UNIVERSITAT DE BARCELONA

Treball final de grau

GRAU DE MATEMÀTIQUES

Facultat de Matemàtiques i Informàtica
Universitat de Barcelona

Teoria de cues

Autor: Lluc Tresserras Pujadas

Director: Dr. Carles Rovira Escofet
Realitzat a: Departament de matemàtiques i informàtica

Barcelona, 20 de juny de 2019

Abstract

As we know queues are a common every-day experience. Queueing theory is the mathematical study of waiting lines or queues of a system.

This work first of all gives a brief introduccion of Markov process, birth-death process, and study some of the most important queueing models, such as the $M/M/1$ queue. Then deals with queueing networks to see how different queues act when they are interconnected. Finally, it shows a simulation of a simple queueing model with the aim of checking the analytic results observed.

Resum

Com sabem les cues són una experiència comuna del dia a dia. La teoria de cues és l'estudi matemàtic de les línies d'espera o cues dins d'un sistema.

Aquest treball primer de tot dona una breu introducció dels processos de Markov, dels processos de naixement i mort, i estudia alguns dels models de cues més importants com pot ser el model de cues $M/M/1$. Seguidament tracta les xarxes de cues per tal de veure com responen diferents cues interconnectades. Finalment mostra una simulació d'un model de cues simple amb l'objectiu de comprovar els resultats observats de manera analítica.

2010 Mathematics Subject Classification. 60K25, 90B22

Agraïments

Vull agrair al professor Carles Rovira per haver-me recolzat i donat un cop de mà durant tot el procés de creació del treball. També volia agrair a tota la meva família, companys i amics per haver-me animat i acompanyat durant tot el treball. Voldria fer un agraïment especial a la meva mare Montserrat, al meu pare Rafel, als meus dos germans Jan i Cesc, i a la Marina, que han estat al meu costat aquests últims anys, aconsellant-me i donant-me suport moral en tot moment.

Índex

1 Introducció

El progrés en la teoria de probabilitat aconseguit en l'últim segle respon a l'increment del paper de la probabilitat en el pensament científic.

Actualment, el camp modern de la informació tecnològica requereix una constant innovació basada en la modelització, anàlisis, disseny i finalment implementació de sistemes nous. En totes aquestes parts del procés és necessària la col·laboració d'un equip de treball format per enginyers, matemàtics, informàtics, físics, i molts d'altres.

La xarxa de telecomunicacions és un dels sistemes més complexos a estudiar, on l'eficiència i fiabilitat dels components juga un paper molt important. Per tenir un coneixement més clar del comportament dels diferents processos involucrats en aquests tipus de sistemes és necessari construir models matemàtics que descriguin els serveis estocàstics d'arribades aleatòries de clients. La teoria de cues és una de les eines matemàtiques més utilitzades a l'avaluació i modelització d'aquests sistemes.

Els processos estocàstics han jugat un paper clau en la modelització de sistemes i fenòmens que semblen variar de manera aleatòria per tal de poder estudiar-los i arribar a conclusions clares i necessàries per a la innovació constant.

Un procés estocàstic és bàsicament un procés que descriu l'evolució en el temps d'un fenomen aleatori. Des del punt de vista de les matemàtiques la teoria dels processos estocàstics es va assentar al voltant del 1950. Des d'aleshores, els processos estocàstics han esdevingut una eina molt comuna per als matemàtics, físics, enginyers, i el seu camp d'aplicació va des de la modelització del preu d'una acció fins a la geometria diferencial.

Les aplicacions i l'estudi de sistemes han inspirat a diversos matemàtics a proposar nous processos estocàstics. Alguns dels exemples més rellevants inclouen els procés de Weiner o moviment Brownià, usat per Louis Bachelier per tal d'estudiar la variació dels preus a la borsa de París, o el procés de Poisson usat per A. K. Erlang per estudiar el nombre de trucades en una central telefònica en un cert període de temps.

Aquests dos processos estocàstics són considerats com els més importants i rellevants a la teoria de processos estocàstics.

D'altra banda la teoria de probabilitats, els esforços dels probabilistes de la primera meitat del segle vint van ser dedicats a l'estudi de la independència: la suma de variables aleatòries i la seva distribució. Després d'haver treballat la independència, el tipus més simple de procés que podem estudiar és el procés de Markov. La propietat de Markov, assegura que l'estat futur d'un procés només dependrà de l'estat actual d'aquest. Tota informació anterior no és utilitzada.

El matemàtic Adrey Markov dona nom a aquests tipus de processos per haver sigut el primer a parlar-ne. Aquest va estudiar els processos de Markov a principis del segle vint publicant el seu primer article el 1906. El moviment brownià i el procés de Poisson són en concret dos processos de Markov.

La teoria de processos de Markov es va dividir en dos sub-teories, depenent de si el temps és discret o continu. En el cas de processos de Markov a temps discret es van anomenar cadenes de Markov. Aquestes tenen moltes aplicacions com a processos estocàstics en processos del món real, per exemple l'estudi del control de velocitat en vehicles de motor, teoria de cues o el creixement de la població d'una certa espècie animal.

Un model de cues és un cas particular de procés de Markov, en el qual es modelitzen de manera estocàstica les entrades i sortides de clients en un cert sistema que ofereix un servei.

El camp de la teoria de cues va començar a explotar-se a principis del segle vint de la mà d'Erlang, que va ser el primer en tractar amb problemes de congestió en el camp de la telefonia i telecomunicacions.

L'objectiu d'aquest treball és presentar els models més bàsics de la teoria de cues per l'anàlisi d'aquests, i intentar comparar-los per tal de veure com s'hauria de distribuir un servei per tal que fos el més eficient possible. El propòsit és veure com els models poden ser construïts i com poden ser analitzats. Cal destacar, que els models que tractarem estan basats en la suposició que les variables aleatòries involucrades en el sistema que modelitzarem són exponencialment distribuïdes i independents les unes de les altres. Cal destacar que complexitat dels sistemes treballats en el camp de la teoria de cues és bastant gran.

Aquest treball es distribueix en tres parts clarament diferenciades.

Primerament donarem una breu introducció als processos de Markov, on donarem algunes propietats importants d'aquests, ja que aquestes seran la base de l'estudi dels diferents models de cues que tractarem. Farem un incís important en els processos de naixement i mort, un cas particular de processos de Markov. És important destacar-los, ja que els sistemes de cues estan modelitzats com a processos de naixement i mort, on un naixement serà l'arribada d'un client i la mort en serà la sortida.

Seguidament ens endinsem en la part principal del treball en la qual donarem a conèixer alguns dels models de cues més importants i clàssics, juntament amb les propietats més bàsiques d'aquests. Cal destacar el model de cues $M/M/1$, ja que serà amb el que ens centrarem més. També intentarem comparar diferents models per veure de quines diferents maneres es pot distribuir un servei per tal que sigui el més òptim i eficaç possible.

Finalment parlarem de les xarxes de cues, uns sistemes concrets en els quals trobem cues simples interconnectades entre si. Intentarem donar resultats útils i buscar propietats que ens puguin descriure el procés i com fer-lo eficient.

2 Processos de Markov

En la teoria de probabilitat, un procés de Markov, és un tipus especial de procés estocàstic que té la propietat que la probabilitat de comportament futur està totalment definida si es coneix l'estat actual. Donada la situació present el futur és independent del passat i el procés manca de memòria. Definim aquest procés d'una manera més formal.

2.1 Definició d'un procés de Markov i propietats:

Considerem E un espai d'estats numerable, i $\{Y_t, t \geq 0\}$ una família de variables aleatòries tal que $\forall t \in [0, +\infty)$ aleshores $Y_t : \Omega \longrightarrow E$.

Definició: El procés estocàstic $\{Y_t, t \geq 0\}$ a valors en un espai d'estats E es diu que és un procés de Markov si:

$$P(Y_{t+s} = j | Y_u : u \leq t) = P(Y_{t+s} = j | Y_t) \, \forall t, s \geq 0.$$

Si a més, $P(Y_{t+s} = j | Y_t = i)$ és independent de $t \geq 0$, per tot $i, j \in E$ i $s \geq 0$, aleshores direm que $\{Y_t, t \geq 0\}$ és homogeni en temps. En aquest cas, podem definir el concepte de probabilitats de transició, que representen la probabilitat de passar d'un estat a un altre en un determinat temps. Utilitzarem la següent notació per definir-les: $P_s(i, j) = P(Y_{t+s} = j | Y_t = i)$.

Podem definir la família de matrius $P_t = \{P_t(i, j)\}_{(i,j) \in E \times E}$ la qual anomenarem matriu de probabilitats de transició. D'aquesta matriu en podem destacar tres propietats fonamentals que ens serviran per definir conceptes importants sobre els processos de Markov:

1. $P_t(i, j) \in [0, 1] \quad \forall i, j \in E,$
2. $\sum_{k \in E} P_t(i, k) = 1 \quad \forall i \in E,$
3. $\sum_{k \in E} P_t(i, k) P_s(k, j) = P_{t+s}(i, j) \quad \forall i, j \in E,$
4. $\lim_{t \to 0} P_t(i, j) = \delta_{ij} = \begin{cases} 0 & si\, j \neq i, \\ 1 & si\, j = i. \end{cases}$

Aquesta última propietat és una condició de continuïtat a la funció de transició.

2.1.1 Equació diferencial Kolmogorov

En aquest apartat, intentarem obtenir dos conjunts d'equacions diferencials que ens serviran per definir $P_t(i, j)$. Començarem veient algunes observacions interessants.

Observació: Per tot $i, j \in E$, la funció:

$$P_{\cdot}(i, j) : [0, \infty) \longrightarrow \mathbb{R},$$

és continua.

Lema 2.1. *Sigui $H : \mathbb{R} \longrightarrow \mathbb{R}$ una funció subadditiva amb $H(t) = 0$, $\forall t \leq 0$, llavors es compleix que:*

$$\lim_{t\to 0} \frac{H(t)}{t} = \sup_{0\leq t<\infty} \frac{H(t)}{t} = c,$$

on $c \in [o, \infty]$.

Demostració. Remarquem que una funció $H : \mathbb{R} \longrightarrow \mathbb{R}$ és subadditiva si i només si $H(t+s) \leq H(t) + H(s)$, $\forall s, t \in \mathbb{R}$.

Comprovem que $H(t) \geq 0$ per qualsevol $t \geq 0$. Suposem doncs, que existeix un $t_0 > 0$ que compleixi que $H(t_0) < 0$. Aleshores:

$$0 = H(-t_0) = H(t_0 + (-2t_0)) \leq H(t_0) + H(-2t_0) < 0,$$

i per tant tenim una contradicció.

Considerem ara $c' < c := \sup_{0\leq t<\infty} \frac{H(t)}{t}$. Per tant, $\exists s_0 > 0$ tal que $c' < \frac{H(s_0)}{s_0}$. Per qualsevol $t > 0$, tenim la descomposició $s_0 = nt + \delta$ amb $n \in \mathbb{N}$, $0 \leq \delta < t$.

Aleshores, tenim que:

$$c' < \frac{H(s_0)}{s_0} \leq \frac{H(nt) + H(\delta)}{s_0} \leq \frac{nH(t)}{s_0} + \frac{H(\delta)}{s_0} = \frac{nt}{s_0}\frac{H(t)}{t} + \frac{H(\delta)}{s_0}.$$

Si seguidament el següent pas que fem és:

$$t \longrightarrow 0,$$

$$n \longrightarrow \infty,$$

$$\delta \longrightarrow 0,$$

i prenem que $\frac{nt}{s_0} \longrightarrow 1$, obtenim que:

$$c' < \liminf_{t\to 0} \frac{H(t)}{t} \leq \limsup_{t\to 0} \frac{H(t)}{t} \leq \sup_{0\leq t<\infty} \frac{H(t)}{t} = c.$$

Per tant, $\exists \lim_{t\to 0} \frac{H(t)}{t} = c$. □

Proposició 2.2.

$$(a)\forall i \in E \ , \ \lim_{t\to 0} \frac{1 - P_t(i,j)}{t} := q_i \in [0, \infty],$$

$$(b)\forall i, j \in E, \ i \neq j \ , \ \lim_{t\to 0} \frac{P_t(i,j)}{t} := q_{i,j} < \infty.$$

Demostració. Per donar una idea de la demostració provarem el primer apartat.

Construïm una funció $H : \mathbb{R} \longrightarrow \mathbb{R}$ de manera que està definida com:

$$H_i(t) = \begin{cases} 0 & \text{si } t < 0, \\ -\ln P_t(i,i) & \text{si } t \geq 0. \end{cases}$$

Veiem clarament, que $H_i(0) = -\ln P_0(i,i) = 0$, i usant la tercera propietat dels processos de Markov que hem definit anteriorment, veiem que $P_{s+t}(i,i) \geq P_s(i,i)P_t(i,i)$. Per tant, tenim que:

$$H_i(s+t) \leq H_i(s) + H_i(t)$$

Per tant, aplicant el Lema tècnic anterior, veiem que $\exists \lim_{t\to 0} \frac{H_i(t)}{t} := q_i$.
Aleshores,

$$\frac{1-P_t(i,i)}{t} = \frac{1-e^{-H_i(t)}}{t} = \frac{1-e^{-H_i(t)}}{H_i(t)} \frac{H_i(t)}{t},$$

Finalment, tenim que:

$$\lim_{t\to 0} \frac{1-e^{-H_i(t)}}{H_i(t)} = 1,$$

$$\lim_{t\to 0} \frac{H_i(t)}{t} = q_i,$$

Per tant, obtenim que:

$$\lim_{t\to 0} \frac{1-P_t(i,i)}{t} = q_i.$$

□

Observació: Es pot observar, que:

$$q_i \geq \sum_{j\neq i} q_{ij}.$$

Donat $i \in E$ diem que és estable(regular) si $q_i < \infty$ i $q_i = \sum_{i\neq j} q_{ij}$.

Definició: La matriu $A = \lim_{t\to 0^+} \frac{P_t - Id}{t}$ s'anomena el generador infinitesimal del procés de Markov.

Proposició 2.3. *Sigui $E = \mathbb{N}$ i suposem que tots els estats són regulars:*
$\forall i,j \in E$ *i* $t \geq 0$,

$$P_t'(i,j) = \sum_{k\neq i} q_{ik}P_t(k,j) - q_iP_t(i,j) = \sum_{k\in E} P_{0^+}'(i,k)P_t(k,j).$$

Aquesta fórmula s'anomena ***Kolmogorov's backward equation(KBE)****, i es pot escriure com:*

$$\frac{d}{dt}P_t = AP_t.$$

Si suposem que $\sum_{k\in E} P_t(i,k)q_k < \infty$, tenim que:

$$P_t'(i,j) = \sum_{k\neq i} P_t(i,k)q_{kj} - P_t(i,j)q_j = \sum_{k\in E} P_t(i,k)P_{0^+}'(k,j).$$

En aquest cas, la fórmula s'anomena ***Kolmogorov's forward equation (KFE)*** *i es pot escriure com:*

$$\frac{d}{dt}P_t = P_tA.$$

2.2 Exemples de processos de Markov

Podem destacar alguns exemples interessants de processos de Markov que ens poden donar diversos resultats interessants.

2.2.1 Procés de Poisson de paràmetre λ

Un procés estocàstic de comptatge $\{N(t), t \geq 0\}$, és un procés de Poisson de paràmetre λ si:

- $N(0) = 0$.
- El procés té increments independents.
- $\forall s, t \geq 0$ es compleix que: $N(t+s) - N(t) \sim Poiss(\lambda s)$.

Aquesta última propietat, ens permet dir que un procés de Poisson té increments estacionaris.

Seguidament, considerem com a espais d'estats $E = \mathbb{N}$. En aquest espai d'estats el fet que els increments d'aquest procés siguin independents, ens permet dir que:

$$P(N(t+s) = j | N(u); u \leq t, N(t) = k) = P(N(t+s) = j | N(t) = k).$$

Per tant, hem comprovat que aquest procés té la propietat de Markov, i per tant, que un procés de Poisson és un cas particular de procés de Markov.

