TEORÍA DE COLAS

- Elementos básicos de los sistemas de colas.
- Patrones de servicio y llegadas.
- Análisis de sistemas de colas simples. Procesos de llegada poissonianos. Procesos de nacimiento y muerte.

INTRODUCCIÓN

El objetivo de la Teoría de Colas es el estudio matemático del comportamiento de las colas o líneas de espera, se enmarca en el área de la Investigación Operativa.

Las Colas son un desequilibrio temporal cuando la demanda de un servicio es superior a la capacidad del sistema. En la formación de colas se habla de clientes. Esta teoría estudia factores como el tiempo de espera medio en las colas o la capacidad de trabajo en el sistema sin que se llegue a colpasar.

Se encuentran en una amplia variedad de situaciones: comercio, computación, industria, informatica, ingenierías, internet, logística, negocios, telefonía, transporte, telecomunicaciones.

La Teoría de Colas proporciona información para la toma de decisiones de capacidad óptima y diseño de sistemas.

TEORÍA DE COLAS

Una Cola se presenta con frecuencia cuando se solicita un servicio por parte de una serie de clientes y tanto el servicio como los clientes son de tipo probabilístico.

La primera aplicación de teoría de Colas se debe al matemático danés Erlang sobre conversaciones telefónicas en 1909 para el cálculo de tamaño de las centralitas.

La Teoría de Colas es una disciplina de Investigación Operativa que se encarga de poponer modelos para el manejo eficiente de *Líneas de Espera*.

Una *Línea de Espera* es una hilera formada por uno o varios clientes que aguardan para recibir un servicio. Los clientes pueden ser personas, objetos, máquinas que requieren un mantenimiento, contenedores de mercancías para ser embarcados, elementos de inventario para ser utilizados, etc.

Una *Línea de Espera* se forma por un desequilibrio temporal entre la demanda de un servicio y la capacidad del sistema para gestionarlo.

Los *Modelos de Líneas de Espera* son muy útiles para determinar cómo operar un sistema de colas de la manera más eficaz, permiten encontrar un balance adecuado entre el costo de servicio y la cantidad de espera: Proporcionar demasiada capacidad de servicio para operar el sistema implica costos excesivos. De otra parte, si no se cuenta con suficiente capacidad de servicio surgen esperas excesivas con desafortunadas consecuencias.

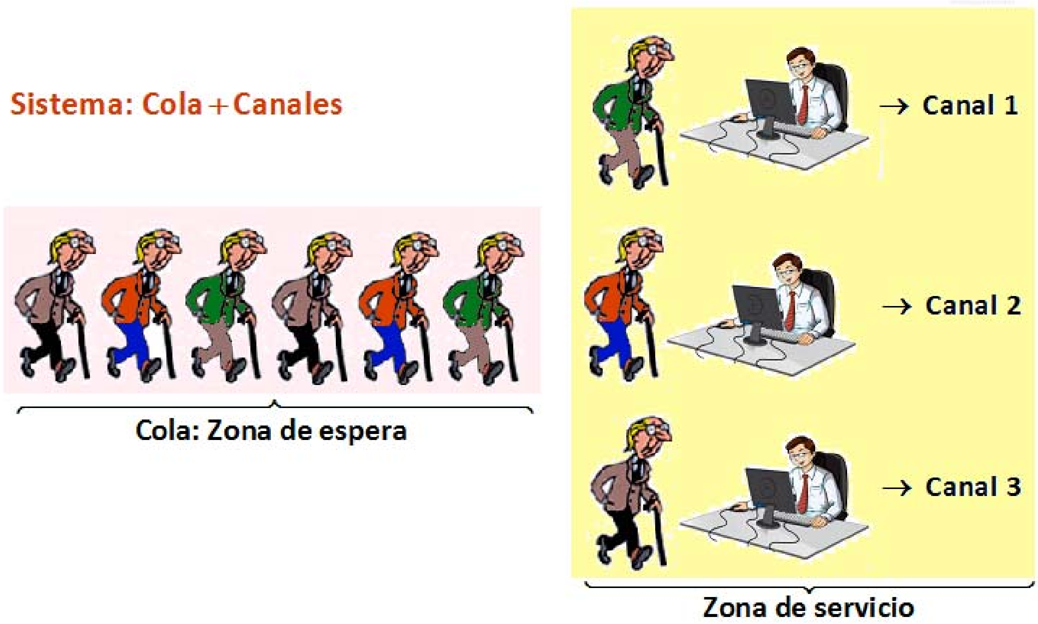
ESTRUCTURA DE LOS PROBLEMAS DE LÍNEAS DE ESPERA: Aunque cada situación específica tiene características diferentes, cuatro elementos son comunes a toda *Línea de Espera*:

- Una población de clientes que genera clientes potenciales.
- Una línea o fila de espera formada por los clientes.
- La instalación del servicio, formada por una persona (o un equipo), una máquina (o grupo de máquinas) que se requiere para proveer el servicio que el cliente solicita.
- Una regla de prioridad para seleccionar al siguiente cliente que será atendido por la instalación de servicio.

El término 'cliente' se utiliza en un sentido general, pudiendo ser una persona, piezas esperando su turno para ser procesadas, una lista de trabajo esperando para ser impresas en una impresora de red, etc.

PROCESO BÁSICO DE COLAS

Cola (Zona de espera): En todo sistema los flujos de entrada y salida no están sincronizados. La cola es una acumulación de clientes (personas, productos, objetos) que están a la espera de ser servidos.
Una cola puede evitarse:
(a) Tasa media de llegadas $<$ Capacidad de servicio
(b) Cuando se tiene un control sobre la dispersión de los tiempos de llegadas y la de los tiempos de servicio.



Básicamente la mayoría de los modelos de colas consiste: Los *clientes* que requieren un servicio se generan en el tiempo en una *fuente de llegada*, después entran al *sistema* y se unen a una *cola*.

En determinado momento se selecciona a un cliente de la cola para proporcionarle el servicio mediante alguna regla conocida como *disciplina de la cola*. Se lleva a cabo el servicio que el cliente requiere mediante un *mecanismo de servicio*, y después el cliente sale del sistema de colas.

- En el sistema se puede actuar en las siguientes características:

(a) Ley que rige las llegadas.
(b) Disciplina de la cola.
(c) Ley que determina el servicio (elección entre tipo y número de canales).

Consideracions sobre los sistemas:

(a) Cuando hay varios canales en paralelo es conveniente mantener una cola única.
(b) Cuando hay tiempos de servicios muy dispares para los diferentes clientes que forman la cola conviene establecer canales separados.
(c) Si la cola aumenta hasta cierto limite conviene aumentar la capacidad de los canales.

