

TEORIA DELLE CODE

Processi di nascita e morte

prof.ssa C. De Francesco

prof. L. De Giovanni

Università degli Studi di Padova

Dipartimento di Matematica “Tullio Levi-Civita”

Descrizione analitica di code come processi di nascita e morte (birth-death)

Consideriamo un sistema di code con:

- m serventi identici
- capacità infinita della coda
- disciplina FIFO
- se n è il numero di utenti nel sistema (coda + servizio):
 - a) arrivi “poissoniani” (o markoviani) con tasso λ_n
 - b) serviti “poissoniani” (o markoviani) con tasso μ_n (tasso di servizio combinato dei serventi attivi)

arrivi \sim nascite / serviti \sim morti

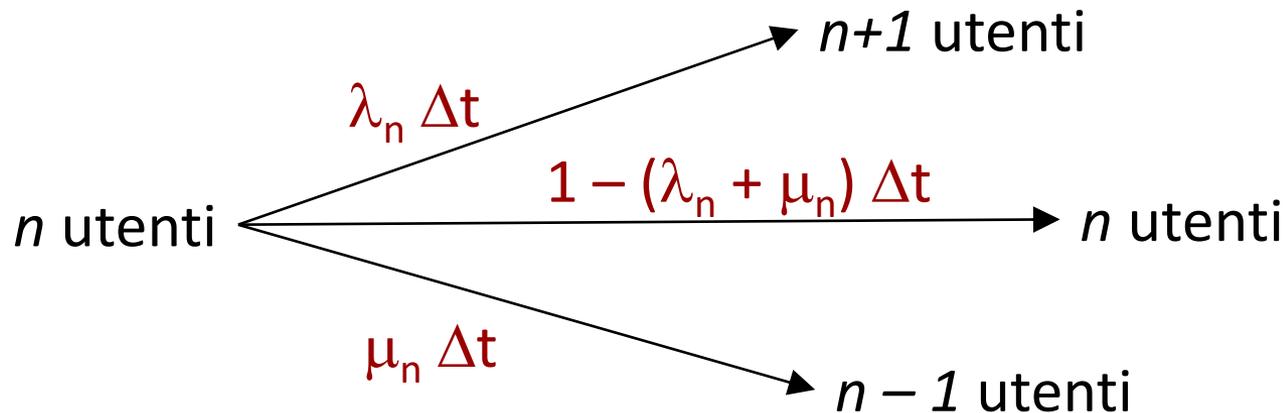
Relazioni fondamentali

Consideriamo le **probabilità di transizione** in un intervallo $(t, t+\Delta t]$

$$P[N(t+\Delta t)=n+1 | N(t)=n] = \lambda_n \Delta t + o(\Delta t)$$

$$P[N(t+\Delta t)=n-1 | N(t)=n] = \mu_n \Delta t + o(\Delta t)$$

$$P[N(t+\Delta t)=n | N(t)=n] = 1 - (\lambda_n + \mu_n) \Delta t + o(\Delta t)$$



$$P_n(t+\Delta t) = P_{n-1}(t) \lambda_{n-1} \Delta t + P_{n+1}(t) \mu_{n+1} \Delta t + P_n(t) [1 - (\lambda_n + \mu_n) \Delta t] + o(\Delta t)$$

Sistema di equazioni differenziali

La relazione precedente può essere riscritta come

$$\begin{aligned} [P_n(t+\Delta t) - P_n(t)]/\Delta t &= \\ &= P_{n-1}(t) \lambda_{n-1} + P_{n+1}(t) \mu_{n+1} - P_n(t) (\lambda_n + \mu_n) + o(\Delta t)/\Delta t \end{aligned}$$

Per $\Delta t \rightarrow 0$, e considerando $\lambda_{-1} = \mu_0 = 0$, si ottiene il **sistema di equazioni differenziali**:

$$d P_n(t)/dt = \lambda_{n-1} P_{n-1}(t) + \mu_{n+1} P_{n+1}(t) - (\lambda_n + \mu_n) P_n(t) \quad n = 1, 2, 3 \dots$$

$$d P_0(t)/dt = \mu_1 P_1(t) - \lambda_0 P_0(t)$$

Studio in condizioni di equilibrio

- Risolvendo (se possibile) il sistema di equazioni differenziali, otteniamo le $P_n(t)$, quindi *l'evoluzione temporale* del sistema
- Siamo interessati allo studio in condizioni di equilibrio, quando (e se), per $t \rightarrow \infty$, le diverse $P_n(t)$ sono “stabili”: $P_n(t) \rightarrow P_n$
- Studio in **condizioni di equilibrio**:
 - si ottiene (sotto condizioni) per $t \rightarrow \infty$
 - $P_n(t)$ indipendenti da t : $\lim_{t \rightarrow \infty} P_n(t) = P_n$ (costanti)
 - $d P_n(t) / dt = 0$, $n = 0,1,2,3 \dots$

Il sistema di equazioni differenziali diventa un sistema di equazioni (**equazioni di equilibrio**).