Finalment, anem a veure quines són les probabilitats de transició, que usarem per construir la matriu de transicions:

$$P_s(i,j) = P(N(t+s) = j | N(t) = i) = \left\{ \begin{array}{cc} 0 & si\, j<i \\ e^{-\lambda s}\frac{(\lambda s)^{j-i}}{(j-i)!} & si\, j \geq i \end{array} \right. .$$

Alguns altres exemples de processos de Markov són els processos de naixement i mort, dels quals parlarem seguidament de manera més extensa, ja que un cas particular d'aquests són els sistemes de cues.

3 Processos de Naixement i Mort

Un cas particular dels processos de Markov és els processos de naixement i mort, l'espai d'estats del qual és $\mathbb{N}$. Aquests processos estan caracteritzats per la propietat que si hi ha una transició, llavors aquesta porta a un estat veí.

La descripció d'aquest procés és la següent: El procés es manté en un estat i durant un cert temps, seguint una distribució exponencial de paràmetre $\lambda_i + \mu_i$. Quan deixa aquest estat, el procés entra a l'estat $(i+1)$ o $(i-1)$.

La teoria de cues és una àrea d'aplicació força important dels processos de naixement i mort. S'ha provat que és molt útil en un rang molt ampli de disciplines des de xarxes informàtiques i telecomunicacions fins a la cinètica química i epidemiologia.

3.1 Definició de un procés de naixement i mort

Considerem un procés de Markov $\{Y_t, t \geq 0\}$ amb espais d'estats $E = \mathbb{N}$ amb probabilitats de transició $P_t(i,j)$.

A més, assumim que les probabilitats de transició compleixen les següents propietats:

1. $P_h(i,i+1) = \lambda_i h + o(h)$, $\forall i \in \mathbb{N}$.
2. $P_h(i,i-1) = \mu_i h + o(h)$, $\forall i \in \mathbb{N}$.
3. $P_h(i,i) = 1 - (\lambda_i + \mu_i)h + o(h)$.
4. $\mu_0 = 0$, $\lambda_0 \geq 0$, $\lambda_i, \mu_i \geq 0 \, \forall i \in \mathbb{N}$.

Podem veure, doncs, que la propietat (3) és equivalent a que $P_h(i,j) = o(h)$ si $|i-j| \geq 2$.

Els paràmetres λ_i i μ_i s'anomenen respectivament els índexs infinitesimals de naixement i mort del procés.

Tenint en compte com hem definit els processos de Markov a l'apartat anterior, utilitzant aquestes quatre propietats podem obtenir fàcilment:

$$q_i = \lim_{t \to 0} \frac{P_t(i,i) - 1}{t} = \lambda_i + \mu_i.$$

$$q_{ij} = \lim_{t \to 0} \frac{P_t(i,j)}{t} = \begin{cases} \lambda_i & \text{si } j = i+1, \\ \mu_i & \text{si } j = i-1, \\ 0 & \text{si } |i-j| \geq 2. \end{cases}$$

Clarament, tots els estats $i \in \mathbb{N}$ són regulars, ja que:

$$q_i = \sum_{i \neq j} q_{i,j} \ , \forall i.$$

Per tant, utilitzant la fórmula Kolmogorov's backward equation, podem escriure:

$$P_t'(0,j) = \lambda_0 P_t(1,j) - \lambda_0 P_t(0,j).$$

$$P_t'(i,j) = \mu_i P_t(i-1,j) - (\lambda_i + \mu_i) P_t(i,j) + \lambda_i P_t(i+1,j) \ , \forall i \geq 0.$$

Assumint que $\sum_k p_t(i,k) q_{kk} < \infty$, també podem escriure la Kolmogorov's forward equation.

$$P_t'(i,0) = -\lambda_0 P_t(i,0) + \mu_1 P_t(i,1),$$

$$P_t'(i,j) = \lambda_{j-1} P_t(i,j-1) - (\lambda_j + \mu_j) P_t(i,j) + \mu_{j+1} P_t(i,j+1).$$

3.1.1 Distribució estacionària

Donat un procés de Markov $\{Y_t, t \geq 0\}$ discret i irreductible qualsevol, aleshores $\lim_{t\to\infty} P(Y(t) = j) = \pi_j$ existeix i és independent de $Y(0)$.

El vector π de probabilitat (finit o infinit numerable) es diu estacionari per a aquest procés, en un temps discret si qualsevol transició d'acord amb la matriu de probabilitats de transició P_t, verifica:

1. $\pi P_t = \pi$,
2. $\sum_{j=0}^{\infty} \pi_j = 1$.

Aquest vector de probabilitat estacionari π també s'anomena distribució estacionaria o distribució d'equilibri.

Aquesta distribució estacionaria ens serà molt útil més endavant per tal de definir diferents sistemes de cues que seran determinats a partir de la distribució estacionaria d'uns processos estocàstics concrets.

Considerem el cas en què tenim un procés de naixement i mort i volem construir la distribució estacionaria d'aquest.

Volem trobar $\pi_j \in [0,1]$. $\sum_{j=0} \pi_j = 1$. Sabem també que: $\sum_{j=0} \pi_j = 1$ i que $\pi P_t = \pi$.

Utilitzant a propietats determinem el vector π de manera que:

Per la propietat (1):

$$\pi_n = \frac{\lambda_{n-1} ... \lambda_0}{\mu_n ... \mu_1} \pi_0.$$

Per la propietat (2):

$$\pi_0 = \left(1 + \sum_{n=0}^{\infty} \frac{\lambda_{n-1} ... \lambda_0}{\mu_n ... \mu_1} \right)^{-1}.$$

És interessant, conèixer aquesta distribució estacionaria, ja que la majoria dels càlculs i raonaments que farem més endavant tenen com a base aquest estat estacionari.

En general, donada una distribució estacionaria d'un procés de naixement i mort, és important conèixer sota quines condicions la distribució existeix. Per garantir l'existència s'ha de satisfer que:

$$S_1 = 1 + \sum_{n=1}^{\infty} \frac{\lambda_{n-1} ... \lambda_0}{\mu_n ... \mu_1} < \infty \tag{3.1}$$

Per tant, amb els diversos λ_n, μ_n que es tindran depenent del model en què estiguis treballant, l'equació 3.1 servirà per buscar les condicions sota les quals existeix la solució d'equilibri π_n.

És important conèixer i entendre els processos de naixement i mort, ja que un servei amb possibilitat de formar una cua, s'interpreta com un procés de naixement i mort, on el naixement consisteix en l'entrada d'un client al servei, i la mort, la sortida d'un client. D'aquesta manera, un el procés es manté en un estat de k clients durant un cert temps fins que entra un client amb una certa probabilitat, o surt un client també amb certa probabilitat.

Seguidament, anem a estudiar alguns d'aquests tipus de processos i a diferenciar el seu comportament entre ells.

4 Conceptes fonamentals en teoria de cues

4.1 Introducció

La teoria de cues és un camp de les matemàtiques que s'encarrega de detallar i gestionar utilitzant mètodes matemàtics analítics, el fenomen de les cues o línies d'espera d'un sistema.

Els primers problemes plantejats sobre teoria de cues van aparèixer en la investigació d'Agner Krarup Erlang a principis del segle XX. Aquest, va crear models per la central telefònica de Copenhage, on ell treballava. El seu primer article va ser publicat el 1909.

El treball d'Erlang va inspirar a enginyers i matemàtics a tractar amb problemes de cues usant mètodes probabilístics. La teoria de cues és una branca viva de les ciències on els experts publiquen molts d'articles sobre el tema. Molts dels resultats obtinguts s'han utilitzat en ciència informàtica, telecomunicacions, teoria de la relativitat, i molts més.

4.2 Descripció de un sistema de cues

Un sistema de cues es pot descriure com un conjunt de çlients"que arriba a un sistema en cerca d'un servei, esperen si aquest no és immediat i abandonen el sistema un cop han sigut atesos. El terme çlient"s'usa amb un sentit general i no implica que hagi de ser un humà, poden significar peces esperant a ser processades o un llistat de documents esperant a ser impresos.

Un sistema de cues està determinat per unes certes característiques essencials que ens serviran per diferenciar la gestió i el comportament d'un sistema amb un altre.

A partir d'aquestes característiques, de manera general, podem descriure qualsevol procés de cues. Evidentment, podem trobar una gran quantitat de problemes complexos, on les característiques que esmentarem seguidament ens queden curtes per tal d'estudiar-los, però és important conèixer aquesta base per tal de descriure bé i amb rigor qualsevol sistema de cues.

4.2.1 Patró d'arribada dels clients

Per caracteritzar el sistema de cues ens cal identificar les propietats probabilístiques del flux d'entrada de clients al servei. El proces d'arribada està caracteritzat per la distribució dels temps entre dues arribades consecutives de clients. Normalment es parteix del fet que aquests temps són variables aleatòries independents i idènticament distribuïdes.

És possible, que el flux d'arribada de clients varïi amb el temps. És a dir, que la variable aleatòria de la qual hem parlat abans depengui del temps. Si aquest patró, es manté constant, parlarem d'un cas estacionari.

4.2.2 Patró de servei dels servidors

De la mateixa manera que el procés d'arribada, els temps de servei, també estan caracteritzats per una variable aleatòria.

En aquest cas, també suposarem que aquestes variables són independents i idènticament distribuïdes. De la mateixa manera, els temps de servei i els temps d'arribada consecutius de clients també suposarem que són independents.

4.2.3 Disciplina de servei

La disciplina de servei ens descriu en quin ordre són atesos els clients d'una cua d'un servei. Tenim diferents tipus de disciplines;

- FIFO *(first in, first out)*: En aquest tipus de disciplina, s'atén primer al primer client que ha arribat al servei.
- LIFO*(last in , first out)*: En aquest tipus de disciplina, s'atén primer a l'últim client que ha arribat al servei.
- RSS *(random selection of service)*: En aquest cas, el servei és aleatori.
- Prioritat: Si un client arriba a la cua amb un ordre de prioritat superior al del client que està sent atès, aquest es retira deixant passar al client que té més necessitat.

4.2.4 Capacitat del sistema

Podem classificar els sistemes de cues en dos tipus: Les cues que poden acumular un nombre limitat de clients les quals direm que són finites, i les que poden reunir un nombre infinit de clients, és a dir, les que són infinites.

4.2.5 Número de canals del servei

Un sistema de cues dependrà molt fortament del nombre de canals del servei d'aquest sistema. Podem parlar de sistemes mono-canals o sistemes multicanals del qual destacarem els sistemes amb canals de serveis paral·lels.

4.3 Notació Kendall

Abans de començar a investigar les propietats i característiques dels diferents sistemes de cues ens cal introduir una notació desenvolupada per David G. Kendal el 1953 per descriure un sistema de cues. Aquesta es basa amb les característiques detallades en l'apartat anterior.

Usant aquesta notació, un sistema de cues es representa de la següent manera:

$$A/B/m/K/n/D$$

on

A: Llei de distribució de probabilitat dels temps entre dues arribades consecutives.

B: Llei de distribució de probabilitat dels temps de servei.

m: Nombre de servidors.

K: Capacitat del sistema. Màxim nombre de clients que pot assumir el sistema, incloent-hi els que estan sent servits.

n: Mida de la població.

D: Disciplina de servei.

Les variables aleatòries que segueixen una llei exponencial usarem la notació M per descriure-les.

En el cas en què la mida de la població i la capacitat siguin infinites, i que la disciplina de servei sigui FIFO, llavors s'omet la seva notació.

5 Models de cues amb font infinita.

Els sistemes de cues es poden classificar en dos casos, el cas en què consideres que la font de clients, és a dir la població, sigui finita i el cas en què consideres que és infinita.

En el cas que la font sigui finita, el flux d'arribada de clients dependrà fortament de l'estat en què es trobi el sistema en aquell moment, fet que fa que els càlculs siguin molt més complicats. En canvi, en el cas que la font sigui infinita, el flux d'arribada és independent del nombre de clients que hi hagi en aquell moment.

Seguidament desenvoluparem i descriurem diversos models de cues tenint en compte que la mida de la població és infinita.

Aquest apartat té com a objectiu donar diverses propietats i característiques sobre cadascun dels models presentats per tal de treure unes certes conclusions, poder predir com evolucionarà el servei al llarg del temps i la gestió que cal fer-ne d'aquest.

5.1 Sistema de cues $M/M/1$

Es tracta del sistema de cues no-trivial més simple que compte tan sols amb un únic servidor, suposa que la capacitat del sistema és infinita i que la disciplina de servei és FIFO. Els clients arriben seguint una distribució Poisson de paràmetre λ, que d'aquí es pot deduir que els temps entre dues arribades consecutives de clients són variables aleatòries independents i segueixen una llei exponencial de paràmetre λ. Els temps de servei, també es considera que són independents, i segueixen una llei exponencial de paràmetre μ. És més, totes les variables aleatòries presents en el sistema són independents entre elles.

Considerem el procés $N(t)$, el qual ens determinarà el nombre de clients del sistema a temps t. Podem veure, que les variables aleatòries que determinen aquest procés estan distribuïdes exponencialment, fet que fa que $N(t)$ tingui la propietat de Markov. Veiem doncs, que $N(t)$ és un procés de Markov amb espai d'estats $\mathbb{N}$

En ser un procés de Markov, ens interessa, per tal de treballar amb el procés $N(t)$, estudiar les seves probabilitats de transició. Veiem que:

$$P_h(k,k+1) = (\lambda h + o(h))(1 - (\mu h + o(h))) + \sum_{i=2}^{\infty} (\lambda h + o(h))^i (\mu h + o(h))^{i-1}.$$

Aquesta igualtat surt del fet que la probabilitat de que en un temps h, el sistema passi de l'estat k a el $k+1$ és la suma de dos termes: El primer és la probabilitat que en un temps h arribi un client i cap servei hagi acabat. El segon és la probabilitat que durant un temps h hagin arribat n clients, i al mateix temps hagin servit a $n-1$ clients per $n \geq 2$.

Si ho simplifiquem, obtenim que:

$$P_h(k,k+1) = \lambda h + o(h).$$

Utilitzant un raonament semblant a l'anterior podem construir la probabilitat que en un temps h, el servei passi d'un estat k a un estat $k-1$. Finalment, obtenim que:

$$P_h(k,k-1) = \mu h + o(h).$$

A més, veiem que per estats no veïns, tenim que $P_h(k,j) = o(h)$ si $|k-j| \geq 2$.

Veiem doncs, que el procés $N(t)$ és un procés de naixement i mort amb índexos infinitesimals: $\lambda_k = \lambda$ i $\mu_k = \mu \;\; \forall k \in \mathbb{N}$.

Seguidament ens interessa i és important saber si es pot conèixer, la distribució estacionaria d'aquest procés de naixement i mort, ja que ens servirà per donar algunes propietats d'aquest tipus de sistema de cues.

És a dir, és important comprovar sota quines condicions aquesta distribució estacionaria existeix. Notarem $\rho = \frac{\lambda}{\mu}$ i anomenarem a ρ com la intensitat de trànsit del sistema, i veurem, tal com hem vis en l'apartat anterior, si es compleix la condició 3.1. Aquesta serà doncs:

$$S_1 = \sum_{n=0}^{\infty} \rho^n < \infty.$$

La sèrie S_1 convergeix si i només si $\rho < 1$. La segona condició (S_2) se satisfà si $\rho \leq 1$.

Llavors, la condició necessària i suficient per tal que un model $M/M/1$ tingui solució d'equilibri, és que $\rho < 1$, que és la condició d'estabilitat.

Per tant, si $\rho > 1$ aleshores el sistema diem que es sobrecarrega. La interpretació que li podem donar a aquest fet és que si $\rho > 1$ els clients arriben més ràpid que com són servits.

Considerem el cas en què $\rho < 1$. Anteriorment ja havíem definit quina era la distribució estacionaria d'un procés de naixement i mort general. En el nostre cas, obtenim que:

$$\pi_k = \pi_0 \rho^k$$

per $k \geq 0$,

$$\pi_0 = \left(1 - \sum_{k=1}^{\infty} \rho^k\right)^{-1} = 1 - \frac{\lambda}{\mu}.$$

Aquesta última igualtat es dóna pel fet de que $\rho < 1$.

