienti λ e μ costanti, che descrive la stazionarietà del nostro sistema a coda, diventa:

$$\sum_{k=1}^{\infty} \frac{\prod_{i=0}^{k-1} \lambda_i}{\prod_{j=1}^{k} \mu_j} < \infty \quad \Rightarrow \quad \sum_{k=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^k < \infty.$$

La serie nella parte destra dell'implicazione converge se e solo se $\dfrac{\lambda}{\mu} < 1$, da cui segue $\lambda < \mu$. Supposta vera questa condizione, si ricava rispettivamente dalle equazioni 2.6 e 2.7 che:

$$p_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^k} = \frac{1}{\sum_{k=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^k} = \frac{1}{\frac{1}{1-\frac{\lambda}{\mu}}} = 1 - \frac{\lambda}{\mu} \tag{3.21}$$

$$p_k = \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right) \quad k \geq 1 \tag{3.22}$$

dove p_0 rappresenta la frazione di tempo in cui il servitore è inattivo e p_k rappresenta la probabilità a regime di trovare k elementi nel sistema.

Verifichiamo ora che la soluzione in transitorio $P_k(t)$ tende per $t \to \infty$ alla soluzione di equilibrio p_k trovata. Per dimostrare quanto appena detto, mostriamo prima che vale la seguente espressione:

$$\sum_{k=0}^{\infty} P_k(t) = 1 = e^{-(\lambda+\mu)t} \exp\left[\sqrt{\lambda\mu}\left(\sqrt{\frac{\lambda}{\mu}} + \sqrt{\frac{\lambda}{\mu}}\right)t\right]. \tag{3.23}$$

Cominciamo considerando la seguente somma delle probabilità

$$\begin{aligned}\sum_{k=0}^{\infty} P_k(t) = e^{-(\lambda+\mu)t}\Bigg[&\sum_{m=-i}^{\infty}\left(\sqrt{\frac{\lambda}{\mu}}\right)^m I_m(2\sqrt{\lambda\mu}\,t)\\ &+\left(\sqrt{\frac{\lambda}{\mu}}\right)^{-2(i+1)}\sum_{m=i+1}^{\infty}\left(\sqrt{\frac{\lambda}{\mu}}\right)^m I_m(2\sqrt{\lambda\mu}\,t)\\ &+\left(1-\frac{\lambda}{\mu}\right)\sum_{k=0}^{\infty}\sum_{j=k+i+2}^{\infty}\left(\frac{\lambda}{\mu}\right)^k\left(\sqrt{\frac{\mu}{\lambda}}\right)^j I_j(2\sqrt{\lambda\mu}\,t)\Bigg]. \end{aligned} \tag{3.24}$$

Ora, scambiando la sommatoria e sommando rispetto a k, l'ultima espressione diventa:

$$\begin{aligned}\left(1-\frac{\lambda}{\mu}\right)&\sum_{j=i+2}^{\infty}\sum_{k=0}^{j-i-2}\left(\frac{\lambda}{\mu}\right)^k\left(\sqrt{\frac{\mu}{\lambda}}\right)^j I_j(2\sqrt{\lambda\mu}\,t)\\ &=\sum_{j=i+2}^{\infty}\left[1-\left(\frac{\lambda}{\mu}\right)^{j-i-1}\right]\left(\sqrt{\frac{\mu}{\lambda}}\right)^j I_j(2\sqrt{\lambda\mu}\,t)\\ &=\sum_{m=i+2}^{\infty}\left(\sqrt{\frac{\mu}{\lambda}}\right)^m I_m(2\sqrt{\lambda\mu}\,t)\\ &\quad-\left(\sqrt{\frac{\lambda}{\mu}}\right)^{-2(i+1)}\sum_{m=i+2}^{\infty}\left(\sqrt{\frac{\lambda}{\mu}}\right)^m I_m(2\sqrt{\lambda\mu}\,t). \end{aligned} \tag{3.25}$$

