Capitolo 10

Traffico e file d'attesa

10.1 Il traffico di una sorgente

Il traffico generato da una sorgente viene caratterizzato attraverso una funzione binaria $z(t)$, $(z = 0, 1)$ che rappresenta i suoi stati di attività, $z = 1$ rappresentando lo stato per cui la sorgente è attiva.

Il traffico medio della sorgente su una finestra temporale T viene rappresentato come

$$z_T = \frac{1}{T}\int_0^T z(t)dt$$

e il valor medio assoluto come

$$\overline{z(t)} = \lim_{T\to\infty}\frac{1}{T}\int_0^T z(t)dt$$

e si dice misurato in *Erlang*.

Se l'attività della sorgente è periodica di periodo $\tau = x + y$ dove x e y rappresentano gli intervalli in cui la sorgente è, e non è, attiva , il traffico medio coincide con la media calcolata su τ:

$$\overline{z(t)} = \frac{x}{x+y} \tag{10.1}$$

In realtà, data la sua natura aleatoria, il traffico viene spesso modellato come un processo casuale $Z(t)$. Particolare importanza assume il caso in cui il traffico è un **processo rigenerativo** dove i punte di rigenerazione sono gli istanti di inizio (o fine) attività. Processi di questo genere sono **ergodici** e la media temporale (10.1) viene a coincidere con la media d'insieme $E[Z]$. Posto che la V.C X rappresenti la durata del periodo di inattività e la V.C. Y la durata del periodo di attività , da distribuzione del primo ordine stazionaria è, per quanto detto nel precedente Capitolo, il traffico medio diventa

$$S = E[Z] = P(Z = 1) = \frac{m_X}{m_X + m_Y} \tag{10.2}$$

che estende la (10.1).

Attiriamo l'attenzione sul processo **arrivi**, definito come il processo degli istanti in cui la sorgente diventa attiva. Nel modello appena visto tale processo rientra nella categorai degli **Eventi di Rinnovo** e il tasso d'arrivo risulta

$$\lambda = \frac{1}{m_X + m_Y} \tag{10.3}$$

e il traffico può essere espresso come

$$S = \lambda m_X \tag{10.4}$$

10.2 Il traffico multisorgente

Negli apparati trasmissivi di multiplazione il traffico generato dalle varie sorgenti si somma e in corrispondenza il processo traffico totale risulta la somma dei processi di traffico monosorgente componenti. Il traffico composto è un processo mutistato rappresentante il numero di sorgenti attive contemporaneamente.

Nel caso in cui processi componenti $Z_i(t)$ sono indipendenti, assunzione che faremo sempre, e hanno lo stesso valor medio di traffico, il traffico composto $Z(t) = \sum Z_i(t)$ risulta seguire una distribuzione istantanea binomiale. Infatti, la probabilità che ci siano k sorgenti attive su n che compongono il flusso è ugiale alla probabilità di successo in n prove con probabilità di successo pari a $\frac{m_X}{m_X+m_Y}$. Dunque

$$P(Z = k) = \binom{n}{k} \left(\frac{m_X}{m_X + m_Y}\right)^k \left(\frac{m_Y}{m_X + m_Y}\right)^{n-k} \qquad (0 \le k \le n) \tag{10.5}$$

Il valor medio è ovviamente la somma dei valori medi e risulta

$$E[Z] = n\frac{m_X}{m_X + m_Y} \tag{10.6}$$

Anche i flussi rappresentanti gli arrivi si sommano. In particolare si ha:

Teorema: (10.7)
Il tasso istantaneo λ di Eventi ottenuti dalla composizione di eventi di Rinnovo indipendenti, è la somma dei tassi degli Eventi componenti, ossia:

$$\lambda = \sum_{i=1}^{n} \lambda_i$$

Dimostrazione
La dimostrazione si basa sulla definizione di tasso ed è uguale alla dimostrazione della validità del Primo Assioma nella composizione di Eventi di Poisson. Si noti che la dimostrazione vale anche per eventi componenti non stazionari. ♣

In base a quanto visto la (10.6) può essere riscritta come

$$E[Z] = \lambda m_X \tag{10.8}$$

Supponiamo ora di far crescere il numero n degli eventi componenti, e di scalare l'asse dei tempi in modo inversamente proporzionale a n, in modo tale che il tasso totale resti invariato. Ciò significa che il tasso λ_i di ciascuna componente tende a zero in modo tale che:

$$\lim_{n\to\infty} \frac{n\lambda_i}{\lambda} \;=\; 1 \tag{10.9}$$

Si ha allora:

Teorema: (10.10)
Gli Eventi ottenuti dalla composizione di infiniti Eventi ricorrenti stazionari nel modo sopra specificato, costituiscono degli Eventi di Poisson.

Dimostrazione
Occorre solo mostrare che ciò che accade in due o più intervallini di misure $\Delta t_1,\ \Delta t_2, \ldots$ e non sovrapponentisi, è statisticamente indipendente (terzo assioma di Poisson, dato che i primi due sono già validi).

Ciò è di immediata intuizione se si considera che, nel caso di n finito, la dipendenza è dovuta alla dipendenza stessa che sussiste in ciascuno dei processi componenti, ma quando la frequenza media dei singoli componenti tende a zero, nessuna componente contribuisce, se contribuisce, con più di un evento, e ciò elimina la dipendenza.

La prova formale lungo questa strada è però complicata dal fatto che occorrono le d.d.p. congiunte almeno del secondo ordine.

Una dimostrazione più semplice è quella che calcola la d.d.p. del tempo d'interarrivo del prossimo evento composto e mostra che è esponenziale negativa, confermando così la proprietà di non memoria degli eventi composti.

Detto U il tempo d'attesa cercato, si ha

$$U = \min(U_1, U_2, \dots U_n)$$

dove U_i sono i tempi del paradosso dei tempi d'attesa relativi agli eventi componenti. I tempi di interarrivo sono costituiti dalla somma del tempo di attività X_i col tempo Y_i. Per valori λ_i sufficientemente piccoli i tempi X_i sono trascurabili mentre la $f_Y(y)$ tende a zero (uniforme) al tendere di $m_Y = 1/\lambda_Y$ a infinito:

$$f_Y(y) \simeq \frac{1}{m_Y} = \lambda_Y$$

Grazie al Teorema 5.41 si può allora scrivere,

$$f_{U_i}(u) = \lambda_i(1 - F_{Y_i}(u)) = \lambda_i - O(\lambda_i)u$$

essendo $O(\lambda_i)$ un termine tale che $\lim_{\lambda_i \to 0} O(\lambda_i)/\lambda_i = 0$. Inoltre

$$F_{Y_i}(u) = \lambda_i u - O(\lambda_i)u^2$$

E' poi:

$$P(U > u) = \prod P(U_i > u) = \prod (1 - F_{U_i}(u)) = \prod (1 - \lambda_i u - O(\lambda_i)u^2))$$

Grazie al limite (10.9) possiamo sostituire λ_i con λ/n e si ha:

$$P(U > u) = (1 - \frac{\lambda u}{n} - O(\lambda/n)u^2)^n$$

che nel limite fornisce

$$P(U > u) = e^{-\lambda u} \quad \clubsuit$$

Il modello appena visto, generato da un numero infinito di contributi, viene chiamato **modello di popolazione infinita**.

Nota (10.11)
La distribuzione del traffico con popolazione infinita è ancora una V.C. di Poisson con valor medio λm_X. Ciò deriva dal fatto che per ciascun flusso m_Y tende all'infinito secondo la (10.9) e dal fatto,

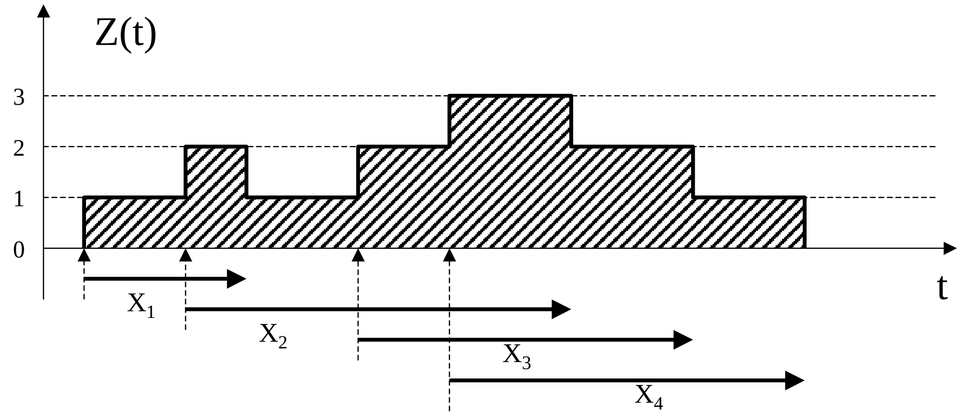

Figura 10.1:

già visto più volte, che la (10.5) con n che tende all'infinito con $E[Z] = \lambda m_X$ costante tende alla Poisson

$$P(Z = k) = \frac{(\lambda m_X)^k}{k!} e^{-\lambda m_X} \tag{10.12}$$

Il risultato (10.8) è generalizzabile a qualunque tipo di traffico grazie al

Teorema: Il risultato di Little (10.13)
Per un qualunque traffico che sia ergodico, si ha

$$E[Z] = \lambda m_X$$

Dimostrazione
Ci limitamo qui a dare la dimostrazione nel caso in cui $Z(t)$ sia un processo rigenerativo ricorrente positivo, che è ergodico (9.92), in cui gli istanti di rigenerazione sono quelli in cui $Z(t)$ si azzera. In realtà la dimostrazione ha portata più estesa, almeno quanta ne ha la (9.69).

