UNIVERSIDAD AUTONOMA DE GUADALAJARA

Incorporada a la Universidad Nacional Autónoma de México

Escuela de Matemáticas

"Introducción a la Teoría de Colas"

TESIS PROFESIONAL

que para obtener el título de:

MATEMATICO

presenta:

ERNESTO ARMANDO PACHECO VELAZQUEZ

Guadalajara, Jalisco 1986

INDICE

INTRODUCCION

CAPITULO I

I.1.-INTRODUCCION.-

En el mundo moderno, las filas de espera han pasado a ser parte de nuestro vivir, y en ocasiones, pasamos gran parte del dia en estas filas de espera.

A medida que la población aumenta, las pequeñas colonias van desarrollandose, y así se convierten en comunidades urbanas de muy diversos tamaños. Así pues, la vida de los hombres ha tenido que organizarse, pues depende en mucho de los servicios prestados por otros para satisfacer sus necesidades. Si se adoptara una filosofía de "servicio inmediato al ser solicitado", el resultado sería muy malo, pues no solo sería antieconómico, sino que también se desperdiciaría una gran cantidad de potencial humano (o maquinario), además de originar problemas muy difíciles de superar. Por lo tanto, las filas de espera o "colas" han pasado a ser parte de la vida cotidiana.

La teoría de las colas trata del estudio matemático de éstas filas de espera, mediante el estudio de esta materia es posible deducir datos como: promedio de longitud de las filas de espera, promedio de tiempo que una unidad (persona, máquina o suceso) debe esperar para ser atendido, número de elementos que se calculo en el sistema total, y otras muchas cosas.

Cabe mencionar que una fila de espera no necesariamente es formada por seres humanos, hay diversos y muy variados ejemplos de colas, por ejemplo: máquinas que "esperan" ser reparadas por un mecánico (en este caso las que forman la línea de espera son las máquinas), vehículos que esperan el "siga" de un semáforo, barcos que esperan cruzar cierto canal, ropa que espera el servicio de una tintorería, etc.

La teoría de la formación de colas tiene mucha importancia, y es necesario estudiarse, pues las filas de espera forman una molestia y desorganizan muchos aspectos de la vida. Los conocimientos que nos brinda la teoría de colas nos ayuda a modificar el proceso de decisiones con respecto a las filas de espera, y con ello mejorar las perspectivas de que la decisión sea acertada.

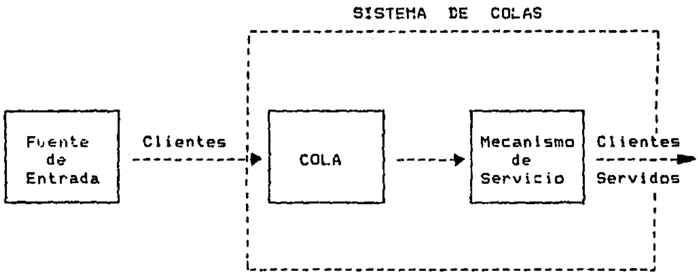
La teoría básica de las colas puede dividirse en dos estudios matemáticos, uno trata de distribuciones de tipo específico de los cuales se derivan fórmulas matemáticas, mientras que el otro trata de distribuciones clásicas, empíricas, o hipotetizadas que siempre se analizan mediante métodos de simulación. El primero de los dos enfoques exige comunmente que la entrada (o abastecimiento de población) siga una distribución de Poisson, y que la prestación de servicios siga una distribución exponencial. El segundo enfoque (simulación) permite un estudio mas amplio, aunque es un poco mas complicado desarrollar dicho estudio. Con ambos métodos puede estudiarse facilmente muchas combinaciones en la formación de colas.

I.2.- ESTRUCTURA BASICA DE LOS MODELOS DE COLAS.-

El proceso básico que se supone en la mayor parte de los modelos de colas es el siguiente: "Clientes" que requieren de cierto servicio, éstos clientes se generan en el tiempo mediante una "fuente de entrada". Los clientes entran al sistema de colas y se unen a una cola. En diversos momentos se selecciona a algún cliente formado para darle servicio. El "mecanismo de servicio" es lo que proporciona al cliente el servicio demandado, después de lo cual el cliente sale del sistema.

En la siguiente figura se muestra el esquema de este proceso:

Figura I.1



Fuente de Entrada

Uno de los datos que nos proporciona la fuente de entrada es el "tamaño", ésto quiere decir el número de clientes potenciales distintos que pueden requerir el servicio y puede suponerse finita o infinita.

Los cálculos para el caso finito son mas complicados, pues la entrada de un cliente al sistema afecta al número de clientes potenciales que se encuentran fuera del sistema, los cálculos para el caso infinito son mas fáciles en virtud de que la población potencial no es afectada por el número de clientes que se encuentran en el sistema, y a menudo pues, se hace la suposición de un tamaño infinito aun cuando el tamaño real sea cierto número finito muy grande (en adelante se considerará que el tamaño es infinito como una hipótesis implícita en cualquier modelo de colas, a menos que se especifique lo contrario).

El patrón estadístico mediante el cual se generan las entradas en el sistema es un proceso de Poisson generalmente, esto es, el número de clientes generados para cualquier instante tiene (en la gran mayoría de los casos) una distribución de Poisson.

También debe especificarse cualquier anomalia en el comportamiento de los clientes. Un ejemplo es la "contrariedad" que significa que el cliente rechaza entrar al sistema (y por lo tanto se pierde) si la cola es demasiado larga. Otro ejemplo es la "renuncia" que significa que el cliente pierde la paciencia y abandona la línea de espera.

Cola

La cola se caracteriza por el número máximo admisible de clientes que puede contener. Las colas son finitas o infinitas. La disciplina de una cola especifica el orden en que se selecciona a los clientes (en la cola) para recibir el servicio deseado. Por ejemplo, el que llega primero recibe primero el servicio (que es lo más común), aleatorio, procedimiento de prioridad, etc.

Mecanismo de Servicio

El mecanismo de servicio consiste en uno o mas medios de servicio los cuales contienen uno o mas canales paralelos de servicio llamados "servidores". Los sistemas de colas deben especificar la disposición de los medios y el número de servidores. Los modelos elementales establecen un medio de servicio con un número fijo de servidores.

El tiempo que transcurre desde que para un cliente se inicia el servicio hasta su terminación se denomina tiempo de servicio. Un modelo de colas debe especificar la distribución de probabilidad de los tiempos de servicio para cada servidor (y posiblemente para los diferentes tipos de clientes), aun cuando es más común [illegible] distribución para todos los servidores. La distribución mas frecuente es la distribución exponencial, aunque hay otras como la distribución de Poisson, la distribución de Erlang, y la distribución "degenerada" (tiempo de servicio constante).

I.3.- EJEMPLOS DE PROBLEMAS QUE PUEDEN SER RESUELTOS MEDIANTE LA TEORIA DE COLAS.-

Problema 1.-

Una pequeña compañía de cigarros tiene quince máquinas empacadoras de los cigarrillos que esta produce, el problema de esta pequeña compañía es que estas máquinas se descomponen muy frecuentemente, por tanto, esta compañía ha contratado a solo doce personas que pueden hacerse cargo de una de estas máquinas, de tal forma que se cuenta con tres máquinas de reserva para poder ser usadas cuando las otras sufren algun desperfecto. Así pues, la compañía trabaja con doce máquinas empacadoras, siempre y cuando no existan mas de tres máquinas que esten esperando ser reparadas, y el número de máquinas que trabajan se reduce en uno por cada máquina adicional que espere ser reparada.

Debemos decir además que la pequeña fabrica ha trabajado hasta hoy con un solo mecánico para reparar este tipo de máquinas, y los datos que la fábrica ha obtenido respecto a las máquinas es que el tiempo que transcurre hasta que cualquiera de las máquinas en trabajo se descompone sigue una distribución exponencial con una media de 45 dias, y el tiempo que transcurre desde que el mecánico empieza a trabajar en una máquina hasta terminar su reparación, tiene también una distribución exponencial con una media de 3 dias. Esto hace que en ocasiones la compañía trabaje con menos de 12 máquinas, y por tanto, la fábrica piensa en la contratación de un nuevo mecánico con el fin de poder reparar dos máquinas a la misma vez.

Cada reparador le cuesta a la compañía $ 7,000.00 diarios, pero la utilidad perdida estimada al dejar de trabajar una empacadora adicional a las tres de reserva es de $ 11,000.00 diarios. El problema consiste pues, en la toma de una decisión para la contratación de un nuevo mecánico.

Problema 2.-

Una oficina de gobierno ofrece cierto servicio para los trámites del Registro Federal de Causantes (R.F.C.). Debido al incremento de personas que solicitan hacer estos trámites, la oficina ha pensado en contratar a un muchacho para ayudar a la persona encargada del trabajo de revisar los datos, sellar y archivar la solicitud.

La oficina ha descubierto mediante algunos dias de prueba, que cuando la persona encargada de éstos trámites trabaja sola, los tiempos de servicio siguen una distribución exponencial con una media de 4 minutos por persona, mientras que cuando es ayudada por el muchacho, sigue también una distribución exponencial, pero con una media de 2.4 de minutos por persona.

La oficina ha descubierto también, mediante un estudio estadístico, que los solicitantes de los trámites del R.F.C., ocurren a la oficina siguiendo una distribución de Poisson con una media de 12 personas por hora (esto es, una persona por cada 5 minutos), durante todas las horas que se presta este servicio. Notemos pues, que la oficina no piensa en ofrecer un nuevo servidor, sino ofrece una mejora en el servicio, lo cual desde luego afectará el costo del servicio, pero en ambas situaciones al hacer el estudio nos debemos remitir al modelo básico que se estudiará en la sección IV.1.

Consideremos que en el caso de que siguiera trabajando tan solo una persona, la oficina gastaría en el costo del servicio $ 650.00, mientras que con el nuevo ayudante gastaría $ 1,000.00 (ambos gastos considerados por hora de servicio).

Además de ésto, la oficina ha hecho una encuesta con los solicitantes del R.F.C., y haciendo una estimación, los clientes han considerado en promedio, que lo que les cuesta a ellos esperar una hora en esa oficina para poder hacer los trámites, es $ 150.00. Además de ésto, se les pidio que estimaran el costo adicional que les provocaba la ruptura de la continuidad en sus labores, a lo que ellos respondieron que una ruptura de media hora les provocaba un costo de $ 30.00, mientras que si tuviesen que esperar una hora, el costo ascendía a $ 120.00

Con todo esto, y considerando que el jefe de esa oficina es una persona que quiere tomar una decisión justa, ¿se debe hacer una mejora en el servicio?, es decir, ¿se contratará a un ayudante para la persona encargada de los trámites del R.F.C.?

Problema 3

La Compañía "Maderas de México" esta planeando la construcción de una nueva sucursal, en el diseño de esta nueva sucursal se estima necesario incluir uno o mas depósitos de herramientas en el area de la carpintería para poder almacenar todos los utencilios que necesitan sus trabajadores. Las herramientas son manejadas por los empleados del almacen, y se les otorgan a los carpinteros cuando éstas se necesitan, pero cuando son acabadas de utilizar son devueltas al almacen.

En las otras tres sucursales que tiene esta compañía, los inspectores se han quejado de que los carpinteros desperdician mucho tiempo en ir a los depósitos y esperar ser servidos, así pues, solicitan más almacenes y mas personas que trabajen en estos. Pero los socios de esta compañía quieren reducir los gastos en la nueva sucursal. Todo esto ha creado situaciones muy conflictivas para el Presidente de "Maderas de México", y por tanto se esta haciendo un estudio en las otras sucursales tratando de obtener datos que ayuden en la toma de una decisión que pueda favorecer a la compañía.

Los datos que se obtienen son los siguientes: cada almacen forma un sistema de colas, ademas, el tiempo utilizado en atender a un carpintero tiene una distribución exponencial con media de un minuto. Los carpinteros ocurren al almacen de una manera aleatoria pero con una tasa media de llegadas de 60 carpinteros cada hora (bajo estos datos se decide usar el modelo básico que se estudiará en la sección IV.1). Además el costo total por cada almacenista es de $ 500.00 por hora, mientras que el valor del trabajo de un carpintero se calcula en $ 1,200.00 por hora. Con estos datos, y con otros que serán presentados en la sección VI.4, ¿cual sería la decisión del Presidente de la compañía?

Hasta aquí, pues, hemos visto una pequeña introducción a la teoría de colas, y hemos visto algunas de la aplicaciones que esta tiene, en el siguiente capítulo veremos la notación que se usará en el desarrollo de los modelos de colas.

NOTACION

CAPITULO II

II.1.- NOTACION.-

En este capítulo estudiaremos algunas de las princiaples variables que influyen en un sistema de colas, y algunas de las relaciones que existen entre ellas.

Definición.-

Llamaremos "cola" al conjunto de unidades (o clientes) que llegan y esperan ser servidas en las instalaciones que proporcionan el servicio que solicitan.

Definición.-

Llamaremos "estado del sistema" al número de clientes en el sistema de colas.

Definición.-

Llamaremos "longitud del sistema" al número de clientes que esperan recibir el servicio.

Nota: El estado del sistema y la longitud de la cola son diferentes. por ejemplo, supongamos que en una taquilla de boletos hay dos servidores y ocho clientes, por tanto dos clientes estan recibiendo el servicio, en este caso, el estado del sistema es de ocho clientes, mientras que la longitud de la cola es de seis clientes.

De aquí obtenemos que:

Longitud de la cola = Estado del sistema - Número de clientes recibiendo servicio.

Definicion.-

Llamaremos:

$N(t)$ a el número de clientes en el sistema de colas en el instante t ($t > 0$).

s a el número de servidores en el sistema de colas.

λ_n a la tasa de llegadas (número esperado de llegadas por unidad de tiempo) de clientes nuevos cuando se encuentran n clientes en el sistema.

μ_n a la tasa media de servicio en el sistema (número esperado de clientes a los que se les proporciona el servicio en unidad de tiempo) cuando se encuentran n clientes en el sistema.

Nota:

Si $\lambda_i = \lambda_j \quad \forall i,j \in \mathbb{N}$ entonces $\lambda_n = \lambda$

Si $\mu_i = \mu_j \quad \forall i,j \in \mathbb{N}$ entonces $\mu_n = \mu$

En estos casos $1/\lambda$ será el tiempo esperado de llegadas y $1/\mu$ será el tiempo esperado de servicio.

Si definimos a $\rho = \lambda / s\mu$, entonces ρ significará el tiempo esperado en que los servidores estarán brindando servicio por unidad de tiempo, esto es, la razón entre el número esperado de clientes y la capacidad de servicio de un negocio. Obviamente es de esperar que ρ sea menor que 1, pues de lo contrario, significará que el estado del sistema tiende a crecer continuamente en la medida que el tiempo transcurre, en este caso en que la cola crece sin medida decimos que la cola "estallaría".

II.2.- ESTADO ESTACIONARIO.-

Cuando un sistema de cola inicia su servicio (o empieza a funcionar), el estado del sistema será en gran parte afectado por el estado inicial y el tiempo transcurrido desde que comenzó el servicio, cuando esto sucede decimos que el estado se encuentra en una condición transitoria.

Despues de un buen tiempo de servicio, el estado del sistema se vuelve independiente del estado inicial y del tiempo transcurrido, en este caso podemos decir que el sistema ha alcanzado un estado estacionario.