Sistema delle equazioni di equilibrio

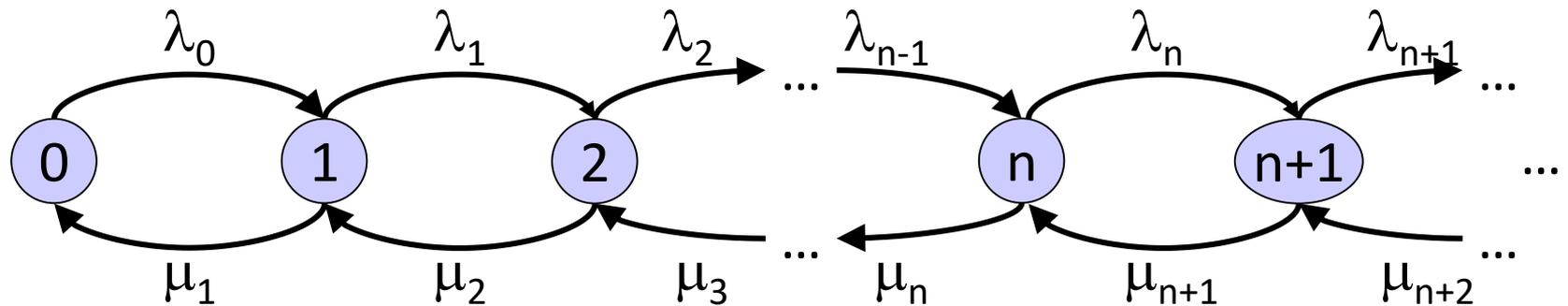
$$\lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1} = (\lambda_n + \mu_n) P_n, \quad n = 1, 2, 3 \dots$$
$$\mu_1 P_1 = \lambda_0 P_0$$

Risolvendole, si ottengono forme chiuse per P_n , L , L_q , \bar{W} , \bar{W}_q

Il sistema di equazioni ottenuto si può ricavare anche dall'analisi del diagramma delle transizioni di stato...

Diagramma delle transizioni di stato

Lo studio delle condizioni di equilibrio può essere fatto con una rappresentazione tramite diagramma delle transizioni di stato

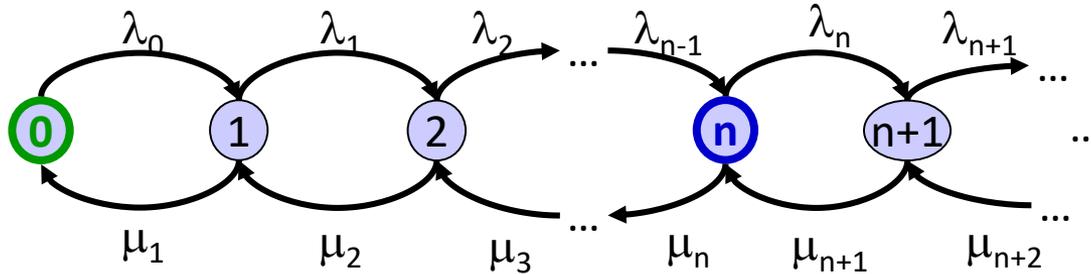


Stato: numero di utenti nel sistema

Sistema delle equazioni di equilibrio come relazioni di “conservazione del flusso”

Stato ergodico $\Rightarrow P_n$ stazionarie

\Rightarrow prob. ingresso = prob. uscita da uno stato ($\equiv dP_n(t)/dt = 0$)



$$P_0 \lambda_0 = P_1 \mu_1$$

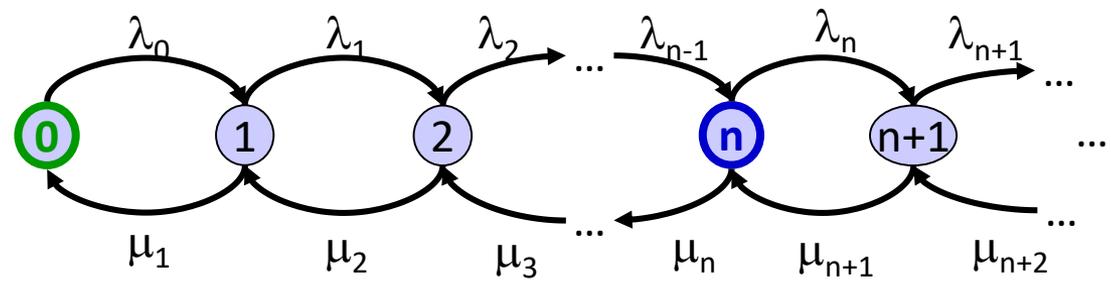
$$P_n (\lambda_n + \mu_n) =$$

$$= P_{n+1} \mu_{n+1} + P_{n-1} \lambda_{n-1}$$

Sistema delle equazioni di equilibrio come relazioni di “conservazione del flusso”

Stato ergodico $\Rightarrow P_n$ stazionarie

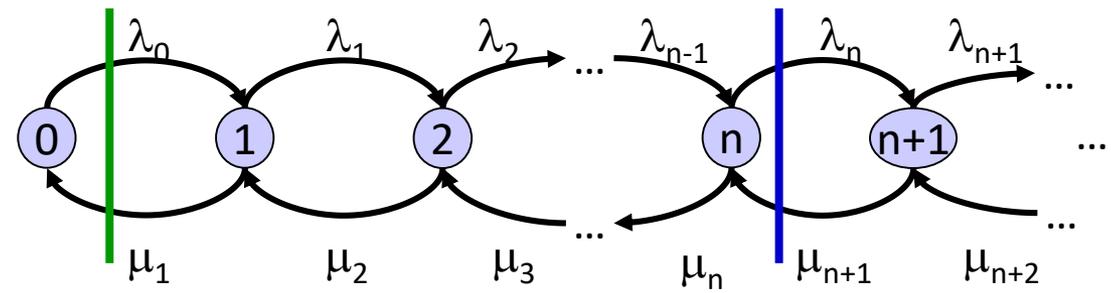
\Rightarrow prob. ingresso = prob. uscita da uno stato ($\equiv dP_n(t)/dt = 0$)



$$P_0 \lambda_0 = P_1 \mu_1$$

$$P_n (\lambda_n + \mu_n) = P_{n+1} \mu_{n+1} + P_{n-1} \lambda_{n-1}$$

\Rightarrow prob. transizioni in un verso = prob. transizioni verso opposto



$$P_0 \lambda_0 = P_1 \mu_1$$

$$P_n \lambda_n = P_{n+1} \mu_{n+1}$$

Risoluzione del sistema

$$P_1 = \lambda_0 / \mu_1 P_0$$

$$P_2 = \lambda_1 / \mu_2 P_1 = \lambda_1 \lambda_0 / (\mu_2 \mu_1) P_0$$

...