La dimostrazione si basa sul fatto che per il processo $Z(t)$ si può scrivere

$$\sum_r rY_r = \int_C Z(t)dt, \tag{10.14}$$

$$\int_C Z(t)dt = \sum_{k=1}^{N} X_k \tag{10.15}$$

essendo C un qualunque ciclo del processo e N il numero di arrivi nel ciclo. La dimostrazione delle (10.14) e (10.15) è di natura deterministica e si ricava facilmente osservando la figura 10.1

Utilizzando la definizione di media e la (9.69) si ha

$$E[Z(t)] = \sum_r r\pi_r = \sum_r r\frac{E[Y_r]}{E[C]} = \frac{E\left[\sum_r rY_r\right]}{E[C]}$$

sostituendo le (10.14)e (10.15) si ottiene

$$E[Z(t)] = \frac{E\left[\sum_{k=1}^{N} X_k\right]}{E[C]} = \frac{E[N]E[X]}{E[C]}.$$

Osservando che $E[N]/E[C]$ definisce la frequenza media degli arrivi λ (va mostrato a partire dalla definizione osservando che anche il processo degli arrivi è rigenerativo a istanti discreti) si ottiene la tesi. ♣

10.3 Generalità sulle File d'attesa

Un sistema a fila d'attesa è costituito da un sistema dispensatore di servizi al quale giungono richieste di servizio (clienti) che vengono evase in base alla capacità del sistema. Le richieste in attesa di servizio possono attendere in file d'attesa finchè verranno processate, dopodichè abbandonano il sistema. Le richieste di servizio possono essere evase secondo le modalità più disparate, che verranno descritte quando necessario.

Le denominazioni usate sono mutuate dalla moderna trattazione della materia, che però si è sviluppata con la nascita del sistema telefonico dove i clienti sono le chiamate telefoniche e il servizio è il tempo di occupazione della chiamata.

Non tutti gli arrivi entrano necessariamente nel sistema. Esso può avere capacità infinita o finita, nel qual caso, i clienti che non possono essere accomodati vengono in genere ignorati.

Per le classificazioni dei sistemi, ci atterremo per quanto possibile alla notazione di Kendall del tipo $X_1/X_2/m/c$ dove X_1 indica il tipo di arrivi o le d.d.p. degli interarrivi:

$X = M$ (Markov) : interarrivi esponenziali (fra cui Poisson).
$X = E_r$: Erlang-k
...
$X = GI$: General Independent.

X_2 indica la ddp del tempo di servizio di significato come X_1; m indica il numero di serventi identici in parallelo, che lavorano a velocità costante; c indica il massimo numero di utenti ammissibili nel sistema (se finito).

Grandezze di interesse in un sistema di file d'attesa sono:

- Il processo occupazione $N(t)$, ossia il numero di clienti nel sistema al tempo t.
- Il processo occupazione della coda $N_c(t)$, ossia il numero di clienti in coda al tempo t.
- Il processo $N_s(t) = N(t) - N_c(t)$ numero di clienti nel servizio al tempo t.
- Il processo $N(t_j)$, campionamento di $N(t)$ negli istanti d'arrivo t_j^-.
- I processi V_i e W_i rispettivamente tempi trascorsi nel sistema e in coda del cliente numero i. Ovviamente $V_i = W_i + X_i$.
- Il processo C_i, tempo fra due istanti consecutivi in cui il sistema comincia a riempirsi dopo essersi svuotato (tempo di ciclo) e i processi B_i e I_i, durate degli intervalli per cui, nel ciclo i, il sistema è occupato (*busy period*) e vuoto (*idle period*) rispettivamente.
- Nel caso di capacità finita, interessano grandezze chiamate rispettivamente *time congestion* e *call congestion*. La prima è la probabilità che $N(t)$ sia al suo massimo valore e quindi riflette la percentuale di tempo per cui il sistema non può accettare altri clienti. La seconda la probabilità che $N(t_j)$ sia al suo massimo valore e quindi riflette la percentuale di clienti che trovano il sistema saturo (e quindi vengono rifiutati).

Descritto il sistema, i processi in gioco possono essere ergodici, cioè stabili, stazionari, o meno, in funzione della scelta di parametri quali λ e il tempo medio di servizio m_x. Un quesito importante è dunque per quali valori di questi parametri i processi in gioco sono ergodici.

10.3.1 Il risultato di Little

Come prima osservazione importante, notiamo che in un qualunque sistema in cui arrivino utenti alla frequenza media λ e ciascuno di essi si trattiene nel sistema un tempo V_i, se accade che i parametri sono tali che l'occupazione $N(t)$ è un processo ergodico, allora tali grandezze son legate dal *risultato di Little*:

$$E[N] = \lambda E[V]$$

La dimostrazione coicide con quella già fornita per il Teorema 10.13.

10.3.2 Il fattore di utilizzo

Chiamiamo fattore di utilizzo ρ del servente, la frazione di tempo (probabilità) per cui il servente lavora. Si noti che questa definizione coincide con la definizione di *traffico* già fornita in precedenza. Il fattore di utilizzo coincide dunque con il traffico smaltito dal servente.

Si noti che, posto che il servente serva un cliente alla volta, ρ è anche il numero medio di clienti che si trovano nel servizio e, grazie al Little's Result e considerando che il processo binario rappresentante le epoche di lavoro del servitore è sempre ergodico, si ha

$$\rho = \lambda_s m_x \tag{10.16}$$

dove λ_s è la frequenza media dei clienti serviti.

Per un sistema in cui si abbia un solo servente e in condizioni di ergodicità, si ha

$$\pi_0 = 1 - \rho \tag{10.17}$$

Si dimostra osservando che finchè il sistema non è vuoto, il servitore lavora. Si noti che questo corollario ci permette di esprimere direttamente la costante di normalizzazione π_0 e di ricavarla nel caso di coda infinita, per cui $\lambda_s = \lambda$.

Si ha inoltre:

$$\rho = 1 - \pi_0 = \frac{E[B]}{E[I] + E[B]} \tag{10.18}$$

essendo I e B le durate dell' idle period e del busy period rispettivamente.

Nel caso in cui gli arrivi siano di Poisson l' idle period ha valor medio $1/\lambda$ e, in base alle (10.18) e (10.16), è possibile esprimere il valor medio del busy period come:

$$E[B] = \frac{\lambda_s}{\lambda} \frac{m_x}{1 - \rho}.$$

immediatamente ottenibile nel caso di coda infinita. Ancora nel caso di coda infinita e m serventi equivalenti si ha:

$$\rho = \frac{\lambda m_x}{m}$$

e ciò perchè mediamente ogni servente riceve in media una frazione $1/m$ di clienti.

10.3.3 Il processo di nascita e morte generalizzato

Se consideriamo sistemi in cui gli arrivi e le uscite sono rigorosamente singoli, il processo occupazione $N(t)$ si comporta come un *processo di nascita e morte generale* (ossia non necessariamente markoviano) anche se il nome *processo di nascita e morte* è riservato solo a quelli markoviani.

Anche nei processi di nascita e morte generale, quando esiste una soluzione asintotica stazionaria π_i, si definiscono le grandezze λ_i^*, il tasso stazionario di passaggio del sistema da ia $i+1$, μ_i^*, il tasso stazionario di passaggio del sistema da ia $i-1$ (l'asterisco ricorda che non sono dati noti come nel caso markoviano). In generale con questi parametri si possono fare molte delle considerazioni già fatte per i processi di nascita e morte markoviani nel Capitolo 9 e per i processi markoviani in generale.

In particolare, se esiste una distribuzione asintotica stazionaria $\mathbf{\Pi}$, questa soddisfa (per definizione di distribuzione stazionaria) il sistema delle condizioni generali di equilibrio nella forma

$$\mathbf{\Pi} = \mathbf{\Pi P}^*(\Delta t) \tag{10.19}$$

o nella forma

$$\mathbf{\Pi Q}^* = 0 \tag{10.20}$$

dove le matrici con l'asterisco si riferiscono ai tassi di transizione sopra definiti.

Le condizioni di equilibrio possono essere poi tradotte nell'equilibrio dei flussi probabilistici, ed allora, per un processo di nascita e morte generale, deve ancora valere

$$\pi_i \; \lambda_i^* = \pi_{i+1} \; \mu_{i+1}^* \tag{10.21}$$

e sommando su tutta la distribuzione

$$\lambda_{in} = \lambda_{out} \tag{10.22}$$

ossia che il tasso medio di uscita deve uguagliare il tasso medio d'ingresso.

10.3.4 La distribuzione all'ingresso e all'uscita

Consideriamo la catena tempo discreta $N(t_j^+)$ dove t_j sono gli istanti in cui un cliente lascia il sistema. Dunque $N(t_j^+)$ è il numero dei clienti che restano dopo l'uscita.

