



273

Universidad Nacional Autónoma de México

FACULTAD DE CIENCIAS

CADENAS DE MARKOV Y PROCESOS DE
RAMIFICACION: MODELOS Y APLICACIONES

T E S I S

Que para obtener el título de

A C T U A R I O

presenta

ISABEL PATRICIA AGUILAR JUAREZ

México, D. F.

TESIS CON
FALLA DE ORIGEN

1989



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

CONTENIDO

Resumen

Introducción

I	Conceptos de Probabilidad y Procesos Estocásticos	1
1.1	Definiciones Básicas	2
1.2	Funciones Generatrices y Característica	9
1.3	Teorema del Límite Central	12
1.4	Procesos Estocásticos	17
II	Cadenas de Markov	23
2.1	Conceptos Fundamentales	24
2.2	Clasificación de Estados	29
2.3	Algunos Resultados Importantes	35
III	Procesos de Ramificación Univariados	39
3.1	Introducción Histórica del Proceso de Galton-Watson	40
3.2	El Proceso de Galton-Watson: Definición y Conceptos Básicos	43
3.3	Comportamiento Asintótico de Z_n	50
3.4	Estacionaridad de Z_n	53
IV	Algunas Aplicaciones a los Procesos de Markov y de Ramificación	55
4.1	Caminatas Aleatorias	55
4.2	El Modelo de Difusión de Ehrenfest	58
4.3	Teoría del Aprendizaje	60
4.4	El Modelo Abierto de Leontief	65
4.5	Teoría de Colas	70
4.6	Control de Inventarios	82
	Conclusiones	

RESUMEN

Una clase importante de problemas de la vida real requieren para su análisis de la consideración de su comportamiento a través del tiempo, es decir, considerar eventos que pueden ocurrir de manera aleatoria, con cierta probabilidad, y que afectarían el desarrollo del problema. Para el análisis de estos problemas en forma dinámica se necesita una extensión de la probabilidad, denominada Procesos Estocásticos.

Existen diversos tipos de Procesos Estocásticos: aquellos con incrementos independientes, los estacionarios, las martingalas; pero unos de importancia fundamental son, sin duda, las Cadenas de Markov, pues han favorecido el amplio y ágil desarrollo de los Procesos Estocásticos tanto en el campo de la teoría como en el de la práctica. Sus áreas de aplicación son muchas y muy variadas: la Física, Biología, Economía e Ingeniería, entre otras.

El objetivo de este trabajo es describir y analizar las Cadenas de Markov y en particular los Procesos de Ramificación Univariados, con algunas aplicaciones de ellos a diversos campos de la Ingeniería, Administración, Física e Investigación de Operaciones, entre otros.

Para ello, en primer término se revisan algunos conceptos básicos indispensables para la comprensión y análisis del comportamiento de esos procesos, y en seguida, aunque sin profundizar en ellos, algunos modelos ya desarrollados de solución de problemas reales y conocidos que de alguna manera utilizan la teoría mencionada.

INTRODUCCION

Uno de los enigmas de la historia de la humanidad es que el hombre parece haber reconocido la importancia del cambio desde los tiempos más remotos, aunque no fué sino hasta muy recientemente que lo volvió objeto de un estudio formal. El juego es tan viejo como la humanidad. En sitios arqueológicos de gran antigüedad se han encontrado, ya, dispositivos de contingencia lo cual hace pensar que conjuntamente con ellos, existía una intuición de las probabilidades involucradas en su uso y de la forma en que más tarde afectarían las decisiones que un jugador necesitaría tomar en el curso de cada juego.

El uso de mecanismos de contingencia para otro tipo de propósitos como jurídicos o de culto, ha sido también universal, aunque el azar era concebido con cualidades divinas o demoniacas, y parecen haber tenido influencia sobre la evolución y sobrevivencia, de algunas tribus y culturas. Aún ahora se utilizan mecanismos de azar o contingencia en situaciones en que no puede existir algo que sugiera duda en la equidad de una selección, por ejemplo: comúnmente los jurados se eligen por sorteo; y en la mayoría de las competencias deportivas se decide mediante un volado el grupo que iniciará el partido.

Los sociólogos han señalado la confianza en el azar en prácticamente todas las situaciones en que la ignorancia de los elementos que afectan el resultado de una decisión las hacen muy difíciles de sostener.

A pesar de este embrollo universal y el significado social de el azar, los pensadores con mayor influencia antes de la era

científica negaron la existencia del azar o, aun cuando su existencia era reconocida, lo excluían como objeto válido de discurso racional. Así Aristóteles definió al azar como la clase de todo aquello que es indefinido e inescrutable al intelecto humano, y prácticamente todos los pensadores de la escuela escolástica consideran que la existencia del azar puede ser compatible con la existencia de un Dios; para Thomas Aquinas un evento aleatorio es solamente una coincidencia de dos o más causas. En la misma línea de pensamiento, Spinoza pensaba que "todas las cosas están determinadas por la necesidad divina de existir y actuar en determinada forma".

Aún los filósofos clásicos se negaban a reconocer la existencia del azar por sí mismo, aún cuando utilizaban una vaga noción de predictibilidad como medida de conocimiento práctico de eventos inciertos. El predominio de esta forma de pensamiento provocó el retraso en el desarrollo de las ciencias empíricas en general y en particular de la estadística y la probabilidad, cuyo interés y análisis aumentaron fuertemente más tarde, ligado con el ascenso del método empírico, y por la necesidad de expresar el grado de conocimiento necesariamente incompleto, ganado mediante la experimentación repetida, y por muchos años se identificaba su aplicabilidad únicamente con los juegos de azar.

Aún cuando parece haber existido algunos escritos sobre la probabilidad desde el siglo XVI, estos eran más bien informales y aparentemente no ejercieron influencia alguna sobre los autores más serios de escritos de probabilidad que aparecieron posteriormente.

Al igual que en otras áreas de las matemáticas modernas, los trabajos iniciales sobre la teoría de la probabilidad se atribuyen a matemáticos franceses del siglo XVII. Se conocen cartas entre Pascal y Fermat que contienen varios ejemplos de argumentos

básicos de análisis combinatorio y su aplicación en el cálculo de probabilidades sencillas. El primer trabajo que examina los métodos que fundamentan los cálculos de la probabilidad fué escrito por el científico alemán Christian Huygens bajo el título de "De Ratiociniis in Aleae Ludo" (Sobre el razonamiento acerca del juego de dados) que apareció en 1657 y tuvo gran influjo llegando a mantenerse durante varias décadas como el tratado de probabilidad, siendo, tal vez, uno de los instrumentos que estimularon el interés en dicha disciplina, lo que ocurrió alrededor de 1700.

En los inicios del siglo XVIII aparecen publicaciones clásicas de autores tales como James Bernoulli a quien se deben la Ley débil de los Grandes Números para lo que conocemos como ensayos repetidos de Bernoulli, la idea de una sucesión ilimitada de ensayos repetidos y la distinción entre probabilidad de un evento y la frecuencia relativa de su ocurrencia en una sucesión de ensayos repetidos. Influenciados por él y por los nacientes métodos de análisis, aparecen trabajos de De Montmort y De Moivre que definió eventos independientes, esperanza y probabilidad condicional y obtuvo las aproximaciones Normal y de Poisson a las probabilidades binomiales.

Más tarde, durante la segunda mitad del mismo siglo XVIII, también aparecieron resultados importantes como la regla de Bayes y el problema de la aguja de Buffon, entre otros, muchos de los cuales anunciaban lo que se aproximaba. En los trabajos de Bayes este indica como las probabilidades "a posteriori" de las "causas" pueden ser calculadas a partir de las probabilidades "a priori" tomando en cuenta la ocurrencia de un evento. De Buffon, por su parte, discutió un gran número de problemas de juegos y pensaba que el conocimiento de la probabilidad sería "un poderoso antídoto contra el mal epidémico de la pasión del jugador". Buffon es recordado principalmente por su problema de la aguja y por ser el

precursor de la probabilidad geométrica y de las técnicas de simulación.

Finalmente, y como para complementar la era en este campo, aparece Pierre-Simon de Laplace con su tratado "La teoría analítica de Probabilidad" (*Théorie Analytique des Probabilités*), que es indudablemente el más grande trabajo de probabilidad antes del siglo XX. Los más reciente historiadores de las matemáticas afirman que Laplace fué influenciado por varios autores tanto anteriores como contemporáneos a él, sin embargo cualquier juicio que se hiciera sería subjetivo, pero lo que si es claro es que éste desarrolló muchos problemas e ideas de gente como Bernoulli, De Montmort y De Moivre. Desde él y hasta el siglo XX, la mayoría de las nuevas ideas que aparecieron fueron relacionadas con los métodos estadísticos, el análisis de errores de observación y las matemáticas actuariales, más que con la teoría de la probabilidad por sí.

Dado que el siglo XIX fué una época de evolución sin precedentes de las ciencias empíricas, y de considerables reformas de la estructura económica y social, no resulta sorprendente que una gran variedad de aplicaciones se adelantaran a los desarrollos teóricos en la mayoría de las ramas de las matemáticas y, en particular, en la probabilidad. En parte como una reacción a este crecimiento y en respuesta al descubrimiento de cantidad de incompatibilidades y paradojas en la teoría de conjuntos, el interés predominante de la investigación matemática en la primera mitad del siglo XX fué el tratamiento abstracto y la axiomatización en varias de las áreas, entre ellas la teoría de conjuntos, y más tarde la teoría de la probabilidad, la cual se revisó desde el punto de vista del método axiomático y desde aquel de su relevancia en varias ciencias físicas, lo que marcó el inicio de la era moderna de la probabilidad.

Los matemáticos de principios de este siglo eran más cautelosos de los elementos experimentales en la teoría de la probabilidad que en ramas más antiguas de las matemáticas como la geometría o el análisis. Aún la noción de "la probabilidad de un evento" parecía vaga y solamente definida en términos de la "frecuencia relativa de la ocurrencia de un evento en una larga sucesión de ensayos".

Aún en situaciones simples como los modelos de urnas, las probabilidades se asignaban sobre las bases de total asignación, elección "aleatoria" y simetría, y no era claro que dicha asignación se pudiera verificar empíricamente, o bien que la teoría matemática pudiera ser independiente de tales elementos extraños y subjetivos. Acerca de ello la discusión más completa desde el punto de vista clásico la hizo Richard von Mises, sin embargo él no pudo proponer un conjunto satisfactorio de axiomas para la probabilidad. Después de que un grupo de matemáticos interesados en generalizar el concepto de integral, entre los que se encontraba Lebesgue, desarrollaron la noción de función de conjunto y de medida, A. N. Kolmogorov descubrió que la probabilidad podía ser desarrollada axiomáticamente dentro del marco de la teoría de la medida. Su libro "The Foundations of Probability Theory" (Fundamentos de la Teoría de Probabilidad) es la piedra angular de los tratamientos modernos de la probabilidad. El resultado que dió mayor libertad a la teoría matemática, de los elementos experimentales y subjetivos fué la Ley Fuerte de los Grandes Números, la cual implica que las probabilidades postuladas pueden ser medidas en un sentido preciso, al menos para experimentos que pueden ser repetidos independientemente e indefinidamente siempre. Esto tiene el mismo efecto libertador para las teorías de la probabilidad como la definición de cantidades geométricas independientes de la actual, objetos físicos y medidas. Este resultado fundamental fué probado para ensayos de Bernoulli por Emile Borel y, en general, para sucesiones de variables aleatorias independientes por Kolmogorov.

La mayoría de las investigaciones desde mediados de los 30's se han concentrado en el estudio de dependencia. Los precursores de este trabajo son De Moivre, Laplace con su "problema de ruina del jugador" y más notablemente A. A. Markov con la teoría de la "cadenas" que llevan su nombre. El estudio de varios tipos de "dependencia" se cubre casi completamente con la teoría de procesos estocásticos o procesos aleatorios que ocurren en el tiempo, la cual presenta una gran cantidad de nuevas ideas y dificultades matemáticas. Su rango y variedad de excitantes aplicaciones parece, aún ahora, ser inagotable. Entre los distintos modelos de procesos estocásticos y en particular de las Cadenas de Markov, se encuentran los Procesos de Ramificación en el que se supone, en el caso más simple, que la población inicial consta de un solo elemento que se reproducirá y el número de descendientes que tenga tendrá una distribución de probabilidad determinada. El desarrollo de dicho modelo se debe en principio a F. Galton y H. W. Watson quienes en el siglo XVIII motivados por un problema demográfico conocido como el "Problema de Extinción de Familias", llegaron a su planteamiento aunque no a su solución. No fué sino hasta varios años más tarde, en que el modelo volvió a ser utilizado, ahora por otros matemáticos y con otras aplicaciones, que la teoría alrededor del mismo obtuvo un gran desarrollo al igual que sus posibilidades de utilización.

La aparición de la teoría de la probabilidad y la inferencia estadística ha revolucionado realmente la ciencia. Prácticamente todos los modelos determinísticos estáticos del siglo XIX han sido reemplazados por los llamados modelos estadísticos. Ahora y en un nivel misterioso y avanzado de la física, ha surgido nuevamente la controversia acerca de si la naturaleza obedece fundamentalmente a leyes determinísticas o estocásticas, pero lo que sí nadie cuestiona es lo apropiado de los modelos probabilísticos para fenómenos que involucren demasiadas partículas para caer a una descripción cinemática exacta tal como en la teoría de dinámica de

gases. El problema surge cuando se cuestiona si las interacciones elementales en átomos, tales como las estudiadas en mecánica cuántica, obedecen leyes probabilísticas o determinísticas. Aunque no se han unificado los puntos de vista al respecto de los físicos atómicos, sus argumentos no afectan la validez de la teoría matemática, sino únicamente el hecho de que sea apropiada para la descripción de ciertos fenómenos físicos especiales.

Los modelos probabilísticos se han vuelto comunes en el estudio de muchos y diversos fenómenos tales como flujo turbulento, crecimiento de población, ingeniería telefónica, fluctuaciones en precios de stock, y otros. El mecanismo preciso en cada uno de estos casos es demasiado complejo, tanto que excede los límites de la imaginación, o totalmente impredecible en cualquier sentido determinístico. La gran cantidad de interacciones en dichos sistemas hacen el desarrollo de un individuo o partícula extremadamente complicado. Las propiedades de interés práctico son típicamente el resultado del desarrollo conjunto de una gran población. Dichas propiedades, como tales, pueden ser formalizadas en términos estadísticos. Finalmente lo que importa es nuestra habilidad para predecir el desarrollo de , o al menos de obtener un modelo satisfactorio para, tales sistemas físicos, biológicos o económicos.