En nuestro estudio de la Teoría de Colas consideraremos únicamente los sistemas dentro de un estado estacionario, pues el caso en el que el sistema esta en una condición transitoria es mucho mas difícil de considerar desde el punto de vista analítico.

La siguiente notación supondra el sistema dentro de un estado estacionario:

n significará el número de clientes en el sistema de colas.

P significará la probabilidad de que se encuentren exactamente n clientes en el sistema de colas o que el estado del sistema sea n.

L significará el número esperado de clientes en el sistema de colas.

Lq significará la longitud esperada de la cola.

$\mathcal{W}$ significará el tiempo de espera en el sistema de colas para cada cliente por separado (incluye tiempo de servicio).

W significará el valor esperado de $(E(\mathcal{W}))$.

$\mathcal{W}_q$ significará el tiempo de espera en la cola para cada cliente por separado (excluye tiempo de servicio).

Wq significará el valor esperado de $(E(\mathcal{W}_q))$.

II.3.- RELACIONES ENTRE L, W, Lq, Wq.-

Supongamos que $\lambda_n = \lambda \quad \forall n \in \mathbb{N}$. Se ha demostrado (1) que en un sistema de colas que se encuentra dentro de un estado estacionario: $L = \lambda W$ (Esto es, el número esperado de clientes es igual a la tasa media de llegadas por la media de el tiempo de espera en el sistema).

Además, en esta misma demostración se prueba también que: $Lq = \lambda Wq$.

Nota: Si las λ_n no son todas iguales, entonces λ será reemplazada por $\bar{\lambda}$ que significará la tasa promedio de llegadas.

(1) Ver el libro Stidham S., (1980), "A last word on $L = \lambda W$" páginas 170-183, Mc Graw Hill.

Si suponemos que $\mu_n = \mu \; \forall \, n \in N$, significará que el tiempo medio de servicio es una constante $(1/\mu)$. De aquí puede deducirse que:

$$W = Wq + 1/\mu$$

Estas relaciones seran muy importantes ya que encontrando el valor de una de ellas, es fácil deducir los valores de las otras variables, y esto nos representa una gran ventaja, pues si éstas relaciones no existieran, nos veríamos frecuentemente en graves problemas para determinar ciertos valores que nos serán indispensables para la toma de una decisión, como pueden ser el valor esperado de el tiempo de espera, el costo de la espera, el costo del servicio, etc.

Después de haber visto en este capítulo la notación que usaremos, analizaremos en el siguiente capítulo un proceso llamado "Nacimiento-Muerte" que será de gran importancia en el desarrollo de nuestro estudio.

EL PROCESO

"NACIMIENTO

MUERTE"

CAPITULO III

III.1.- EL PROCESO NACIMIENTO MUERTE.-

Una gran parte de los modelos de los sistemas de colas ocurren basados en un proceso llamado "Proceso Nacimiento-Muerte".

Con el término nacimiento entenderemos la llegada de un cliente nuevo que se incorpora al sistema de colas, y por el término muerte entenderemos la salida del sistema de un cliente que ya ha recibido el servicio.

Las hipótesis necesarias para que un proceso se considere como un proceso Nacimiento-Muerte, son las siguientes:

a) Si $N(t) = n$, la distribución de probabilidad de tiempo hasta el siguiente nacimiento es exponencial con parámetro λ_n .

b) Si $N(t) = n$, la distribución de probabilidad de tiempo hasta la siguiente muerte es exponencial con parámetro μ_n .

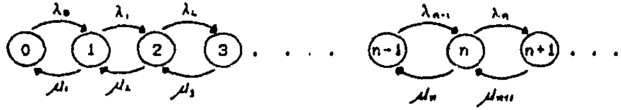
c) En un instante cualquiera solo puede ocurrir un nacimiento o una muerte.

El siguiente diagrama nos ilustra las tasas de entrada y de salida en el proceso Nacimiento Muerte.

Figura III.1.-

Diagrama de tasas en el proceso nacimiento-muerte.

Estado



III.2.- EL PRINCIPIO "TASA DE ENTRADA = TASA DE SALIDA".-

Analicemos ahora el siguiente proceso en un sistema de colas. Supongamos que una librería pequeña con un solo servidor empieza a brindar servicio a las 9:00 A.M., y que la entrada y salida de clientes ocurre de la siguiente manera:

9:00 No hay clientes esperando (el estado del sistema es 0)

9:03 Llega un cliente.

9:04 Llega un cliente.

9:08 Sale un cliente "servido".

9:09 Llega un cliente.

9:11 Sale un cliente "servido".

9:12 Llega un cliente.

9:13 Llega un cliente.

9:14 Sale un cliente "servido".

9:17 Sale un cliente "servido".

9:23 Sale un cliente "servido".

9:23 Llega un cliente.

9:25 Llega un cliente.

9:30 Sale un cliente "servido".

....

Por tanto los cambios en el estado del sistema se han realizado de la siguiente manera:

Sale del estado 0 e ingresa al estado 1

Sale del estado 1 e ingresa al estado 2

Sale del estado 2 e ingresa al estado 1

Sale del estado 1 e ingresa al estado 2

Sale del estado 2 e ingresa al estado 1

Sale del estado 1 e ingresa al estado 2

Sale del estado 2 e ingresa al estado 3

Sale del estado 3 e ingresa al estado 2

Sale del estado 2 e ingresa al estado 1

Sale del estado 1 e ingresa al estado 0

Sale del estado 0 e ingresa al estado 1

Sale del estado 1 e ingresa al estado 2

Sale del estado 2 e ingresa al estado 1

....

Así pues nos damos cuenta de que el número de veces que el proceso entra y el número de veces que el proceso sale de un estado determinado es el mismo, o se mira afectado por tan solo uña unidad. Pero esta diferencia de uno, al final se puede considerar como despreciable en las tasas promedio (número total de ocurrencias por unidad de tiempo), por tanto, a "largo plazo", estas tasas de entrada y de salida deben de ser iguales.

Esto nos conduce a la siguiente conclusión:

Principio de "La tasa de entrada es igual a la tasa de salida".-

Para cualquier estado n del sistema, la tasa media de entrada es igual a la tasa media de salida.

Analizando un poco el sistema en el estado 0, nos daremos cuenta que el proceso entra a este estado únicamente desde es estado 1, la tasa media para entrar al estado 0 es μ_1. Pero la probabilidad de que el proceso efectivamente se encuentre en el estado 1 es P_1. Por lo tanto, la tasa de entrada al estado 0 es "claramente" $\mu_1 P_1$.

Siguiendo el mismo razonamiento, obtendremos que la tasa de salida del estado 0 es $\lambda_0 P_0$, y con el principio anterior, obtenemos que: $\mu_1 P_1 = \lambda_0 P_0$.

En los demás estados, las transiciones posibles son dos (tanto de entrada como de salida).

De aqui podemos resumir nuestros resultados de la siguiente forma:

Estado	Tasa de entrada = Tasa de salida
0	$\mu_1 P_1 = \lambda_0 P_0$
1	$\lambda_0 P_0 + \mu_2 P_2 = (\lambda_1 + \mu_1) P_1$
2	$\lambda_1 P_1 + \mu_3 P_3 = (\lambda_2 + \mu_2) P_2$
...	....
n-1	$\lambda_{n-2} P_{n-2} + \mu_n P_n = (\lambda_{n-1} + \mu_{n-1}) P_{n-1}$
n	$\lambda_{n-1} P_{n-1} + \mu_{n+1} P_{n+1} = (\lambda_n + \mu_n) P_n$
n+1	$\lambda_n P_n + \mu_{n+2} P_{n+2} = (\lambda_{n+1} + \mu_{n+1}) P_{n+1}$

Aplicando esto podemos llegar a los siguientes resultados:

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

$$P_2 = \lambda_1/\mu_2 \; P_1 + 1/\mu_2 \; (\mu_1 P_1 - \lambda_0 P_0) = \lambda_1/\mu_2 \; P_1 = \lambda_1 \lambda_0/\mu_2 \mu_1 \quad P_0$$

$$P_3 = \lambda_2/\mu_3 \; P_2 + 1/\mu_3 \; (\mu_2 P_2 - \lambda_1 P_1) = \lambda_2/\mu_3 \; P_2 = \lambda_2 \lambda_1 \lambda_0/\mu_3 \mu_2 \mu_1 \quad P_0$$

....

$$P_n = \lambda_{n-1}/\mu_n \; P_{n-1} + 1/\mu_n \; (\mu_{n-1} P_{n-1} - \lambda_{n-2} P_{n-2}) = \lambda_{n-1}/\mu_n \; P_{n-1} = \lambda_{n-1} \cdots \lambda_1 \lambda_0/\mu_n \cdots \mu_2 \mu_1 \; P_0$$

Simplificando la notación hagamos

$$C_n = \lambda_{n-1} \cdots \lambda_1 \lambda_0/\mu_n \cdots \mu_1 \mu_0$$

Por lo tanto $P_n = C_n \; P_0$

Con estos resultados es fácil verificar el siguiente teorema:

Teorema IV.1

En un sistema de colas, la probabilidad de que el sistema se encuentre en el estado 0 es:

$$P_0 = \frac{1}{1 + \sum_{n=1}^{\infty} C_n}$$

Demostración.-

Tenemos que $\sum_{n=0}^{\infty} P_n = 1 \implies P_0 + \sum_{n=1}^{\infty} P_n = 1$

$$== \quad P_0 + P_0 \sum_{n=1}^{\infty} C_n = 1$$

$$== \quad P \; (1 + \sum_{n=1}^{\infty} C_n) = 1$$

$$== \quad P = 1/(1 + \sum_{n=1}^{\infty} C_n)$$

Además, "claramente" podemos ver que el numero esperado de clientes en el sistema es:

$$L = \sum_{n=0}^{\infty} n\, P_n$$

Y si existen s servidores, entonces s representa el número de clientes que están siendo atendidos (y por lo tanto eliminados de la cola), por lo que:

$$Lq = \sum_{n=s}^{\infty} (n-s)\, P_n$$

Y por las relaciones dadas en el capítulo anterior, obtenemos que:

$$W = L/\bar{\lambda} \qquad\qquad y \qquad\qquad Wq = Lq/\bar{\lambda}$$

donde $\bar{\lambda}$ es la tasa promedio de llegadas, esto es, $\bar{\lambda} = \sum_{n=0}^{\infty} P_n \lambda_n$

Una última cosa que quisiera hacer notar, es que debido a que las tasas medias λ_0, λ_1, . . . , λ_n, . . y μ_1, μ_2, . . , μ_n, . . . pueden tomar cualquier valor no negativo, esto nos ofrece una gran flexibilidad en los sistemas de filas de espera, con lo que los modelos basados en este proceso (y que estudiaremos en el siguiente capítulo), son los mas usados comunmente. Además, debido a la hipótesis a) y b) que hemos visto en la sección III.1, se dice que estos modelos tienen un entrada de Poisson y tiempos de servicios exponenciales.

MODELOS BASICOS

DE

COLAS

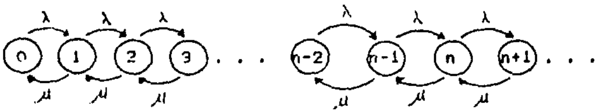
CAPITULO IV

IV.1 EL MODELO BASICO.-

Es muy común encontrar en los sistemas de colas que las tasas medias de llegadas y de salida permanescan constantes (esto es, $\lambda_n = \lambda$ y $\mu_n = \mu$ $\forall n \in \mathbb{N}$), por tanto, cuando el sistema tiene un solo servidor, el diagrama de las tasas resultantes se desarrolla como se muestra en la figura IV.1.

Figura IV.1.-

Estado



En este modelo es de suponerse que la tasa media de llegadas es menor que la tasa media de salidas, ya que de otro modo la cola tendería a crecer sin medida, o dicho de otro modo, la cola "estallaría".

Basados en estos fundamentos, podemos deducir facilmente los siguientes teoremas:

Teorema IV.1.-

Si la tasa media de llegadas y la tasa media de servicio son constantes, λ y μ respectivamente, y si s=1, entonces:

$P = (1 - \rho)$ donde $\rho = \lambda/\mu$

Demostración.-

Tenemos que si $\lambda_n = \lambda$ y $\mu_n = \mu$, entonces tendremos que $C_n = (\lambda/\mu)^n$, por tanto, si hacemos $\rho = \lambda/\mu$ vamos a obtener lo siguiente:

$$P_0 = 1/(1+\sum_{n=1}^{\infty} C_n)$$

$$\Rightarrow \quad P_0 = 1/(1+\sum_{n=1}^{\infty} \rho^n)$$

$$\Rightarrow \quad P_0 = 1/(\sum_{n=0}^{\infty} \rho^n)$$

$$\Rightarrow \quad P_0 = (\sum_{n=0}^{\infty} \rho^n)^{-1}$$

$$\Rightarrow \quad P_0 = [1/(1-\rho)]^{-1}$$

$$P_0 = 1 - \rho \qquad \text{donde } \rho = \lambda/\mu$$

Teorema IV.2.-

Con las condiciones del teorema anterior, $L = \lambda/(\mu - \lambda)$

Demostración.-

Tenemos que $L = \sum_{n=0}^{\infty} n P_n \Rightarrow L = \sum_{n=0}^{\infty} n(1-\rho)\rho^n$

$$\Rightarrow \quad L = (1-\rho)\rho \sum_{n=1}^{\infty} n \rho^{n-1}$$

$$\Rightarrow \quad L = (1-\rho)\rho \sum_{n=0}^{\infty} \frac{d(\rho^n)}{d\rho}$$

$$\Rightarrow \quad L = (1-\rho)\rho \frac{d}{d\rho}\left(\sum_{n=0}^{\infty} \rho^n\right)$$

$$\Rightarrow \quad L = (1-\rho)\rho \frac{d}{d\rho}\left(1/(1-\rho)\right)$$

$$\Rightarrow \quad L = (1-\rho)\rho(1-\rho)^{-2}$$

$$\Rightarrow \quad L = \rho/(1-\rho)$$

$$\Rightarrow \quad L = \lambda/(\mu-\lambda)$$

Teorema IV.3.-

Con las condiciones del teorema IV,1 tenemos que:

$$Lq = \lambda^2 / [\mu (\mu - \lambda)]$$

Demostración.-

Tenemos que $Lq = \sum_{n=1}^{\infty} (n-1) P_n \implies Lq = \sum_{n=0}^{\infty} nP_n - \sum_{n=1}^{\infty} P_n$

$$\implies Lq = L - (1 - P_0)$$

$$\implies Lq = L - \rho$$

$$\implies Lq = \lambda / (\mu - \lambda) - \mu / \lambda$$

$$\implies Lq = \lambda^2 / [\mu (\mu - \lambda)]$$

Corolario IV.1.-

a) $Wq = \lambda / [\mu (\mu - \lambda)]$

b) $W = 1 / (\mu - \lambda)$

Demostracion.-

a) Tenemos que $Wq = Lq/\lambda \implies Wq = (1/\lambda) \; [\lambda^2 / \mu (\mu - \lambda)]$

$$\implies Wq = \lambda / [\mu (\mu - \lambda)]$$

b) Tenemos que $W = L/\lambda \implies W = [\lambda / (\mu - \lambda)] \; (1/\lambda)$

$$\implies W = 1/(\mu - \lambda)$$

Los cálculos para obtener W y W_q son un poco mas complicados, para obtener la distribución de probabilidad del tiempo de espera en el sistema (incluyendo el servicio) W, para una llegada aleatoria, supongamos que el cliente que llega encuentra n clientes en el sistema, por tanto, tendrá que esperar $(n+1)$ tiempos de servicios exponenciales. Si R_1, R_2, . . . , R_n, R_{n+1}, son variables aleatorias independientes de los tiempos de servicio, y estos tienen una distribución exponencial, con parámetro μ, y definimos $Y_{n+1} = R_1 + R_2 + \ . \ \ . \ \ . \ + R_n + R_{n+1}$, entonces, Y_{n+1} representa el tiempo de espera condicional cuando se tienen n clientes en el sistema. Además, Y_{n+1} tiene una distribución gamma, por tanto se puede concluir que:

$$P(\mathcal{W} > t) = \sum_{n=0}^{\infty} P_n \ P(Y_{n+1} > t)$$

Después de algunos pasos algebraicos, esto lo podemos reducir a (2):

$$P(\mathcal{W} > t) = e^{-\mu(1-\rho)t}$$

Por tanto, $\mathcal{W}$ tiene una distribución exponencial con parámetro $\mu(1 - \rho)$.

