Universitat de Barcelona
Facultat de Matemàtiques i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

TEORIA DE CUES

Autor: Gabriel Jaume Martín

Director: Dr. Carles Rovira Escofet
Realitzat a: Departament de Matemàtiques i Informàtica
Barcelona, 24 de gener de 2022

Abstract

Queueing Theory is used in a wide variety of areas, so the computation of optimal configurations of queueing systems is a very interesting problem. We will detail what is a queueing process and the characteristics that are necessary to describe this kind of process. In order to be able to properly study the properties of different models, we will first establish the stochastic processes foundations needed to do it. Then, we will compute these properties, focusing on the performance measures, for the $M/M/1$, $M/M/n$, $M/M/1/K$ and $M/M/n/K$ models. Finally, we will provide the design of a computer program that helps us find the optimal configurations of the systems, cost-wise.

Resum

La Teoria de Cues és utilitzada en una gran diversitat d'àrees, de manera que trobar les configuracions òptimes dels sistemes de cues és un problema interessant. En aquest treball detallarem en què consisteix un procés de cues i les característiques necessàries per poder descriure processos d'aquest tipus. Per tal de poder estudiar correctament les propietats dels models, primer donarem les bases de processos estocàstics necessàries per a poder-ho fer. Posteriorment, calcularem aquestes propietats, centrant-nos en les mesures d'efectivitat, per als models $M/M/1$, $M/M/n$, $M/M/1/K$ i $M/M/n/K$. I per acabar, proporcionarem el disseny d'un programa que ens permeti trobar les configuracions que facin òptim el sistema a nivell de cost.

2020 Mathematics Subject Classification. 60J28, 60K30, 60K25

Agraïments

M'agradaria començar donant les gràcies al meu tutor, el Dr. Carles Rovira Escofet, per la seva implicació en el treball i per l'ajuda que ha aportat durant tot el procés. A continuació, m'agradaria agrair a la meva família pel suport constant que sempre rep i per haver fet possible que hagi pogut arribar fins aquí. Finalment, voldria agrair a tots els meus amics els ànims que m'han donat i haver compartit aquesta experiència amb mi.

Índex

1 Introducció

1.1 Motivació

Les assignatures que s'estudien durant tot el transcurs del Grau de Matemàtiques són de caràcter purament teòric. Rara vegada es veu alguna aplicació directa que podrien tenir els coneixements adquirits a algun cas pràctic, com pugui ser al funcionament d'una empresa. Per aquest motiu, volia utilitzar aquest treball per poder aprofundir un poc més en l'àrea de les matemàtiques aplicades.

A part d'això, el semestre anterior vaig cursar l'optativa de Processos Estocàstics, impartida pel Dr. Carles Rovira, i em va resultar molt interessant. En particular, em vaig sentir atret pels processos de naixement i mort i els models de cues. Així doncs, em vaig decidir per fer el Treball Final de Grau de Matemàtiques sobre aquesta temàtica.

1.2 Objectius

L'objectiu principal d'aquest treball serà donar les condicions per al disseny òptim d'un sistema de cues, basant-nos en la minimització de la funció de cost. Per tal de poder assolir aquesta finalitat, explicarem en què consisteix un procés de cues i quines són les característiques generals que s'utilitzaran per a descriure un procés d'aquest tipus.

A més, donarem les bases teòriques de processos estocàstics necessàries per a poder estudiar un procés de cues matemàticament, amb la finalitat d'estudiar alguns models concrets.

Així, podrem calcular les diferents característiques i mesures de rendiment d'aquests, i usar-les per a satisfer el propòsit principal d'optimitzar el cost.

1.3 Estructura de la Memòria

L'estructura que s'ha donat a aquesta memòria és la que considero més adequada perquè una persona sense cap noció sobre Teoria de Cues sigui capaç d'entendre el treball sense dificultat.

Per aquest motiu, en la **Secció 2: Descripció d'un procés de cues**, es dona una idea general sobre què és un sistema de cues, la notació que utilitzarem al llarg de total la memòria i l'explicació d'en què consisteix l'optimització d'un sistema de cues.

A la **Secció 3: Processos estocàstics en Teoria de cues**, expliquem alguns conceptes generals sobre processos estocàstics i fem un estudi detallat sobre els processos de Markov, arribant al cas particular dels processos de naixement i mort, ja que tots els models estudiats seguiran un procés d'aquest tipus.

La **Secció 4: Models de Teoria de Cues**, és un recull de les propietats que hem considerat més importants de quatre dels models més usats d'aquesta teoria: $M/M/1$, $M/M/n$, $M/M/1/K$ i $M/M/n/K$. Al final d'aquesta secció, podem trobar un resum de totes les fórmules que hem deduït per a les mesures d'efectivitat de cada un dels models.

A continuació tenim la **Secció 5: Disseny òptim d'un sistema de cues**. En aquesta secció, proporcionem la descripció d'un programa que hem dissenyat per al càlcul del mínim de la funció de cost en funció de diferents paràmetres, i exposem diferents exemples d'ús d'aquest programa. També proporcionem una aproximació del càlcul òptim de n per al model $M/M/n$, que ens servirà en el cas que vulguem optimitzar la relació cost/qualitat del servei.

Finalment, ens trobem amb la secció de **Conclusions**, on fem un resum dels resultats obtinguts.

1.4 Referències utilitzades

Al final d'aquesta memòria, podem trobar a l'apartat de **Referències** totes les referències que s'han utilitzat per a l'elaboració d'aquesta memòria.

Cal destacar que els tres llibres que apareixen en aquest apartat: [1], [2] i [3], s'han fet servir al llarg de tota la memòria i en cada una de les seccions. No obstant això, en el cas de la secció 3, la secció 4 i la secció 5 ens ha estat necessari complementar la informació proporcionada per aquests llibres amb altres fonts.

D'aquesta manera, per a la secció 3 hem fet servir [4], hem fet servir [5] per a la secció 4 i hem fet servir [6] i [7] per a la secció 5.

2 Descripció d'un procés de cues

Haver d'esperar en una cua és una cosa típica de la vida quotidiana. Tothom ha experimentat alguna vegada frustració per haver d'estar una quantitat de temps molt gran en una cua, per gaudir d'un servei que, en moltes ocasions, dura una petita fracció del temps que ha estat esperant. Igualment, és molt comú trobar-se en llocs, com pugui ser un supermercat, on hi ha diverses cues, i plantejar-se quina és la cua que ens permetrà sortir abans de l'establiment. Doncs tots aquests problemes, i molts més, són els que intenta estudiar la Teoria de Cues.

De fet, aquesta branca de les matemàtiques va molt més enllà d'estudiar cues tal com les coneixem, formades per clients i servidors. Existeixen molts processos que es comporten d'una manera molt similar, com ara elements que es fabriquen en una cadena de producció, avions esperant per enlairar-se o fins i tot programes d'ordinador esperant a ser executats. Per aquest motiu, poden ser descrits per la mateixa estructura i, en conseqüència, estudiats dins la mateixa teoria. En particular, la Teoria de Cues va nàixer a principis del segle XX amb la publicació del matemàtic danès A. K. Erlang "The Theory of Probabilities and Telephone Conversations"on estudiava els problemes de congestió de tràfic telefònic.

Per tal de poder entendre quins problemes poden ser plantejats a partir d'aquesta teoria, així com poder conèixer l'estructura sobre la qual es fonamenten les propietats matemàtiques que descriurem en apartats posteriors, a continuació descriurem les característiques principals d'un procés de cues.

2.1 Característiques d'un procés de cues

L'estructura general d'un procés de cues consisteix en: un conjunt de clients arriben a un sistema buscant un servei, esperen, si aquest no és immediat, i abandonen el sistema una vegada han estat atesos. En alguns models es poden tenir en compte consideracions extres, com per exemple que els clients abandonin el sistema cansats d'esperar. Cal destacar, com ja hem mencionat abans, que no s'ha d'entendre client necessàriament com a una persona, sinó com qualsevol element d'un sistema que necessiti consumir un servei.

Per poder descriure aquest tipus procés, normalment ens serà suficient amb determinar sis característiques bàsiques: el patró d'arribada dels clients, el patró de servei dels servidors, la disciplina de la cua, la capacitat del sistema, el nombre de canals de servei i el nombre de fases de servei.

És molt important escollir adequadament quines característiques tindrà el nostre model per tal que s'adeqüi a la situació que estem intentant descriure. Com en qualsevol cas en què s'intenti modelar un sistema matemàticament, es requerirà la idealització del sistema que estem estudiant. Així i tot, és convenient no excedir-se en la idealització i dedicar el temps suficient en concretar els paràmetres que utilitzarem en el nostre model.

2.1.1 Patró d'arribada dels clients

Per tal de poder caracteritzar el sistema en general, és necessari trobar les propietats probabilístiques del flux entrant de sol·licituds. El procés d'arribada, que en general serà estocàstic, vindrà descrit per la distribució dels temps d'arribada dels clients.

Addicionalment, cal conèixer les característiques dels clients que entren al sistema:

- Nombre de clients per arribada: És important saber si els clients arribaran individualment o si, per contra, aquests poden aparèixer en grups. A més, en el segon cas, s'hauria de conèixer quina funció de probabilitat descriu la mida del grup.
- Reacció dels clients: Direm que un client és pacient en cas que aquest es mantingui a la cua independentment de la resta de factors. Casos com que el client no entrés al sistema a causa de la grandària de la cua, sortís cansat d'esperar o es canviés a una altra cua que pensés que és més ràpida, són exemples de clients que anomenarem impacients. Ens referirem a aquests casos com patrons d'arribada depenents de l'estat.
- Dependència en el temps del patró d'arribada: En cas que el patró d'arribada sigui independent del temps direm que aquest és estacionari. En cas contrari, direm que el patró és no estacionari.

2.1.2 Patró de servei dels servidors

El patró de servei es refereix a la forma en la qual els serveis són processats. Igual que abans, es requereix una distribució de probabilitat per poder descriure la seqüència que segueixen els temps de servei als clients. De la mateixa manera, els serveis poden ser individuals o en grup, poden ser estacionaris o no estacionaris i també poden dependre de l'estat del sistema.

Per tal que el patró no sigui estacionari cal que depengui del temps. Hi ha molts d'exemples de factors que poden fer que això es doni, com podria ser que el sistema millorés amb el temps, fent que els serveis cada cop siguin més eficients i més ràpids.

Per contra, la dependència amb l'estat del sistema té a veure amb la quantitat de clients que hi ha a la cua. Hi ha sistemes que funcionen més de pressa com més gent hi ha esperant o, contràriament, sistemes que van més lent com més gent hi ha a la cua.

És convenient tenir present la diferència entre dependència en el temps i dependència en l'estat. Seguint amb els exemples d'abans, un sistema que acceleri quan la cua és molt gran serà més eficient quan hi hagi més gent, independentment de quan es doni aquesta condició. Així i tot, pot haver-hi sistemes que siguin no estacionaris i dependents de l'estat alhora.

Observem que la probabilitat de distribució de la longitud de la cua serà el resultat de dos processos separats, arribades i sortides, ja que els clients entren i surten del sistema en intervals irregulars de temps. Normalment, considerarem aquests processos independents un de l'altre.

2.1.3 Disciplina de la cua

La disciplina de la cua fa referència a les regles de selecció de clients que s'utilitzen. Les més usuals són:

- FIFO ("First In First Out"): Com indica el mateix nom, en aquest tipus de sistema el primer a entrar és el primer a sortir.

- FCFS ("First Come First Served"): Seguint aquesta disciplina, el primer a arribar és el primer a ser servit.
- LIFO ("Last Come First Out"): Seguint aquest tipus de disciplina surt primer a l'últim client que ha arribat.
- LCFS ("Last Come First Served"): En aquests sistemes el primer a ser atès és el darrer que ha arribat.
- RS ("Random Service"): En aquest tipus de sistema la selecció de client a ser atès es fa de forma aleatòria, independentment de l'ordre d'arribada. Altres notacions per serveis aleatoris poden ser SIRO o RSS.
- Sistemes amb prioritat: En aquest tipus de sistema existeixen clients que tenen prioritat de servei respecte d'altres i, per tant, se serveix primer al de major nivell de prioritat, independentment del moment en què entri a la cua. Hi ha dos tipus principals de sistemes amb prioritat:
  - HOL ("Head Of Line"): També anomenada prioritat sense dret preferent de compra ("Priority without Preemption"). En aquest cas els clients amb major prioritat es mouen al principi de la cua, però no són atesos fins que hi hagi un servidor lliure, amb independència del nivell de prioritat de la persona que utilitza el servidor.
  - Prioritat amb dret preferent de compra ("Priority with Preemption"): En aquest cas, els clients amb major prioritat obtenen el servei de forma immediata, encara que el servidor estigui sent utilitzat per un client de menor prioritat.
- PS ("Processor Sharing"): Aquest tipus de disciplina és utilitzada normalment en sistemes d'ordinadors amb un cert nombre de terminals, de manera que tots els çlients"reben la mateixa amplitud de banda del servidor, independentment de la longitud de la tasca que necessiti realitzar cada client.

2.1.4 Capacitat del sistema

La capacitat del sistema determina la quantitat total de clients que pot haver-hi. Així doncs, si la capacitat del sistema és infinita la cua podrà tenir un nombre indefinit de clients. Per contra, poden existir limitacions físiques que impossibilitin aquesta opció. Exemples d'això podrien ser la mida del taller d'un mecànic o la quantitat de llocs disponibles en una sala d'espera. En cas que la capacitat sigui finita serà necessari conèixer la quantitat de places disponibles per a la cua i la quantitat de llocs disponibles per a ser servits.

2.1.5 Nombre de canals de servei

Quan parlem de canals de servei ens referim a les estacions que ens permeten servir a diferents clients simultàniament.

Cada cop que un client entri en el sistema, aquest haurà de ser assignat a un dels canals. Existeixen diferents formes d'assignar cada nou client a un dels canals, algunes de les més comunes són:

- Que hi hagi una cua diferent per a cada un dels servidors, de manera que cada client escull lliurement quin dels servidors vol utilitzar. Aquest és molt típic en supermercats.
- Que hi hagi una cua diferent per a cada servidor i es vagi assignant als clients a cada un dels servidors segons un determinat ordre.
- Que hi hagi una única cua per a tots els servidors i es vagi assignant a cada element de la cua als diferents servidors a mesura que aquests vagin quedant lliures. Aquest cas és el més eficient de tots i és molt utilitzat en bancs.

2.1.6 Nombre de fases de servei

Quan es demana un servei aquest pot ser proporcionat en un sol pas o, per contra, poden existir diferents fases per les quals s'han de passar abans que sigui completat. Un exemple de servei d'un sol pas podria ser un supermercat, on, un cop seleccionats els articles que es volen comprar, és suficient amb pagar per a completar el servei. D'altra banda, un exemple de servei en múltiples fases podria ser la fabricació d'un producte en cadena, on un producte necessita passar per diferents estats abans d'estar acabat.

En el cas d'un sistema amb diverses fases, el pas pels diferents estats pot ser lineal o no. Existeixen sistemes en els quals, un cop has passat per un estat i has avançat al pròxim, poden retornar-te a un anterior. Aquest seria el cas d'un producte que, en una cadena de producció, fallés un dels controls de qualitat i fos retornat a una de les fases prèvies per solucionar els problemes.

2.2 Notació

A continuació descriurem la notació utilitzada generalment en la descripció de processos de cues, introduïda pel matemàtic D. G. Kendall el 1953. Seguint aquesta notació, un sistema vindrà determinat especificant les seves característiques de la manera:

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

On:

- A és la funció de distribució dels temps entre arribades.
- B és la funció de distribució dels temps de servei.
- m és el nombre de servidors.
- K és la capacitat del sistema.
- n és el tamany de la població.
- D és la disciplina.

És important destacar que és una pràctica comuna només utilitzar els tres primers símbols A/B/m. Quan això passi se suposarà que tant la capacitat del sistema com el tamany de la població són infinits i la disciplina és la FCFS. Durant aquest treball, quan una d'aquestes característiques no es correspongui amb el que acabem de dir, especificarem només aquella que en difereixi. Per exemple, si tenim un sistema amb capacitat finita l'anomenarem A/B/m/K, i si tenim un amb capacitat infinita però disciplina FIFO A/B/m/FIFO.

La simbologia utilitzada per a les funcions de distribució més comunes és:

- M: Exponencial.
- E_k: Distribució d'Erlang.
- H: Distribució hiperexponencial.
- PH: Distribució de tipus fase.
- D: Determinista.
- G: Distribució arbitrària.

Es pot observar que no hem mencionat cap mena de símbol per referir-nos a característiques com el nombre de clients per arribada, la seva reacció o el nombre de fases que té un sistema. En general suposarem que tenim el cas més senzill (una sola fase, clients pacients i arribant d'un en un, etc.), i en cas contrari serà especificat.

2.3 Optimització de sistemes de cues

L'objectiu principal a l'hora de modelitzar un sistema no és més que entendre millor el sistema, per tal de poder donar una solució òptima a un problema. En aquest treball, el nostre objectiu serà intentar donar una configuració que optimitzi la relació costos/beneficis d'un sistema donat. Això ho farem optimitzant l'esperança de la funció de cost per unitat de temps, tal i com es proposa a [2].

Abans de presentar l'equació de cost a optimitzar, introduïrem la notació que utilitzarem per anomenar els diferents tipus de costs que pot tenir un sistema:

- CS: cost de servei per unitat de temps.
- CWS: cost d'espera en el sistema per client per unitat de temps.
- CI: cost de tenir el servidor buit per unitat de temps.
- CSR: cost del ritme de servei, μ, per unitat de temps.
- CLC: cost de perdua de clients per client per unitat de temps.
- R: benefici per client servit.

També introduïrem la notació que utilitzarem per denotar algunes de les diferents mesures d'efectivitat del sistema, és a dir, diferents variables aleatòries que descriuen el comportament del sistema i els seus valors mitjans:

- n: nombre de servidors.
- $\bar{n}$: nombre mitjà de servidors en ús.