- Con relación a la *disciplina de la cola* hay que considerar:

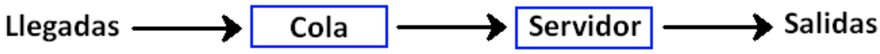
(a) Llegadas individuales o en grupos.
(b) Dependencia del múmero de clientes que pretenden incorporarse al sistema en función del número de clientes que se encuentran en el mismo.
(c) Disuasión de los clientes que pretenden incorporarse al sistema en función de la longitud de la cola (con una cierta probabilidad).

- En relación con el servicio hay que considerar:

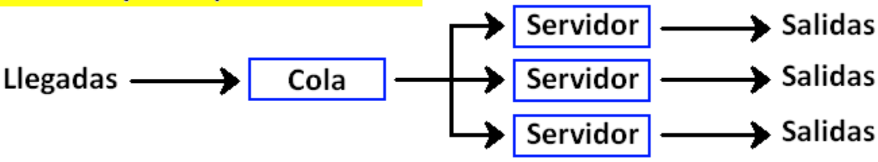
(a) Ley de servicio única a lo largo del tiempo y para todos los clientes.
(b) Ley de servicio variable en función del tipo de cliente y longitud de la cola.

TIPOS DE SISTEMAS DE COLAS

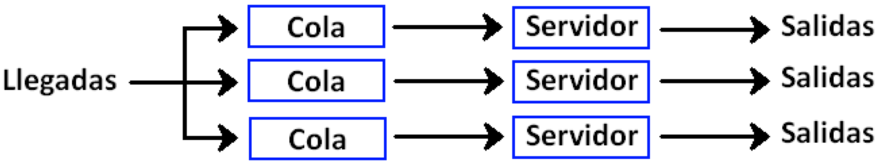
Una Cola y un Servidor



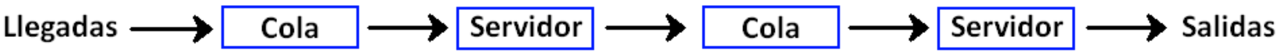
Una Cola y múltiples Servidores



Varias Colas y múltiples Servidores



Una Cola y Servidores secuenciales



CARACTERÍSTICAS DE LOS SISTEMAS DE COLAS

Un sistema de colas se describe adecuadamente con seis característica:

- Fuente de llegada de clientes.
- Patrón de servicio de servidores.
- Disciplina de cola.
- Capacidad del sistema.
- Número de canales de servicio.
- Número de etapas de servicio.

Fuente de llegada de clientes: En situaciones de colas habituales, la llegada de clientes es estocástica, esto es, depende de una variable aleatoria, con lo que se necesita conocer la distribución probabilística entre dos llegadas sucesivas de clientes.
La fuente de llegada puede variar con el tiempo, cuando se mantiene constante se dice estacionaria, si varía (por ejemplo, con las horas del día) se llama no-estacionaria.
Pueden contemplarse distintas situaciones: Clientes que llegan independiente o simultáneamente (llegan lotes), en este último caso hay que definir su distribución probabilística. Clientes que abandonan la cola por ser demasiado larga o que tras esperar mucho abandonan.
El supuesto normal, para un modelo básico de colas, es que la llegada de clientes hasta un momento específico sigue una distribución de Poisson, aunque no sea la única distribución que puede considerarse.

Patrón de servicio de servidores: Pueden presentar un tiempo de servicio variable, en cuyo caso hay que asociar una función de probabilidad. Pueden atender en lotes o de modo individual.
El tiempo de servicio puede variar con el número de clientes en la cola, trabajando más rápido o más lento, en este caso se conoce como patrones de servicio dependientes.
El patrón de servicio puede ser no-estacionario variando con el tiempo transcurrido.

Disciplina de cola: Es la regla en el orden que se van a seleccionar los clientes que se encuentran a la espera de ser atendidos en la cola, existen varias reglas, entre las más comunes se pueden encontrar:

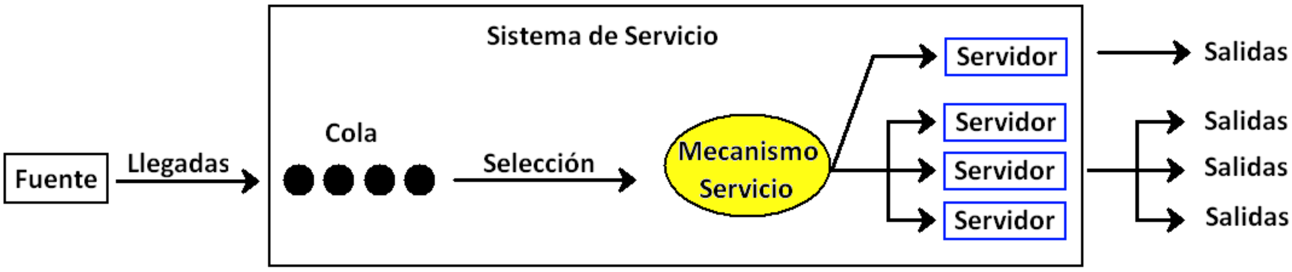
♠ FIFO (first in first out): Se atiende al cliente en el orden que llegan a la cola, el primero en llegar será el primero en ser atendido. En los modelos básicos de colas se supone como normal la disciplina de primero en entrar, primero en salir, a menos que se establezca de otra manera.
♠ LIFO (last in first out): Consiste en atender primero al que ha llegado de último, también se le conoce como 'pila'.
♠ RSS (random selection of service): Se selecciona a los clientes de una cola de forma aleatoria, con algún procedimiento de prioridad o algún otra preclasificación.
♠ Processor Sharing: Todos los clientes experimentan con eficacia el mismo retraso, ya que comparten entre todos los clientes de la cola la capacidad del sistema atendiendo a todos por igual.

Capacidad del sistema: Es el número máximo de clientes que pueden estar dentro del sistema haciendo cola antes de ser atendidos para recibir el servicio, al igual que la fuente de llegada este número puede ser finito o infinito.

Número de canales de servicio: Es preferible utilizar sistemas multiservicios con una única línea de espera para todos que con una cola por servidor. Al hablar de canales de servicio paralelo se trata generalmente de una cola que alimenta a varios servidores.

Número de etapas de servicio: Puede ser unietapa o multietapa, en este último el cliente puede pasar por un número de etapas mayor que uno. En algunos sistemas multietapa se admite la vuelta atrás o reciclado, modo habitual en sistemas productivos como controles de calidad y reprocesos.

MECANISMO DE SERVICIO



Se caracteriza por las *estaciones de servicio* (servidores o dependientes) y por los *canales de servicio* que desembocan en cada uno de los servidores.

Una única cola puede desembocar en varios servidores que van siendo ocupados de acuerdo a una disciplina de selección, el caso habitual de asignación al primer servidor que queda libre.