Además, $P(\mathcal{W}_q > t) = \sum_{n=1}^{\infty} P_n \ P(Y_n > t)$

$$\Longrightarrow \quad P(\mathcal{W}_q > t) = \sum_{n=1}^{\infty} (1-\rho)\rho^n \ P(Y_n > t)$$

$$\Longrightarrow \quad P(\mathcal{W}_q > t) = \rho \sum_{n=0}^{\infty} P_n \ P(Y_{n+1} > t)$$

$$\Longrightarrow \quad P(\mathcal{W}_q > t) = \rho \ P(\mathcal{W} > t)$$

$$\Longrightarrow \quad P(\mathcal{W}_q > t) = \rho \ e^{-\mu(1-\rho)t}$$

Por último, también podemos decir que

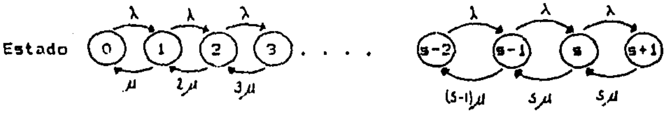
$$P(\mathcal{W}_q = 0) = P_0 = 1 - \rho$$

(2) Harris W., (1981), "Introducción a la Teoría de Colas", Mc Graw Hill, pp. 153-158

IV.2.- EL MODELO BASICO CON VARIOS SERVIDORES.-

Cuando el modelo de colas tiene más de un servidor, el diagrama de tasas es como en el modelo siguiente:

Figura IV.2.-



Además, podemos deducir facilmente los siguientes teoremas:

Teorema IV.4.-

Si la tasa media de llegadas y la tasa media de servicio son constantes, λ y μ respectivamente, y si el número de servidores es s, entonces:

$$P = \frac{1}{\sum_{n=0}^{s-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!} \frac{1}{1-\lambda/s\mu}}$$

Demostración.-

Si el número de servidores es s, entonces

$$C_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!} & \text{si } n < s \\[2ex] \dfrac{(\lambda/\mu)^s}{s!} \left(\dfrac{\lambda}{\mu s}\right)^{n-s} = \dfrac{(\lambda/\mu)^n}{s!\, s^{n-s}} & \text{si } n > s \end{cases}$$

Por lo tanto,

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

y si $\lambda < s\mu$

$$P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!} \frac{1}{(1-\lambda/s\mu)}}$$

Teorema IV.5.-

Con las mismas condiciones que el teorema anterior

$$P_n = \begin{cases} \frac{(\lambda/\mu)^n}{n!} P_0 & \text{si } 0 \leqslant n \leqslant s \\ \frac{(\lambda/\mu)^n}{s!\ s^{n-s}} P_0 & \text{si } n \geqslant s \end{cases}$$

Demostración.-

La demostración es trivial, pues $P_n = C_n P_0$

Teorema IV.6.-

Con las condiciones del teorema IV.4,

$$Lq = \frac{P_0 (\lambda/\mu)^s \rho}{s!\ (1-\rho)^2} \qquad \text{donde} \quad \rho = \lambda/s\mu$$

Demostración.-

Definamos a ρ como $\rho = \lambda/s\mu$

Además $Lq = \sum_{n=s}^{\infty} (n-s)\ P_n$

$$\Longrightarrow \quad Lq = \sum_{i=0}^{\infty} i\, P_{i+s}$$

$$\Longrightarrow \quad Lq = \sum_{i=0}^{\infty} i\, \frac{(\lambda/\mu)^{i+s}}{s!\ s^{i+s-s}} P_0$$

$$\Longrightarrow \quad Lq = \sum_{i=0}^{\infty} i\, \frac{(\lambda/\mu)^s}{s!} \rho^i P_0$$

$$\Longrightarrow \quad Lq = P_0 \frac{(\lambda/\mu)^s}{s!} \rho \sum_{i=0}^{\infty} i\, \rho^{i-1}$$

$$===> \quad Lq = P_0 \frac{(\lambda/\mu)^s}{s!} \rho \sum_{i=0}^{\infty} \frac{d}{d\rho}(\rho^i)$$

$$===> \quad Lq = P_0 \frac{(\lambda/\mu)^s}{s!} \rho \frac{d}{d\rho}\left(\sum_{i=0}^{\infty} \rho^i\right)$$

$$===> \quad Lq = P_0 \frac{(\lambda/\mu)^s}{s!} \rho \frac{d}{d\rho}\left(\frac{1}{1-\rho}\right)$$

$$\therefore \quad Lq = \frac{P_0 (\lambda/\mu)^s \rho}{s! (1-\rho)^2}$$

Corolario IV.2.-

a) $Wq = \dfrac{P_0 (\lambda/\mu)^s \rho}{\lambda \; s! \; (1-\rho)^2}$

b) $W = \dfrac{P_0 (\lambda/\mu)^s \rho}{\lambda \; s! \; (1-\rho)^2} + \dfrac{1}{\mu}$

c) $L = \dfrac{P_0 (\lambda/\mu)^s \rho}{s! \; (1-\rho)^2} + \dfrac{\lambda}{\mu}$

Demostración.-

a) Tenemos que $Wq = Lq/\lambda$, por tanto, la demostración es trivial.

b) Si consideramos que $W = Wq + 1/\mu$, llegamos al resultado.

c) También tenemos que $L = \lambda W = \lambda (Wq + 1/\mu) = Lq + \lambda/\mu$ y con esto concluimos la demostración.

El metodo para encontrar la distribución de probabilidad para los tiempos de espera es similar que en el caso de un solo servidor.

El resultado en este caso es el siguiente (3):

$$P(W > t) = e^{-\mu t}\left[1 + \frac{P_0 (\lambda/\mu)^s}{s! (1-\rho)} \cdot \frac{1 - e^{-\mu t(s-1-\lambda/\mu)}}{s - 1 - \lambda/\mu}\right]$$

$$P(Wq > t) = [1 - P(Wq = 0)] \, e^{-s\mu(1-\rho)t} \quad \text{donde} \quad P(Wq = 0) = \sum_{n=0}^{s-1} P_n$$

(3) Si $s-1-\lambda/\mu = 0$, entonces $P(W > t) = e^{-\mu t}\left[1 + \dfrac{P_0 (\lambda/\mu)^s}{s! (1-\rho)} \mu t\right]$

IV.3.- MODELOS BASICOS CON COLAS LIMITADAS.-

En el análisis que hemos hecho en la sección anterior, hemos visto colas que pueden crecer (por así decirlo) ilimitadamente, aunque la probabilidad de que lleguen a un número M "muy grande" es muy pequeña.

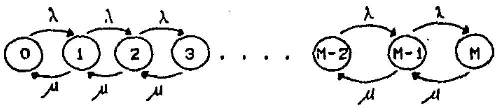
En esta sección estudiaremos colas que aun teniendo una fuente de entrada infinita, sea una "cola limitada". Esto significa que el número de clientes en el sistema nunca será mayor que cierto número M especificado.

Los casos en que esta condición se cumple son muy variados, puede ser que los clientes que lleguen no entren al sistema y "lleven sus negocios a otras partes" cuando observan que en la cola hay demasiados clientes (M), que el "negocio" no permita la entrada al sistema cuando el estado del sistema es muy grande (M), que el "negocio" no pueda tener mas de algún número muy grande (M) de clientes en la cola, etc.

Por ejemplo, supongamos que un camión urbano recorre la Ciudad de México, que los clientes son los pasajeros, los servidores son los asientos y que el servicio que los clientes solicitan es el poderse sentar. Si el camión tiene 30 asientos (excluyendo el asiento del chofer del camión), a treinta personas a la vez se le estará prestando el servicio, sin embargo, la cola será limitada, pues la cola la forman las personas que están en el camión y viajan paradas, y aunque continuamente suban y bajan personas, y aunque la fuente de entrada es "ilimitada" (en este caso es una cantidad muy grande), podemos estar seguros de que el estado del sistema nunca será mayor de 300 clientes, así pues, aunque la fuente de entrada tiene una población "infinita", la cola será limitada.

Cuando el número de servidores es uno, el diagrama de tasas es similar al que se muestra en la figura IV.1, excepto que el modelo se suspende en el estado M, esto es:

Figura IV.3.-

Estado 

De aquí, pues, resulta fácil obtener los siguientes teoremas:

Teorema IV.7.-

Cuando la cola es limitada, y la tasa media de llegadas y de salidas se mantienen constantes hasta que la cola alcanza el estado M (donde M es la limitación de la cola), y s=1, entonces:

$$P_0 = \frac{1 - \rho}{1 - \rho^{M+1}}$$

Demostración.-

Tenemos que

$$C = \begin{cases} (\lambda/\mu)^n & n \leq M \\ 0 & n > M \end{cases}$$

entonces

$$P_0 = \frac{1}{\sum_{n=1}^{M} (\lambda/\mu)^n} \implies P_0 = \frac{1}{\left[1-(\lambda/\mu)^{M+1}\right]/1-\lambda/\mu}$$

$$\implies P_0 = \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{M+1}}$$

$$\implies P_0 = \frac{1-(\lambda/\mu)}{1-(\lambda/\mu)^{M+1}}$$

y si definimos $\rho = \lambda/\mu$ $\implies P_0 = \dfrac{1-\rho}{1-\rho^{M+1}}$

Teorema IV.8.-

Con las condiciones del teorema anterior,

$$P_n = \frac{1-\rho}{1-\rho^{M+1}} \rho^n \qquad \text{si } 0 \leq n \leq M$$

Demostración.-

$$P_n = C_n P_0 = (\lambda/\mu)^n \frac{1-\rho}{1-\rho^{M+1}} \implies P_n = \frac{1-\rho}{1-\rho^{M+1}} \rho^n$$

Teorema IV.9.-

Con las mismas condiciones del teorema anterior,

$$L = \frac{\rho}{1-\rho} - \frac{(M+1)\rho^{M+1}}{1-\rho^{M+1}}$$

Demostración.-

$$\text{Tenemos que} \quad L = \sum_{n=0}^{M} n\, P_n \quad \Longrightarrow \quad L = \sum_{n=0}^{M} n \left(\frac{1-\rho}{1-\rho^{M+1}} \right) \rho^n$$

$$\Longrightarrow \quad L = \frac{1-\rho}{1-\rho^{M+1}}\, \rho \sum_{n=0}^{M} n\, \rho^{n-1}$$

$$\Longrightarrow \quad L = \frac{1-\rho}{1-\rho^{M+1}}\, \rho \sum_{n=0}^{M} \frac{d}{d\rho}\, \rho^n$$

$$\Longrightarrow \quad L = \frac{1-\rho}{1-\rho^{M+1}}\, \rho\, \frac{d}{d\rho} \sum_{n=0}^{M} \rho^n$$

$$\Longrightarrow \quad L = \frac{1-\rho}{1-\rho^{M+1}}\, \rho\, \frac{d}{d\rho} \left(\frac{1-\rho^{M+1}}{1-\rho} \right)$$

$$\Longrightarrow \quad L = \frac{1-\rho}{1-\rho^{M+1}}\, \rho\, \frac{-(M+1)\,\rho^M\,(1-\rho) + 1 - \rho^{M+1}}{(1-\rho)^2}$$

$$\Longrightarrow \quad L = \frac{1-\rho}{1-\rho^{M+1}}\, \rho\, \frac{-(M+1)\,\rho^M + (M+1)\,\rho^{M+1} + 1 - \rho^{M+1}}{(1-\rho)^2}$$

$$\Longrightarrow \quad L = \rho\, \frac{-(M+1)\,\rho^M + M\,\rho^{M+1} + 1}{(1-\rho^{M+1})\,(1-\rho)}$$

$$\Longrightarrow \quad L = \rho\, \frac{(M+1)(\rho^{M+1} - \rho^M) + 1 - \rho^{M+1}}{(1-\rho^{M+1})\,(1-\rho)}$$

$$\Longrightarrow \quad L = \rho\, \frac{(M+1)\,\rho^M\,(\rho - 1)}{(1-\rho)(1-\rho^{M+1})} + \rho\, \frac{(1-\rho)^{M+1}}{(1-\rho)(1-\rho^{M+1})}$$

$$\Longrightarrow \quad L = \frac{\rho}{1-\rho} - \frac{(M+1)\,\rho^{M+1}}{1-\rho^{M+1}}$$

Corolario IV.3.-

$$Lq = L - (1 - P_0)$$

Demostración.-

Trivial.-

Como las λ_n no son iguales, entonces $L \neq \lambda W$ y $Lq \neq \lambda Wq$ pero podemos obtener los tiempos esperados de permanencia de los clientes que entran al sistema por las fórmulas:

$$W = L/\bar{\lambda} \qquad\qquad Wq = Lq/\bar{\lambda}$$

donde $\bar{\lambda} = \sum_{n=0}^{\infty} \lambda_n P_n = \sum_{n=0}^{M-1} \lambda_n P_n = \lambda(1 - P_M)$

Nota: Observese que estos resultados no requieren que $\lambda < \mu$

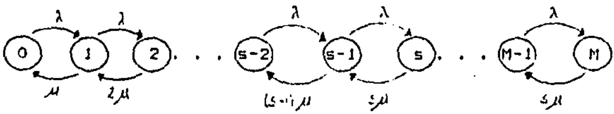
IV.4.- EL MODELO DE COLAS LIMITADAS CON VARIOS SERVIDORES.-

Como en este modelo no se permiten mas de M clientes en el sistema, es lógico pensar que el número máximo de servidores es precisamente M, es decir, $s \leq M$.