Così, utilizzando il fatto che $I_n(z) = I_{-n}(z)$ e addizionando la prima espressione a destra di 3.25 con la prima espressione dentro le parentesi di 3.24 e la seconda di 3.25 con la seconda espressione dentro le parentesi di 3.24, si ha

$$\begin{aligned}\sum_{k=0}^{\infty} P_k(t) = e^{-(\lambda+\mu)t}\Bigg[&\sum_{m=-i}^{\infty}\left(\sqrt{\frac{\lambda}{\mu}}\right)^m I_m(2\sqrt{\lambda\mu}\,t)\\ &+\sum_{m=-\infty}^{-(i+2)}\left(\sqrt{\frac{\lambda}{\mu}}\right)^m I_m(2\sqrt{\lambda\mu}\,t)\\ &+\left(\sqrt{\frac{\lambda}{\mu}}\right)^{-(i+1)} I_{-(i+1)}(2\sqrt{\lambda\mu}\,t)\Bigg] \end{aligned} \tag{3.26}$$

Per terminare questa dimostrazione definiamo dapprima le seguenti notazioni generali:

$$e^{(x/2)(y+1/y)} = \sum_{k=-\infty}^{\infty} y^k I_k(x) = \sum_{k=0}^{\infty} y^k I_k(x) + \sum_{k=1}^{\infty} y^{-k} I_k(x). \tag{3.27}$$

Dunque, la 3.26 per quanto appena detto, diventa:

$$\begin{aligned} \sum_{k=0}^{\infty} P_k(t) &= e^{-(\lambda+\mu)t} \sum_{m=-\infty}^{\infty} \left(\sqrt{\frac{\lambda}{\mu}}\right)^m I_m(2\sqrt{\lambda\mu}\, t) \\ &= e^{-(\lambda+\mu)t} \exp\left[\sqrt{\lambda\mu}\left(\sqrt{\frac{\lambda}{\mu}} + \sqrt{\frac{\lambda}{\mu}}\right) t\right] = 1. \end{aligned} \tag{3.28}$$

Mostriamo ora che per $\lambda < \mu$ la soluzione in transitorio tende a quella di equilibrio per $t \to \infty$. Per facilitare i calcoli assumiamo innanzitutto che $i = 0$.

Ora, poiché in generale per la funzione di Bessel di primo tipo di ordine ν, $I_\nu(z)$, vale

$$I_\nu(z) = \frac{e^z}{\sqrt{2\pi z}} \qquad z \to \infty;$$

allora per λ e μ fissati e per $t \to \infty$ si scrive

$$I_k(2\sqrt{\lambda\mu}\, t) \sim \frac{\exp(2\sqrt{\lambda\mu}\, t)}{\sqrt{2\pi(2\sqrt{\lambda\mu})\, t)}} \tag{3.29}$$

Procediamo andando a sostituire questa espressione nella soluzione in transitorio 3.20. Usando il fatto che $\lambda + \mu > 2\sqrt{\lambda\mu}$ in quanto abbiamo assunto che $\lambda/\mu < 1$ si ha che i primi due termini del secondo membro della 3.20 tendono a zero per $t \to \infty$.

Per quanto riguarda invece il comportamento al limite dell'ultimo fattore sempre della 3.20; si nota dapprima che

$$\begin{aligned} \sum_{j=k+2}^{\infty} \left(\sqrt{\frac{\mu}{\lambda}}\right)^j I_j(2\sqrt{\lambda\mu}\, t) &= \sum_{j=0}^{\infty} \left(\sqrt{\frac{\mu}{\lambda}}\right)^j I_j(2\sqrt{\lambda\mu}\, t) \\ &\quad - \sum_{j=0}^{k+1} \left(\sqrt{\frac{\mu}{\lambda}}\right)^j I_j(2\sqrt{\lambda\mu}\, t). \end{aligned} \tag{3.30}$$