$$P_n = \lambda_{n-1} / \mu_n P_{n-1} = \lambda_{n-1} \lambda_{n-2} \dots \lambda_0 / (\mu_n \mu_{n-1} \dots \mu_1) P_0$$

Posto: $\pi_0 := 1$ e $\pi_n := \lambda_{n-1} \lambda_{n-2} \dots \lambda_0 / (\mu_n \mu_{n-1} \dots \mu_1)$ per $n \geq 1$

$$\Rightarrow P_n = \pi_n P_0$$

Deve valere $\sum_{n=0}^{\infty} P_n = 1 \Rightarrow (\sum_{n=0}^{\infty} \pi_n) P_0 = 1 \Rightarrow P_0 = \frac{1}{\sum_{n=0}^{\infty} \pi_n}$

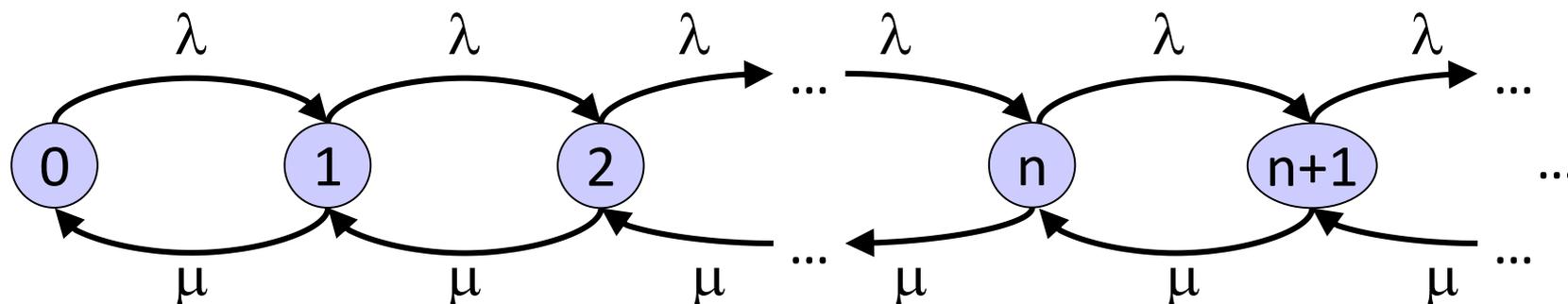
Condizione di ergodicit : $\sum_{n=0}^{\infty} \pi_n < \infty$

Note le P_n , ricaviamo le altre quantit , ad es. $L = \sum_{n=0}^{\infty} n P_n$

La formula di Little si pu  applicare con $\lambda = \sum_{n=0}^{\infty} \lambda_n P_n$

Modello M/M/1

- Tempi di interarrivo: v.a. $\text{Exp}(\lambda)$ ($\lambda_n = \lambda$) [arrivi di Poisson]
- Tempi di servizio: $\text{Exp}(\mu)$ ($\mu_n = \mu$)
- Coda illimitata, 1 servente, FIFO



- Dalle precedenti relazioni ($\rho = \lambda/\mu$):

$$\pi_n = (\lambda/\mu)^n = \rho^n \quad \mathbf{P}_n = \rho^n P_0$$

condizione di ergodicità: $\sum_{n=0}^{\infty} \rho^n = 1/(1-\rho) < \infty$ **se $\rho < 1$**

- Quindi, se $\rho < 1$, il sistema converge verso la distribuzione:

$$\mathbf{P}_0 = 1 - \rho \quad \mathbf{P}_n = \rho^n (1 - \rho) \quad \text{per } n \geq 0$$

M/M/1: derivazione di L, Lq, W, Wq

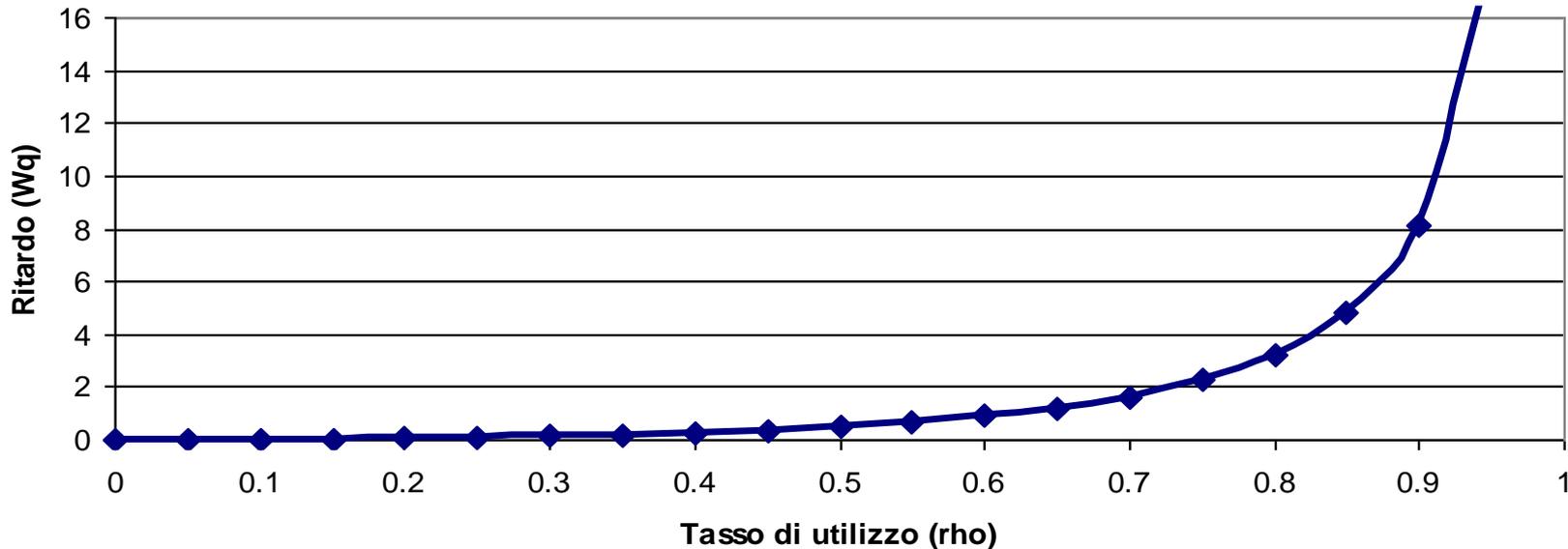
- $$\begin{aligned} L &= \sum_{n=0}^{\infty} n P_n = \sum_{n=0}^{\infty} n \rho^n (1-\rho) = (1-\rho) \sum_{n=0}^{\infty} n \rho^n = (1-\rho) \rho \sum_{n=0}^{\infty} n \rho^{n-1} = \\ &= (1-\rho) \rho \frac{d}{d\rho} \left(\sum_{n=0}^{\infty} \rho^n \right) = (1-\rho) \rho \frac{d}{d\rho} \left(\frac{1}{1-\rho} \right) = (1-\rho) \rho \frac{1}{(1-\rho)^2} = \\ &= \frac{\rho}{1-\rho} = \frac{\lambda/\mu}{1-\lambda/\mu} = \frac{\lambda}{\mu-\lambda} \end{aligned}$$