En la ahora clásica "Teoría de Juegos y Desarrollo Económico" de John von Neumann y Oskar Morgenstern, la probabilidad recibió una nueva interpretación, que es de gran importancia para las ciencias sociales y la toma de decisiones. En la teoría de decisiones, se consideran un conjunto de alternativas para cada protagonista y un orden entre ellas que refleja las preferencias del mismo. Esta idea se complementa con la asignación de probabilidades a subconjuntos del conjunto de alternativas y se obtiene una estrategia mixta, con ello se puede desarrollar una teoría equilibrada para elecciones racionales.

La teoría de juegos fué originalmente utilizada en aplicaciones económicas, sin embargo, resultó incidentalmente en un ensanchamiento del objeto de la estadística.

Existen pequeñas diferencias o ninguna, entre las propiedades de las variantes modernas de la probabilidad y la probabilidad clásica. La diferencia está en la interpretación cualitativa dada al conjunto específico de funciones que son introducidas.

El objetivo de este trabajo es describir y analizar las cadenas de Markov y en particular los procesos de Ramificación, con algunas aplicaciones a diversos campos de la Ingeniería, Administración, Física, e Investigación de Operaciones, entre otros.

La estructura que se eligió para la presentación del material es la siguiente: El primer capítulo contiene algunos conceptos de la Teoría de Probabilidad que se utilizan en los capítulos subsecuentes; El segundo trata sobre los Procesos de Markov. El tercer capítulo analiza los Procesos de Ramificación con un solo tipo así como algunos resultados relacionados con ellos, y en el cuarto se presentan aplicaciones de lo expuesto en los capítulos uno, dos y tres, y algunos ejemplos resueltos.

CAPÍTULO I

CONCEPTOS DE PROBABILIDAD Y PROCESOS ESTOCÁSTICOS

La teoría de la probabilidad es una rama de las matemáticas que encuentra aplicación en diversas áreas del conocimiento ayudando a evaluar incertidumbre y minimizar errores de observación y permitiendo predecir comportamientos futuros. La materia de estudio son los fenómenos aleatorios, algunos estáticos y otros dinámicos; muchos de estos fenómenos tienen como parámetro al tiempo y se les llama Procesos Estocásticos. Dichos procesos representan la forma dinámica de la probabilidad y su aplicación en problemas industriales, económicos e ingenieriles, es usual.

El objetivo de este capítulo es presentar los conceptos fundamentales de la teoría de Probabilidad y los Procesos Estocásticos. En la primera sección se introducen los conceptos básicos de la probabilidad (variables aleatorias, y características matemáticas de ellas, funciones de densidad y distribución). En la segunda y la tercera secciones se incluyen algunos resultados importantes de la probabilidad como son funciones característica y generatrices, así como el teorema del límite central. La cuarta sección es una introducción a los Procesos Estocásticos.

1.1 DEFINICIONES BASICAS

En una diversidad de situaciones cotidianas se observa que al repetirse un fenómeno bajo ciertas condiciones fijas, no siempre produce la misma respuesta. Un fenómeno aleatorio es aquel que al repetirse las mismas condiciones iniciales, no se puede conocer, con certeza, que resultado se producirá.

La teoría de la probabilidad tiene por objeto desarrollar modelos matemáticos para el estudio de los fenómenos mencionados. La traducción formal de fenómeno aleatorio, en el contexto de dicha teoría, es la siguiente:

Un Espacio de Probabilidad es una terna (Ω, A, P) , en donde:

- 1) Ω es un conjunto arbitrario diferente del vacío.
- 2) A es un σ -campo de subconjuntos de Ω , es decir, una familia de subconjuntos de Ω que satisface las siguientes condiciones:

i) $\emptyset \in A$

ii) Si $A \in A$ entonces $A^c \in A$

iii) Si A_1, A_2, \dots son elementos de A , su unión también lo es.

- 3) P es una función cuyo dominio y rango son A y el intervalo cerrado $[0,1]$ respectivamente, que cumple:

i) $P(\Omega) = 1$

ii) Si $A \in A$, $P(A) > 0$

iii) Si A_1, A_2, \dots son elementos de A cuyas intersecciones por parejas, son vacías, entonces

$$P(\cup A_i) = \sum P(A_i)$$

Intuitivamente se puede decir que Ω es el conjunto de todos los resultados posibles del fenómeno; A es una familia de subconjuntos de Ω lo suficientemente rica como para poder agrupar cualquier conjunto de resultados, y nos dice las propiedades de la estructura objeto de estudio; sus elementos se denominan eventos. De hecho A constituye el dominio "natural" de definición de P . P es una medida llamada probabilidad, y nos dice como es estudiado el objeto por la teoría.

Sean A y B dos elementos de A . La probabilidad condicional de A dado B es:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

En muchas ocasiones, la información de que B ha ocurrido cambia la probabilidad de ocurrencia de A . Cuando $P(A \cap B) = P(A)P(B)$, la información sobre B no permite hacer ninguna inferencia acerca de la ocurrencia de A , entonces se dice que A y B son independientes. Esto es, dos eventos A y B son independientes, si:

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

Esta definición también se acepta si $P(B) = 0$. Así mismo, los eventos A_1, A_2, \dots, A_n se llaman mutuamente independientes si, para todas las combinaciones $1 \leq i < j < k < \dots \leq n$, se aplican las siguientes reglas de multiplicación:

$$\begin{aligned} P(A_1 \cap A_j) &= P(A_1)P(A_j) \\ P(A_1 \cap A_j \cap A_k) &= P(A_1)P(A_j)P(A_k) \\ \dots & \\ P(A_1 \cap A_2 \cap \dots \cap A_n) &= P(A_1)P(A_2) \dots P(A_n) \end{aligned}$$

Un ejemplo de la utilización de este concepto:

Supongase que una baraja de 52 cartas consta de 4 ases y 48 no-ases. ¿Cuál es la probabilidad de que, de 5 cartas extraídas con reemplazo, dos exactamente sean ases?

Sea A_i el evento de extraer un as en la i -ésima extracción. $i = 1, 2, 3, 4, 5$. La probabilidad de que los dos ases salgan en las dos primeras extracciones, es:

$$\begin{aligned}
 P(A_1 \cap A_2 \cap A_3^c \cap A_4^c \cap A_5^c) &= P(A_1 \cap A_2 \cap A_3^c \cap A_4^c) \cdot P(A_5^c | B) \\
 &= \dots\dots\dots \\
 &= P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3^c | A_1 \cap A_2) \cdot \\
 &\quad P(A_4^c | A_1 \cap A_2 \cap A_3^c) \cdot P(A_5^c | B) \\
 &= (4/52)(4/52)(48/52)(48/52)(48/52) \\
 &= (4/52)^2 (48/52)^3 = 0.00465
 \end{aligned}$$

Hasta aquí puede observarse que los eventos A_i son mutuamente independientes. Ahora, la probabilidad calculada hasta este momento, corresponde a una forma particular de extraer dos ases y cinco no-ases, pero existen diez formas distintas de obtenerlos (combinaciones de cinco elementos tomados de dos en dos), cada una de ellas con la misma probabilidad, entonces:

$$P(\text{dos ases y tres no-ases}) = 10 (0.00465) = 0.0465$$

Obsérvese que se ha hablado de eventos, y si bien la teoría de la probabilidad se desarrolla alrededor de ellos y de sus probabilidades, las expresiones que se obtendrían utilizando esta terminología serían bastante complejas, por lo que en adelante se hará referencia a ellos en términos de variables aleatorias.

Se dice que X es una variable aleatoria real sobre (Ω, \mathcal{A}, P) , si es

una función de Ω en \mathbb{R} tal que si $x \in \mathbb{R}$, $\{ \omega \in \Omega \mid X(\omega) \leq x \} \in \mathcal{A}$.

Obsérvese que $X(\omega)$ es la pregunta aplicada al elemento $\omega \in \Omega$, y que el valor que toma, es la respuesta. Las variables aleatorias se clasifican, dependiendo del número de valores que pueden tomar, como sigue: Aquellas que solamente pueden tomar un número finito o numerable de valores, se llaman variables aleatorias discretas, y a las que pueden tomar un número infinito no-numerable de valores, se les llama variables aleatorias continuas.

Lo que interesa ahora, es conocer el comportamiento de la variable aleatoria X , probabilísticamente hablando, y esto se puede lograr a través de la función de distribución de la variable, la cual dirá como se distribuyen las respuestas a la pregunta que interesa sobre el fenómeno aleatorio. En estas condiciones, para cualquier variable aleatoria X , se define:

La función de distribución de la variable aleatoria X en x es

$$F_X(x) = P\{ \omega \in \Omega ; X(\omega) \leq x \}.$$

Observe que gracias a la definición de variable aleatoria esta función siempre está definida. Para la misma variable aleatoria X , resulta de utilidad para muchos cálculos, el uso de otra función llamada función de densidad de X , la cual proporciona información acerca del comportamiento probabilístico de la variable aleatoria y guarda una estrecha relación con la función de distribución de la misma variable. Dicha función de densidad se define como sigue:

$$f(x) = \frac{dF_X(x)}{dx} \quad \text{si } X \text{ es continua,}$$

$$\text{y} \quad f(x) = P\{X = x\} \quad \text{si } X \text{ es discreta.}$$

Así como las variables aleatorias se clasifican en discretas y

continuas, es posible también, clasificar de igual manera a las funciones de distribución que describen a unas y a otras.

Entre las distribuciones discretas se encuentran:

La Binomial con parámetros "n" y "p", denotada como bin(n,p), con

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \dots, n.$$

La Poisson con parámetro λ , que se denota como P(λ), en la que

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \lambda > 0, \quad x = 0, 1, 2, \dots$$

La uniforme con parámetro n denotada por U(n), con

$$f(x) = \frac{1}{n} \quad x = 0, 1, \dots, n.$$

entre otras. De ellas, probablemente la más importante, por sus aplicaciones estadísticas, es la Binomial.

En el caso continuo, se pueden mencionar:

La distribución normal con parámetros μ (media de X) y σ^2 (varianza de X), para la cual

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x - \mu)^2}{2\sigma^2}} \quad \begin{array}{l} \sigma \geq 0 \\ -\infty < x < \infty \\ \mu \in \mathbb{R} \end{array}$$

La distribución gamma con parámetros α y β , en la que

$$f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha) \beta^\alpha} x^{\alpha-1} e^{-x/\beta} \quad \begin{array}{l} \alpha \geq 0 ; \beta \geq 0 \\ 0 \leq x < \infty \end{array}$$

como ejemplos de las más importantes. Dentro de esta clase de funciones de distribución, la que se considera como la más importante, es la distribución normal, pues es de gran utilidad en

aplicaciones estadísticas, así como dentro de la misma teoría de probabilidades por resultar una buena aproximación para otras distribuciones tanto continuas como discretas, como se verá en la sección 1.3.

Comúnmente, cuando se habla de que una variable aleatoria Y se distribuye conforme a una normal con media μ y varianza σ^2 , se hace en los siguientes términos:

$$Y \sim N(\mu, \sigma^2)$$

Una variable aleatoria posee también ciertas características numéricas, que proporcionan información adicional sobre el comportamiento de la variable. Dichas características son:

La esperanza de una variable aleatoria X , con función de distribución $F(x)$, es:

Para el caso discreto

$$E(X) = \sum x f(x)$$

y para el continuo

$$E(X) = \int x dF(x)$$

La varianza de X , si $\mu = E(X)$, se define como

$$\text{Var}(X) = E((X - \mu)^2) = E(X^2) - \mu^2$$

En general el k -ésimo momento de X es:

Si X es discreta

$$E(X^k) = \sum x^k f(x)$$

y si es continua

$$E(X^k) = \int x^k dF(x)$$

Ejemplo: Supóngase una variable aleatoria X con función de densidad $f(x) = 2(1-x)$, $0 < x < 1$. El k -ésimo momento de X está definido como

$$\begin{aligned} E(X^k) &= \int x^k dF(x) \\ &= \int_0^1 x^k (2(1-x)) dx \\ &= 2 \int_0^1 x^k (1-x) dx \\ &= 2 \int_0^1 (x^k - x^{k+1}) dx \\ &= \left[\frac{2x^{k+1}}{k+1} - \frac{2x^{k+2}}{k+2} \right]_0^1 \end{aligned}$$

entonces la esperanza de X es

$$E(X) = E(X^1) = \frac{2}{2} - \frac{2}{3} = \frac{1}{3}$$

y la varianza de X

$$\begin{aligned} \text{Var}(X) &= E(X^2) - E^2(X) \\ &= \frac{2}{3} - \frac{2}{4} - \frac{1}{9} \\ &= \frac{1}{18} \end{aligned}$$

1.2 FUNCIONES GENERATRICES Y CARACTERISTICA

En muchas ocasiones interesa el comportamiento probabilístico, no solo de una variable aleatoria, sino de combinaciones de varias de ellas. Un paso importante para lograr dicho conocimiento, es el cálculo de la función de densidad de la suma de las variables. Si las variables son independientes y solo toman valores enteros no-negativos, el cálculo de la densidad de probabilidad de la suma, se simplifica utilizando la función generatriz de probabilidad.

Se define, entonces, la función generatriz de probabilidad de una variable aleatoria X no-negativa que toma valores enteros, como

$$f(s) = \sum_{k=0}^{\infty} P(X = k) s^k \quad -1 \leq s \leq 1$$

En general, los momentos definidos en la sección anterior, se pueden conocer a través de una función llamada función generatriz de momentos $\psi_X(t)$, la cual está definida como

$$\psi_X(t) = E(e^{tX}),$$

sin embargo, esta función no está definida en todos los casos, lo cual hace necesario utilizar una nueva función denominada función característica de la misma variable aleatoria X , la cual está definida como

$$\varphi_X(t) = E(e^{itX}).$$

Es importante notar que la información que da la función característica, es la misma que la que da la función de

distribución, es decir, si se conoce F_X , entonces se conoce ϕ_X , y a la inversa, si se conoce ϕ_X , se conoce F_X . Esto significa que existe una biyección entre el conjunto de funciones características y el de funciones de distribución.

Por lo tanto si se desea conocer ϕ_X a partir de F_X basta calcular

$$E(e^{itx}) = \int e^{itx} dF_X(x).$$

Conocida la función característica de X , se utiliza el siguiente teorema para obtener la correspondiente función de distribución.