Observem clarament que $\pi_k = (1-\rho)\rho^k$ per $k \geq 0$. Per tant, segueix una distribució geomètrica amb paràmetre $1-\rho$.

Un cop trobada la distribució estacionària podem dedicar-nos a estudiar diferents conceptes que ens poden donar informació per tal d'avaluar aquest sistema de cues.

5.1.1 Mesures de rendiment

Notem amb la variable N el nombre de clients al sistema en el cas estacionari. Calculem $E(N)$.

Com que tenim que N segueix una distribució geomètrica, tenim que:

$$E(N) = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}.$$

Notem que $E(N)$, com a funció de ρ, té una asímptota vertical a $\rho = 1$. Això ens indica el comportament dramàtic del nombre mitjà de clients al sistema a mesura que ens anem apropant cap a $\rho = 1$, és a dir, cap a la violació de la condició d'estabilitat.

A més a més de l'esperança de la variable N, també podem obtenir la seva variància, a partir de la distribució geomètrica:

$$Var(N) = \sum_{k=0}^{\infty}(k - E(N))^2\pi_k = \frac{\rho}{(1-\rho)^2}.$$

Tenint en compte que en el sistema $M/M/1$ només hi ha un servidor, si notem com el nombre de clients a la cua com Q, podem calcular $E(Q)$ com:

$$E(Q) = \sum_{k=1}^{\infty}(k-1)\pi_k = \sum_{k=1}^{\infty} k\pi_k - \sum_{k=1}^{\infty}\pi_k = E(N) - (1 - \pi_0) = E(N) - \rho = \frac{\rho^2}{1-\rho}.$$

Anem a veure un petit exemple per tal de veure com es calculen les diferents propietats esmentades anteriorment i quin tipus d'informació ens poden donar.

Exemple 1: En un petit servidor el temps de processament per treball es distribueix exponencialment amb un temps mitjà de 3 minuts. Els treballs arriben aleatòriament cada 4 minuts de mitjana. Els treballs es processen amb la disciplina FIFO.

Primerament, veiem que les **taxes de naixement i mort** són $\lambda = \frac{1}{4}$ treballs per minut i $\mu = \frac{1}{3}$ treballs per minut respectivament.

Veiem que $\rho = \frac{3}{4} < 1$, que ens indica que existeix una solució d'equilibri.

Algunes probabilitats que poden ser d'interès són:

$$P(\text{treball hagi d'esperar a ser processat}) = 1 - \pi_0 = \rho = 0.75$$

$$P(\text{treball trobi cua}) = P(N \geq 2) = 1 - P(N < 2) = 1 - (P(N = 0) + P(N = 1)) =$$
$$= 1 - (1 - \rho + (1 - \rho)\rho) = \rho^2 = 0.5625$$

És a dir, que només el 25% dels treballs passaran immediatament a ser processats, i el 56,25% trobaran cua a l'arribar.

També ens pot interessar el **nombre mitjà de treballs en el sistema, i en la cua**.

$$E(N) = \frac{\rho}{1-\rho} = \frac{0.75}{0.25} = 3$$
$$E(Q) = \frac{\rho^2}{1-\rho} = 2.25$$

Veiem, que el valor de $E(N)$ va augmentant a mesura que ρ tendeix cap a 1, pel fet que augmenta la taxa d'arribades o bé que disminueix la taxa de servei.

Així, si la taxa d'arribades augmenta un 25% sent ara, 18.75 treballs per hora o 0.3125 treballs per minut, eleva ρ fins a 0.9375 i en conseqüència, tenim que el nombre mitjà de treballs al sistema són 15 treballs, que són 5 cops més dels que es tenien anteriorment.

5.1.2 Temps d'espera

Considerem ara la variable aleatòria W, que ens denotarà el temps invertit d'un client a la cua en l'estat estacionari. La definim de la següent manera:

$$W = \begin{cases} 0 & \text{si } N = 0, \\ S_1' + S_2 + ... + S_n & \text{si } N = n \ , \ n \geq 1. \end{cases}$$

On $S_2, S_3, ..., S_n$ són els temps de servei del segon client fins a l'enèsim client respectivament, que estan a la cua.

La variable aleatòria S_1' denota el temps de servei restant del client que està al servidor. Anem a veure quina distribució segueix.

Sabem que $S_i \sim exp(\mu)$ $\forall i$. Utilitzant la propietat de Markov de la variable exponencial, deduïm que $S_1' \sim exp(\mu)$.

Passem a calcular la funció distribució F_W associada a la variable W.

Tenim que $F_W(0) = P(W \leq 0) = P(N = 0) = 1 - \rho$.

Si busquem ara $F_W(w)$ amb $w \neq 0$, aleshores, usant el teorema de la probabilitat total es té que:

$$F_W(w) = P(W \leq w) = \sum_{n=0}^{\infty} P(W \leq w | N = n) P(N = n) =$$

$$= P(W \leq w | N = 0) P(N = 0) + \sum_{n=1}^{\infty} P(W \leq w | N = n) P(N = n) =$$

$$= 1(1 - \rho) + \sum_{n=1}^{\infty} P(S_1' + ... + S_n \leq w)(1 - \rho)\rho^n =$$

$$= 1 - \rho + \sum_{n=1}^{\infty} (1 - \rho)\rho^n \int_0^w e^{-\mu t} \frac{\mu^n t^{n-1}}{(n-1)!} dt = (1 - \rho) + (1 - \rho) \int_0^w e^{-\mu t} \sum_{n=1}^{\infty} \frac{\lambda^n t^{n-1}}{(n-1)!} dt =$$

$$= (1 - \rho) + (1 - \rho)\lambda \int_0^w e^{-\mu t} e^{\lambda t} dt = (1 - \rho) + (1 - \rho)\lambda \left[\frac{e^{(\lambda - \mu)t}}{\lambda - \mu} \right]_0^w = 1 - \rho e^{(\lambda - \mu)w}.$$

Observem, doncs, que:

$$F_W(w) = 1 - \rho e^{(\lambda - \mu)w} \; w \geq 0.$$

En la construcció de $F_W(w)$ hem usat el fet que $S_{1'} + ... + S_n \sim \Gamma(n, \mu)$.

Hem d'incidir sobre el fet que W no és una variable aleatòria absolutament continua, per tant no podem parlar de la densitat de la variable aleatòria. Si ens fixem en com hem definit la variable, podem observar que trobem una part discreta, que correspon al fet que $W = 0$ i una part absolutament continua.

Si suposéssim que la variable W és absolutament continua, tindríem que la densitat d'aquesta és de la següent forma:

$$f_W(w) = -\rho(\mu - \lambda)e^{(\lambda - \mu)w},$$

obtinguda pel fet de que $f_W(w) = \frac{d}{dx}F(x)$. Si fem els càlculs, veiem que:

$$\int_0^\infty f_W(w)dw = \rho < 1.$$

fet que contradiu amb una de les propietats de la densitat d'una variable aleatòria. Tot hi això, aquesta densitat ens serveix per a definir la part absolutament continua de la variable.

Seguidament, aquests últims resultats, ens serveixen per trobar l'esperança i variància de la variable:

$$E(W) = 0P(W = 0) + \int_0^\infty t f_W(t)dt = \frac{\rho}{\mu(1-\rho)} = \frac{1}{\mu}E(N),$$
$$Var(W) = E(W^2) - (E(W))^2 = \rho\frac{2}{(\mu(1-\rho))^2} - \frac{\rho^2}{(\mu(1-\rho))^2}.$$

Un cop hem estudiat el temps que inverteix un client a la cua (W), podem considerar ara la variable aleatòria notada per T definida com:

$$T = W + S$$

, on S correspon al temps de servei.

És fàcil veure que la variable aleatòria T fa referència al temps total que un client inverteix al sistema.

Podem trobar la distribució de la variable aleatòria T utilitzant un procediment anàleg al que hem utilitzat per trobar la funció de distribució de la variable aleatòria W. Si quan arriba un client al sistema ja hi ha n clients esperant, aquest, haurà de romandre en el sistema un temps total igual a suma de n+1 variables aleatòries independents idènticament distribuïdes que segueixen una llei exponencial de paràmetre μ, notades com $\{S_i\}_{i\in\{1,\ldots n+1\}}$.

Així, usant que $S_1 + \ldots S_{n+1} \sim \Gamma(n+1, \mu)$ i el teorema de la probabilitat total, tenim que:

$$F_T(t) = 1 - e^{(\lambda-\mu)t}.$$

A partir de la definició de T, podem definir algunes propietats d'aquesta variable aleatòria, com l'esperança i la variància d'aquesta:

$$E(T) = E(W + S) = E(W) + E(S) = \frac{1}{\mu}E(N) + \frac{1}{\mu} = \frac{1}{\mu(1-\rho)},$$
$$Var(T) = \rho\frac{2}{(\mu(1-\rho))^2} - \frac{\rho^2}{(\mu(1-\rho))^2} + \frac{1}{\mu^2} = \frac{1}{(\mu(1-\rho))^2},$$

on hem usat el fet que les variables W i S són independents entre si.

Indagant una mica més en les esperances estudiades anteriorment, sempre en el cas estacionari, podem observar les següents relacions:

$$\lambda E(T) = \lambda\frac{1}{\mu(1-\rho)} = \frac{\rho}{1-\rho} = E(N),$$

$$\lambda E(W) = \lambda \frac{\rho}{\mu(1-\rho)} = \frac{\rho^2}{1-\rho} = E(Q).$$

Aquestes dues relacions s'anomenen les fórmules de Little o la llei de Little.

Exemple 2: Considerem les mateixes premisses de l'exercici d'abans, però ara anem a analitzar els temps en el sistema i en la cua.

Usant les fórmules de Little podem trobar el **temps mitjà d'espera invertit pels programes en el sistema** i el **temps mitja d'espera a la cua pels programes**:

$$E(T) = \frac{E(N)}{\lambda} = \frac{3}{0.25} = 12 \text{ minuts}$$

$$E(W) = \frac{E(Q)}{\lambda} = \frac{2.25}{0.25} = 9 \text{ minuts}$$

Podem estudiar algunes probabilitats com per exemple la **probabilitat que un treball hagi d'esperar a la cua més de 20 minuts**:

$$P(W > 20) = 1 - P(W \leq 20) = 1 - (1 - \rho^{(\lambda-\mu)20}) = 0.75e^{-\frac{20}{12}} = 0.142$$

o la **probabilitat de romandre en el sistema més de 20 minuts**

$$P(T > 20) = 1 - P(T \leq 20) = 1 - (1 - e^{(\lambda-\mu)20} = e^{-\frac{20}{12}} = 0.188$$

Llavors, més del 14% dels programes estaran en la cua més de 20 minuts i quasi el 20% dels programes no sortiran del sistema en menys de 20 minuts.

Seguidament decidim augmentar la capacitat del nostre servidor en el cas que la càrrega de treball arribi a un nivell tal que el temps mitjà que inverteix un treball en el sistema arriba als 30 minuts. Volem saber per quin valor de λ ens passarà això. Sabem el valor de $E(T)$, i la capacitat actual del servidor, és a dir, el valor de μ. Per trobar λ utilitzarem:

$$E(T) = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu - \lambda},$$

de manera que:

$$30 = \frac{3}{1-3\lambda} \text{ i per tant, } \lambda = 0.3 \text{ treballs per minut.}$$

5.1.3 Distribució a l'arribada i sortida de clients

Hem vist, anteriorment, que un sistema de cues és un procés de naixement (arribades) i mort (serveis). Anem a veure una propietat dels processos de naixement i mort que ens pot servir a l'hora de treballar models de cues, i en particular per treballar xarxes de cues de les quals parlarem més endavant.

Tal com hem dit abans, si tenim que $\rho = \frac{\lambda}{\mu} < 1$ podem definir una distribució estacionaria de la següent manera:

$$\pi_n = \frac{\lambda_{n-1}...\lambda_0}{\mu_n...\mu_1}\pi_0.$$

Anem a veure quina és la distribució en els moments de les arribades i sortides de clients. Aquest resultat ens serà útil més endavant.

Denotem amb N_a l'estat en què es troba el procés a l'instant en què arriba un client del sistema, i N_s l'estat en què es troba en el moment en què un client surt, i denotem amb $\Pi_k = P(N_a = k)$ i $D_k = P(N_s = k)$, per $k = 0, 1, 2, ...$, les seves distribucions.

Aplicant el teorema de Bayes és fàcil veure que:

$$\Pi_k = \lim_{h \to 0} \frac{(\lambda_k h + o(h))\pi_k}{\sum_{j=0}^{\infty}(\lambda_j h + o(h))\pi_j} = \frac{\lambda_k \pi_k}{\sum_{j=0}^{\infty} \lambda_j \pi_j}.$$

De manera similar tenim que:

$$D_k = \lim_{h \to 0} \frac{(\mu_{k+1} h + o(h))\pi_{k+1}}{\sum_{j=1}^{\infty}(\mu_j h + o(h))\pi_j} = \frac{\mu_{k+1} \pi_{k+1}}{\sum_{j=1}^{\infty} \mu_j \pi_j}.$$

Però com que sabem que $\pi_{k+1} = \frac{\lambda_k \pi_k}{\mu_{k+1}}$ $\forall k$ obtenim que:

$$D_k = \frac{\lambda_k \pi_k}{\sum_{j=0}^{\infty} \lambda_j \pi_j} = \Pi_k \ , \forall k. \tag{5.1}$$

En altres paraules, la informació anterior indica que les distribucions en l'estat estable en el moment dels naixements i les morts són les mateixes.

Això ens indica que la distribució dels temps de sortida de clients segueix una exponencial de paràmetre λ i no μ com seria de suposar.

Significa doncs, que el procés de sortida dels clients és un procés de Poisson de paràmetre λ.

5.1.4 Cicles d'ocupació i desocupació

Podríem analitzar, també els cicles d'ocupació i desocupació del servei.

Considerem dues noves variables aleatòries que anomenarem T_o i T_d de manera que representen la longitud d'un cicle d'ocupació i desocupació respectivament.

Es pot veure que $T_d \sim exp(\lambda)$, ja que el servidor està desocupat fins que arriba un nou client seguint una distribució exponencial de paràmetre λ, i com que π_0 representa la proporció de temps que el servidor està desocupat (o el que és el mateix, el sistema està vuit) en estat estacionari, llavors:

$$\frac{E(T_d)}{E(T_d) + E(T_o)} = \pi_0 = 1 - \rho.$$

I llavors, en podem deduir que:

$$E(T_o) = \frac{1}{\lambda} \frac{\rho}{1 - \rho} = \frac{1}{\mu - \lambda}.$$

5.2 Sistema de cues $M/M/n$

El sistema de cues $M/M/n$, o model Erlang-C, és un model de cues amb multiservidors. És una variació de la cua clàssica que hem estudiat anteriorment on tenim n servidors funcionant independentment els uns dels altres. La notació de Kendall descriu un sistema on les arribades a una única cua segueixen una llei Poisson de paràmetre λ, hi ha n servidors, i els temps de servei de cadascun dels servidors estan distribuïts exponencialment amb paràmetre μ. Igual que en el cas del model de cues $M/M/1$, suposem que la capacitat del sistema és infinita i seguirem una disciplina de servei FIFO.

Aquesta és una modificació natural del sistema clàssic, ja que si la taxa d'arribada de clients és més gran que la de serveis, el sistema es desestabilitza i necessitem més servidors per tal que el sistema torni a ser eficient.

En el cas que nosaltres estudiarem seguidament, on tenim diferents servidors en paral·lel, no hi ha vàries cues, sinó que hi haurà una única cua i els clients s'aniran situant als servidors que quedin lliures.