- $$\bar{W} = \frac{L}{\lambda} = \frac{1}{\mu-\lambda} \quad \text{da (1)}$$

- $$\bar{W}_q = \bar{W} - \frac{1}{\mu} = \frac{1}{\mu-\lambda} - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu-\lambda)} \quad \text{da (3)}$$

- $$L_q = \lambda \bar{W}_q = \frac{\lambda^2}{\mu(\mu-\lambda)} \quad \text{da (2)}$$

Tasso di utilizzo e ritardo



- **Alta** sensitività del ritardo ad **alti** livelli di utilizzo
- **Bassa** sensitività del ritardo a **bassi** livelli di utilizzo
- In generale, per bassi livelli di utilizzo, la coda M/M/1 è **robusta** (bassa sensitività delle grandezze a piccole variazioni di ρ)

Ulteriori proprietà di M/M/1

- Le fdp di diverse grandezze per un sistema M/M/1 possono essere derivate analiticamente
- Ad esempio: per il tempo totale nel sistema (W) la fdp è (vedi slide successiva)

$$f_W(t) = (\mu - \lambda)e^{-(\mu - \lambda)t} \quad \text{per } t \geq 0$$

$$\Rightarrow W \sim \text{Exp}(\mu - \lambda)$$

Fdp di W

La fdp di W, tempo totale nel sistema, è data dalla

$$f_W(t) = (\mu - \lambda)e^{-(\mu - \lambda)t} \quad \text{per } t \geq 0$$

Dim:

Sia A_n l'evento: un cliente arriva e trova $n \geq 0$ clienti nel sistema

$$P[A_n] = P[N = n] = P_n = \rho^n(1 - \rho)$$

Quanto tempo trascorrerà nel sistema tale cliente?

La somma di $n + 1$ v.a. $Exp(\mu)$ indipendenti:

- il tempo necessario a finire il servizio del cliente che sta attualmente ricevendo il servizio,
- il tempo necessario a servire gli $n - 1$ clienti in coda davanti al cliente in considerazione,
- il tempo di servizio del cliente in considerazione.

$$f_{W|A_n}(t|A_n) = \frac{\mu^{n+1} t^n \cdot e^{-\mu t}}{n!} \quad t \geq 0$$

Quindi:

$$\begin{aligned} f_W(t) &= \sum_{n=0}^{\infty} P_n \cdot f_{W|A_n}(t|A_n) = \sum_{n=0}^{\infty} (1 - \rho) \rho^n \frac{\mu^{n+1} t^n \cdot e^{-\mu t}}{n!} \\ &= (1 - \rho) \mu e^{-\mu t} \sum_{n=0}^{\infty} \frac{\lambda^n t^n}{n!} = (1 - \rho) \mu e^{-\mu t} \cdot e^{\lambda t} = (\mu - \lambda) e^{-(\mu - \lambda)t} \end{aligned}$$

Ulteriori proprietà di M/M/1

- Le fdp di diverse grandezze per un sistema M/M/1 possono essere derivate analiticamente
- Ad esempio: per il tempo totale nel sistema (W) la fdp è

$$f_W(t) = (\mu - \lambda)e^{-(\mu - \lambda)t} \quad \text{per } t \geq 0$$

$$\Rightarrow W \sim \text{Exp}(\mu - \lambda)$$

- La fdp del tempo in coda (W_q) è mista:

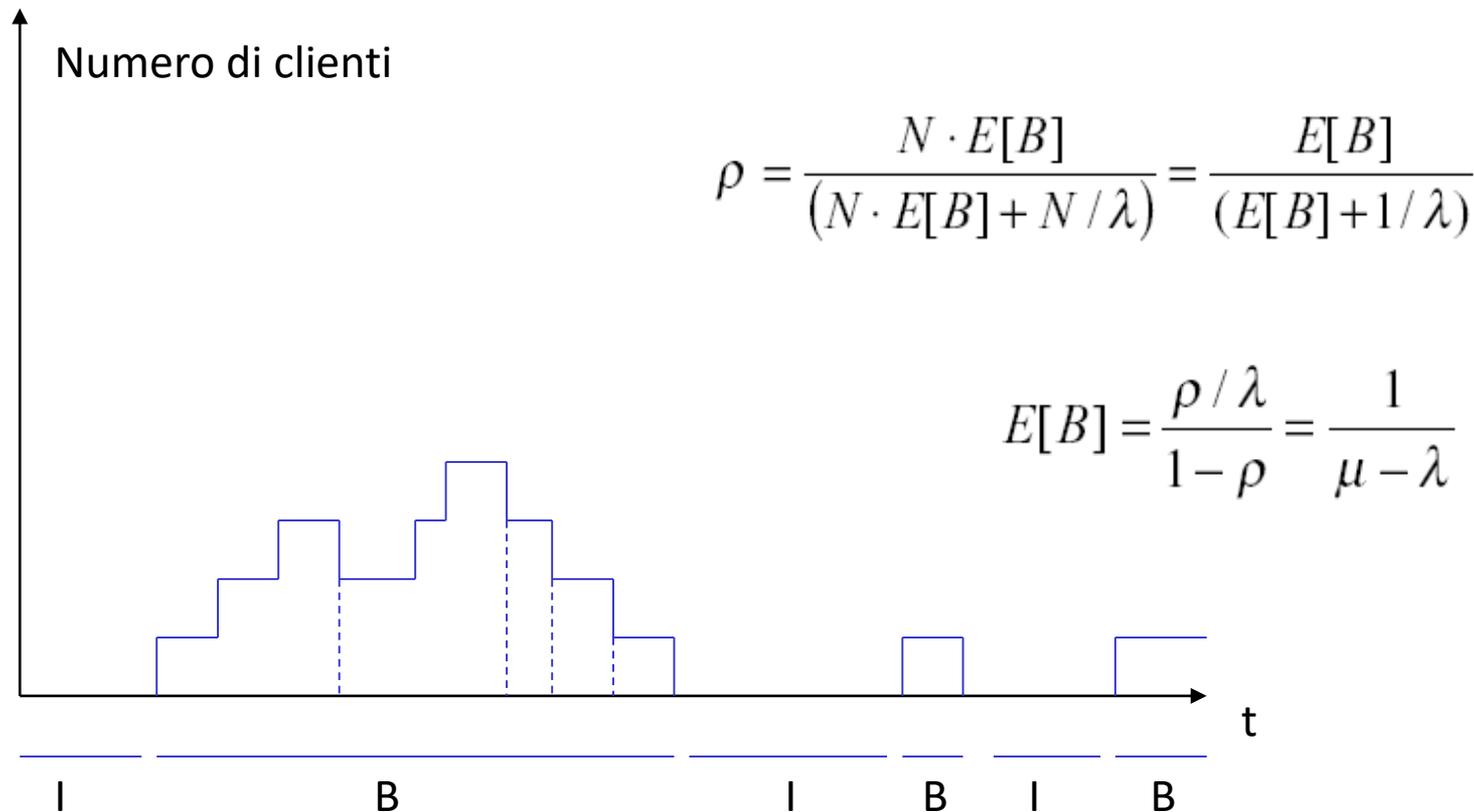
$$F_{W_q}(t) = \begin{cases} 1 - \rho & \text{se } t = 0 \\ 1 - \rho e^{-(\mu - \lambda)t} & \text{se } t > 0 \end{cases}$$

- Gli scarti quadratici medi di N , N_q , W , W_q sono tutti proporzionali a $1/(1-\rho)$, come i loro valori attesi (risp. L , L_q , \bar{W} , \bar{W}_q)

Periodi inattivi e periodi occupati – E[B]

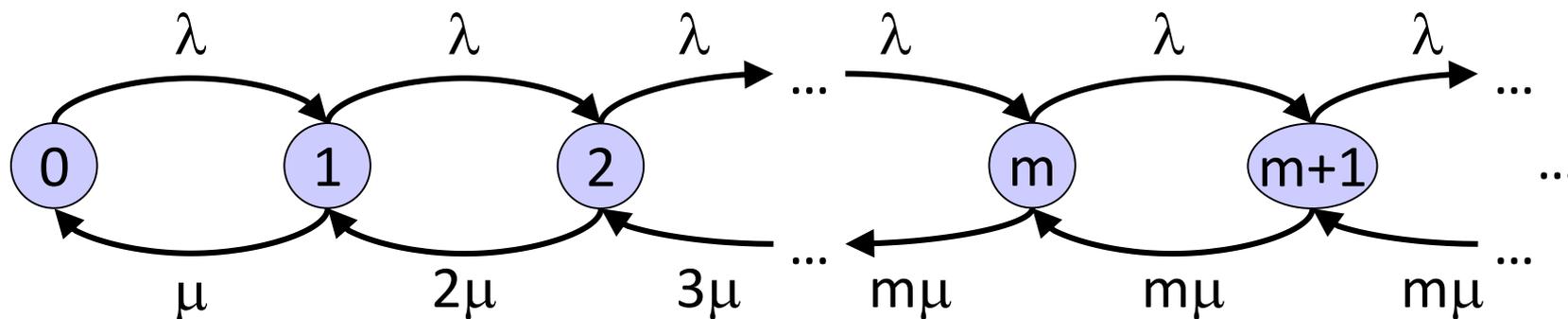
La lunghezza media del *busy period*, $E[B]$, è uguale a $1/(\mu-\lambda)$:

si osservi un gran numero, N , di *busy periods* (mediamente aspetto $1/\lambda$ perché arrivi un nuovo utente e torni in stato B)



Modello M/M/m, capacità infinita della coda

- $\lambda_n = \lambda$ $\mu_n = \min(n\mu, m\mu)$



- Ricordando che $P_n = \pi_n P_0$, in condizioni stazionarie si ha:

$$P_n = \frac{(\lambda/\mu)^n}{n!} P_0 \quad \text{per } n = 0, 1, \dots, m-1$$

$$P_n = \frac{(\lambda/\mu)^n}{m^{n-m} m!} P_0 \quad \text{per } n = m, m+1, \dots$$

