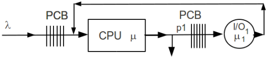
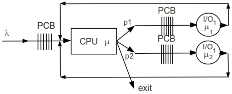
Modelli a code d'attesa dei sistemi operativi

Definizioni Preliminari

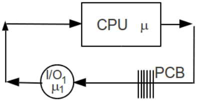
Topologie Tandem (1 dispositivo I/O



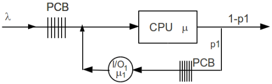
Tandem (2 dispositivi I/O)



Coda chiusa



Coda aperta



Definizioni Preliminari

- variabili aleatorie: il risultato di un esperimento dall'esito incerto
- probabilitá condizionate:

$$P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)$$

Formula di Bayes

$$P(A) = \frac{P(A \mid B)}{P(B \mid A)} P(B)$$

- Funzioni di probabilitá **Funzione cumulativa di distribuzione (Funzione di ripartizione)** :

$$F(X) = \mathrm{Prob}\{X \leq x\}$$

Quindi

$$\mathrm{Prob}\{x_1 \leq X \leq x_2\} = \int_{x_1}^{x_2} f(\xi)\, d\xi$$

Funzione densitá di probabilitá $f(x)$

$$\mathrm{F}(X) = \int_{-\infty}^{x} f(\xi)\, d\xi$$

Quindi
$\int_{-\infty}^{+\infty} f(\xi)\, d\xi = 1$
$\int_{x_1}^{x_2} f(\xi)\, d\xi = \int_{-\infty}^{x_2} f(\xi)\, d\xi - \int_{-\infty}^{x_1} f(\xi)\, d\xi = \mathrm{F}(X_2) - \mathrm{F}(X_1)$

- momenti statistici
  momento del k-esimo ordine attorno alla variabile w:

$$\int_{-\infty}^{+\infty} (x - w)^k f(x)\, dx$$

speranza matematica della variabile x:

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \lambda_x$$

momento del secondo ordine attorno a λ_x:

$$\mathrm{var}^2(X) = \int_{-\infty}^{+\infty} (x - \lambda_x)^2 f(x)\, dx = \sigma_X^2$$

scarto quadratico medio: $\sigma_X = \sqrt{\mathrm{var}^2(X)}$

La distribuzione di Poisson

Rappresenta la probabilitá che ci siano n arrivi all'istante t:

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Alcune caratteristiche:

Proprietá

Poisson implica che il tempo tra due arrivi sia distribuito con la distribuzione esponenziale $\mathrm{F}(t) = \mathrm{Prob}\,(T \leq t) = 1 - e^{-\lambda t}$.

Proprietá

La confluenza di n arrivi Poissoniani con frequenze d'arrivo $\lambda_1, \lambda_2, \ldots, \lambda_n$ *é ancora Poissoniana con frequenza* $\sum_{i=1}^{n} \lambda_i$.

Proprietá

La decomposizione di un processo Poissoniano con frequenza λ *in n rami d luogo a n processi Poissoniani con frequenza* $p_i \lambda$ *dove* p_i *é la probabilitá del ramo i-esimo.*

Proprietá

$E(n) = \lambda t$, $\sigma^2(n) = \lambda t$

La distribuzione esponenziale

$$f(x) = \lambda e^{-\lambda x}$$

Quindi

$$\mathrm{F}(X) = \int_{-\infty}^{x} f(\xi)\, d\xi = 1 - e^{-\lambda x}$$

Proprietá

La distribuzione esponenziale é l'unica distribuzione continua che dipende solo dall'istante iniziale

Proprietá

La distribuzione esponenziale gode della proprietá Markoviana, cioé l'assenza di memoria.

Proprietá

$E(n) = \frac{1}{\lambda}$, $\sigma^2(n) = \frac{1}{\lambda^2}$

La distribuzione gaussiana

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\lambda x}{\sigma}\right)^2}$$

Proprietá

La somma di n variabili aleatorie indipendenti e identicamente distribuite (con la stessa densitá), per n che tende a infinito (nella pratica per n sufficientemente grande), é una variabile aleatoria con distribuzione gaussiana. (teorema del limite centrale).