TEOREMA 1.1) Sean ϕ_X y F_X la función característica y la función de distribución respectivamente, de una variable aleatoria X . Si x_1 y x_2 son puntos de continuidad de la función F_X , entonces

$$F_X(x_2) - F_X(x_1) = (1/2\pi) \lim_{c \rightarrow 0} \int_{-c}^c \frac{e^{-itx_1} - e^{-itx_2}}{it} \phi_X(t) dt.$$

Para la demostración del teorema se parte del lado derecho de la igualdad y se sustituye $\phi_X(t)$ por su equivalente en términos de la función de distribución $F_X(x)$ como sigue:

$$(1/2\pi) \lim_{c \rightarrow 0} \int_{-c}^c \frac{e^{-itx_1} - e^{-itx_2}}{it} \phi_X(t) dt =$$

$$(1/2\pi) \lim_{c \rightarrow 0} \int_{-c}^c \frac{1}{it} [e^{-it(z-x_1)} - e^{-it(z-x_2)}] dF(z) dt$$

e intercambiando el orden de integración y desarrollando las

exponenciales complejas, se llega a

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \int_0^c \left[\frac{\sin t(c-x_1)}{t} - \frac{\sin t(c-x_2)}{t} \right] dt dF(z)$$

descomponiendo la integral con respecto a $F(z)$ en varias integrales, se llega a que el límite superior del lado derecho es:

$$F(x_2 - \delta) - F(x_1 + \delta) + R_1(\delta, x_1, x_2)$$

y el límite inferior es:

$$F(x_2 + \delta) - F(x_1 - \delta) + R_2(\delta, x_1, x_2)$$

con $|R_1(\delta, x_1, x_2)| < 2(F(x_1 + \delta) - F(x_1 - \delta) + F(x_2 + \delta) - F(x_2 - \delta))$ $i=1,2$
 $\delta > 0$.

Si $\delta \rightarrow 0$ $F(x_1 + \delta)$ y $F(x_1 - \delta) \rightarrow F(x_1)$
 y $F(x_2 + \delta)$ y $F(x_2 - \delta) \rightarrow F(x_2)$

por lo tanto $\lim_{c \rightarrow 0} (1/2\pi) \int_{-c}^c \frac{e^{-itx_1} - e^{-itx_2}}{it} \varphi_X(t) dt = F(x_2) - F(x_1)$

(Para la demostración formal consultar Gnedenko, B. V.: "Theory of Probability", Ed. Chelsea Publishing Company).

La importancia de la fórmula de inversión radica en su utilidad no tanto práctica sino teórica, pues ayuda a demostrar resultados de mayor empleo en la práctica tales como el teorema de unicidad, el cual afirma que "una función de distribución está únicamente determinada por su función característica".

1.3 TEOREMA DEL LIMITE CENTRAL

En relación a un fenómeno aleatorio, se pueden hacer no solo una, sino varias preguntas simultáneamente, y en esos casos, la teoría presentada anteriormente no resulta suficiente para responderlas, por lo que se hace necesaria una teoría más general, y se habla ahora, de vectores aleatorios y función de distribución conjunta, cada uno de ellos con interpretación muy semejante a sus equivalentes en el caso de una sola pregunta. Así,

X es un vector aleatorio n-dimensional, si $X = (X_1, X_2, \dots, X_n)$ sobre (Ω, A, P) , es tal que para $i=1, 2, \dots, n$, X_i es variable aleatoria sobre (Ω, A, P) , y la función de distribución conjunta de X en $x = (x_1, x_2, \dots, x_n)$ es

$$F_X(x) = P(\omega \in \Omega ; x_1(\omega) \leq x_1, x_2(\omega) \leq x_2, \dots, x_n(\omega) \leq x_n) .$$

En muchas ocasiones interesa conocer el comportamiento de los resultados de varias repeticiones sucesivas de un mismo fenómeno aleatorio. Lo que se obtiene de esta necesidad son las sucesiones de variables aleatorias, y la teoría desarrollada acerca de ellas está fundamentada en un caso particular, que son las sucesiones de ensayos de Bernoulli, es decir, sucesiones de ensayos repetidos e independientes, en los que solo hay dos resultados posibles (éxito y fracaso), y con las mismas probabilidades en todos los ensayos.

La relación entre los ensayos de Bernoulli y la teoría de las variables aleatorias se aclara cuando se considera la dependencia que existe entre el número S_n de éxitos y el número n de ensayos. En cada ensayo, S_n se incrementa en 1 o 0, lo cual se puede escribir como

$$S_n = X_1 + X_2 + \dots + X_n$$

donde la variable X_k es igual a uno si el k -ésimo ensayo resulta un éxito y cero en el caso contrario. Por lo tanto, S_n es una suma de n variables aleatorias mutuamente independientes y con distribución arbitraria, cada una de las cuales toma el valor uno o cero con probabilidades p y q respectivamente. La Ley débil de los grandes números establece que, para n grande, es probable que la proporción promedio de éxitos S_n/n se aproxime a p , lo cual es un caso particular de la Ley de los Grandes Números, que establece que:

Si $\{X_k\}$ es una sucesión de variables aleatorias mutuamente independientes y con igual distribución, y si existe la esperanza $\mu = E(X_k)$, entonces para toda $\varepsilon > 0$, cuando $n \rightarrow \infty$

$$P \left\{ \left| \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} - \mu \right| > \varepsilon \right\} \rightarrow 0;$$

en otras palabras, la probabilidad de que el promedio S_n/n difiera de la esperanza en menos que una ε dada, arbitrariamente pequeña, tiende a uno. La generalización conceptual de este resultado, se puede plantear como sigue:

Una sucesión de variables aleatorias X_n converge en probabilidad a una variable aleatoria X ($X_n \xrightarrow{P} X$), si para cualquier $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P\{|X - X_n| < \varepsilon\} = 1.$$

La ley de los grandes números establece que bajo ciertas condiciones, la suma

$$(1/n) \sum_{k=1}^n (X_k - EX_k),$$

converge, en probabilidad, a cero, es decir,

$$(1/n) \sum_{k=1}^n (X_k - EX_k) \xrightarrow{P} 0.$$

Otros conceptos relacionados con la misma idea de convergencia de variables aleatorias se presentan a continuación.

Se dice que una sucesión de variables aleatorias X_n converge en distribución a una variable aleatoria X , si para toda x que sea punto de continuidad de F_X ,

$$F_{X_n} \rightarrow F_X$$

y se denota como

$$X_n \xrightarrow{D} X$$

TEOREMA 1.2) Sean X_n , $n \geq 1$ y X variables aleatorias tales que

$$\lim_{n \rightarrow \infty} \varphi_{X_n}(t) = \varphi_X(t), \quad -\infty < t < \infty,$$

entonces

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x),$$

para todos los puntos x que son puntos de continuidad de X . es decir, la biyección que se mencionaba entre funciones de distribución y funciones características es continua.

Para estudiar las leyes que caracterizan a las sumas de un número grande pero finito de variables aleatorias, cada una de las cuales tiene solo un efecto pequeño en la suma, se debe considerar una sucesión de sumas que tengan siempre un número creciente de términos y tomar como solución al problema la función de distribución que es el límite de la sucesión de funciones de distribución para las sumas. Para que las funciones de distribución de las sumas de las variables aleatorias converjan a una distribución normal, se deben imponer ciertas condiciones a dichas variables aleatorias, condiciones que están contenidas en el Teorema del Límite Central.

TEOREMA 1.3 (Teorema del Límite Central) Sea $\{X_k\}$ una sucesión de variables aleatorias mutuamente independientes e idénticamente distribuidas. Supóngase que existen

$\mu = E(X_k)$ y $\sigma^2 = \text{Var}(X_k)$ y sean $S_n = \sum_{i=1}^n X_i$. Entonces, para toda n :

fija

$$\frac{\sum_{i=1}^n (X_i - E(X_i))}{\sqrt{\text{Var} \sum_{i=1}^n X_i}} \xrightarrow{D} Y, \quad Y \sim N(0,1).$$

En palabras el teorema dice que la suma de una sucesión de variables aleatorias estandarizadas tiene una función de distribución acumulativa que converge a aquella de una variable normal estándar. El teorema tiene muchas generalizaciones, por ejemplo, no es necesario que las variables aleatorias X_i tengan la misma función de distribución acumulativa. Note que este teorema es más fuerte que la ley de los grandes números que únicamente establece que bajo ciertas condiciones, la media muestral converge a la media real. Así mismo da una función de distribución acumulativa a la cual converge la función de distribución acumulativa de una suma estandarizada. En la práctica siempre que se sabe que una variable aleatoria es la suma de un gran número de variables bien determinadas, se tiene una justificación para suponer que su suma es normalmente distribuida.

Un ejemplo de la aplicación de este teorema, es el siguiente:

Sea X_1, X_2, \dots, X_n una sucesión de variables aleatorias con distribución $\text{bin}(1, p)$, entonces $\mu = p$ y $\sigma^2 = p(1-p)$. Si $Y_n =$

$X_1 + X_2 + \dots + X_n$, entonces $Y_n \sim \text{bin}(n, p)$, $\mu = np$ y $\sigma^2 = np(1-p)$. Supóngase que $n = 100$ y $p = 1/2$, y se desea calcular

$P(Y = 48, 49, 50, 51, 52)$. Como Y es una variable aleatoria discreta, es equivalente a calcular $P(47.5 < Y < 52.5)$; $\mu = 50$, $\sigma^2 = 25$. Así,

$$\begin{aligned}P(47.5 < Y < 52.5) &= P\left(\frac{47.5 - 50}{5} < \frac{Y - 50}{25} < \frac{52.5 - 50}{5}\right) \\&= P(-0.5 < \frac{Y - 50}{25} < 0.5)\end{aligned}$$

Aplicando el teorema del límite central, se sabe que la distribución de $(Y - 50)/\sqrt{25}$ se comporta, aproximadamente, como una Normal con media 0 y varianza 1, entonces, utilizando tablas de distribución normal (0,1), se encuentra que:

$$\begin{aligned}P(47.5 < Y < 52.5) &= P(-0.5 < \frac{Y - 50}{25} < 0.5) \\&= P\left(\frac{Y - 50}{25} < 0.5\right) - P\left(\frac{Y - 50}{25} < -0.5\right) \\&= 0.691 - 0.309 \\&= 0.382\end{aligned}$$

1.4 PROCESOS ESTOCÁSTICOS

Cuando se definió el concepto de vector aleatorio, se interpretó como varias preguntas simultáneas acerca de un mismo fenómeno aleatorio, sin embargo, puede representar, también, una sola pregunta acerca del mismo fenómeno, pero hecha en tiempos distintos, si el resultado del mismo varía en el tiempo. A este tipo de fenómenos aleatorios que dependen del tiempo, se les llama Procesos Estocásticos.

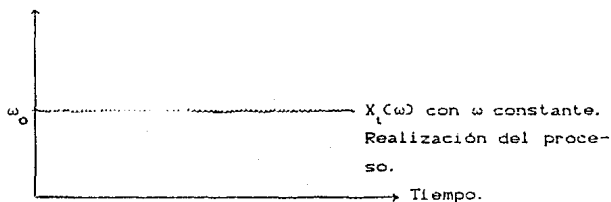
Sea (Ω, A, P) un espacio de probabilidad y T un conjunto arbitrario. Un Proceso Estocástico es una función X que tiene como dominio al conjunto T y como contradominio a un conjunto de variables aleatorias sobre (Ω, A, P) , es decir, X es tal que para toda $t \in T$ fija, $X(t, \omega)$, es una variable aleatoria sobre (Ω, A, P) .

Los principales elementos distintivos de los procesos estocásticos están en la naturaleza del espacio de estados (S), el conjunto de índices de los parámetros (T), y las relaciones de dependencia entre las variables aleatorias X_t .

El espacio de estados S , es el conjunto de posibles valores de X_t . En el caso en que $S = \{0, 1, 2, \dots\}$, se dice que el proceso toma valores en los enteros, o alternativamente, que es un proceso con espacio de estados discreto. Si S es la recta real $(-\infty, \infty)$, entonces X_t es un proceso estocástico con valores reales. Se dice que x es un posible estado de un proceso estocástico X , si para $h > 0$ $P(x-h < X(t) < x+h) > 0$. Si S es no-numerable, a X_t se le llama Proceso, y si es discreto Cadena.

El espacio parametral o espacio de índices de los parámetros (T), puede ser numerable o no-numerable, de hecho basta con que T sea

no vacío. Si $T = \{0, 1, \dots\}$, a X_t se le conoce como proceso estocástico a tiempo discreto. Cuando T sea discreto, se escribirá X_n en lugar de X_t . Si $T = [0, \infty)$ entonces X_t se llama proceso a tiempo continuo. Obsérvese que: Si se fija t , X_t es una variable aleatoria; Si se fijan t y ω , $X_t(\omega)$ es un número real; Si se fija ω y se deja a t variable, lo que se obtiene es una función que, al variar t , se le asigna $X_t(\omega)$. A la función $t \rightarrow X_t(\omega)$, se le llama trayectoria asociada a ω , trayectoria del proceso o realización del proceso; Finalmente, si t y ω son variables, entonces $X_t(\omega)$ es un proceso estocástico.



Ejemplo: Supóngase un experimento que consiste en observar la velocidad de un automóvil durante los primeros sesenta segundos de una carrera. Si se denota por Y_t a la velocidad que lleva al tiempo t , el conjunto $\{Y_t ; 0 \leq t \leq 60\}$ es un proceso estocástico a tiempo continuo, con espacio de estados $S = [0, \infty)$.

Existen varios tipos de procesos, entre los más importantes están:

1) Procesos estacionarios: Un proceso estocástico X_t para t en T , es llamado estacionario, si las funciones de distribución conjunta de las familias de variables aleatorias $\{X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h}\}$ y $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ son las mismas para toda h y elecciones arbitrarias de t_1, t_2, \dots, t_n de T . Esto significa que no importa en que momento (t) se está estudiando el

comportamiento del proceso, pues este es el mismo siempre que se haga el mismo número de observaciones y a la misma distancia. La condición que debe de cumplir un proceso para ser estacionario, afirma que en esencia dicho proceso está en equilibrio probabilístico y que los tiempos particulares en los cuales se examine el proceso, no son relevantes.

Ejemplo: Proceso de Poisson

Supóngase un fenómeno aleatorio tal que, en el tiempo:

- i) La probabilidad de que al menos un evento ocurra en un intervalo de tiempo h es:

$$P(h) = ah + o(h) \quad 1$$
- ii) La probabilidad de dos ocurrencias del evento en un intervalo de tiempo h es $o(h)$.
- iii) El número de ocurrencias en intervalos ajenos de tiempo, son variables aleatorias independientes.
- iv) El proceso es estacionario.

Sean: X_t = Número de ocurrencias hasta el instante t .

$P(X_t = m)$ = Probabilidad de que hasta el instante t , hayan ocurrido m eventos.
 $= P_m(t)$.