- $\bar{S}$: nombre mitjà de servidors aturats.
- N: nombre de clients en el sistema.
- L: valor esperat de N.
- N_q: nombre de clients que es troben a la cua.
- L_q: valor esperat de N_q
- T: temps total en el sistema.
- W: valor esperat de T.
- T_q: temps d'espera a la cua.
- W_q: valor esperat de T_q
- P_B: probabilitat que un client no pugui entrar en el sistema.

Ara sí, el cost total esperat per unitat de temps serà:

$$E[Cost\ total] = n \cdot CS + L \cdot CWS + \bar{S} \cdot CI + n \cdot CSR + \lambda \cdot P_B \cdot CLC - \lambda \cdot (1 - P_B) \cdot R. \quad (2.1)$$

Per tal de calcular les mesures d'efectivitat i poder estudiar l'equació (2.1), utilitzarem propietats de la teoria de cues clàssica.

3 Processos estocàstics en Teoria de Cues

Notem que no ens és de gran interès l'estudi de sistemes dels que coneixem els temps concrets en què es donaran les entrades i sortides al sistema. En la majoria de casos reals ens trobarem amb factors que impossibilitaran que els temps d'arribada i servei siguin previament coneguts. Per aquest motiu ens centrarem en l'estudi de sistemes en què almenys un dels dos segueixi un procés estocàstic.

Així doncs, quan estudiem un procés de cues ens centrarem en la seva descripció estocàstica i en paràmetres relacionats amb la mesura del rendiment del sistema. Alguns d'aquests paràmetres seran la longitud de la cua, el temps d'espera o la durada del període en què els servidors estan ocupats.

D'aquesta manera començarem descrivint que és un procés estocàstic i anirem avançant fins a arribar al tipus de procés més comú en Teoria de Cues, el procés de naixement i mort.

3.1 Conceptes generals

Els processos estocàstics neixen de la voluntat de descriure processos empírics que estan dominats per lleis probabilístiques. Per tal de poder-los tractar de forma matemàtica els definirem de la següent manera:

Definició 3.1. *Un procés estocàstic amb espai d'estats S és una família $\{X_t, t \in T\}$ de variables aleatòries $X_t : \Omega \to S$ indexades en un conjunt T, que anomenarem espai de paràmetres.*

Notem que, en general, aquests processos intenten descriure l'evolució aleatòria d'un fenomen en el temps. Per tant, podem entendre T com el conjunt de temps en el qual es dona aquesta evolució i X_t com cada un dels estats que es recorren a cada temps t. Notem també que T pot ser discret, que en aquest cas parlarem d'un procés amb paràmetre discret, o pot ser no numerable, normalment considerarem $\mathbb{R}$ o un interval de $\mathbb{R}$, que en aquest cas parlarem d'un procés amb paràmetre continu. A més, suposarem que estem en el marc d'un cert espai de probabilitat $(\Omega, \mathcal{F}, P)$.

Com que un procés estocàstic és una família de variables aleatòries, cada una amb la seva llei de probabilitat, ens interessarà estendre aquest concepte de llei de probabilitat per al procés.

Definició 3.2. *Donat un procés estocàstic $\{X_t, t \in T\}$ definim les seves distribucions conjuntes en dimensió finita com la família de lleis multidimensionals dels vectors aleatoris $(X_{t_1}, ..., X_{t_m})$ per tot $t_1, ..., t_m \in T$ i per tot $m \geq 1$.*

De fet, el Teorema de Kolmogorov ens assegura que podrem definir un procés estocàstic donades només les seves distribucions conjuntes en dimensió finita.

Teorema 3.3. *(Teorema de Kolmogorov) Considerem una família:*

$$\{P_{t_1,...,t_n}; t_1 < t_2 < ... < t_n, n \geq 1, t_i \in T \ \forall i\}, \tag{3.1}$$

on:

1. *$P_{t_1,...,t_n}$ és una probabilitat sobre $\mathbb{R}^n$,*

2. Si $\{t_{i_1} < t_{i_2} < ... < t_{i_m}\} \subset \{t_1 < t_2 < ... < t_n\}$, la distribució de probabilitat $P_{t_{1_1},...,t_{i_m}}$ és la llei marginal de $P_{t_1,...,t_n}$.

Aleshores existeix un procés estocàstic $\{X_t, t \in T\}$, definit en el mateix espai de probabilitat, tal que (3.1) són les seves distribucions conjuntes en dimensió finita. És a dir, el vector $(X_{t_1}, ..., X_{t_m})$ té llei $P_{t_1,...,t_m}$.

Tal com hem definit procés estocàstic, per saber quin element de l'espai d'estats S valdrà cada un dels X_t necessitem conèixer el valor de $t \in T$ i el valor de $\omega \in \Omega$. Això ens permet veure el procés de dues maneres:

- Com una funció de t on, per cada t, tenim una variable aleatòria X_t.
- Com una funció de ω on, per cada ω, tenim un valor concret de X_t per cada t.

Aquesta segona forma d'entendre el procés ens porta a la següent definició.

Definició 3.4. *Fixat $\omega \in \Omega$, la funció $t \to X_t(\omega)$ definida a l'espai de paràmetres T, s'anomena una realització o una trajectòria (sample path) del procés.*

Finalment, plantejarem algunes consideracions sobre el concepte de continuïtat.

Definició 3.5. *Un procés estocàstic a valors reals $\{X_t,\ t \in T\}$ on T és un interval de $\mathbb{R}$, es diu que és continu en probabilitat si, per a tot $\epsilon > 0$ i cada $t \in T$:*

$$\lim_{s \to t} P(|X_t - X_s| > \epsilon) = 0.$$

Definició 3.6. *Fixem $p \geq 1$. Sigui $\{X_t,\ t \in T\}$, on T és un interval de $\mathbb{R}$, un procés estocàstic a valors reals tal que $E(|X_t|^p) < \infty$ per a tot $t \in T$. El procés $\{X_t,\ t \in T\}$ és continu en mitjana d'ordre p si per tot $t \in T$:*

$$\lim_{s \to t} E(|X_t - X_s|^p) = 0.$$

Notem que la continuïtat en mitjana d'ordre p implica la continuïtat en probabilitat, però no ens garanteix que les trajectòries siguin contínues. Per a això utilitzarem el criteri de continuïtat de Kolmogorov.

Criteri de continuïtat de Kolmogorov: Sigui $\{X_t,\ t \in T\}$ un procés estocàstic a valors reals i T un interval finit de $\mathbb{R}$. Suposem que existeixen constants $\alpha > 1$ i $p > 0$ tal que:

$$E(|X_t - X_s|^p) \leq C_T |t - s|^\alpha,$$

per tot $s, t \in T$ on C_t és una constant. Aleshores, existeix una versió del procés $\{X_t,\ t \in T\}$ amb trajectòries contínues.

Definició 3.7. *Un procés estocàstic $\{X_t,\ t \in T\}$ es diu que és una versió d'un altre procés estocàstic $\{Y_t,\ t \in T\}$ si per a tot $t \in T$ se satisfà que:*

$$P(X_t = Y_t) = 1.$$

També podem dir que $\{X_t,\ t \in T\}$ i $\{Y_t,\ t \in T\}$ són equivalents.

3.2 Processos de Markov

Existeixen moltes situacions en les quals, tant el patró d'arribada com el patró de sortida d'un sistema de cues seguiran un procés de Poisson. Més concretament, podem dir que, en aquestes situacions, els processos de cues són processos de naixement i mort. Donat que aquests dos processos mencionats són processos de Markov, en aquesta secció definirem i donarem algunes propietats d'aquesta família de processos.

Per tal d'estudiar els processos de Markov a temps continu, que seran els que ens interessaran per a la Teoria de Cues, primer necessitarem estudiar les anomenades cadenes de Markov a temps discret.

3.2.1 Cadenes de Markov a temps discret

Suposem que observem l'estat d'un sistema en un conjunt discret d'instants de temps $n = 0, 1, 2, ...$ Cada una de les observacions fetes als diferents instants de temps definiran una família de variables aleatòries $\{X_n, n \geq 0\}$ que assumirem que podran prendre valors S_i on $0 \leq i \leq m$ i m determina la quantitat d'estats possibles. Per tant, aquesta família serà un procés estocàstic amb espai d'estats, S, i espai de paràmetres discrets.

Definició 3.8. *Direm que un procés estocàstic $\{X_n, n \geq 0\}$ s'anomena cadena de Markov, si per cada $x_i \in S$:*

$$P(X_n = x_n | X_{n-1} = x_{n-1}, ..., X_0 = x_0) = P(X_n = x_n | X_{n-1} = x_{n-1}). \tag{3.2}$$

donat que el primer membre estigui definit.

Podem intuir de l'equació (3.2) que donat l'estat inicial del sistema, el futur serà independent del passat.

Definició 3.9. *Per a un cert temps n, anomenarem probabilitat de transició de l'estat i a l'estat j a la probabilitat condicionada:*

$$p_{ij}(n) = P(X_n = j | X_{n-1} = i).$$

Definició 3.10. *Direm que una cadena de Markov és homogènia o temporalment homogènia si $p_{ij}(n)$ no depèn del temps n. És a dir, si:*

$$P(X_n = j | X_{n-1} = i) = P(X_{n+m} = j | X_{n+m-1} = i),$$

per $m \geq -(n-1)$. En aquest cas només direm p_{ij}.

Notem que p_{ij} ens dona la probabilitat de passar d'un estat a un altre en el pas d'una sola unitat de temps. La probabilitat de transició en n unitats de temps serà:

$$p_{ij}^{(n)} = P(X_{r+n} = j | X_r = i). \tag{3.3}$$

Per tant: $p_{ij} = p_{ij}^{(1)}$.

Tal com ho hem definit, tenim que: $p_{ij}^{(0)} = 1$ si $i = j$, i $p_{ij}^{(0)} = 0$ si $i \neq j$. Així, l'equació (3.3) queda definida per tot n i podrem denotar el vector de probabilitat inicial, $\pi(0)$, com:

$$\pi_i = P(X_0 = i) \quad i \quad \pi(0) = \{\pi_0, \pi_1, ..., \pi_m\}.$$

Es fàcil veure que, donats p_{ij} i $\pi(0)$, podrem determinar:

$$P(X_0 = x_0, X_1 = x_1, ..., X_n = x_n) = \pi_0 p_{x_0,x_1} p_{x_1,x_2} ... p_{x_{n-1},x_n}.$$

Definició 3.11. *Ara podrem definir la que anomenarem matriu de transició o matriu de probabilitat de transició com $P = (p_{ij})$ $i, j \in S$. Aquesta serà una matriu quadrada no negativa, que satisfarà:*

- $0 \leq p_{ij} \leq 1$.
- $\sum_j p_{ij} = 1$ *per cada* $i \in S$.

Les matrius quadrades que satisfacin aquestes propietats les anomenarem matrius estocàstiques.

Observació. El producte de matrius estocàstiques és una altra matriu estocàstica.

Demostració. Siguin $A = (a_{ij})$ i $B = (b_{ij})$ dues matrius estocàstiques del mateix ordre. Notem que podem estudiar només el cas AB, ja que els resultats per BA seran els mateixos intercanviant a i b.

Els elements de la matriu AB seran de la forma:

$$c_{ij} = \sum_k a_{ik} b_{kj}.$$

Donat que A i B són estocàstiques, se satisfà:

$$0 \leq b_{kj} \leq 1,$$

$$\sum_k a_{ik} = \sum_k b_{ik} = 1 \ \forall k \in S.$$

Aleshores:

$$0 = \sum_k a_{ik} \cdot 0 \leq \sum_k a_{ik} b_{kj} \leq \sum_k a_{ik} = 1.$$

A més:

$$\sum_j c_{ij} = \sum_j \sum_k a_{ik} b_{kj} = \sum_k \left(a_{ik} \sum_j b_{kj} \right) = \sum_k a_{ik} = 1.$$

En conseqüencia queda demostrat que AB és estocàstica. $\square$

Proposició 3.12. *Per cada matriu estocàstica $P = (p_{ij})$, $i, j = 0, 1, ...$ existeix una cadena de Markov homogènia $\{X_n, n \geq 0\}$ amb espai d'estats $S = \{S_i, i \geq 0\}$ i probabilitat de transició en una unitat de temps p_{ij}, $i, j \in S$.*

Proposició 3.13. *Sigui P una matriu estocàstica associada a una cadena de Markov homogènia $\{X_n, n \geq 0\}$ amb espai d'estats $S = \{S_i, i \geq 0\}$ i probabilitat de transició en n unitats de temps $p_{ij}^{(n)}$, $i, j \in S$. Llavors, P^n també serà una matriu estocàstica i $P^n = (p_{ij}^{(n)})$, és a dir, serà la matriu de transició en n unitats de temps de la cadena de Markov associada a P. No obstant això, no tota matriu estocàstica serà la matriu de transició en n unitats de temps d'una cadena de Markov.*

Demostracio. Aplicant l'observació anterior, és senzill veure que, donada P una matriu estocàstica, llavors P^n, la potència n-èssima de P, també ho serà.

A més, per recursió tenim que:

$$p_{ij}^{(2)} = \sum_{k \in S} p_{ik} p_{kj} \ \forall i, j \in S,$$

$$p_{ij}^{(n)} = \sum_{k \in S} p_{ik}^{(n-1)} p_{kj} \ \forall i, j \in S.$$

Que es correspón amb el valor de cada un dels elements de la matriu P^n. □

3.2.2 Cadenes de Markov a temps continu

Molts dels sistemes que intentarem modelar mitjançant Teoria de Cues es comporten de manera tal que estudiar el temps com un conjunt discret no és adient. Per aquest motiu, ens interessarà estudiar les cadenes de Markov a temps continu.

A diferència d'abans, una cadena de Markov a temps continu i espai d'estats $S = \{S_i, i \geq 0\}$ serà un procés estocàstic de la forma: $\{X(t),\ 0 \leq t \leq \infty\}$. A més, assumint homogeneïtat temporal, la funció de probabilitat de transició donada per:

$$p_{ij}^{(t)} = P(X(t+u) = j | X(u) = i),\ t > 0,\ i, j \in S,$$

és independent de $u \geq 0$. Aquesta $p_{ij}(t)$ satisfà, per tot t, que:

$$0 \leq p_{ij}^{(t)} \leq 1, \qquad \sum_j p_{ij}^{(t)} = 1,\ \forall j \in S.$$

Així podrem definir la matriu de probabilitats de transició com:

$$P(t) = (p_{ij}^{(t)}),\ i, j \in S, \qquad P(0) = I,$$

on I denota la matriu identitat.

Denotarem la probabilitat que el sistema es trobi a l'estat j en el temps t com: $\pi_j(t) = P(X(t) = j)$. I denotarem el vector de probabilitat dels estats del sistema a temps t com: $\pi(t) = \{\pi_1(t), \pi_2(t) ...\}$.

És fàcil veure que: $\pi_j(t) = \sum_j p_{ij}(t)\pi_i(0)$. Això ens permetrà obtenir el vector de probabilitat dels estats del sistema a temps t només sabent el vector de probabilitat inicial i la funció de probabilitat de transició: $\pi(t) = \pi(0)P(t)$.

3.2.3 Temps d'estada

Ens interessarà conèixer el temps que necessitarem per canviar d'un estat i a un altre estat. Aquest temps es una variable aleatòria i l'anomenarem temps d'estada, τ_i.

Per les propietats de les cadenes de Markov se satisfà que:

$$P(\tau_i > s + t | X(0) = i) = P(\tau_i > s + t | X(0) = i, \tau_i > s) P(\tau_i > s | X(0) = i),\ t \geq 0. \quad (3.4)$$

Si denotem: $F_i(u) = P(\tau_i > u | X(0) = i),\ u \geq 0$, l'equació (3.4) la podem escriure com: $F_i(t + s) = F_i(t) F_i(s),\ s, t \geq 0$.

F(.) és una funció continua per la dreta. L'única solució de l'equació funcional anterior continua per la dreta és: $F_i(u) = e^{-a_i u},\ u \geq 0, a_i > 0$, on a_i és una constant.

Això ens diu que τ_i és una exponencial de paràmetre a_i. A més, τ_i i τ_j són independents.

En general, tindrem que, per qualsevol $t, s \geq 0$:

$$p_{ij}^{(t+s)} = \sum_k p_{ik}^{(t)} p_{kj}^{(s)},\ \forall i, j, k \in S,$$

que, si escrivim en forma matricial, obtenim l'equació de **Champman-Kolmogorov**:

$$P(t + s) = P(t)P(s).$$

3.2.4 Generador infinitesimal

Donat que $p_{ij}^{(t)}$ és contínua i derivable per la dreta en t, calcularem la seva derivada:

$$q_{ij} = \lim_{h\to 0} \frac{p_{ij}^{(h)} - p_{ij}^{(0)}}{h} = \lim_{h\to 0} \frac{p_{ij}^{(h)}}{h},\ i \neq j,$$

$$q_{ii} = \lim_{h\to 0} \frac{p_{ii}^{(h)} - p_{ii}^{(0)}}{h} = \lim_{h\to 0} \frac{p_{ii}^{(h)} - 1}{h}.$$

Denotarem: $q_{ii} = -q_i$. Aquest resultat ens permetrà escriure, per t propers a 0:

$$p_{ij} = q_{ij}t + o(t),$$

$$p_{ii} = 1 - q_i t + o(t).$$

Observem que per a qualsevol $S_0 \subset S$ finit, se satisfà que:

$$1 - p_{ii}^{(t)} = \sum_{i\neq j} p_{ij}^{(t)} \geq \sum_{j\in S_0} p_{ij}^{(t)}.$$

De manera que dividint per t i aplicant $\lim\limits_{t\to 0}$ obtindrem:

$$q_i = \lim_{t\to 0} \frac{1 - p_{ii}^{(t)}}{t} \geq \lim_{t\to 0} \sum_{j\in S_0} \frac{p_{ij}^{(t)}}{t} = \sum_{j\in S_0} q_{ij}(t).$$

I com això és cert per tot S_0, arribem a que:

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

En particular, direm que un estat $i \in S$ és estable o reglar quan tinguem igualtat.