La distribuzione geometrica

$$\mathrm{Prob}(X = n) = P_n = \rho^n(1-\rho) \qquad\qquad 0 < \rho < 1$$

$$\mathrm{E}(n) = \sum_{n=0}^{\infty} nP_n = \frac{\rho}{1-\rho}$$

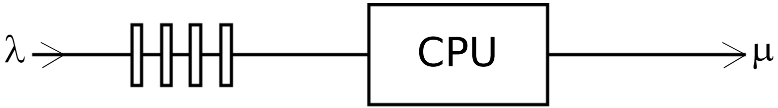
Processi di nascita e morte

- se i tempi di interarrivo hanno una distribuzione esponenziale, $f(x) = \lambda e^{-\lambda x}$ allora $Prob(n \text{ processi in } t) = P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$
- se i tempi di esecuzione hanno una distribuzione esponenziale, $f(x) = \lambda e^{-\mu x}$ allora $Prob(n \text{ esecuzioni in } t) = P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$

da cui

- $Prob(0 \text{ processi in } \Delta t) = 1 - \lambda \Delta t$
- $Prob(1 \text{ processi in } \Delta t) = \lambda \Delta t$
- $Prob(0 \text{ esecuzioni in } \Delta t) = 1 - \mu \Delta t$
- $Prob(1 \text{ esecuzione in } \Delta t) = \mu \Delta t$

Legge della conservazione del flusso

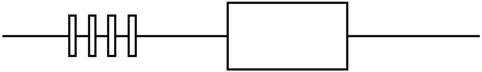


Frequenza arrivi $= \lambda$ [arrivi/secondo]
Frequenza partenze $= \mu$ [partenze/secondo]
Tempo medio di interarrivi $= 1/\lambda$ Tempo medio di interpartenze $=1/\mu$
Ipotesi: Tempo medio di servizio $\leq$ tempo medio di interarrivo

Allora: **frequenza partenze (μ) = frequenza arrivi (λ)**

I processi non vengono creati o distrutti e i tempi di attesa e servizio sono limitate
Nota: in caso di coda pesantemente caricata il tempo medio di servizio (esecuzione CPU) é l'inverso della frequenza partenze $= 1/\mu$

Teorema di Little



Frequenza di arrivi, tempo di interarrivo

Proprietá

Teorema di Little: Il numero medio di processi nel sistema operativo uguale alla frequenza degli arrivi moltiplicata per il tempo di attesa medio:

$$\overline{n} = \lambda T$$

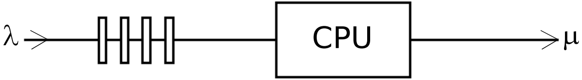
Tempo di attesa medio:

$$T = \frac{1}{\lambda}\,\overline{n}$$

coefficiente di utilizzazione del sistema a coda:

$$\rho = \frac{\mathrm{E}(CPU)}{\mathrm{E}(interarrivi)} = \frac{tempo\_medio\_di\_CPU}{tempo\_medio\_di\_interarrivi} = \frac{\lambda}{\mu}$$

Dimostrazione intuitiva di Little



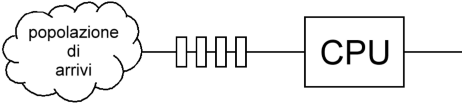
Numero medio di processi nel sistema = N; T = tempo attesa medio nel sistema; W = tempo medio di interpartenza=$1/\mu$

L'(N+1)-esimo processo arrivato deve aspettare che tutti gli altri N escano dal sistema:

$T = N * W = N * 1/\mu = N * 1/\lambda$

Cioé $T=N/\lambda$ ovvero $N = \lambda * T$

Modello M/M/1 di un Sistema Operativo



Prima Domanda: Qual'é la probabilitá che ci siano n processi nel sistema all'istante $t + \Delta t$?