El diagrama de tasas para este modelo, es el siguiente:

Figura IV.4.-

Estado



Por tanto, podemos obtener los siguientes teoremas:

Teorema IV.10.-

En el modelo de colas limitadas con s servidores,

$$P_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!} P_0 & n \leqslant s \\ \dfrac{(\lambda/\mu)^n}{s!\, s^{n-s}} P_0 & s \leqslant n \leqslant M \\ 0 & n > M \end{cases}$$

Demostración.-

$$\text{Tenemos que } C_n = \begin{cases} (\lambda/\mu)^n/n! & n \leqslant s \\ \dfrac{(\lambda/\mu)^s}{s!} (\lambda/s\mu)^{n-s} & s \leqslant n \leqslant M \\ 0 & n > M \end{cases}$$

De aqui obtenemos que $P_0 = \dfrac{1}{\sum\limits_{n=0}^{s} (\lambda/\mu)^n/n! + \sum\limits_{n=s+1}^{M} \left(\frac{\lambda}{s\mu}\right)^{n-s}}$

$$\text{Y por lo tanto } P_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!} P_0 & n < s \\ \dfrac{(\lambda/\mu)^n}{s!\, s^{n-s}} P_0 & s < n < M \\ 0 & n > M \end{cases}$$

Teorema IV.11.-

Con las condiciones del teorema anterior

$$Lq = \frac{P_0 (\lambda/\mu)^s \rho}{s!\, (1-\rho)^2} \left[1 - \rho^{M-s} - (M-s)\, \rho^{M-s} (1-\rho)\right]$$

Demostración.-

Tenemos que $Lq = \sum\limits_{n=s}^{\infty} (n-s)\, P_n \implies Lq = \sum\limits_{n=s}^{M} (n-s)\, P_n$

$\implies Lq = \sum\limits_{n=s}^{M} (n-s)\, P_n$

$$\Longrightarrow \quad L_q = \sum_{j=0}^{M-S} j\, P_{j+s}$$

$$\Longrightarrow \quad L_q = \sum_{j=0}^{M-S} j\, \frac{(\lambda/\mu)^{j+s}}{s!\, s^j}\, P_0$$

$$\Longrightarrow \quad L_q = \sum_{j=0}^{M-S} \frac{(\lambda/\mu)^s}{s!}\, j \left(\lambda/s\mu\right)^j P_0$$

$$\Longrightarrow \quad L_q = \frac{(\lambda/\mu)^s}{s!} P_0 \left(\lambda/s\mu \right) \sum_{j=0}^{M-S} j \left(\lambda/s\mu\right)^{j-1}$$

Hagamos ahora $\rho = \lambda / s\mu$

$$\Longrightarrow \quad L_q = \frac{(\lambda/\mu)^s}{s!} P_0\, \rho \sum_{j=0}^{M-S} \frac{d}{d\rho} (\rho^j)$$

$$\Longrightarrow \quad L_q = \frac{P_0 (\lambda/\mu)^s}{s!} \rho\, \frac{d}{d\rho} \sum_{j=0}^{M-S} \rho^j$$

$$\Longrightarrow \quad L_q = \frac{P_0 (\lambda/\mu)^s \rho}{s!}\, \frac{d}{d\rho} \left(\frac{1-\rho^{M-S+1}}{1-\rho} \right)$$

$$\Longrightarrow \quad L_q = \frac{P_0 (\lambda/\mu)^s \rho}{s!} \left[\frac{-(1-\rho)(M-S+1)\,\rho^{M-S} + 1 - \rho^{M-S+1}}{(1-\rho)^2} \right]$$

$$\Longrightarrow \quad L_q = \frac{P_0 (\lambda/\mu)^s \rho}{s!\,(1-\rho)^2} \left(1 - \rho^{M-S+1} - (M-S+1)\,\rho^{M-S} + (M-S+1)\,\rho^{M-S+1} \right)$$

$$\Longrightarrow \quad L_q = \frac{P_0 (\lambda/\mu)^s \rho}{s!\,(1-\rho)^2} \left[1 - \rho^{M-S} - (M-S)\,\rho^{M-S} (1-\rho) \right]$$

Teorema IV.12.-

Con las mismas condiciones que el teorema anterior,

$$L = \sum_{n=0}^{s-1} n\, P_n + L_q + s \left(1 - \sum_{n=0}^{s-1} P_n \right)$$

Demostracion.-

Tenemos que $L = \sum_{n=0}^{M} n\, P_n \quad \Longrightarrow L = \sum_{n=0}^{M} n\, P_n$

$$\Longrightarrow \quad L = \sum_{n=0}^{s-1} n\, P_n + \sum_{n=s}^{M} n\, P_n - \sum_{n=s}^{M} s\, P_n + \sum_{n=s}^{M} s\, P_n$$

$$\Longrightarrow \quad L = \sum_{n=0}^{s-1} n\, P_n + \sum_{n=s}^{M} (n-s)\, P + \sum_{n=s}^{M} s\, P_n$$

$$\Longrightarrow \quad L = \sum_{n=0}^{s-1} n\, P_n + L_q + s \left(1 - \sum_{n=0}^{s-1} P_n \right)$$

Al igual que en el modelo anterior, en este modelo podemos obtener W y Wq como

$$W = L/\bar{\lambda} \qquad\qquad Wq = Lq/\bar{\lambda}$$

donde $$\bar{\lambda} = \sum_{n=0}^{\infty} \lambda_n P_n = \sum_{n=0}^{M-1} \lambda_n P_n = \lambda (1 - P_M)$$

IV.5.- EL MODELO BASICO CON UNA FUENTE DE ENTRADA LIMITADA.-

En estos casos, se supone que la fuente de entrada es limitada, esto es, que tiene una población finita de donde abastecerse. Este caso se da sobre todo en fábricas, donde un mecánico tiene la responsabilidad de mantener en buen estado cierto número de máquinas, o en casos similares, tiene tanto auge en este tipo de problemas que a veces recibe el nombre de "modelo de colas de reparación de máquinas".

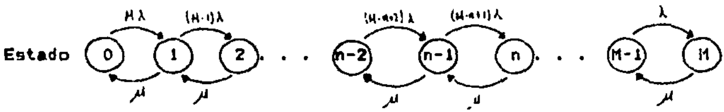
En este modelo llamaremos "M" a el tamaño de la población potencial, y es obvio que cuando un número n de clientes estén recibiendo el servicio en el sistema ($0 \leq n \leq M$), solo quedarán M-n clientes potenciales.

La población potencial alterna entre estar "fuera" o "dentro" del sistema de colas. Supondremos que el tiempo que transcurre desde que un miembro sale del sistema hasta su regreso a él tiene una distribución exponencial con parámetro λ.

Cuando n miembros están "dentro", (M-n) miembros se encontrarán "fuera", entonces la distribución de probabilidad del tiempo restante hasta la siguiente llegada al sistema seguirá una distribución exponencial con parámetro $\lambda_n = (M-n)\lambda$.

El siguiente diagrama nos ilustra un poco este modelo:

Figura IV.5.-



Nota: Observese que $\lambda_n = 0$ si $n \geq M$

De aqui podemos deducir los siguiente teoremas:

Teorema IV.13.-

Si la fuente de entrada es limitada en el modelo de colas básico, entonces:

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

Demostración.-

En este modelo observamos que:

$$C_n = \begin{cases} M\,(M-1)\,(M-2)\,.\,.\,.\,(M-n+1) & n \leq M \\ \\ 0 & n > M \end{cases}$$

$$==> \qquad P = \frac{1}{1 + \sum_{n=1}^{M} C_n}$$

$$==> \qquad P = \frac{1}{\sum_{n=0}^{M}\left(\frac{M!}{(M-n)!}\left(\lambda/\mu\right)^n\right)}$$

Teorema IV.14.-

Con las condiciones del teorema anterior,

$$P = \begin{cases} \dfrac{M!}{(M-n)!}\left(\dfrac{\lambda}{\mu}\right)^n P_0 & n \leq M \\ \\ 0 & n > M \end{cases}$$

Demostración.-

Dado que $P_n = C_n P_0$, la demostración es trivial.-

Teorema IV.15.-

Con las condiciones del teorema IV.13,

$$Lq = M - \frac{\lambda + \mu}{\lambda} (1 - P_0)$$

Demostración.-

Tenemos que $\quad Lq = \sum_{n=1}^{\infty} (n-1)\ P_n$

$$==> \qquad Lq = \sum_{n=1}^{M} (n-1)\ P_n$$

$$==> \qquad Lq = \sum_{n=1}^{M} n\ P_n - \sum_{n=1}^{M} P_n$$

$$==> \qquad Lq = M\ P_0 + \sum_{n=1}^{M} n\ P_n - \sum_{n=1}^{M} P_n - M\ P_0$$

$$\Longrightarrow \quad Lq = M\left(1 - \sum_{n=1}^{M} P_n\right) + \sum_{n=1}^{M} n\, P_n - (1 - P_0) - M\, P_0$$

$$\Longrightarrow \quad Lq = M - M \sum_{n=1}^{M} P_n + \sum_{n=1}^{M} n\, P_n - (1 - P_0) - M\, P_0$$

$$\Longrightarrow \quad Lq = M - \sum_{n=1}^{M} (M-n)\, P_n - (1 - P_0) - M\, P_0$$

$$\Longrightarrow \quad Lq = M - \sum_{n=1}^{M} \frac{(M-n)\, M!}{(M-n)!}\, P_0 - (1 - P_0) - M\, P_0$$

Obsérvese que el M-ésimo término de la sumatoria es cero.

$$\Longrightarrow \quad Lq = M - \sum_{n=1}^{M-1} \frac{(M-n)\, M!}{(M-n)!} \left(\frac{\lambda}{\mu}\right)^{n} P_0 - (1 - P_0) - M\, P_0$$

$$\Longrightarrow \quad Lq = M - \sum_{n=1}^{M-1} \frac{M!}{(M-n-1)!} \left(\frac{\lambda}{\mu}\right)^{n} P_0 - (1 - P_0) - M\, P_0$$

$$\Longrightarrow \quad Lq = M - \left(\frac{\mu}{\lambda}\right) \sum_{n=1}^{M-1} \frac{M!}{(M-(n+1))!} \left(\frac{\lambda}{\mu}\right)^{n+1} P_0 - (1 - P_0) - M\, P_0$$

Cambiando el indizador $m = n+1$

$$\Longrightarrow \quad Lq = M - \left(\frac{\mu}{\lambda}\right) \sum_{m=2}^{M} \frac{M!}{(M-m)!} \left(\frac{\lambda}{\mu}\right)^{m} P_0 - (1 - P_0) - M\, P_0$$

$$\Longrightarrow \quad Lq = M - \left(\frac{\mu}{\lambda}\right) \sum_{m=2}^{M} P_m - (1 - P_0) - M\, P_0$$

$$\Longrightarrow \quad Lq = M - \frac{\mu}{\lambda} (1 - P_1 - P_0) - (1 - P_0) - M\, P_0$$

$$\Longrightarrow \quad Lq = M - \frac{\mu}{\lambda} (1 - P_0) + \frac{\mu}{\lambda} P_1 - (1 - P_0) - M\, P_0$$

Y como $P_1 = M \frac{\lambda}{\mu} P \quad \Longrightarrow P_1 \frac{\mu}{\lambda} = M\, P_0$

$$Lq = M - \frac{\mu + \lambda}{\mu} (1 - P_0)$$

Teorema IV. 16.-

Con las condiciones del teorema IV.13

$$L = M - \frac{\mu}{\lambda} (1 - P_0)$$

Demostración.-

Tenemos que $L = \sum_{n=0}^{M} n\, P_n \implies L = \sum_{n=1}^{M} n\, P_n$

$$\implies \quad L = \sum_{n=1}^{M} (n-1)\, P_n + \sum_{n=1}^{M} P_n$$

$$\implies \quad L = Lq + (1 - P_0)$$

$$\implies \quad L = M - \frac{\mu}{\lambda}\,(1 - P_0)$$

Por último podemos decir que $W = L/\bar{\lambda}$ y $Wq = Lq/\bar{\lambda}$

donde $\bar{\lambda} = \sum_{n=0}^{\infty} \lambda_n\, P_n$

$$\implies \quad \bar{\lambda} = \sum_{n=0}^{M} \lambda_n\, P_n$$

$$\implies \quad \bar{\lambda} = \sum_{n=0}^{M} (M-n)\,\lambda\, P_n$$

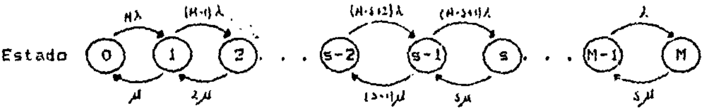
$$\implies \quad \bar{\lambda} = \lambda \sum_{n=0}^{M} (M-n)\, P_n$$

$$\implies \quad \bar{\lambda} = \lambda\, (M-L)$$

IV.6.- EL MODELO CON UNA FUENTE DE ENTRADA FINITA Y VARIOS SERVIDORES.-

En este caso, el diagrama de tasas se desarrolla como en la siguiente figura:

Figura IV.6.-



Teorema IV.17.-

En el modelo de colas con una fuente de entrada limitada y s servidores,

$$P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{M!}{(M-n)!\ n!} \left(\frac{\lambda}{\mu}\right)^n + \sum_{n=s}^{M} \frac{M!}{(M-n)!\ s!\ s^{n-s}} \left(\frac{\lambda}{\mu}\right)^n}$$

Demostración.-

Tenemos que

$$C_n = \begin{cases} \dfrac{M!}{(M-n)!\ \ n!} \left(\dfrac{\lambda}{\mu}\right)^n & n \leqslant s \\ \dfrac{M!}{(M-n)!\ \ s!\ \ s^{n-s}} \left(\dfrac{\lambda}{\mu}\right)^n & s \leqslant n \leqslant M \\ 0 & n > M \end{cases}$$

$$\therefore \quad P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{M!}{(M-n)!\ n!} \left(\frac{\lambda}{\mu}\right)^n + \sum_{n=s}^{M} \frac{M!}{(M-n)!\ s!\ s^{n-s}} \left(\frac{\lambda}{\mu}\right)^n}$$

Corolario IV.3.-

En el modelo de colas con una fuente de entrada limitada y s servidores

$$P_n = \begin{cases} P_0 \dfrac{M!}{(M-n)!\ \ n!} \left(\dfrac{\lambda}{\mu}\right)^n & 0 \leqslant n \leqslant s \\ P_0 \dfrac{M!}{(M-n)!\ \ s!\ \ s^{n-s}} \left(\dfrac{\lambda}{\mu}\right)^n & s \leqslant n \leqslant M \\ 0 & n > M \end{cases}$$

Demostración.-

Trivial

En este caso, los cálculos para L y Lq no se pueden simplificar, pero podemos poner a L en función de Lq como lo expresa el siguiente teorema:

Teorema IV.18.-

En el modelo de colas básico con una fuente de entrada limitada y s servidores

$$L = \sum_{n=0}^{s-1} n P_n + Lq + s\left(1 - \sum_{n=0}^{s-1} P_n\right)$$

Demostracion.-

$$\text{Tenemos que } L = \sum_{n=0}^{\infty} n P_n \implies L = \sum_{n=0}^{M} n P_n$$

$$\implies L = \sum_{n=0}^{s-1} n P_n + \sum_{n=s}^{M} n P_n$$

$$\implies L = \sum_{n=0}^{s-1} n P_n + \sum_{n=s}^{M} (n-s) P_n + \sum_{n=s}^{M} s P_n$$

$$\implies L = \sum_{n=0}^{s-1} n P_n + Lq + s\left(1 - \sum_{n=0}^{s-1} P_n\right)$$

Los cálculos de W y Wq son los mismos que en el modelo para un solo servidor, $W = L/\bar{\lambda}$ y $Wq = Lq/\bar{\lambda}$ donde $\bar{\lambda} = \sum_{n=0}^{M} \lambda_n P_n$

Después de ver los principales teoremas de los modelos de colas, estudiaremos otros modelos de colas en donde λ y μ varian dependiedo del estado del sistema.

COEFICIENTES

DE

PRESION

CAPITULO V

V.1.- TASA DE SERVICIO QUE DEPENDE DEL ESTADO.-

Hasta esta parte del estudio de la teoría de las colas, hemos supuesto modelos con una tasa de servicio constante siempre. Pero no siempre sucede esto en los sistemas reales de colas, generalmente cuando el estado del sistema es grande sucede que los servidores tienden a trabajar mas rapidamente que cuando el estado del sistema es pequeño o cero. Esto puede deberse a que los servidores incrementan su esfuerzo, o porque se obtiene ayuda en ciertas partes del servicio, o también por una mejor calidad del servicio (esto es, una mayor rapidez sin lastimar por esto la eficiencia del servicio).