És fàcil provar, a partir d'arguments utilitzats en apartats anteriors, que el sistema de cues $M/M/n$ és un procés de naixement i mort i que les seves probabilitats de transició següents:

$$
\begin{aligned}
P_h(k,k+1) &= (\lambda h + o(h))(1-(\mu_k h + o(h))) + o(h) = \lambda h + o(h),\\
P_h(k,k-1) &= (\mu_k h + o(h))(1-(\lambda h + o(h))) + o(h) = \lambda h + o(h),
\end{aligned}
$$

on $\mu_k = \begin{cases} k\mu \text{ , si } 0 \leq k \leq n, \\ n\mu \text{ , si } n < k. \end{cases}$

En el cas d'aquest sistema, la nostra condició d'estabilitat que ens permet descriure una distribució estacionaria, serà $a = \frac{\lambda}{n\mu} = \frac{\rho}{n} < 1$.

Si ens trobem sota la condició d'estabilitat, la distribució estacionària està definida en dos casos.

- Si $k < n$, aleshores es té que :

$$\pi_k = \pi_0 \prod_{i=0}^{k-1} \frac{\lambda}{(i+1)\mu} = \pi_0 \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!}.$$

- Si $k \geq n$, aleshores es té que:

$$\pi_k = \pi_0 \prod_{i=0}^{n-1} \frac{\lambda}{(i+1)\mu} \prod_{j=n}^{k-1} \frac{\lambda}{n\mu} = \pi_0 \left(\frac{\lambda}{\mu}\right)^k \frac{1}{n!n^{k-n}} = \pi_0 \frac{a^k n^n}{n!}.$$

Finalment, veiem que:

$$\pi_0 = \left(1 + \sum_{k=1}^{n-1} \frac{\rho^k}{k!} + \sum_{k=n}^{\infty} \frac{\rho^k}{n!n^{k-n}}\right)^{-1}.$$

A més, utilitzant l'expressió de a podem desenvolupar la sèrie

$$\sum_{k=n}^{\infty} \frac{\rho^k}{n!n^{k-n}} = \frac{1}{n!} \sum_{i=0}^{\infty} \frac{\rho^{i+n}}{n^i} = \frac{\rho^n}{n!} \sum_{i=0}^{\infty} a^i = \frac{\rho^n}{n!(1-a)}, \tag{5.2}$$

per tant:

$$\pi_0 = \left(1 + \sum_{k=1}^{n-1} \frac{\rho^k}{k!} + \frac{\rho^n}{n!(1-a)}\right)^{-1}.$$

Un cop definida la distribució estacionària, podem passar a comprovar algunes propietats importants del procés i obtenir resultats interessants que ens serviran per analitzar i observar el comportament d'aquesta cua. Concretem que utilitzarem la mateixa notació que en la secció anterior, és a dir, la notació utilitzada per descriure diferents elements del sistema de cues $M/M/1$

5.2.1 Fórmula Erlang C

D'aquest procés, ens pot interessar calcular la probabilitat que un client que arriba al sistema es vegi forçat a unir-se a la cua, és a dir, que els n servidors estiguin ocupats. Aquesta serà:

$$P(client\, espera) = \sum_{k=n}^{\infty} \pi_k = \sum_{k=n}^{\infty} \pi_0 \frac{\rho^k}{n! n^{k-n}},$$

que a partir de (3) i per l'expressió de π_0 podem escriure-ho com:

$$P(client\, espera) = \frac{\frac{\rho^n}{n!(1-a)}}{\sum_{k=0}^{n-1} \frac{\rho^k}{k!} + \frac{\rho^n}{n!(1-a)}} = \frac{\frac{\rho^n n}{n!(n-\rho)}}{\sum_{k=0}^{n-1} \frac{\rho^k}{k!} + \frac{\rho^n n}{n!(n-\rho)}} = C(n, \rho).$$

Aquesta probabilitat és utilitzada freqüentment en diferents problemes pràctics, com per exemple en sistemes de telefonia. És també una formula bastant famosa anomenada **fórmula Erlang C**.

La fórmula rep el nom del seu creador Agner Krarup Erlang, pioner en la teoria de cues, i va ser desenvolupada particularment per l'expressió de la probabilitat que un client que arriba hagi de fer cua en el sistema que estem treballant.

Quan Erlang va desenvolupar aquesta fórmula, es va desenvolupar en un conjunt de suposicions. Aquestes, són bastant precises en la majoria de casos, però tot i això en el cas que hi hagi una congestió de clients extremadament alta, la fórmula Erlang C no pot donar resultats molt precisos.

Això és bastant lògic, ja que la fórmula Erlang C està construïda a partir de la distribució estacionaria del procés, i ja hem vist anteriorment que si la taxa d'arribades dels clients és molt superior a la de servei, el sistema es congestiona i la distribució estacionaria deixa de ser una eina útil i eficaç.

Aquesta fórmula determinarà molts dels resultats que veurem en aquesta secció, per tant serà una peça clau en la descripció i estudi del sistema de cues $M/M/n$.

5.2.2 Mesures de rendiment

De la mateixa manera que la secció anterior, podem estudiar el nombre mitjà de clients a la cua, és a dir, la longitud mitjana de la cua:

$$E(Q) = \sum_{k=n}^{\infty}(k-n)\pi_k = \sum_{i=0}^{\infty} i\pi_{n+i} = \sum_{i=0}^{\infty} \frac{i\rho^n}{n!}a^i\pi_0 = \pi_0\frac{\rho^n}{n!}a\sum_{i=0}^{\infty}\frac{\partial a^i}{\partial a} =$$
$$= \pi_0\frac{\rho^n}{n!}a\frac{\partial}{\partial a}\sum_{i=0}^{\infty}a^i = \pi_0\frac{\rho^n}{n!}\frac{a}{(1-a)^2} = \frac{a}{1-a}C(n,\rho) = \frac{\rho}{n-\rho}C(n,\rho).$$

Ens pot ser d'utilitat conèixer el nombre mitjà de servidors ocupats. Denotarem amb L el nombre de servidors ocupats en un sistema de cues. Per tal de desenvolupar aquest càlcul necessitem el resultat que hem obtingut en 5.2. L'aplicarem a la tercera igualtat:

$$E(L) = \sum_{k=0}^{n-1} k\pi_k + \sum_{k=n}^{\infty} n\pi_k = \sum_{k=0}^{n-1} k\pi_0\frac{\rho^k}{k!} + \sum_{k=n}^{\infty} n\frac{\rho^n}{n!n^{k-n}} =$$
$$= \pi_0\left(\rho\sum_{k=0}^{n-2}\frac{\rho^k}{k!} + \frac{\rho^n}{(n-1)!}\frac{1}{1-a}\right) = \rho\left(\sum_{k=0}^{n-2}\frac{\rho^k}{k!} + \frac{\rho^{n-1}}{(n-1)!} + \frac{\rho^{n-1}}{(n-1)!}\left(\frac{1}{1-a}-1\right)\right)\pi_0.$$

Desenvolupem el darrer terme:

$$\frac{\rho^{n-1}}{(n-1)!}\left(\frac{1}{1-a}-1\right) = \frac{\rho^{n-1}}{(n-1)!}\sum_{i=1}^{\infty}a^i = \frac{1}{(n-1)!}\sum_{i=1}^{\infty}\frac{\rho^{i+n-1}}{n^i} =$$
$$= \frac{1}{(n-1)!}\sum_{k=n}^{\infty}\frac{\rho^k}{n^{k-n+1}} = \sum_{k=n}^{\infty}\frac{\rho^k}{n!n^{k-n}} = \frac{\rho^n}{n!(1-a)}.$$

Per tant, tornant al desenvolupament de $E(L)$:

$$E(L) = \rho\left(\sum_{k=0}^{n-1}\frac{\rho^k}{k!} + \frac{\rho^n}{n!(1-a)}\right)\pi_0 = \rho\frac{1}{\pi_0}\pi_0 = \rho.$$

El nombre mitjà de clients en el sistema $E(N)$ també pot ser una informació rellevant en alguns problemes i, a diferència del cas $M/M/1$, el calculem ara i no primerament, ja que utilitzarem els càlculs anteriors per tal de desenvolupar $E(N)$ de la següent manera.

$$E(N) = \sum_{k=0}^{\infty} k\pi_k = \sum_{k=0}^{n-1} k\pi_k + \sum_{k=n}^{\infty}(k-n)\pi_k + \sum_{k=n}^{\infty} n\pi_k =$$
$$= E(L) + E(Q) = \rho + \frac{\rho}{n-\rho}C(n,\rho).$$

De fet, la igualtat $E(N) = E(L) + E(Q)$ té bastant de sentit, ja que un client està a la cua del servei o essent servit.

Després de l'estudi d'aquests tres elements, podriem arribar a altres diferents conclusions, com el càlcul del nombre mitjà de servidors desocupats (ho notarem amb D).

$$E(D) = n - E(L).$$

D'aquí en podem treure fórmules que ens poden ser interessants i útils:

$$E(N) - n = E(Q) - E(D).$$

5.2.3 Temps d'espera

De la mateixa manera que en el sistema $M/M/1$, ens interessa estudiar quina llei segueix el temps invertit per un client a la cua. Intentem construir la següent llei per passos. Un client s'haurà d'esperar a la cua si en el seu moment d'arribada hi ha com a mínim n clients dins el sistema. En el cas que quan arribi no hi ha ningú a la cua, però tots els servidors estan ocupats, el temps d'espera del client a la cua serà el temps de servei d'un dels servidors, per tant estarà distribuït exponencialment amb paràmetre $n\mu$. En el cas que al sistema hi hagi $n+j$ clients al sistema, llavors el temps d'espera del client a la cua seguirà una distribució Erlang amb paràmetres $(j+1, n\mu)$.

El que intentarem trobar és una funció que ens pugui definir com es distribueix el temps d'espera dels clients a la cua. El problema que tenim, de la mateixa manera que ens hem trobat quan hem estudiat el model $M/M/1$, és que la variable W que defineix el temps d'espera d'un client a la cua, no és absolutament continua.

Per tant, no podem construir una densitat com a tal, sinó que podrem trobar una funció semblant a una densitat, però que ens serà útil per poder definir la distribució de W i trobar alguns resultats importants per la modelització d'aquest procés.

Observació: La distribució Erlang és una distribució, de probabilitat continua amb dos paràmetres (k, λ) que és equivalent a la distribució gamma amb paràmetres $k \in \mathbb{N}$ i $\lambda > 0$. Opera en: $x \in [0, \infty)$. La seva funció densitat és la següent:

$$f_{E(k,\lambda)}(x) = \lambda^k \frac{x^{k-1}}{(k-1)!} e^{-\lambda x}.$$

Per la llei de probabilitat total, que diu que si tenim una partició A_i amb $i \in I$ de l'espai mostral i un esdeveniment B, llavors $P(B) = \sum_{i\in I} P(B|A_i)P(A_i)$, que en el nostre cas, l'aplicarem per funcions de densitat, tenim que la funció densitat del temps d'espera d'un client a la cua és:

$$f_W(w) = \sum_{k=n}^{\infty} \pi_k f_{E(n+1-k,\mu n)}(w) = \sum_{j=0}^{\infty} \pi_{n+j} f_{E(j+1,\mu n)}(w) = \sum_{j=0}^{\infty} \pi_{n+j} (n\mu)^{j+1} \frac{x^j}{j!} e^{-n\mu x}.$$

Podem treballar una mica amb aquest resultat, i així obtenir una expressió de la funció més útil. Substituint la distribució π_{n+j}

$$f_W(x) = \sum_{j=0}^{\infty} \pi_0 \frac{\rho^n}{n!} a^j (n\mu)^{j+1} \frac{x^j}{j!} e^{-n\mu x} = \pi_0 \frac{\rho^n}{n!} n\mu e^{-n\mu x} \sum_{j=0}^{\infty} \frac{(an\mu x)^j}{j!} =$$

$$= \pi_0 \frac{\rho^n}{n!} n\mu e^{-(n\mu-\lambda)x} = \pi_0 \frac{\rho^n}{n!} \frac{1}{1-a} n\mu(1-a) e^{-n\mu(1-a)x}.$$

Podem observar que:

$$\pi_0 \frac{\rho^n}{n!} \frac{1}{1-a} = \pi_0 \frac{\rho^n}{n!} \sum_{k=0}^{\infty} a^j = \pi_0 \frac{\rho^n}{n!} \sum_{k=n}^{\infty} a^{k-n} = \sum_{k=n}^{\infty} \pi_0 \frac{\rho^k}{n! n^{k-n}} = P(client\, espera). \quad (5.3)$$

Per tant, retornant a la densitat de W, tenim que:

$$f_W(x) = P(client\, espera)n\mu(1-a)e^{-n\mu(1-a)x}.$$

A partir de la funció densitat, podem construir la funció de distribució de la següent manera:

$$F_W(x) = P(W \le x) = 1 - P(W > x) = 1 - \int_x^\infty f_W(y)dy =$$

$$1 - \Big[P(client\, espera)e^{-n\mu(1-a)y}\Big]_x^\infty = 1 - C(n,\rho)e^{-\mu(n-\rho)x}.$$

Si ens fixem com hem construït la funció Erlang C, veiem que el que hem intentat calcular és la probabilitat que un client que arriba s'hagi d'esperar.

Després de treballar la distribució del temps d'espera d'un client a la cua (denotat per la variable W), podem tornar a estudiar aquesta probabilitat de manera que:

$$P(client\, espera) = P(W > 0) = 1 - P(W \le 0) =$$

$$= 1 - (1 - C(n,\rho)e^{-\mu(1-\rho)0}) = C(n,\rho).$$

Utilitzant la versió de la funció densitat també podem determinar l'esperança de W.

$$E(W) = \int_0^\infty x f_W(x)dx = \frac{\rho^n}{n!}\pi_0\frac{1}{1-a}\int_0^\infty x(n\mu(1-a))e^{-n\mu(1-a)x}dx.$$

Ens centrarem amb aquesta última integral, i com que veiem que correspon a l'esperança d'una variable aleatòria que segueix una distribució exponencial de paràmetre $n\mu(1-a)$, tenim que:

$$\int_0^\infty x(n\mu(1-a))e^{-n\mu(1-a)x}dx = \frac{1}{(1-a)n\mu}.$$

Recuperant el càlcul de $E(W)$, i utilitzant el desenvolupament 5.3, tindrem:

$$E(W) = \frac{\rho^n}{n!}\pi_0\frac{1}{1-a}\frac{1}{(1-a)n\mu} = \frac{1}{\mu(n-\rho)}C(n,\rho).$$

Anteriorment, doncs, hem considerat el cas en què un client arriba al sistema i es troba cua, és a dir, que al sistema hi ha n clients o més. En aquest cas, el temps que inverteix el client al sistema és la suma del temps d'espera del client a la cua més el temps de servei d'aquest. En el cas contrari, el servei del client en qüestió comença immediatament un cop ha arribat.

Usarem T com la variable aleatòria que defineix el temps que inverteix un client a dins el sistema, d'ençà que entra, fins que surt.

Ens interessa trobar $E(T)$ que ens servirà per comparar sistemes i treure resultats útlis per la pràctica. És clar, que $T = W + S$, ja que el temps invertit al sistema serà el temps invertit a la cua i el temps que dura el servei del client.

Per això:

$$E(T) = E(W+S) = E(W) + E(S) = \frac{1}{\mu(n-\rho)}C(n,\rho) + \frac{1}{\mu}.$$

En el sistema de cues $M/M/n$ també es compleixen les fòrmules de Little, i poden ser provades per simples càlculs:

$$\lambda E(T) = \lambda \left(\frac{1}{\mu} + \frac{1}{n\mu} \frac{\rho^n}{n!} \pi_0 \frac{1}{(1-a)^2} \right) =$$

$$\rho + \frac{\rho^n}{n!} \pi_0 \frac{a}{(1-a)^2} = E(N).$$

Tenim que:

$$E(N) = \lambda E(T).$$

Fent el mateix procediment és fàcil veure també que:

$$E(Q) = \lambda E(W).$$

5.3 Comparació models M/M/1 i M/M/n

En els últims apartats hem estat estudiant dos dels models de cues més importants i bàsics de la teoria de cues. Després d'haver-ho fet, pot ser interessant veure quin dels dos models és més eficaç, és a dir, si tenim el mateix nombre de servidors en cadascun dels models, per exemple n, quin sistema serà més eficaç, un que té n cues on cadascuna segueix un model $M/M/1$ o un que segueix un model $M/M/n$.