Risposta: $P_n(t + \Delta t) = P_n(t)[1 - (\lambda + \mu)\Delta t] + P_{n+1}(t)\mu\Delta t + P_{n-1}(t)\lambda\Delta t$

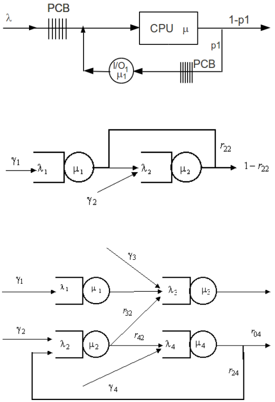
Seconda Domanda: Qual'é la probabilitá che ci siano n processi nel sistema all'istante t?

Risposta: $P_n(t) = \rho^n P_0(t)$

Terza Domanda: Qual'é la probabilitá che ci siano 0 processi nel sistema all'istante t?

Risposta: $P_0(t) = 1 - \rho$. Infatti $P_0(t) = 1/\sum \rho^n = 1/(1/(1-\rho))$ per $\rho < 1$. Quindi: $P_n(t) = \rho^n(1-\rho) \Rightarrow \mathrm{E}(n) = \frac{\rho}{(1-\rho)} \Rightarrow w = \frac{1}{\mu(1-\rho)}$ (w=tempo d'attesa medio dei processi).

Reti di code d'attesa: reti aperte



Reti di Jackson

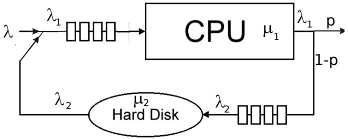
- La rete é composta da N nodi, $L_1 \dots L_N$, dove ogni nodo corrisponde ad una coda singola
- Gli arrivi dall'esterno alla rete ad un nodo i sono poissoniani con parametro λ_i
- I tempi di servizio di ciascun servente sono indipendentiI tempi di servizio di ciascun servente sono indipendenti e distribuiti esponenzialmente con parametro μ_i
- Dop aver completato il servizio presso il nodo i-esimo, ciascun cliente puó passare al nodo j-esimo con probabilitá $p_{i,j}$

Teorema di Jackson

In una rete aperta di Jackson di m nodi, la distribuzione di probabilitá a regime dello stato é data da
$Prob\{L_1 = n_1, L_2 = n_2, \dots, L_N = n_N\} = p_{n1} p_{n2} \dots p_{nm}$ dove
$p_{nj} = \rho_j^{nj}(1 - \rho_j)$.

Questo teorema afferma che la rete puó essere analizzata scomponendola in m sistemi a coda M/M/1 indipendenti, ovvero la rete si comporta come se ciascun nodo fosse un nodo isolato indipendente M/M/1 con frequenze medie di arrivo λ_i e frequenze medie di servizio pari a μ_i

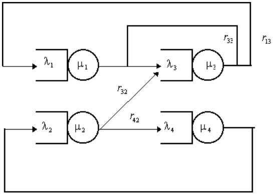
Esempio

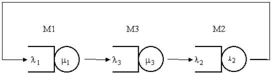


I processi possono peró uscire dal sistema con probabilitá p.
Se gli arrivi sono Poissoniani possiamo dire che:
λ é l'ingresso e l'uscita complessiva dal sistema (conservazione del flusso). Inoltre
$\lambda_1 = \lambda + \lambda_2 = \lambda + \lambda_1(1-p) \to \lambda_1 = \frac{\lambda}{p} \to \rho_1 = \frac{\lambda_1}{\mu_1} = \frac{\lambda}{p\mu_1}$.
$\lambda_2 = (1-p)\lambda_1 = \lambda\frac{1-p}{p} \to \rho_2 = \frac{\lambda_2}{\mu_2} = \frac{\lambda(1-p)}{p\mu_2}$.