$P_0(t + h)$ = Probabilidad de que no ocurra ningún evento hasta $t + h$.
 $=$ Probabilidad de que no ocurra ningún evento hasta t y ninguno entre t y $t + h$. (por iii)
 $=$ Probabilidad de que no ocurra ningún evento hasta t y ninguno hasta h . (por iv)
 $= P_0(t) P_0(h)$. (por iii)

Partiendo de esto, es posible obtener una expresión analítica para la probabilidad $P_0(t)$, como sigue:

¹ Si $g(h) = o(h) \implies \lim_{h \rightarrow 0} [g(h)/h] = 0$

$$\frac{P_0(t+h) - P_0(t)}{h} = \frac{P_0(t)P_0(h) - P_0(t)}{h}; \text{ pero } P_0(h) = 1 - P(h).$$

$$\begin{aligned} \text{entonces } \frac{P_0(t)(1-P(h)) - P_0(t)}{h} &= \frac{-P_0(t)P(h)}{h} \\ &= \frac{-P_0(t)(ah + o(h))}{h} \end{aligned}$$

$$\lim_{h \rightarrow 0} \frac{P_0(t+h) - P_0(t)}{h} = -aP_0(t) = -P_0'(t)$$

por lo tanto, $P_0(t) = ce^{-at}$

$P_0(0) = 1$, lo cual implica que $ce^{-a \cdot 0} = 1$, y se obtiene $c = 1$
entonces $P_0(t) = e^{-at}$

Recursivamente:

$$P_m(t+h) = P_m(t)P_0(h) + P_{m-1}(t)P_1(h) + o(h)$$

pues los términos que seguirían a $P_{m-1}(t)P_1(h)$ serían:

$$P_{m-2}(t)P_2(h) + P_{m-3}(t)P_3(h) + \dots \dots \dots = o(h)$$

$\underbrace{\hspace{1.5cm}}_{o(h)} \quad \underbrace{\hspace{1.5cm}}_{o(h)} \quad (\text{por iii})$

Pero $P_1(h) = P(h) + o(h)$

entonces:

$$\begin{aligned} P_m(t+h) - P_m(t) &= -aP_m(t)h + aP_{m-1}(t)h + o(h) \\ &= P_m(t)(1 - P(h)) + P_{m-1}(t)ah + o(h) - P_m(t) \\ &= P_m(t) - P_m(t)(ah + o(h)) + P_{m-1}(t)ah + o(h) \\ &\quad - P_m(t) \end{aligned}$$

Así, $P_m'(t) = -aP_m(t) + aP_{m-1}(t)$. Sea $Q_m(t) = P_m(t)e^{at}$,
derivando se obtiene $Q_m'(t) = aQ_{m-1}(t)$; $Q_0(t) = 1$. En
particular, $Q_1'(t) = aQ_0(t) = a$, de donde

$$Q_1(t) = at + c \quad \text{por lo que } c = 0$$

$$Q_2'(t) = a Q_1(t) = a^2 t$$

entonces

$$Q_2(t) = [(a^2 t^2)/2] + c, \quad \text{y } c=0$$

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

...

entonces,

$$Q_m'(t) = a^m t^{m-1}$$

$$Q_m(t) = (a^m t^m / m!)$$

$$P_m(t) = Q_m(t) e^{-at} / m!$$

sea $at = \lambda$, entonces:

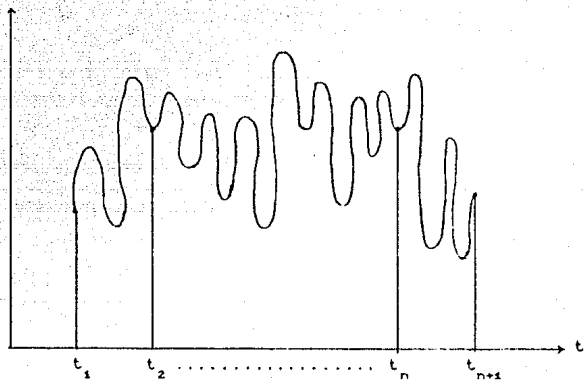
$P_m(t) = \lambda^m e^{-\lambda} / m!$ esto es: las ocurrencias de eventos en un intervalo de tiempo, se comportan como una Poisson.

2) Martingalas: X es una Martingala si para cualquier conjunto t_1, t_2, \dots, t_{n+1} tales que $t_1 < t_2 < \dots < t_{n+1}$, $ECX_{t_{n+1}} | X_{t_1} = a_1, \dots, X_{t_n} = a_n = a_n$. Las martingalas se pueden considerar como modelos apropiados para juegos, en el sentido que X_t significa la cantidad de dinero que un jugador tiene al tiempo t .

3) Procesos de Ramificación: Considérese una partícula que puede producir nuevas partículas de la misma clase. Una sola partícula forma la generación original, o generación cero. Cada partícula tiene probabilidad p_k ($k = 0, 1, 2, \dots$) de originar exactamente k partículas nuevas; los descendientes directos de la n -ésima generación forman la $(n + 1)$ -ésima generación. Las partículas de cada generación actúan de manera independiente una de otra. Lo que interesa es el tamaño de las generaciones sucesivas.

4) X es un Proceso de Markov si para todo conjunto de valores t_1, \dots, t_{n+1} con $t_1 < t_2 < \dots < t_{n+1}$,

$PC X_{t_{n+1}} = x_{n+1} | X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}, X_{t_n} = x_n) =$
 $= PC X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n)$, esto es, el futuro solo
 depende del presente (Propiedad de Markov).



CAPITULO II

CADENAS DE MARKOV

La teoría de los Procesos de Markov es una de las más conocidas y desarrolladas dentro de los Procesos Estocásticos. Su importancia radica en la diversidad de aplicaciones que tiene, en Física, Biología, Investigación de Operaciones, Ingeniería, Comercio, y Ciencias Sociales, lo cual responde a la amplia variedad de modelos que dentro de dicha teoría se han desarrollado. Ejemplo de ellos son: Proceso de Nacimiento-Muerte, Procesos de Renovación, y Procesos de Ramificación, entre otros.

El presente capítulo pretende presentar en términos generales, los conceptos fundamentales que permitan definir y analizar una Cadena de Markov, conceptos tales como espacio de estados, probabilidades de transición, tipos de estados, periodo, probabilidades estacionarias y de absorción, para lo cual el capítulo se integra como sigue: La primera sección está dedicada a los conceptos fundamentales (espacio de estados, tipos de estados, probabilidades de transición, etc.), la segunda presenta conceptos sobre la clasificación de los estados, y la tercera algunos resultados importantes acerca de tiempos de estancia, y absorción, principalmente.

2.1 CONCEPTOS FUNDAMENTALES

En los Procesos Estocásticos existen varios tipos y modelos, cada uno con su propia importancia y capacidad de solución de problemas específicos. Sin embargo, quizás los más importantes por su versatilidad y utilidad en la simulación y solución de problemas en diversos campos de la práctica profesional, sean los Procesos Markovianos.

Indiscutiblemente, para que un proceso pueda ser considerado Markoviano, debe satisfacer ciertas condiciones, las cuales se indican en la siguiente definición:

Un proceso estocástico X es de Markov si para todo conjunto de valores t_1, \dots, t_{n+1} tales que $t_1 < t_2 < \dots < t_{n+1}$, ocurre que

$$P(X_{t_{n+1}} = x_{n+1} \mid X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}, X_{t_n} = x_n) =$$

$$P(X_{t_{n+1}} = x_{n+1} \mid X_{t_n} = x_n) .$$

Es decir, el estado futuro del proceso depende únicamente de su situación en el presente, sin importar la manera como haya llegado a él, esto es, no importa el desarrollo pasado del proceso.

Sea R un intervalo de la recta real. La función:

$$P(x, s; t, R) = P(X_t \in R \mid X_s = x), \quad t > s,$$

se llama función de Probabilidad de transición, y determina la estructura de los procesos de Markov, ya que gracias a la propiedad de Markov, permite conocer las distribuciones de dimensión finita.

Los Procesos de Markov se pueden clasificar de acuerdo a la naturaleza del parámetro (discreto o continuo), o bien, a la naturaleza del espacio de estados; a aquellos con espacio de estados discreto, se les llama Cadenas de Markov.

Ejemplo: Supóngase que se tiene una máquina con dos posibles estados: 0 si está descompuesta, 1 si funciona, y sea X_n la variable que denota el estado de la máquina el día n . Si el estado de la máquina solo depende del día anterior, es decir, si se sabe que:

$$\begin{aligned} PC X_n = 0 | X_{n-1} = 0 &= 1 - \alpha, \\ PC X_n = 1 | X_{n-1} = 0 &= \alpha, \\ PC X_n = 1 | X_{n-1} = 1 &= 1 - \beta, \\ PC X_n = 0 | X_{n-1} = 1 &= \beta, \end{aligned}$$

se puede preguntar por la probabilidad de que la máquina funcione todos los días.

Sean $\pi_0(0) = PC X_0 = 0$ y $\pi_0(1) = PC X_0 = 1$, entonces $\pi_0(0) + \pi_0(1) = 1$

por lo tanto

$$\begin{aligned} PC X_0 = 1, X_1 = 1, \dots, X_n = 1 &= PC X_0 = 1 \cdot PC X_1 = 1 | X_0 = 1 \cdot PC X_2 = 1 | X_0 = 1, X_1 = 1 \cdot \\ &PC X_3 = 1 | X_0 = 1, X_1 = 1, X_2 = 1 \cdot \dots \cdot \\ &PC X_n = 1 | X_0 = 1, \dots, X_{n-1} = 1 \\ &= PC X_0 = 1 \cdot PC X_1 = 1 | X_0 = 1 \cdot \dots \cdot PC X_n = 1 | X_{n-1} = 1 \\ &= \pi_0(1) (1-\beta)^n \end{aligned}$$

Una cadena de Markov está completamente definida por su matriz de probabilidades de transición y la especificación de una distribución de probabilidad del proceso al tiempo cero.

Las probabilidades de transición se pueden acomodar en un arreglo cuadrangular cuya dimensión es igual al número de estados del proceso, a dicho arreglo se le llama Matriz de probabilidades de Transición y tanto sus renglones como sus columnas son los posibles estados del proceso, así, cada una de sus entradas (i,j)

es la probabilidad de que el proceso pase del estado i al j .

En el análisis de una cadena de Markov es necesario calcular las probabilidades de las posibles realizaciones del proceso, y para ello son de gran importancia las matrices de probabilidades de transición en n pasos. $P^n = || P_{ij}^n ||$. Aquí P_{ij}^n denota la probabilidad de que el proceso vaya del estado i al estado j en n transiciones. Formalmente: $P_{ij}^n = P\langle X_{n+m} = j | X_m = i \rangle$. Es importante observar que se está tratando solamente con procesos temporalmente homogéneos, con probabilidades de transición estacionarias, pues de otra manera $P\langle X_{n+m} = j | X_m = i \rangle$ dependería también de m .

La suposición Markoviana permite expresar inmediatamente a P_{ij}^n en términos de $||P_{ij}||$ como sigue:

TEOREMA 2.1) (Ecuación de Chapman-Kolmogorov) Sea $P = ||P_{ij}||$ la matriz de probabilidades de transición de una cadena de Markov entonces:

$$P_{ij}^n = \sum_{k=0}^{n-1} P_{ik}^r P_{kj}^s, \quad (2.1)$$

para cualquier par de enteros no-negativos r y s , que satisfacen que $r+s = n$. Para $n = 0$ se define

$$P_{ij}^0 = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Observe que la relación (2.1) es, exactamente, la fórmula para multiplicación de matrices, esto significa que $P^{(n)} = P^n$, en otras palabras, los números P_{ij}^n pueden ser considerados como las entradas en la matriz P^n , la n -ésima potencia de P .

Ejemplo (de aplicación a la genética) Considérese el caso de la

transmisión de un determinado carácter hereditario en una población cerrada. La sangre de un ser humano está constituida por glóbulos blancos y rojos. Los glóbulos rojos son los encargados de transportar oxígeno, en ellos existe un pigmento proteínico llamado hemoglobina tipo A representada por un par de genes AA. Cuando en la hemoglobina ocurren alteraciones genéticas, se producen genes del tipo SS que cambian la estructura de los glóbulos rojos propiciando su destrucción, lo que trae como consecuencia una anemia.

Considerando una población cerrada que se genera a partir de dos individuos, se pueden obtener seis posibles tipos de unión diferentes, los cuales pueden producir tres tipos diferentes de genotipos (AA, AS, SS). En la siguiente tabla se muestran estos tipos de unión y las probabilidades de cada genotipo:

TIPO DE UNIÓN	PROBABILIDADES DEL GENOTIPO HIJO		
	AA	AS	SS
1 AA con SS	0	1	0
2 AA con AS	1/2	1/2	0
3 AS con AS	1/4	1/2	1/4
4 AS con SS	0	1/2	1/2
5 AA con AA	1	0	0
6 SS con SS	0	0	1

Supóngase que se tiene una población en la que (como en algunos grupos aislados) se restringe la reproducción de la especie a efectuarse únicamente entre descendientes. Partiendo de un determinado tipo de unión se puede obtener la probabilidad de cada uno de los tipos de unión que se pueden producir en el siguiente paso, lo cual significa que se puede representar el comportamiento de esta población a través de una cadena de Markov, en donde cada tipo de unión representa un posible estado de la cadena.

La matriz de probabilidades de transición asociada al problema, es la siguiente:

$$P = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \left(\begin{array}{cccccc} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/4 & 0 & 1/4 & 0 \\ 1/8 & 1/4 & 1/4 & 1/4 & 1/16 & 1/16 \\ 0 & 0 & 1/4 & 1/2 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right) \end{matrix}$$

La matriz de probabilidades de transición en dos pasos (P^2), de acuerdo al teorema 2.1, se puede calcular como sigue:

$$P^2 = P \times P$$

$$P^2 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \left(\begin{array}{cccccc} 1/8 & 1/4 & 1/4 & 1/4 & 1/16 & 1/16 \\ 1/32 & 5/16 & 3/16 & 1/16 & 25/64 & 1/64 \\ 1/32 & 3/16 & 5/16 & 3/16 & 9/64 & 9/64 \\ 1/32 & 1/16 & 3/16 & 5/16 & 1/64 & 25/64 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right) \end{matrix}$$

Frecuentemente es interesante analizar el desarrollo asintótico de un proceso, es decir, encontrar el comportamiento de P^n_{ij} cuando $n \rightarrow \infty$. Para ello es necesario introducir algunos principios de clasificación de estados de una cadena de Markov.