A partir d'aquí, podrem definir el generador infinitesimal de procés de Markov a temps continu com la matriu:

$$Q := \lim_{t\to 0} \frac{P(t) - I}{t} = \frac{d}{dt}(P(t))_{|t=0^+}.$$

En el cas que $S = \{0, 1, 2, ..., m\}$ sigui un subconjunt finit dels naturals, Q tindrà la forma:

$$Q = \begin{pmatrix} -q_0 & q_{01} & \cdots & q_{0m} \\ q_{10} & -q_1 & \cdots & q_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m0} & q_{m1} & \cdots & -q_m \end{pmatrix}.$$

3.2.5 Equacions de Kolmogorov

Igual que en l'apartat anterior, quan utilitzem la notació de derivada en $t = 0$ ens referirem concretament a la derivada per la dreta.

Teorema 3.14. ***Kolmogorov's backward equation KBE:*** *Suposem que tots els elements de S són regulars. Aleshores, $\forall i, j \in S$ i $t \geq 0$:*

$$p_{ij}'^{(t)} = \sum_{k \neq i} q_{ik} p_{kj}^{(t)} - q_i p_{ij}^{(t)} = \sum_{k \in S} p_{ik}'^{(0^+)} p_{kj}^{(t)}.$$

En notació matricial ho podríem escriure com: $\frac{d}{dt}P(t) = QP(t)$.

Teorema 3.15. ***Kolmogorov's forward equation KFE:*** *Suposem que $i \in S$ és regular i que $\sum_{k \in S} p_{ik}^{(t)} q_k < \infty$. Aleshores, per a $t \geq 0$:*

$$p_{ij}'^{(t)} = \sum_{k \neq i} p_{ik}^{(t)} q_{kj} - p_{ij}^{(t)} q_j = \sum_{k \in S} p_{ik}^{(t)} p_{kj}'^{(0^+)}.$$

En notació matricial ho podríem escriure com: $\frac{d}{dt}P(t) = P(t)Q$.

A més, l'equació matricial: $\frac{d}{dt}P(t) = QP(t)$ té per solució:

$$P(t) = e^{Qt} = \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!} = \sum_{n=0}^{\infty} Q^n \frac{t^n}{n!}.$$

3.2.6 Distribució estacionària

Si considerem el procés només observant els salts, és a dir, estudiant només els canvis d'estat sense tenir en compte el temps que ha passat entre ells, podrem considerar el procés de Markov a temps continu com un procés a temps discret.

Notem que la probabilitat que el procés salti d'un estat i a un estat j, donat que els temps d'estada segueixen una exponencial de paràmetre q_i, serà $p_{ij} = \frac{q_{ij}}{q_i}$. En particular, $p_{ii} = 0$, donat que només estem estudiant els salts.

A la cadena de Markov vista d'aquesta manera l'anomenarem Embedded Markov Chain, o cadena de Markov associada, i la seva matriu de transició la denotarem com $P = p_{ij}$.

Considerem una cadena de Markov associada i suposem que:

$$\lim_{t \to \infty} p_{ij}^{(t)} = \pi_j, \ \forall i.$$

Això significa que després d'un temps suficientment gran, la probabilitat que el procés es trobi en l'estat j és independent de l'estat i inicial. Anomenarem a aquesta $\{\pi_j\}$ la distribució límit de la cadena de Markov. Recordem que, en aquest cas, quan parlem de temps ens referim al nombre de salts que ha realitzat el procés.

Com hem vist en apartats anteriors: $\pi_j(t) = \sum_i \pi_i(0) p_{ij}^{(t)}$. Per tant:

$$\lim_{t \to \infty} \pi_j(t) = \lim_{t \to \infty} \sum_i \pi_i(0) p_{ij}^{(t)} = \sum_i \pi_i(0) \lim_{t \to \infty} p_{ij}^{(t)} = \sum_i \pi_i(0) \pi_j = \pi_j \sum_i \pi_i(0) = \pi_j.$$

D'aquí obtenim que: $\lim_{t\to\infty} p_{ij}^{(t)} = \lim_{t\to\infty} \pi_j(t) = \pi_j$ i que, a més, no depèn del temps.

Denotem per $\pi = (\pi_0, \pi_1, ...)$ a la distribució límit. Notem que, si existeix, se satisfan les dues condicions següents:

$$\pi(t) = \pi(t-1)P, \qquad \lim_{t\to\infty} \pi(t) = \lim_{t\to\infty} \pi(t-1) = \pi.$$

Llavors:

$$\pi = \lim_{t\to\infty} \pi(t) = \lim_{t\to\infty} \pi(t-1)P = \pi P.$$

Aquest resultat, juntament amb la condició de contorn: $\sum_j \pi_j = 1$, són les anomenades **equacions estacionàries** de la cadena de Markov:

$$\left\{ \begin{array}{c} \pi = \pi P \quad o \quad 0 = \pi Q \\ \sum_j \pi_j = 1 \end{array} \right\}. \tag{3.5}$$

És important tenir en compte que aquestes equacions podrien tenir solució en alguns casos en què la distribució límit no existeix. Les equacions estacionàries ens permetran trobar la distribució límit en el cas que existeixi, i en aquest cas l'anomenarem distribució estacionaria, però no implicaran la seva existència.

El següent teorema ens indicarà una forma de saber si existeix distribució límit:

Teorema 3.16. *Sigui un procés de Markov irreduïble, aleshores la distribució límit existeix i és independent de la condició inicial del procés. A més, els valors π_j seran idènticament iguals a 0 o tots positius formant una distribució de probabilitat satisfent la condició de contorn.*

Direm que un procés de Markov és irreduïble si la matriu de transició de la cadena de Markov associada, P, és irreduïble. Intuïtivament això ens diu que es pot arribar des d'un estat inicial qualsevol a qualsevol altre en un nombre finit de salts.

3.3 Processos de naixement i mort

Un cop hem explicat que és una cadena de Markov a temps continu, podem procedir a explicar que són els processos de naixement i mort. Aquests seran el tipus de procés estocàstic que seguiran la majoria de models de teoria de cues. En particular, serà el cas de tots els models $M/M/n$ per $n \geq 1$, independentment de la capacitat del sistema, el tamany de la població i la disciplina escollida.

Un procés de naixement i mort és un cas concret de les cadenes de Markov a temps continu amb espai d'estats $S = \mathbb{N}$, on cada un dels estats denota la quantitat d'individus que hi ha en el sistema. Les transicions d'un estat a un altre només es poden donar a partir de "naixements"o de "morts", de manera que, si ens trobem en un estat n, només ens podrem moure a l'estat $n+1$ o $n-1$ respectivament. Més en concret, el temps que es tardarà a donar-se un canvi d'estat vindrà determinat per una llei exponencial de paràmetres que anomenarem λ_n per al cas dels naixements i μ_n per al cas de les morts. Notem que per cada estat podem tenir paràmetres diferents i que també poden ser diferents el d'increment i el de disminució de la població.

Utilitzant les propietats estudiades en l'apartat anterior, podem definir un procés de naixement i mort assumint que:

- $p_{i,i+1}^{(h)} = \lambda_i h + o(h)$.
- $p_{i,i-1}^{(h)} = \mu_i h + o(h)$.
- $p_{i,i}^{(h)} = 1 - (\lambda_i + \mu_i)h + o(h)$.

De manera que, com:

$$p_{i,i-1}^{(h)} + p_{i,i}^{(h)} + p_{i,i+1}^{(h)} = 1 + o(h),$$

i s'ha de satisfer que:

$$\sum_{j \in S} p_{ij}^{(t)} = 1,$$

ja tenim la propietat que el sistema només es podrà moure al següent estat o al previ.

Ara podrem calcular fàcilment el generador infinitesimal:

$$q_i = \lim_{h \to 0} \frac{p_{ii}^{(h)} - 1}{h} = \lambda_i + \mu_i,$$

$$q_{ij} = \lim_{h \to 0} \frac{p_{ij}^{(h)}}{h} = \begin{Bmatrix} \lambda_i \ si \ i - j = -1 \\ \mu_i \ si \ i - j = 1 \\ 0 \ si \ |i - j| > 1 \end{Bmatrix},$$

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & \cdots & \cdots & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdots & \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

És fàcil comprovar que se satisfan les següents condicions:

$$q_i = \sum_{j \neq i} q_{ij} \quad i \quad \sum_{k \in S} p_{ik}^{(t)} q_k < \infty \ \ \forall i.$$

Així doncs, podrem escriure les equacions KBE i KFE, encara que només escriurem la segona. Llavors, $\forall i \geq 1$:

$$p_{i0}^{\prime(t)} = -\lambda_0 p_{i0}^{(t)} + \mu_1 p_{i1}^{(t)},$$

$$p_{ij}^{\prime(t)} = \lambda_{j-1} p_{i,j-1}^{(t)} - (\lambda_j + \mu_j) p_{ij}^{(t)} + \mu_{j+1} p_{i,j+1}^{(t)} \ \forall j \geq 1.$$

Notem que la solució de la KFE no depèn de i, de manera que podríem reescriure-la substituint $p_{ij}^{(t)}$ per $p_t^{(j)}$ representant la probabilitat que el sistema es trobi en l'estat j en un temps t.

Finalment, comprovarem l'existència de la distribució estacionària i la calcularem. Per tal de comprovar l'existència és suficient amb notar que, donats λ_i i μ_i positius i no nuls, llavors la cadena de Markov és irreduïble. Ara utilitzant el teorema (3.16) ja tenim existència.

Imposant les equacions estacionàries (3.5), tenim les relacions:

$$0 = -\lambda_0 \pi_0 + \mu_1 \pi_1,$$

$$0 = \lambda_{n-1} \pi_{n-1} - (\lambda_n + \mu_n) \pi_n + \mu_{n+1} \pi_{n+1}.$$

Resoldrem aquestes equacions recursivament. La primera relació ens dona:

$$\pi_1 = \frac{\lambda_0}{\mu_1}\pi_0.$$

Per $n = 1$:

$$(\lambda_1 + \mu_1)\pi_1 = \lambda_0\pi_0 + \mu_2\pi_2 \implies \pi_2 = \frac{\lambda_0\lambda_1}{\mu_1\mu_2}.$$

Seguint la recursió és fàcil arribar a que:

$$\pi_n = \pi_0 \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}}.$$

De fet, suposant que això sigui vàlid per n també serà vàlid per $n + 1$:

$$(\lambda_n + \mu_n)\pi_n = \lambda_{n-1}\pi_{n-1} + \mu_{n+1}\pi_{n+1}$$

$$\implies \pi_{n+1} = \frac{\lambda_n}{\mu_{n+1}}\pi_n = \pi_0 \prod_{i=0}^{n} \frac{\lambda_i}{\mu_{i+1}}.$$

Utilitzant ara la condició de contorn, tenim:

$$\pi_0 + \sum_{i=0}^{\infty} \left(\prod_{n=0}^{i} \frac{\lambda_n}{\mu_{n+1}} \right) \pi_0 = 1.$$

Arribant a:

$$\pi_0 = \left(1 + \sum_{i=0}^{\infty} \left(\prod_{n=0}^{i} \frac{\lambda_n}{\mu_{n+1}} \right) \right)^{-1}.$$

Concloent que la distribució límit serà no nul·la sempre que:

$$\sum_{i=0}^{\infty} \left(\prod_{n=0}^{i} \frac{\lambda_n}{\mu_{n+1}} \right) < \infty.$$

3.3.1 Procés de Poisson

Podem dividir els processos de naixement i mort en tres grans grups:

- Procés de naixement pur: En aquest cas els $\mu_i = 0$ i $\lambda_i \geq 0 \ \ \forall i$.
- Procés de mort pur: En aquest cas els $\mu_i \geq 0$ i $\lambda_i = 0 \ \ \forall i$.
- Procés de naixement i mort: En aquest cas els $\mu_i \geq 0$ i $\lambda_i \geq 0 \ \ \forall i$.

Un procés de Poisson consisteix en un procés de naixement i mort on $\mu_i = 0 \ \forall i$ i $\lambda_i = \lambda \ \forall i$, és a dir, un procés de naixement pur de paràmetre constant λ. Aquest és un procés de comptatge i té un gran nombre d'aplicacions a l'hora de modelitzar fenòmens estocàstics. A continuació presentem algunes de les seves propietats:

- Utilitzant les KFE és fàcil veure que:

$$P(X(t+s) = i+k|X(t) = i) = \frac{(\lambda t)^k}{k!}e^{-kt} \quad \forall i, k \in S; \ \forall t, s \geq 0.$$

- Propietat additiva: La suma de n processos de Poisson independents de paràmetres $\lambda_i; 1 \leq i \leq n$, és un procés de Poisson de paràmetre $\lambda = \lambda_1 + ... + \lambda_n$.

- Propietat de descomposició: Un procés de Poisson sempre es podrà escriure com la suma d'un nombre qualsevol de processos de Poisson independents, sempre que la suma dels paràmetres d'aquests processos sigui igual al de l'original.

- El temps entre dues arribades consecutives en un procés de Poisson de paràmetre λ ve determinat per una exponencial de paràmetre λ.

- Propietat de Markov: el temps entre dues arribades és estadísticament independent de l'anterior temps entre arribades i també segueix una distribució exponencial de paràmetre λ.

- Propietat d'aleatorietat: Els esdeveniments en un procés de Poisson són purament aleatoris.

4 Models de Teoria de Cues

Fins ara, hem explicat que és la Teoria de cues, els paràmetres i la notació que es requereix per fer un model, i la base matemàtica necessària per poder-ne estudiar el comportament i les propietats bàsiques. A partir d'ara utilitzarem tota aquesta informació, prèviament donada, per estudiar alguns dels models més comuns.

En aquesta secció, per a cada un dels models, estudiarem les condicions en que existeix distribució estacionaria i la calcularem. En addició, calcularem les mesures d'efectivitat mencionades a l'apartat 2.3, i en alguns casos alguna més, en aquestes condicions.

En el cas concret dels models $M/M/1$ i $M/M/n$, també determinarem les distribucions dels temps d'espera i estudiarem els seus processos de sortida.

4.1 Model M/M/1

Tal com vam explicar a l'apartat de notació, model $M/M/1$ vol dir que tenim un sol servidor, que els temps entre arribades i els temps de servei segueixen una distribució exponencial de paràmetres λ i μ respectivament. Això ens diu que ens trobem en el cas d'un procés de naixement i mort de paràmetres $\lambda_i = \lambda, \forall i$ i $\mu_i = \mu \; i \geq 1$. Cal destacar que, com només s'indiquen tres paràmetres, estem en el cas de sistema amb disciplina FCFS i capacitat i tamany de la població infinits. A més, només estudiarem aquest model en les condicions que tinguem distribució estacionaria, com veurem més endavant $\rho < 1$, i assumirem que totes les variables són independents les unes de les altres.

Utilitzarem $N(t)$ per denotar el **nombre de clients** que es troben en el sistema a temps t, i direm que el sistema es troba en un estat n si hi ha n individus en un moment determinat, $N(t) = n$.

Denotarem la **intensitat de trafic** com:

$$\rho = \frac{\lambda}{\mu},$$

és a dir, la intensitat de tràfic serà la relació de la mitjana del temps de servei, entre la mitjana del temps entre arribades. Aquesta variable ens donarà una idea de com de ràpid es serveix als clients en funció de la freqüència entre arribades. Això ens porta a pensar que la condició necessària perquè el sistema tingui distribució estacionaria és $\rho < 1$, que és exactament el que obtenim a l'aplicar els resultats de processos de naixement i mort:

$$\pi_0 = \left(1 + \sum_{i=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^i\right)^{-1} = 1 - \frac{\lambda}{\mu} \geq 0 \implies \frac{\lambda}{\mu} = \rho < 1$$

També sabem, utilitzant resultats anteriors, que aplicant les equacions estacionaries obtenim:

$$\pi_n = \pi_0 \prod_{i=1}^{n} \frac{\lambda}{\mu} = \pi_0 \rho^n = (1-\rho)\rho^n$$

Per tant, hem vist que la solució per a la distribució estacionaria d'un model $M/M/1$, cas en que $\rho < 1$, és la distribució de probabilitat geomètrica:

$$\pi_n = (1-\rho)\rho^n$$

4.1.1 Mesures d'efectivitat

Un cop plantejat el model, de les primeres coses que ens podem demanar és com és d'efectiu el sistema. Per aquest motiu, estudiarem diferents paràmetres que ens indicaran com es comporta en mitjana.