M/M/m : esistenza dello stato stazionario

- Ergodicità se $\sum_{n=0}^{\infty} \pi_n$ converge

Sostituzione $j = n - m$

$$\sum_{n=0}^{\infty} \pi_n = \sum_{n=0}^{m-1} \frac{(\lambda/\mu)^n}{n!} + \sum_{n=m}^{\infty} \frac{(\lambda/\mu)^n}{m^{n-m} m!} = \sum_{n=0}^{m-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^m}{m!} \sum_{j=0}^{\infty} \frac{(\lambda/\mu)^j}{m^j}$$

- Se $\rho = \lambda/(\mu m) < 1$ (condizione di ergodicità)

$$\sum_{n=0}^{\infty} \pi_n = \sum_{n=0}^{m-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^m}{m!} \frac{1}{1 - \lambda/(m\mu)}$$

$$P_0 = \left[\sum_{n=0}^{\infty} \pi_n \right]^{-1} = \left[\sum_{n=0}^{m-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^m}{m!} \frac{1}{1 - \lambda/(m\mu)} \right]^{-1}$$

$$P_n = \pi_n P_0$$

M/M/m : alcuni risultati

- Probabilità di dover attendere:

$$C = P[N \geq m] = \sum_{n=m}^{\infty} P_n = \sum_{n=m}^{\infty} \pi_n P_0 = P_0 \frac{(\lambda/\mu)^m}{m! [1 - \lambda/(m\mu)]}$$

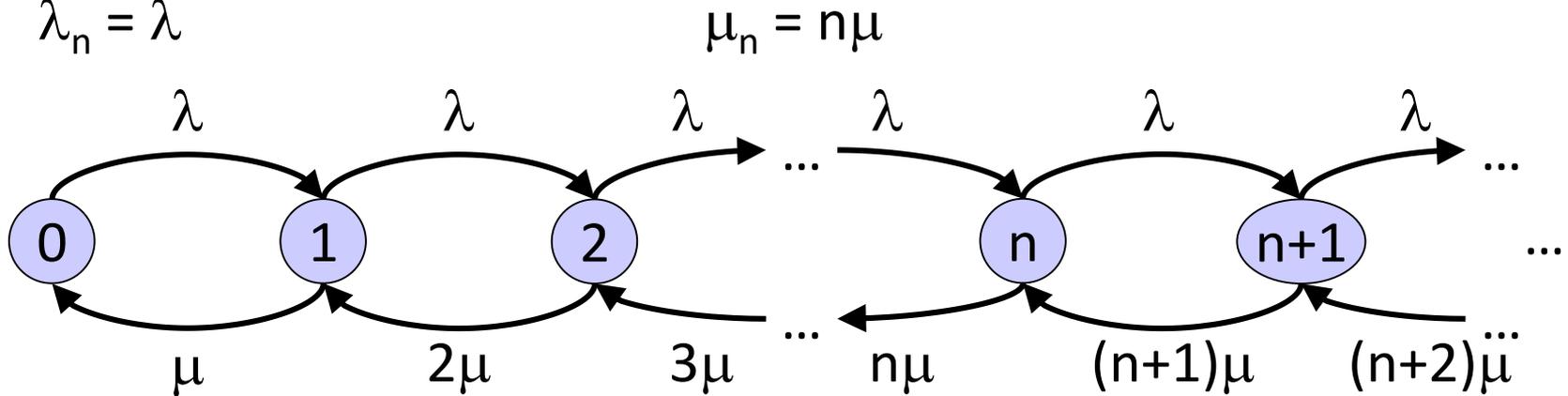
Posto $\rho = \lambda/(\mu m)$:

- $L = m\rho + \rho C/(1-\rho)$
- $\bar{W} = 1/\mu + C/[m\mu(1-\rho)]$
- $L_q = [\rho/(1-\rho)]C$
- $\bar{W}_q = C/[m\mu(1-\rho)]$

... vedi libro di testo [in lettura]...

Modello M/M/∞

- $\lambda_n = \lambda$



- $$\sum_{n=0}^{\infty} \pi_n = \sum_{n=0}^{\infty} \frac{(\lambda/\mu)^n}{n!} = e^{\lambda/\mu}$$

[si raggiunge sempre l'equilibrio]

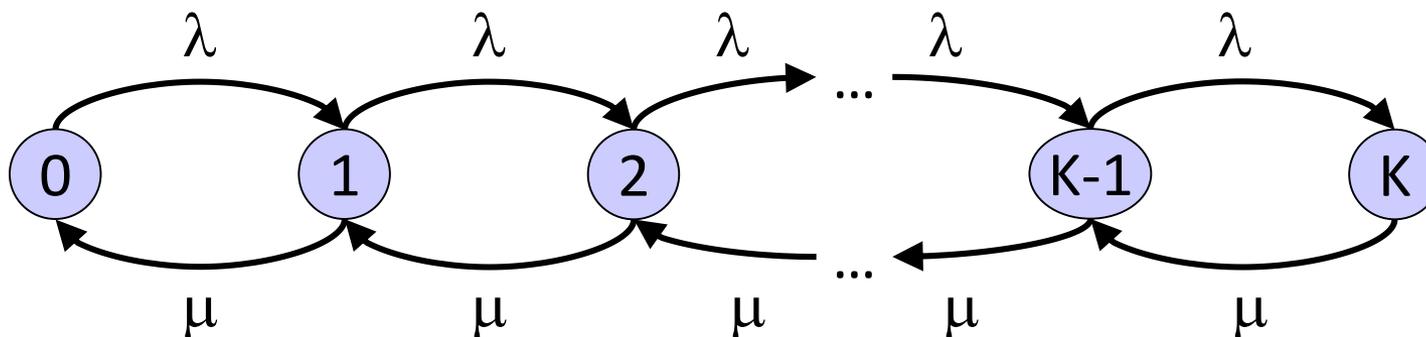
- $$P_0 = \left[\sum_{n=0}^{\infty} \pi_n \right]^{-1} = \left[\sum_{n=0}^{\infty} \frac{(\lambda/\mu)^n}{n!} \right]^{-1} = e^{-\lambda/\mu}$$