2.2 CLASIFICACION DE ESTADOS

Un estado j es llamado accesible del estado i si para algún entero $n \geq 0$, $P_{ij}^n > 0$, es decir, el estado j es accesible del estado i , si hay probabilidad positiva de que en un número finito de transiciones el estado j pueda ser alcanzado empezando en el estado i .

Dos estados i, j son llamados comunicantes si cada uno es accesible del otro, y se denota como $i \leftrightarrow j$.

El concepto de comunicación es una relación de equivalencia mediante la cual se puede particionar el espacio de estados, en clases de equivalencia.

Los estados en una clase de equivalencia son aquellos que son comunicantes con cada uno de los otros.

Se dice que una cadena de Markov es irreducible si la relación de equivalencia induce una sola clase. En otras palabras, un proceso es irreducible si todos sus estados son comunicantes con los otros.

Se define el periodo del estado i ($d(i)$) como el máximo común divisor de todos los enteros $n \geq 1$ para los cuales $P_{ii}^n > 0$.

Si $P_{ii}^n = 0$ para todo $n \geq 1$, se define $d(i) = 0$. Nótese que el periodo es una propiedad de clase, es decir: Si $i \leftrightarrow j$, entonces $d(i) = d(j)$.

Una cadena de Markov se llama aperiódica si cada uno de sus estados tiene periodo 1.

Supóngase un estado i arbitrario pero fijo. Se define para cada entero n

$$f_{ii}^n = P(X_n = i, X_v \neq i \quad v = 1, 2, \dots, n-1 \mid X_0 = i)$$

esto es, f_{ii}^n es la probabilidad de que, partiendo del estado i , el primer regreso al mismo estado ocurra en la n -ésima transición. Claramente $f_{ii}^1 = P_{ii}$ y f_{ii}^n puede ser calculada recursivamente como sigue:

$$P_{ii}^0 = \sum_{k=0}^n f_{ii}^k P_{ii}^{n-k} \quad n \geq 1 \quad (2.2)$$

en donde se define $f_{ii}^0 = 0$ para todo valor de i .

De acuerdo a lo visto en el capítulo anterior, la función generatriz de probabilidad de la sucesión de probabilidades de transición $\langle P_{ij}^n \rangle$, es

$$P_{ij}(s) = \sum_{n=0}^{\infty} P_{ij}^n s^n \quad \text{para } |s| < 1$$

De manera similar se define la función generatriz de probabilidad de la sucesión $\langle f_{ij}^n \rangle$:

$$F_{ij}(s) = \sum_{n=0}^{\infty} f_{ij}^n s^n \quad \text{para } |s| < 1.$$

en donde f_{ij}^k es la probabilidad de que, partiendo del estado i , la cadena visite por primera vez el estado j en la k -ésima transición. Partiendo de la siguiente propiedad:

$$\text{Si } A(s) = \sum_{k=0}^{\infty} a_k s^k \quad \text{y} \quad B(s) = \sum_{l=0}^{\infty} b_l s^l,$$

entonces:

$$\begin{aligned}
 A(s)B(s) &= \left(\sum_{k=0}^{\infty} a_k s^k \right) \left(\sum_{l=0}^{\infty} b_l s^l \right) \\
 &= a_0 b_0 + (a_1 b_0 + b_1 a_0) s + (a_2 b_0 + a_1 b_1 + a_0 b_2) s^2 + \dots \\
 &= \sum_{k=0}^{\infty} s^k \left(\sum_{j=0}^k a_j b_{k-j} \right) \\
 &= \sum_{k=0}^{\infty} c_k s^k = C(s)
 \end{aligned}$$

donde $C_k = a_0 b_k + a_1 b_{k-1} + \dots + a_k b_0$ (2.3)

Si ahora se identifican las a_k 's con las f_{ii}^k 's y las b_l 's con las P_{ii}^l 's, entonces comparando las ecuaciones (2.2) y (2.3),

$$P_{ii}^n = \sum_{k=0}^n f_{ii}^k P_{ii}^{n-k} \quad n > 1 \quad \text{y} \quad C_r = a_0 b_r + a_1 b_{r-1} + \dots + a_r b_0,$$

se obtiene:

$$F_{ii}(s) P_{ii}(s) = P_{ii}(s) - 1 \quad \text{para} \quad |s| < 1$$

o bien

$$P_{ii}(s) = \frac{1}{1 - F_{ii}(s)} \quad \text{para} \quad |s| < 1.$$

Esto significa que basta conocer la función generatriz de probabilidad de la sucesión de probabilidades de primer regreso $\langle f_{ii}^n \rangle$ a un estado, para a partir de ella obtener la función generatriz de probabilidad de las probabilidades de transición $\langle P_{ii}^n \rangle$.

Un estado i es recurrente si, partiendo la cadena de dicho estado, la probabilidad de que regrese a él en algún momento, es 1, es decir,

$$\sum_{n=1}^{\infty} f_{ii}^n = 1.$$

Un estado i es transitorio si, partiendo la cadena de dicho estado, existe una probabilidad positiva de que nunca regrese a él, en otros términos.

$$\sum_{n=1}^{\infty} f_{ii}^n < 1.$$

Una clase de equivalencia (o simplemente una clase) es recurrente si todos los estados en dicha clase, lo son.

Un estado i es absorbente si $P_{ii} = 1$. Esto significa que si la cadena llega a un estado absorbente, ya nunca sale de él.

Un estado absorbente es también recurrente. En el caso de estados absorbentes las probabilidades de primera visita se conocen como probabilidades de absorción.

Ejemplo:

1) Considérese una cadena de Markov con matriz de transición

	0	1	2	3	4	5
0	1	0	0	0	0	0
1	1/4	1/2	1/4	0	0	0
2	0	1/5	2/5	1/5	0	1/5
3	0	0	0	1/3	1/3	1/3
4	0	0	0	1/2	0	1/2
5	0	0	0	1/4	0	3/4

para clasificar los estados e identificar las clases en la cadena, se utiliza una matriz auxiliar, anotando en ella "0" cuando no es posible la transición y "+" cuando si lo es en 1 o más pasos.

0	1	2	3	4	5
0	1	2	3	4	5
1	2	3	4	5	6
2	3	4	5	6	7
3	4	5	6	7	8
4	5	6	7	8	9
5	6	7	8	9	0

De aquí se observa lo siguiente:

Estados absorbentes: 0

Estados Recurrentes: 3,4,5

Estados Transitorios: 1,2

Clases cerradas e irreducibles: $C_1 = \{0\}$, $C_2 = \{3,4,5\}$.

$d = \text{máx. c. d. } \langle n \geq 1 \mid P_{11}^n > 0 \rangle = \text{máx. c. d. } \langle 1,2,3 \rangle = 1$

entonces el periodo de la cadena es 1, es decir, la cadena es aperiódica.

2) Ejemplo de genética (continúa): Para clasificar los estados e identificar las clases en la cadena, se utiliza la siguiente matriz auxiliar:

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	0
5	6	7	8	9	0	1
6	7	8	9	0	1	2

de ella se observa lo siguiente:

Estados absorbentes: 5, 6

Estados recurrentes: 5, 6

Estados transitorios: 1, 2, 3, 4

Clases cerradas e irreducibles: $C_1 = \{5\}$, $C_2 = \{6\}$

$d = \text{máx. c. d. } \langle n \geq 1 \mid P_{11}^n > 0 \rangle = 1$, entonces la cadena es

aperiódica.

De aquí se desprende que los descendientes, tarde o temprano alcanzarán los genes tipo AA o SS.

2.3 ALGUNOS RESULTADOS IMPORTANTES

Supóngase una Cadena de Markov con espacio de estados S y un conjunto de estados transitorios T . Llámese N' al tiempo que pasa la cadena en estados transitorios, entonces

$$N' = \sum_{i \in T} N_i(\omega)$$

en donde $N_i(\omega)$ es el número de veces que la cadena pasa por el estado i en un número infinito de pasos.

Así, $N = N' + 1$ es el tiempo de absorción.

El tiempo medio de absorción de la cadena está, por tanto, determinado por el valor esperado de N

$$m_j = E \langle N \mid X_0 = j \rangle = 1 + E \langle N' \mid X_0 = j \rangle$$

m_j = tiempo medio de absorción, y j estado transitorio. Los m_j 's se pueden calcular resolviendo el siguiente sistema de ecuaciones lineales:

$$m_j = 1 + \sum p_{jk} m_k \quad j \in T.$$

Si para una cadena irreducible existe un conjunto de números π_i tales que

$$1) \pi_i > 0 \quad \text{para todo valor de } i \in S$$

$$2) \sum_{i=0}^{\infty} \pi_i = 1$$

$$3) \pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij} \quad \text{para todo valor de } j$$

entonces se dice que $\langle \pi_i \rangle$ es la distribución estacionaria de la

cadena.

TEOREMA 2.2) Si X es una cadena de Markov irreducible, recurrente positiva y aperiódica, con distribución estacionaria (π_i) , entonces

$$\pi_i = \lim_{n \rightarrow \infty} P_{ii}^n$$

Nótese que si: i es un estado recurrente

$$\lim_{n \rightarrow \infty} P_{ii}^n = \pi_i.$$

Si i es transitorio

$$\lim_{n \rightarrow \infty} P_{ii}^n = 0$$

pues $\sum_{n=1}^{\infty} P_{ii}^n < \infty$

(Esto implica que una cadena nunca permanecerá indefinidamente en una clase transitoria, sino que siempre, en algún momento, pasará a alguna clase recurrente sin regresar nuevamente a la transitoria).

Si i es recurrente y j transitorio

$$\lim_{n \rightarrow \infty} P_{ij}^n = 0.$$

y, finalmente, si i, j son transitorios

$$\lim_{n \rightarrow \infty} P_{ij}^n = 0.$$

Ahora, ¿Qué ocurre cuando i es transitorio y j recurrente?

Si j pertenece a una clase recurrente C , la cadena será absorbida por C , con una probabilidad $\pi_1(C)$, en donde, si se denota por $\pi_1^n(C)$ a la probabilidad de absorción en el n -ésimo paso:

$$\pi_1(C) = \sum_{n=1}^{\infty} \pi_1^n(C)$$

y

$$\pi_1^1(C) = \sum_{j \in C} P_{1j}$$

entonces:

$$\pi_1^n(C) = \sum_{j \in T} P_{1j} \pi_j^{n-1}(C) \quad n \geq 2$$

$$\begin{aligned} \pi_1(C) &= \pi_1^1(C) + \sum_{n=2}^{\infty} \pi_1^n(C) \\ &= \pi_1^1(C) + \sum_{n=2}^{\infty} \sum_{j \in T} P_{1j} \pi_j^{n-1}(C) \\ &= \pi_1^1(C) + \sum_{j \in T} P_{1j} \sum_{n=2}^{\infty} \pi_j^{n-1}(C) \\ &= \pi_1^1(C) + \sum_{j \in T} P_{1j} (\pi_j^1(C) + \pi_j^2(C) + \dots) \\ &= \pi_1^1(C) + \sum_{j \in T} P_{1j} \pi_j(C) \end{aligned}$$

Ejemplo: Sea X una cadena de Markov con espacio de estados $S = \{1, 2, 3\}$ y matriz de transición

$$P = \begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.6 & 0 & 0.4 \\ 0 & 0.4 & 0.6 \end{bmatrix}$$

se puede observar fácilmente que todos los estados son recurrentes y aperiódicos. Se desea calcular la distribución estacionaria, es decir, números π_1, π_2, π_3 que satisfagan

$$1) \pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}$$

$$2) \sum_{i=1}^{\infty} \pi_i = 1$$

Se tiene que

$$\pi_1 = \pi_1(0.3) + \pi_2(0.6) \dots\dots\dots (1)$$

$$\pi_2 = \pi_1(0.5) \dots\dots\dots + \pi_3(0.4) \dots\dots\dots (2)$$

$$\pi_3 = \pi_1(0.2) + \pi_2(0.4) + \pi_3(0.6) \dots\dots\dots (3)$$

si se supone $\pi_1 = 6$ entonces $\pi_2 = 7$ (de la ec. 1)

$\pi_3 = 10$ (de la ec. 2)

Para que satisfagan además

$$\sum_{i=1}^{\infty} \pi_i = 1$$

se dividen los valores obtenidos para $\{\pi_i\}_{i=1}^3$ entre 23, que es la suma de 10, 7, y 6, y entonces se obtiene que

$$\pi = (\pi_1, \pi_2, \pi_3) = (6/23, 7/23, 10/23)$$

es la distribución estacionaria de X.

Ejemplo de genética (continúa): Para obtener ahora la distribución estacionaria de la cadena, se tiene:

$$\begin{aligned} \pi_1 &= \pi_3 \\ \pi_2 &= \pi_2/2 + \pi_3/4 + \pi_5/4 \\ \pi_3 &= \pi_1/6 + \pi_2/4 + \pi_3/4 + \pi_4/4 + \pi_5/16 + \pi_6/16 \\ \pi_4 &= \pi_3/4 + \pi_4/2 + \pi_6/4 \\ \pi_5 &= \pi_5 \\ \pi_6 &= \pi_6 \end{aligned}$$

Sean $\pi_1 = \pi_5 = \pi_6 = 1/6$, entonces:

$$\pi_3 = 1/6$$

$$\pi_2 = 1/6$$

$$\pi_4 = 1/6$$

y por lo tanto $\pi = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$ es la distribución estacionaria de la cadena.

CAPITULO III

PROCESOS DE RAMIFICACION UNIVARIADOS

La teoría de los Procesos de Ramificación, desarrollada en los últimos tiempos, ha sido utilizada en el estudio de fenómenos genéticos, químicos y físicos, principalmente. Su origen se remonta al siglo XVIII cuando F. Galton y H. W. Watson crean un modelo matemático (Modelo de Galton-Watson) para estudiar un problema demográfico. El caso más sencillo de los Procesos de Ramificación es el univariado o con un solo tipo, en el que se analiza el desarrollo de una población basándose en una característica común a todos los individuos que la integran.

El modelo de Galton-Watson también conocido como Proceso de Galton-Watson, no obstante ser un caso particular de los Procesos de Ramificación, ha sido fundamental para el desarrollo tanto teórico como práctico de los mismos, y por ello es indispensable su estudio para la comprensión del comportamiento de los casos generales.

El objetivo del presente capítulo es presentar el Proceso de Galton-Watson univariado, y sus principales resultados característicos, para lo cual se divide en cuatro secciones. En la primera de ellas se hace un bosquejo histórico del surgimiento del Proceso de Galton-Watson; en la segunda sección se formaliza la definición de dicho proceso y se revisan algunas de sus características como la distribución de probabilidad, la función generatriz de probabilidad, media y varianza, y probabilidad de extinción asociada. En la tercera sección se analiza el comportamiento asintótico del proceso, y en la cuarta se presentan las probabilidades estacionarias del mismo así como algunas propiedades de ellas.