Poniamo, in condizioni di stazionarietà:

$$r_i = P(N(t_j^+) = i)$$

Teorema: (10.23)
Le distribuzioni r_i e π_i sono legate dalla relazione:

$$r_i = \frac{\lambda_i^*}{\lambda_{in}} \; \pi_i = \frac{\mu_{i+1}^*}{\lambda_{in}} \; \pi_{i+1} \tag{10.24}$$

Dimostrazione
Si ha

$$r_i = P(N(t_j^+) = i) = P(N(t) = i + 1/\text{partenza in } [t, t + \Delta t])$$

e, applicando Bayes

$$r_i = P(\text{partenza in } [t, t+\Delta t]/N(t) = i+1) \; \frac{P(N(t) = i+1)}{P(\text{partenza in } [t, t+\Delta t])}$$

In condizioni di stazionarietà si ha:

$$P(N(t) = i+1) = \pi_{i+1}.$$

E' poi

$$P(\text{partenza in } [t, t+\Delta t]) = \lambda_{out}\Delta t = \lambda_{in}\Delta t$$

$$P(\text{partenza in } [t, t+\Delta t]/N(t) = i+1) = \mu^*_{i+1}\Delta t$$

Dunque si ottiene

$$r_i = \frac{\mu^*_{i+1}}{\lambda_{in}} \; \pi_{i+1}$$

e sfruttando le relazioni di equilibrio (10.21) si completa la prova della tesi. ♣

Analogamente posto

$$q_i = P(N(t_k^-) = i)$$

dove t_k sono gli istanti di entrata in coda e $N(t_k^-)$ sono i clienti visti da chi arriva, si ha il

Teorema: (10.25)
Le distribuzioni q_i e π_i sono legate dalla relazione:

$$q_i = \frac{\lambda^*_i}{\lambda_{arr}} \; \pi_i \tag{10.26}$$

Dimostrazione
Infatti è:

$$\begin{aligned} q_i &= P(N(t) = i/\text{arrivo in } [t, t+\Delta t]) = \\ &= P(\text{arrivo in } [t, t+\Delta t]/N(t) = i) \; \frac{P(N(t) = i)}{P(\text{arrivo in } [t, t+\Delta t])} \end{aligned}$$

e, per quanto già detto nella dimostrazione del teorema precedente, scende la tesi. ♣

Si noti che se non tutti i clienti che arrivano sono ammessi coda si ha $\lambda_{arr} \neq \lambda_{in}$. Altrimenti,

Corollario (10.27)
se tutti gli arrivi sono ammessi nel sistema si ha:

$$r_i = q_i$$

Corollario (10.28)
Se il sistema ha dimensione finita pari a Q e rifiuta quelli in eccesso si ha

$$r_i = \frac{q_i}{1 - q_Q} \qquad i \leq Q - 1$$

e discende dal confronto fra le due distribuzioni (10.24) e (10.26) ed osservando che $\lambda_{in} = \lambda_{arr}(1 - q_Q)$. In particolare si noti che la distribuzione sopra è anche quella vista da coloro che arrivano ed *entrano nel sistema*.

Nel caso gli arrivi siano di Poisson, ciscuno di questi è come un punto di ispezione casuale, in cui il sistema è stazionario e pertanto

Corollario (10.29)
Se gli arrivi sono di Poisson si ha

$$q_i = \pi_i$$

Si noti che gli istanti di tempo t_k^-, definiti come gli istanti di arrivo dei clienti, formano una corrispondenza uno a uno con il numero d'ordine d'arrivo dei clienti k. Pertanto l'interpretazione ergodica della q_i è la frazione *dei clienti* che trovano verificato al loro arrivo l'evento $\{N(t_k^-) = i\}$. Analoghe considerazioni valgono per il processo $N(t_k^+)$ considerato agli istanti di uscita dei clienti.

Il Teorema 10.25 fornisce poi il legame fra la *call congestion* q_Q e la *time congestion* π_Q nella forma:

$$q_Q = \frac{\lambda_Q^*}{\lambda_{arr}} \pi_Q \tag{10.30}$$

e mostra che nel caso di arrivi di Poisson, le due sono uguali.

10.4 I Sistemi Markoviani

10.4.1 M/M/1/Q

I sistemi M/M/1 sono caratterizzati da arrivi di Poisson con tasso globale λ e tempi di servizio indipendenti con valor medio

Come già visto nel capitolo precedente, $N(t)$ costituisce un sistema Markoviano di nascita e morte con $\lambda_i = \lambda$ e $\mu_i = \mu$. Si ha inoltre λarrivi $= \lambda$. La distribuzione ergodica è la geometrica troncata, e posto $\rho = \lambda/\mu$, si ha

$$\pi_i = \frac{1-\rho}{1-\rho^{Q+1}}\,\rho^i \qquad i = 0, 1, \ldots, Q$$

ed esiste per ogni ρ. Il fattore ρ è usato per comodità e non coincide qui con il fattore di utilizzazione del servente. E' anche chiamato fattore di utilizzo offerto, che è diverso da quello effettivo poichè $\lambda_{in} = \lambda(1-\pi_Q)$. L'utilizzazione effettiva è:

$$\rho_{in} = 1 - \pi_0 = \rho\frac{(1-\rho^Q)}{1-\rho^{Q+1}} = \rho(1-\pi_Q)$$

10.4.2 M/M/1/∞

Come il caso precedente ma con numero infinito di stati e dunque $\rho = \lambda/\mu$ è il vero fattore di utilizzo. Sappiamo che è ergodica per $\rho < 1$ ed allora

$$\pi_i = (1-\rho)\rho^i \qquad i = 0, 1, \ldots$$

Si ricava così, facilmente, che

$$E[N] = \frac{\rho}{1-\rho}$$

e con il Little's Result

$$E[V] = \frac{m_x}{1-\rho} = \frac{1}{\mu}\,\frac{1}{1-\rho}.$$

Poi

$$E[W] = E[V] - m_x$$

dunque

$$E[W] = m_x\frac{\rho}{1-\rho} = \frac{1}{\mu}\,\frac{\rho}{1-\rho}$$

Dalle formule viste si vede che è conveniente mettere in comune le risorse.

Esempio, se si hanno 2 code identiche, per ciascuna di queste si ha

$$E[N] = \frac{\rho}{1-\rho}$$

$$E[V] = \frac{1}{\mu}\,\frac{1}{1-\rho}$$

$$E[W] = \frac{1}{\mu}\,\frac{\rho}{1-\rho}$$

Se si fondono le code e i servizi ottenendo un servizio veloce il doppio si ha

$$E[N] = \frac{\rho}{1-\rho}$$

$$E[V] = \frac{1}{2\mu}\,\frac{1}{1-\rho}$$

$$E[W] = \frac{1}{2\mu}\,\frac{\rho}{1-\rho}$$

Si vede che benchè il numero di clienti nel sistema sia in media lo stesso i tempi d'attesa sono dimezzati.

Vediamo ora le d.d.p. dei tempi di attesa in coda e nel sistema.

Teorema: (10.31)
Il tempo di attesa nel sistema è una V.C. esponenziale negativa con media

$$E[V] = \frac{1}{\mu}\,\frac{1}{1-\rho}$$

Dimostrazione
Grazie al teorema della probabilità totale

$$f_V(x) = \sum_{i=0}^{\infty} q_i E_{i+1}(x)$$

dove $E_i(x)$ è la ddp Erlang-i somma di i tempi d'attesa esponenziali. Sostituendo si ha:

$$\begin{aligned} f_V(x) &= \sum_{i=0}^{\infty}(1-\rho)\,\rho^i\frac{(\mu x)^i}{i!}\mu\, e^{-\mu x} = \mu(1-\rho)e^{-\mu x}\sum_{i=0}^{\infty}\frac{(\mu\rho x)^i}{i!} = \\ &= \mu(1-\rho)e^{-\mu x}\, e^{\mu\rho x} = \mu(1-\rho)e^{-\mu(1-\rho)x} \qquad x \geq 0 \;\clubsuit \end{aligned}$$

Analogamente, è lasciato al lettore di mostrare che

Teorema: (10.32)
per il tempo d'attesa in coda si ha

$$f_W(x/W > 0) = \mu(1-\rho)e^{-\mu(1-\rho)x} \quad \clubsuit$$

10.4.3 Sistema M/M/m/Q

Serventi multipli indipendenti, ciascuno di tasso μ, e dimensioni massime del sistema Q.

Si tratta sempre di un processo di nascita e morte con

$$\lambda_i = \lambda$$

$$\mu_i = \begin{cases} i\ \mu & i \leq m \\ m\ \mu & i \geq m \end{cases}$$

Posto $\lambda/m\mu = \rho$ (fattore di utilizzo offerto al singolo servente), con facili passaggi si vede che la soluzione generale è:

$$\pi_i = \pi_0 (m\rho)^i / i! \qquad i \leq m$$

$$\pi_i = \pi_0\ \rho^i m^m / m! \qquad m \leq i \leq Q$$

10.4.4 Sistema M/M/m/m

Come il precedente con $Q = m$. Indicato il traffico offerto totale come

$$A = m\rho$$

si ottiene

$$\pi_i = \frac{A^i/i!}{\sum_{k=0}^{m} A^k/k!} \qquad i = 0, 1, \ldots, m.$$

La distribuzione sopra è una Poisson Troncata. La probabilità di rifiuto è $q_m = \pi_m$ con

$$\pi_m = \frac{A^m/m!}{\sum_{k=0}^{m} A^k/k!} = B_m(A) \tag{10.33}$$

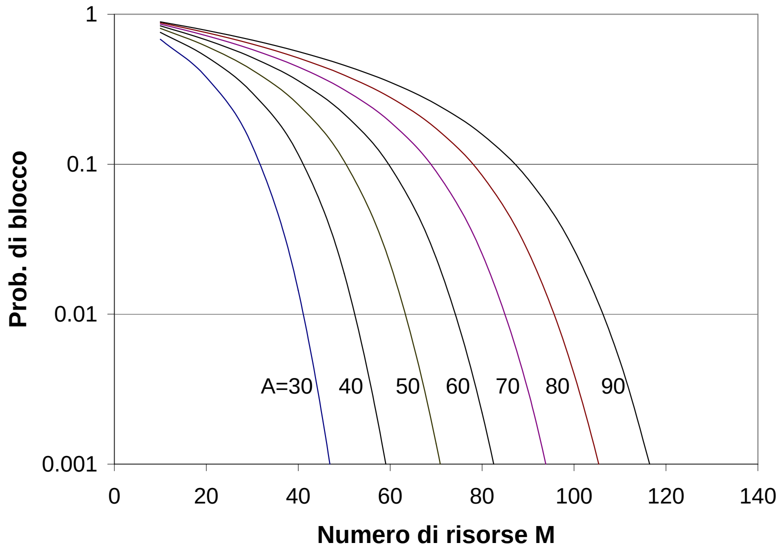

Figura 10.2:

ed è chiamata formula B di Erlang.

Si dimostra che vale la seguente formula ricorsiva

$$B_m(A) = \frac{A\ B_{m-1}(A)}{m + A\ B_{m-1}(A)} \tag{10.34}$$

Enunciamo qui l' importantissimo risultato, che non dimostreremo, che i risultati qui trovati, distribuzione di Poisson troncata e formula di Erlang, continuano a valere per sistemi M/G/m/m !