Per començar, calcularem el valor esperat i la variància de del nombre de clients en el sistema, $L = E[N]$ i $Var(N)$ respectivament. Abans, observem que:

$$\sum_{n=0}^{\infty} \rho^n = \frac{1}{1-\rho},$$

$$\sum_{n=0}^{\infty} n\rho^n = \rho \sum_{n=0}^{\infty} n\rho^{n-1} = \rho \frac{d}{d\rho} \left(\sum_{n=0}^{\infty} \rho^n \right) = \rho \frac{d}{d\rho} \left(\frac{1}{1-\rho} \right) = \frac{\rho}{(1-\rho)^2}.$$

Llavors:

$$L = E[N] = \sum_{n=0}^{\infty} n\pi_n = (1-\rho) \sum_{n=0}^{\infty} n\rho^n = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda}.$$

En addició:

$$\begin{aligned}
Var(N) &= \sum_{n=0}^{\infty} (n-L)^2 \pi_n = (1-\rho) \sum_{n=0}^{\infty} (n^2 + L^2 - 2nL)\rho^n \\
&= (1-\rho) \left(\sum_{n=0}^{\infty} n^2\rho^n + \sum_{n=0}^{\infty} L^2\rho^n - \sum_{n=0}^{\infty} 2nL\rho^n \right) \\
&= (1-\rho) \left(\sum_{n=0}^{\infty} n(n-1)\rho^n + \sum_{n=0}^{\infty} n\rho^n \right) - \frac{\rho^2}{(1-\rho)^2} \\
&= (1-\rho)\rho^2 \frac{d^2}{d\rho^2} \left(\sum_{n=0}^{\infty} \rho^n \right) + \frac{\rho}{1-\rho} - \frac{\rho^2}{(1-\rho)^2} \\
&= 2\frac{\rho^2}{(1-\rho)^2} - \frac{\rho^2}{(1-\rho)^2} + \frac{\rho}{1-\rho} = \frac{\rho}{(1-\rho)^2}.
\end{aligned}$$

Anomenarem $N_q(t)$ al nombre de clients que hi ha fent cua a l'instant t. Notem que podem definir: $N_q(t) = \max\{0, N(t) - 1\}$. Aprofitant els resultats obtinguts per als sumatoris que ens han sortit durant els càlculs de L i $Var(N)$, podrem calcular fàcilment l'esperança i la variància d'aquesta variable:

$$L_q = E[N_q] = \sum_{n=1}^{\infty} (n-1)\pi_n = \sum_{n=1}^{\infty} n\pi_n - \sum_{n=1}^{\infty} \pi_n = L - (1-\rho) \sum_{n=0}^{\infty} \rho^n = L - \frac{(1-\rho)\rho}{1-\rho} = \frac{\rho^2}{1-\rho},$$

$$\begin{aligned}
Var(N_q) &= \sum_{n=1}^{\infty} (n-1)^2 \pi_n - L_q^2 = (1-\rho) \left(\sum_{n=1}^{\infty} n^2\rho^n + \sum_{n=1}^{\infty} \rho^n - 2\sum_{n=1}^{\infty} n\rho^n \right) - \frac{\rho^4}{(1-\rho)^2} \\
&= (1-\rho) \left(\frac{2\rho^2}{(1-\rho)^3} + \frac{\rho}{(1-\rho)} - \frac{\rho}{(1-\rho)^2} \right) - \frac{\rho^4}{(1-\rho)^2} = \frac{\rho^2(1+\rho-\rho^2)}{(1-\rho)^2}.
\end{aligned}$$

També podria ser interessant conèixer el valor esperat de les cues si ignoréssim els casos en què la cua és buida. Això és:

$$L'_q = E[N_q|N_q \neq 0] = \sum_{n=1}^{\infty}(n-1)p'_n = \sum_{n=2}^{\infty}(n-1)p'_n.$$

On p'_n és la probabilitat que hi hagui n individus en el sistema, donat que s'ha de satisfer que $n \geq 2$. Utilitzant la fórmula de Bayes tenim:

$$p'_n = \frac{P(N = n \ i \ N \geq 2)}{P(N \geq 2)}.$$

Observem que se satisfà que:

$$P(N \geq n) = \sum_{k=n}^{\infty}(1-\rho)\rho^k = (1-\rho)\rho^n \sum_{k=n}^{\infty} \rho^{k-n} = \rho^n.$$

I per tant:

$$p'_n = \frac{\pi_n}{\rho^2}$$

$$L'_q = \sum_{n=2}^{\infty}(n-1)\frac{\pi_n}{\rho^2} = \frac{1-\rho}{\rho^2}\left(\sum_{n=2}^{\infty} n\rho^n - \sum_{n=2}^{\infty}\rho^n\right) = \frac{L-\rho}{\rho^2} - 1 = \frac{1}{1-\rho}.$$

Ara, voldrem conèixer el nombre mitjà de servidors en ús, $\bar{n}$, i de servidors aturats, $\bar{S}$. Com només tenim un servidor, $\bar{n}$ serà la probabilitat que hi hagi un o més individus i $\bar{S}$ la de que no n'hi hagi cap:

$$\bar{n} = \sum_{i=1}^{\infty}\pi_i = 1 - \pi_0 = \rho,$$

$$\bar{S} = 1 - \bar{n} = \pi_0 = 1 - \rho.$$

Per acabar, ens interessarà trobar els valors esperats pel temps d'espera, W, i pel temps d'espera a la cua, W_q. Per això, ens serà molt útil les fórmules de Little.

Teorema 4.1. *En condicions d'estat estacionari, sigui T_q el temps que un client tarda esperant a la cua abans de ser atès, sigui T el temps total que està un client en el sistema i sigui S el temps de servei. Sigui $W_q = E[T_q]$ i $W = E[T]$. Llavors, les fórmules de Little són:*

$$L = \lambda W,$$

$$L_q = \lambda W_q.$$

Demostració. Demostrarem només $L = \lambda W$, ja que l'altra fórmula es demostra de forma anàloga.

Suposem que el sistema ha estat operant el temps suficient per a arribar a la condició d'estat estacionari. Considerem un interval de temps T en el que pot no haver-se donat cap període ocupat o poden haver-se'n donat més d'un. Sigui:

- $A(T)$= nombre total d'arribades durant el temps T.

- $B(T)$= temps d'espera total de tots els clients que han arribat durant el període de temps T.

Llavors, definim els valors mitjans durant T de:

- El nombre d'arribades:
$$\lambda(T) = \frac{A(T)}{T}.$$
- El temps d'espera total en el sistema:
$$W(T) = \frac{B(T)}{A(T)}.$$
- El nombre de clients en el sistema:
$$L(T) = \frac{B(T)}{T}.$$

Per tant:
$$L(T) = \frac{B(T)}{T} = \frac{B(T)}{A(T)}\frac{A(T)}{T} = W(T)\lambda(T).$$

Suposem que existeixen els límits quan $T \to \infty$ i que venen donats per:
$$\lim_{T\to\infty} \lambda(T) = \lambda, \qquad \lim_{T\to\infty} W(T) = W.$$

Llavors, també existiria el límit de L(T) per T tendint a ∞, L, arribant finalment a:
$$L = \lambda W.$$

□

Notem que la demostració que hem realitzat no depèn de les distribucions dels temps d'arribada i serveis, del nombre de servidors del sistema, ni la disciplina de la cua, per tant podrem utilitzar les fórmules de Little en models que estudiarem posteriorment.

Usant aquestes fórmules tenim:
$$W = \frac{L}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu-\lambda},$$
$$W_q = \frac{L_q}{\lambda} = \frac{\rho}{\mu(1-\rho)} = \frac{\rho}{\mu-\lambda}.$$

4.1.2 Distribució del temps d'espera

Sigui T_q el temps d'espera a la cua, llavors $F_{W_q}(t)$ serà la seva funció de distribució. Cal destacar que en aquest apartat els resultats dependran de la disciplina escollida, que en el cas d'aquest model en particular serà la FCFS, com havíem comentat abans.

Observem que $F_{W_q}(t)$ té probabilitat no nul·la que el temps d'espera sigui zero, però per temps majors que zero es comporta com una variable aleatòria a temps continu. Sigui q_n la probabilitat que un individu es trobi n clients en el sistema, tenim:
$$F_{W_q}(0) = P(T_q \leq 0) = P(sistema\ buit\ a\ l'arribada) = q_0 = \pi_0.$$

Notem que, en general, no serà cert que $\pi_n = q_n$, però, com ens trobem en el cas en què les arribades segueixen un procés de Poisson, aquesta igualtat es donarà en aquest model.

Només ens falta trobar la probabilitat que un client hagi d'esperar un temps menor o igual que t, $F_{W_q}(t)$. Suposant que hi hagi n clients en el moment de l'arribada, el temps que tardarà el nou client en ser servit serà el temps que es tardarà a servir els n clients que hi havia en el sistema prèviament a la seva arribada. Donat que el temps de servei és exponencial, i per tant tenim la propietat de la falta de memòria, la distribució del temps necessari per fer els n serveis serà independent del temps d'arribada del nou client i serà la convolució de n variables aleatòries exponencials. Això és conegut com una distribució d'Erlang de tipus n i la seva funció de densitat de probabilitat és:

$$f(t) = \frac{\mu^k}{(k-1)!} t^{k-1} e^{-\mu t}, \quad (0 < t \leq \infty).$$

Notem que la distribució d'Erlang de tipus n no és més que un cas concret d'una distribució Gamma. Per tant, totes les propietats que siguin vàlides per la distribució Gamma també ho seràn per la d'Erlang de tipus n.

Procedim a calcular $F_{W_q}(t)$:

$$\begin{aligned}
F_{W_q}(t) = P(T_q \leq t) = F_{W_q}(0) + \sum_{n=1}^{\infty} P(&n \; serveis \; en \; temps \leq t | n \; clients \; a \; l'arribada)\pi_n \\
&= 1 - \rho + (1-\rho)\sum_{n=1}^{\infty} \rho^n \int_0^t \frac{\mu(\mu x)^{n-1}}{(n-1)!} e^{-\mu x} dx \\
&= 1 - \rho + \rho \int_0^t \mu(1-\rho) e^{-\mu x} \sum_{n=1}^{\infty} \frac{(\mu x \rho)^{n-1}}{(n-1)!} dx \\
&= 1 - \rho + \rho \int_0^t \mu(1-\rho) e^{-\mu x(1-\rho)} dx \\
&= 1 - \rho + \rho(1 - e^{-\mu(1-\rho)t}).
\end{aligned}$$

De manera que arribem a:

$$F_{W_q}(t) = 1 - \rho e^{-(\mu-\lambda)t}.$$

Utilitzant aquest resultat, podem comprovar el valor esperat trobat anteriorment:

$$W_q = \int_0^{\infty} [1 - F_{W_q}(t)] dt = \int_0^{\infty} \rho e^{-(\mu-\lambda)t} dt = \frac{\rho}{\mu - \lambda}.$$

Sigui T el temps total en el sistema, el càlcul de la funció de distribució del temps total en el sistema, $F_{W(t)}$, es farà de la mateixa manera, llevat que en un temps t s'hauran de servir $n+1$ clients en lloc de n. Els resultats és els següents:

$$F_W(t) = 1 - e^{-(\mu-\lambda)t}, \quad (t \geq 0).$$

Notem que el temps total dins el sistema, T, segueix una funció de distribució exponencial de paràmetre $(\mu - \lambda)$, així:

$$W = \frac{1}{(\mu - \lambda)},$$

$$Var(T) = \frac{1}{(\mu - \lambda)^2}.$$

Sigui S la variable aleatòria temps d'un servei, de manera que és una exponencial de paràmetre μ. A més, sabem que $T = T_q + S$, per tant:

$$Var(T) = Var(T_q) + Var(S).$$

Aleshores:

$$Var(T_q) = \frac{1}{(\mu - \lambda)^2} - \frac{1}{\mu^2} = \frac{\rho(2-\rho)}{(\mu - \lambda)^2}.$$

4.1.3 Procés de sortida

Per acabar amb el model M/M/1 comentarem la peculiaritat que, en estat estacionari, el procés de sortida, és a dir, el procés que descriu com els clients surten del sistema, segueix una distribució de Poisson del mateix paràmetre que el procés de Poisson de les arribades. Aquest fenomen va ser observat per primera vegada per Morse l'any 1955 i la primera demostració va ser proporcionada per Burke el 1956.

Teorema 4.2. *Sigui un sistema $M/M/1$ en estat estacionari, els temps entre sortides són variables aleatòries exponencials independentment i idènticament distribuïdes amb mitjana $\frac{1}{\lambda}$, on λ és el paràmetre del procés de Poisson de les arribades.*

Demostració. Sigui $N(t)$ el nombre de clients en el sistema a temps t, i siguin $T'_1, T'_2, ..., T'_n, ...$ els successius instants de sortida, de manera que $R = T'_{n+1} - T'_n$ és el n-èssim interval de temps entre sortides. Sigui la distribució de probabilitat conjunta de $N(t)$ i $T'_{n+1} - T'_n$, definida com:

$$F_k(t) = P\left(N(T'_n + t) = k,\ T'_{n+1} - T'_n > t\right),\ t > 0,\ k = 0, 1, 2...$$

Com els procesos d'arribada i servei són de Poisson, la probabilitat que hi hagi k persones en el sistema abans d'una arribada és la mateixa que la probabilitat que es quedin k persones en el sistema després d'un servei. Així:

$$F_k(0) = \pi_k = (1-\rho)\rho^k,\ k = 0, 1, 2...$$

Per un interval de temps infinitesimal de longitud dt se satisfà:

$$F_0(t + dt) = F_0(t)(1 - \lambda dt) + o(dt),$$

$$F_k(t + dt) = F_k(t)[1 - \lambda dt - \mu dt] + F_{k-1}(t)\lambda dt + o(dt).$$

En conseqüència tenim:

$$F'_0(t) = -\lambda F_0(t),$$

$$F'_k(t) = \lambda F_{k-1}(t) - (\lambda + \mu)F_k(t), \quad k = 1, 2, 3, ...$$

Imposant la condició inicial $F_k(0) = \pi_k$, passem a resoldre el següent sistema d'equacions diferencials per inducció. Per $k = 0$, està clar que:

$$F_0(t) = \pi_0 e^{-\lambda t}.$$

Llavors, suposem que $F_k(t) = \pi_k e^{-\lambda t}$ i passem a veure que val $F_{k+1}(t)$:

$$F'_{k+1}(t) = \lambda \pi_k e^{-\lambda t} - (\lambda + \mu) F_{k+1}(t).$$

Resolent l'equació homogènia tenim:

$$F_{k+1}(t) = C(t) e^{-(\mu+\lambda)t}.$$

Derivant i substituint a l'equació anterior:

$$C'(t) e^{-(\mu+\lambda)t} = \lambda \pi_k e^{-\lambda t} \iff C'(t) = \lambda \pi_k e^{\mu t}.$$

Finalment, trobem que $C(t) = \rho \pi_k e^{\mu t}$, de manera que:

$$F_{k+1}(t) = \rho \pi_k e^{-\lambda t} = \pi_{k+1} e^{-\lambda t}.$$

I així, hem obtingut l'única solució del sistema:

$$F_k(t) = \pi_k e^{-\lambda t} = (1-\rho)\rho^k e^{-\lambda t}.$$

Observem que:

$$\begin{aligned} P(N(T'_{n+1}+0) = k,\ t \leq T'_{n+1} - T'_n < t + dt) &= F_{k+1}(t) P(\textit{servei acabat en } (t, t+dt)) \\ &= F_{k+1}(t)[\mu dt + o(dt)] \\ &= (1-\rho)\rho^{k+1} e^{-\lambda t} \mu dt + o(dt) \\ &= (1-\rho)\rho^k \lambda e^{-\lambda t} dt + o(dt). \end{aligned}$$

I, per tant, queda demostrada la independència de $N(T'_{n+1}+0)$ i $(T'_{n+1} - T'_n)$. D'aquí es dedueix que $R = T'_{n+1} - T'_n$ segueix una exponencial de paràmetre λ.

Per acabar la demostració, hem de veure la independència entre els temps entre sortides. Sigui Λ el conjunt de longituds d'un nombre arbitrari d'intervals entre sortides posteriors a l'interval de longitud R. Sigui $P(.)$ la funció de probabilitat de la variable a dins el (). La propietat de Markov ens dona que: $P(\Lambda|N(R)) = P(\Lambda|N(R), R)$. I com hem vist que $N(R)$ i R són independents:

$$P(N(R), R) = P(N(R))P(R).$$

Utilitzant les dues expressions anteriors i la fórmula de Bayes tenim:

$$\begin{aligned} P(R, N(R), \Lambda) &= P(\Lambda|N(R))P(N(R), R) = P(\Lambda|N(R))P(N(R))P(R) \\ &= P(\Lambda, N(R))P(R). \end{aligned}$$

I com:

$$P(R, \Lambda) = \sum_{N(R)=0}^{\infty} P(\Lambda, N(R))P(R) = P(R)P(\Lambda).$$

Ens queda demostrada la independència mútua entre tots els intervals. $\square$

4.2 Model M/M/n

El model $M/M/n$ consisteix en un sistema de n servidors, operant independentment entre ells, on els temps entre arribades i els temps de servei segueixen una llei exponencial de paràmetres λ_i i μ_i respectivament. Recordem que, en addició, suposarem que tenim disciplina $FCFS$ i capacitat i tamany de població infinits. És fàcil veure doncs, que ens tornem a trobar en el cas d'un procés de naixement i mort, però aquesta vegada $\lambda_i = \lambda$ i, sigui μ el paràmetre del procés de Poisson que segueix cada un dels servidors de manera individual, tindrem que:

$$\mu_i = \begin{Bmatrix} i\mu \ (1 \leq i \leq n) \\ n\mu \ (i \geq n) \end{Bmatrix}.$$

Notem que aquest model sorgeix com una modificació natural del model anterior, com a mesura per evitar sortir de l'estat estacionari. Si ens posem en la situació en què $\lambda > \mu$, és evident que augmentar el nombre de servidors modificarà la condició d'estacionarietat per al nou procés, i que, si augmentem n suficientment, arribarem a una situació en què el nou sistema es trobi en estat estacionari. Passem doncs a obtenir les probabilitats π_k de l'estat estacionari, així com la condició d'estabilitat. Per això utilitzarem els resultats obtinguts a l'apartat de processos de naixement i mort:

$$\pi_0 = \left(1 + \sum_{i=0}^{\infty}\left(\prod_{j=0}^{i}\frac{\lambda_j}{\mu_{j+1}}\right)\right)^{-1} = \left(1 + \sum_{i=1}^{n-1}\frac{\lambda^i}{i!\mu^i} + \sum_{i=n}^{\infty}\frac{\lambda^i}{n^{i-n}n!\mu^i}\right)^{-1}.$$

Definint: $a = \dfrac{\lambda}{n\mu} = \dfrac{\rho}{n}$, notem que la condició d'estabilitat serà $\dfrac{\rho}{n} < 1$, i:

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

També tenim que:

$$\pi_k = \pi_0\prod_{i=0}^{k-1}\frac{\lambda_i}{\mu_{i+1}} = \pi_0\prod_{i=0}^{n-1}\frac{\lambda}{\mu(i+1)}\prod_{i=n}^{k-1}\frac{\lambda}{\mu n}.$$

Que ho reescriurem per claredat com:

$$\pi_k = \begin{Bmatrix} \pi_0\dfrac{\rho^k}{k!} \ (k \leq n) \\ \pi_0\dfrac{a^k n^n}{n!} \ (k > n) \end{Bmatrix}.$$

Recordant que el procés d'arribades és de Poisson i la propietat d'aleatorietat d'aquest procés, la distribució del sistema en el moment de les arribades serà la mateixa que la distribució del sistema en un instant aleatori. Així, la probabilitat que un client que arribi hagi d'esperar és:

$$P(esperar) = \sum_{k=n}^{\infty}\pi_k = \sum_{k=n}^{\infty}\pi_0\frac{\rho^k}{n!n^{k-n}} = \frac{\dfrac{\rho^n}{(1-a)n!}}{\displaystyle\sum_{i=0}^{n-1}\frac{\rho^i}{i!} + \frac{\rho^n}{n!}\frac{1}{1-a}}.$$

Aquesta probabilitat apareixerà freqüentment en els següents apartats i és coneguda com la fórmula C d'Erlang o fórmula de retard d'Erlang. La denotarem com:

$$C(n, \frac{\lambda}{\mu}) = C(n, \rho) = \frac{\dfrac{n\rho^n}{(n-\rho)n!}}{\displaystyle\sum_{i=0}^{n-1} \frac{\rho^i}{i!} + \frac{n\rho^n}{(n-\rho)n!}}.$$

4.2.1 Mesures d'efectivitat

Seguint amb la mateixa dinàmica que amb el model $M/M/1$, a continuació calcularem diferents paràmetres que ens aportaran informació sobre l'efectivitat del sistema i el seu comportament en mitjana.