És per això que compararem dos sistemes, un sistema on tenim n cues independents que segueixen un model $M/M/1$ i un sistema on tenim una sola cua que segueix un model $M/M/n$. En els dos casos, tenim n servidors, i suposarem que el flux d'entrada i sortida dels dos sistemes és el mateix.

Considerarem que els dos sistemes tenen una capacitat infinita i que la disciplina de servei és FIFO.

Per tal d'entendre l'argument que farem servir, treballarem primer en un cas particular.

Suposarem que tenim en aquest cas, tan sols dos servidors a tots dos sistemes, i creem una situació hipotètica on a cadascun dels dos sistemes hi ha només dos clients. En el model de sistema d'una sola cua i dos servidors independents, és a dir, el model $M/M/2$, cada client es distribuirà a un dels dos sistemes, ja que el primer client pot triar a quin sistema vol anar, i el segon, seguint les regles que regeixen aquest sistema en concret, se situarà al servidor que queda lliure.

Ens interessa saber, quin és el temps total que inverteixen els dos clients al sistema, és a dir, quant de temps passarà des que els dos clients entren, fins que ja no en queda cap.

Per tal d'estudiar-ho, podem observar quant de temps passa des que els clients comencen a ser servits, fins que el segon client abandona el sistema.

Després d'haver estudiat el model de cues $M/M/n$, sabem que quan hi ha dos clients al sistema, el servei del primer client que surt del sistema segueix una exponencial de paràmetre 2μ. Un cop servit el primer client, només queda el segon per ser servit, i el temps de servei d'aquest segueix una exponencial de paràmetre μ.

En el cas que tinguem dues cues independents cadascuna amb el seu propi servidor, un cop els clients entren al servei, podem trobar-nos amb tres situacions diferents. Que els dos clients se situïn a la cua del primer servidor, que se situïn a la cua del segon servidor, o que es situïn cadascun d'ells a una cua diferent (aquest últim cas coincidiria amb la distribució dels clients del model $M/M/2$).

Si ens fem la mateixa pregunta que abans, els temps de servei del primer client, i del segon client, en els dos primers casos, seguiran una distribució exponencial de paràmetre μ, en canvi el temps de servei del primer client en el tercer cas, seguirà una distribució exponencial de paràmetre 2μ, i el servei del segon client segueix una exponencial de paràmetre μ, tal com hem vist en el cas del sistema $M/M/2$.

Coneixent la distribució exponencial, podem deduir fàcilment, que si els dos clients estan distribuïts un a cada servidor, el temps de servei del primer client a ser servit és menor que en els altres dos casos, pel fet que els dos clients estan sent servits a l'hora, per tant, sortirà primer el client amb el servidor més ràpid. Si els dos clients comparteixen servidor, anul·lem aquesta possibilitat.

Podem observar que en aquests dos casos, els clients abandonarien abans el sistema si el servei segueix el model $M/M/2$. En el millor dels casos, que seria el cas en què en el sistema amb dues cues independents els clients es distribueixen un a cada cua, no hi hauria distinció amb el model $M/M/2$.

Aquest mateix plantejament, el podem traslladar per sistemes amb n servidors, i k clients. L'argument per tal de fer la comparació seria el mateix.

En el cas del sistema amb n cues independents, podríem distribuir els k clients de diverses maneres diferents, només una, coincidiria amb la distribució de clients del sistema $M/M/n$. La resta, observem que el temps que inverteixen els clients al sistema és major al temps invertit pels clients en el sistema $M/M/n$, de la mateixa manera que hem observat en el cas particular.

Per aquesta raó, podem concloure que el sistema amb model $M/M/n$ és més eficient i útil que n cues independents cadascuna de les quals segueix un model $M/M/1$.

El que també podríem estudiar per comprovar el que hem assegurat anteriorment, és les mesures de rendiment dels dos sistemes, si considerem que estan sota les mateixes condicions.

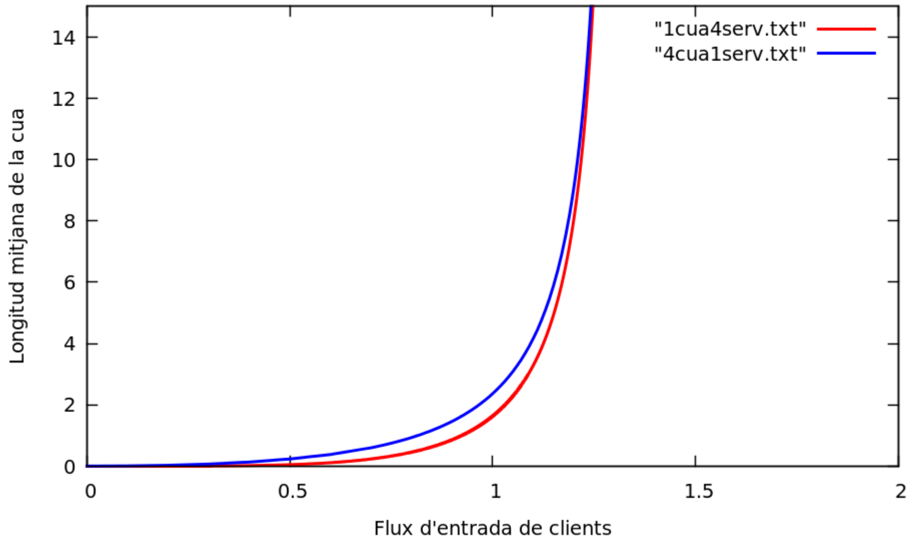
Considerarem que als dos sistemes, els clients arriben seguint una distribució Poisson de paràmetre λ, i que els temps de servei segueixen una exponencial de paràmetre μ. Igual que abans, al primer sistema hi haurà n cues on cada cua té el seu propi servidor (cada cua considerarem que és un sistema independent seguint un model $M/M/1$), i en el segon sistema, hi haurà una sola cua amb n servidors (el segon sistema segueix un model $M/M/n$).

Pel nostre problema, el que farem, és comparar el nombre mitjà de clients d'una de les cues del primer sistema amb el nombre mitjà de clients de la cua del segon sistema.

Suposarem que a cada cua independent del primer sistema, els clients arriben seguint una distribució Poisson de paràmetre $\frac{\lambda}{n}$.

Compararem el nombre mitjà de clients a la cua d'un sistema $M/M/1$ amb flux d'entrada $\frac{\lambda}{n}$ i flux de sortida μ i d'un sistema $M/M/n$ amb flux d'entrada λ i flux de sortida μ.

La demostració teòrica no és molt agradable i fàcil de tractar. Per això hem utilitzat un programa en C per veure quina és la llargada mitjana de la cua en cada sistema depenent del flux d'entrada de clients a cada sistema. Podem trobar el codi utilitzat a l'Annex. Fixarem, per exemple, $\mu = 0.33$ i $n = 4$ i deixarem com a variable independent el flux d'entrada λ.



En l'anterior gràfic, es pot observar la longitud mitjana de la cua en funció del flux d'entrada de clients (λ). El gràfic blau representa el sistema que segueix un model $M/M/1$ i el gràfic vermell un sistema que segueix un model $M/M/4$.

Podem observar que a mesura que anem augmentant el flux d'entrada de clients, el sistema que segueix un model de cues $M/M/4$ té en tot moment una longitud mitjana de la cua inferior al del sistema amb quatre cues, cadascuna de les quals segueix un model $M/M/1$.

Aquest gràfic ens mostra, que el temps mitjà que invertirà un client al sistema amb 4 cues independents i cada cua amb el seu servidor, és superior al temps mitjà que inverteix en el sistema amb model de cues $M/M/n$. Anem-ho a veure:

Notem amb Q_1 el nombre de clients a la cua del sistema que segueix un model $M/M/4$, i amb Q_2 el nombre de clients a una de les cues del sistema amb quatre cues independents (cada cua segueix un model $M/M/1$). Sigui λ el flux d'entrada de clients als dos sistemes, hem vist de moment que $\forall \lambda < 4\mu$, on μ és el flux de sortida de cada servidor, tenim que:

$$E(Q_1) < E(Q_2)$$

I utilitzant el teorema de Little, observem que:

$$E(Q_1) = \lambda E(W_1) \text{ i } E(Q_2) = \frac{\lambda}{4} E(W_2)$$

on W_1 i W_2 denoten el temps que inverteix un client a cadascuna de les dues cues. Per tant, tornant a la inequació d'abans veiem que:

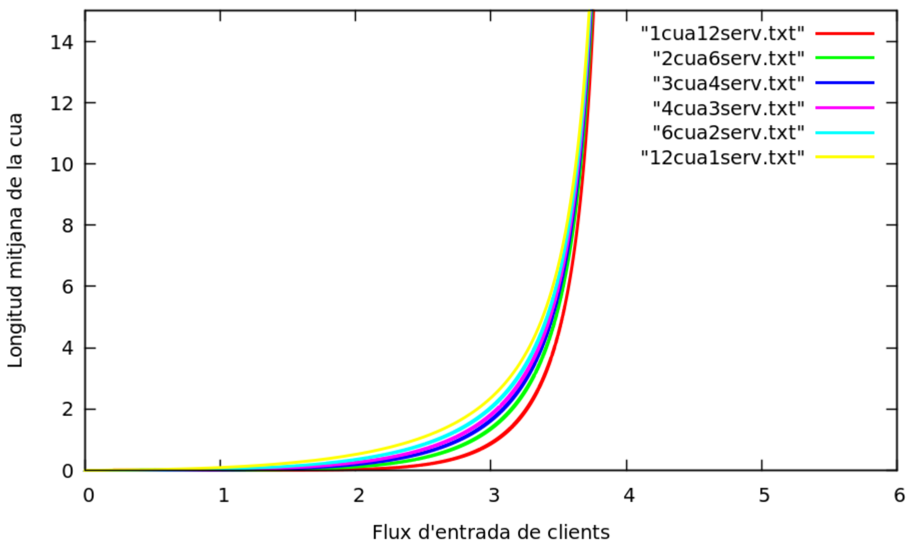
$$\lambda E(W_1) < \frac{\lambda}{4} E(W_2) \text{ , per tant } E(W_1) < \frac{1}{4} E(W_2)$$

i això ens mostra que $E(W_1) < E(W_2)$ cosa que ens reafirma l'argument que havíem donat abans.

Podem fer moltes altres proves, per exemple, si tenim un sistema amb 12 servidors, ens podria interessar veure de quina manera hem de distribuir aquests 12 servidors per tal que el sistema sigui el màxim d'òptim possible. Podem distribuir els 12 servidors en el nostre sistema de sis formes diferents:

1. Dotze cues independents amb un servidor a cada cua.
2. Sis cues independents amb dos servidors a cada cua.
3. Quatre cues independents amb tres servidors a cada cua.
4. Tres cues independents amb quatre servidors a cada cua.
5. Dues cues independents amb sis servidors a cada cua.
6. Una cua amb dotze servidors.

D'aquesta manera podem veure, utilitzant el mateix programa d'abans, el següent:



Aquest gràfic, com era d'esperar ens mostra que la distribució més optima és tenir una sola cua amb dotze servidors, veiem que a mesura que augmentem el flux d'entrada de clients, la longitud mitjana de la cua va augmentant, però en tot moment, el sistema amb una sola cua i dotze servidors té un nombre mitjà de clients a la cua inferior a la resta de distribucions. Aquest fet ens fa arribar a una conclusió més general.

Podem observar, que si tu vols dissenyar un sistema de capacitat i infinita, seguint una disciplina FIFO, amb n servidors i que els clients arriben seguint una distribució Poisson de paràmetre , i els temps de servei estan distribuïts exponencialment amb paràmetre μ, el model de servei més eficient serà $M/M/n$. La resta de models, factibles, que hem pogut treballar abans són poc eficaços.

5.4 Sistema de cues $M/M/\infty$

El sistema de cues $M/M/\infty$ és un model que suposa que el sistema té infinits servidors. Aquest sistema, de la mateixa manera que el sistema $M/M/n$ tindria una única cua on els clients d'aquesta es col·locarien als servidors que quedessin lliures. Tot hi això, la idea d'aquest model de cues, és un sistema on els clients arriben i no han d'esperar.

Per treballar-lo utilitzarem les mateixes notacions i hipòtesis dels apartats anteriors. Els índexs d'arribades, de serveis, disciplina i capacitat.

De la mateixa manera que els altres sistemes utilitzats, el nombre de clients al sistema $(N(t), t \leq 0)$ és un procés de naixement i mort. En el cas d'aquest sistema, els índexs infinitesimals seran els següents:

$$\lambda_k = \lambda \ \forall k \in \mathbb{N},$$

$$\mu_k = k\mu \ \forall k \in \mathbb{N}.$$

Construïts a partir del mateix argument que el model $M/M/n$, el que passa és que a l'haver-hi servidors il·limitats no hi ha el problema que els servidors quedin tots ocupats.

El que ens interessa ara un cop trobats els índexs infinitesimals és veure quina és la condició d'estabilitat per tal de poder definir la distribució estacionaria d'aquest procés.

Hem de veure quan:

$$S_1 = 1 + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \frac{\lambda_k}{\mu_{k+1}} < \infty.$$

Per tant, sabent els valors de λ_k i μ_k, i recordant que $\rho = \frac{\lambda}{\mu}$ tenim que:

$$1 + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \frac{\lambda}{(k+1)\mu} = 1 + \sum_{n=1}^{\infty} \frac{\rho^n}{n!} = e^{\rho} < \infty.$$

Veiem que $e^{\rho} < \infty$ per tot λ i per tot μ donats.

Podem veure doncs, que el sistema no es congestiona, i això és degut al fet de tenir infinits servidors.

Veiem doncs que per tot λ i μ podem definir una distribució estacionaria de la manera següent:

$$\pi_k = \pi_0 \prod_{n=0}^{k-1} \frac{\lambda}{(n+1)\mu} = \pi_0 \frac{\rho^k}{k!},$$

$$\pi_0 = 1 + \sum_{k=1}^{\infty} \frac{\rho^k}{k!} = e^{\rho},$$

per tant, tenim que:

$$\pi_k = \frac{\rho^k}{k!} e^{-\rho}.$$

Veiem doncs, que N que denota el nombre de clients al sistema en el cas estacionari, segueix una distribució Poisson de paràmetre ρ.

Les mesures de rendiment en aquest model concret són bastant senzilles de calcular i trobar. Analitzarem primer el nombre mitjà de clients al sistema, sabem que N segueix una distribució Poisson, per tant:

$$E(N) = \rho.$$

En aquest sistema, tal com hem dit abans, no hi ha cua pel fet que hi hagi infinits servidors, per tant, tenim que $E(Q) = 0$. En conseqüència tenim també que $E(W) = 0$, ja que els clients no inverteixen temps a la cua, van directe a un servidor. Com a conclusió podem dir, doncs, $E(T) = E(S) = \frac{1}{\mu}$.

5.5 Model de cues $M/M/n/n$.

Aquest sistema és el més vell i més famós de la teoria de cues. L'origen de la teoria de trànsit o teoria de la congestió va començar investigant aquest sistema.

Com és d'esperar, els clients arriben seguint una llei Poisson i els temps de servei estan distribuïts exponencialment. En aquest sistema es suposa que si els n servidors estan ocupats quan un nou client arriba, el client es perd.

Considerem el procés $(N(t), t \leq 0)$, que ens denota, com els altres sistemes estudiats, el nombre de clients que hi ha al sistema a l'instant t, o cosa que és el mateix, el nombre de servidors ocupats a l'instant t. Utilitzant els mateixos raonaments que als apartats anteriors, és fàcil demostrar que es tracta d'un procés de naixement i mort amb índexs infinitesimals:

$$\lambda_k = \begin{cases} \lambda & \text{si } k < n, \\ 0 & \text{si } k \geq n, \end{cases}$$

$$\mu_k = k\mu, \; k = 1, 2, ..., n.$$

Veiem clarament, que el sistema no s'arriba a congestionar mai pel fet que ens trobem dins un sistema amb capacitat finita. És per això que per qualsevol índex d'arribada i sortida podem definir una distribució estacionaria del sistema.