Anche per la proprietà appena citata, la B di Erlang ha sempre avuto un'importanza fondamentale per il dimensionamento delle giunzioni nei sistemi telefonici, che sono sistemi a perdita, in cui l'utente non può fare coda.

La Figura 10.2 mostra le probabilità di blocco in funzione del numero di risorse disponibili per diversi valori del traffico offerto A. Si vede che se si vuole una probabilità di blocco di 0.01 con $A = 30$ occorre predisporre un numero di risorse pari a 40, mentre con $A = 90$ le risorse necessarie sono 107, con un guadagno di 13 risorse rispetto al caso in cui i 90 Erlang fossero affrontati in tre gruppi disgiunti da 30. Come si vede, conviene mettere le risorse in comune. Questa è la ragione del perchè in telefonia si tende a concentrare le centrali telefoniche e le relative giunzioni, fino a raggiungere 50.000 − 100.000 utenze.

10.4.5 Sistema M/M/m/∞

Si ricava dal M/M/m/Q con $Q \to \infty$.

Affinchè sia ergodico si richiede che converga la $\sum \pi_i$ e in particolare la

$$\sum_{i=m}^{\infty} \pi_i = \pi_0 \frac{(m\rho)^m}{m!} \sum_{k=0}^{\infty} \rho^k$$

ne consegue la verifica che la condizione di ergodicità è $\rho < 1$.

In questo caso

$$\sum_{i=m}^{\infty} \pi_i = \pi_0 \, \frac{(m\rho)^m}{m!} \, \frac{1}{1-\rho}$$

e dunque

$$\pi_0 = 1/(\sum_{k=0}^{m-1} (m\rho)^k/k! + \frac{(m\rho)^m/m!}{1-\rho})$$

$$\begin{aligned} \pi_i &= \pi_0 \; (m\rho)^i/i! \qquad && i \leq m \\ \pi_i &= \pi_0 \; \rho^i m^m/m! \qquad && i \geq m \end{aligned} \tag{10.35}$$

La probabilità di trovare tutti i servitori occupati, e dunque di dover attendere, è, utilizzando $A = m\rho$:

$$C_m(A) = \sum_{i=m}^{\infty} \pi_i = \frac{A^m/m!}{(1-\rho) \sum_{k=0}^{m-1} A^k/k! + A^m/m!} \tag{10.36}$$

ed è chiamata formula C di Erlang. E' legata alla B di Erlang dalla formula

$$C_m(A) = \frac{B_m(A)}{1 - \frac{A}{m}(1 - B_m(A))} \tag{10.37}$$

La d.d.p. del tempo d'attesa, condizionata all'esistenza dell'attesa e cioè al caso $i \geq m$, si trova osservando che i servizi multipli equivalgono ad un sistema con servizio unico pari al minimo fra m esponenziali uguali a valor medio $1/\mu$. Tale servizio equivalente è ancora esponenziale con media ridotta di m, ma il fattore di utilizzo del servizio equivalente è lo stesso e dunque:

$$f_W(w/W > 0) = \mu m(1-\rho)e^{-\mu m(1-\rho)w}$$

$$E[W/W > 0] = \frac{1}{\mu \; m(1-\rho)}$$

Si vede che conviene comunque, rispetto alle due code separate, mettere in comune code e servizi anche se non si fondono i servizi.

10.4.6 Sistema M/M/∞/∞

Si ottiene dal caso precedente con $m \to \infty$. L' utilizzo ρ del singolo servente tende a zero e si ha

$$\pi_i = \frac{A^i}{i!}\, e^{-A} \qquad i = 0, 1, \ldots \tag{10.38}$$

ed il numero di persone nel sistema è distribuito secondo Poisson.

In realtà noi abbiamo già provato che la distribuzione è Poisson anche se il tempo di servizio è di tipo generale (10.12).

Si noti che il risultato precedente non può assolutamente essere esteso a un sistema GI/G/∞ perchè non è possibile scomporre il flusso degli arrivi in infiniti flussi elementari *indipendenti.*

10.4.7 Sistema M/M/1/ con popolazione finita M

Ciascun elemento della popolazione accede alla coda con un tasso λ ed ha un servizio indipendente con tasso μ. Il numero di clienti nel sistema è Markoviano ed ergodico per ogni λ e μ (sistema a dimensione finita). Più precisamente costituisce un processo di nascita e morte con

$$\begin{aligned} \lambda_i &= (M-i)\lambda \\ \mu_i &= \mu \end{aligned}$$

ed ha per soluzione (posto $\alpha = \lambda/\mu$)

$$\pi_i = \pi_0\ \alpha^i \frac{M!}{(M-i)!} \qquad i = 0, 1, \ldots, M$$

ricavando π_0 si ottiene

$$\pi_i = \frac{\alpha^i/(M-i)!}{\sum_{k=0}^{M} \alpha^k/(M-k)!} \qquad i = 0, 1, \ldots, M$$

nota come distribuzione di Engset.

Poichè gli arrivi al sistema, pur senza memoria, non costituiscono un processo di Poisson, le q_i saranno diverse dalle π_i. Dal Teorema 10.25 essendo $\lambda_i^* = (M-i)\ \lambda$ si ha

$$q_i = \frac{(M-i)\ \lambda}{\lambda_o}\ \pi_0\ \alpha^i \frac{M!}{(M-i)!} \qquad i = 0, 1, \ldots, M-1$$

Esprimendo in funzione di

$$q_0 = \frac{M\ \lambda}{\lambda_o}\ \pi_0$$

e rinormalizzando si ottiene infine

$$q_i = \frac{\alpha^i/(M-i-1)!}{\sum_{k=0}^{M-1} \alpha^K/(M-k-1)!} \qquad i = 0, 1, \ldots, M-1$$

Si noti che coincidono con le π_i di popolazione $M-1$.

10.4.8 Sistema M/M/m con popolazione finita $M > m$

E' ancora un processo di nascita e morte con

$$\lambda_i = (M-i)\ \lambda$$

$$\mu_i = \begin{cases} i\ \mu & i \leq m \\ m\ \mu & i \geq m \end{cases}$$

ponendo ancora $\alpha = \lambda/\mu$, si ha

$$\begin{array}{ll} \pi_i = \pi_0 \binom{M}{i} \alpha^i & i \leq m \\ \pi_i = \pi_0 \binom{M}{i} \alpha^i \frac{i!}{m!} m^{m-i} & i \geq m \end{array}$$

Si noti che, se $M = m$, la distribuzione stazionaria è la Binomiale:

$$\pi_i = \binom{m}{i} \left(\frac{\alpha}{1+\alpha}\right)^i \left(\frac{1}{1+\alpha}\right)^{m-i}$$

Si ha poi, la facile prova è lasciata al lettore,:

$$\begin{array}{ll} q_i = q_0 \binom{M-1}{i}\ \alpha^i & i \leq m \\ q_i = q_0 \binom{M-1}{i}\ \alpha^i \frac{i!}{m!} m^{m-i} & i \geq m \end{array}$$

ancora coincidono con le π_i di popolazione $M-1$.

10.4.9 Sistema M/M/m/m con popolazione finita $M > m$

Dal caso precedente, nel sistema senza coda a rifiuto si ha

$$q_i = \frac{\binom{M-1}{i} \alpha^i}{\sum_{k=0}^{m-1} \binom{M-1}{k} \alpha^k} \qquad i = 0, 1, \ldots, m-1$$

e la probabilità di rifiuto

$$q_m = \frac{\binom{M-1}{m} \alpha^m}{\sum_{k=0}^{m-1} \binom{M-1}{k} \alpha^k} \tag{10.39}$$

nota come formula di ENGSET. Anche per questi sistemi, come per gli analoghi con arrivi di Poisson questa, mostreremo che i risultati restano validi anche con servizi di tipo generale.

10.4.10 Sistema con arrivi scoraggiati

E' un processo di nascita e morte in cui il tasso di arrivo dipende dallo stato del sistema: il cliente entra nel servizio con probabilità $1/(n+1)$ se n è il numero dei clienti.

$$\lambda_n = \frac{\lambda}{n+1}$$

$$\mu_n = \mu$$

risulta

$$\pi_n = \left(\frac{\lambda}{\mu}\right)^n e^{-\lambda/\mu} \frac{1}{n!}$$

E' una distribuzione di Poisson con media λ/μ. Il fattore di utilizzazione è

$$\rho_s = 1 - \pi_0 = 1 - e^{-\lambda/\mu} =$$

e da questo possiamo ottenere la frequenza media degli ingressi

$$\lambda_{in} = \rho_s/m_x = \mu(1 - e^{-\lambda/\mu})$$

L'ultimo risultato ci consente di calcolare la distribuzione vista da chi entra come

$$q_i = \frac{\lambda_i}{\lambda_{in}}\pi_i = \left(\frac{\lambda}{\mu}\right)^{i+1} \frac{1}{(i+1)!} \frac{e^{-\lambda/\mu}}{1 - e^{-\lambda/\mu}}$$

Si noti invece che la distribuzione vista da chi arriva è coincidenta con la π_i dal momento che gli arrivi sono di Poisson.

10.4.11 M/M/1/∞ con arrivi in gruppo

Gli istanti di arrivo sono sempre di Poisson con tasso λ, però ad ogni istante di arrivo si presentano A clienti con probabilità $P(A = i) = g_i$[1].