Puede haber multiplicidad en el número de servidores, es posible encontrar múltiples colas que surtan clientes a un único o a múltiples servidores.

En una determinada estación de servicio, el cliente entra en uno de estos canales y el servidor le presta el servicio completo.

Los modelos de colas deben especificar el número de estaciones de servicio (canales de servicio en serie) y el número de servidores (canales paralelos) en cada una de ellas. Los modelos elementales se componen de una estación, ya sea con un servidor o con un número finito de ellos.

La variable más importante que caracteriza el mecanismo de servicio es el tiempo de servicio es el *tiempo de servicio*. Se denomina *tiempo de servicio* el que transcurre desde el inicio del servicio para el cliente hasta su terminación en una estación.

El modelo de un sistema de colas debe especificar la distribución de probabilidad de los *tiempos de servicio* de cada servidor, siendo habitual suponer la misma distribución para todos los servidores.

La distribución de servicio que más se emplea en la práctica (por ajustarse a un gran número de situaciones como por su simplicidad en el cálculo) es la *distribución exponencial*. Otras distribuciones qie se utilizan son la *distribución degenerada* (para tiempos de servicio constantes) y la *distribución de Erlang* (*Gamma*) para combinaciones de distribuciones exponenciales.

COSTES DE COLA Y SERVICIO

Los costes asociados a dos componentes fundamentales del sistema de colas:

COLA COSTE DE ESPERA	$\Rightarrow$	SERVICIO COSTE DE SERVICIO

COSTE DE ESPERA: Coste del tiempo que se produce cuando los clientes tienen que esperar, como un activo que no está siendo utilizado repercute negativamente, afectando a la imagen de la empresa o sistema, por lo que se considera de forma económica.

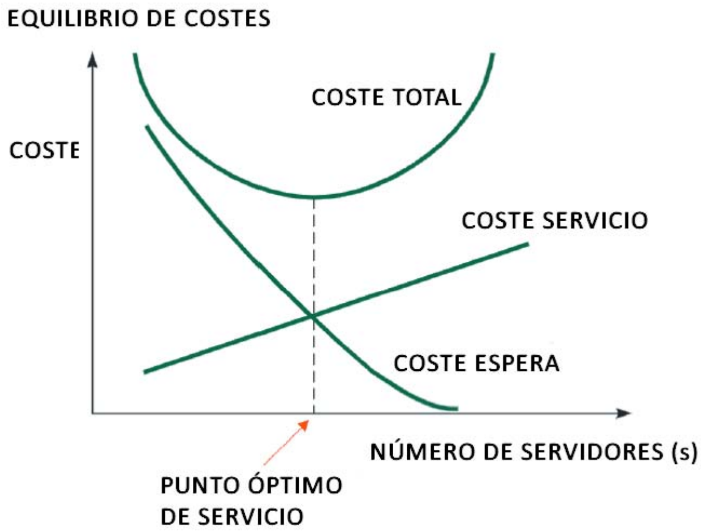
COSTE DE SERVCIO: Se integra en el conjunto de gastos de funcionamiento del servicio (coste de insatalaciones, maquinaria, mantenimiento, personal, etc.).
Se calcula multiplicando el coste de servicio por el número de servidores que componen el sitema: Coste total de servicio $= s \,.\, C_s$

EQUILIBRIO DE COSTES

La Teoría de Colas tiene como objetivo que estos costes sean los mínimos posibles, con tasas de servicio bajas, los costes de espera son altos.

Por el contrario, con tasas de servicio altas, los costes de espera son bajos.

La finalidad del estudio de costes es encontrar una combinación adecuada donde el coste total sea el mínimo.



Siendo, $c_q \equiv$ coste espera cola y $c_s \equiv$ coste servicio, se tiene:

Coste total de espera: $CQ = c_q \,.\, L_q$

Coste total servicio: $CS = c_s \,.\, s$

Coste total en sistema: $CTS = c_q \,.\, L_s + c_s \,.\, s$

DISTRIBUCIONES DE PROBABILIDAD DEL MECANISMO DE SERVICIO

Las fuentes de variación en los problemas de colas o filas de espera provienen del carácter aleatorio de la llegada de clientes y de las variaciones que se registran en los distintos tiempos de servicio. Generalmente cada una de esas fuentes suele describirse mediante una distribución de probabilidad.

La mayor parte de los modelos de colas estocásticas asumen que el tiempo entre diferentes llegadas de clientes siguen una distribución exponencial $\text{Exp}(\lambda)$, o lo que es equivalente, que el ritmo de llegadas sigue una distribución de Poisson.

También es habitual admitir que el ritmo de atención al cliente cuando el servidor está ocupado sigue una distribución de Poisson y la duración de la atención al cliente sigue una distribución exponencial.

La distribución de llegadas: $p_n(t) = \dfrac{(\lambda t)^n}{n!}\, e^{-\lambda t} \quad , \; n = 0, 1, 2, \cdots$ clientes

$p_n(t) \equiv$ Probabilidad de que *n* clientes estén en el sistema en el tiempo t

El tiempo entre llegadas, se define como la probabilidad de que no llegue ningún cliente, es decir: $p_0(t) = \dfrac{(\lambda t)^0}{0!}\, e^{-\lambda t} = e^{-\lambda t}$, siendo una distribución exponencial.

El uso del patrón de servicio (llegada) tiene, entre otras, las propiedades:

- Los datos que definen un proceso de Poisson vienen dados por el número medio de llegadas.

- Si el número de llegadas sigue una distribución de Poisson $P(\lambda)$, el tiempo entre llegadas sigue una distribución exponencial de media $(1/\lambda)$ y lo contrario, es decir, $p_n(t) = \dfrac{(\lambda t)^n}{n!}\, e^{-\lambda t} \;\Leftrightarrow\; p_0(t) = e^{-\lambda t}$

- La distribución exponencial tiene amnesia: La probabilidad de que falten t unidades para que llegue el siguiente cliente es independiente de cuanto tiempo transcurra sin que llegue ningún cliente: $P_r\left(T \leq 1 \;\middle|\; T \geq t_0\right) = P_r(0 \leq T \leq t_1 - t_0)$

- El número de llegadas en intervalos de tiempo no superpuestos es estadísticamente independiente.
- La probabilidad de que una llegada ocurra en el tiempo t y $(t + \Delta t)$ es $\lambda t + o(\Delta t)$ donde $\lambda \equiv$ tasa de llegada y $o(\Delta t)$ verifica que $\lim\limits_{\Delta t \to 0} \dfrac{o(\Delta t)}{\Delta t} = 0$, donde $o(\Delta t)$ se puede tomar como la probabilidad de que llegue más de un cliente.