Aquesta està definida de la següent manera:

$$\pi_k = \begin{cases} \pi_0 \left(\frac{\lambda}{\mu}\right) \frac{1}{k!} & si\, k < n, \\ 0 & si\, k \geq n. \end{cases}$$

on,

$$\pi_0 = \left(\sum_{k=0}^{n} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} \right)^{-1}.$$

Es té, que la mesura més important del sistema és el càlcul de la probabilitat que un client es perdi, és a dir, la probabilitat que els n servidors estiguin ocupats. Aquesta, tenint en compte com els altres sistemes estudiats que $\rho = \frac{\lambda}{\mu}$, doncs, és:

$$\pi_n = \frac{\frac{\rho^n}{n!}}{\sum_{k=0}^{n} \frac{\rho^k}{k!}} = B(n, \rho).$$

Aquesta fòrmula va ser introduida per Erlang i va ser anomanada com **fórmula Erlang B**.

Una manera de construir $B(n,\rho)$ és de forma recursiva. De la següent manera:

$$B(n,\rho) = \frac{\frac{\rho^n}{n!}}{\sum_{i=0}^{n} \frac{\rho^i}{i!}} = \frac{\frac{\rho}{n}\frac{\rho^{n-1}}{(n-1)!}}{\sum_{i=0}^{n} \frac{\rho^i}{i!} + \frac{\rho}{n}\frac{\rho^{n-1}}{(n-1)!}} =$$

$$= \frac{\frac{\rho}{n}B(n-1,\rho)}{1+\frac{\rho}{n}B(n-1,\rho)} = \frac{\rho B(n-1,\rho}{n+\rho B(n-1,\rho)}.$$

Usarem $B(1,\rho) = \frac{\rho}{1+\rho}$ com a valor inicial per tal de començar la recurssió. És important utilitzar aquest mètode per tal de calcular $B(n,\rho)$, ja que el càlcul directe d'aquest element pot causar algun problema degut al valor del factorial.

Anem a veure alguns resultats sobre aquest sistema de cues. Estudiem les mesures de rendiment del model.

Podem calcular el nombre mig de clients dins el sistema:

$$E(N) = \sum_{k=0}^{n} k\pi_k = \sum_{k=0}^{n} k\frac{\rho^k}{k!}\pi_0 = \rho\sum_{k=0}^{n-1}\frac{\rho^k}{k!}\pi_0 = \rho(1-\pi_n).$$

Podem veure clar, que en aquest sistema en concret, el nombre de clients dins el sistema coincideix amb el nombre de servidors ocupats, ja que és un model que té la particularitat que no es forma cap cua dins el sistema. És per això, que podem dir que:

$$E(N) = E(n_o).$$

on n_o correspon a la variable aleatoria que ens defineix el nombre de servidors ocupats del sistema.

Podem parlar sobre els cicles d'ocupació i de desocupació del sistema. Tal com hem vist quan hem estudiat el sistema $M/M/1$, es té que:

$$1-\pi_0 = \frac{E(T_o)}{E(T_d)+E(T_o)}.$$

on T_d i T_o representen la longitud d'un cicle de desocupació i ocupació del servei respectivament.

Veiem clar que $E(T_d)=\frac{1}{\lambda}$, ja que un servidor està ocupat fins que arriba un nou client al servei. És a dir, que $T_d \sim exp(\lambda)$.

Partint d'això, podem veure que:

$$E(T_o) = \frac{1-\pi_0}{\lambda\pi_0} = \frac{\sum_{i=1}^{n}\frac{\rho^i}{i!}}{\lambda\left(1+\sum_{i=1}^{n}\frac{\rho^i}{i!}\right)}.$$

Igual que en el model de cues $M/M/\infty$ que hem treballat anteriorment, veiem clar que no és de cap utilitat estudiar la llargada mitjana de la cua o estudiar el temps invertit d'un client a la cua i al sistema, a causa del fet que en aquests tipus de models, no hi ha la possibilitat que s'hi formi una cua.

Un client pot trobar-se amb dues situacions: Que quan arriba al sistema, hi ha servidors lliures, i per tant, en aquest cas, és servit directament, o, que quan arriba al sistema, es troba tots els servidors ocupats, i, en comptes d'afegir-se a la cua, tal com hem suposat quan hem treballat el sistema $M/M/n$, el sistema rebutja el client, i per tant el perdem (degut a que la capacitat del sistema és de n clients, que coincideix amb el nombre de servidors).

És per això, que és fàcil deduir, que si un client entra dins el sistema el temps que invertirà en ell, serà el temps que durarà el seu servei. Per tant, $E(T) = E(S) = \frac{1}{\mu}$.

6 Models de cues amb font finita.

Els models de cues tractats amb anterioritat són sistemes on les arribades segueixen una llei Poisson procedents d'una font infinita de clients. En aquest nou apartat ens centrarem en algun model de cues amb població finita. Aquests models són importants des d'un punt de vista pràctic, ja que en moltes situacions la font de clients és finita. Centrarem l'estudi d'aquests models investigant un exemple del que s'anomena problema d'interferència de la màquina tractat per molts experts.

Hem de considerar n màquines que operen independentment les unes de les altres. Després que una màquina falli és reparada per un o multiples reparadors segons la disciplina en la qual ens trobem. Un cop reparada la maquina torna a ser operativa de nou i es torna a començar el procés.

Aquests models simples tenen moltes aplicacions en diversos camps, per exemple en fabricació, ciència computacional, ciència de la gestió, etc.

Seguidament anirem a tractar un d'els models de cues finit per tenir una petita idea de com funcionen i es desenvolupen aquests tipus de models.

6.1 Sistema de cues $M/M/r/r/n$.

Podem veure fàcilment, que aquest model funciona de la mateixa manera que el model $M/M/r/r$, estudiat anteriorment, però amb la condició afegida que la font de clients és finita, en aquest cas, una població de cardinal n. Depenent de la capacitat del sistema, denotada per r, un client pot trobar el sistema ple. En el cas de font infinita, hem pogut observar, que si un client troba tots els servidors plens, aquest client es perd. Però en el cas del model que estudiarem seguidament, aquest client, retorna a la font de clients i s'hi queda un cert temps, seguint una distribució exponencial. Pel fet que les variables amb les quals treballarem estan distribuïdes exponencialment, tenim que el nombre de clients dins el sistema és un procés de naixement i mort amb índex infinitesimals:

$$\lambda_k = (n-k)\lambda \ , \ 0 \leq k < r,$$

$$\mu_k = k\mu \ , \ 1 \leq k \leq r.$$

Utilitzant el mateix ranoment que quan hem estudiat el model $M/M/r/r$, veiem que per tot $r \geq n$ el sistema no es pot congestionar mai. Per tant podem definir i utilitzar sense cap problema la distribució estacionaria del sistema.

Veiem que és la següent:

$$\pi_k = \binom{n}{k}\rho^k \pi_0 \quad 0 \leq k \leq r,$$

$$\pi_0 = \left(\sum_{i=0}^{r} \binom{n}{i} \rho^i \right)^{-1}.$$

Aquesta distribució s'anomena una binominal truncada o **distribució Engset**. En el cas especial en què $r = n$ que correspon al cas en què tot client té el seu propi servidor té una distribució pròpia que ens pot semblar interessant:

$$\pi_k = \frac{\binom{n}{k}\rho^k}{\sum_{i=0}^{n}\binom{n}{i}\rho^i} = \frac{\binom{n}{k}\rho^k}{(1+\rho)^n} =$$

$$= \binom{n}{k} \left(\frac{\rho}{1+\rho}\right)^k \left(1 - \frac{\rho}{1+\rho}\right)^{n-k}.$$

Veiem que correspon a una distribució binomial amb paràmetre $p = \frac{\rho}{1+\rho}$.

De la mateixa manera que abans, és bastant senzill d'estudiar quines són les mesures de rendiment del sistema.

Com ja hem fet abans comprovarem quina és la mitjana de clients del sistema:

$$E(N) = \sum_{k=0}^{r} k\pi_k,$$

i veiem que es compleix que, per les mateixes raons donades quan s'ha estudiat el model $M/M/n/n$, tenim que:

$$E(r_o) = E(N),$$

on la variable aleatòria r_o determina el nombre de servidors que estan sent utilitzats.

Podem parlar, també del nombre mitjà de clients que est troben a la font, és a dir, fora el sistema. Denotarem amb la variable m el nombre de clients fora el sistema:

$$E(m) = n - E(N).$$

També podria ser interessant centrar-nos en la proporció esperada de servidors ocupats denotada per la variable U_s i la proporció esperada de clients a la font, denotada per la variable U_f.

$$U_s = \frac{E(r)}{r},$$

$$U_f = \frac{E(m)}{n} = \frac{E(T_f)}{E(T_f) + \frac{1}{\mu}},$$

on $E(T_f)$ denota el temps mitjà que un client destina a la font. Aquesta esperança, doncs, es pot obtenir de la següent manera:

$$E(T_f) = \frac{U_f}{\mu(1 - U_f)}.$$

Parlem seguidament de la probabilitat que un client en arribar es trobi el sistema de capacitat n i amb r servidors ple (notarem aquesta probabilitat com $P_B(n,r)$. Utilitzarem la fórmula de Bayes per tal de calcular aquesta probabilitat.

Si considerem que la variable A_i fa referència a l'esdeveniment que en el sistema hi hagi i client, i B fa referència a l'esdeveniment que un client de la font accedeixi al sistema, aleshores, la probabilitat que volem calcular serà:

$$P_B(n,r) = P(A_r|B) = \frac{P(B|A_r)P(A_r)}{\sum_{i=0}^{r} P(B|A_i)P(A_i)}.$$

Si denotem seguidament amb $\pi_k(n,r)$ la probabilitat que a un sistema $M/M/r/r/n$ hi hagi k clients, tenim:

$$P_B(n,r) = \lim_{h \to 0} \frac{((n-r)\lambda h + o(h))\pi_k(n,r)}{\sum_{i=0}^{r}((n-i)\lambda h + o(h))\pi_i(n,r)} = \frac{(n-r)\pi_r(n,r)}{\sum_{i=0}^{r}(n-i)\pi_i(n,r)} =$$

$$= \frac{(n-r)\binom{n}{r}\rho^r}{\sum_{i=0}^{r}(n-i)\binom{n}{i}\rho^i} = \frac{(n-r)\frac{n!}{r!(n-r)!}\rho^r}{\sum_{i=0}^{r}(n-i)\frac{n!}{i!(n-i)!}\rho^i} = \frac{\frac{(n-1)!}{r!(n-1-r)!}\rho^r}{\sum_{i=0}^{r}\frac{(n-1)!}{i!(n-1-i)!}\rho^i} =$$

$$= \frac{\binom{n-1}{r}\rho^r}{\sum_{i=0}^{r}\binom{n-1}{i}\rho^i} = \pi_r(n-1,r).$$

Veiem, per tant, que la probabilitat que un client que vol entrar es trobi el sistema (de capacitat n i r servidors) ple és igual a la probabilitat que en un sistema de capacitat $n-1$ i r servidors hi hagi r clients dins el sistema.

Aquesta probabilitat la denotarem per $E(n,r,\rho)$ que s'anomena **fórmula de pèrdua Engest**. Seguidament mostrarem com es pot obtenir aquesta fórmula de manera recursiva:

$$E(n,r,\rho) = \frac{\binom{n-1}{r}\rho^r}{\sum_{i=0}^{r}\binom{n-1}{i}\rho^i} = \frac{\binom{n-1}{r-1}\frac{n-r}{r}\rho^r}{\sum_{i=0}^{r-1}\binom{n-1}{i}\rho^i + \binom{n-1}{r-1}\frac{n-r}{r}\rho^r} =$$

$$= \frac{\frac{\binom{n-1}{r-1}\rho^r}{\sum_{i=0}^{r-1}\binom{n-1}{i}\rho^i}\frac{n-r}{r}}{\frac{\sum_{i=0}^{r-1}\binom{n-1}{i}\rho^i + \binom{n-1}{r-1}\frac{n-r}{r}\rho^r}{\sum_{i=0}^{r-1}\binom{n-1}{i}\rho^i}} = \frac{\frac{n-r}{r}\rho E(n,r-1,\rho)}{1+\frac{n-r}{r}\rho E(n,r-1,\rho)} = \frac{(n-r)\rho E(n,r-1,\rho)}{r+(n-r)\rho E(n,r-1,\rho)}.$$

El valor inicial que prendrem per construir la recurrència:

$$E(n,1,\rho) = \frac{\frac{(n-1)!}{1!(n-1-1)!}\rho}{\sum_{i=0}^{1}\frac{(n-1)!}{i!(n-1-i)!}\rho^i} = \frac{(n-1)\rho}{1+(n-1)\rho}.$$

Podríem ara desenvolupar les diferents mesures de rendiment estudiades abans en el cas particular en què $r=n$. En aquest cas hem vist que N segueix una distribució binomial de paràmetre $\frac{\rho}{1+\rho}$, per tant tenim que:

$$E(N) = n\frac{\rho}{1+\rho}.$$

A partir d'aquest resultat, podem desenvolupar tots els altres resultats treballats anteriorment.

$$E(m) = \frac{n}{1+\rho},\ U_s = \frac{\rho}{1+\rho},\ U_f = \frac{1}{1+\rho},\ E(T_f) = \frac{1}{1+\rho}.$$

També podem veure clarament que si la capacitat del sistema és de n clients, tots els clients podran accedir al sistema. Per tant, no hi ha la possibilitat que un client de la font es quedi fora del sistema, per tant es té que $P_B(n,n)=0$.

Tornant al cas general, veiem que en aquest sistema és impossible que s'hi formi una cua, ja que està dissenyat per tal que si tots els servidors estan ocupats, no pugui entrar cap més client al sistema fins que un dels servidors quedi lliure. És per això, doncs, que el temps que un client inverteix al sistema correspon al temps que inverteix un client a ser servit. Per tant, obtenim el següent resultat:

$$E(T) = \frac{1}{\mu}.$$

7 Xarxes de cues

En aquest capítol es realitzarà una introducció al tema de les xarxes de cues. Aquesta és una àrea de gran interès investigador i d'aplicació, amb problemes molt complicats de plantejar i de resoldre.

Les xarxes de cues es poden descriure com un grup de nodes (suposarem que són k) connectats entre si, on cada node representa una instal·lació de servei. Quan un client abandona un node, es pot unir a un altre o pot abandonar la xarxa.

Cada node pot constar de c_i servidors on $i = 1, ..., k$. En el cas més general, els clients poden entrar per qualsevol node i, moure's per la xarxa i sortir per qualsevol node.

No necessàriament tots els clients han d'entrar i sortir pels mateixos nodes, ni han d'agafar el mateix camí un cop han entrat. Els clients poden tornar a un node visitat anteriorment, saltar-se alguns nodes, i fins i tot romandre en un mateix node per sempre.

Les xarxes de cues més simples i amb resultats més significatius s'anomenen les **Xarxes de Jackson**. En aquest capítol ens centrarem a estudiar-les.

Aquestes xarxes tenen les següent característiques:

1. Les arribades des de l'exterior al node i segueixen un procés de Poisson de paràmetre γ_i.
2. Els temps de servei en cada node i són independents i segueixen una distribució exponencial de paràmetre μ_i, que podria ser depenent de l'estat.
3. La probabilitat que un client que hagi completat el seu servei en un node i, vagi al node j és r_{ij} amb $i = 1, ..., k$, $j = 0, ..., k$
4. $r_{i,0}$ indica la probabilitat que un client abandoni la xarxa des del node i.