Si hanno le seguenti equazioni di equilibrio ai nodi

$$\begin{aligned} &(\lambda + \mu)\pi_k = \mu\pi_{k+1} + \sum_{i=0}^{k} \pi_i \lambda g_{k-i} && k \geq 1 \\ &\lambda\pi_0 = \mu\pi_1 && k = 0 \end{aligned} \tag{10.40}$$

Per risolvere questa equazione ricorsiva alle differenze usiamo le trasformate z o funzioni generatrici

$$\Pi(z) = \sum \pi_k z^k \quad ; \qquad G(z) = \sum g_i z^i.$$

Moltiplichiamo le (10.40) per z^k e sommiamo per ogni k ottenendo

$$\lambda\pi_0 + (\lambda + \mu) \sum_{k=1}^{\infty} \pi_k z^k = \mu \sum_{k=0}^{\infty} \pi_{k+1} z^k + \lambda \sum_{k=0}^{\infty} \left(\sum_{i=0}^{k} \pi_i g_{k-i} \right) z^k$$

Sostituendo le espressioni delle trasformate si ottiene

$$\lambda\pi_0 + (\lambda + \mu)[\Pi(z) - \pi_0] = \frac{\mu}{z}[\Pi(z) - \pi_0] + \lambda\Pi(z) \cdot G(z)$$

e risolvendo per $\Pi(z)$ si ha infine

$$\Pi(z) = \pi_0 \frac{\mu(1 - z)}{\mu(1 - z) - \lambda z[1 - G(z)]} \tag{10.41}$$

[1] Supponiamo che $g_o = 0$; se così non fosse potremmo sempre porre $\lambda' = \lambda(1 - g_o)$ e $g'_i = \frac{g_i}{1-g_o}$ ed otterremo la stessa cosa

π_0 si determina imponendo $\Pi(1) = 1$. Con la regola di L'Hospital si ha

$$1 = \pi_0 \frac{-\mu}{-\mu + \lambda G'(1)}$$

dove $G'(1)$ corrisponde al numero medio di arrivi $E[A]$ nel gruppo e dunque

$$\pi_0 = 1 - \frac{\lambda G'(1)}{\mu} = 1 - \rho$$

E' lasciata al lettore la verifica che il caso con arrivi di Poisson rientra nel presente caso con $G(z) = z$

Nel caso di arrivi in gruppo di dimensione costante $r > 0$, per cui $G(z) = z^r$, si ricava che:

$$\Pi(z) = \left(1 - r\frac{\lambda}{\mu}\right) \frac{1 - z}{\frac{\lambda}{\mu} z^{r+1} - \left(1 + \frac{\lambda}{\mu}\right) z + 1}$$

Seguendo la falsariga di quanto detto in 5.2.4 si vede che la distribuzione vista dal primo cliente (primo ad entrare in coda) in ogni gruppo eguaglia la π_i. Analogamente si ha

$$q_{ik} = \frac{\pi_{i-k+1}}{\sum_{r=k-1}^{\infty} \pi_r} \qquad i \geq k-1;\ k \geq 1$$

dove q_{ik} è la probabilità che il k-esimo cliente del gruppo veda i clienti in coda davanti a sè. Ricordando i risultati visti in 3.8, la distribuzione vista dal generico cliente risulta essere

$$q_i = \sum_{k=1}^{\infty} \frac{\pi_{i-k+1}}{\sum_{r=k-1}^{\infty} \pi_r} \frac{\sum_{n=k}^{\infty} g_n}{E[A]}$$

In particolare il numero medio di clienti trovato nel sistema all' ingresso del generico cliente è legato al valor medio dell'occupazione dalla relazione

$$\sum_i i q_i = E[N] + \frac{E[A^2]}{2E[A]} - \frac{1}{2}$$

10.4.12 M/M/1/∞ con servizi in gruppo di tipo esaustivo

In questo caso il servente prende un numero massimo di r di clienti dalla coda, quando possibile, e li serve in gruppo mettendoci un tempo esponenziale con media $1/\mu$. Se nella coda ce ne sono meno di r alla partenza del nuovo periodo di servizio, il servitore continua ad accettare nel servizio nuovi arrivi finchè il numero di clienti nel servizio raggiunge r oppure il servizio cessa, svuotando il sistema.

E' molto simile al caso precedente con arrivi in gruppo di dimensione costante; la soluzione analitica non è però altrettanto semplice per quanto riguarda i passaggi matematici, ma porta ad un risultato estremamente semplice. Si può infatti far vedere che

$$\Pi(z) = \frac{1 - 1/z_0}{1 - z/z_0}$$

dove z_0 è l'unico zero del polinomio

$$\frac{\lambda}{\mu} z^{r+1} - \left(1 + \frac{\lambda}{\mu}\right) z^r + 1 \tag{10.42}$$

tale che $|z_0| > 1$. La distribuzione del numero di utenti nel sistemi è allora geometrica:

$$\pi_k = \left(1 - \frac{1}{z_0}\right) \left(\frac{1}{z_0}\right)^k$$

Nel caso di arrivi in gruppo e code finite può interessare la probabilità di avere almeno un rifiuto R o la frequenza media dei rifiuti λ_R:

$$R = \sum_{n=0}^{Q} \pi_n \sum_{Q-n+1}^{\infty} g_i = \sum_{n=0}^{Q} \pi_n \left(1 - \sum_{i=0}^{Q-n} g_i\right) = 1 - \sum_{n=0}^{Q} \pi_n \sum_{0}^{Q-n} g_i$$

$$\lambda_R = \lambda \sum_{n=0}^{Q} \pi_n \sum_{i=Q-n+1}^{\infty} g_i [i - Q + n]$$

10.4.13 M/M/1/∞ con servizi in gruppo di valore costante

E' la variante rispetto al precedente caso in cui il servizio non parte finchè il numero di clienti nel servizio non è r.

La soluzione viene derivata da quella del caso precedente con una considerazione degna di nota. Infatti, scrivendo le condizioni di equilibrio attorno agli stati di dimensione $r, r+1, r+2, \ldots$, si vede che vi compaiono *solo* le probabilità $\pi_{r-1}, \pi_r, \pi_{r+1}, \ldots$, e che tali equazioni sono le *stesse* che nel caso precedente. Ne consegue che le soluzione devono essere proporzionali, ossia:

$$\pi_k = C \left(\frac{1}{z_0}\right)^k \qquad k \geq r - 1$$

dove z_0 è lo stesso che nel caso precedente. Scrivendo le restanti condizioni di equilibrio, utilizzando in queste l'espressione sopra trovata ed imponendo la condizione di congruenza si trova, dopo qualche passaggio e sfruttando il fatto che z_0 è lo zero del polinomio (10.42), che:

$$\pi_j = \begin{cases} \frac{1}{r}\left(1 - (\frac{1}{z_o})^{j+1}\right) & 0 \leq j \leq r-1 \\ \\ \frac{\lambda}{r\mu}\left(1 - \frac{1}{z_o}\right)\left(\frac{1}{z_o}\right)^{j-r} & j \geq r-1 \end{cases} \tag{10.43}$$

Va notato che le due espressioni coincidono in $j = r - 1$, sfruttando ancora il fatto che z_0 è lo zero del polinomio (10.42).

10.4.14 M/E_r/1/∞

E' un sistema con arrivi di Poisson e servizi con d.d.p. Erlang-r. In questo caso il processo $N(t)$ non è markoviano perchè il servizio Erlang-r con $r > 1$ non ha tasso di uscita costante.

Però la distribuzione Erlang-r è equivalente ad r servizi esponenziali in cascata (stadi) ciascuno con valor medio m_x/r. E ciascuno di questi stadi ha tasso di uscita costante. Il tutto ridiventa dunque markoviano se si ingloba nello stato del sistema il numero di stadi di servizio $S(t)$ che il cliente ha già ottenuto al tempo t. Il processo da studiare è dunque il processo bidimensionale {N(t),S(t)}.

La pratica sopra illustrata è molto comune, nel senso che è spesso possibile allargare lo spazio degli stati della variabile di sistema in modo da far diventare markoviano il processo allargato.

Nel caso in oggetto non è però necessario ricorrere a una variabile di stato bidimensionale; infatti si può avere uno stato monodimensionale rinumerando opportunamente gli stati bidimensionali nel seguente modo:

$$J(t) = rN(t) - S(t)$$

La variabile $J(t)$ rappresenta il lavoro ancora da fare nel sistema, misurato col numero di stadi che restano ancora da completare, al tempo t, per esaurire le richieste (se nel frattempo non arriva più nessuno). Ogni cliente porta nel sistema r stadi che vengono esuriti singolarmente.

Il diagramma delle transizioni di questo sistema coincide con quello del sistema M/M/1/∞ con arrivi in gruppi di dimensione costante r e dove i tassi di discesa sono $r\mu$ (tasso di uscita dallo stadio). Ponendo nella soluzione (10.41) $G(z) = z^r$, si ottiene allora

$$\Pi'(z) = \pi'_0 \frac{r\mu(1-z)}{\lambda z^{r+1} - (\lambda + r\mu)z + r\mu}$$

Si trova poi $\pi'_0 = 1 - \lambda/\mu = 1 - \rho$.

I legami fra la distribuzione degli stadi trovata e la distribuzione dei clienti sono poi:

$$\pi_0 = \pi'_0$$

$$\pi_n = \sum_{j=(n-1)r+1}^{nr} \pi'_j$$

10.4.15 E_r/M/1/∞

Anche in questo sistema il processo N(t) non è Markoviano. Si può però considerare che il cliente in arrivo debba percorrere r stadi esponenziali prima di giungere al sistema. Introducendo la variabile supplementare $S(t)$ numero degli stadi già completati dal cliente in arrivo, la descrizione {N(t),S(t)} risulta markoviana.

Anche qui, come nel caso precedente, la variabile di stato può essere resa monodimensionale ponendo

$$J(t) = rN(t) + S(t)$$

Ciò equivale a supporre un cliente composto da r parti che raggiungono il sistema separatamente al tasso $r\lambda$. Ma il cliente non può essere servito finchè tutte le sue r parti non sono giunte. Allora J rappresenta l'ammontare delle parti presenti nel sistema. Quando un cliente ha completato il servizio, al tasso μ, allora tutte le sue r parti lasciano il sistema. La catena $J(t)$ è allora analoga alla $N(t)$ di un sitema $M/M/1$ con servizi a gruppi di valore costante r, la cui distribuzione, qui indicata con P_k è la (10.43), dove però λ è rimpiazzato da $r\lambda$.