M/M/∞ : alcuni risultati

- $P_n = \frac{(\lambda/\mu)^n e^{-\lambda/\mu}}{n!}$ per $n = 0, 1, 2, \dots$
- N è una distribuzione poissoniana di parametro λ/μ
- $L = \lambda/\mu$ $\bar{W} = 1/\mu$ $L_q = 0$ $\bar{W}_q = 0$
- Applicazioni: buona **approssimazione** di sistemi dove è alta la probabilità di trovare un server libero, perchè $\rho \ll 1$

Modelli M/M/1/K

- Capacità del sistema finita e pari a K (num. max utenti nel sistema): gli utenti che trovano il sistema occupato lo abbandonano
- $\lambda_n = \lambda$ per $n = 0, 1 \dots K-1$, $\mu_n = \mu$ per $n = 1, 2 \dots K$ [0 altrimenti]



- Condizione di **ergodicità sempre** raggiunta (numero finito di stati)

$$P_0 = \left[\sum_{n=0}^K \pi_n \right]^{-1} = \left[\sum_{n=0}^K (\lambda/\mu)^n \right]^{-1} = \left[\frac{1-\rho^{K+1}}{1-\rho} \right]^{-1} \quad \text{per } \rho = \lambda/\mu \neq 1$$

$$P_n = \pi_n P_0 = \rho^n P_0 = \frac{\rho^n (1-\rho)}{1-\rho^{K+1}} \quad \text{per } n = 1, 2, 3 \dots K \text{ e } \rho \neq 1$$

M/M/1/K : alcuni risultati

- Se $\rho=1$, $P_n=1/(K+1)$, per ogni $n = 0, \dots, K$ (da cui ricavare, ad es. L)
- Probabilità che un **utente lasci** il sistema $\equiv P_K$ (se P_K è trascurabile, il sistema è approssimabile con capacità infinita)
- Si può ricavare (ad esempio)

$$L_q = \frac{\rho}{1-\rho} - \frac{\rho(1+K\rho^K)}{1-\rho^{K+1}} \quad \text{per } \rho \neq 1$$

- Leggi di Little valide ma attenzione: bisogna contare solo gli utenti che effettivamente entrano nel sistema, cioè il **tasso effettivo degli arrivi** $\lambda' = \lambda(1-P_K)$

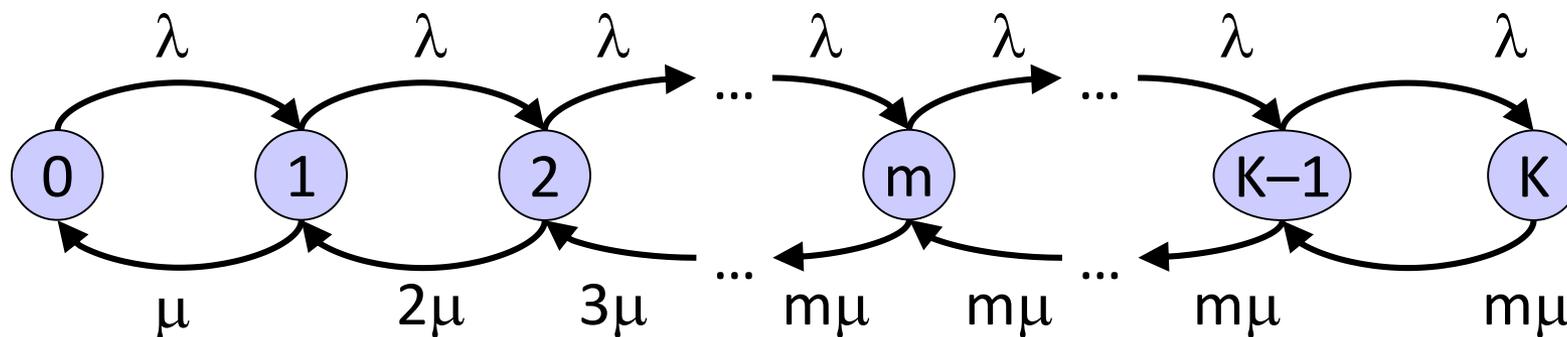
$$L = \lambda' W$$

$$L_q = \lambda' \bar{W}_q$$

$$\bar{W} = \bar{W}_q + 1/\mu$$

Modelli M/M/m/K

- $K \geq m$ $\lambda_n = \lambda$ per $0 \leq n \leq K-1$, $\mu_n = \min(n\mu, m\mu)$ per $1 \leq n \leq K$