Les xarxes les quals compleixen que $\gamma_i = 0$, $\forall i \in 1, ..., k$, la qual cosa significa que no entra cap client a dins el sistema des de l'exterior, i que $r_{i0} = 0$, $\forall i$, és a dir, que cap client abandona el sistema, s'anomenen **xarxes de Jackson tancades**. El cas general, que hem descrit anteriorment és referit com **xarxes Jackson obertes**.

En aquest capítol Ens podríem centrar a estudiar xarxes on cada node i consta de $c_i > 1$ servidors, però l'estudi d'aquestes seria bastant complex i . Com que l'objectiu d'aquest apartat és donar una breu introducció de la teoria de xarxes de cues i uns petits resultats, simplificarem l'estudi del comportament d'aquestes xarxes en el cas que cada node i tingui tan sols un servidor i la capacitat d'aquest node sigui infinita.

Primer de tot anirem a descriure un cas particular de les xarxes obertes.

7.1 Cues tàndem

Les cues tàndem tenen les següents particularitats:

$$\gamma_i = \begin{cases} \lambda & \text{si } i = 1, \\ 0 & \text{si } i \neq 1. \end{cases}$$

i

$$r_{ij} = \begin{cases} 1 & \text{si } j = i+1 \text{ i } 1 \le i \le k-1, \\ 1 & \text{si } i = k \text{ i } j = 0, \\ 0 & \text{en els altres casos }. \end{cases}$$

Els nodes es poden veure com una sèrie de sistemes on els clients circulen en una sola direcció de node a node. Els clients poden entrar al sistema des de fora només pel node 1 i sortir només pel node k.

Treballarem el cas en què les arribades a la xarxa segueixen una distribució Poisson de paràmetre λ i els temps de servei de cada node estan distribuïts exponencialment amb paràmetre μ. També considerarem únicament el cas en què la capacitat entre cada etapa de la xarxa és infinita.

El primer node, per tant, és un sistema que segueix un model $M/M/1$. Per tal de veure com es comporta aquest tipus de xarxa i poder modelitzar-la, és necessari obtenir la distribució de sortida (la distribució dels temps de servei) per tal de trobar quina serà la distribució d'arribada a la següent estació.

Anteriorment, quan hem estudiat el model de cues $M/M/1$, hem observat que la igualtat 5.1 ens mostra que la distribució dels temps de sortida dels clients d'un sistema de cues que segueix model $M/M/1$ és idèntica a la dels temps d'arribada dels clients al mateix sistema, és a dir, que els temps d'arribada i els temps de servei dels clients segueixen una distribució exponencial de paràmetre λ. Per tant, podem assumir que dins aquest sistema tots els nodes són models $M/M/1$ independents.

7.2 Xarxes de Jackson obertes

En aquest tipus de xarxes, com hem dit abans, compleixen:

1. L'arribada externa a qualsevol node i segueix una distribució Poisson de paràmetre γ_i.
2. El servidor de cada etapa té un servei exponencial de paràmetre μ_i.
3. De cada etapa i un client canvia a una altra etapa amb una probabilitat r_{ij} i a l'exterior amb una probabilitat r_{i0}

Introduïm la notació següent que ens servirà per treballar el model que volem estudiar.

Considerarem N_i el nombre de clients al node i.

Sense cap problema podríem assumir que la taxa de serveis depèn de l'estat en què es troba, és a dir $\mu_i = \mu_i(N_i)$. Per simplificar la notació suposarem a continuació una taxa de serveis constant μ_i.

Denotarem també λ_i com la taxa d'arribades de clients al node i.

Encara que les arribades externes als nodes segueixen una distribució Poisson, això no ens assegura que les arribades als nodes des de l'interior de la xarxa segueixin també una distribució Poisson. En general no ho faran.

El valor del flux λ_i es troba considerant que està compost pel flux d'arribada de clients al node i des de l'exterior de la xarxa i des dels altres nodes. Està definit de la següent manera:

$$\lambda_i = \gamma_i + \sum_{j=1}^{k} \lambda_j r_{ij} \text{ per } i = 1, ..., k.$$

Aquestes equacions s'anomenen les equacions de trànsit, i podem considerar:

$$\vec{\lambda} = (\lambda_1, ..., \lambda_k),$$

$$\vec{\gamma} = (\gamma_1, ..., \gamma_k),$$

$$R = \{r_{ij}\}_{(i,j)\in\{1,...,k\}\times\{1,...,k\}},$$

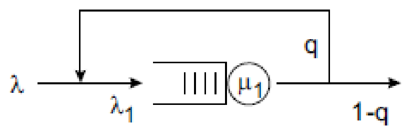
de manera que llavors tenim que:

$$\vec{\lambda} = \vec{\gamma} + \vec{\lambda}R,$$

i per tant, es té que:

$$\vec{\lambda} = \vec{\gamma} + (I - R)^{-1}.$$

Exemple: Suposem que tenim una xarxa de cues que segueix el següent diagrama:



Volem calcular el flux d'entrada al servei. Tenim que:

$$\lambda_1 = \lambda + q\lambda_1,$$

$$\lambda_1 = \frac{\lambda}{1-q}.$$

7.2.1 Distribució estacionaria

En aquest apartat donarem un resultat que ens pot ser interessant a l'hora de treballar amb xarxes de Jackson obertes. Definirem, la distribució estacionaria del sistema i sota quines condicions es dóna.

Primer de tot suposarem que el nombre de clients N_i als diferents nodes, $i = 1, ..., k$, són independents. Considerarem també, que tot node i es comporta com si les arribades a aquest seguissin un procés de Poisson de paràmetre λ_i.

Direm que la xarxa està en equilibri si per tots els k sistemes de cues que configuren que la formen estan en equilibri.

Recordem que treballarem el cas en què cada node està format únicament per un sol servidor.

Per tractar aquest problema suposarem que cada node i de la xarxa es comporta com si seguís un model de cues $M/M/1$, i per tant, pel fet d'haver estudiat aquest model amb anterioritat, sabem que existeix una solució d'equilibri per aquest sistema si $\rho_i = \frac{\lambda_i}{\mu_i} < 1$.

Si es dóna aquesta condició per tot $i = 1, ..., k$ podem construir una distribució estacionaria per estudiar el sistema.

Considerem, també, alguna notació que ens serà útil. L'estat de la xarxa està determinat pel vector $N = (N_1, ..., N_k)$. Els possibles valors d'aquest estat els denotem per $n = (n_1, ..., n_k)$.

Direm que la xarxa està en l'estat n si $N = n$, és a dir, $N_1 = n_1$,...,$N_k = n_k$.

La distribució estacionaria del sistema, estarà definida com:

$$\pi(n) = P(N = n).$$

de manera que compleix que

$$\pi(n) = 0 \text{ si per alguna } i \text{ , } n_i < 0.$$

La distribució estacionaria d'una xarxa, en general, és un càlcul bastant complex, però en el cas que tots els nodes es comportin com a models de cues $M/M/1$, existeix un teorema, el teorema de Jackson, que ens descriu la distribució estacionaria en aquest cas. No ens centrarem en la demostració del teorema, ja que no és l'objectiu d'aquest apartat. La demostració d'aquest es pot trobar en (5).

Teorema 7.1 (Teorema de Jackson). *En una xarxa de Jackson oberta de k nodes, cada node i segueix un model de cues $M/M/1$ amb $\rho_i < 1$, podem definir, la distribució estacionaria d'aquest sistema com:*

$$\pi(n) = \pi(n_1, ..., n_k) = \prod_{i=1}^{k} \pi_i(n_i) \text{ on } \pi_i(n_i) = (1 - \rho_i)\rho_i^{n_i}.$$

7.2.2 Mesures de rendiment

Assumint que ens trobem en l'estat estacionari del model de xarxes de Jackson obertes, podem trobar resultats importants i significatius per tal de treballar i estudiar el comportament de la xarxa.

Hem comentat abans, que cada node i té un sol servidor, es comporta com un model de cues $M/M/1/$, per tant, té sentit, que les mesures de rendiment de dins de cada node seran les corresponents a un model de cues $M/M/1$. Per exemple, tenim doncs que:

$$E(N_i) = \frac{\rho_i}{1 - \rho_i}.$$

Podem utilitzar els resultats estudiats en apartats anteriors per trobar també quin serà el valor de $E(N_i)$.

És evident, que un cop trobat $E(N_i)$ per $i = 1, ..., k$, podem trobar $E(N)$ de la següent manera:

$$E(N) = \sum_{i=1}^{k} E(N_i).$$

Utilitzant la llei de Little podem trobar el temps mitjà que inverteixen els clients a cadascun dels nodes. Notem el temps mitjà que inverteixen els clients al node i com:

$$E(T_i) = \frac{E(N_i)}{\lambda_i}.$$

Si estudiem ara el temps total que inverteix un client a la xarxa, tenim que:

$$E(T) = \frac{E(N)}{\lambda}.$$

Si definim W_i com el temps que un client espera al node i abans de ser servit, podem considerar que:

$$E(W_i) = E(T_i) - \frac{1}{\mu_i}.$$

7.2.3 Eficiència de la xarxa

Donada una xarxa de Jackson oberta, ens interessa que aquesta sigui el màxim eficient possible. Anem a veure una manera de fer eficient la xarxa.

Suposem que la xarxa es comporta com si estigués composta per k sistemes de cues independents que segueixen un model $M/M/1$, i que tots els nodes estan en equilibri (podem definir una distribució estacionaria).

El que ens interessa és minimitzar el temps mitjà que els clients inverteixen a la xarxa, i per tant, el nombre mitjà de clients a la xarxa.

Fixarem els valors de λ_i i assumim que podem escollir lliurement els valors de μ_i , $\forall i$ excepte per la restricció $\sum_{i=1}^{k} \mu_i = C$.

Definim doncs, una funció N tal que:

$$N(\mu_1, ..., \mu_k) = \sum_{i=1}^{k} \frac{\rho_i}{1-\rho_i} = \sum_{i=1}^{k} \frac{\lambda_i}{\mu_i - \lambda_i}.$$

Volem trobar el mínim d'aquesta funció.

Utilitzant el mètode dels multiplicadors de Lagrange, definim una funció H de manera que:

$$H(\mu_1, ..., \mu_k, x) = \sum_{i=1}^{k} \frac{\lambda_i}{\mu_i - \lambda_i} + x \left(\sum_{i=1}^{k} \mu_i - C \right).$$

Busquem els punts crítics d'aquesta funció. Per trobar-los busquem les solucions del sistema:

$$\begin{cases} \frac{\partial H}{\partial \mu_1} = 0, \\ \quad ... \\ \frac{\partial H}{\partial \mu_k} = 0, \\ \sum_{i=1}^{k} \mu_i - C = 0. \end{cases}$$

Tenim que:

$$\frac{\partial H}{\partial \mu_i} = -\frac{\lambda_i}{(\mu_i - \lambda_i)^2} + x = 0 \;\Rightarrow\; \mu_i = \lambda_i + \left(\frac{\lambda_i}{x} \right)^{1/2}.$$

Utilitzant $\sum_{j=1}^{k} \mu_j - C = 0$, tenim que:

$$\sum_{j=1}^{k} \lambda_j + \left(\frac{\lambda_j}{x} \right)^{1/2} - C = 0 \;\Rightarrow\; \frac{1}{x^{1/2}} = \frac{C - \sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{k} \lambda_j^{1/2}}.$$

I per tant, tenim que μ_i adients que hem de prendre per tal que la nostra xarxa sigui el màxim adient possible és:

$$\mu_i = \lambda_i + \frac{\lambda_i^{1/2}}{\sum_{j=1}^{k} \lambda_j^{1/2}}(C - \sum_{j=1}^{k} \lambda_j).$$

7.3 Xarxes de Jackson tancades

Com hem dit abans, si una xarxa de Jackson oberta compleix que $\gamma_i = 0$ i $r_{i0} = 0$ $\forall i$ tenim un una xarxa de Jackson tancada. A diferència de les xarxes obertes, una xarxa tancada no té una font exterior, i els clients no abandonen mai la xarxa. Per tant, podem dir que hi ha una població constant de M clients dins la xarxa.

De la mateixa manera que les xarxes obertes, tenim que cada node i és un sistema de cues amb disciplina FIFO, on els temps de serveis són independents i segueixen una distribució exponencial de paràmetre μ_i. Aquests paràmetres poden dependre de l'estat en què es trobi el sistema, és a dir, $\mu_i = \mu_i(n_i)$.

Un client que abandona un node i, va cap al node j amb una probabilitat r_{ij}.

En aquest tipus de xarxes, el flux d'entrada de clients a un node, és similar al de les xarxes obertes, però en aquest cas, no arriben clients des de la font exterior, per tant:

$$\lambda_i = \sum_{j=1}^{k} \lambda_j r_{ji} \ , \ i = 1, ..., k.$$

Aquestes igualtats constitueixen un sistema d'equacions lineals homogènies.

El problema en què ens trobem és que qualsevol d'aquestes equacions és linealment dependent de totes les altres.

Suposem que $(\hat{\lambda_1}, ..., \hat{\lambda_k})$ és una solució. Sabem que donat $\alpha \in \mathbb{R}$, $\alpha(\hat{\lambda_1}, ..., \hat{\lambda_k})$ també serà solució del sistema- Tot hi això, més endavant veurem que donada una solució particular del sistema $(\hat{\lambda_1}, ..., \hat{\lambda_k})$, existeix un $\hat{\alpha}$ que ara mateix no està determinat, però que correspon al valor de la constant que juntament amb la solució particular del sistema podem obtenir la solució general, que ens donarà els fluxos d'entrada reals, que seran de la forma:

$$\lambda_i = \hat{\alpha}\hat{\lambda_i}.$$

Denotarem $\hat{\rho_i} = \frac{\hat{\lambda_i}}{\mu_i}$.

7.3.1 Anàlisis del valor mig

Quan hem estudiat les xarxes obertes, tots els resultats que hem obtingut han sigut assumint que els nodes es comportaven com uns certs models de cues, però, el problema, és que aquestes suposicions fan que els resultats que obtenim siguin només una aproximació del resultat real.

En aquest capítol intentarem trobar els resultats reals de les mesures de rendiment treballades anteriorment.

Donat un node i qualsevol, el nostre objectiu serà trobar el nombre mig real de clients $E(N_i[M])$ i el temps mitjà real invertit pel client dins el sistema $E(T_i[M])$, i també el valor real dels fluxos d'entrada de clients al sistema λ_i.

Abans de començar considerem el següent teorema que utilitzarem seguidament. No el demostrarem pel fet que és tant sols una eina utilitzada i no és el nostre objectiu.

Teorema 7.2 (Teorema d'arribades a una xarxa tancada)**.** *A una xarxa tancada amb M clients, la probabilitat que un client, en el moment que entra a qualsevol node, es trobi el sistema en un estat n, és la mateixa que la probabilitat d'equilibri $\pi(n)$ a una xarxa amb $M-1$ clients.*

Aquest teorema pot ser provat també de la mateixa manera per esperances i per xarxes obertes. Els detalls d'aquesta els ometrem.

Ens centrarem ara a estudiar el temps mitjà que inverteix un client al node i.

Aquest temps serà la suma entre el temps que el client està sent servit, i el temps que tarden a ser servits els clients que estan per davant seu. Per tant:

$$E(T_i[M]) = \frac{1}{\mu_i} + E(N_i^*[M])\frac{1}{\mu_i},$$

on $E(N_i^*[M])$ consisteix en el nombre mitjà de clients que es troba el client en el moment que arriba a la cua i.

Pel teorema d'arribades podem dir $E(N_i^*[M]) = E(N_i[M-1])$ Finalment, utilitzant aquest canvi, es té que:

$$E(T_i[M]) = (1 + E(N_i[M-1]))\frac{1}{\mu_i}.$$

Anem a determinar ara, el nombre mitjà de clients al node i.