La distribuzione del numero di clienti è allora:

$$\pi_n = \sum_{j=rn}^{r(n+1)-1} P_j$$

con

$$\pi_0 = 1 - \rho$$

Vediamo ora di calcolare le q_i. Conviene fare riferimento alla classificazione degli stati bidimensionale ponendo:

$$\pi_{ij} = P(N(t) = i, S(t) = j)$$

e

$$\pi_i = P(N(t) = i) = \sum_{j=0}^{r-1} \pi_{ij}$$

Con riferimento al Teorema 10.25, si ha $\lambda_{arr} = \lambda_{in} = \lambda$ e

$$\lambda_i^* = r\lambda \frac{\pi_{i,r-1}}{\pi_i}$$

Si ottiene dunque

$$q_i = r\pi_{i,r-1}$$

E' lasciata al lettore l' interpretazione del fatto che $\sum_i \pi_{i,r-1} = 1/r$.

10.4.16 Servizi Markoviani più generali

Una distribuzione E_r può essere utilizzata per approssimare un'altra distribuzione di pari valor medio $1/\mu$ e varianza $1/(r(\mu)^2)$.

Il rapporto

$$c = \frac{\sigma_x}{m_x} \tag{10.44}$$

prende il nome di coefficiente di variazione ed indica la variabilità della V.C. X. Per l'esponenziale $c = 1$, per l'Erlang-r $c = 1/\sqrt{r}$ e quindi la variabilità dell'Erlang-r è minore di 1; essa quindi è adatta ad approssimare d.d.p. con $c < 1$.

Una generalizzazione potrebbe essere quella di considerare più stadi esponenziali ciascuno però con valor medio diverso

$$B^*(s) = \frac{\mu_1}{s+\mu_1} \cdot \frac{\mu_2}{s+\mu_2} + \ldots \frac{\mu_r}{s+\mu_r}$$

ma si può vedere che il c è ancora < 1.

Un servizio con stadi esponenziali ma $c > 1$ lo si può ottenere mettendo in parallelo v servizi di durata media $1/\mu_i, i = 1, 2, \ldots, v$ tali che il nuovo cliente che entra nel servizio sceglie lo stadio i con probabilità α_i. La corrispondente d.d.p. è data da:

$$H_v(x) = \sum_{i=1}^{v} \alpha_i \; \mu_i \; e^{-\mu_i x}$$

e prende il nome di iperesponenziale. I sui parametri del primo e second'ordine sono:

$$m_x = \sum_{i=1}^{v} \frac{\alpha_i}{\mu_i}$$

$$E[X^2] = 2\sum_{i=1}^{v} \frac{\alpha_i}{\mu_i^2}$$

Un sistema M/H_r/1/∞ può essere descritto dalla coppia di numeri (n, i) dove n è il numero di utenti nel sistema ed $i, i = 1, 2, \ldots, r$ è lo stadio scelto dall'utente che si trova nel servizio.

Con ulteriori generalizzazioni del servizio con stadi serie-parallelo esponenziali è poi possibile approssimare a piacere una qualunque d.d.p.. File d'attesa con servizi di questo tipo possono essere studiate mediante l'uso di variabili supplementari.

10.5 Altri esempi Markoviani

Esempio *Ripetizioni di nuovi servizi* (10.45)
Consideriamo il sistema M/M/1 con la variante che, scaduto il tempo di servizio, il cliente decide di uscire con probabilità p, mentre con probabilità $1-p$ decide di ripetere il servizio con un nuovo tempo X indipendente dal precedente e la procedura si ripete.

Si tratta di un sistema Markoviano di nascita e morte con $\lambda_i = \lambda$ e $\mu_i = \mu p$. E' dunque come un sistema M/M/1 in cui il servizio ha durata $1/\mu p$. Lo stesso risultato può essere ottenuto osservando che effettivamente la somma di esponenziali negativi a tasso μ in numero casuale con distribuzione Geometrica con probabilità di successo p è ancora esponenziale negativa con tasso μp.

Cosa cambia se il cliente, con probabilità $1-p$ ritorna all'ingresso della coda? E se il cliente ripete il servizio con la stessa durata del precedente ?

Esempio *Servitore con vacanza* (10.46)
Consideriamo un sistema M/M/1 con la variante che il servitore alterna, a periodi di lavoro esponenziali negativi con tasso α, periodi di riposo esponenziali negativi con tasso β in cui ogni servizio viene sospeso, per essere ripreso nello stesso punto quando il servente ridiventa attivo.

Il processo occupazione non è markoviano perchè tale non è il tempo di servizio (la vacanza aggiunge un intervallo). Introducendo però una variabile ausiliaria che indichi se il servizio lavora (indice aggiunto l) oppure è a riposo (indice aggiunto v), si ottiene un processo bidimensionale markoviano con il diagramma degli stati mostrato in Figura 10.3.

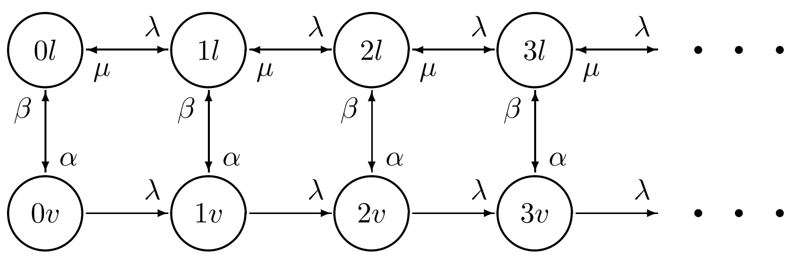

Figura 10.3: Diagramma di coda con servitore con vacanza

Il calcolo delle probabilità stazionarie di questo sistema è piuttosto complicato nonostante l'apparente semplicità del diagramma degli stati. Supponendo note queste probabilità stazionarie è immediato il calcolo della distribuzione del numero di utenti nel sistema. Infatti basta sommare le probabilità π_{il} e π_{iv} per ottenere la probabilità che ci siano i utenti nel sistema. Il valore μ_∞ che determina le condizioni di equilibrio è μ per la frazione di tempo per cui il servente lavora, ossia

$$\mu_\infty = \mu \frac{\beta}{\alpha + \beta}$$

Esempio *Serventi con velocità diversa* (10.47)
Consideriamo una variante al sistema M/M/2 in cui i serventi abbiano velocità diversa (tassi μ_1 e μ_2 con $\mu_1 > \mu_2$) e i clienti quando trovano entrambi i serventi liberi scelgono il più veloce. Il diagramma degli stati è mostrato in Figura 10.4.

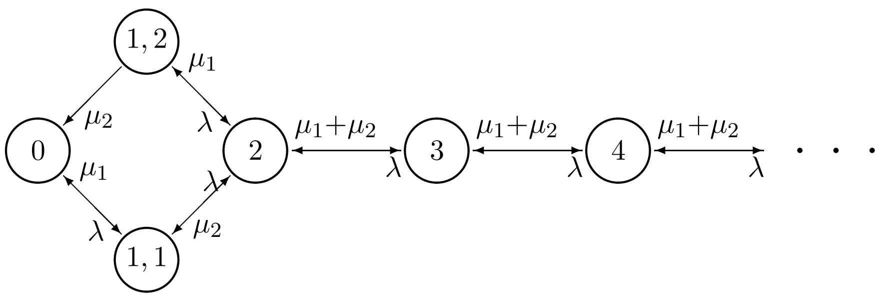

Figura 10.4: Diagramma di coda con servizi a velocità diversa

Gli stati $(1, 1)$ e $(1, 2)$ rappresentano il caso di un solo utente nel sistema che si trova nel servizio 1 o 2 rispettivamente. Si noti che dallo stato 2 in su il sistema è un sistema di nascita e morte, per cui si può scrivere:

$$\pi_i = \pi_2 \left(\frac{\lambda}{\mu_1 + \mu_2} \right)^{i-2}, \qquad i = 2, 3, \dots$$

Dall'equilibrio ai nodi si possono poi ricavare le espressioni di π_{11}, π_{12} e π_2 in funzione di π_0. I coefficienti di utilizzo dei due serventi si ottengono come

$$\rho_1 = 1 - \pi_0 - \pi_{12}$$
$$\rho_2 = 1 - \pi_0 - \pi_{11}$$

La frequenza media dei clienti serviti dai due servitori la si calcola dal Little's result come

$$\lambda_1 = \rho_1 \mu_1$$
$$\lambda_2 = \rho_2 \mu_2$$

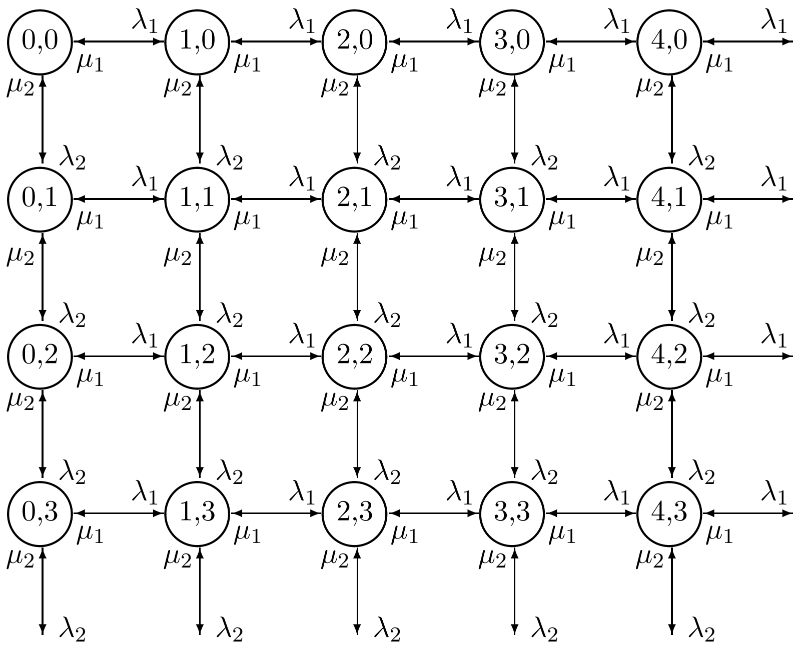

Figura 10.5: Diagramma a stati della coda condivisa

e, sempre dal Little's result si ha

$$E[X] = \frac{\rho_1 + \rho_2}{\lambda}$$

Si osservi che se è $\mu_1 = \mu_2$, allora gli stati $(1,1)$ e $(1,2)$ soddisfano le condizioni per essere trattati come un unico macrostato e la catena coincide allora con quella di un sistema M/M/2. In effetti, il processo $N(t)$ è lo stesso nei due casi.