En esta fase del estudio se ve la conveniencia de desarrollar un nuevo patrón que represente el incremento de las tasas reales de servicio cuando aumenta el tamaño de la cola. Este nuevo modelo nos debe dar una aproximación razonable del patrón real, y debe de ser muy sencillo para que se pueda poner en práctica facilmente.

Lo único que haremos será definir una nueva constante , a la que llamaremos coeficiente de presión, y recordar algunas de las definiciones vistas en el Capítulo II.

Plantearemos el caso para un solo servidor de la siguiente manera. Haremos $\mu_n = n^a \mu_1$ donde

n significa el número de clientes en el sistema (estado de el sistema).

μ_n significa la tasa media de servicio cuando hay n clientes en el sistema.

$1/\mu_1$ significa la tasa media de servicio cuando hay un cliente en el sistema (tiempo "normal" de servicio esperado).

a significa el "coeficiente de presión" (constante positiva que indica como se ve afectada la tasa de servicio en el sistema por el estado del mismo).

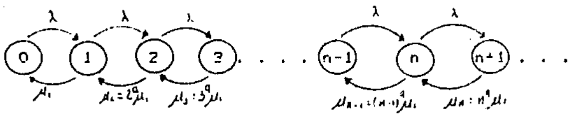
Nota: Los modelos de colas del capítulo anterior suponían que $a = 0$, si suponemos $a = 1$, nos queda que la tasa de servicio es directamente proporcional a el estado del mismo.

Si suponemos ahora que las entradas siguen una distribución de Poisson con $\lambda_n = \lambda$, y tiempos de servicios exponenciales μ_n, esto resulta un caso especial en el proceso "Nacimiento-Muerte".

Ademas, el diagrama de tasas con la nueva costante ya establecida, es el siguiente:

Figura V.1.-

Estado



Analizando esto, llegamos a las siguientes ecuaciones de balance:

Estado	Tasa de entrada	=	Tasa de salida
0	$\mu_1 P_1$	=	λP_0
1	$\lambda P_0 + \mu_2 P_2$	=	$(\lambda + \mu_1) P_1$
2	$\lambda P_1 + \mu_3 P$	=	$(\lambda + \mu_2) P_2$
...	....		
...	....		
n-1	$\lambda P_{n-2} + \mu_n P_n$	=	$(\lambda + \mu_{n-1}) P_{n-1}$
n	$\lambda P_{n-1} + \mu_{n+1} P_{n+1}$	=	$(\lambda + \mu_n) P_n$
...	....		

Por lo tanto,

$$P_1 = \frac{\lambda}{\mu_1} P_0$$

$$P_2 = \frac{\lambda}{\mu_2} P_1 + \frac{\mu_1 P_1 - \lambda P_0}{\mu_2} = \frac{\lambda}{\mu_2} P_1 = \frac{\lambda^2}{\mu_2 \mu_1} P_0 = \frac{\lambda^2}{2^a \mu_1 \mu_1} P_0$$

$$P_3 = \frac{\lambda}{\mu_3} P_2 + \frac{\mu_2 P_2 - \lambda P_1}{\mu_3} = \frac{\lambda}{\mu_3} P = \frac{\lambda^3}{\mu_3 \mu_2 \mu_1} P_0 = \frac{\lambda^3}{3^a \mu_1 \, 2^a \mu_1 \, \mu_1} P_0$$

....

....

$$P_n = \frac{\lambda}{\mu_n} P_{n-1} + \frac{\mu_{n-1} P_{n-1} - \lambda P_{n-2}}{\mu_n} = \frac{\lambda}{\mu_n} P_{n-1} = \frac{\lambda^n}{\mu_n \cdots \mu_2 \mu_1} P_0 = \frac{\lambda^n}{n^a \mu_1 \cdot \ldots \cdot 2^a \mu_1 \, \mu_1} P_0$$

....

De aquí podemos obtener que:

$$C_n = \frac{\lambda^n}{\mu_n \cdot \ldots \, \mu_2 \cdot \mu_1}$$

$$C_n = \frac{\lambda^n}{n^a \mu_1 \cdot \ldots \cdot 2^a \mu_1 \cdot \mu_1}$$

$$C_n = \frac{(\lambda/\mu_1)^n}{(n!)^a}$$

Teorema V.1 -

$$\text{a) } P = \frac{1}{\sum_{n=0}^{\infty} \frac{(\lambda/\mu_1)^n}{(n!)^a}}$$

$$\text{b) } P = \frac{(\lambda/\mu_1)^n}{(n!)^a} P_0$$

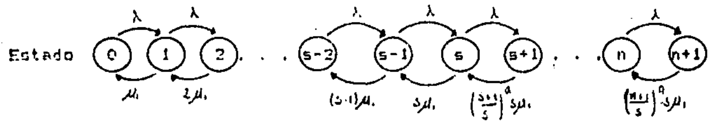
Demostración.-

Trivial.

En este caso, no se cuenta con expresiones "sencillas" para obtener las sumas que nos determinan a L y Lq, pero podemos obtener una buena aproximación mediante el cálculo de número grande de terminos, lo cual es fácil hacerlo por medio de una computadora (1).

Para el caso de varios servidores, el diagrama de tasas es el siguiente:

Figura V.2.-



(1) Los interesados pueden recurrir a Gross, Harris, 1974, "Fundamentals of Queueing Theory", Wiley, Nueva York, pp. 193-195.

Por lo tanto, μ_n variará de la siguiente forma:

$$\mu_n = \begin{cases} \mu_1 n & n \leq s \\ (n/s)^a \; s\mu_1 & n \geq s \end{cases}$$

Teorema V.2.-

En el caso de varios servidores, en una tasa de servicio que depende del estado, C_n varía de la siguiente forma:

$$C_n = \begin{cases} (\lambda/\mu_1)^n/n! & n < s \\ \dfrac{(\lambda/\mu_1)^n}{s! \left(\frac{n!}{s!}\right)^a \; s^{(1-a)(n-s)}} & n \geq s \end{cases}$$

Demostración.-

Tenemos que si $n < s$, la demostración en este caso es trivial.

Si $n \geq s$, entonces

$$C = \frac{\lambda^n}{\mu_n \mu_{n-1} \cdot \ldots \cdot \mu_{s+1} \mu_s \cdot \ldots \cdot \mu_2 \mu_1}$$

$$\Longrightarrow \quad C = \frac{\lambda^n}{\left(\frac{n}{s}\right)^a s\mu_1 \left(\frac{n-1}{s}\right)^a s\mu_1 \cdot \ldots \cdot \left(\frac{s+1}{s}\right)^a s\mu_1 \cdot s\mu_1 (s-1)\mu_1 \cdot \ldots \cdot 2\mu_1 \mu_1}$$

$$\Longrightarrow \quad C = \frac{\lambda^n}{\left(\frac{n}{s}\right)^a \left(\frac{n-1}{s}\right)^a \cdot \ldots \cdot \left(\frac{s+1}{s}\right)^a \; s^{n-s} \; s! \; (\mu_1)^n}$$

$$\Longrightarrow \quad C = \frac{(\lambda/\mu_1)^n}{\left(\frac{n!}{s!}\right)^a \left(1/s^{n-s}\right)^a \; s^{n-s} \; s!}$$

$$\Longrightarrow \quad C = \frac{(\lambda/\mu_1)^n}{s! \left(\frac{n!}{s!}\right)^a \; s^{(n-s)(1-a)}}$$

Teorema V.2.-

En el caso de varios servidores, con una tasa de servicio que depende del estado:

a) $$P_0 = \frac{1}{1 + \sum_{n=1}^{s-1} \frac{(\lambda/\mu)^n}{n!} + \sum_{n=s}^{\infty} \frac{(\lambda/\mu_1)^n}{s!\,(n!/s!)^a\; s^{(1-a)(n-s)}}}$$

b)

$$P_n = \begin{cases} \dfrac{(\lambda/\mu_1)^n}{n!}\, P_0 & n \leq s \\[2ex] \dfrac{(\lambda/\mu_1)^n}{s!\,(n!/s!)^a\; s^{(1-a)(n-s)}}\, P_0 & n \geq s \end{cases}$$

Demostración.-

Trivial.

Aunque al parecer, el determinar el coeficiente de presión puede ser difícil, en la práctica resulta mucho mas fácil de lo que imaginamos, solo es necesario observar el patrón de comportamiento del servicio, y tener algunos conocimientos básicos de las propiedades de los logaritmos. Un ejemplo muy claro de esta determinación del coeficiente de presión es el siguiente:

En un supermercado, el gerente descubre que el tiempo que un cajero emplea en dar servicio a un cliente disminuye a medida que el número de clientes aumenta, en parte porque el cajero trabaja más rapidamente, y en parte porque no empaca todas las cosas, sino que pide al cliente que embolse las cosas más sencillas. Los nuevos datos que el gerente obtiene son los siguientes, si no hay clientes esperando la tasa media de servicio es de ocho minutos por cliente, mientras que el cajero consume cinco minutos cuando hay seis clientes esperando, por lo tanto

$\mu_1 = 15/2$ clientes por hora

$\mu_6 = 12$ clientes por hora

$$\mu_6 = \mu_1\, 6^a \quad \Longrightarrow \quad 6^a = 12/(15/2) = 8/5$$

$$\Longrightarrow \quad a = \ln(8/5) \;/\; \ln 6 = .47 \;/\; 1.7917 = .262312$$

V.2.- TASA DE LLEGADA QUE DEPENDE DEL ESTADO.-

Hay ocasiones en que en lugar de variar la tasa de servicio, lo que puede variar es la tasa de llegadas, esto puede deberse a que se puede desviar a los clientes a otro medio de servicio, o porque los clientes renuncian a ingresar a la cola, o por otras razones.

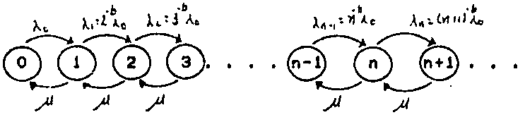
En este caso, el modelo en lugar de hacer variar la tasa de servicio, debe hacer variar la tasa de llegadas y esto puede hacerse de la siguiente forma:

Definamos ahora $\lambda_n = (n+1)^{-b}\lambda_0$

donde b es una constante cuya interpretación es análoga a a (esto es en el caso de un solo servidor).

El diagrama de tasas en este caso es el siguiente:

Figura V.3.-

Estado 

Por tanto C_n nos queda de la forma

$$C_n = \frac{\lambda_0 \cdot \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_{n-1}}{\mu^n}$$

$$\Longrightarrow \quad C = \frac{\lambda_0 \; 2^{-b}\lambda_0 \; 3^{-b}\lambda_0 \cdot \ldots \cdot n^{-b} \; \lambda_0}{\mu^n}$$

$$\Longrightarrow \quad C = \frac{(\lambda_0/\mu)^n}{(n!)^b}$$

Esto da lugar al siguiente teorema.

Teorema V.4.-

En un modelo de colas en que la tasa de llegada depende del estado, con un solo servidor,

a) $P_0 = \dfrac{1}{\sum_{n=0}^{\infty} \dfrac{(\lambda_0/\mu)^n}{(n!)^b}}$

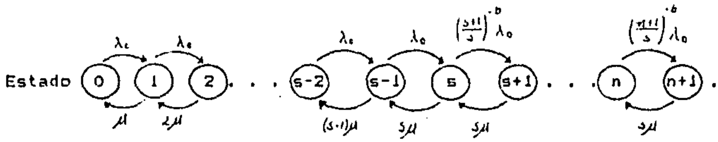
b) $P_n = \dfrac{(\lambda_0/\mu)^n}{(n!)^b} \, P_0$

Demostración.-

Trivial.

En el caso de varios servidores, el diagrama de tasas nos queda de la siguiente forma:

Figura V.4.-



Por lo tanto, C_n nos queda muy similar a el obtenido en el modelo en que la tasa de servicio depende del estado,

$$C_n = \begin{cases} \dfrac{(\lambda_0/\mu)^n}{n!} & n \leq s \\[2ex] \dfrac{(\lambda_0/\mu)^n}{s!\ (n!/s!)^b\ s^{(1-b)(n-s)}} & n \geq s \end{cases}$$

Con esto podemos concluir el siguiente teorema,

Teorema V.5.-

En el caso de varios servidores, en un modelo de colas en que la tasa de llegadas depende del estado,

$$\text{a) } P = \frac{1}{1 + \sum_{n=1}^{s-1} \frac{(\lambda_0/\mu)^n}{n!} + \sum_{n=s}^{\infty} \frac{(\lambda_0/\mu)^n}{s! \; (n!/s!)^b \; s^{(1-b)(n-s)}}}$$

$$\text{b) } P = \begin{cases} \dfrac{(\lambda_0/\mu)^n}{n!} P_0 & n < s \\[2ex] \dfrac{(\lambda_0/\mu)^n}{s! \; (n!/s!)^b \; s^{(1-b)(n-s)}} P_0 & n \geq s \end{cases}$$

Demostración.-

Trivial.

V.3.- MODELOS DE COLAS EN QUE LAS TASAS DE SERVICIO Y DE LLEGADAS DEPENDEN DEL ESTADO.-

Este modelo nos ofrece una visión mas general y mas real de el problema, es muy similar a los dos patrones anteriores, pues lo que sucede en este modelo es que hace variar tanto a la tasa de servicio como a la tasa de llegadas de la siguiente forma:

Si $\mu_n = n^a \mu$, y $\lambda_n = (n+1)^{-b} \lambda_0$, definimos una constante nueva $c = a + b$, y por tanto en el caso de un solo servidor,

$$C_n = \frac{(\lambda_0/\mu)^n}{(n!)^c}$$

y por tanto,

Teorema V.6.-

En el modelo de colas en el que varía la tasa de llegadas y la de servicio dependiendo del estado, con un solo servidor,

$$P = \frac{1}{\sum_{n=0}^{\infty} \frac{(\lambda_0/\mu_1)^n}{(n!)^c}}$$

$$P_n = \frac{(\lambda_0/\mu_1)^n}{(n!)^c} P$$

Donde $c = a + b$, y a y b son definidos como en las dos secciones anteriores.

Demostración.-

Trivial.

En el caso de varios servidores, los resultados son como a continuación se describen:

$$C_n = \begin{cases} \dfrac{(\lambda_0/\mu_1)^n}{n!} & 1 \leqslant n \leqslant s \\[2em] \dfrac{(\lambda_0/\mu_1)^n}{s! \; (n!/s!)^c \; s^{(1-c)(n-s)}} & n \geqslant s \end{cases}$$

y por lo tanto,

Teorema V.7.-

En el caso de varios servidores en el modelo de colas en que tanto la tasa de servicio como la de llegadas dependen del estado,

a) $$P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{(\lambda_0/\mu_1)^n}{n!} + \sum_{n=s}^{\infty} \frac{(\lambda_0/\mu_1)^n}{s! \; (n!/s!)^c \; s^{(1-c)(n-s)}}}$$

$$
\text{b) } P_n = \begin{cases} \dfrac{(\lambda_0/\mu_1)^n}{n!} \, P_0 & 0 \leq n \leq s \\[2ex] \dfrac{(\lambda_0/\mu_1)^n}{s! \; (n!/s!)^c \; s^{(1-c)(n-s)}} \, P_0 & n \geq s \end{cases}
$$

Demostración.-

Trivial.-

Hasta aquí, hemos visto otros modelos de colas que se encuantran dentro del proceso Nacimiento-Muerte, aunque la tasa de servicio y la de llegadas no son constantes ya que dependen de el estado, con esto tenemos ya material suficiente para poder resolver los problemas que se nos presentan en la sección I.3, solo nos falta analizar como se aplica la teoría de decisiones en los modelos de colas, cosa que veremos en el siguiente capítulo.