Començarem calculant la mida esperada de la cua, L_q, donat que és més fàcil de calcular que L, al només haver de tenir en compte $\pi_i, \ i \geq n$.

$$\begin{aligned} L_q &= \sum_{i=n+1}^{\infty} (i-n)\pi_i = \sum_{j=0}^{\infty} j\pi_{n+j} = \sum_{j=0}^{\infty} \frac{j\rho^{n+j}}{n^j n!}\pi_0 = \pi_0 \frac{\rho^n}{n!} a \sum_{j=0}^{\infty} j(a^{j-1}) \\ &= \pi_0 \frac{\rho^n}{n!} a \frac{d}{da} \left(\sum_{j=0}^{\infty} a^j \right) = \pi_0 \frac{\rho^n}{n!} \frac{a}{(1-a)^2} = \frac{\rho}{n-\rho} C(n, \rho). \end{aligned}$$

El nombre mitjà de servidors en ús és:

$$\begin{aligned} \bar{n} &= \sum_{i=0}^{n-1} i\pi_i + \sum_{i=n}^{\infty} n\pi_i = \pi_0 \rho \sum_{i=0}^{n-2} \frac{\rho^i}{i!} + \pi_0 \frac{\rho^n}{(1-a)(n-1)!} \\ &= \pi_0 \rho \sum_{i=0}^{n-2} \frac{\rho^i}{i!} + \pi_0 \rho \left(\frac{\rho^n}{(n-1)!} + \frac{\rho^n}{(n-1)!} \left(\frac{a}{1-a} \right) \right) \\ &= \pi_0 \rho \sum_{i=0}^{n-1} \frac{\rho^i}{i!} + \pi_0 \rho \frac{\rho^n}{(1-a)n!} = \frac{\pi_0 \rho}{\pi_0} = \rho. \end{aligned}$$

El nombre mitjà de persones en el sistema és:

$$L = \sum_{i=0}^{\infty} i\pi_i = \sum_{i=0}^{n-1} i\pi_i + \sum_{i=n}^{\infty} (i-n)\pi_i + \sum_{i=n}^{\infty} n\pi_i = \bar{n} + L_q = \rho + \frac{\rho}{n-\rho} C(n, \rho).$$

Sigui $\bar{S}$ el nombre mitjà de servidors aturats, és fàcil adonar-se que:

$$\bar{S} = n - \bar{n} = n - \rho.$$

De manera que:

$$L = n - \bar{S} + L_q \implies L - n = L_q - \bar{S}.$$

Per acabar, utilitzant les fórmules de Little, calcularem els valors esperats de W i W_q, el temps d'espera en el sistema i el temps d'espera a la cua respectivament:

$$W_q = \frac{L_q}{\lambda} = \frac{1}{\mu(n-\rho)}C(n,\rho),$$
$$W = \frac{L}{\lambda} = \frac{1}{\mu} + \frac{1}{\mu(n-\rho)}C(n,\rho).$$

Notem que només amb les fórmules de Little i el valor de L_q haguéssim pogut obtenir W, W_q, L i $\bar{n}$, ja que $W_q = \frac{L_q}{\lambda}$, $W = W_q + \frac{1}{\mu}$, $L = \lambda W$ i $\bar{n} = L - L_q$.

Podem observar també que, sigui $F_{W_q}(0)$ la probabilitat que un client no hagi d'esperar a ser atès, o, equivalentment, la probabilitat que hi hagi menys de n persones en el sistema:

$$F_{W_q}(0) = \sum_{i=0}^{n-1} \pi_i = \pi_0 \sum_{i=0}^{n-1} \frac{\rho^i}{i!}.$$

De l'expressió de π_0 podem deduir que:

$$\sum_{i=0}^{n-1} \frac{\rho^i}{i!} = \frac{1}{\pi_0} - \frac{n\rho^n}{n!(n-\rho)}.$$

I com era d'esperar:

$$F_{W_q}(0) = 1 - C(n,\rho).$$

4.2.2 Distribució del temps d'espera

Per trobar les distribucions dels temps d'espera F_W i F_{W_q} per al model $M/M/n$ farem servir arguments molt similars als utilitzats per al model $M/M/1$. Perquè un client hagi d'esperar a la cua quan arriba al sistema, el nombre mínim de clients que hi ha d'haver en aquell moment és n. En aquest cas, el temps de servei d'un client ve donat per una distribució exponencial de paràmetre $n\mu$, de manera que si tenim $n+j$ clients en el sistema, el temps d'espera a la cua vindrà caracteritzat per una distribució d'Erlang amb paràmetres $(j+1,\ n\mu)$, que com ja hem comentat abans es tracta d'un cas concret de distribució Gamma. Aplicant el teorema de la probabilitat total tenim:

$$\begin{aligned}
w_q(t) &= \sum_{j=0}^{\infty} \pi_{n+j}(n\mu)^{j+1}\frac{t^j}{j!}e^{-n\mu t} = \sum_{j=0}^{\infty} \pi_0 \frac{\rho^n}{n!}a^j(n\mu)^{j+1}\frac{t^j}{j!}e^{-n\mu t} \\
&= \pi_0 \frac{\rho^n}{n!} n\mu e^{-n\mu t} \sum_{j=0}^{\infty} \frac{(an\mu t)^j}{j!} = \pi_0 \frac{\rho^n}{n!} n\mu e^{-n\mu(1-a)t} \\
&= \pi_0 \frac{1}{1-a}\frac{\rho^n}{n!} n\mu(1-a)e^{-n\mu(1-a)t} \\
&= P(esperar)n\mu(1-a)e^{-n\mu(1-a)t}.
\end{aligned}$$

Recordem que T_q és el temps d'espera a la cua. D'aquesta manera, és fàcil veure que:

$$P(T_q > t) = \int_t^{\infty} w_q(x)dx = P(esperar)e^{-n\mu(1-a)t} = C(n,\rho)e^{-n\mu(1-a)t}.$$

La funció de distribució del temps d'espera a la cua serà:

$$F_{W_q}(t) = 1 - P(T_q > t) = 1 - C(n,\rho)e^{-n\mu(1-a)t}.$$

Calcularem la mitjana per comprovar que es correspon amb l'expressió obtinguda amb les fórmules de Little. Per això, primer notem que T_q és una variable mixta amb una part discreta i una part continua, donat que la probabilitat de $T_q = 0$ és diferent de 0. Com podem veure a [5], l'esperança d'una variable mixta d'aquest tipus vindrà donada per:

$$E[X] = \sum_{x \in E_X} xP(X = x) + \int_{\mathbb{R}} xh_X(x)dx.$$

Donat que l'únic valor amb probabilitat no nul·la és $T_q = 0$, el sumatori serà 0. El mateix ens passarà quan calculem $E[X^2]$. En conseqüència:

$$W_q = \int_0^\infty xw_q(x)dx = P(esperar)\frac{n\mu(1-a)}{(n\mu(1-a))^2} = \frac{C(n,\rho)}{n\mu(1-a)}.$$

També ens interessarà calcular la variància del temps d'espera a la cua. Per això calculem:

$$\begin{aligned}
\int_0^\infty x^2 w_q(x)dx &= P(esperar)n\mu(1-a)\int_0^\infty x^2 e^{-n\mu(1-a)}dx \\
&= P(esperar)n\mu(1-a)\frac{2}{n\mu(1-a)}\int_0^\infty xe^{-n\mu(1-a)}dx \\
&= \frac{2C(n,\rho)}{(n\mu(1-a))^2}.
\end{aligned}$$

Aleshores:

$$Var(T_q) = \int_0^\infty x^2 w_q(x)dx - W_q^2 = \frac{C(n,\rho)(2 - C(n,\rho))}{(n\mu(1-a))^2}.$$

Passem ara a trobar la funció de densitat de probabilitat de W. Sabem que si un client arriba i hi ha menys de n individus en el sistema, llavors només haurà d'esperar el temps que el servidor tardi a acabar el servei, d'altra banda, si hi ha més de n clients en el sistema, haurà d'esperar el temps a què li toqui el seu torn més el temps a ser servit. Això ho podem escriure com:

$$w(t) = P(no\ esperar)\mu e^{-\mu t} + f_{W_q+S}(t).$$

Utilitzant $w_q(t)$ donat prèviament:

$$\begin{aligned}
f_{W_q+S}(z) &= \int_0^z w_q(t)\mu e^{-\mu(z-t)}dt \\
&= P(esperar)n\mu(1-a)\mu\int_0^z e^{-n\mu(1-a)t}e^{-\mu(z-t)}dt \\
&= P(esperar)n\mu(1-a)\mu e^{-z\mu}\int_0^z e^{-\mu(n-\rho-1)t}dt \\
&= P(esperar)n\mu(1-a)\frac{1}{n-\rho-1}e^{-z\mu}(1 - e^{-\mu(n-\rho-1)z}).
\end{aligned}$$

Substituint:

$$w(t) = P(no\ esperar)\mu e^{-\mu t} + P(esperar)\frac{n\mu(1-a)}{n-\rho-1}e^{-t\mu}(1-e^{-\mu(n-\rho-1)t}).$$

Igual que abans ens interessarà calcular:

$$\begin{aligned}
P(T>t) &= \int_t^\infty w(x)dx \\
&= P(no\ esperar)\mu\int_t^\infty e^{-\mu x}dx \\
&\qquad + P(esperar)\frac{n\mu(1-a)}{n-\rho-1}\int_t^\infty e^{-x\mu}(1-e^{-\mu(n-\rho-1)x})dx \\
&= P(no\ esperar)e^{-\mu t} + P(esperar)\frac{1}{n-\rho-1}((n-\rho)e^{-\mu t}+e^{-\mu(n-\rho)t}) \\
&= e^{-\mu t}\left[P(no\ esperar)+P(esperar)\frac{1}{n-\rho-1}((n-\rho)+e^{\mu(n-\rho-1)t})\right].
\end{aligned}$$

Per tant, arribem a què:

$$F_W(t) = 1 - P(T>t).$$

A més, podrem calcular W:

$$\begin{aligned}
W &= \int_0^\infty tw(t)dt \\
&= P(no\ esperar)\mu\int_0^\infty te^{-\mu t}dt \\
&\qquad + P(esperar)\frac{\mu(n-\rho)}{n-\rho-1}\int_0^\infty te^{-t\mu}(1-e^{-\mu(n-\rho-1)t})dt \\
&= P(no\ esperar)\frac{1}{\mu} + P(esperar)\frac{(n-\rho)^2-1}{\mu(n-\rho)}\frac{1}{n-\rho-1} \\
&= \frac{1-C(n,\rho)}{\mu} + \frac{C(n,\rho)((n-\rho)^2-1)}{\mu(n-\rho)(n-\rho-1)} \\
&= \frac{1}{\mu} + \frac{C(n,\rho)}{\mu(n-\rho)}.
\end{aligned}$$

Que mirant resultats anteriors, ho podem reescriure tal com esperàvem:

$$W = \frac{1}{\mu} + W_q.$$

Finalment, calcularem la variància de T:

$$\begin{aligned}
E[T^2] &= \int_0^\infty t^2 w(t)dt \\
&= P(no\ esperar)\mu\int_0^\infty t^2e^{-\mu t}dt \\
&\qquad + P(esperar)\frac{\mu(n-\rho)}{n-\rho-1}\int_0^\infty t^2e^{-t\mu}(1-e^{-\mu(n-\rho-1)t})dt \\
&= P(no\ esperar)\frac{2}{\mu^2} + P(esperar)\frac{(n-\rho)^3-1}{\mu^2(n-\rho)^2}\frac{2}{n-\rho-1} \\
&= \frac{2}{\mu^2} + \frac{2C(n,\rho)((n-\rho)^2-1)}{\mu^2(n-\rho)^2(n-\rho-1)}.
\end{aligned}$$

En conseqüència, arribem al resultat que volíem:

$$
\begin{aligned}
Var(T) = E[T^2] - W^2 &= \frac{1}{\mu^2} + \frac{2C(n,\rho)}{\mu^2(n-\rho)^2} - \frac{C(n,\rho)^2}{\mu^2(n-\rho)^2} \\
&= \frac{1}{\mu^2} + \frac{C(n,\rho)(2-C(n,\rho))}{\mu^2(n-\rho)^2} = Var(S) + Var(T_q).
\end{aligned}
$$

4.2.3 Procés de sortida

Igual que vam comentar per al model M/M/1, el procés de sortida per al model M/M/n també segueix un procés de Poisson de paràmetre λ, és a dir, un procés de Poisson amb el mateix paràmetre que les arribades. La demostració d'aquesta propietat es farà de manera similar.

Teorema 4.3. *En un sistema de cues M/M/n en estat estacionari on el temps d'arribada segueix una distribució exponencial de paràmetre λ i els temps entre serveis una exponencial de paràmetre μ_i, els temps entre sortides són independents i idènticament distribuïts com una variable aleatòria exponencial de mitjana $\frac{1}{\lambda}$. En altres paraules, en les condicions mencionades, el procés de sortida és de Poisson de paràmetre λ.*

Demostració. Sigui $N(t)$ el nombre de clients en el sistema a temps t, i siguin $T_1', T_2', ...$ els successius instants en què es produeix la sortida d'un client del sistema, de manera que $R = T_{c+1}' - T_c'$ és el c-èssim interval entre sortides. Sigui:

$$F_k(t) = P(N(T_c' + t) = k, R > t), \ t > 0, k \geq 0.$$

Donat que el procés d'arribades és de Poisson:

$$F_k(0) = \pi_k.$$

Per un interval infinitesimal de temps dt:

$$F_0(t + dt) = F_0(t)[1 - \lambda dt] + o(dt),$$

$$F_k(t + dt) = F_k(t)[1 - \lambda dt - j\mu dt] + F_{k-1}\lambda dt + o(dt), \ k \geq 1.$$

On $j = k$ si $0 \leq k \leq n$ i $j = n$ si $k \geq n$. Podem transformar les equacions com:

$$F_0'(t) = -\lambda F_0(t),$$

$$F_k'(t) = \lambda F_{k-1}(t) - (\lambda + j\mu)F_k(t), \ k \geq 1.$$

Donat que la condició inicial és $F(0) = \pi_k$, l'única solució del sistema anterior ve donada per:

$$F_k(t) = \pi_k e^{-\lambda t}.$$

Una altra vegada, voldrem demostrar que $N(T_{c+1}' + 0)$ i $T_{c+1}' - T_c'$ són independents. Per això, voldrem calcular:

$$P(N(T_{c+1}' + 0) = k, \ t \leq T_{c+1}' - T_c' < t + dt) = F_{k+1}(t)P(\mathit{servei\ acabat\ en}\ (t, t + dt)).$$

Per $k+1 \leq n$, tenim $k+1$ canals de servei ocupats i un patró de servei $(k+1)\mu$. Per $k+1 \geq n$ tenim tots els servidors ocupats i un patró de servei de $n\mu$. D'aquesta manera tenim:

$$F_{k+1}(t)(k+1)\mu dt + o(dt),\ \ k+1 \leq n,$$
$$F_{k+1}(t)n\mu dt + o(dt),\ \ k+1 > n.$$

Estudiem cada cas per separat:

Si $k+1 \leq n$:

$$\begin{aligned} F_{k+1}(t)[(k+1)\mu dt] + o(dt) &= \pi_{k+1}e^{-\lambda t}[(k+1)\mu dt] + o(dt) \\ &= \frac{\left(\frac{\lambda}{\mu}\right)^{k+1}}{(k+1)!}\pi_0 e^{-\lambda t}(k+1)\mu dt + o(dt) \\ &= \pi_k \lambda e^{-\lambda t}dt + o(dt). \end{aligned}$$

Si $k+1 > n$:

$$\begin{aligned} F_{k+1}(t)[n\mu dt] + o(dt) &= \pi_{k+1}e^{-\lambda t}[n\mu dt] + o(dt) \\ &= \frac{\left(\frac{\lambda}{\mu}\right)^{k+1}}{n!n^{k+1-n}}\pi_0 e^{-\lambda t}n\mu dt + o(dt) \\ &= \frac{\left(\frac{\lambda}{\mu}\right)^{k}}{n!n^{k-n}}\pi_0 \lambda e^{-\lambda t}dt + o(dt) \\ &= \pi_k \lambda e^{-\lambda t}dt + o(dt). \end{aligned}$$

En conseqüència:

$$P(N(T'_{c+1}+0) = k,\ t \leq T'_{c+1} - T'_c < t + dt) = \pi_k \lambda e^{-\lambda t}dt + o(dt),$$

provant la independència entre $N(T'_{n+1}+0)$ i $T'_{n+1} - T'_n$ i se'n dedueix que $R = T'_{n+1} - T'_n$ segueix una distribució exponencial de paràmetre λ.