- La distribución del número de llegadas en intervalos de tiempo iguales es equivalente

$$p_n(t-s)=\frac{(\lambda[t-s])^n}{n!}\,e^{-\lambda[t-s]} \quad \forall\, t>s \;,\; t,s\geq 0$$

- Si el proceso de llegada sigue una distribución de Poisson $P(\lambda)$, los tiempos de llegada son aleatorios con una función de probabilidad uniforme sobre el período analizado: $f_T\left(t_1, t_2, \ldots, t_k \,\middle|\, k \text{ llegadas en } [0,T]\right)=\frac{k!}{T^k}$

- λ puede variar con el tiempo, con lo que $p_n(t)=e^{-m(t)}\,\frac{\left(m(t)\right)^n}{n!}$ donde $m(t)=\int_0^t \lambda(s)\,ds$

- Llegadas múltiples: En cada evento de llegada aparecen i clientes, siendo $\lambda=\sum_{i=1}^{n}\lambda_i$

OTRAS DISTRIBUCIONES DE PROBABILIDAD TIPO DISCRETO:

BERNOUILI: Si la probabilidad de cada ocurrencia es diferente, la más sencilla es la distribución de Bernouilli donde la variable sólo puede tomar dos valores excluyentes (verdadero, falso) con cierta probabilidad p de que ocurra un suceso (suele denominarse éxito) y una probailidad $q=1-p$ de que ocurra el suceso contrario (fracaso). La media de una distribución de Bernouilli es p y la varianza pq

BINOMIAL: La distribución Binomial, se denota $B(n,p)$, representa la probabilidad de obtener k éxitos con probabilidad p, a partir de n intentos. En consecuencia, es la suma de n pruebas de Bernouilli. La media de una distribución binomial es np y la varianza npq, siendo $p+q=1$

GEOMÉTRICA: La distribución Geométrica, se denota $G(p)$, representa la probabilidad de obtener la primera ocurrencia de un suceso en *n* pruebas o ensayos. La función de cuantía es $P(X=n)=q^{n-1}\,p$, la media es $\mu_X=1/p$ y la varianza $\sigma_X^2=q/p^2$ siendo $p+q=1$

BINOMIAL NEGATIVA: La distribución Binomial negativa, se denota por $BN(n,p)$, representa el número de pruebas en las que aparece un suceso hasta la *n-ésima* aparición del suceso contrario. Es un modelo adecuado para tratar aquellos procesos en los que se repite un determinado ensayo o prueba independiente, con resultados excluyentes, hasta conseguir un número determinado de resultados favorables (por vez primera).

La función de cuantía es $P(X=k)=\binom{n+k-1}{k} p^n\, q^k$

La media de una distribución Binomial negativa es $(n\,q\,/\,p)$ y la varianza $n\,q\,/\,p^2$

Una distribución Binomial negativa $BN(1,p)$ es una distribución geométrica.

<u>OTRAS DISTRIBUCIONES DE PROBABILIDAD TIPO CONTINUO</u>:

ERLANG: Es una distribución de probabilidad continua de parámetros k y λ, suma de k variables aleatorias independientes exponenciales de parámetro λ, cuya función de densidad es $f(x)=\dfrac{(\lambda\,x)^{k-1}}{(k-1)!}\lambda\, e^{-\lambda x}\quad \forall x,\ \lambda>0$

La distribución de Erlang es el equivalente de la distribución Gamma con el parámetro $k=1,2,\ \ldots$ y $\lambda=1/\theta$. Para $k=1$ es la distribución Exponencial.

Para describir el tiempo de espera hasta el suceso número *k* en un proceso de Poisson se toma la distribución de Erlang con un parámetro $\theta=\dfrac{1}{\lambda}$

$$E(X)=\frac{k}{\lambda}\qquad V(X)=\frac{k}{\lambda^2}$$

WEIBULL: La distribución de Weibull se utiliza habitualmente para describir el tiempo que transcurre entre dos averías consecutivas de la misma máquina.

El modelo Weibull generaliza al modelo exponencial.

La media de una distribución de Weibull es $E(T)=n\,\Gamma\left(\dfrac{1}{\beta}+1\right)$, la varianza

$\sigma_T^2=\eta\left[\Gamma\left(\dfrac{2}{\beta}+1\right)-\Gamma\left(\dfrac{1}{\beta}+1\right)^2\right]$ y la función de fiabilidad es $R(t)=e^{-(t/\eta)^{\beta}}$

La función de densidad y función de distribución son:

$$f(t)=\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} e^{-(t/\eta)^{\beta}}\qquad F(t)=1-e^{-(t/\eta)^{\beta}}\qquad t\geq 0\ ,\ \beta>0\ \text{ y }\ \eta>0$$

denotando $\eta\equiv$ parámetro de escala y $\beta\equiv$ parámetro de forma

$$\text{Gamma}\ \ \Gamma(p)=\int_0^{\infty} x^{p-1}\,e^{-x}\,dx\qquad \Gamma(p)=(p-1)!\quad \Gamma(1/2)=\sqrt{\pi}\quad \Gamma(p)\Gamma(1-p)=\frac{\pi}{\operatorname{sen} p\pi}$$

$$\text{Beta}\ \ \beta(p,q)=\int_0^{1} x^{p-1}\,(1-x)^{q-1}\,dx\quad \beta(p,q)=\beta(q,p)=\frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}\quad \beta(p,q)=\int_0^{\infty}\frac{t^{p-1}}{(1+t)^{p+q}}\,dt$$

LOG-NORMAL: La distribución Log-Normal se utiliza para describir el tiempo que se utiliza para la reparación de máquinas.

Una variable X se dice que tiene una distribución log-Normal si los logaritmos neperianos de sus valores están normalmente distribuidos, es decir,

si la variable $\eta = \operatorname{Ln} X$ es $N(\mu, \sigma)$

La función de densidad es: $f(x) = \dfrac{1}{x\,\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\frac{(\operatorname{Ln} x - \mu)^2}{\sigma^2}} \qquad x > 0$

La media y varianza son: $\mu_X = e^{\mu + \sigma^2/2} \qquad \sigma_X^2 = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$

PROCESOS DE POISSON

Si los tiempos entre llegadas/servicios de clientes se distribuyen según una exponencial $\operatorname{Exp}(\lambda)$, el número de llegadas/servicios de clientes hasta un cierto tiempo es un proceso de Poisson.

- Sea la variable $X \sim \operatorname{Exp}(\lambda)$ la variable aleatoria entre llegadas o tiempo de servicio,

su función de densidad $f(x) = \begin{cases} \lambda\, e^{-\lambda x} & \text{para } x \geq 0 \\ 0 & \text{en otro caso} \end{cases}$ es estrictamente decreciente.