LA TEORIA DE DECISIONES EN LOS MODELOS DE COLAS

CAPITULO VI

VI.1.- LA TOMA DE DECISIONES EN MODELOS DE COLAS.-

Las situaciones en que dentro del análisis del problema es necesario tomar cierto tipo de decisiones, es de gran variedad, y por tanto, no es posible tener un solo procedimiento que nos resuelva todo este tipo de situaciones, pero si podemos plantear un procedimiento para las decisiones mas comunes como son:

i) Número de servidores en un medio de servicio

ii) Eficiencia de los servidores

iii) Número de los medios de servicio

Los problemas vistos en la sección I.3, nos ilustran situaciones a las que debemos responder con alguna de estas soluciones.

Como ya sabemos, la disyuntiva de todos los sistemas de colas surge en estas dos consideraciones:

1) Reducir los costos de servicio (esta consideración, por tanto recomienda un mínimo nivel de servicio) y,

2) Reducir los largos tiempos de espera (lo cual recomienda un alto nivel de servicio).

Estas dos consideraciones son las que provocan los conflictos en la toma de decisiones, y la meta en este caso es obtener el mejor balance posible, entre el costo de servicio y la reducción de los tiempos de espera.

Las siguientes gráficas nos ilustran un poco las consideraciones hechas anteriormente.

Figura VI.1.-

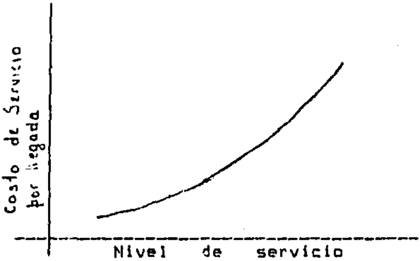


Figura VI.2.-

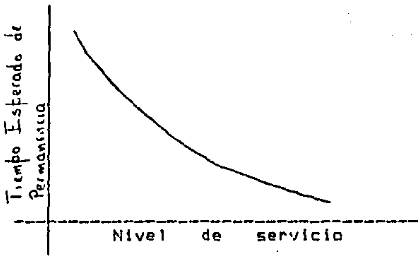
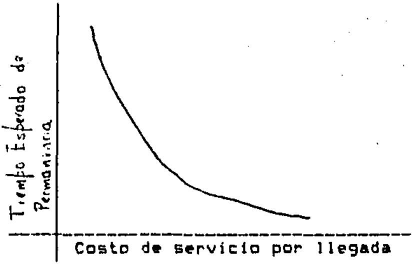


Figura VI.3.-



Nuestro problema en estos casos es poder medir en términos de costos, los tiempos de espera por unidad de tiempo, o lo que es lo mismo, hacernos la siguiente pregunta: " ¿ a cuanto desembolso es equivalente para un cliente la espera de una unidad de tiempo ? ".

Este modo de medir es muy subjetivo en situaciones en que tal cliente es externo a el "negocio" que presta el servicio, ya que aqui el modo de poder medir las pérdidas por la espera de una unidad de tiempo se basa tan solo en una estimación que puede parecer intangible.

Pongamos el caso de un cliente que llega a un supermercado, el costo de la espera de este cliente consiste en la utilidad perdida, ya sea porque el cliente se vaya demasiado irritado y decida no volver mas (a esto se le llama una pérdida a futuro), o bien, simplemente porque el cliente acabe por perder la paciencia y lleve sus "negocios" a otro lado.

El problema de estimar los costos de espera en clientes que pertenecen a la organización que presta el servicio, es mas fácil de estimar, pues podemos identificar estos costos de espera con la ociosidad de los clientes.

Pongamos ahora el caso de un maquinista en una fábrica de chocolates, los momentos en que el maquinista va a obtener a cierta parte de la empresa los materiales que el necesita para poner a trabajar su máquina (estos materiales pueden ser azúcar, cacao, grasas, etc.), y realiza dentro de la misma empresa una cola para poder obtener estos materiales, son momentos en que su maquinaria estará sin trabajar, y por tanto, dejará de producir cierta cantidad de utilidades, en estos casos, la productividad perdida nos dará como resultado una utilidad perdida, que es medible facilmente en los costos. Cabe mencionar también que esta cantidad de tiempo desperdiciado se empieza a contar desde que la máquina está parada hasta que es puesta a trabajar de nuevo, esto es, desde que el operador de la máquina empieza su recorrido hacia el lugar donde almacenen los materiales, su tiempo de espera en la cola, y el tiempo que tarda en volver a su máquina.

En los casos en que los "clientes" no pertenecen a la organización que brinda el servicio, debemos pues, obtener costos tangibles y poder ofrecer (aunque talvez en forma muy imprecisa) consideraciones de costos para poder utilizar la ayuda de un análisis matemático mas claro.

Despues de haber hecho este análisis, es "claro" que nuestro objetivo se basa en poder minimizar el costo esperado de servicio, y el costo esperado de la espera para ese servicio.

Esto es: Minimizar $E(CT) = E(CS) + E(CE)$

donde $E(CT)$ significa el valor esperado del costo total.

$E(CS)$ significa el valor esperado del costo de servicio.

$E(CE)$ significa el valor esperado del costo de la espera.

VI.2.- EL VALOR ESPERADO DE LA ESPERA.-

Para poder expresar el valor esperado del costo de espera, esto es $E(CE)$, plantearemos una función de costo de espera que nos describa la variación real del costo de la espera en que se incurre.

Si consideramos el caso en que los clientes en el sistema pertenecen a la organización que presta el servicio, observamos que la tasa de rendimiento perdido es (en la gran mayoría de las veces) proporcional al número de clientes en el sistema de colas, esto lo podemos observar en el problema 1 y 3 de la sección I.3. Si observamos el problema 1, podemos observar que a veces es posible mantener algunos miembros de la población potencial inactivos, pero un incremento grande en el número de estos miembros ociosos puede lesionar en forma grave los intereses económicos de la empresa.

En estos casos. lo que es necesario analizar es el número N de clientes en el sistema de colas, y por tanto la forma de plantear las funciones de costo de espera es poder construir una función que dependa del estado, a esta función la denominaremos por c(N).

Esta función nos debe estimar la tasa de los costos de espera cuando ocurre que N = n, y calculando las probabilidades P_n, podemos obtener E(CE) mediante el cálculo del valor esperado de una función de una variable aleatoria discreta, esto es:

$$E(CE)=\sum_{n=0}^{\infty} c(n)\ P_n$$

Nota: Cuando c(n) es una función lineal, es decir, la tasa del costo de espera es proporcional a N (c(n) = Kn), entonces, la fórmula para obtener E(CE) es,

$$E(CE) = K\sum_{n=0}^{\infty} n\ P_n = K\ L$$

donde K está representando el costo de la espera por unidad de tiempo para cada cliente.

En el problema 1, tenemos 3 máquinas de reserva, por tanto, no tendremos ninguna pérdida mientras que el número de máquinas que requieren reparación no sea mayor que 3. Mientras que si el número de clientes en el sistema es mayor que 3, la utilidad perdida será de \$ 11,000.00 diarios por cada "cliente adicional", y así pues,

$$c(n) = \begin{cases} 0 & 0 \leqslant n \leqslant 3 \\ \\ 11{,}000\ (n-3) & 4 \leqslant n \leqslant 15 \end{cases}$$

Por lo tanto podemos [illegible] tabla

		s = 1		s = 2	
N = n	c(n)	P_n	c(n) P_n	P_n	c(n) P_n
0	0	.1805	0	.35800	0
1	0	.1805	0	.35800	0
2	0	.1680	0	.16706	0
3	0	.1460	0	.07240	0
4	11 000	.1170	1 287	.02896	310
5	22 000	.0860	1 892	.01062	234
6	33 000	.0570	1 881	.00354	117
7	44 000	.0340	1 496	.00106	47
8	55 000	.0180	990	.00028	15
9	66 000	.0090	594	.00006	4
10	77 000	.0030	231	.00001	1
11	88 000	.0010	88	2×10^{-6}	0
12	99 000	.0003	30	3×10^{-7}	0
13	110 000	6×10^{-5}	7	3×10^{-8}	0
14	121 000	8×10^{-6}	1	2×10^{-9}	0
15	132 000	3×10^{-7}	0	6×10^{-11}	0
C(CE) =			\$ 8 497		\$ 737

Analizando ahora el caso en el problema 2, en que los clientes son externos a la organización que proporciona el servicio, cosa que ocurre con mucha frecuencia en sistemas de servicios comerciales, servicios de transporte y sistemas que brindan un servicio social, los costos de espera se deben considerar como un costo social, y la magnitud de este costo, es afectada por la magnitud de la espera experimentada por los clientes, además, en sistemas de servicios comerciales, y en servicios de transportes, se deben considerar también los "negocios futuros perdidos", de los que ya hemos hablado en la sección anterior.

En ambos casos pues, el costo de la espera está en función de $\mathcal{W}$, por tanto, nuestro objetivo es poder construir una función que dependa de $\mathcal{W}$, a esta función la denotaremos por $a(\mathcal{W})$.

Una manera de construir la función $a(\mathcal{W})$, es estimar el costo de la espera para un cliente cuando $\mathcal{W}=\omega$ ($a(\omega)$), para diferentes valores de ω y ajustar un polinomio a estos puntos, y por tanto el valor esperado de esta función (notemos que es una función continua) podemos expresarla como

$$E(a(\omega)) = \int_0^{\infty} a(\omega)\ f_{\mathcal{W}}(\omega)\ d\omega$$

donde $f_{\mathcal{W}}(\omega)$ es la función densidad de probabilidad para $\mathcal{W}$.

Ahora, notemos que $E(a(\omega))$ es el costo esperado de la espera por cliente, y $E(CE)$ es el costo esperado de la espera por unidad de tiempo, y estas cantidades son diferentes. Por tanto, lo que nos falta hacer es multiplicar $E(a(\omega))$ por el número esperado de clientes por unidad de tiempo. Por ejemplo, si la tasa media de llegadas es una constante λ, entonces

$$E(CE) = \lambda\ E(a(\omega)) = \lambda \int_0^{\infty} a(\omega)\ f_{\mathcal{W}}(\omega)\ d\omega$$

En el problema 2 de la sección I.3, podemos deducir que si un cliente llega y es atendido de inmediato, el costo de la espera es cero, mientras que si espera una hora, el costo es de $ 150.00, por tanto, la primera componente de $a(\omega)$ es un polinomio que pase por (0,0) y (1,150) y lo podemos expresar como $150\ \omega$ (donde ω es expresado en horas).

La segunda consecuencia de la espera, es la ruptura de la continuidad en sus labores, y aquí debemos considerar tres puntos, (0,0), (1/2,30) y (1,120), por tanto esta componente la podemos expresar como $120\,\omega^2$.

Por lo tanto, $a(\omega) = 150\ \omega + 120\ \omega^2$

Además, ajustandonos a la versión del modelo básico con un solo servidor (visto en la sección IV.1), podemos deducir que:

$$f_W(\omega) = \mu\ (1-\rho)\ e^{-\mu(1-\rho)\omega}$$

Por lo tanto,

$$E(a(\omega)) = \int_0^\infty (150\ \omega + 120\,\omega^2)\ \mu\ (1-\rho)\ e^{-\mu(1-\rho)\omega}\, d\omega$$

Además, $\mu\ (1-\rho) = \mu - \mu\ \frac{\lambda}{\mu} = \mu - \lambda$, entonces nos queda que

$$\mu(1-\rho) = \begin{cases} 3 & \text{para la persona encargada (sola)} \\ 13 & \text{para la persona encargada y el ayudante} \end{cases}$$

Por lo tanto,

$$E(a(\omega)) = \begin{cases} 76.7 & \text{para la persona encargada (sola)} \\ 113 & \text{para la persona encargada y el ayudante} \end{cases}$$

Y como $\lambda = 12$, entonces

$$E(CE) = \lambda\, E(a(\omega)) = \begin{cases} 920 & \text{para la persona encargada (sola)} \\ 156 & \text{para la persona encargada y el ayudante} \end{cases}$$

Si $a(\omega)$ es una función lineal, esto es

$$a(\omega) = K\,\omega$$

donde K es el costo de la espera por unidad de tiempo, entonces $E(CE) = \lambda E(K\omega) = \lambda\, K\, E(\omega) = K\lambda\omega = K\, L$

Nota: Observese que tanto en el análisis de funciones discretas como en estos casos, si la función obtenida es una función lineal, $E(CE) = K\,L$, esto es, el costo de la espera es directamente proporcional al tiempo total de espera.

En el problema 3, el valor promedio de producción de un carpintero trabajando es \$ 1,200.00 por hora, por lo tanto:

$$a(\omega) = 1200\omega = K\omega \qquad \text{donde } K = 1200$$

y así pues,

$$E(CE) = 1200\, L$$

VI.3.- PLANTEAMIENTO DE LOS MODELOS DE DECISION.-

En la sección VI.1, hemos visto que las tres variables que comunnmente requieren de la toma de una decisión son: el número de servidores (s), la eficiencia de estos servidores (μ), y el número de estaciones de servicio (observemos que la población potencial respecto a cada medio de servicio, variará respecto a este número de estaciones de servicio). En esta sección se plantearán los modelos para la toma de estas decisiones.

Modelo a) s desconocida

Este modelo se diseña para el caso en que μ y λ son fijas, y debe tomarse una decisión para poder obtener el número optimo de servidores que deben trabajar en un negocio.

Definiremos en este caso:

CS Costo marginal de un servidor por una unidad de tiempo.

Por tanto, conociendo μ , λ y CS debemos obtener s, tal que se minimice el valor esperado del costo total (E(CT)).

En este caso, E(CT) = s CS + E(CE), y como normalmente solo necesitamos algunos valores alternativos de s, la manera en que debemos tratar de resolver este modelo, es calcular E(CT) para estos valores alternativos y seleccionar el que minimiza a E(CT).

Para el problema 1 de la sección I.3, cada mecánico le cuesta a la compañía $ 7,000.00 diarios, y con los datos obtenidos en la sección anterior, podemos obtener la siguiente tabla:

s	s CS	E(CE)	E(CT)
1	7 000	8 497	15 497
2	14 000	737	14 737
$\geq$ 3	$\geq$ 21 000	$\geq$ 0	$\geq$ 21 000

Por lo tanto, la decisión en este caso es contratar a otro mecánico, y de este modo tener dos mecánicos en la companía.

Modelo b) μ y s desconocidas

Este modelo esta diseñado para el caso en que necesitamos seleccionar la eficiencia en el servicio (μ) y el número de servidores (s). En este caso es posible que se disponga de diferentes valores de μ pues podemos elegir la calidad de los servidores, por ejemplo: fábricas que dispongan de diferentes máquinas que realicen el mismo trabajo en cantidades de tiempos distintas, o ajustar una misma máquina a una velocidad diferente cambiando la cantidad de energía, otro tipo de ejemplo es la selección del número de miembros de una brigada que realicen una tarea determinada considerando a toda la brigada como un solo servidor. En estos casos los valores de μ pueden alternar (estas alternativas para μ son generalmente finitas).