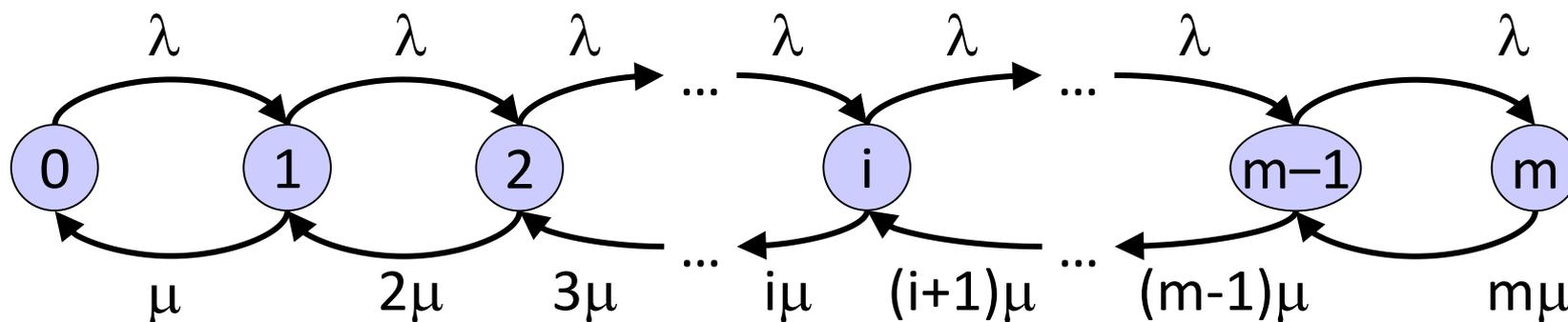


A partire da

$$P_0 = \left[\sum_{n=0}^m \frac{(\lambda/\mu)^n}{n!} + \sum_{n=m+1}^K \frac{(\lambda/\mu)^n}{m^{n-m} m!} \right]^{-1}$$

si possono ottenere espressioni in forma chiusa per L , L_q , \bar{W} , \bar{W}_q .
 Modello più generale e di maggiore utilità pratica: ad esempio, decidere la capacità della coda e il numero di serventi, dati limiti sul tasso di abbandono e sul tempo medio di attesa.

M/M/m/K : caso speciale K = m



- Si può studiare come caso particolare oppure (storicamente) risolvendo direttamente le equazioni di equilibrio, ottenendo

$$P_n = \frac{(\lambda/\mu)^n / n!}{\sum_{j=0}^m (\lambda/\mu)^j / j!} \quad \text{per } n=0, 1, 2, \dots, m$$

- La formula ottenuta per $P_m \equiv$ prob. del sistema pieno \equiv prob. di abbandono è nota come «**formula di perdita di Erlang**»

Possibili estensioni dei modelli M/M/.

- Tassi di arrivo e/o di servizio **dipendenti dallo stato** del sistema
 - Ad esempio: $\lambda_n = c_n \lambda$, $\mu_n = d_n \mu$ [$c_n=(n+1)^{-a}$, $d_n=n^b$, $a,b > 0$]
 - Permettono di modellare *balking* (abbandoni in funzione dello stato), o *reneging* (abbandoni dopo un timeout di attesa), o di aumento del tasso di servizio per smaltire lunghe code
- Processi **non** di nascita e morte
- Chiave per l'analisi: la **rappresentazione dello stato** (es. M/E_k/1)
- Raramente possono essere ottenute espressioni in forma chiusa per P_n (e L , \bar{W} etc.). In alternativa:
 - **soluzione numerica** delle equazioni di equilibrio
 - **simulazione**

Esercizi

1. Una postazione della Guardia di Finanza dotata di un unico sportello opera all'ingresso merci di un aeroporto. Il finanziere di turno ispeziona, in ordine di arrivo, i documenti dei camion in ingresso, con un tempo esponenziale di media 3 minuti. Secondo recenti indagini statistiche, il flusso dei camion è assimilabile a un processo di Poisson con tasso 15 camion/ora. Considerando che lo spazio è sufficiente per la sosta in attesa di tutti i camion in arrivo, determinare:
 - a. il numero medio di camion complessivamente presenti all'ingresso merci,
 - b. il numero medio di camion in attesa dell'ispezione,
 - c. la percentuale dei camion ispezionati senza attesa,
 - d. il tempo medio di attesa in coda prima dell'ispezione,
 - e. il tempo medio complessivamente speso dai camion all'ingresso.

Esercizi

2. Un segmento di ormeggio di un terminal marittimo per container è attrezzato con due gru identiche. Il carico/scarico di una nave ha una durata esponenziale con valore atteso 48 ore. Il tempo tra due arrivi consecutivi segue una legge esponenziale con media di 30 ore. Sapendo che una nave è servita da una sola gru, determinare:
- la probabilità che nessuna gru sia in funzione,
 - il numero medio di gru occupate,
 - Il fattore di utilizzazione di una singola gru e la corrispondente produttività giornaliera,
 - la probabilità che una nave non possa essere servita appena arrivata,
 - il tempo medio di attesa prima dell'inizio delle operazioni di carico/scarico,
 - il numero medio di navi presenti sul segmento di ormeggio.

Esercizi

3. Un servizio “dial-a-ride” riceve, mediamente, una chiamata all’ora. I mezzi impiegati per soddisfare ciascuna chiamata rientrano al deposito centrale dopo un tempo medio di servizio di 4 ore. Ipotizzando distribuzioni esponenziali per la frequenza delle chiamate e per i tempi di servizio, e che il numero di mezzi sia sufficiente per soddisfare tutte le chiamate senza attesa, determinare la probabilità che, su un lungo periodo di osservazione, in un dato istante, il numero di mezzi in circolazione sia superiore a 3.

Esercizi

4. Un DSP (Digital Signal Processor) elabora dei messaggi composti da pacchetti. Il tempo di interarrivo tra due messaggi consecutivi è esponenziale con valore atteso pari a 60 ms. Il DSP dispone di un buffer per 3 messaggi, che elabora uno per volta alla velocità di 20 pacchetti/ms, mentre i messaggi che arrivano quando il buffer è saturo vengono persi (il buffer contiene sia i messaggi in attesa, sia quelli in elaborazione). Sapendo che i messaggi contengono un numero di pacchetti distribuito esponenzialmente con media 960, determinare:

- a. la probabilità che il DSP sia in stato “idle”,
- b. la probabilità di perdere messaggi,
- c. il numero medio di messaggi nel buffer in attesa di essere elaborati,
- d. il numero medio di messaggi nel DSP,
- e. il fattore di utilizzazione del DSP (inteso come percentuale di tempo in cui il DSP elabora messaggi).