Proposició 7.3. *Donada una solució $(\hat{\lambda_1}, ..., \hat{\lambda_k})$ del sistema d'equacions $\lambda_i = \sum_{j=1}^{k} \lambda_j r_{ji}$ i el temps mitjà que inverteix un client a el node i, es compleix que el nombre mitjà de clients enl node és:*

$$E(N_i[M]) = M\frac{\hat{\lambda_i} E(T_i[M])}{\sum_{j=1}^{k} \hat{\lambda_j} E(T_j[M])}.$$

Demostració. El flux d'arribada real de clients és $\lambda_i = \hat{\alpha}\hat{\lambda_i}$.Usant la llei de Litlle tenim que:

$$M\frac{\hat{\lambda_i} E(T_i[M])}{\sum_{j=1}^{k} \hat{\lambda_j} E(T_j[M])} = M\frac{\lambda_i[M] E(T_i[M])}{\sum_{j=1}^{k} \lambda_j[M] E(T_j[M])} = M\frac{E(N_i[M])}{\sum_{j=1}^{k} E(N_j[M])} =$$

$$= M\frac{E(N_i[M])}{M} = E(N_i[M]).$$

□

Usant la llei de Little, tenim també que:

$$\lambda_i[M] = \frac{E(N_i[M])}{E(T_i[M])}.$$

Algoritme: Donat una xarxa tancada, els resultats dels elements treballats anteriorment es poden trobar seguint el següent mètode iteratiu.

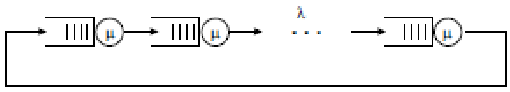
Començarem la recurssió amb $E(N_i[0]) = 0$. Podem assegurar-ho, ja que en una xarxa buida, el nombre mitjà de clients a qualsevol node és zero.

Una iteració consistirà en el següent:

1. $E(T_i[M]) = (1 + E(N_i[M-1]))\frac{1}{\mu_i}$.
2. $E(N_i[M]) = M\frac{\hat{\lambda}_i E(T_i[M])}{\sum_{j=1}^{k}\hat{\lambda}_j E(T_j[M])}$.
3. $\lambda_i[M] = \frac{E(N_i[M])}{E(T_i[M])}$.

A la segona equació, $(\hat{\lambda_1}, ..., \hat{\lambda_k})$ és qualsevol solució del sistema d'equacions $\lambda_i = \sum_{j=1}^{k}\lambda_j r_{ji}$.

Exemple: Considerem la següent xarxa amb k servidors.



Compleix que $\mu_i = \mu\,\forall i \in \{1, ..., k\}$.

Podem veure fàcilment que les equacions del flux d'arribades són les següents:

$$\lambda_i = \lambda_{i-1}\ ,\ i = 2, ..., k.$$

$$\lambda_1 = \lambda_k.$$

Tenim doncs, que una solució del sistema serà:

$$\lambda_1 = ... = \lambda_k = 1.$$

Començant amb el valor inicial $E(N_i[0]) = 0$, podem resoldre progressivament els valors mitjans de la xarxa per poblacions més grans.

$$\left\{\begin{array}{c} E(T_i[1]) = \frac{1}{\mu} \\ E(N_i[1]) = \frac{1}{k} \\ \lambda_i[1] = \frac{1}{k}\mu \end{array}\right. ,\ \left\{\begin{array}{c} E(T_i[2]) = \frac{k+1}{k\mu} \\ E(N_i[2]) = \frac{2}{k} \\ \lambda_i[1] = \frac{3}{k+2}\mu \end{array}\right. \ ...\ \left\{\begin{array}{c} E(T_i[M]) = \frac{k+M-1}{k\mu} \\ E(N_i[M]) = \frac{M}{k} \\ \lambda_i[1] = \frac{M}{k+M-1}\mu \end{array}\right.$$

8 Simulació d'un sistema de cues que segueix un model M/M/1

Ens podem trobar, que a l'hora de treballar alguns sistemes de cues, alguns problemes no es poden resoldre utilitzant mètodes analítics. Entre moltes altres raons podríem destacar el fet que no existeixin uns fluxos d'entrada i sortida específics, una gran complexitat del sistema o la disciplina de la cua.

A més, quan hem estudiat tots els sistemes de cues anteriors, hem considerat que ens trobàvem dins un sistema en equilibri on podem definir una distribució estacionària.

Tal com hem vist, no sempre es pot donar aquest cas, i per això les simulacions és una eina que ens pot ser útil en aquests casos a l'hora de trobar resultats.

En aquest capítol intentarem fer alguna simulació per tal de comprovar alguns dels resultats més importants obtinguts en els capítols anteriors.

Treballarem el model de cues més simple, el model $M/M/1$.

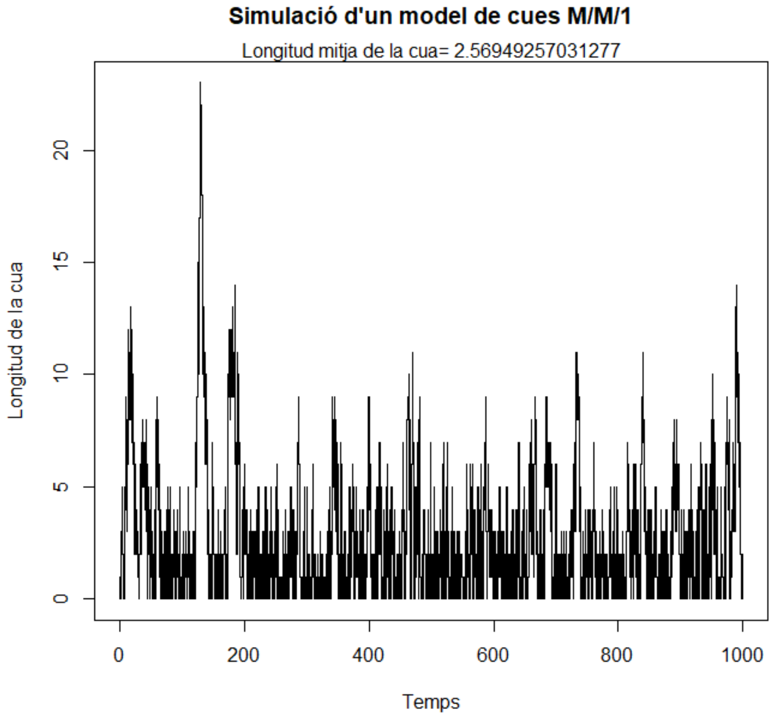
Per fer la simulació del sistema hem utilitzat l'entorn i el llenguatge de programació R, ja que aquest és un llenguatge amb un enfocament a l'anàlisi estadístic. La base de la nostra simulació serà la funció $rexp()$ del llenguatge R que generarà de manera aleatòria valors de la distribució exponencial, és a dir, el temps entre arribades i els temps de servei dels clients.

En la nostra simulació donarem un temps total que prendrem com a controlador de la simulació de manera que la simulació s'aturi quan arribem a aquest temps total. Donarem un flux d'entrada i de sortida de clients i observarem quina és la longitud de la cua en tot moment i la longitud mitjana esperada del sistema. Jugarem en tot moment que donat l'estat del sistema en què ens trobem, veure si es dóna abans una arribada d'un client o el servei del client que està essent servit.

S'han fet dues simulacions per tal de comprovar el comportament que hem observat anteriorment de manera analítica de la cua.

L'objectiu d'aquest apartat serà l'estudi de la longitud de la cua en cadascun dels diferents sistemes. Ens centrarem el resultat obtingut en estudiar la longitud mitjana de la cua de manera analítica i utilitzant la simulació.

Primerament hem considerat un sistema on el flux d'entrada i de sortida de clients són $\lambda = 3$ i $\mu = 4$ respectivament, hem obtingut el següent resultat:



Podem comprovar que aquest gràfic ens dóna informació de quina és la longitud de la cua en tot moment i en quins moments hi ha un canvi d'estat, és a dir, quan entra un client i quan en surt un.

També hem demanat que ens faci el càlcul de la longitud mitjana de la cua del sistema. En aquest cas, hem obtingut que aquesta és $E(Q) = 2.5695$.

Seguidament podríem comparar els resultats obtinguts en la simulació amb els que obtenim en utilitzar els mètodes analítics estudiant anteriorment.

Considerem un sistema de cues que segueix un model $M/M/1$, de manera que els clients arriben seguint una variable Poisson de paràmetre $\lambda = 3$ i els temps de servei estan distribuïts exponencialment amb paràmetre $\mu = 4$. Obtenim que $\rho = \frac{\lambda}{\mu} < 1$, per tant, podem definir una distribució estacionaria dins del sistema. Si la variable Q representa el nombre de clients a la cua, tenim que:

$$E(Q) = \frac{\rho^2}{1-\rho} = 2,25$$

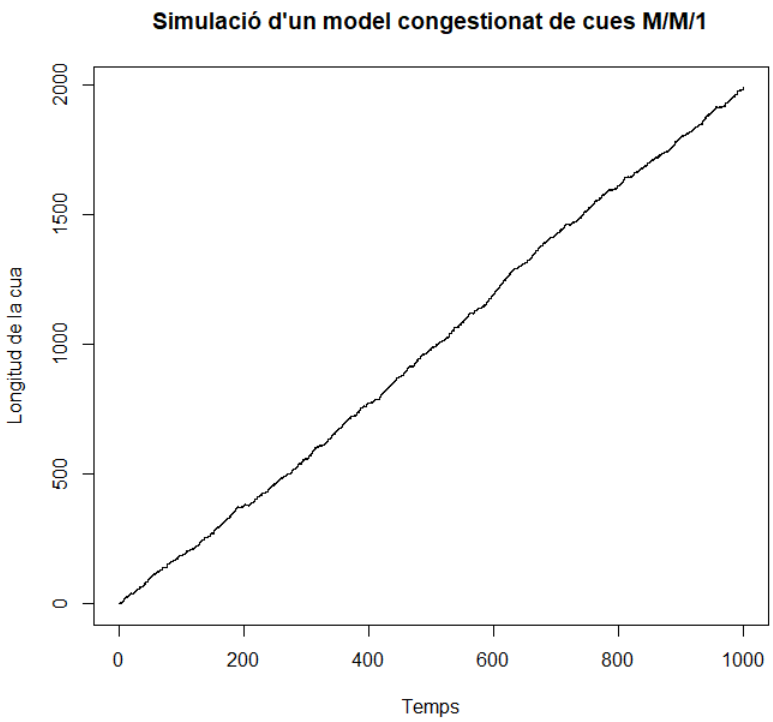
Per tant observem que els resultats són similars als obtinguts a la simulació.

Podríem ara, intentar simular un sistema de cues de manera que el flux d'entrada de clients és superior al flux de sortida. En aquest cas, obtenim que $\rho > 1$, de manera que tal com hem dit en capítols anteriors, el sistema es congestiona.

Aquesta simulació ens servirà per a observar què és el que passa en el moment que

$\lambda > \mu$, ja que tot l'estudi analític que hem fet en els apartats anteriors, hem suposat que el sistema està en equilibri de manera que poguem definir una distribució estacionaria. D'aquesta manera no hem vist cap resultat en el cas en què el sistema es congestioni. Anem a veure-ho:

Utilitzarem com a hipòtesis que el flux d'entrada de clients és $\lambda = 3$ i $\mu = 1$, i anem a veure quins són els resultats.



En aquest cas podem observar que la longitud de la cua del sistema creix de manera constant, per tant el sistema tendeix a congestionar-se i a explotar.

Aquest resultat ens reafirma el que ja pensàvem que ja suposàvem que passaria, el sistema va acumulant clients i no és òptim.

Anem a veure ara L'algoritme que hem utilitzat per tal de modelitzar el sistema.

Començarem donant les variables necessàries per a la simulació:

```
> lambda <- x
> mu <- y
> tempsT<-1000
> s <- 0
> vectest <- 0
> vecttemps <- 0
> estat <- 0
> temps <- 0
> tue <- temps
> ta <- 0
```

```
> ts <- tempsT
> N <- 0
```

En la nostra simulació fixarem els valors de **lambda** i **mu**, que correspondran al flux d'entrada i sortida de clients del servei. El **tempsT** correspon al temps total que durarà la simulació, quan sobrepassem aquest temps la simulació s'aturarà. La variable **s** correspon al sumatori del producte entre un estat del sistema i el temps en què el sistema es troba en aquest estat (ens serà útil per trobar la longitud mitjana de la cua del sistema). Les variables **vectest** i **vecttemps** corresponen a dos vectors on hi guardem la longitud de la cua dins el sistema i els temps en què hi ha un canvi d'estat respectivament (ens serà útil per poder dibuixar la gràfica que ens mostri la longitud de la cua en tot moment). L'**estat** i el **temps** corresponen a dos variables que ens mostren a quin estat ens trobem i a quin temps respectivament. La variable **tue** representa el temps que passa des de l'últim esdeveniment. Les variables **ta** i **ts** representen en temps que falta per la pròxima arribada i el pròxim servei. Finalment, la variable **N** ens donarà el nombre mitjà de clients a la cua del sistema modelitzat.

Seguidament podem observar l'algoritme utilitzat per tal de simular un model de cues $M/M/1$ amb flux d'entrada λ i de sortida μ.

```
while(temps <tempsT){
    if(ta < ts){
        temps <- ta
        s <- s + estat*(temps - tue)
        vectest <- append(vectest, estat)
        estat <- estat + 1
        vecttemps<- append(vecttemps, temps)
        tue <- temps
        ta <- temps + rexp(1, lambda)
        if(estat==1){
            ts <- temps + rexp(1,mu)
        }
    }else{
        temps <- ts
        s <- s + estat*(temps-tue)
        vectest <- append(vectest, estat)
        estat <- estat -1
        vecttemps <- append(vecttemps, temps)
        tue <- temps
        if(estat > 0){
            ts <- temps + rexp(1,mu)
        }else{
            ts<-tempsT
        }
    }
}
> N <- s/temps
```

Referències

[1] Gross, D., Shortle, J. F., Thompson, J. M., and Harris, C. M. *Fundamentals of Queueing Theory*, 4th ed. Wiley-Interscience, New York, NY, USA, 2008.

[2] Gunther, N. Simulating a queue in r. *The Pith of Performance* (Maig 2010).

[3] Sabater, J. P. G. Aplicando teoría de colas en dirección de operaciones, 2015-2016.

[4] Sztrik, J. *Basic Queueing Theory: Foundations of System Performance Modeling.* GlobeEdit, 2016.

[5] Virtamo, J. Queueing theory. http://www.netlab.tkk.fi/opetus/s383143/kalvot/english.shtml.

9 Annex

Codi en C per trobar el nombre mig de clients de una cua respecte el flux d'entrada de clients del sistema. Per tal de definir el sistema de cues el programa et demana el nombre de cues en paral·lel que té el sistema, el nombre de servidors que gestiona cada cua, i quin és el flux de sortida de cada servidor del sistema.

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#define DIM 20

int main(void){
        int i, n, m, fac1=1, fac2 = 1;
        double p, L, sum=1, lambda=0, mu;
        char sor[DIM];
        FILE *fp;

        printf("Quin  s  el fitxer de sortida?\n");
        scanf("%s", sor);
        fp = fopen(sor, "w");

        printf("Entra el numero de cues en paral lel i el flux de sortida\n")
        scanf("%d", &m);
        scanf("%le", &mu);
        printf("Entra el numero de servidors que gestiona cada cua\n");
        scanf("%d",&n);
        p = lambda/(m*mu);
        while(p<n){
                for(i=1; i<=n; i++){
                        fac1 = fac1*i;
                }


                L = (pow(p,n)*n)/(fac1*(n-p));

                for(i=1;i<n;i++){
                        fac2=fac2*i;
                        sum = sum + pow(p,i)/fac2;
                }
                L = L/(sum + L);
                L = L*p/(n-p);
                fprintf(fp, "%le %le\n", lambda, L);

                if(p < 0.5){
                        lambda = lambda + 1e-1;
                }
                if(0.5<=p && p<0.9){
                        lambda = lambda + 1e-2;
                }
```

```
            if(0.9<=p){
                lambda = lambda + 1e-4;
            }
            fac1=1;
            fac2=1;
            sum=1;
            p = lambda/(m*mu);
        }
        fclose(fp);
        return 0;
}
```