La demostració de la independència dels intervals entre sortides és exactament la mateixa que vàrem fer per al model M/M/1, i per aquest motiu no la repetirem. $\square$

4.3 Model M/M/1/K

Fins ara hem estat estudiant models amb capacitat infinita, però, en moltes situacions en què vulguem analitzar un sistema mitjançant Teoria de Cues, ens trobarem amb limitacions físiques que impedeixen que el sistema pugui créixer tant com volguem. Per aquest motiu, a continuació presentarem les propietats del model $M/M/1/K$, on K representa la capacitat del sistema, és a dir, el màxim nombre de clients en el sistema incloent els que estan sent servits.

Ens tornem a trobar en el cas d'un procés de naixement i mort. En aquest cas tenim que: $\lambda_i = \lambda$ $, 0 \leq i < K$, mentre que en el cas $i = K$, $\lambda = 0$. Igual que en el cas de capacitat infinita, $\mu_i = \mu$ per tot $1 \leq i \leq K$. Aplicant les equacions estacionàries obtenim:

$$\pi_0 = \left(1 + \sum_{i=1}^{K}\left(\frac{\lambda}{\mu}\right)^i\right)^{-1} = \left\{\begin{array}{l} \frac{1}{K+1},\ \rho = 1 \\ \frac{1-\rho}{1-\rho^{K+1}},\ \rho \neq 1 \end{array}\right\}.$$

I, a més, recordem que:

$$\pi_i = \pi_0 \rho^i.$$

Es pot veure que aquest model és estable sempre que $\rho > 0$, donat que K és finit. No obstant això, quan $K \to \infty$ la condició d'estabilitat tornarà a ser $\rho < 1$, ja que aquest model convergeix al model $M/M/1$. Una forma de veure aquesta convergència és notar que si $\rho < 1$ i $K \to \infty$, llavors, $\rho^K \to 0$ i, en conseqüència, $\pi_0 \to 1 - \rho$, que és el π_0 propi del model $M/M/1$.

4.3.1 Mesures d'efectivitat

Ara passarem a calcular com es veuen modificats els comportaments en mitjana del model $M/M/1$ pel fet que el sistema tingui capacitat finita. El primer que calcularem és el valor del nombre mitjà de clients en el sistema:

Cas $\rho = 1$:

$$L = \sum_{i=0}^{K} i\pi_i = \sum_{i=0}^{K} \frac{i}{K+1} = \frac{K}{2}$$

Cas $\rho \neq 1$:

$$L = \sum_{i=0}^{K} i\pi_i = \sum_{i=0}^{K} \frac{i\rho^i(1-\rho)}{1-\rho^{K+1}} = \frac{(1-\rho)\rho}{1-\rho^{K+1}} \sum_{i=0}^{K} i\rho^{i-1} = \frac{\rho(1-(K+1)\rho^K + K\rho^{K+1})}{(1-\rho^{K+1})(1-\rho)}$$

Notem que la probabilitat que el servidor estigui en ús és: $U_S = 1 - \pi_0$, ja que és la probabilitat que hi hagi alguna persona en el sistema. En particular, el valor esperat de servidors en ús serà $\bar{n} = 1 - \pi_0$ i, conseqüentment, el valor esperat servidors aturats serà $\bar{S} = \pi_0$.

El valor esperat de la longitud de la cua és:

$$L_q = \sum_{i=1}^{K} (k-1)\pi_k = \sum_{i=1}^{K} i\pi_i - \sum_{i=0}^{K} \pi_i = L - U_S$$

En el cas de capacitat finita, el procés d'arribades no és el mateix que el procés d'entrades en el sistema. Podria donar-se el cas que una persona arribés al sistema però no pogués entrar degut a que ja hi ha el nombre màxim de persones. En altres paraules, tenim un procés d'arribades d'individus buscant un servei i un procés d'individus que aconsegueixen entrar en el sistema per a ser servits. En concret, denotarem a_n a la probabilitat que, en un interval de temps infinitesimal $(t, t+h)$ en el que hi hagi n clients, una arribada suposi una entrada al sistema . Això ho farem aplicant el teorema de Bayes:

$$\begin{aligned}
a_n &= P(n\ el\ sistema|arribada\ a\ punt\ d'ocórrer) \\
&= \frac{P(arribada\ a\ punt\ d'ocórrer|n\ el\ sistema)\pi_n}{\displaystyle\sum_{k=0}^{K-1} P(arribada\ a\ punt\ d'ocórrer|k\ el\ sistema)} \\
&= \lim_{h\to 0} \frac{(\lambda h + o(h))\pi_n}{\displaystyle\sum_{k=0}^{K-1} (\lambda h + o(h))\pi_k} = \frac{\lambda\pi_n}{\displaystyle\sum_{k=0}^{K-1} \lambda\pi_k} = \frac{\pi_n}{1-\pi_K}.
\end{aligned}$$

Sigui $\bar{\lambda}$ el valor esperat de clients que entren el sistema per unitat de temps, passem a calcular-ho. Cal tenir en compte que un individu entrarà en el sistema si el nombre de clients és menor que K, per tant:

$$\bar{\lambda} = \sum_{i=0}^{K-1} \lambda \pi_i = \lambda(1 - \pi_K).$$

Llavors, el nombre mitjà de clients perduts és:

$$\lambda - \bar{\lambda} = \lambda \pi_K.$$

Mitjançant les fórmules de Little, també sabem que:

$$W = \frac{L}{\bar{\lambda}} = \frac{L}{\lambda(1 - \pi_K)},$$

$$W_q = \frac{L_q}{\bar{\lambda}} = \frac{L - U_S}{\lambda(1 - \pi_K)} = W - \frac{1}{\mu}.$$

La probabilitat π_K juga un paper important a l'hora de calcular els diferents valors del comportament del sistema en mitjana. En concret, π_K, en ser la probabilitat que hi hagi K persones en el sistema, també ho és que arribi un client al sistema i es perdi. Anomenarem a aquesta: probabilitat de bloqueig o de pèrdua, i, amb la intenció de mantenir la notació que s'usa típicament en la literatura de Teoria de Cues per aquesta variable, la denotarem per P_B.

Com hem vist abans, P_B només dependrà de K i de ρ, de manera que escriurem:

$$P_B(K, \rho) = \frac{\rho^K}{\sum_{k=0}^{K} \rho^k}.$$

És senzill calcular que:

$$P_B(1, \rho) = \frac{\rho}{1 + \rho},$$

$$P_B(K, \rho) = \frac{\rho \rho^{K-1}}{\rho \rho^{K-1} + \sum_{k=0}^{K-1} \rho^k} = \frac{\rho P_B(K-1, \rho)}{1 + \rho P_B(K-1, \rho)}.$$

Per tant, podrem calcular la probabilitat de bloqueig per qualsevol valor de K de manera recursiva. És evident que, si agafem $\rho < 1$, aquesta probabilitat decreix a mesura que augmentem el valor de K. En particular, tendeix a 0 quan K tendeix a infinit. Per aquest motiu, donat un valor límit P^* per a la probabilitat de bloqueig, sempre trobarem un valor de K tal que $P_B(K, \rho) < P^*$.

4.4 Model M/M/n/K

També ens interessarà conèixer el comportament del model $M/M/n/K$ amb n servidors i capacitat finita K. De la mateixa manera que tots els altres models vists, es tracta d'un procés de naixement i mort, però en aquest cas:

$$\lambda_i = \begin{Bmatrix} \lambda, \ i < K \\ 0, \ i = K \end{Bmatrix}, \ \mu_i = \begin{Bmatrix} i\mu, \ i \leq n \\ n\mu, \ n < i \leq K \end{Bmatrix}.$$

Aplicant les equacions estacionàries obtenim:

$$\pi_0 = \left(1 + \sum_{i=0}^{K} \left(\prod_{c=0}^{i} \frac{\lambda_c}{\mu_{c+1}}\right)\right)^{-1} = \left(\sum_{i=0}^{n-1} \frac{\lambda^i}{i!\mu^i} + \sum_{i=n}^{K} \frac{\lambda^i}{n^{i-n}n!\mu^i}\right)^{-1}.$$

Utilitzant la mateixa notació que abans, $a = \rho/n$, ens queda:

$$\pi_0 = \begin{Bmatrix} \left(\dfrac{\rho^n(1-a^{K-n+1})}{(1-a)n!} + \displaystyle\sum_{i=0}^{n-1} \frac{\rho^i}{i!}\right)^{-1}, \ a \neq 1 \\ \left(\dfrac{\rho^n(K-n+1)}{n!} + \displaystyle\sum_{i=0}^{n-1} \frac{\rho^i}{i!}\right)^{-1}, \ a = 1 \end{Bmatrix}.$$

I repetint els càlculs del model $M/M/n$, tenim:

$$\pi_i = \begin{Bmatrix} \pi_0 \dfrac{\rho^i}{i!} \ (i \leq n) \\ \pi_0 \dfrac{a^i n^n}{n!} \ (n < i \leq K) \end{Bmatrix}.$$

Notem que a igual que en el model $M/M/1/K$, pel fet de tenir un nombre finit de persones que poden ser admeses dins el sistema, sempre ens trobarem en condicions estacionàries.

Podem veure que, en el cas de $\rho < 1$, per $K \to \infty$ tenim $a^{K-n+1} \to 0$, i recuperem el model $M/M/n$. A més, si fem $n = 1$, recuperem el model $M/M/1/K$. Dit d'una altra manera, podrem considerar aquest model com un cas més general de tots els que hem vist anteriorment.

4.4.1 Mesures d'efectivitat

A continuació, calcularem com es modifiquen els valors esperats dels diferents paràmetres que ens interessa conèixer d'aquest model, respecte al model $M/M/n$. Començarem calculant el valor esperat de gent esperant a la cua:

$$\begin{aligned} L_q &= \sum_{i=n+1}^{K} (i-n)\pi_i = \sum_{i=n+1}^{K} (i-n) \frac{\rho^i \pi_0}{n^{i-n} n!} = \frac{\pi_0 \rho^n}{n!} \sum_{i=n+1}^{K} (i-n) a^{i-n} \\ &= \frac{\pi_0 \rho^n a}{n!} \sum_{j=1}^{K-n} j a^{j-1} = \frac{\pi_0 \rho^n a}{n!} \frac{d}{da} \left(\sum_{j=1}^{K-n} a^j\right) \\ &= \frac{\pi_0 \rho^n a}{n!} \frac{d}{da} \left(\frac{1-a^{K-n+1}}{1-a}\right). \end{aligned}$$

Arribant a:

$$L_q = \frac{\pi_0 \rho^n a}{n!(1-a)^2} [1 - a^{K-n+1} - (1-a)(K-n+1)a^{K-n}].$$

Per trobar el valor en el cas concret en què $a = 1$, utilitzarem dues vegades la regla de l'Hopital:

$$\begin{aligned}\lim_{a\to 1} L_q &= \lim_{a\to 1} \frac{\pi_0\rho^n}{n!}\frac{1 - a^{K-n+1} - (1-a)(K-n+1)^2 a^{K-n}}{-2(1-a)} \\ &= \lim_{a\to 1} \frac{\pi_0\rho^n}{n!2}\{-(K-n+1)a^{K-n} \\ &\qquad - (K-n+1)^2[-a^{K-n} + (1-a)(k-n)a^{K-n-1}]\} \\ &= \frac{\pi_0\rho^n}{n!2}[(K-n+1)^2 - (K-n+1)] \\ &= \frac{\pi_0\rho^n}{n!2}(K-n+1)(K-n).\end{aligned}$$

El valor esperat de clients que entren en el sistema per unitat de temps serà:

$$\bar{\lambda} = \sum_{i=0}^{K-1} \lambda\pi_i = \lambda(1-\pi_K).$$

A més, el nombre mitjà de servidors ocupats serà:

$$\begin{aligned}\bar{n} &= \sum_{i=0}^{n-1} i\pi_i + \sum_{i=n}^{K} n\pi_i = \pi_0\rho\left(\sum_{i=0}^{n-2}\frac{\rho^i}{i!} + \frac{\rho^{n-1}}{(n-1)!}\sum_{i=0}^{K-n} a^i\right) \\ &= \pi_0\rho\left(\sum_{i=0}^{n-1}\frac{\rho^i}{i!} + \frac{\rho^{n-1}}{(n-1)!}\frac{a(1-a^{K-n})}{1-a}\right) \\ &= \pi_0\rho\left(\sum_{i=0}^{n-1}\frac{\rho^i}{i!} + \frac{\rho^n}{n!}\frac{(1-a^{K-n+1})}{1-a} + \frac{\rho^n a^{K-n}}{n!}\right) \\ &= \pi_0\rho\left(\pi_0^{-1} + \frac{a^K n^n}{n!}\right) = \rho(1-\pi_K).\end{aligned}$$

Sabent el valor de $\bar{n}$ podem calcular fàcilment $\bar{S}$ i L:

$$\bar{S} = n - \bar{n} = n + \rho\pi_K - \rho,$$

$$L = L_q + \bar{n} = L_q + \rho(1-\pi_K).$$

Ara, utilitzant les fórmules de Little obtenim:

$$W_q = \frac{L_q}{\bar{\lambda}} = \frac{L_q}{\lambda(1-\pi_K)},$$

$$W = \frac{L}{\bar{\lambda}} = W_q + \frac{1}{\mu}.$$

Observem que, agafant exactament els mateixos càlculs que vam fer per al model $M/M/1/K$, la probabilitat que un client entri en el sistema en un moment en el qual hi ha n clients és:

$$a_n = \frac{\pi_n}{1-\pi_K}.$$

Finalment, podem veure que la probabilitat de bloqueig és $P_B = \pi_K$, però, a diferència del model anterior, en aquest cas també dependrà del paràmetre n. Això farà que la seva expressió sigui molt més complicada que abans.

4.4.2 Model M/M/n/n

En aquesta subsecció ens centrarem en el cas concret en que la capacitat màxima del sistema és igual al nombre de servidors, és dir, el cas en que $K = n$. Aquest model té una gran importancia per ser un dels més utilitzats i també per ser el model amb que es va originar la teoria de cues.

El procés estarà en estat k si hi ha k persones en el sistema o, equivalentment, si hi ha k servidors ocupats. A més, els paràmetres del procés de naixement i mort que segueix el sistema són:

$$\lambda_k = \begin{Bmatrix} \lambda, & k < n \\ 0, & k \geq n \end{Bmatrix}, \; \mu_k = \begin{Bmatrix} k\mu, & k \leq n \\ 0, & n < k \end{Bmatrix}.$$

Com ja haviem comentat abans, pel fet que el sistema només pot admetre un nombre finit de persones, el sistema sempre es troba en condicions estacionaries.

Aplicant les equacions estacionàries obtenim:

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

$$\pi_k = \begin{Bmatrix} \pi_0 \dfrac{\rho^k}{k!}, & k \leq n \\ 0, & k > n \end{Bmatrix}.$$

La mesura més important d'aquest sistema és π_n, que serà la probabilitat de bloqueig. La seva fórmula ve donada per:

$$\pi_n = \frac{\dfrac{\rho^n}{n!}}{\displaystyle\sum_{k=0}^{n} \frac{\rho^k}{k!}}.$$

Aquesta és l'anomenada fórmula B d'Erlang, que denotarem per $B(n, \rho)$, i és la probabilitat que una persona que arribi al sistema no pugui entrar com a client.

En el cas particular que n sigui gran i ρ sigui petita podrem aproximar π_0 per $e^{-\rho}$, que s'aproxima a una distribució de Poisson. No obstant això, per n i ρ grans aquesta aproximació no serà vàlida.

Donat que per valors grans de n el càlcul d'aquest paràmetre pot ser complicat, donarem una fórmula recursiva que ens faciliti aquest càlcul:

$$\begin{aligned} B(n,\rho) &= \frac{\dfrac{\rho^n}{n!}}{\displaystyle\sum_{i=0}^{n} \frac{\rho^i}{i!}} = \frac{\dfrac{\rho}{n}\dfrac{\rho^{n-1}}{(n-1)!}}{\displaystyle\sum_{i=0}^{n-1} \frac{\rho^i}{i!} + \frac{\rho}{n}\frac{\rho^{n-1}}{(n-1)!}} \\ &= \frac{\dfrac{\rho}{n} B(n-1,\rho)}{1 + \dfrac{\rho}{n} B(n-1,\rho)} = \frac{\rho B(n-1,\rho)}{n + \rho B(n-1,\rho)} \end{aligned} \tag{4.1}$$

Es fàcil veure que $B(1, \rho) = \frac{\rho}{1+\rho}$. Per tant, ho podem usar com a valor inicial per a calcular $B(n, \rho)$ per a qualsevol n, sense necessitat d'haver d'utilitzar la funció factorial.

Recordant la fórmula de retard de Erlang, notem que podrem escriure $C(n,\rho)$ com una funció de $B(n,\rho)$ de la següent manera:

$$\begin{aligned}
C(n,\rho) &= \frac{\dfrac{n\rho^n}{(n-\rho)n!}}{\displaystyle\sum_{i=0}^{n-1}\frac{\rho^i}{i!}+\frac{n\rho^n}{(n-\rho)n!}} = \frac{\dfrac{\rho^n}{n!}}{\left(1-\dfrac{\rho}{n}\right)\displaystyle\sum_{i=0}^{n-1}\frac{\rho^i}{i!}+\frac{\rho^n}{n!}} \\
&= \frac{B(n,\rho)}{(1-B(n,\rho))\left(1-\dfrac{\rho}{n}\right)+B(n,\rho)} \\
&= \frac{B(n,\rho)}{1-\dfrac{\rho}{n}(1-B(n,\rho))} \\
&= \frac{nB(n,\rho)}{n-\rho+\rho B(n,\rho)}.
\end{aligned} \tag{4.2}$$

De l'equació anterior es fa evident que $B(n,\rho) < C(n,\rho)$. A més, també podrem escriure $C(n,\rho)$ de forma recursiva, la qual cosa ens serà útil pels mateixos motius que $B(n,\rho)$, com mostrem a continuació.