Esempio *Coda condivisa* (10.48)
Consideriamo due sistemi M/M/1 distinti con servizi ed arrivi indipendenti e parametri $\lambda_1, \lambda_2, \mu_1, \mu_2$. E' possibile studiare il processo congiunto $(N_1(t), N_2(t))$ attraverso la catena di Markov bidimensionale mostrata in Figura 10.5.

D'altra parte, l'indipendenza delle code ci dice che la distribuzione congiunta è il prodotto delle due distribuzioni marginali. Si ha pertanto

$$\pi_{ij} = \pi_i \pi_j = (1-\rho_1)(1-\rho_2)\rho_1^i \rho_2^j \tag{10.49}$$

con $\rho_1 = \lambda_1/\mu_1$ e $\rho_2 = \lambda_2/\mu_2$.

Il lettore è sollecitato a verificare che effettivamente la soluzione sopra scritta garantisce l'equilibrio dei flussi. Questo, d'altra parte, è implicito nel fatto che lungo le colonne varia la sola π_j e che questa, come soluzione di un processo di nascita e morte, garantisce l'equilibrio dei flussi verticali nei due sensi. Analogamente per le righe e dunque l'equilibrio ai nodi è garantito dalla distribuzione vista, che è dunque la soluzione.

Si supponga ora che i due servizi abbiano la fila d'attesa in comune e che questa sia illimitata. La descrizione markoviana continua a essere quella precedente e così la soluzione.

Esempio (10.50)
Dei clienti arrivano secondo un processo di Poisson a due sistemi di code distinte $M/M/1$ con servitori indipendenti di velocità μ_1 e μ_2 rispettivamente. Al momento di entrare in una delle code, ciascun cliente sceglie quella col minor numero di clienti nel sistema, oppure sceglie a caso se le code sono uguali. Inoltre appena un cliente trova davanti a se un numero di persone maggiore di quelle contenute nell'altra coda, si trasferisce in quest'ultima.

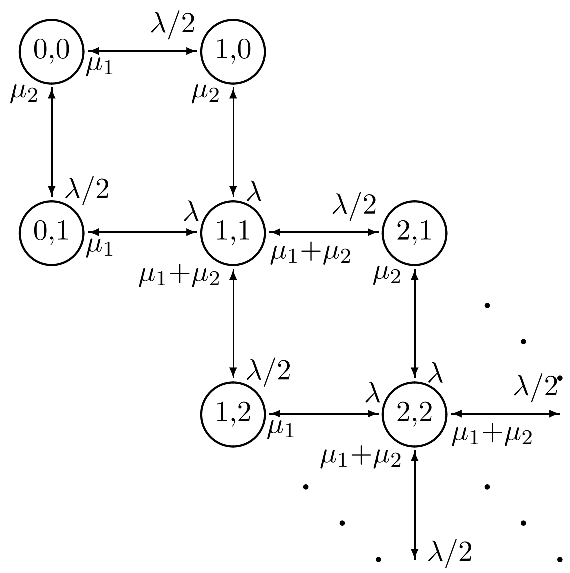

Figura 10.6: Catena delle code

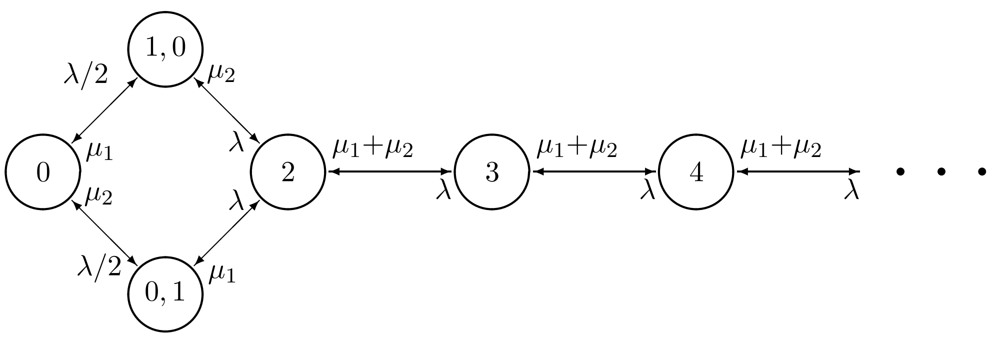

Figura 10.7: Catena delle code ridotta

E' possibile studiare il processo occupazione congiunta $(N_1(t), N_2(t))$ attraverso la catena di markov bidimensionale mostrata in Figura 10.6 La determinazione delle probabilità stazionarie di questo sistema può venire semplificata in quanto gli stati, tranne due, possono essere raggruppati in macrostati, rappresentanti il numero complessivo di clienti, nel modo mostrato in Figura 10.7. Gli stati $(0,1)$ e $(1,0)$ non possono venire conglobati in un unico stato in quanto le transizioni verso lo stato $(0,0)$ hanno tasso diverso. Le probabilità stazionarie possono essere ricavate come nell'esercizio precedente.

Le probabilità stazionarie degli stati della catena originale che non coincidono con la catena ridotta, possono essere ottenute con le equazioni di equilibrio di uno di questi stati insieme con le probabilità stazionarie della catena ridotta o con la relazione: $\pi_{(i,i+1)} + \pi_{(i+1,i)} = \pi_{2i+1}$.

10.6 Cenni sui sistemi M/G/1

Nei sistemi M/G/1, il tempo di servizio ha una d.d.p. $b(x)$ nota ma non vincolata a essere di un particolare tipo. In questi sistemi il processo $N(t)$ non è markoviano perchè i tassi di transizione non sono senza memoria. La probabilità che $N(t)$ scenda di uno, cioè che un servizio finisca, nel prossimo Δt, dipende da quanto tempo il cliente è stato nel servizio.

Anche questi sitemi sono studiabili in modo completo, anche se la distribuzione di equilibrio e quella dei ritardi è ricavabile spesso solo in forma di trasformata.

Vale la pena però di indicare un modo veloce ed istruttivo per ricavare il tempo medio d'attesa in coda.

Si considerino degli eventi rigenerativi con d.d.p. del tempo di intercorrenza pari alla d.d.p. $b(x)$ del tempo di servizio. Tali eventi corrispondono alle uscite dei clienti dal sistema in un sistema sempre pieno. Sia Z la V.C. *tempo d'attesa alla prossima uscita* o anche *tempo di servizio restante* a partire da un istante di ispezione casuale. Tale V.C. è quella descritta nel paradosso degli eventi di rinnovo. Essa è anche il tempo d'attesa alla prossima uscita sperimentato da un osservatore (o un cliente) di Poisson. Ricordiamo che si ha

$$E[Z] = \frac{m_x}{2} + \frac{\sigma_x^2}{2m_x}$$

Teorema: (10.51)
Il tempo W_i d'attesa in coda prima di ottenere il servizio, con disciplina FCFS, ha media pari a:

$$E[W] = \frac{\rho E[Z]}{1-\rho} \tag{10.52}$$

Dimostrazione
Il tempo medio d'attesa in coda di un cliente ha due componenti. La prima è dovuta al tempo $E[Z]$ di cui abbiamo già parlato, mediato però dal fatto che va conteggiato solo se il sistema non è vuoto. La seconda è dovuta ai clienti trovati in coda all'arrivo.

Detto N_c il numero di clienti trovati nella coda all'arrivo, si ha dunque:

$$E[W] = \rho E[Z] + E[N_c]m_x.$$

Usando il Little's result

$$E[N_c] = \lambda E[W]$$

si ottiene un' equazione in $E[W]$ che risolta conferma la tesi. ♣

Si noti che, essendo ρ la probabilità di trovare il servizio occupato, il tempo medio d'attesa condizionato la fatto che si attende è

$$E[W/W > 0] = \frac{E[Z]}{1-\rho} \tag{10.53}$$

Esplicitando $E[Z]$ ed utilizzando il coefficiente di variazione $c = \sigma_x/m_x$ si ottiene:

$$E[W] = \frac{\rho(1+c^2)}{2(1-\rho)}m_x \tag{10.54}$$

$$E[V] = m_x + \frac{\rho(1+c^2)}{2(1-\rho)}m_x \tag{10.55}$$

$$E[N] = \rho + \frac{\rho^2(1+c^2)}{2(1-\rho)} \tag{10.56}$$

Esempio *Coda M/D/1* (10.57)
Applicando i risultati trovati a una coda M/D/1 in cui la varianza del tempo di servizio è nulla, si trova applicando la (10.54)

$$E[W] = \frac{\rho}{2(1-\rho)}m_x \tag{10.58}$$

Si noti che questo tempo è esattamente la metà del caso M/M/1 con egual ρ.

Esempio (10.59)
In una coda M/G/1 arrivano due flussi di clienti. Il primo arriva con tasso λ_1 e richiede un tempo di servizio con media m_1 e varianza σ_1^2 mentre il secondo arriva con tasso λ_2 e richiede un tempo di servizio con media m_2 e varianza σ_2^2. Si calcoli il fattore di uilizzo ρ del servente, la probabilità ρ_1 di trovare nel servizio un cliente di tipo 1 e il valor medio del tempo d'attesa in coda. Si calcoli anche il valor medio del tempo trascorso nel sistema dai clienti di tipo 1 e il valor medio indistinto.