Definiremos en este ...

$e(\mu)$ Costo marginal del servidor por unidad de tiempo cuando la tasa de servicio es .

U Conjunto de valores factibles de .

Por tanto, dados λ, $e(\mu)$ y U debemos encontrar μ y s tal que se minimice el valor esperado del costo total.

Ahora tenemos que $E(CT) = s\, e(\mu) + E(CE)$ donde $\mu \in U$

En el problema 2 de la sección I.3, $\mu = 15$ para la persona encargada (sola) y $\mu = 25$ para la persona encargada y su ayudante. Por tanto, $U = \{15, 25\}$

Ademas, con los datos de este ejemplo

$$e(\mu) = \begin{cases} 650 & \text{si } \mu = 15 \\ 1\,000 & \text{si } \mu = 25 \end{cases}$$

En este ejemplo también tenemos restringido el número de servidores a s=1, por lo tanto:

μ	s	$s\,e(\mu)$	E(CE)	E(CT)
15	1	650	920	1 570
25	1	1 000	156	1 156

Por tanto, la decisión es contratar el ayudante para la persona encargada del R.F.C.

Modelo c) s y λ desconocidas

Este modelo se diseña para el caso en que es necesario seleccionar el número de medios de servicio y el número de servidores (s). Esto ocurre mucho en plantas industriales, donde una población de empleados solicita cierto servicio indispensable para el trabajo, por tanto, debe tomarse una decisión sobre la proporción de empleados que debe asignarse a cada medio de servicio (notese que al hablar de esta proporción estamos involucrando implicitamente a λ).

Con el fin de simplificar este modelo, se requirirá que λ y s sean iguales para todos los medios de servicio.

Definiremos ahora

CS Costo marginal de un servidor por unidad de tiempo.

CF Costo fijo del servicio por cada planta de servicio por unidad de tiempo.

λ_p Tasa media de llegadas de la poblacion.

n Numero de medios de servicio.

Nota: Observese que $\lambda_p = n\lambda$, esto es $\lambda = \lambda_p / n$

Por tanto, dados μ, CS, CF y λ_p, debemos encontrar λ y s, tratando de este modo de minimizar E(CT).

La primera impresión es que E(CT) = n ((CF + s CS) + E(CE)), donde E(CE) es el valor esperado del costo de espera por unidad de tiempo para cada medio de servicio, lo cual es un error muy grande, pues, lo que esto implicaría es que n debe de ser 1, ya que de no ser así, en algunos momentos habría servidores ociosos en uno de los medios de servicio, mientras que en otros medios habría clientes esperando, de este modo la tasa media de servicio sería menor que si todos los clientes tuviesen la acceso a todos los servidores en un solo medio de servicio, esto en sí, ya nos indica que algo debe de estar equivocado. Además, estamos ignorando un dato que es la clave de nuestro error, ese dato es el tiempo que un cliente emplea en ir a uno de los medios de servicio. Al principio de este capítulo habiamos hecho la observación que la utilidad perdida es la suma del tiempo de espera en la cola mas el tiempo empleado en el trayecto (ir y regresar) a uno de los medios de servicio.

Si denotamos este tiempo empleado en este recorrido por la variable aleatoria R, el tiempo total perdido por un cliente es en realidad $\mathcal{W} + R$. Por tanto, si llamamos CR a el costo del tiempo que se emplea en el recorrido para cada cliente por cada unidad de tiempo y si separamos el tiempo total desperdiciado en el costo del tiempo de espera mas el costo del tiempo del recorrido, tendremos que

$$E(CT) = n\,((CF + s\,CS) + E(CE) + \lambda\,CR\,E(R))$$

donde λ es el número esperado de llegadas por unidad de tiempo para cada medio de servicio.

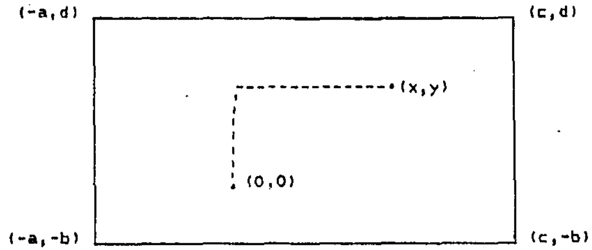
Por tanto, si podemos evaluar E(R), podremos resolver este modelo para los diversos valores de s y n, obteniendo de esta forma el valor mínimo para E(CT). En la siguiente sección veremos como poder evaluar E(R).

VI.4.- E(R).-

A el valor E(R), lo podemos considerar como el tiempo promedio que utilizan los clientes para trasladarse a un medio de servicio y regresar de él. Para poder evaluar E(R) consideraremos que la población asignada al medio de servicio esta uniformemente distribuida en toda el área asignada, que la velocidad promedio de recorrido no depende de la distancia del trayecto, y que todos los recorridos siguen trayectorias rectilineas y que son paralelas a los lados principales al área sobre la que se está haciendo el estudio. De esta forma, desarrollaremos un procedimiento general para evaluar E(R).

Consideraremos un área rectangular y recorridos rectilineos como en la siguiente figura.-

Figura VI.4.-



Definiremos en este caso:

R	Tiempo de recorrido (ir y regresar) por llegada.
v	Velocidad promedio de los clientes en ir a un medio de servicio.
a,b,c,d	Distancias respectivas desde el medio de servicio hasta la frontera del área asignada para tal medio (observe la figura VI.4).

Por tanto, dados v, a, b, c, d nuestro objetivo es encontrar E(R) (el valor esperado de R).

Si usamos un sistema de coordenadas ortogonales (x,y), donde (x,y) representa la localización de un cliente, resultará que estas coordenadas son en realidad variables aleatorias X, Y, de donde proviene una llegada aleatoria. donde X varía desde -a hasta c, y Y varía desde -b hasta d, y como la distancia recorrida por ir y regresar en una llegada aleatoria es

$$D = 2\ (E(|X|) + E(|Y|))$$

podemos decir entonces que:

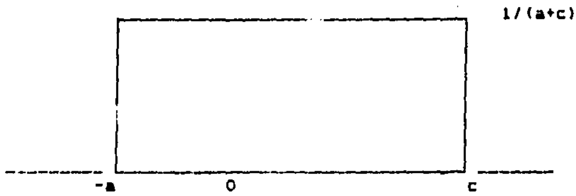
$$E(R) = (2/v)\ (E(|X|) + E(|Y|))$$

así pues, nuestro objetivo ahora es identificar las distribuciones de probabilidad para |X| y |Y| y calcular sus medias.

Analicemos en primer lugar a |X| . La distribución de probabilidad para |X| se puede obtener a partir de la distribución de probabilidad de X, ya que con el supuesto de que los clientes estan distribuidos uniformemente en el área asignada, X debe tomar una distribución uniforme entre -a y c, tal como se muestra en la siguiente figura.

Figura VI.5.-

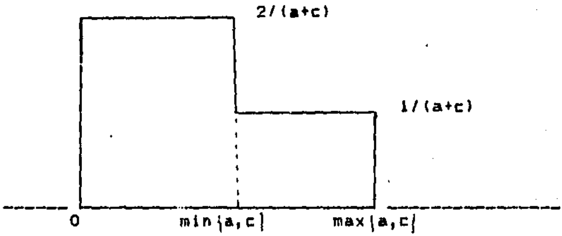
Función densidad de probabilidad de X



La altura "claramente" es 1/(a+c)

Por lo tanto, como $|-x| = |x|$, entonces se llega a que la distribución de probabilidad de $|X|$ es como se muestra en la siguiente figura.

Figura VI.6.-



Así pues,

$$E(|X|) = \int_0^{\max\{a,c\}} x\, f_{|X|}(x)\, dx$$

$$= \int_0^{\min\{a,c\}} \frac{2x}{a+c}\, dx + \int_{\min\{a,c\}}^{\max\{a,c\}} \frac{x}{a+c}\, dx$$

$$\frac{(\min\{a,c\})^2}{a+c} + \frac{1}{2}\;\frac{(\max\{a,c\})^2 - (\min\{a,c\})^2}{a+c}$$

$$= \frac{1}{2\,(a+c)}\left((\max\{a,c\})^2 + (\min\{a,c\})^2\right)$$

$$= \frac{a^2 + c^2}{2\,(a+c)}$$

El análisis de $|Y|$ es similar, por tanto, el resultado es:

$$E(\,Y\,) = \frac{b + d}{2\,(b+d)}$$

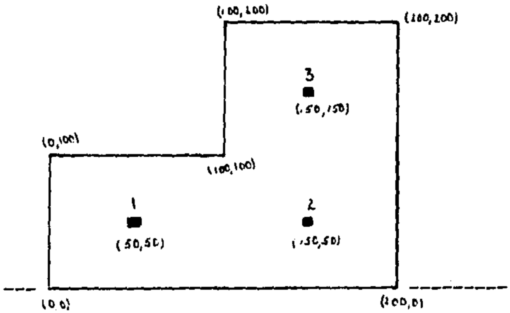
y

$$E(R) = \frac{1}{v}\quad\frac{a+c}{a+c} + \frac{b+d}{b+d}$$

Analicemos por último, los nuevos datos del problema 3 de la sección I.3 que se obtienen de un análisis de los planos.

Los planos de la nueva fábrica proporcionados por los arquitectos son los siguientes:

Figura VI.6,-



Los sitios 1, 2, 3 son las localizaciones posibles para los depósitos de herramientas, las coordenadas se dan en metros, los carpinteros estan distribuidos de manera uniforme en toda la fábrica, y a cada carpintero se le asignará el depósito mas cercano, la velocidad promedio de un carpintero para trasladarse a las plantas de servicio es v = 5 Km/hr = 5 000 m/hr.

Las tres alternativas que tiene el Presidente de "Maderas de México" son:

Alternativa 1: Tener 3 depósitos (Plantas 1, 2, 3)
Alternativa 2: Tener 1 depósito (Planta 2)
Alternativa 3: Tener 2 depósitos (Plantas 1 y 3)

Analizando cada alternativa obtenemos

Con 3 depósitos:

Si se instalan tres depósitos, cada uno de ellos estaría dando servicio a un área de 100×100 m². Por tanto, en este caso analizando una de estas áreas, obtendremos que $a = c = 50$, y $b = d = 50$.

$$E(R) = \frac{1}{5000}\left[\frac{(50)^2 + (50)^2}{50 + 50} + \frac{(50)^2 + (50)^2}{50 + 50}\right]$$

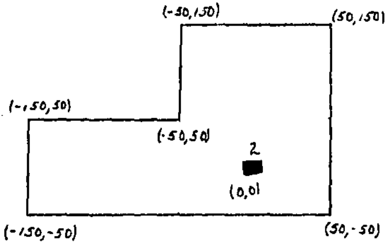
$$= \frac{1}{5000}\left[\frac{5000}{100} + \frac{5000}{100}\right]$$

$$= 1/50 = .02$$

Con 1 depósito

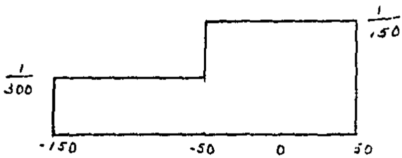
La obtención de E(R) en este caso es mas complicada. En primer lugar remarcaremos al sitio 2 con las coordenadas (0,0) y de este modo el plano queda de la siguiente forma:

Figura VI.7.-



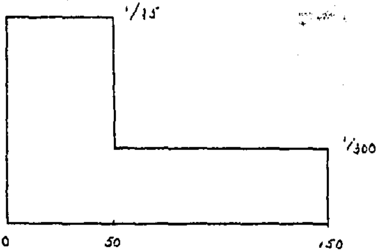
Así pues, dividiendo la altura entre el área total (de tal forma que el área bajo la curva de la función densidad sea uno), nos quedará como se ilustra a continuación.

Figura VI.8.-



Y por lo tanto, la función de distribución para |X| es como en la siguiente figura.

Figura VI.9.-



Y por lo tanto,

$$E(|X|) = \int_0^{150} x \, f_{|X|}(x)\,dx$$

$$= \int_0^{50} x \frac{1}{15}\,dx + \int_{50}^{150} x \frac{1}{300}\,dx$$

$$= \frac{x^2}{150}\Big|_0^{50} + \frac{x^2}{600}\Big|_{50}^{150}$$

$$= \frac{2500}{150} + \frac{20000}{600}$$

$$= 50$$

"Claramente", si utilizamos el mismo procedimiento para obtener E(| Y |), encontraremos que la distancia para | Y | es idéntica a la de |X|, y así E(|Y|) = 50

Por tanto,

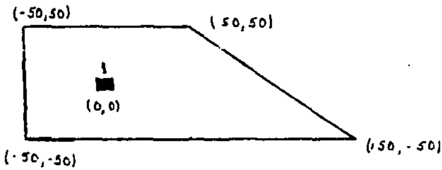
$$E(R) = \frac{2}{5000}\,(50 + 50) = .04$$

Con 2 depósitos

Con dos depósitos las áreas asignadas a los sitios 1 y 3 se dividirán por un segmento rectilíneo entre (100,100) y (200,0). Además, las dos áreas asignadas a los depósitos y los mismos depósitos se localizan simetricamente respecto a este segmento rectilíneo, y por lo tanto, E(R) es el mismo para los dos depósitos.

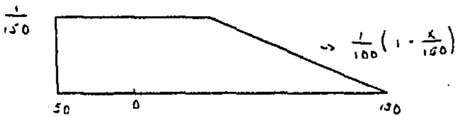
Remarcando ahora el sitio 1 como (0,0) resultarán las siguientes posiciones.

Figura VI.10.-



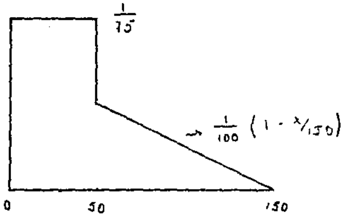
Por tanto, la funcion densidad de probabilidad para X es como se muestra a continuación.

Figura VI.11.-



Y por lo tanto, la función densidad de probabilidad para |X| es:

Figura VI.12.-



Por lo tanto

$$E(|X|) = \frac{1}{75}\int_0^{50} x\,dx + \frac{1}{100}\int_{50}^{150}\left(x - x^2/150\right)dx$$

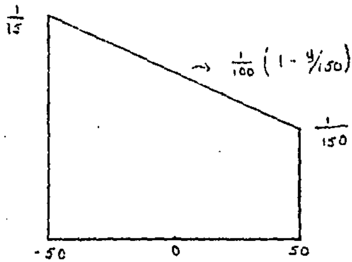
$$= \left.\frac{x^2}{150}\right|_0^{50} + \frac{1}{100}\left.\left(\frac{x^2}{2} - \frac{x^3}{450}\right)\right|_{50}^{150}$$

$$= \frac{2500}{150} + \frac{25000}{900}$$

$$= \frac{400}{9}$$

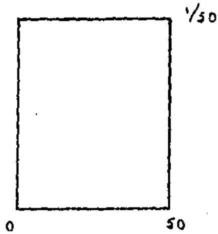
La función densidad de probabilidad para Y se obtiene similarmente utilizando el ancho del área asignada y despues dividiendo entre el tamaño del área, como se muestra en la siguiente figura.