La función de distribución: $F(x) = \displaystyle\int_{-\infty}^{x} f(t)\, dt = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

$X \sim \operatorname{Exp}(\lambda) \;\; \rightarrow \;\; E(X) = \dfrac{1}{\lambda} \quad V(X) = \dfrac{1}{\lambda^2}$

La distribución de probabilidad del tiempo que falta para que ocurra el evento es siempre la misma independientemente del tiempo que haya pasado.

$$P(X > t + s \mid X > s) = \frac{P\big[(X > t + s) \cap (X > s)\big]}{P(X > s)} = \frac{P(X > t + s)}{P(X > s)} = P(X > t)$$

En consecuencia, la distribución exponencial $\exp(\lambda)$ carece de memoria y es la única distribución continua con tal propiedad, ya que

$$P(X > t + s) = e^{-\lambda(t+s)} = e^{-\lambda t}\, e^{-\lambda s} = P(X > t)\, P(X > s)$$

- La suma de procesos de entrada de Poisson es también un proceso de Poisson siendo la tasa la suma de las tasas respectivas.

- Si las llegadas a un sistema son de tipo Poisson con tasa λ y cada llegada es encaminada a un subsistema s_i con una probabilidad p_i el proceso de llegada a cada subsistema es también de Poisson con tasa $\lambda\, p_i$

MODELO GENERAL: PROCESOS DE NACIMIENTO Y MUERTE

En la mayor parte de los modelos elementales de colas es común que las entradas y salidas del sistema ocurran según un proceso de nacimiento y muerte. El proceso explica cómo varía el estado del sistema $N(t)$ al aumentar t.
En este contexto,

$N(t)$ estado del sistema en tiempo $t \equiv$ Número de cliente en el sistema
Nacimiento $\equiv$ Llegada de clientes al sistema
Muerte $\equiv$ Salida de clientes una vez servidos

- La distribución del tiempo que falta para la llegada es exponencial $\text{Exp}(\lambda_n)$ con $n = 0, 1, 2, \ldots$, siendo λ_n la tasa de llegada de clientes al sistema.

- La distribución del tiempo que falta para la salida es exponencial $\text{Exp}(\mu_n)$ con $n = 0, 1, 2, \ldots$, siendo μ_n la tasa de salida de clientes del sistema.

- Hay independencia entre el tiempo hasta próxima llegada y tiempo hasta próxima salida.

La transición del estado será: $\begin{cases} n \;\; \rightarrow \;\; n+1 \;\; \text{un nacimiento} \\ n \;\; \rightarrow \;\; n-1 \;\; \text{una muerte} \end{cases}$

Tomando estos supuestos, el proceso es un tipo especial de cadena de Markov de tiempo continuo.

Los parámetros λ_n y μ_n son tasas medias en la distribución exponencial, en ocasiones estos valores son constantes $\forall\, \mu$

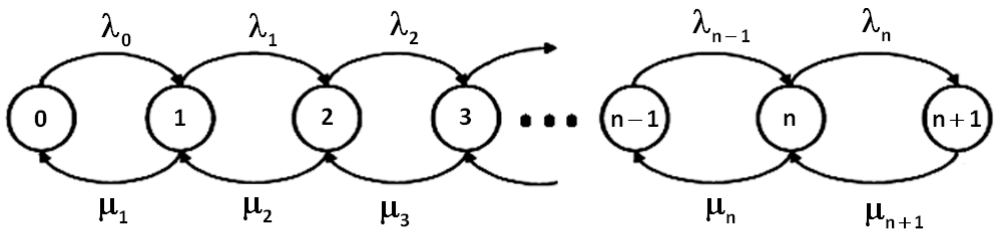
La llegada como la salida son procesos de Poisson e independientes, luego de un estado dado se puede pasar a dos posibles estados.

Tasa media de llegada al estado n $\equiv \lambda_{n-1}\, p_{n-1} + \mu_{n+1}\, p_{n+1}$

Tasa media de salida del estado n $\equiv \lambda_n\, p_n + \mu_n\, p_n$

$p_n \equiv$ Probabilidad de que haya n clientes en el sistema de manera estacionaria

Por ser el sistema estacionario, la tasa media de llegada es igual a la tasa media de salida para cualquier estado n, es decir, $\lambda_{n-1}\, p_{n-1} + \mu_{n+1}\, p_{n+1} = \lambda_n\, p_n + \mu_n\, p_n$



$$\boxed{\lambda_{n-1}\, p_{n-1} + \mu_{n+1}\, p_{n+1} = \lambda_n\, p_n + \mu_n\, p_n}$$

Diagrama de tasas del proceso de nacimiento y muerte

Ecuaciones del balance o de equilibrio:

$$n=0: \quad \mu_1 p_1 = \lambda_0 p_0 \quad \rightarrow \quad p_1 = \frac{\lambda_0}{\mu_1}\, p_0$$

$$n=1: \quad \lambda_0\, p_0 + \mu_2 p_2 = \lambda_1\, p_1 + \mu_1\, p_1 \quad \rightarrow \quad p_2 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2}\, p_0$$

$$n=2: \quad \lambda_1\, p_1 + \mu_3 p_3 = \lambda_2\, p_2 + \mu_2\, p_2 \quad \rightarrow \quad p_3 = \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3}\, p_0$$

por recurrencia, $p_n = \dfrac{\lambda_0 \lambda_1 \lambda_2 \, \ldots \, \lambda_{n-1}}{\mu_1 \mu_2 \mu_3 \, \ldots \, \mu_n}\, p_0 \quad \mapsto \quad \displaystyle\sum_{n=0}^{\infty} p_n = 1$

$$\sum_{n=0}^{\infty} \frac{\lambda_{n-1} \cdots \lambda_2 \lambda_1 \lambda_0}{\mu_n \cdots \mu_3 \mu_2 \mu_1}\, p_0 = 1 \quad \rightarrow \quad p_0 = \frac{1}{\displaystyle\sum_{n=0}^{\infty} \frac{\lambda_{n-1} \cdots \lambda_1 \lambda_0}{\mu_n \cdots \mu_2 \mu_1}} = \frac{1}{1 + \displaystyle\sum_{n=1}^{\infty} \frac{\lambda_{n-1} \cdots \lambda_1 \lambda_0}{\mu_n \cdots \mu_2 \mu_1}}$$

El procedimiento para resolver estas ecuaciones no es otro que despejar todas las variables en términos de una de ellas, siendo la más conveniente p_0.