Se moltiplichiamo ora quest'ultima espressione per $e^{-(\lambda+\mu)t}$, si ha che la somma finita di destra tende a zero per $t \to \infty$ poiché ogni termine tende a zero separatamente. Mentre per risolvere il limite dell'altra serie, cioè del primo termine a destra di 3.30, torniamo a considerare la scrittura 3.27.
Si vede che l'ultima somma a destra di 3.27 con $y = \sqrt{\frac{\mu}{\lambda}}$ e $x = \sqrt{\lambda\mu}$ tende a zero per $t \to \infty$. Ora, risulta che la prima somma di destra di 3.27 è proprio uguale alla prima serie a destra della 3.30 che a noi interessa studiare.
Si può terminare la dimostrazione scrivendo per quanto visto nelle espressioni 3.27 e 3.28 quanto segue

$$e^{-(\lambda+\mu)t} \sum_{k=0}^{\infty} \left(\sqrt{\frac{\mu}{\lambda}} \right)^k I_k(2\sqrt{\lambda\mu}\, t)$$
$$= e^{-(\lambda+\mu)t} \exp \left[\sqrt{\lambda\mu} \left(\sqrt{\frac{\lambda}{\mu}} + \sqrt{\frac{\lambda}{\mu}} \right) t \right] = 1.$$

Pertanto

$$P_k(t) \to \left(\frac{\lambda}{\mu} \right)^k \left(1 - \frac{\lambda}{\mu} \right) \qquad \text{per } t \to \infty.$$

Bibliografia

[1] Leonard Kleinrock, *Sistemi a coda : introduzione alla teoria delle code.* U. Hoepli, 1992

[2] Leonard Kleinrock, Richard Gail *Queueing systems : problems and solutions.* Wiley, 1996

[3] Thomas L. Saaty, *Elements of queueing theory : with applications.* McGraw-Hill, 1961.

[4] F. Biagini, M. Campanino, *Elementi di probabilità e statistica.* Springer, 2006.

[5] Walter Rudin, *Analisi reale e complessa.* Boringhieri, 1988.

Ringraziamenti

In primo luogo vorrei ringraziare il professore Massimo Campanino che mi ha permesso di eseguire questo elaborato e mi ha seguito durante la sua stesura con grande disponibilità.

I ringraziamenti più sentiti vanno a coloro la cui tesi è dedicata, Mamma, Papà e Silvia. Ai miei genitori dico grazie per avermi sempre sostenuto, economicamente e non, e per aver sempre creduto in me anche in quei momenti in cui il percorso si è fatto più difficoltoso. A mia sorella invece, dico grazie per essere stata il mio grande punto di riferimento e di forza senza il quale forse non avrei raggiunto questo traguardo. Un ringraziamento anche ai mie nonni, grande esempio di molti valori oggi un po' affievoliti.
Poi c'è Lui, Roberto, lo ringrazio per aver alleggerito quelle giornate troppo uguali l'una con l'altra passate sopra ai libri con le nostre chiacchierate e per aver capito quali erano i momenti in cui avevo più il bisogno di sentirlo vicino e gli altri in cui invece avevo più la necessità di concentrare le mie forze sugli esami.

Un grazie di cuore lo rivolgo Lucia e Lauxite, compagne di tutte i giorni. Con loro ho condiviso tutto, la suoneria più simpatica per iniziare la giornata, l'ascolto e il supporto che ci siamo date nei momenti più o meno gioiosi, i pranzi e le cene fatte con un pizzico di fantasia, i viaggi in treno, l'ansia per affrontare gli esami, le serate nei locali Bolognesi a cui fanno da sfondo *le passeggiate* lungo i portici di questa città.

Grazie a Simona e Fosca, coinquiline selezionate con accortezza prima ma amiche in poco tempo le quali hanno portato ognuna a suo modo interessanti novità e con le quali si è creato fin da subito un equilibrio di serena convivenza in casa.

Un ringraziamento generale va infine a parenti e amiche che con me hanno condiviso rispettivamente i tradizionali giorni di festa in famiglia e i piacevoli aperitivi, cene o coffee time.