Figura VI.13.-



Y por tanto, la distribución densidad de probabilidad es uniforme para |Y|, como se ilustra a continuación.

Figura VI.14.-



Y así,

$$E(Y) = \frac{1}{50}\int_0^{50} y \, dy$$

$$= \frac{1}{100} \, y^2 \Big|_0^{50}$$

$$= 25$$

Por lo tanto

$$E(R) = \frac{2}{5000}\left[\frac{400}{9} + 25\right]$$

$$= .278$$

Ahora que ya hemos obtenido E(R) para las tres alternativas, analicemos los datos que podemos obtener de la sección I.3,

$\mu = 60$ $\lambda_p = 60$

$CS = 500$ $C(R) = 1200$

Lo único que nos falta obtener es el costo fijo, y el Presidente estima que el costo fijo por cada almacen por únidad de tiempo es de $ 400.00.

Por lo tanto,

$$E(CT) = n\ (\ (400 + 500\ s) + 1200\ L + (60/n)\ 1200\ E(R)\ )$$

La siguiente tabla nos indica cual es la alternativa que nos ofrece el minimo global:

n	λ	s	L	E(R)	CF+s CS	E(CE)	CR E(R)	E(CT)
1	60	1	∞	.04	900	∞	2 880	∞
1	60	2	1.333	.04	1 400	1 600	2 880	5 880
1	60	3	1.045	.04	1 900	1 254	2 880	6 034
2	30	1	1.000	.0278	1 300	1 200	708	6 416
2	30	2	.533	.0278	1 800	640	708	6 296
2	30	$\geq$ 3	$\geq$ 0	.0278	$\geq$ 2 300	$\geq$ 0	$\geq$ 708	$\geq$ 6 016
3	20	1	.500	.02	1 700	600	480	8 340
3	20	2	.343	.02	2 200	312	480	8 970
3	20	$\geq$ 3	$\geq$ [illegible]	.02	$\geq$ 2 700	$\geq$ 0	$\geq$ 480	$\geq$ 9 540

Por lo tanto la decisión que el Presidente de "Maderas de México" ha tomado es tener un depósito con dos servidores.

Despues de haber analizado en este capítulo la aplicación de la teoría de decisiones en los modèlos de colas, estudiaremos por último otros modelos de colas que no se basan en el proceso Nacimiento-Muerte.

OTROS MODELOS DE COLAS

CAPITULO VII

VII.1.- INTRODUCCION.-

En los capitulos anteriores, hemos visto modelos de colas que se basan en el proceso "Nacimiento-Muerte", por tanto se requiere que tanto sus tiempos de llegadas como los tiempos de servicio, tengan una distribución exponencial. Esta distribución nos proporciona un ajuste muy favorable para la gran mayoría de los sistemas de colas, pues la suposición de tiempos exponenciales entre llegadas, implica que las llegadas ocurren aleatoriamente (un proceso de entrada de Poisson), lo cual es razonable en muchos casos, pero no lo es cuando las llegadas se programan o regulan cuidadosamente. Del mismo modo, la distribución real de tiempos de servicio se desvía mucho de la forma exponencial cuando los clientes solicitan servicios bastante semejantes. Así pues, es necesario desarrollar otros modelos de colas que usen distribuciones alternativas.

Por desgracia, el estudio de estos modelos de colas con distribuciones no exponenciales es muy difícil y solo ha sido posible obtener algunos resultados para unos cuantos de estos modelos. El análisis de estos modelos no los trataremos en este trabajo (1), pero se resumirán los modelos y se describirán los resultados disponibles.

(1) Para los interesados en estos modelos pueden consultar el libro de Gross, Harris (1974), "Fundamentals of Queueing Theory", Wiley, Nueva York.

VII.2.-ENTRADA DE POISSON Y TIEMPOS DE SERVICIO INDEPENDIENTES.-

Supongamos que un sistema de colas tiene un proceso de entrada de Poisson con una tasa media de llegadas. Si suponemos (como en los anteriores casos) que los clientes tienen tiempos de servicios independientes, con una misma distribución, pero sin poner ahora restricción alguna para el tipo de distribución del tiempo de servicio, necesitamos únicamente conocer (o estimar) la media $1/\mu$ y la varianza σ de esta distribución.

Entonces, si $\rho = \lambda/\mu < 1$, si los resultados para este modelo general son los siguientes:

$$P_0 = 1 - \rho$$

$$Lq = \frac{\lambda^2 \sigma^2 + \rho^2}{2\,(1-\rho)}$$

$$L = \rho + Lq$$

$$Wq = Lq/\lambda$$

$$W = Wq + 1/\mu$$

Notese que Lq, L, Wq, W, aumentan conforme σ^2 se incrementa, esto nos indica que además de la velocidad promedio, la consistencia influye mucho sobre el rendimieto del medio de servicio (2).

(2) Los esfuerzos para deducir resultados semejantes aplicables al caso de varios servidores hasta el momento no han tenido exito.

VII.3.- ENTRADA DE POISSON Y TIEMPOS DE SERVICIO CONSTANTES.-

Cuando el servicio consiste en desarrollar una misma rutina para todos los clientes, tiende a haber poca variación en el tiempo de servicio, y de esta forma supondremos que todos los tiempos son iguales a una constante fija. Esto, junto con la suposición de un proceso de entrada de Poisson y un solo servidor, será un caso especial de el modelo anterior con $\sigma^2=0$, y así pues

$$L_q = \frac{\rho^2}{2\ (1-\rho)}$$

L, Wq, y W se obtienen a partir de Lq como se vió en la sección anterior.

Notese que Lq y Wq son la mitad del caso en que el tiempo de servicio es exponencial, en donde $\sigma^2 = 1/\mu^2$, así pues, decreciendo σ^2 mejora mucho las medidas de rendimiento de un sistema de colas.

VII.4.- ENTRADA DE POISSON Y TIEMPOS DE SERVICIO DE ERLANG.-

En el modelo anterior hemos visto la suposición de $\sigma=0$ en los tiempos de servico, mientras que en la distribución exponencial suponiamos que $\sigma = 1/\mu$, lo cual es suponer un alto grado de variabilidad. Entre estos dos casos $(0 < \sigma < 1/\mu)$, existe otra clase de distribución de tiempos de servicio llamada distribución Erlang (3).

(3) Nombre dado en honor de uno de los primeros investigadores de la Teoría de Colas

La función de densidad de probabilidad de la distribución Erlang es

$$f(t) = \frac{(\mu k)^k}{(k-1)!} t^{k-1} e^{-k\mu t} \quad t \geqslant 0$$

donde $\mu > 0$ y $k \in \mathbb{N}$.

Su media y su desviación estandar son:

$$\text{Media} = 1/\mu$$

$$\text{Desviación Estandar} = \frac{1}{\sqrt{k}} \frac{1}{\mu}$$

Por tanto k es el parámetro que especifica el grado de variabilidad de los tiempos de servicio con respecto a la media.

La distribución Erlang es muy importante en la teoría de las colas, pues si suponemos que T_1, T_2, . . . , T_k son k variables aleatorias independientes con una distribución exponencial, y cuya media es $1/k$, entonces su suma:

$$T = T_1 + T_2 + \ldots + T_k$$

tiene una distribución Erlang con parámetros μ y k. Esto se utiliza mucho cuando se dá la ocasión en que el servicio total requerido por un cliente hace que el servidor realize una misma tarea k veces.

Los resultados para el caso de un solo servidor con $\sigma^2 = 1/k\mu^2$ son los siguientes (4):

(4) Si se desea mas información ver paginas 275-284 de el libro citado en (1).

$$L_q = \frac{\lambda^2/k\mu^2 + \rho^2}{2(1-\rho)} = \frac{1+k}{2k} \cdot \frac{\lambda^2}{\mu(\mu-\lambda)}$$

$$W_q = \frac{1+k}{2k} \cdot \frac{\lambda}{\mu(\mu-\lambda)}$$

$$W = W_q + 1/\mu$$

$$L = \lambda W$$

VII.5.- MODELOS DE COLAS SIN ENTRADA DE POISSON.-

Hasta ahora hemos tratado colas que suponen una entrada de Poisson, pero este supuesto seria violado si de alguna forma se consiguiera que se pudiesen programar o regular estas llegadas, o evitando de alguna forma una ocurrencia aleatoria.

En este caso, si los tiempos de servicio siguieran una distribución exponencial, obtendriamos tres modelos invirtiendo en los modelos anteriormente vistos en este capítulo los tiempos de llegada y de servicio. Desgraciadamente no se han podido desarrollar expresiones sencillas como las ya dadas en los anteriores modelos, pero se han tabulado algunas tablas que han sido el resultado de extensos trabajos (5).

(5) Si se desea más información ver páginas 321-329 de el libro citado en (1).

VII.6.- MODELOS DE COLAS CON DISCIPLINA DE PRIORIDAD.-

Estos son modelos en que la disciplina de la cola se basa en un sistema de prioridad, y por tanto, el orden en que se seleccionan los clientes se basa en un sistema de prioridad. Esto sucede a menudo en muchos sistemas de colas, los trabajos más urgentes se realizan antes que otros, o clientes importantes a los que se les da precedencia sobre otros, o en hospitales cuando se atiende primero a pacientes de mayor gravedad, etcetera. Por tanto, el uso de modelos de prioridad nos brinda una mejoría que resulta de mucha ayuda dentro de la Teoría de las Colas.

Por desgracia, la inclusión de prioridad ha hecho que el análisis matemático se complique bastante y que los resultados que hasta el momento se han podido obtener sean muy limitados, y casi todos se restringen al caso de un solo servidor.

El modelo que a continación estudiaremos, es para el caso de varios servidores, el modelo supondra que se tiene P clases de prioridades (la prioridad 1 supondra la prioridad mas alta y la P la mas baja), que siempre que un servidor queda libre empieza a atender a otro cliente de la cola, el cliente seleccionado será aquel cuya prioridad sea la mas alta dentro de la cola y que haya esperado el mayor tiempo. Para cada clase de prioridad se supone una entrada de Poisson y tiempos de servicio exponenciales. Por último, el modelo supone la hipótesis un tanto restrictiva que la tasa media de servicio $1/\mu$, es la misma para todas las clases de prioridad, aunque se permite que la tasa media de llegadas difiera entre las diferentes clases de prioridad.

Dividiremos nuestro estudio en dos ramas:

a) Prioridades No Aseguradas.

Esto significa que un cliente que este recibiendo el servicio no puede ser desplazado si entra al sistema de colas un cliente de prioridad mayor, esto es, si un servidor ha comenzado a servir a un cliente, el servicio debe completarse sin interrupción.

En este caso, W_K nos determinará el tiempo esperado de permanencia en el sistema (incluyendo tiempo de servicio) para un cliente de la clase k, y esta dada por:

$$W_K = \frac{1}{A\ B_{K-1}\ B_K} + \frac{1}{\mu} \qquad k=1,2,\ldots,P$$

en donde:

$$A = s!\ \left[(s\mu - \lambda)/\rho^s\right] \sum_{j=0}^{s-1} \rho^j/j! + s\mu$$

$$B = 1$$

$$B = 1 - \frac{\sum_{i=1}^{k} \lambda_i}{s} \qquad k=1,2,\ldots,P$$

y

s = número de servidores

μ = tasa media de servicio por servidor ocupado

λ_i = tasa media de llegadas para la prioridad de clase i

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

$\rho = \lambda/\mu$

(En estos resultados se supone que $\sum_{i=1}^{k}\lambda_i < \mu$ de modo que la prioridad de clase k puede llegar a alcanzar una condición de estado estacionario).

Ademas, $L_k = \lambda_k W_k$

Con el fin de determinar el tiempo esperado de permanencia en la cola (excluyendo el tiempo de servicio para la prioridad de clase k, se restará $1/\mu$ de W_k; y para obtener la longitud esperada de la cola correspondiente se multiplica por λ_k.

b) Prioridades Aseguradas.

Esto quiere decir que si un cliente de prioridad inferior está recibiendo servicio será desplazado siempre que entre a la cola un cliente de prioridad superior, además, cuando un servidor tiene "éxito" en terminar un servicio, el nuevo cliente a quien empezará a prestar servicio será seleccionado de la manera anteriormente descrita, de modo que un cliente que es desplazado regresará a recibir el servicio, y despues de un número suficiente de intentos terminará por recibirlo.

En este caso:

$$W_k = \frac{1/\mu}{B_{k-1} \quad B_k} \qquad \text{si } k=1,2,\ldots,P$$

$$L_k = \lambda_k W_k$$

Los resultados correspondientes para el tiempo de permanencia en la cola (excluyendo el tiempo de servicio), y la longitud esperada de la cola, se obtienen a partir de W_k y L_k como ya se describio anteriormente (6).

(6) Si se desea más información consultar páginas 207-221 de el libro citado en (1).

CONCLUSIONES

CONCLUSIONES

Los sistemas de colas predominan en toda la sociedad, y lo adecuado que estos sistemas se encuentren puede tener un efecto sobre la calidad de la la productividad y el mejoramiento de nuestro modo de vivir.

La teoría de colas estudia los sistemas de colas planteando modelos matemáticos de su operación, y a continuación usa estos modelos con el fin de obtener mejores medidas de rendimiento. Este estudio proporciona información vital para diseñar con efectividad sistemas de colas que logren un balance apropiado entre el costo de proporcionar un servicio y el costo de la espera por ese servicio.

En este trabajo se han presentado la mayor parte de los modelos de colas, para los cuales se dispone de resultados útiles, sin embargo, estos modelos están muy lejos de agotar la lista. Aún quedan muchos modelos de colas potencialmente útiles que, hasta la fecha no han podido tratarse matemáticamente. Cuando no se puede obtener un modelo que proporcione una representación razonable del sistema de colas bajo estudio, un procedimiento común es obtener los datos apropiados sobre el rendimiento y utilizar el método de simulación.

Hemos visto también la importancia de la aplicación de la teoría de decisiones dentro de la teoría de colas, y se han presentado los casos mas comunes de decisión, no obstante, faltaron por analizar otras varibles de decisión mas complicadas como es el caso de calcular el tamaño de una sala para un sistema de colas, el diseño de un sistema de colas con prioridad, etc.

Los modelos recorrido-tiempo también pueden ser muy útiles cuando los servidores deben ir hasta el cliente desde el medio de servicio (como es el caso de camiones de bomberos o ambulancias), asi como también en otros modelos.

Otra área potencialmente útil para la aplicación de la teoría de colas es el desarrollo de políticas para controlar los sistemas de colas, por ejemplo, para ajustar dinamicamente el número de servidores o la tasa de servicio, con el fin de compensar los cambios en el número de clientes en el sistema. Se esta conduciendo una investigación considerable en esta área, de manera que pronto la teoría y los resultados de cálculo deben haber progresado lo suficiente como para poder hacer uso de ellos con amplitud.

La teoría de colas ha probado ser una herramienta muy útil y podemos estar seguros de que su uso seguirá creciendo, como reconocimiento a las muchas maneras en que se estan desarrollando los sistemas de colas.

BIBLIOGRAFIA

BIBLIOGRAFIA

Gross, Harris,
"Fundamentals of Queuing Theory";
Wiley
Nueva York, 1974.

Newell, Gordon F.,
"Applications of Queuing Theory";
Chapman and Hall
Londres, 1971.

Cooper, Robert B.
"Introduction to Queuing Theory";
Mc Millan,
Nueva York, 1978.

Harris W.,
"Introduccion a la Teoria de Colas";
Mc Graw Hill,
Nueva York, 1981.