Il flusso d'arrivo può essere considerato in modo indistinto con frequenza d'arrivo $\lambda = \lambda_1 + \lambda_2$. In questo flusso, un utente scelto a caso è di tipo 1 con probabilità

$$\alpha_1 = \frac{\lambda_1}{\lambda}$$

Il tempo medio e la varianza del servizio risultano

$$m = \frac{\lambda_1}{\lambda} m_1 + \frac{\lambda_2}{\lambda} m_2$$

$$\sigma^2 = \frac{\lambda_1}{\lambda} \sigma_1^2 + \frac{\lambda_2}{\lambda} \sigma_2^2$$

Il fattore di utilizzo è

$$\rho = \lambda m = \lambda_1 m_1 + \lambda_2 m_2 = \rho_1 + \rho_2$$

Ricordando che il fattore di utilizzo è anche la probabilità di trovare il servente occupato da un cliente, si vede che la probabilità di trovare nel servizio un cliente di tipo 1 è ρ_1, e la la probabilità di trovare nel servizio un cliente di tipo 1 dato che il servizio è occupato è

$$\beta_1 = \frac{\rho_1}{\rho} = \frac{\lambda_1 m_1}{\lambda m}$$

diversa da α_1 se m_1 è diverso da m.

Il tempo medio d'attesa medio in coda non dipende dal tipo cliente che attende ed è dato dalla

$$E[W] = \frac{\rho E[Z]}{1 - \rho}$$

con

$$E[Z] = \frac{m}{2} + \frac{\sigma^2}{2m}$$

Il tempo medio trascorso nel sistema è

$$E[V] = \frac{\rho E[Z]}{1 - \rho} + m$$

mentre

$$E[V_1] = \frac{\rho E[Z]}{1 - \rho} + m_1$$

Si noti che risulta, ovviamente

$$[E[V] = \frac{\lambda_1}{\lambda} E[V_1] + \frac{\lambda_2}{\lambda} E[V_2] \tag{10.60}$$

10.7 Problemi da risolvere

P.10.1 Un sistema di attesa funziona nel seguente modo. Il servizio è inoperoso finché il numero di clienti diventa $N + 1$. Allora immediatamente inizia il servizio, esponenziale a valor medio $1/\mu$, e vengono rifiutati tutti i nuovi clienti finché il sistema si è svuotato e tutto ricomincia daccapo. Nell'ipotesi che i clienti arrivino secondo un processo di Poisson con frequenza media λ si calcoli:

a) la probabilità che nel sistema ci siano i persone,

b) la percentuale di tempo per cui il sistema è vuoto,

c) la percentuale di tempo per cui il servizio lavora.

P.10.2 Ad un posteggio di taxi, arrivano clienti secondo un processo di Poisson con frequenza media λ formando una coda se non trovano taxi disponibili. Questi ultimi arrivano al posteggio secondo un processo di Poisson con frequenza media $\mu > \lambda$ formando, se non trovano clienti, una coda finita di capacità N. Calcolare:

a) la probabilità che un cliente attenda,

b) il numero medio di clienti e taxi nelle rispettive code,

c) il tempo medio di attesa dei clienti.

P.10.3 Ad una fila di attesa i clienti arrivano secondo un processo di Poisson con frequenza media λ. Il servitore incomincia a funzionare solo quando ci sono N clienti in coda e cessa quando il sistema si è svuotato, riprendendo nuovamente solo quando i clienti in attesa sono N. Tracciare la catena di Markov che descrive il sistema e calcolarne le probabilità stazionarie.

P.10.4 Si consideri un sistema $M/M/1$ con la seguente variante. Nel caso in cui il cliente sia solo nel sistema questi continua a ripetere il servizio. Quando arriva un nuovo cliente quello in servizio esce, ma solo dopo aver completato il servizio stesso. Si chiede di calcolare:

a) la probabilità stazionaria del numero di clienti nel sistema,

b) il numero medio di persone presenti nel sistema,

c) il numero medio di ripetizioni del servizio del cliente che si trova da solo nel sistema.

P.10.5 Un sistema di pulitura rapida di vestiti da uomo è costituito da una sala d'attesa che si suppone infinita e da un servizio a cui i clienti accedono singolarmente quando è vuoto e che funziona nel seguente modo: due inservienti puliscono in parallelo uno la giacca l'altro i pantaloni impiegando tempi che sono v.c. indipendenti con la medesima d.d.p. esponenziale a valor medio $1/\mu$. Il cliente lascia il servizio dopo che entrambi i capi sono stati puliti. Supposto che i clienti arrivino secondo un processo di Poisson con frequenza media λ, si chiede di:

a) tracciare il diagramma degli stati markoviani del sistema,

b) dire qual è il massimo valore consentito di λ affinché il sistema sia stabile (cioè ammetta probabilità stazionarie),

c) rispondere ai punti a) e b) nel caso che i due inservienti abbiano rispettivamente $1/\mu_1$ e $1/\mu_2$ come valori medi dei tempi di pulitura.

P.10.6 N macchine si rompono indipendentemente ciascuna con tasso costante ed uguale a λ. Ogni riparazione viene compiuta in due fasi che si succedono secondo un preciso ordine e sono operate da due addetti, uno per ciascuna fase. Una macchina, oltre che funzionante, può essere in uno dei seguenti stati: in attesa della prima fase, ricevendo la prima fase, in attesa della seconda fase, ricevendo la seconda fase. Posto che le fasi abbiano durata esponenziale ed indipendente con tassi eguali a μ si dica come si può ottenere la distribuzione del numero di macchine rotte, la frequenza media di rottura, la probabilità che una macchina sia rotta.

P.10.7 Si consideri una coda M/M/1 in cui i clienti in arrivo entrano nel sistema con probabilità pari a

$$\frac{Q-i}{Q}$$

dove i rappresenta il numero di clienti trovati nel sistema. Si calcoli

1. la distribuzione stazionaria del numero di clienti
2. la distribuzione stazionaria del numero di clienti lasciati nel sistema da un cliente che esce.
3. il fattore di utilizzo del sistema.

P.10.8 Un cliente arriva ad una coda M/M/1. Si calcolino le probabilità che

1. sia il primo in un Busy Period;
2. sia l'ultimo in un Busy Period;
3. sia l'ultimo, essendo stato il primo, in un Busy Period.

P.10.9 (*Code in cascata*) Si considerino due sistemi M/M/1 con servizi indipendenti, dove i clienti che escono dal primo sistema entrano nel secondo. Detto λ il flusso d'ingresso nella prima coda, si mostri che la soluzione è sempre data dalla (10.49) con $\lambda_1 = \lambda_2 = \lambda$.

P.10.10 (*Code con priorità interruttiva*) Due flussi di clienti, con frequenza λ_1 e λ_1 richiedono servizi esponenziali negativi con tassi μ_1 e μ_2 rispettivamente. Essi arrivano ad un sistema con un unico servente dove fanno code distinte. I clienti della coda 1 hanno priorità interruttiva su quelli della coda 2. Ossia, i clienti della coda 2 non possono essere serviti finchè esiste un cliente nella coda 1. Inoltre, se arriva un cliente del primo tipo mentre il servizio ne serve uno del secondo tipo, quest'ultimo si ferma e torna in coda, lasciando il servizio al primo, e riprendendo il servizio solo quando la coda 1 si è svuotata.

1. Si tracci il diagramma degli stati della catena N_1, N_2 che descrive il numero di clienti dei due tipi nel sistema;
2. Si mostri che gli insiemi degli stati che hanno N_1 costante sono dei macrostati e si derivi la riapettiva distribuzione stazionaria.
3. Si trovino le probabilità stazionarie nel caso in cui esista solo la coda 1 e possa accomodare al più un cliente. I clienti che non possono essere accomodati spariscono.

4. Si calcoli la probabilità di trovare nel servizio un cliente di tipo 2.

P.10.11 Si consideri la variante di una coda con arrivi di Poisson e servizio Erlang-2 in cui vengono ammessi solo i clienti che arrivano mentre il servitore è vuoto o si trova nel primo stadio della Erlang-2. Gli atri vengono rifiutati.

1. Si trovi la distribuzione stazionaria del numero di clienti nel sistema.
2. Si calcoli il tempo medio trascorso nel sistema.
3. Si calcoli la distribuzione stazionaria del numero di clienti nel sistema lasciati dietro di sè da chi esce.

P.10.12 Una fila d'attesa $M/M/1$ è tale che la macchina che esegue il servizio deve essere rivista per manutenzione dopo aver servito N clienti. Il periodo di manutenzione, durante il quale il servizio non può venire erogato, è una V.C. esponenziale con valore medio $1/\alpha$. Si tracci il diagramma markoviano dello stato del sistema e si determini qual è la massima frequenza degli arrivi ammessa in condizioni di equilibrio (per questa domanda non serve la soluzione stazionaria della catena).

P.10.13 Si confronti il tempo medio d'attesa in una coda M/G/1 nei due casi in cui

1. il servizio sia una ripetizione di servizi esponenziali negativi indipendenti e p sia la probabilità di ripetizione;
2. il servizio sia una ripetizione di tempi di servizio tutti uguali al primo e questo sia estratto da una famiglia esponenziale negativa uguale al caso precente e p sia la probabilità di ripetizione.

P.10.14 Due macchine si rompono con tasso λ_1 e λ_2. L'addetto alle riparazioni è unico e, per le riparazioni delle due macchine, impega un tempo esponenziale negativo con tassi μ_1 e μ_2 rispettivamente. Se una macchina si rompe mentre l'addetto è occupato, l'altra macchina attende finchè si libera. Si calcolino le probabilità asintotiche che la prima macchina sia funzionante, in riparazione, in attesa di riparazione. Si ripeta il conto nel caso in cui $\mu_1 = \mu_2$.