Para simplificar la notación se denota por C_n al multiplicador de p_0:

$$C_n = \frac{\lambda_{n-1} \lambda_{n-2} \cdots \lambda_1 \lambda_0}{\mu_n\, \mu_{n-1} \cdots \mu_2 \mu_1}$$

Para $n=0$ se define $C_n = 1$

La expresión de la probabilidad de estado estable es: $p_n = C_n \times p_0$

$$\sum_{n=0}^{\infty} p_n = 1 \quad \rightarrow \quad \left(\sum_{n=0}^{\infty} C_n \right) \times p_0 = 1 \quad \Rightarrow \quad p_0 = \frac{1}{\displaystyle\sum_{n=0}^{\infty} C_n}$$

Estos resultados son de estado estable pues al aplicar el limite cuando t tiende a infinito hay el mayor alejamiento del momento inicial y se desarrollan bajo el supuesto de que los parámetros λ_n y μ_n tienen valores con los cuales el proceso puede alcanzar la condición de estado estable.

En un puesto del mercado se compran y venden mentiras, los clientes entran según un proceso de Poisson de tasa 4 clientes/minuto y la permanencia en el puesto es un tiempo exponencial con un tiempo medio de 10 minutos. Suponiendo que la capacidad del puesto es infinita para recibir a clientes, se solicita:

a) Probabilidades de equilibrio
b) Número medio de clientes en el puesto del mercado

Solución:

a) Sea $X(t) \equiv$ Número de personas en el mercado en el instante t

$\{X(t), t \geq 0\}$ es un proceso de nacimiento y muerte

Tasas de entrada (nacimientos): $\lambda_j = 4 \quad \forall j \geq 0$ clientes/minuto

Tasas de salida (muertes): $\mu_j = \left(\frac{1}{10}\right) x\, j \quad \forall j \geq 0$ clientes/minuto

Ecuaciones de balance o equilibrio: $\lambda_{n-1}\, p_{n-1} + \mu_{n+1}\, p_{n+1} = \lambda_n\, p_n + \mu_n\, p_n$

$$n = 0: \quad \mu_1\, p_1 = \lambda_0\, p_0 \quad \rightarrow \quad p_1 = \frac{\lambda_0}{\mu_1}\, p_0 = \frac{4}{1/10}\, p_0 = 40\, p_0$$

$$p_2 = \frac{\lambda_0\, \lambda_1}{\mu_1\, \mu_2}\, p_0 = \frac{4 \times 4}{\frac{1}{10} \times \frac{2}{10}}\, p_0 = \frac{1}{2} 40^2\, p_0 = \frac{1}{2!}\, 40^2\, p_0$$

$$p_3 = \frac{\lambda_0\, \lambda_1\, \lambda_2}{\mu_1\, \mu_2\, \mu_3}\, p_0 = \frac{4 \times 4 \times 4}{\frac{1}{10} \times \frac{2}{10} \times \frac{3}{10}}\, p_0 = \frac{1}{6} 40^3\, p_0 = \frac{1}{3!} 40^3\, p_0$$

por recurrencia, $p_n = \frac{1}{n!} 40^n\, p_0 \quad \forall n \geq 1$

$$\sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty} \frac{1}{n!} 40^n\, p_0 = 1 \quad \rightarrow \quad p_0 = \frac{1}{\sum_{n=0}^{\infty} \frac{1}{n!} 40^n} = \frac{1}{1 + \sum_{n=1}^{\infty} \frac{1}{n!} 40^n} = e^{-40}$$

Se considera el desarrollo de Taylor $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!}$

$$e^{40} = 1 + \frac{40}{1!} + \frac{40^2}{2!} + \cdots + \frac{40^n}{n!} \quad \rightarrow \quad \sum_{n=0}^{\infty} \frac{1}{n!} 40^n = e^{40}$$

Finalmente, $p_n = \dfrac{1}{n!} 40^n \, p_0 = \dfrac{40^n}{n!} \; e^{-40} \quad n \geq 0 \quad \rightarrow$ distribución de Poisson $P(\lambda = 40)$

b) El número medio de clientes en el mercado:

$$L = \sum_{n=0}^{\infty} n\, p_n = \sum_{n=1}^{\infty} n \left(\frac{40^n}{n!} e^{-40} \right) = e^{-40} \sum_{n=1}^{\infty} \cancel{n} \frac{40^n}{\cancel{n}\,(n-1)!} = 40\, e^{-40} \sum_{n=1}^{\infty} \frac{40^{n-1}}{(n-1)!} =$$

$$= 40\, e^{-40} \sum_{n=1}^{\infty} \frac{40^{n-1}}{(n-1)!} = 40\, e^{-40} \sum_{k=0}^{\infty} \frac{40^k}{k!} = \; 40\, e^{-40}\, e^{40} = 40$$

🖫 En una oficina bancaria rural un empleado atiende a los clientes que llegan según porceso de Poisson de parámetro λ. Los clientes son reacios a esperar demasiado tiempo, de modo que si hay k personas esperando a ser atendidas con probabilidad $q(k)$ se quedan en otro caso se van. Una vez que entran en la cola esperan independientemente según su paciencia o abandonan un tiempo exponencial de tasa ϕ. Por otra parte, el tiempo de servicio del empleado es exponencial de tasa μ.
Se quiere modelar el número de personas en la cola como un proceso de nacimiento y muerte, calulando tasas y probabilidades de la variable estado.

Solución:

$X(t) \equiv$ Número de personas en la cola de espera

Los clientes llegan a la cola (nacimientos) según una distribución de Poisson de parámetro $\lambda\, q(k)$

Los clientes salen atendidos (muertes) con distribución de Poisson de tasa μ o por impaciencia con distribución de Poisson de tasa ϕ

Suponiendo que la variable $X(t)$ se comporta como una cadena de Markov de tiempo continuo, las probabilidades de transición se rigen por la ecuación:

$$\frac{dp_{ij}(t)}{dt} = \sum_{k \neq j} p_{ik}(t)\, \upsilon_k\, p_{kj} - \upsilon_j\, p_{ij}(t) \quad \text{donde} \quad \begin{cases} \upsilon_k \equiv \text{Tasa de salida del estado k} \\ \upsilon_j \equiv \text{Tasa de salida del estado j} \end{cases}$$

Se supone que existe una distribución estacionaria, basta con imponer que las tasas de muerte sean mayores que las taas de nacimiento.

Con distribuciones estacionarias, las probabilidades de transición no dependen del tiempo ni del estado inicial, es decir, $p_{ij}(t) \quad \to \quad p_j$

Las probabilidades de transición no cambian con el tiempo, la ecuación de equilibrio en cadenas de Markov en tiempo continuo:

$$\frac{dp_{ij}(t)}{dt} = 0 \quad \to \quad \sum_{k \neq j} p_k \upsilon_k p_{kj} - \upsilon_j p_j = 0 \quad \Rightarrow \quad \upsilon_j p_j = \sum_{k \neq j} p_k \upsilon_k p_{kj}$$

En consecuencia, los cambios de la variable $X(t)$ son a estados contiguos. En concreto, sí $X(t) = h$ los cambios inmediatos son a $(h-1)$ y a $(h+1)$.