3.1 INTRODUCCION HISTORICA DEL PROCESO DE GALTON-WATSON.

El análisis matemático del Proceso de Galton-Watson se desarrolló motivado, principalmente, por el deseo de encontrar una solución para un problema demográfico conocido como Problema de extinción de familias: Durante mucho tiempo se ha observado la extinción de familias notables, y que apellidos que en algún tiempo fueron muy comunes, ahora son raros o han desaparecido totalmente. Se hicieron varias conjeturas al respecto, entre ellas, dado que se ha observado que la tendencia es universal, se dedujo rápidamente, como una explicación al fenómeno, que una evolución en bienestar económico y capacidad intelectual, es necesariamente acompañada por una disminución en la "fertilidad". Sin embargo, se ha observado también que, por leyes de cambios naturales, continuamente están desapareciendo una gran proporción de familias, lo cual evidentemente significa que, mientras se desconozca dicha proporción, no se puede afirmar que una disminución en el número de integrantes de una misma familia sea un signo de "fertilidad disminuida".

Alrededor de 1874, Francis Galton y H. W. Watson, se abocaron al estudio del problema. Galton no de acuerdo con la hipótesis de que las familias distinguidas eran más tendientes a desaparecer, optó por estudiarla determinando, primeramente, la probabilidad de que una familia ordinaria desaparezca, utilizando para ello datos de fertilidad de toda la población, y con ello planteó el problema así: Sean P_0, P_1, P_2, \dots las probabilidades de que un hombre tenga 0, 1, 2, ... hijos, respectivamente, y que cada uno de ellos tenga a su vez, iguales probabilidades de tener hijos, y así sucesivamente. ¿Cuál es la probabilidad de que la descendencia de hijos (varones) se extinga después de r generaciones?, o más generalmente, ¿Cuál es la probabilidad de un número prefijado de

descendientes (varones) en una generación dada ?. La solución dada por Watson está basada en estudios subsecuentes, y por desarrollos puramente algebraicos concluye, erróneamente, que toda familia se extinguirá aun cuando su tamaño se incremente de una generación a la siguiente. Separadamente, Galton en 1981, estudió estadísticas de las tasas reproductivas de nobles ingleses concluyendo que un factor en la disminución de dichas tasas es la tendencia de los nobles a contraer matrimonio con herederas. Una heredera que proviene de una familia sin hijos (varones), debería esperar tener, por herencia, una fertilidad menor que la ordinaria. Galton y Watson no llegaron a una solución correcta para el problema; no fué sino hasta los años 30's que se presentó la primera determinación correcta y completa de la probabilidad de extinción para el proceso de Galton-Watson, y fué dada por J. F. Steffensen (1930 - 1932). Mas tarde, en 1938, el mismo problema fué tratado por Kolmogorov, quien determinó la forma asintótica de la probabilidad de que una familia continúe existiendo un número grande de generaciones.

El modelo matemático de Galton y Watson parece haber sido olvidado por muchos años después de su creación, pues el siguiente estudio conocido fué el de R. A. Fisher en 1922 y 1930, quien usó un modelo idéntico para estudiar la sobrevivencia de la progenie de un gene mutante y las variaciones aleatorias en la frecuencia de genes; en 1927, Haldane aplicó el modelo a la genética; La idea de Galton fué utilizada por A. J. Lotka en 1931, quien tomó datos de fertilidad americana para determinar la probabilidad de extinción de una línea de descendientes varones; en 1935, por Semenovoff en los fundamentos de su teoría de cadenas de reacciones químicas, y en 1938 por W. Shockley y J. R. Pierce, en el estudio de la multiplicación de electrones en un dispositivo de detección electrónica. Después de 1940 se incrementó el interés en el modelo por la analogía del comportamiento de las familias con las cadenas de reacciones y por el creciente interés general en las

aplicaciones de la Teoría de la Probabilidad.

El modelo original del proceso de Galton-Watson no solo está relacionado con diversas líneas de desarrollo de los procesos estocásticos como el proceso de nacimiento-muerte, sino también con trabajos de ecuaciones funcionales e iteración de funciones y con la Teoría de radiaciones cósmicas. Si bien algunos de los problemas mencionados han requerido ser estudiados por medio de modelos más elaborados, varios de dichos modelos se pueden considerar como descendientes directos del proceso de Galton-Watson.

3.2 EL PROCESO DE GALTON-WATSON: DEFINICION Y CONCEPTOS BASICOS.

Considérense objetos que pueden generar nuevos objetos de su mismo tipo, por ejemplo: personas, bacterias o neutrones. Un conjunto inicial de objetos, llamado 0-ésima generación, producirá nuevos objetos los cuales a su vez generarán a los que formarán la segunda generación y así sucesivamente. La descripción más simple del proceso de Galton-Watson es la siguiente: Supóngase que se desea conocer el tamaño de cada generación sin importar cuando nació cada individuo. Se denota por Z_0, Z_1, Z_2, \dots al número de individuos en la 0-ésima, primera, segunda, ... generaciones respectivamente (también se podrían interpretar Z_0, Z_1, \dots como el tamaño de la población en una sucesión de puntos en el tiempo). Supóngase también que :

1) Si el tamaño de la n -ésima generación es conocido, la ley de probabilidad que gobierna a las siguientes generaciones, no depende del tamaño de las generaciones anteriores a la n -ésima, es decir, Z_0, Z_1, \dots forman una cadena de Markov (Véase capítulo II).

2) Las probabilidades de transición de la cadena no varían con el tiempo, esto es, el proceso es estacionario (Ver capítulo I).

3) Diferentes objetos de una misma generación, no interfieren entre sí, esto es, el número de individuos que generen miembros distintos de una generación, son variables aleatorias independientes.

4) $Z_0 = 1$, es decir, la población inicial consta de un solo individuo.

Sean Z_n el número de individuos en la n -ésima generación pertenecientes a una misma población o familia. Z_0, Z_1, Z_2, \dots por su interpretación en el proceso, solo toman valores enteros no-negativos y forman la sucesión de variables aleatorias en la cadena de Markov.

La distribución de probabilidad de Z_1 está descrita por la igualdad:

$$P(Z_1 = k) = p_k, \quad k = 0, 1, 2, \dots$$

en donde p_k no depende de n y representa la probabilidad de que un individuo de la n -ésima generación tenga k hijos en la generación $n+1$. Entonces:

$$\sum_{k \in \mathbb{N}} p_k = 1.$$

y la distribución condicional de Z_{n+1} dado $Z_n = k$, por la suposición 3. es la suma de k variables aleatorias independientes cada una de ellas distribuida como Z_1 . Si $Z_n = 0$, Z_{n+1} tiene probabilidad 1 de ser cero, cualquiera que sea el valor de n .

Se definen entonces las probabilidades de transición del Proceso de Galton-Watson como:

$$P_{ij} = P(Z_{n+1} = j \mid Z_n = i), \quad n, i, j = 0, 1, 2, \dots$$

y se pueden calcular de manera sencilla utilizando la función generatriz de probabilidad de Z_n , la cual se puede obtener de manera recurrente mediante el teorema que afirma que dicha función generatriz es el n -ésimo iterado $f_n(s)$, es decir, que $f_0(s) = s$, $f_1(s) = f(s)$, $f_{n+1}(s) = f(f_n(s))$; $n = 1, 2, \dots$

En consecuencia, $f_{m,n} = f_m(f_n(s))$; $m, n = 0, 1, 2, \dots$ y en particular $f_{n+1} = f_n(f(s))$

Este resultado se puede visualizar como sigue: Si se denota por $f_m(s)$ a la función generatriz de Z_m , por la definición de $f(s)$,

se tiene que

$$f(s) = \sum_{k=0}^{\infty} P(Z_1 = k) s^k = f_1(s)$$

en donde si $Z_n = k$, y Y_i es el número de descendientes del i -ésimo elemento de la generación n , entonces $Z_{n+1} = Y_1 + Y_2 + \dots + Y_k$.
Para una k fija,

$$f_{n+1}(s) = f_{\sum_{i=1}^k Y_i}(s)$$

y como las Y_i 's son independientes e idénticamente distribuidas, se obtiene que

$$f_{(n+1)}(s) = \prod_{i=1}^k f_{Y_i}(s) = [f(s)]^k \quad \dots \dots \dots (3.1)$$

aplicando la definición:

$$f_{(n+1)}(s) = \sum_{j=0}^{\infty} P(Z_{n+1} = j | Z_n = k) s^j \quad \dots \dots \dots (3.2)$$

entonces utilizando la fórmula de probabilidad total se llega a

$$\begin{aligned} f_{n+1}(s) &= \sum_{j=0}^{\infty} P(Z_{n+1} = j) s^j \\ &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} P(Z_n = k) P(Z_{n+1} = j | Z_n = k) s^j \\ &= \sum_{k=0}^{\infty} [P(Z_n = k) \sum_{j=0}^{\infty} P(Z_{n+1} = j | Z_n = k) s^j] \end{aligned}$$

Por (1) y (2)

$$= \sum_{k=0}^{\infty} P(Z_n = k) [f(s)]^k = f_{(n)}(f(s))$$

De la misma forma:

$$\begin{aligned} f_{(n)}(f(s)) &= f_{(n-1)}(f(f(s))) = f_{(n-2)}(f(f(f(s)))) \\ &= \underbrace{f(f(f(\dots f(f(f(s)))) \dots))}_{n \text{ veces}} = f_{(n)}(f(s)) \end{aligned}$$

es decir, la función generatriz de probabilidad de Z_n es el n -ésimo iterado de $f(s)$ ($f_n(s)$).

Como Z_1, Z_2, \dots, Z_n , son variables aleatorias, cada una de ellas tiene esperanza y varianza, y se puede encontrar para ellas, una expresión general. Si se denota por m a $EC(Z_1)$ y se parte de que $EC(Z_n) = f_n'(1)$ y $f_n(1) = 1$, es fácil concluir que $EC(Z_n) = m^n$, $n = 0, 1, 2, \dots$. Si se denota por σ^2 a la $Var(Z_1)$, y se sabe que $Var(Z_n) = f_n''(1) + f_n'(1) - [f_n'(1)]^2$, se deriva la ecuación (1) y se obtiene

$$Var(Z_n) = \begin{cases} \sigma^2 m^n (m^n - 1) / (m^2 - m), & m \neq 1 \\ n\sigma^2, & m = 1 \end{cases}$$

Ejemplo 1: Considérese un proceso de ramificación con distribución de probabilidad de Z_1

$$P_k = \frac{1}{2} k + \frac{1}{4}(k-1) + \frac{1}{4}(k-2)$$

y función generatriz de probabilidad

$$f(s) = \frac{1}{2} + \frac{1}{4}s + \frac{1}{4}s^2.$$

Derivando esta función es posible calcular el valor esperado de Z_1 , por lo tanto

$$f'(1) = m = 3/4$$

y

$$\sigma^2 = EC(Z_1^2) - m^2 = 11/16$$

es decir que, la esperanza y la varianza de Z_n son $m^n = (3/4)^n$, y

$$Var(Z_n) = \frac{11}{16} (3/4)^n [(3/4)^n - 1] / [(3/4) - 1].$$

Ejemplo 2: Supóngase un proceso de ramificación con distribución

de probabilidad

$$p_k = \frac{1}{4} k + \frac{1}{4}(k-1) + \frac{1}{2}(k-2)$$

y función generatriz de probabilidad

$$f(s) = \frac{1}{4} + \frac{1}{4}s + \frac{1}{2}s^2$$

El valor esperado de Z_1 es $m = 5/4$, y la variancia es $\sigma^2 = 11/16$, por lo tanto para Z_n los valores correspondientes son $m^n = (5/4)^n$, y $\text{Var}(Z_n) = (11/16)(5/4)^n[(5/4)^n - 1] / [(5/4) - 1]$.

Antes de poder examinar con detalle las propiedades y el comportamiento de este proceso, es necesario clasificar los estados de la cadena. La naturaleza del espacio de estados y la clasificación dependen críticamente de la distribución de los descendientes $\langle P_i \rangle$. El espacio de estados contendrá al estado cero, el cual es un estado absorbente si es alcanzado, pues $p_{00}=1$.

En el caso especial en que $p_0 + p_1 = 1$, el espacio de estados se reduce a $S = \{0, 1\}$ y a lo más habrá un individuo en la n -ésima generación. En particular cuando $p_0 = 1$ y $p_1 = 0$, se tiene que $Z_1 = 1$ y $Z_n = 0$ para $n > 1$, por tanto, el estado 1 es transitorio y el cero es alcanzado con probabilidad 1. Si $p_0 = 0$ y $p_1 = 1$, entonces $Z_n = 1$ para todo n , el estado 1 es absorbente y el cero nunca es alcanzado.

Si $0 < p_0 < 1$ con $p_1 = 1 - p_0$, entonces $Z_0 = 1, \dots, Z_{N-1} = 1$, $Z_n = 0$ para $n \geq N$ cuando N es una variable aleatoria geométrica. En este caso el estado 1 es transitorio y el estado 0 es alcanzado con probabilidad 1.

Observe también que si $p_0 = 0$, entonces la sucesión $\langle Z_n \rangle$ es no-decreciente y entonces el estado cero nunca es alcanzado. En este caso si $0 \leq p_1 \leq 1$ el espacio de estados es $S = \{0, 1, 2, \dots\}$ y

todos los estados excepto el cero, son transitorios.

Si $0 < p_0 \leq p_0 + p_1 < 1$, entonces los estados 1, 2, 3, ... son todos transitorios con el estado cero absorbente. De cualquier manera, el estado cero puede o no ser alcanzado y esta es la pregunta que se quiere considerar con más detalle. En otras palabras, empezando con $Z_0 = 1$, se quiere determinar bajo que condiciones se extinguirá la cadena de ramificación, es decir, la probabilidad de que la cadena se absorba en el estado cero.

Si se define extinción como el evento de que la sucesión aleatoria $\langle Z_n \rangle$ consista de ceros excepto para un número finito de valores de n , como Z_n solo toma valores enteros positivos, extinción es también el evento $Z_n \rightarrow 0$. Además también se tiene que $P\{Z_{n+1} = 0 \mid Z_n = 0\} = 1$, y por la definición de f_n

$$P\{Z_n \rightarrow 0\} = \lim_{n \rightarrow \infty} f_n(0).$$

Si se denota por q a dicha probabilidad, entonces q es la probabilidad de que a partir de una cierta generación (n), la población o familia no tenga ningún elemento. Dependiendo del valor de m , q será 1 o menor que 1, como lo muestra el siguiente teorema. Si $m \leq 1$, la familia seguramente desaparecerá.