Primer, de l'equació (4.2) es pot deduir fàcilment que:

$$\begin{aligned}
C(n,\rho) = \frac{nB(n,\rho)}{n-\rho+\rho B(n,\rho)} &\implies (n-\rho)C(n,\rho) = B(n,\rho)(n-\rho C(n,\rho)) \\
&\implies B(n,\rho) = \frac{(n-\rho)C(n,\rho)}{n-\rho C(n,\rho)}.
\end{aligned} \tag{4.3}$$

Utilitzant l'equació (4.1) a l'equació (4.2) tenim:

$$\begin{aligned}
C(n,\rho) &= \frac{nB(n,\rho)}{n-\rho+\rho B(n,\rho)} = \frac{n\dfrac{\rho B(n-1,\rho)}{n+\rho B(n-1,\rho)}}{n-\rho+\rho\dfrac{\rho B(n-1,\rho)}{n+\rho B(n-1,\rho)}} \\
&= \frac{n\dfrac{\rho B(n-1,\rho)}{n+\rho B(n-1,\rho)}}{\dfrac{n(n-\rho+\rho B(n-1,\rho))}{n+\rho B(n-1,\rho)}} = \frac{\rho B(n-1,\rho)}{n-\rho+\rho B(n-1,\rho)}.
\end{aligned}$$

I ara, fent servir (4.3):

$$\begin{aligned}
C(n,\rho) &= \frac{\rho\dfrac{(n-1-\rho)C(n-1,\rho)}{n-1-\rho C(n-1,\rho)}}{n-\rho+\rho\dfrac{(n-1-\rho)C(n-1,\rho)}{n-1-\rho C(n-1,\rho)}} = \frac{\rho\dfrac{(n-1-\rho)C(n-1,\rho)}{n-1-\rho C(n-1,\rho)}}{\dfrac{(n-1)(n-\rho)-\rho C(n-1,\rho)}{n-1-\rho C(n-1,\rho)}} \\
&= \frac{\rho(n-1-\rho)C(n-1,\rho)}{(n-1)(n-\rho)-\rho C(n-1,\rho)}.
\end{aligned}$$

Notem que $C(1,\rho)=\rho$, valor que podrem utilitzar com a inicial a l'hora de fer el càlcul recursiu.

4.5 Resum de resultats

Per tal de presentar de manera més clara els resultats obtinguts dels diferents models estudiats, presentem la Taula 1 amb els valors de les mesures d'efectivitat de a cada un.

Mesures	$M/M/1$	M/M/n	M/M/1/K	M/M/n/K
π_0	$1-\rho$	$\left(\sum_{i=0}^{n-1}\frac{\rho^i}{i!}+\frac{\rho^n}{n!(1-a)}\right)^{-1}$	$\frac{1-\rho}{1-\rho^{K+1}}$	$\left(\frac{\rho^n(1-a^{K-n+1})}{n!(1-a)}+\sum_{i=0}^{n-1}\frac{\rho^i}{i!}\right)^{-1}$
π_k	$(1-\rho)\rho^k$	$\left\{\begin{matrix}\pi_0\frac{\rho^k}{k!}\ (k\leq n)\\ \pi_0\frac{a^k n^n}{n!}\ (k>n)\end{matrix}\right\}$	$\pi_0\rho^k$	$\left\{\begin{matrix}\pi_0\frac{\rho^k}{k!}\ (k\leq n)\\ \pi_0\frac{a^k n^n}{n!}\ (n<k\leq K)\end{matrix}\right\}$
L	$\frac{\rho}{1-\rho}$	$\rho+\frac{\rho}{n-\rho}C(n,\rho)$	$\frac{\rho(1-(K+1)\rho^K+K\rho^{K+1})}{(1-\rho^{K+1})(1-\rho)}$	$L_q+\rho(1-\pi_K)$
L_q	$\frac{\rho^2}{1-\rho}$	$\frac{\rho}{n-\rho}C(n,\rho)$	$L+\pi_0-1$	$\frac{\pi_0\rho^n a}{n!(1-a)^2}[1-a^{K-n+1}-(1-a)(K-n+1)a^{K-n}]$
W	$\frac{1}{\mu-\lambda}$	$\frac{1}{\mu}+\frac{1}{\mu(n-\rho)}C(n,\rho)$	$\frac{L}{\lambda(1-\pi_K)}$	$\frac{L}{\lambda(1-\pi_K)}$
W_q	$\frac{\rho}{\mu-\lambda}$	$\frac{1}{\mu(n-\rho)}C(n,\rho)$	$\frac{L+\pi_0-1}{\lambda(1-\pi_K)}$	$\frac{L_q}{\lambda(1-\pi_K)}$
P_B	0	0	$\frac{\rho^K}{\sum_{k=0}^{K}\rho^k}$	$\pi_0\frac{a^K n^n}{n!}$
$\bar{n}$	ρ	ρ	$1-\pi_0$	$\rho(1-\pi_K)$
$\bar{S}$	$1-\rho$	$n-\rho$	π_0	$n+\rho\pi_K-\rho$

Taula 1: Mesures d'efectivitat

Cal destacar que només apareixen els casos en que $a<1$, ja que són els casos que estudiarem i l'únic cas en que, simultàniament, els 4 models es troben en condicions en que tenim distribució estacionaria.

A més, recordem que:

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

5 Disseny òptim d'un sistema de cues

Recordem que l'objectiu principal d'aquesta memòria és trobar una configuració que optimitzi la relació costs/beneficis d'un sistema de cues donat, mitjançant l'optimització de l'esperança de la funció de cost per unitat de temps de dit sistema.

Com podem veure a l'equació (2.1), l'esperança de la funció de cost depèn de sis constants que determinen els diferents tipus de costs que pot presentar un sistema i els beneficis que s'obtenen per client. A banda d'això, també depèn de λ, μ, n i K de diferents maneres en funció del model que estem estudiant. Tot això fa que trobar una configuració òptima de forma analítica sigui excessivament complicat.

Per aquest motiu, el que farem serà dissenyar un programa informàtic que, donats els valors de les constants de cost d'un sistema concret i tres de les altres 4 variables, ens retorni el valor de la variable restant que fa òptim el sistema.

En aquesta secció explicarem el funcionament del programa i presentarem exemples d'optimització de sistemes de cues dels diferents models estudiats. També donarem una aproximació del nombre òptim de servidors, n, per un model $M/M/n$, alternativa al càlcul del mínim de la funció de cost i en la que es busqui la millor relació cost/qualitat de servei.

5.1 Funcionament del programa

A continuació, donarem una explicació detallada de com funciona el programa que hem dissenyat per a trobar la configuració òptima. Cal destacar que només donarem aquesta explicació pel programa del model $M/M/n/K$, donat que els altres tenen una estructura anàloga, però canviant les fórmules per als diferents valors esperats i amb un nombre menor de paràmetres a escollir per optimitzar. Per veure el codi font anar a l'Annex A.

Per començar, es pregunta quin paràmetre volem optimitzar i que es donin els valors de: CS, CWS, CI, CSR, CLC, i R. Posteriorment, ens preguntarà pels valors de la resta de paràmetres. Si, per exemple, volguéssim calcular el valor òptim de n, ens preguntaria pels valors de λ, μ i K.

Un cop determinats tots els valors coneguts, el programa seguirà diferents processos segons la variable a optimitzar escollida:

- n: En el cas que vulguem optimitzar n primer calculem el valor de ρ i declarem dues variables que representen el cost mínim i el valor de n per a aquest. Seguidament, fem un bucle for des del primer valor tal que $a < 1$ fins a K, on es vagi calculant el valor de l'esperança de la funció de cost per a cada un. A cada iteració es compara la variable cost mínim amb el nou valor. En cas que el nou valor sigui major o igual al de la variable no es fa res, però en cas que sigui menor aquest es guarda a la variable i també es guarda el nou valor de n. Al final es donen el valor de cost mínim, i el seu n corresponent, i un gràfic amb el valor del cost per cada n estudiat.
- λ: En el cas que vulguem optimitzar λ primer declarem dues variables que representen el cost mínim i el valor de λ per a aquest. Seguidament, fem un bucle for des de y fins a un cert nombre x, i augmentant de y en y, on es vagi calculant el valor de l'esperança de la funció de cost per a cada un. A cada iteració es comprova que $a < 1$, en cas contrari es trenca el bucle, i es compara la variable cost mínim amb el nou valor. En cas que el nou valor sigui major o igual al de la variable no es

fa res, però en cas que sigui menor aquest es guarda a la variable i també es guarda el nou valor de λ. Al final es donen el valor de cost mínim, i el seu λ corresponent, i un gràfic amb el valor del cost per cada λ estudiat.

- μ: En el cas que vulguem optimitzar μ primer declarem dues variables que representen el cost mínim i el valor de μ per a aquest. Seguidament, fem un bucle for des de y fins a un cert nombre x, i augmentant de y en y, on es vagi calculant el valor de l'esperança de la funció de cost per a cada un. A cada iteració es comprova que $a < 1$, i es compara la variable cost mínim amb el nou valor. En cas que el nou valor sigui major o igual al de la variable no es fa res, però en cas que sigui menor aquest es guarda a la variable i també es guarda el nou valor de μ. Al final es donen el valor de cost mínim, i el seu μ corresponent, i un gràfic amb el valor del cost per cada μ estudiat.

- K: En el cas que vulguem optimitzar K primer es comprova que $a < 1$, després declarem dues variables que representen el cost mínim i el valor de K per a aquest. Seguidament, fem un bucle for des de n fins a un cert nombre x, on es vagi calculant el valor de l'esperança de la funció de cost per a cada un. A cada iteració es compara la variable cost mínim amb el nou valor. En cas que el nou valor sigui major o igual al de la variable no es fa res, però en cas que sigui menor aquest es guarda a la variable i també es guarda el nou valor de K. Al final es donen el valor de cost mínim, i el seu K corresponent, i un gràfic amb el valor del cost per cada K estudiat.

Aquí acabaria la descripció del funcionament, però podem fer una sèrie d'observacions:

- Els càlculs de la funció de cost es fan substituint les fórmules de la Taula 1 a l'equació (2.1). És important destacar que, per tal que sigui possible fer els càlculs per valors grans de n i K, serà necessari utilitzar les fórmules recursives per a aquelles variables que n'havem vist. Aquestes fórmules les podem trobar al llarg de la secció 4.

- Podem utilitzar el programa del model $M/M/n$ per optimitzar models $M/M/1$ i el programa del model $M/M/n/K$ per models $M/M/1/K$ només fent $n = 1$, tot i que no serà computacionalment eficient. No obstant això, donat que hi ha una fita al valor màxim que es pugui donar a K en el programa, i a més treballar amb ella serà molt costós, no tindria sentit usar els programa dels models de capacitat finita per optimitzar els de capacitat infinita.

- No té sentit considerar el cas en què el nombre de servidors sigui major que la capacitat màxima del sistema, $n > K$. Aquesta és la condició que ens determina els límits superior i inferior dels bucles per a l'optimització de n i K respectivament.

- Tot i que demanem que $a < 1$ per a tots els models, això només serà necessari als de capacitat infinita. Si ho fem és per poder treballar amb tots de manera unificada, però, donat que els de capacitat finita tenen sempre distribució estacionaria, aquesta condició podria ser eliminada del programa. Així i tot, en la majoria de casos no és interessant que $a > 1$, ja que pot suposar un augment molt considerable dels costos i sistemes molt colapsats.

- Els valors x i y dels bucles on apareixen es poden posar al gust de l'usuari. Per als casos concrets de μ i λ, serà interessant agafar una x prou gran per a estudiar el nombre més gran possible de casos de a, mentre que per al cas de K hem d'anar amb compte d'agafar un valor massa gran perquè es puguin calcular les variables. Pel que fa a la y, serà interessant ajustar-la per obtenir la precisió desitjada. Si y és massa gran, és possible que ens saltem el valor més precís, però si és massa petit, el programa tardarà molt en executar-se.

- Tal com hem plantejat el problema, l'estudi de l'optimització de μ no té gaire sentit. Està clar que fixat CSR, com major sigui μ més rendible serà el sistema. El que tindria més sentit seria calcular els costos ajustant CSR a cada nova μ.
- En els models de capacitat infinita $P_B = 0$, de manera que per aquests $CLC = 0$ sempre.

5.2 Exemples d'optimització

Per tal de comprovar la utilitat del programa que hem dissenyat, a continuació presentem alguns exemples d'optimització de la funció de cost. Podem trobar el codi font a l'Annex A.

Exemple 5.1. Suposem que volem obrir un nou supermercat i hem de decidir el nombre de caixes, n, que volem posar. Després de fer els càlculs, s'estima que entraran uns 3 clients per minut al supermercat, $\lambda = 3$, i que cada caixa tardarà un minut en servir cada client, $\mu = 1$, en mitjana. El cost de tenir un servidor és de $CS = 1$ euro per minut, el cost de tenir clients en el sistema és de $CWS = 25$ euros el minut i el cost de mantenir un servidor desocupat és $CI = 1$ euro el minut. El recinte és suficientment gran com perquè no hi hagi limitacions en la quantitat de persones que hi puguin entrar, per tant $CLC = 0$ euros. Finalment, la mitjana de diners que es deixa un client per compra és de $R = 30$ euros. També voldrem saber si el sistema serà rentable per algun nombre n.

Primer de tot, notem que es tracta d'un model $M/M/n$, donat que la capacitat del sistema es considera infinita, aleshores haurem d'utilitzar el programa d'aquest model. Mitjançant el programa hem trobat que el cost mínim s'obté a $n = 5$ i és, aproximadament, 50.86 euros. A més, a la Figura 1 podem veure la gràfica proporcionada. Notem que el cost és positiu fins i tot en el mínim, això voldrà dir que el sistema no és rendible i que en mitjana només generarà pèrdues.

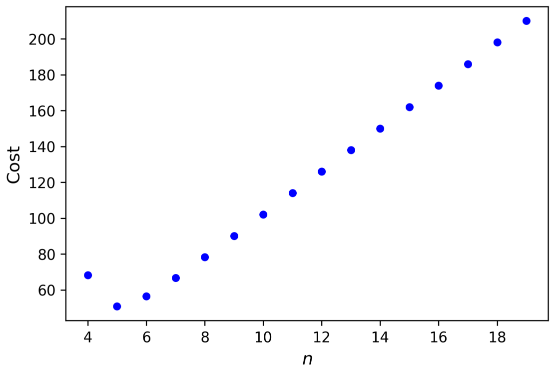

Figura 1: Cost esperat per unitat de temps en funció de n per a l'exemple 5.1

Exemple 5.2. Considerem el mateix sistema de l'exemple anterior amb els mateixos paràmetres de cost i $\mu = 1$. Ara voldrem veure si utilitzant el valor òptim $n = 5$, existeix algun λ que faci que el sistema produeixi beneficis.

Mitjançant el programa hem trobat que el cost mínim s'obté a $\lambda = 2.4$ i és, aproximadament, 48.22. Això voldrà dir que no exitirà tampoc cap λ que faci que el sistema sigui rendible. A més, a la Figura 2 podem veure la gràfica proporcionada. En aquest cas, hem agafat $x = 4.5$, per poder veure amb claredat la regió on apareix el punt crític, i $y = 0.05$, per tenir una precisió prou bona. Notem que per $x \geq 5$ ens sortim de la regió en la que tenim distribució estacionaria i que per valors propers a 5 el cost creix de forma desmesurada.

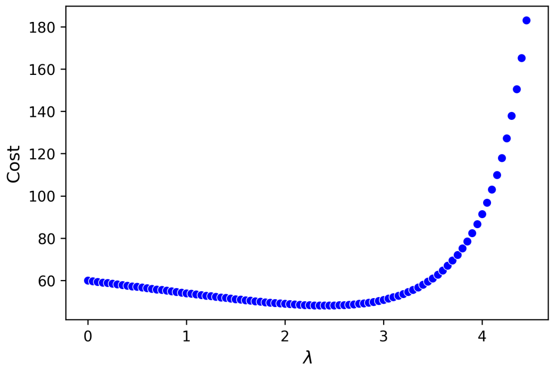

Figura 2: Cost esperat per unitat de temps en funció de λ per a l'exemple 5.2

Exemple 5.3. Seguint amb el mateix sistema d'abans, ara voldrem veure com canvia el cost esperat si considerem que la capacitat sigui finita. Per un model $M/M/n/K$ els paràmetres i costs: $\lambda = 4$, $\mu = 1$, $K = 30$, $CS = 1$, $CWS = 25$, $CI = 1$, $CSR = 10$, $CLC = 2$, $R = 30$. Volem trobar el valor òptim de n.

Mitjançant el programa hem trobat que el cost mínim s'obté a $n = 6$ i és, aproximadament, 62.23. A més, a la Figura 3 podem veure la gràfica proporcionada. Evidentment, el sistema segueix fora produir beneficis en mitjana.

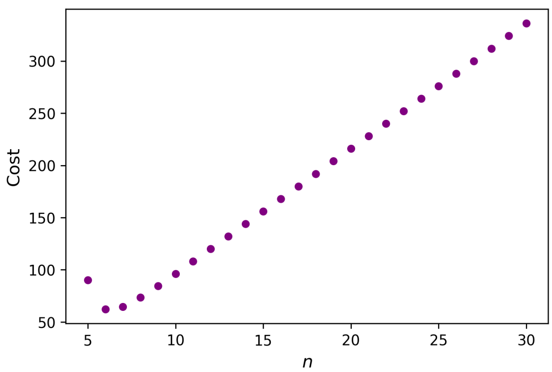

Figura 3: Cost esperat per unitat de temps en funció de n per a l'exemple 5.3

Exemple 5.4. Considerem el mateix model utilitzat el valor de $n = 6$, que ens ha fet mínim el cost, i voldrem trobar el valor òptim de λ.

Mitjançant el programa hem trobat que el cost mínim s'obté a $\lambda = 3.15$ i és, aproximadament, 56.41. A més, a la Figura 4 podem veure la gràfica proporcionada. De la mateixa manera que a l'exemple 5.2, fem $x < 6$, per tal de provar tots els valors que fan $a < 1$, i fem $y = 0.05$ per tenir una precisió prou bona.