Como los cambios son a estados contiguos, se puede modelar $X(t)$ como un proceso de Poisson de nacimiento y muerte.

Tasa de nacimiento: $\lambda(k) = q(k)$

Tasa de muerte: $\mu(k) = \begin{cases} \mu + \phi\, k & k > 0 \\ 0 & k = 0 \end{cases}$

Ecuaciones del balance o de equilibrio:

$$p_1 = \frac{\lambda(0)}{\mu(1)}\, p_0 = \frac{\lambda\, q(0)}{\mu + \theta}\, p_0 = \lambda \frac{q(0)}{\mu + \theta}\, p_0$$

$$p_2 = \frac{\lambda(1)}{\mu(2)}\, p_1 = \frac{\lambda\, q(1)}{\mu + 2\phi}\, p_1 = \frac{\lambda\, q(1)}{\mu + 2\phi}\, \frac{\lambda\, q(0)}{\mu + \theta}\, p_0 = \lambda^2\, \frac{q(0)}{\mu + \phi}\, \frac{\lambda\, q(1)}{\mu + 2\phi}\, p_0$$

$$p_3 = \frac{\lambda(2)}{\mu(3)}\, p_2 = \frac{\lambda\, q(2)}{\mu + 3\phi}\, p_2 = \frac{\lambda\, q(2)}{\mu + 3\phi}\, \frac{\lambda\, q(1)}{\mu + 2\phi}\, \frac{\lambda\, q(0)}{\mu + \theta}\, p_0 = \lambda^3\, \frac{q(0)}{\mu + \theta}\, \frac{q(1)}{\mu + 2\phi}\, \frac{q(2)}{\mu + 3\phi} p_0$$

por recurrencia, $p_n = \lambda^n \prod_{k=0}^{n-1} \frac{q(k)}{\mu + k\theta} p_0 \qquad n \geq 1$

donde $p_0 = \dfrac{1}{1 + \sum_{n=1}^{\infty} \left(\lambda^n \prod_{k=0}^{n-1} \frac{q(k)}{\mu + k\theta} \right)}$

En un almacén con capacidad para N artículos se guardan dos tipos de artículos, cada uno de ellos ejecutados en respectivos procesos. Las llegadas de los dos tipos de artículos siguen un proceso de Poisson con tasas respectivas λ_1 y λ_2. Los tiempos de ejecución de cada tipo de artículo son exponenciales, respectivamente, μ_1 y μ_2. Los artículos que llegan cuando el almacén está completo se derivan a otros almacenes. Se solicita un modelo que permita conocer a largo plazo la proporción de artículos que son derivados a otros almacenes.

Solución:

$X(t) \equiv \text{Número de artículos en el almacén}: \; X(t) = X_1(t) + X_2(t)$

$$\begin{cases} X_1(t) \equiv \text{Número de artículos tipo 1 almacenados} \\ p_k^1 \equiv \text{Probabilidad de que haya k artículos tipo 1 en el almacén} \end{cases}$$

$$\begin{cases} X_2(t) \equiv \text{Número de artículos tipo 2 almacenados} \\ p_{N-k}^2 \equiv \text{Probabilidad de que haya (N-k) artículos tipo 2 en el almacén} \end{cases}$$

La proporción de artículos que se derivan a otros almacenes:

$$p = P\big[X(t) > N\big] = 1 - P\big[X(t) \le N\big] = 1 - \sum_{k=0}^{N} P\Big[\big(X_1(t) = k\big) \cap \big(X_2(t) = N-k\big)\Big] = 1 - \sum_{k=0}^{N} p_k^1\, p_{N-k}^2$$

Suponiendo un estado estacionario, a largo plazo se cumple la ecuación de equilibrio entre las variables $X_1(t)$ y $X_2(t)$:

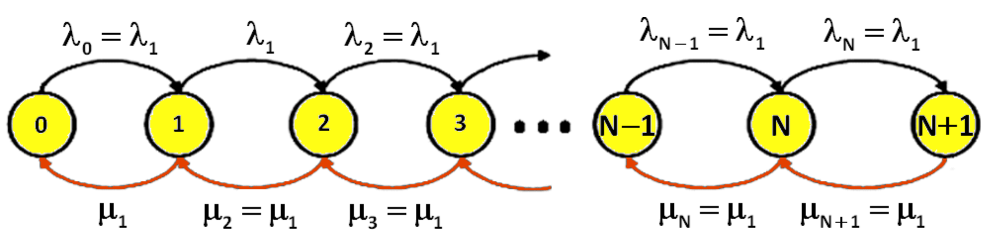
$$\upsilon_j\, p_j = \sum_{k \neq j} p_k\, \upsilon_k\, p_{kj}$$

Como los cambios de estado corresponden solo a los artículos (j) y $(j+1)$ las variables $X_1(t)$ y $X_2(t)$ corresponden a procesos de nacimiento y muerte.

Ecuación de equilibrio: $\big[\lambda_j + \mu_j\big]\, p_j = \lambda_{j-1}\, p_{j-1} + \mu_{j+1}\, p_{j+1}$

Variable $X_1(t)$:

Tasa de nacimiento $\lambda_j = \begin{cases} \lambda_1 & X < N \\ 0 & X \ge N \end{cases}$ Tasa de muerte $\mu_j = \begin{cases} 0 & j = 0 \\ \mu_1 & j \ge 1 \end{cases}$



$$\big[\lambda_j + \mu_j\big]\, p_j = \lambda_{j-1}\, p_{j-1} + \mu_{j+1}\, p_{j+1} \quad \equiv \quad \lambda_j\, p_j + \mu_j\, p_j = \lambda_{j-1}\, p_{j-1} + \mu_{j+1}\, p_{j+1}$$

Ecuaciones de balance o de equilibrio:

$$N=0: \quad \mu_1 p_1^1 = \lambda_0 p_0^1 \quad \rightarrow \quad p_1^1 = \frac{\lambda_0}{\mu_1} p_0^1 = \left(\frac{\lambda_1}{\mu_1}\right) p_0^1$$

$$N=1: \quad \lambda_0 p_0^1 + \mu_2 p_2^1 = \lambda_1 p_1^1 + \mu_1 p_1^1 \quad \rightarrow \quad p_2^1 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0^1 = \left(\frac{\lambda_1}{\mu_1}\right)^2 p_0^1$$

$$N=2: \quad \lambda_1 p_1^1 + \mu_3 p_3^1 = \lambda_2 p_2^1 + \mu_2 p_2^1 \quad \rightarrow \quad p_3^1 = \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3} p_0^1 = \left(\frac{\lambda_1}{\mu_1}\right)^3 p_0^1$$

por recurrencia, $p_N^1 = \left(\frac{\lambda_1}{\mu_1}\right)^N p_0^1 \quad \rightarrow \quad p_0^1 = \dfrac{1}{1+\sum_{k=1}^{\infty}\left(\frac{\lambda_1}{\mu_1}\right)^k}$