TEOREMA 3.1: Si $m = E(Z_1) \leq 1$, la probabilidad de extinción q , es uno. Si $m > 1$, q es la única solución no-negativa menor que uno, de la ecuación $s = f(s)$.

DEMOSTRACION: Se sabe, por inducción, que $f_n(0) < 1$, $n = 0, 1, \dots$ y se ha observado que $0 = f_0(0) \leq f_1(0) \leq f_2(0) \leq \dots \leq q = \lim_{n \rightarrow \infty} f_n(0)$ como $f_{n+1}(0) = f(f_n(0))$ y $f(0) = \lim_{n \rightarrow \infty} f_{n+1}(0) = q$, se obtiene que $q = f(q)$ y $0 \leq q \leq 1$. Si $m \leq 1$, entonces $f(s) < 1$ para $0 \leq s < 1$, y utilizando la ley de la media para expresar $f(s)$ en términos de $f(1)$, se tiene $f(s) > s$, $0 \leq s < 1$, y $q = 1$.

Si $m > 1$, entonces $f(s) < s$ cuando $s < 1$, pero $f(0) > 0$. Por lo tanto la ecuación $s = f(s)$ tiene al menos una solución en el intervalo semiabierto $[0,1)$. Si existen dos soluciones, s_0 y t_0 con $0 \leq s_0 < t_0 < 1$, el teorema de Rolle implica la existencia de ξ y η , $s_0 < \xi < t_0 < \eta < 1$, tales que $f'(\xi) = f'(\eta) = 1$, lo cual es imposible porque f es estrictamente convexa. Ahora, $\lim f_n(0)$ no puede ser uno porque $\{f_n(0)\}$ es una sucesión no decreciente, mientras que $f_{n+1}(0) = f(f_n(0))$ podría ser menor que $f_n(0)$ si $f_n(0)$ fuera ligeramente menor que 1. Por lo tanto, q debe ser la única solución de $s = f(s)$ en $[0,1)$.

Es importante ver que la sucesión $\langle Z_n \rangle$ no permanece positiva y acotada, sino que se va a infinito o a cero, y que cualquiera que sea el valor finito de $m = ECZ_1$, se tiene que:

$$\lim_{n \rightarrow \infty} PCZ_n = k = 0, \quad k = 1, 2, \dots$$

más aún, $Z_n \rightarrow \infty$ con probabilidad $1-q$ y a cero con probabilidad q .

Ejemplo 1 (continúa): Para determinar la probabilidad de extinción del proceso, dado que $m = 3/4 < 1$, el teorema 3.1 afirma que

$$q = 1$$

Ejemplo 2 (continúa): En este caso el valor obtenido para m es $5/4 > 1$, por lo tanto el teorema 3.1, afirma que la probabilidad de extinción asociada a este proceso es la única solución no-negativa y menor que 1 de la ecuación $s = f(s)$, entonces,

$$q = 1/2$$

3.3 COMPORTAMIENTO ASINTOTICO DE Z_n

Para poder conocer con mayor detalle el comportamiento del proceso, es necesario estudiar la distribución límite de Z_n cuando n es grande, y el desarrollo de la sucesión aleatoria Z_1, Z_2, \dots , pero en general no es posible obtenerla en forma explícita excepto cuando $m = 1$, por lo que se intentará analizar el comportamiento de la distribución de Z_n a través de la distribución de una nueva variable aleatoria W_n definida como $W_n = Z_n / m^n$, $n = 0, 1, \dots$, en donde m^n es la esperanza de Z_n .

Obsérvese que aplicando repetidamente el hecho de que

$$E(Z_{n+1} | Z_n) = m Z_n, \quad n = 0, 1, \dots$$

se obtiene

$$E(Z_{n+k} | Z_n) = m^k Z_n, \quad n, k = 0, 1, 2, \dots$$

y por lo tanto

$$E(W_{n+k} | W_n) = W_n, \quad n, k = 0, 1, 2, \dots$$

y aplicando la expresión para la varianza de Z_n encontrada en la sección 3.2, se llega a que si $m > 1$ y $E(Z_1^2) < \infty$, las variables aleatorias W_n convergen en media cuadrática a otra v.a. W , con probabilidad 1. Además se tiene que $E(W) = 1$, y

$$\text{Var}(W) = \frac{\text{Var}(Z_1)}{m^2 - m} > 0.$$

De donde se desprende que

$$E(W^2 | W > 0) - [E(W | W > 0)]^2 > 0$$

De acuerdo a estos resultados, es posible utilizar la distribución de W para estudiar la de Z_n cuando n es grande.

Si se definen ahora las funciones generatrices de momentos de W_n y

W respectivamente como

$$\varphi_n(s) = E(e^{-sW}) = f_n(e^{-s}, m^n)$$

y

$$\varphi(s) = E(e^{-sW}) \quad , \quad \operatorname{Re}(s) \geq 0 \quad ,$$

y si $K(u) = P(W \leq u)$ es la distribución de W. Si además la esperanza de Z_1 es mayor que 1 y la de Z_1^2 es finita, la función generatriz de momentos de mW es la función generadora de probabilidad de Z_n aplicada a la función generatriz de momentos de W, es decir, $\varphi(ms)$ satisface la relación

$$\varphi(ms) = f(\varphi(s)) \quad , \quad \operatorname{Re}(s) \geq 0 \quad , \quad \text{con } \varphi'(0) = 1$$

y la función $\varphi(-it)$, $-\infty < t < \infty$, es la única función característica que satisface dicha relación, y corresponde a una distribución con primer momento 1.

Un resultado importante es que si $m > 1$ y $E(Z_1^2) < \infty$, la distribución de $K(u)$ es absolutamente continua excepto para un salto de magnitud q en $u = 0$. Además si la probabilidad de extinción es mayor que cero y $E(Z_1) > 1$, entonces $P(Z_n = 0)$ cuando $n \rightarrow \infty$, se puede determinar con solo conocer q y la función generatriz de probabilidad de Z_n , como sigue:

$$f_n(0) = q - d[f'(q)]^n + o([f'(q)]^{2n})$$

en donde d es una constante positiva. Empero, ¿qué se puede hacer cuando no es posible conocer directamente la distribución de Z_n ? En 1936, Kolmogorov demostró que si $m = 1$ y $f''(1) < \infty$, entonces $P(Z_n > 0)$ es, aproximadamente, $2/nf''(1)$, y como

$$E(Z_n) = 1 = E(Z_n \mid Z_n \neq 0)P(Z_n \neq 0),$$

entonces

$$E\{Z_n \mid Z_n \neq 0\} \approx \frac{nf''(1)}{2}, \text{ cuando } n \rightarrow \infty.$$

Esto lleva a concluir que si no es posible conocer directamente la distribución de Z_n , se puede intentar aproximarla a través de la distribución condicional de $2Z_n / nf''(1)$ dado $Z_n \neq 0$, utilizando para ello el siguiente teorema.

TEOREMA 3.2: Supóngase $m = 1$ y $f''(1) < \infty$, entonces

$$\lim P\left(\frac{2Z_n}{nf''(1)} > u \mid Z_n \neq 0\right) = e^{-u}, \quad u \geq 0.$$

DEMOSTRACION: La función característica de la variable aleatoria $2Z_n / nf''(1)$, dado que $Z_n \neq 0$, es

$$\varphi_n(t) = \frac{f\left(e^{2it/nf''(1)}\right) - 1}{1 - f''(0)} + i, \quad n = 1, 2, \dots; -\infty < t < \infty.$$

Sea I el intervalo cerrado $[-t_0, t_0]$, excluyendo el punto $t = 0$, donde t_0 es un número positivo arbitrario. Si n es suficientemente grande, entonces utilizando un resultado que en este trabajo no se menciona ¹ se obtiene que $\varphi_n(t) \rightarrow 1/(1-it)$, $t \in I$, y como t_0 es arbitrario, $\varphi_n(t)$ debe converger para cada t , a $1/(1-it)$, que es la función característica de la distribución exponencial.

¹ LEMA: Suponga que $m = 1$ y $f''(1) < \infty$. Sea S el conjunto de puntos s tales que (a) son interiores al círculo unitario, o (b) están en el segmento del círculo unitario $\theta_0 \leq \arg s \leq \theta_0$, excluyendo el punto $s = 1$, donde θ_0 es un número positivo que se especifica en la demostración. Entonces

$$\frac{1}{1-f(s)} = \frac{1}{1-s} + \frac{nf''(1)}{2} + O(\log n), \quad s \in S, n \rightarrow \infty.$$

(Ver demostración en Harris, Theodore E., "The Theory of Branching Processes", Springer-Verlag, 1963., p.20).

3.4 ESTACIONARIDAD DE Z_n

En 2.3 se definió la distribución estacionaria para una cadena de Markov como un conjunto de números π_i , $i = 0, 1, 2, \dots$ con las siguientes características:

$$\left. \begin{aligned} \pi_i &> 0; \\ \sum_{i=0}^{\infty} \pi_i &= 1; \\ \pi_j &= \sum_{i=0}^{\infty} \pi_i P_{ij}, \quad i = 0, 1, 2, \dots \end{aligned} \right\} \quad (3.3)$$

Si se satisfacen las mismas características (3.3), pero se permite que la suma no necesariamente sea finita, se dice que el conjunto de π_i es una medida estacionaria de la cadena de Markov.

Como ya se mencionó, el Proceso de Galton-Watson es un caso particular de Procesos Markovianos, entonces si (P_{ij}) es la matriz de transición para dicho proceso, y si $0 < p_0 < 1$, existen números no-negativos π_1, π_2, \dots , no todos cero, que satisfacen las ecuaciones

$$\pi_i = \sum_{j=0}^{\infty} \pi_j P_{ji}, \quad i = 0, 1, 2, \dots$$

$$\sum \pi_i = \infty$$

llamados medida estacionaria para el Proceso de Galton-Watson, y la función generatriz

$$N(s) = \sum_{j=1}^{\infty} \pi_j s^j$$

es analítica para $|s| < q$, donde q es la probabilidad de

extinción, y satisface la ecuación

$$\Pi(f(s)) = 1 + \Pi(s), \quad |s| < \rho \quad \dots \dots \dots (3.4)$$

(conocida como ecuación de Abel) suponiendo que los números π_i están multiplicados por un factor constante, de tal forma que $\Pi(f(0)) = 1$. Si se toma $\Pi(s)$ como la función $2R(s)/f''(1)$ cuando $m = 1$, para i grande las π_i tienen la forma

$$\pi_i = \frac{2}{f''(1)} - \left(1 - \frac{2}{3} \frac{f'''(1)}{(f''(1))^2}\right) \frac{1}{i} + o\left(\frac{1}{i}\right).$$

Si $p_0 > 0$, $m = 1$ y $f^{(4)}(1) < \infty$ y τ_1, τ_2, \dots con una medida estacionaria, del proceso, la serie $\tau(s) = \sum \tau_i s^i$ converge para $|s| < 1$. Si ahora se normaliza de tal forma que $\tau(f(0)) = 1$, se tiene que τ satisface la ecuación (3.4) para $|s| < 1$ y las τ_i satisfacen que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{\infty} \pi_i = \frac{2}{f''(1)}$$

con π_i reemplazada por τ_i .

CAPITULO IV

ALGUNAS APLICACIONES A LOS PROCESOS DE MARKOV Y DE RAMIFICACION

En muchas ocasiones para analizar el comportamiento de algunos fenómenos es necesario hacer uso de modelos que permitan la representación de los mismos. En la actualidad existen modelos de todo tipo que permiten analizar fenómenos Económicos, Biológicos, Físicos, etc. Muchos de estos modelos utilizan Procesos de Markov o alguno de sus casos especiales, como marco teórico, pues gran parte de los problemas que se plantean diariamente dependen del tiempo o de algún otro parámetro que se pueda manejar de manera similar.

En el caso de los Procesos de Ramificación, no obstante lo reciente de su desarrollo, existen ya aplicaciones muy concretas a cuestiones prácticas, algunas de las cuales se presentan en este capítulo.

El objetivo de esta parte del trabajo es mostrar, a través de la presentación de algunos modelos de problemas prácticos diversas áreas, que tanto los Procesos de Markov en general y los de Ramificación en particular, no son solamente una parte más de la teoría de la probabilidad, sino que constituyen una herramienta importante para la solución de una gran variedad de problemas reales. Para lograrlo, se ha dividido el capítulo en seis secciones cada una de las cuales presenta algún problema específico y el modelo que usualmente se utiliza para resolverlo, así como su relación con los procesos mencionados.

4.1 CAMINATAS ALEATORIAS.

Considérese una caminata aleatoria que se mueve en una línea y tiene estados $0, 1, \dots, n$. La caminata tiene probabilidades p de moverse hacia la derecha (de i a $i+1$) y q de moverse hacia la izquierda (de i a $i-1$), para estados $i = 1, 2, \dots, n-1$, y los estados 0 y n son absorbentes. La matriz de probabilidades de transición asociada al proceso (P) se puede expresar de la forma

$$P = \left(\begin{array}{c|c} I & O \\ \hline R & Q \end{array} \right)$$

en donde I es la matriz identidad; Q es la matriz de probabilidades de transición entre estados transitorios, y R muestra las probabilidades de transición de los estados transitorios a los absorbentes. Se definen una matriz fundamental N como $N = (I - Q)^{-1}$ cuyas entradas n_j corresponden al número total de veces que el proceso está en el estado transitorio j , y la matriz $B = NR$, que tiene como componentes b_{ij} a las probabilidades de que, habiendo iniciado el proceso en un estado transitorio i , termine en uno absorbente j .