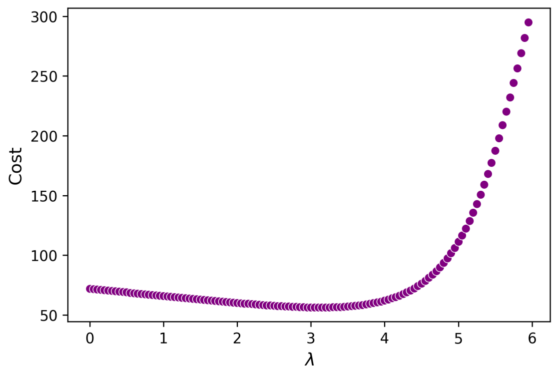

Figura 4: Cost esperat per unitat de temps en funció de λ per a l'exemple 5.4

Exemple 5.5. Finalment, considerarem el mateix sistema per n i λ que ens han fet mínims els costs de 5.3 i 5.4, respectivament. Ara voldrem trobar el valor òptim de K.

Mitjançant el programa hem trobat que el cost mínim s'obté a $K = 6$ i és, aproximadament, 54.63. A més, a la Figura 5 podem veure la gràfica proporcionada. Donat que els valors dels paràmetres de cost són els mateixos que als exemples anteriors, hem fixat $x = 30$ per veure com es relaciona la K que fa mínim el cost amb els valors de K dels exemples previs.

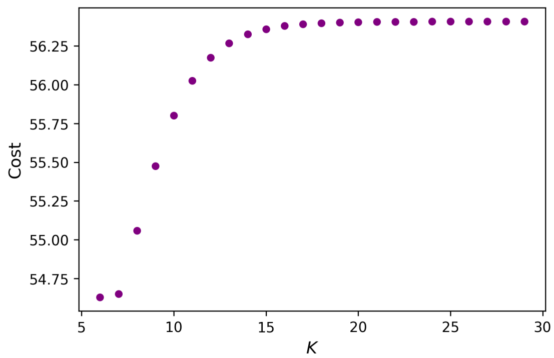

Figura 5: Cost esperat per unitat de temps en funció de K per a l'exemple 5.5

Així, podem concloure dient que per un sistema amb els costs donats no hi ha configuració que no generi pèrdues en mitjana.

5.3 Aproximació de n òptim

A l'hora de dissenyar un sistema de cues, reduir al mínim el seu cost siguin quines siguin les conseqüències no serà una bona pràctica. Mantenir un nivell elevat en la qualitat de servei és també molt important, ja que males opinions dels clients poden afectar negativament als beneficis a mitjan o llarg termini.

Amb freqüència, per tal de tenir una qualitat de servei elevada, el nombre de servidors, n, haurà de ser més elevat del que obtindríem optimitzant l'equació (2.1). Però decidir exactament quants servidors més posar pot ser una tasca complicada. El nostre objectiu serà proporcionar una aproximació que ens permeti escollir un n que faci mínim el cost, mantenint un cert mínim de qualitat prèviament establert per nosaltres. És rellevant destacar que això només ho farem per un model $M/M/n$.

Primer de tot, recordem que perquè existeixi distribució estacionaria necessitem que $a < 1$, o, dit d'una altra manera, que $\rho < n$. Per tant, podrem escriure:

$$n = \rho + \Delta,$$

on Δ es pot entendre com el nombre de servidors en què s'excedeix ρ. Notem que, com això és només una interpretació i no una quantitat real de servidors del sistema, Δ pot no ser enter.

A l'hora de plantejar-nos quin és l'equilibri adequat entre cost i qualitat del servei, tenim principalment tres maneres d'enfocar el problema: emfatitzar la qualitat del servei, que anomenarem domini de qualitat; emfatitzar la reducció dels costs, que anomenarem domini d'eficiència; o buscar un equilibri entre els dos, que anomenarem domini de qualitat i eficiència o QED.

D'aquesta manera, si tenim un sistema inicial que volem redimensionar a causa d'un canvi que s'hagi produït sobre ρ, els diferents dominis buscaran fer constant a, Δ i $1 - F_{W_q}(0)$ respectivament. Donat que usar el domini de qualitat tendeix a redimensionar els sistemes de forma excessivament cara i que usar el domini d'eficiència tendeix a provocar sistemes constantment congestionats, ens interessarà centrar-nos en el QED.

No obstant això, donat que en el QED el que es manté constant és $1 - F_{W_q}(0)$, la fórmula per trobar n no serà simple. Per aquest motiu, ens interessarà tenir una aproximació que ens permeti trobar fàcilment n. Aquesta serà proporcionada per la llei de l'arrel quadrada de ρ i ens diu que, per valors grans de ρ, el valor de n que manté $1 - F_{W_q}(0)$ constant és:

$$n \approx \rho + \beta\sqrt{\rho}, \tag{5.1}$$

on β és una constant. En altres paraules, que per mantenir $1-F_{W_q}(0)$ constant, el nombre extra de servidors haurà de créixer com l'arrel quadrada de ρ.

Recordant que $1 - F_{W_q}(0) = C(n, \rho)$, obtindrem la justificació teòrica de la nostra l'aproximació del següent teorema, tal com es mostra a [6]:

Teorema 5.6. *Considerem una seqüència de cues $M/M/n$ indexades per un paràmetre $c \in \mathbb{N}$. Suposem que la cua c té $n_c = c$ servidors i ρ_c. Llavors:*

$$\lim_{c\to\infty} C(c, \rho_c) = \alpha, \quad 0 < \alpha < 1, \tag{5.2}$$

Si, i només si:

$$\lim_{c\to\infty} \frac{c - \rho_c}{\sqrt{c}} = \beta, \quad \beta > 0, \tag{5.3}$$

On $C(n, \rho)$ és la fórmula C d'Erlang, on α i β són constants relacionades per:

$$\alpha = \frac{\phi(\beta)}{\phi(\beta) + \beta\Phi(\beta)},$$

sent ϕ i Φ són la funció de densitat de probabilitat i la funció de distribució d'una variable aleatòria normal estàndard, respectivament.

L'equació (5.2) es correspon amb la condició que $1 - F_{W_q}(0)$ sigui constant al llarg de la seqüència de cues i de l'equació (5.3) tenim que aquesta condició es satisfà per valors elevats de c quan $c \approx \rho_c + \beta\sqrt{c}$. La llei de l'arrel quadrada de ρ substitueix $\sqrt{c}$ per $\sqrt{\rho_c}$. Els paràmetres α i β podran ser interpretats com a constants que determinen la qualitat del servei, ja que α és exactament la probabilitat que el temps d'espera a la cua sigui diferent de 0 i β està directament relacionat amb α tal com es mostra al teorema.

Aquesta fórmula ens permetrà trobar el nombre c de servidors que ens donen una certa qualitat de servei α. Això ho farem trobant la β que satisfaci la relació amb α donada pel teorema i substituint a l'equació (5.1). També podem utilitzar l'aproximació que β sigui el $(1-\alpha)$ quantil d'una distribució normal estàndard, per facilitar el càlcul de β. Per més informació sobre els detalls de les deduccions d'aquestes aproximacions, consultar [6] i [7].

Per tal d'il·lustrar com s'utilitzaria aquesta fórmula presetem el següent exemple:

Exemple 5.7. Sigui una cua $M/M/n$ amb $\lambda = 250$ i $\mu = 1$, volem trobar el mínim valor de n tal que $C(n, \rho) \leq 0.01$. Sabem que β serà el (1-0.01) quantil d'una distribució normal estàndard, llavors $\beta = 2.326$ i tenim que $n \approx 250 + 2.326\sqrt{250} = 286.78$.

6 Conclusions

En les primeres seccions d'aquesta memòria hem pogut determinar les propietats principals dels models de cues: $M/M/1$, $M/M/n$, $M/M/1/K$ i $M/M/n/K$. En altres paraules, hem explicat en què consisteixen, hem aconseguit donar les bases matemàtiques necessàries per poder-los estudiar i els hem estudiat en el règim en què existeix distribució estacionària. Ens hem centrat principalment, però, en el càlcul de les seves mesures d'efectivitat, que podem trobar resumides a la Taula 1. A més, per als models de capacitat infinita hem vist que, en estat estacionari, el procés de sortida segueix un procés de Poisson del mateix paràmetre que el procés de Poisson de les arribades, λ, i hem estudiat com són les distribucions dels temps d'espera: T i T_q.

Posteriorment, utilitzant aquests resultats, hem pogut dissenyar un programa que retorna el valor mínim de l'esperança de la funció de cost per unitat de temps en funció de: n, λ, μ o K, fixant prèviament els valors dels paràmetres de cost i les variables que no siguin la que volem optimitzar.

Finalment, hem proporcionat una fórmula que ens permet calcular de manera aproximada el valor òptim de n per un model $M/M/n$, en el cas que ens interessi optimitzar la relació cost/qualitat de servei, en lloc de només fer mínim el cost.

A Codi font

A continuació, presentem el codi font que s'ha fet servir a la secció 5 en el càlcul dels exemples. Cal destacar que el llenguatge de programació utilitzat és Python.

Per tal d'evitar redundància, només inclourem el cas del model $M/M/n/K$. El codi dels altres models seria anàleg, però canviant les fórmules de les funcions auxiliars per les pertinents i ajustant els apartats de: paràmetre a optimitzar i declaració de paràmetres coneguts. A més, la versió presentada usa les fórmules tal com apareixen a la Taula 1, donat que hem treballat amb valors de n i K petits al llarg de la memòria. Si calgués, també s'haurien de canviar les fórmules que donessin problemes per les seves versions recursives.

```python
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

"""# Funcions auxiliars"""

def erlangc (n, rho):
  """
  Funcio que donats els valors de n i rho retorna el valor de la funcio
  C de Erlang per aquests parametres.

  Args:
  n: int nombre de servidors
  rho: float intensitat de trafic

  Returns:
  c: valor de la funcio c de erlang
  """

  aux1=n*(rho**n)
  aux2=(n-rho)*np.math.factorial(n)

  aux3=1
  for i in range(1, n, 1):
    aux3+=(rho**i)/np.math.factorial(i)

  c=(aux1/aux2)/(aux3+(aux1/aux2))

  return c

def pi0 (n, K, rho):
  """
  Funcio que donats els valors de n, K i rho retorna el valor de la
  probabilitat que no hi hagui ningu per aquests parametres.

  Args:
  n: int nombre de servidors
  K: capacitat maxima del sistema
  rho: float intensitat de trafic

  Returns:
  pi_0: probabilitat que no hi hagui ningu en el sistema
  """

  aux1=1
  for i in range(1, n, 1):
```

```python
    aux1+=(rho**i)/np.math.factorial(i)

  aux2=np.math.factorial(n)*(1-(rho/n))
  aux3=(rho**n)*(1-(rho/n)**(K-n+1))
  aux4=aux1+aux3/aux2

  pi_0=1/aux4

  return pi_0

def pi (n, K, rho):
  """
  Funcio que donats els valors de n, K i rho retorna un vector amb les
  probabilitats que hi ha i individus en el sistema per aquests parametres
    .

  Args:
  n: int nombre de servidors
  K: capacitat maxima del sistema
  rho: float intensitat de trafic

  Returns:
  pi_k: vector de probabilitats
  """
  pi_k=[]
  pi_k.append(pi0(n, K, rho))

  for i in range(1, K+1, 1):
    if i>n:
      aux1=pi_k[0]*((rho/n)**i)*(n**n)
      aux2=np.math.factorial(n)
      pi_k.append(aux1/aux2)
    else:
      aux1=pi_k[0]*(rho**i)
      aux2=np.math.factorial(i)
      pi_k.append(aux1/aux2)

  return pi_k

from math import factorial
def L(rho, n, K, pi_k):
  """
  Funcio que donats els valors de n i rho retorna el valor del temps mitja
  d'espera en el sistema.

  Args:
  n: int nombre de servidors
  rho: float intensitat de trafic
  K: capacitat maxima del sistema
  pi_k: vector de probabilitats

  Returns:
  l: temps mitja d'espera en el sistema
  """
  lq=pi_k[0]*(rho**n)*(rho/n)*(1-(rho/n)**(K-n+1)-(1-(rho/n))*(K-n+1)*(rho
    /n)**(K-n))
  lq=lq/(np.math.factorial(n)*((1-(rho/n))**2))
  l = lq + rho*(1-pi_k[K])

  return l

```

```python
def S(rho, n, pi_K):
  """
  Funcio que donats els valors de n, rho i pi_K retorna el valor del
    nombre
  mitja de servidors en desus.

  Args:
  n: int nombre de servidors
  rho: float intensitat de trafic
  pi_K: probabilitat de bloqueig

  Returns:
  s: nombre mitja de servidors en desus
  """

  s=n-rho+rho*pi_K
  return s

"""# Parametres

"""

#Declaracio de les constants de cost
CS=float(input("CS: "))
CWS=float(input("CWS: "))
CI=float(input("CI: "))
CSR=float(input("CSR: "))
CLC=float(input("CLC: "))
R=float(input("R: "))

#Parametre a optimitzar
print("Escolliu el parametre a optimitzar:\n 1:n\n 2:lambda\n 3:mu\n 4:K\n
   ")
opti=int(input())

while opti!=1 and opti!=2 and opti!=3 and opti!=4:
  print("Heu de seleccionar una de les opcions: ")
  opti=int(input())

#Declaracio de parametres coneguts
if opti==1:
  lamb = float(input("Escriviu el valor de lambda: "))
  mu = float(input("Escriviu el valor de mu: "))
  K = int(input("Escriviu el valor de K: "))
if opti==2:
  n = float(input("Escriviu el valor de n: "))
  mu = float(input("Escriviu el valor de mu: "))
  K = int(input("Escriviu el valor de K: "))
if opti==3:
  lamb = float(input("Escriviu el valor de lambda: "))
  n = int(input("Escriviu el valor de n: "))
  K = int(input("Escriviu el valor de K: "))
if opti==4:
  lamb = float(input("Escriviu el valor de lambda: "))
  mu = float(input("Escriviu el valor de mu: "))
  n = int(input("Escriviu el valor de n: "))

#calcul del valor optim de n
if opti==1:
  rho=lamb/mu
  cost=[]
```

```
  valors_n=[]
  min = np.inf
  index_min=0
  for i in range(1, K+1, 1):
    if (rho/i)<1:
      pi_k=pi(i, K, rho)
      costi = i*CS+L(rho, i, K, pi_k)*CWS+S(rho, i, pi_k[K])*CI+i*CSR+lamb*pi_k[K]*CLC-lamb*(1-pi_k[K])*R
      cost.append(costi)
      valors_n.append(i)
      if costi<min:
        min=costi
        index_min=i

  print("El minim cost s'obte a n={} i es {}".format(index_min, min))
  aux="$n$"

#calcul del valor optim de lambda
elif opti==2:
  cost=[]
  valors_n=[]
  min = np.inf
  val_lam=0
  for i in range(10000):
    lamb=i*0.05
    rho=lamb/mu
    if (rho/n)<1:
      pi_k=pi(n, K, rho)
      costi = n*CS+L(rho, n, K, pi_k)*CWS+S(rho, n, pi_k[K])*CI+n*CSR+lamb*pi_k[K]*CLC-lamb*(1-pi_k[K])*R
      cost.append(costi)
      valors_n.append(lamb)
      if costi<min:
        min=costi
        val_lam=lamb
  print("El minim cost s'obte a lambda={} i es {}".format(val_lam, min))
  aux="$\lambda$"

#calcul del valor optim de mu
elif opti==3:
  cost=[]
  valors_n=[]
  min = np.inf
  val_mu=0
  for i in range(10000):
    mu=0.05+i*0.05
    rho=lamb/mu
    if (rho/n)<1:
      pi_k=pi(n, K, rho)
      costi = n*CS+L(rho, n, K, pi_k)*CWS+S(rho, n, pi_k[K])*CI+n*CSR+lamb*pi_k[K]*CLC-lamb*(1-pi_k[K])*R
      cost.append(costi)
      valors_n.append(mu)
      if costi<min:
        min=costi
        val_mu=mu
  print("El minim cost s'obte a mu={} i  s  {}".format(val_mu, min))
  aux="$\mu$"

#calcul del valor optim de K
else:
```

```
  rho=lamb/mu
  cost=[]
  valors_n=[]
  min = np.inf
  val_K=0
  if (rho/n)<1:
    for i in range(n, 30, 1):
      pi_k=pi(n, i, rho)
      costi = n*CS+L(rho, n, i, pi_k)*CWS+S(rho, n, pi_k[i])*CI+n*CSR+lamb
   *pi_k[i]*CLC-lamb*(1-pi_k[i])*R
      cost.append(costi)
      valors_n.append(i)
      if costi<min:
        min=costi
        val_K=i
    print("El minim cost s'obte a K={} i es {}".format(val_K, min))
    aux="$K$"
  else:
    print("rho dividit n ha de ser menor que 1")
    aux=0

"""# Grafiques

"""
if aux!=0:
  data={"x": valors_n,
        "y": cost}


  fig=plt.figure()
  sns.scatterplot(data= data, x="x", y="y", color="purple")
  plt.xlabel(aux, size=12)
  plt.ylabel("Cost", size=12)
  plt.savefig("mmnk_opti.png", dpi=300, bbox_inches="tight" )
```

Referències

[1] Gross, D., Shortle, J. F., Thompson, J. M., and Harris, C. M.: *Fundamentals of Queueing Theory*, 5a ed., John Wiley & Sons, 2008.

[2] J. Sztrik: *Basic Queueing Theory*, University of Debrecen, 2011.

[3] J. Medhi: *Stochastic Models in Queueing Theory*, 2a ed., Academic Press, 2003.

[4] Rovira, C.: *Processos Estocàstics*, Universitat de Barcelona, 2021, `http://hdl.handle.net/2445/178195`.

[5] Grupo CDPYE-UGR: Esperanza matemática de una variable aleatoria mixta, `https://www.ugr.es`.

[6] S. Halfin, W. Whitt: Heavy-Traffic Limits for Queues with Many Exponential Servers, *Operations Research*, 29:567-588, 1981.

[7] L. V. Green, P. J. Kolesar, W. Whitt: Coping with Time-Varying Demand When Setting Staffing Requirements for a Service System, *Production and Operations Management*, 16:13-39, 2007.