Variable $X_2(t)$:

Tasa de nacimiento $\lambda_j = \begin{cases} \lambda_2 & X < N \\ 0 & X \geq N \end{cases}$ Tasa de muerte $\mu_j = \begin{cases} 0 & j=0 \\ \mu_2 & j \geq 1 \end{cases}$

$$N=1: \quad \lambda_0 p_0^2 + \mu_2 p_2^2 = \lambda_1 p_1^2 + \mu_1 p_1^2 \quad \rightarrow \quad p_2^2 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} p_0^2 = \left(\frac{\lambda_2}{\mu_2}\right)^2 p_0^2$$

$$N=2: \quad \lambda_1 p_1^2 + \mu_3 p_3^2 = \lambda_2 p_2^2 + \mu_2 p_2^2 \quad \rightarrow \quad p_3^2 = \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3} p_0^2 = \left(\frac{\lambda_2}{\mu_2}\right)^3 p_0^2$$

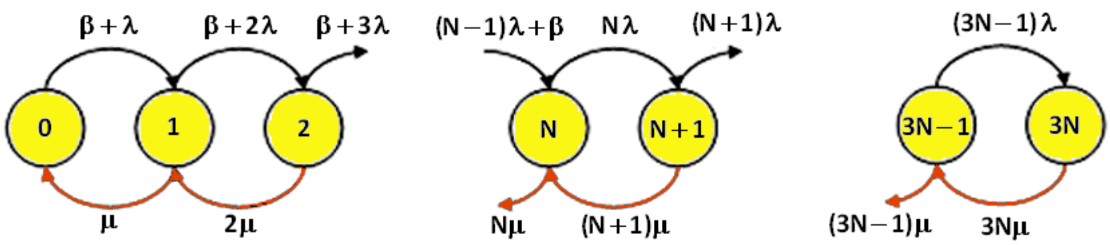
por recurrencia, $p_N^2 = \left(\frac{\lambda_2}{\mu_2}\right)^N p_0^2 \quad \rightarrow \quad p_0^2 = \dfrac{1}{1+\sum_{k=1}^{\infty}\left(\frac{\lambda_2}{\mu_2}\right)^k}$

En consecuencia, $p = 1 - \sum_{k=0}^{N} p_k^1 p_{N-k}^2 = 1 - p_0^1 \, p_0^2 \sum_{k=0}^{N} \left(\frac{\lambda_1}{\mu_1}\right)^k \left(\frac{\lambda_2}{\mu_2}\right)^{N-k}$

Una banda de rock celebra un concierto en un gran estadio. La banda tiene dos tipos de seguidores: El Grupo A llega de acuerdo a un proceso de Poisson de tasa β y se retira indignado si llega al estadio y encuentra su capacidad completa.
El Grupo B responde de forma contraria, se motiva con una gran afluencia de público, tiene un tiempo entre llegadas distribuido exponencialmente dependiendo del número de personas que se encuentran en el estadio, es decir, el tiempo esperado de una de estas personas es $1/\lambda$ cuando solo hay una persona en el estadio, mientras que si hay i personas en el estadio el tiempo esperado de un miembro de éste grupo es $1/\lambda i$. Cómo no le interesa la capacidad del estadio ingresan al recinto hasta que éste a punto de reventar, lo que ocurre cuando el número de personas es $3N$.
Por otra parte, independientemente del fan que se trate cuando se aburre regresa del estadio en un tiempo distribuido exponencialmente de media $1/\mu$.
Para analizar la dinámica del número de personas que acuden al concierto se supone que no acaba nunca. Se solicita:
a) Modelar el número de personas en el recinto como un proceso de nacimiento y muerte. ¿Cuál es la condición de estacionalidad?.
b) Encontrar una expresión que permita calcular las probabilidades estacionarias.
c) Proporción de fans del Grupo 1 que no pueden ingresar al estadio cuando $N=3$, $\lambda=\beta=1$ y $\mu=2$

Solución:

a) De acuerdo con el enunciado se tiene el diagrama de transición:



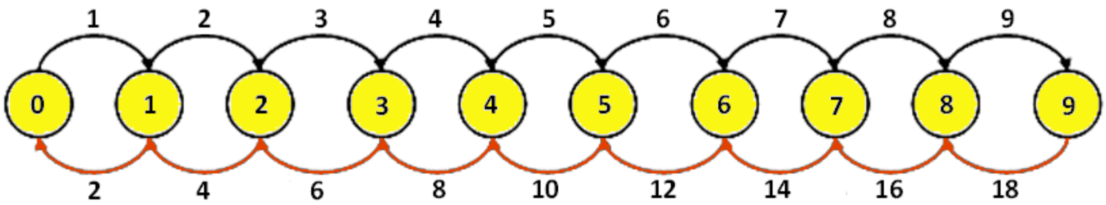
Es una cadena finita por lo que existen probabilidades estacionarias, dado que es irreductible.

b) Al tratarse de un proceso de nacimiento y muerte, utilizando las fórmulas conocidas se tiene:

$$p_i = \begin{cases} \dfrac{\prod_{k=0}^{i-1} (\beta + k\lambda)}{i!\,\mu^i}\, p_0 & 0 < i \le N \\[2ex] \dfrac{\prod_{k=0}^{N-1} (\beta + k\lambda)\, \lambda^{i-N}}{i\,(N-1)!\,\mu^i}\, p_0 & N < i \le 3N \end{cases}$$

donde $$p_0 = \frac{1}{\displaystyle\sum_{i=1}^{N} \frac{\prod_{k=0}^{i-1}(\beta + k\lambda)}{i!\,\mu^i} + \sum_{i=N+1}^{3N} \frac{\prod_{k=0}^{N-1}(\beta + k\lambda)\,\lambda^{i-N}}{i\,(N-1)!\,\mu^i}}$$

c) La cadena que resulta en el Grupo 1 con las características dadas:



Las ecuciones del estado estacionario son: $p_i = \left(\frac{1}{2}\right)^i p_0$

$$\sum_{k=0}^{9} p_i = \sum_{k=0}^{9} \left(\frac{1}{2}\right)^k p_0 = 1 \quad \rightarrow \quad p_0 = \frac{1}{\displaystyle\sum_{k=0}^{9}\left(\frac{1}{2}\right)^k} = \frac{1}{\dfrac{1-\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)}{1-\frac{1}{2}}} = \frac{1-\frac{1}{2}}{1-\left(\frac{1}{2}\right)^{10}} = \frac{\frac{1}{2}}{1-\left(\frac{1}{2}\right)^{10}}$$