Il numero dei processi nel sistema CPU é $N_1 = \frac{\rho_1}{1-\rho_1}$ e nel sistema disco é $N_2 = \frac{\rho_2}{1-\rho_2}$. Il numero complessivo é $N = N_1 + N_2$ e il tempo di attesa nel sistema é $T = \frac{N}{\lambda} = \frac{\rho_1}{\lambda(1-\rho_1)} + \frac{\rho_2}{\lambda(1-\rho_2)}$

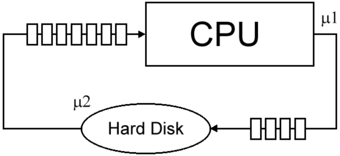
Reti di code d'attesa: reti chiuse





Coda ciclica

Caso piú semplice di coda chiusa. Code chiuse piú complesse si analizzano con i processi di Markov.



Attenzione: in questo caso il numero di processi costante e pari a N.
Attenzione: in questo caso $\rho = \frac{\mu_2}{\mu_1}$

Anche qui $P_n(t) = \rho^n P_0(t)$ ma in questo caso $\boxed{P_0(t) = \dfrac{1-\rho}{1-\rho^{N+1}}}$. Infatti

$P_0(t) = 1/\sum \rho^n = 1/(1-\rho^{N+1})/(1-\rho)$
Quindi $P_n(t) = \rho^n \frac{1-\rho}{1-\rho^{J+1}}$

Inoltre: l'utilizzazione della CPU é: $U = 1 - P_0(t) = \rho \frac{1-\rho^N}{1-\rho^{N+1}}$

Qualche esercizio

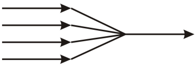
- Esempio 1
  Si consideri un S.O. non pre-emptive, senza iterazioni di I/O. Supponendo che il tempo medio di esecuzione dei processi sia di 0.2 secondi, trovare la frequenza massima di arrivo dei processi tale che il tempo di attesa nel sistema sia minore di 0.5 secondi.
- Esempio 2
  Si consideri un S.O. non pre-emptive, senza iterazioni di I/O. Si supponga che da alcune misure risulti che il tempo medio di esecuzione dei processi sia pari a 100/(frequenza clock CPU in MHz). Se il tempo medio di interarrivo e' di 3 secondi, qual'e' il clock minimo della CPU tale che il tempo d'attesa nel sistema sia minore di 0.5 secondi?

- Esempio 5
  Un Sistema Operativo rappresentabile nel seguente modo:

  La frequenza di elaborazione di CPU1 é di 1 processo/s e di CPU2 di 2 processi/s. In un certo istante la CPU2 é utilizzata e la probabilitá che ci siano PCB nella coda della 2 CPU é del 25%. Determinare il livello di utilizzazione di CPU1. Determinare inoltre il numero di processi in attesa della prima CPU.

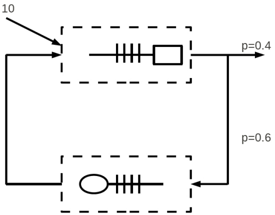
- Esempio 6
  Un Sistema Operativo possiede una CPU (che esegue in media di 50 processi al secondo) ed un disco (in media esegue 40 richieste al secondo). Sapendo che un processo ha uno spazio di indirizzamento medio di 500KB, che il Sistema Operativo non usa memoria virtuale, e che la CPU é utilizzata al 60%, determinare la minima quantit di memoria di cui il Sistema Operativo deve disporre. [ricorda: la probabilit di avere n processi in coda in un sistema a coda ciclica $\rho^n \frac{1-\rho}{1-\rho^{N+1}}$ dove ρ (frequenza esecuzione disco)/(frequenza elaborazione CPU)]

- Esempio 8
  I processi che arrivano ad un sistema operativo provengono da varie sorgenti, tutte poissoniane, come si vede in figura:



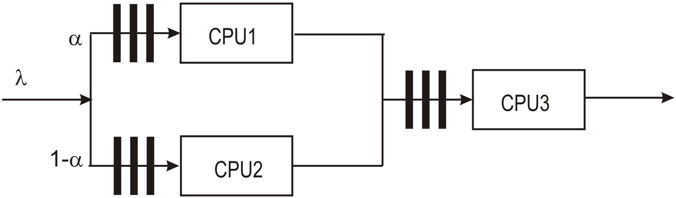
  I tempi medi di interarrivo sono di 10, 5, 3.33 e 2.5 secondi rispettivamente per le varie sorgenti. Qual'é la probabilitá che nel processo degli arrivi risultante arrivino 5 processi in 10 secondi?

- Esempio 10
  Si consideri un sistema di calcolo modellabile come in figura:



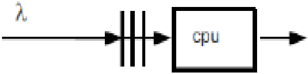
  Il disco ha una frequenza media di risposta μ_2 pari a 18.18 richieste/secondo e la CPU ha una frequenza media di esecuzione μ_1 pari a 33.3 esecuzioni/secondo. Arrivano processi dall'esterno con una frequenza di 10 processi al secondo e la probabilitá d'uscita dal sistema é di 0.4. Calcolare il numero medio di processi nell'intero sistema.

- Esempio 11
  Si consideri un sistema di calcolo modellabile come in figura:

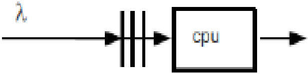


I processi arrivano al sistema e si biforcano con probabilitá α e $1 - \alpha$ sui due sottosistemi di CPU1 e CPU2. Quando i processi terminano l'elaborazione sulle due CPU, confluiscono sulla terza CPU per l'elaborazione finale. Sapendo che la frequenza d'arrivo al sistema é di 1.5 processi al secondo e che la frequenza di esecuzione della terza CPU é di 2 processi al secondo calcolare l'utilizzazione di CPU3. Sapendo che l'utilizzazione della CPU1 e CPU2 sono del 90% e 80% rispettivamente, calcolare la frequenza di servizio di CPU1 e CPU2 e il numero totale medio di processi nelle due code se $\alpha = 0.5$.

- Esempio 12
  In un sistema diskless, modellabile con la coda visualizzata in figura:

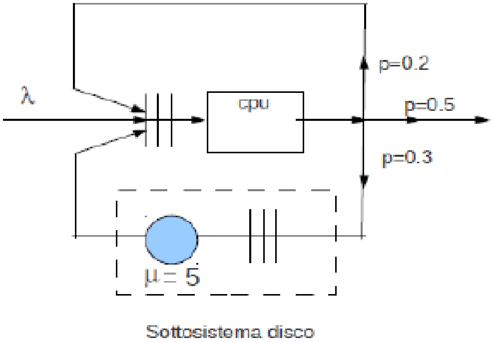


  un processo é dedicato alla visualizzazione del carico della CPU stimato tramite la probabilitá di utilizzazione. Il calcolo della utilizzazione CPU é realizzato attraverso misure della lunghezza della coda processi. Se in certi periodi vengono misurati 2, 18 e 3 numeri medi di processi in coda, quali sono i corrispondenti valori di utilizzazione CPU?
- Esempio 13
  Un server WEB, modellabile come in figura,



  viene letto da processi clienti i cui arrivi sono modellabili con una distribuzione di Poisson con frequenza pari a 9.98 processi al secondo. Una operazione in cpu viene eseguita in media ogni 100 ms. Qual'é il tempo medio di attesa dei clienti nel server?

- Esempio 14
  Un sistema operativo é modellabile con il modello descritto col seguente diagramma:



dove il disco soddisfa alle richieste con una frequenza media di 5 richieste/s e le probabilitá medie di suddivisione dei processi in uscita dalla CPU sono quelle indicate in figura come risulta da misure. Nell'ipotesi che le code non siano pesantemente caricate e in regime di stazionarietá, stimare la minima frequenza di arrivo λ tale che il sottosistema disco sia occupato piú del 60%.