Ejemplo: Si se toma $n = 5$ y $p = 2/3$, se tiene que la matriz de probabilidades de transición es la siguiente:

$$P = \begin{array}{c} \begin{matrix} 0 & 5 & 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 0 \\ 5 \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix} \left(\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 2/3 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 2/3 & 0 \\ 0 & 0 & 0 & 1/3 & 0 & 2/3 \\ 0 & 2/3 & 0 & 0 & 1/3 & 0 \end{array} \right) \end{array}$$

La matriz que muestra el número de veces que, en promedio, estará el proceso en un estado transitorio determinado antes de ser absorbido es

$$N = 1/31 \begin{pmatrix} 45 & 42 & 36 & 24 \\ 21 & 63 & 54 & 36 \\ 9 & 27 & 83 & 42 \\ 3 & 9 & 21 & 45 \end{pmatrix}$$

la matriz de probabilidades de absorción cuando el proceso empieza en un estado transitorio se denota como B y sus entradas b_{ij} son

$$B = 1/31 \begin{pmatrix} 15 & 16 \\ 7 & 24 \\ 3 & 28 \\ 1 & 30 \end{pmatrix} \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix}$$

Este tipo de caminatas aleatorias es frecuentemente conocido como el "problema de la ruina del jugador". En él se piensa en dos personas que juegan un cierto juego repetidamente, teniendo, el jugador A, una probabilidad p de ganar, un capital original de i pesos mientras que el de su oponente es $n-i$. Supóngase que se juega 1 peso en cada juego. Entonces la fortuna de A se comporta de acuerdo a una caminata aleatoria como la descrita anteriormente. La absorción en n significa que A termina teniendo todo el dinero, mientras que la absorción en 0 indica que A ha sido arruinado. Si $p = 1/2$ la probabilidad de ruina es la fracción de las dos fortunas que posee el oponente. Pero si el jugador A tiene alguna ventaja en el juego tendrá una buena oportunidad de ganar aún cuando su oponente tenga un capital mucho mayor que el suyo.

4.2 EL MODELO DE DIFUSION DE EHRENFEST

El modelo de difusión de Ehrenfest es un modelo muy simple de mecánica estadística en el que se considera un gas contenido en un volumen dividido en dos regiones A y B, separadas por una membrana permeable. Supóngase que el gas tiene s moléculas, y que a cada instante una de las s moléculas es elegida al azar y movida a la otra región. Se desea conocer de que manera cambia la composición de las dos regiones, en el tiempo, por ejemplo si se inicia con todas las moléculas en una misma región, cuánto tiempo en promedio pasará para que se encuentren la mitad de las moléculas en cada región?. La respuesta se puede encontrar utilizando los métodos de Cadenas de Markov.

Una Cadena de Markov se puede formar suponiendo en principio que las moléculas son identificables. Considérense como estados un vector $\gamma = (x_1, x_2, \dots, x_s)$ donde x_j es uno si la j -ésima molécula está en la región A, y cero en otro caso. Entonces, conociendo los estados se puede saber la composición de A y de B. Hay 2^s estados, y se puede cambiar de estado cambiando, simplemente, alguna de las coordenadas de γ . De γ el proceso se puede mover a cualquiera de s estados, y cada transición ocurre con probabilidad $1/s$. Se puede pasar de un estado a otro en una sucesión de pasos, pero para regresar al mismo estado el número de pasos tiene que ser par, incluyendo el dos, por lo que el periodo de la cadena es dos. En general para una cadena con s estados, el tiempo promedio de retorno a un estado será 2^s y el tiempo promedio en un estado ó entre ocurrencias de un estado γ , es uno. Se puede calcular el tiempo promedio que toma ir de un estado a otro (m_d) como

$$m_d = \sum_{i=1}^d Q_{s-1}^i, \quad 0 < d \leq s.$$

en donde

$$Q_i^s = \sum_{k=0}^i \frac{\binom{s}{k}}{\binom{s-1}{i}}$$

$i = 0, 1, \dots, s-1$.

este tipo de cadenas se conoce como cadena microscópica, y a partir de ella se puede obtener una cadena macroscópica juntando todos los estados con el mismo número de moléculas en la región A. Los estados de la nueva cadena serán V_0, V_1, \dots, V_s , y cada V_i es el conjunto de todos los estados del proceso microscópico con i moléculas en la región A, y las probabilidades de transición son

$$P_{i,i+1} = 1 - 1/s$$

$$P_{i,i-1} = 1/s$$

$$P_{i,j} = 0 \text{ en cualquier otro caso.}$$

Este nuevo proceso también tiene periodo dos.

4.3 TEORIA DEL APRENDIZAJE.

Existen varios modelos para estudiar el aprendizaje, y uno de ellos es un modelo desarrollado por W. D. Estes. Un caso especial y relativamente simple es el siguiente: En un experimento típico, un individuo se sitúa frente a un par de luces, y se le pide que adivine cual de las luces se encenderá en seguida, la izquierda o la derecha, para lo cual tiene dos respuestas posibles. Se denota por A_0 a la respuesta "izquierda" y por A_1 a la "derecha", y si E_0 significa encender la luz izquierda, y E_1 encender la derecha, cuando el experimentador enciende alguna de las luces guarda tanto la respuesta A como la acción E , y repite el procedimiento un número grande de veces. La teoría intenta predecir el cómo cambiarán en el largo plazo las respuestas del individuo, para un comportamiento dado del experimentador.

En un gran número de experimentos el experimentador elegirá sus acciones con probabilidades fijas, dependiendo solamente de la acción del individuo. Dichas probabilidades pueden estar dadas por la siguiente matriz:

$$\begin{array}{c} A_0 \\ A_1 \end{array} \begin{array}{cc} E_0 & E_1 \\ \left[\begin{array}{cc} 1-v & v \\ w & 1-w \end{array} \right] \end{array}$$

es decir, una respuesta "izquierda" (A_0) es reforzada encendiendo la luz izquierda (E_0) con probabilidad $1-v$; en otro caso la luz derecha es encendida. Una respuesta "derecha" es reforzada con probabilidad $1-w$. Tanto v como w son números entre cero y uno, y se mantienen fijos durante todo el experimento.

El modelo supone que el individuo tiene un cierto número z

desconocido de elementos estimulantes cada uno de los cuales es conectado en cada etapa a una de las dos respuestas posibles A_1 , y las conexiones originales son desconocidas. Los supuestos, entonces, son los siguientes:

1) El individuo muestrea un subconjunto de elementos estimulantes, por medio de un proceso de ensayos independientes, en el cual cualquiera de los estímulos es muestreado con probabilidad t o no muestreado con probabilidad $1-t$.

2) Si en el conjunto muestreado hay k estímulos conectados a A_0 y l a A_1 , el individuo responde A_1 con probabilidad $l/(k+l)$. Cuando ninguno de los estímulos es muestreado se supone que la probabilidad de A_1 es la misma que si todos los estímulos hubieran sido muestreados.

3) Si el experimentador ejecuta E_0 , cualquier elemento estimulante que había sido conectado previamente a A_1 y que había sido muestreado justo antes por el individuo, es reconectado a A_0 . Igualmente, después de E_1 , todos los estímulos muestreados con conectados a A_1 .

El modelo se puede representar como una cadena de Markov con $(s+1)$ estados, en donde el estado s_i ocurre cuando exactamente i estímulos están conectados a A_1 , e $i = 0, 1, \dots, s$. Todas las cantidades de interés dependen solamente de cuantos estímulos están conectados y de que manera, por lo que son funciones de la cadena. Por ejemplo, la probabilidad de la acción A_1 del estado s_i , se obtiene como sigue:

$$P_1[A_1] = \sum_{k=0}^{s-i} \sum_{l=0}^i \binom{s-i}{k} \binom{i}{l} t^m (1-t)^{s-m} \frac{l}{m+l} + (1+t)^2 \frac{i}{s},$$

k, l no ambos cero

donde k es el número de estímulos muestreados de entre aquellos

conectados con A_0 , l de los conectados con A_1 , $m = k+1$, y el último término proviene de la suposición hecha en el caso en que ningún estímulo es muestreado. Si se reescribe la suma en términos de m , se llega a que es igual a l/s .

Para construir la matriz de transición P se deben considerar las cuatro posibilidades de combinaciones de A 's y E 's. La combinación de A_1 y E_0 , con una transición de s_i a s_j tiene probabilidades wX , donde

$$x_{ij} = \begin{cases} \sum_{k=0}^{i-1} \binom{s-1}{k} \binom{i}{i-j} t^{i-j+k} (1-t)^{s-i+j-k} \cdot \frac{i-1}{i-j+k} & \text{si } j < i \\ 0 & \text{si } j \geq i \end{cases}$$

mientras que la transición hacia atrás de s_i a s_j , por medio de la combinación A_0 y E_0 tiene probabilidades $(1-v)(Y-X)$, donde

$$y_{ij} = \begin{cases} \binom{i}{i-j} t^{i-j} (1-t)^j & \text{si } j \leq i \\ 0 & \text{si } j > i \end{cases}$$

Si se considera $x_{ij}^* = x_{s-i, s-j}$, y $y^* = y_{s-i, s-j}$, entonces vX^* y $(1-w)(Y^* - X^*)$ representan las probabilidades de transición hacia adelante, por lo que

$$P = v(X + X^* - Y) + w(X + X^* - Y^*) + (Y + Y^*) - (X + X^*) + (v + w - 1)(1-t)^s I$$

en donde el término final representa los casos en que ningún elemento estimulante es muestreado, los cuales no fueron incluidos en los dos términos anteriores.

Los valores de v y w determinan la naturaleza de la cadena de Markov, así si alguno de ellos es cero se tiene una cadena con un

estado absorbente, y si ambos son cero se tienen dos estados absorbentes. Se prueba que la probabilidad limite de una respuesta A_1 es $v/(v+w)$, y es igual a la probabilidad limite de una acción E_1 del experimentador. Es importante recalcar que el individuo no maximiza el número de respuestas correctas, sin embargo si brinda un equilibrio en el cual él está respondiendo "derecha" con la misma frecuencia con la que la luz derecha se está encendiendo. Como $v/(v+w)$ es el número promedio de respuestas A_1 por ensayo en equilibrio, y como $1/s$ es el número promedio de ocurrencias del estado s_1 , el vector

$$\delta = \left\{ \frac{1}{s} - \left[\frac{v}{v+w} \right] \right\}$$

da la desviación entre el número promedio de respuestas A_1 en un estado dado y en equilibrio. La desviación total del equilibrio es proporcional al vector de desviación δ . Por lo tanto, la desviación total puede ser grande debido a $1/s$ si el estado inicial está lejos del equilibrio $v/(v+w)$, o bien debido a que t sea pequeña.

Considérese el caso de una cadena con un solo estado absorbente en $i = 0$, es decir, $v = 0$ y $w > 0$. En este caso el individuo está siendo condicionado a dar respuestas A_0 . Si responde A_0 , la respuesta es reforzada siempre, pero si responde A_1 también es reforzada ocasionalmente, con probabilidad $1-w$. Si se denota por \bar{r} al vector r sin su primera componente, y $N = (I-Q)$, entonces $N\bar{r}$ es el número total de respuestas incorrectas o A_1 , que para un estado inicial s_1 es $1/w \cdot 1/s$. Entonces puede haber un gran número de errores debido a tres razones: La fracción $1/s$ de elementos estimulantes que necesitaban recondicionamiento fue alto al principio; el parametro de aprendizaje t es bajo; o w es pequeño, es decir, una respuesta A_1 es reforzada frecuentemente.

En ocasiones es razonable suponer que los elementos estimulantes

fueron originalmente conectados aleatoriamente, lo cual proporciona un vector de probabilidades iniciales π cuya expresión es:

$$\pi = \left\{ \frac{1}{2^a} \begin{bmatrix} s \\ i \end{bmatrix} \right\}.$$

Por lo que la media del número total de respuestas incorrectas es

$$\pi \frac{1}{wt} \gamma = \left\{ \frac{1}{wt} \right\} \pi \gamma = \frac{1}{2wt}.$$

En el caso de que $v = w = 0$, donde cada acción del individuo es reforzada, existen dos estados absorbentes ($i = 0, i = s$) y la pregunta más interesante es si el individuo termina siendo condicionado a una respuesta A_0 o A_1 . En este caso $P_{ij} = \gamma$, pero γ tiene una componente 0 para s_0 , y 1 para s_1 , por tanto γ es la probabilidad de absorción en s_0 , y la probabilidad de que sea completamente condicionado a respuestas A_1 es igual a la fracción de elementos estimulantes conectados originalmente a A_1 .

4.4 EL MODELO ABIERTO DE LEONTIEF.

En un modelo de insumo-producto de Leontief considerese una economía en donde hay r industrias, y supóngase que cada industria produce exactamente un solo tipo de bienes. Se acepta que los factores naturales de producción como son la tierra, minerales, etc., son libres y no influyen en el costo final de los bienes. En general las industrias están interconectadas en el sentido de que cada una debe comprar una cierta cantidad (positiva o cero) de los productos de las demás para que funcione su industria. Se definirán los coeficientes tecnológicos como sigue: q_{ij} es la cantidad del producto de la industria j que debe ser adquirido por la industria i para que pueda producir un peso de sus propios bienes. Sea Q una matriz de $r \times r$ con entradas q_{ij} . Es fácil ver que la suma de q_{ij} para i fijo, da el valor total de los insumos que necesita la industria i para producir \$1 de sus propios bienes. Si la industria i es productiva o por lo menos sale sin ganar ni perder, dicha suma debe ser menor o igual que el valor de sus productos, es decir, $q_{i1} + q_{i2} + \dots + q_{ir} \leq 1$. Se dirá que la empresa es rentable si ocurre la desigualdad estricta, y si ocurre la igualdad se dirá que es infructuosa. Se supondrá que todas las empresas son productivas o infructuosas, descartando la posibilidad de que alguna sea improductiva.

Sea x_i el valor monetario de la producción de la industria i , y $\pi = (x_1, x_2, \dots, x_r)$ el vector renglón de producciones. Como la industria i necesita una cantidad $x_i q_{ij}$ de la producción de la industria j , el vector de insumos requeridos por las industrias es πQ , y su j -ésima componente es el valor total de la producción que debe tener la industria j para satisfacer la demanda interindustrial de su producto.

Suponga que la economía requiere para consumo una cantidad c_i del producto de la industria i . Sea $\gamma = (c_1, c_2, \dots, c_r)$ el vector de consumo. se necesita que $\gamma \geq 0$. La necesidad de que el vector de producción de la economía sea ajustado para que tanto las necesidades interindustriales como las necesidades de consumo sean satisfechas es fácil de escribir en forma vectorial como

$$\pi = \pi Q + \gamma,$$

es decir,

$$\pi(I - Q) = \gamma$$

que es un sistema de r ecuaciones con r incógnitas que tiene solución solo si $I-Q$ tiene inversa. Mas aún, las soluciones serán no-negativas para cada γ si y solo si $(I-Q)^{-1}$ tiene todas sus componentes no-negativas.

El problema se puede solucionar utilizando para ello la cadena de Markov asociada a un modelo de insumo-producto, la cual tiene las siguientes propiedades:

1) Los estados son los r procesos del modelo, mas un estado absorbente adicional s_0 , llamado estado banca.

2) La matriz de transición P está definida como:

$$\begin{aligned} p_{00} &= 1 \\ p_{0j} &= 0 & j > 0 \\ p_{i,j} &= q_{i,j} & i, j > 0 \\ p_{i0} &= 1 - \sum_{j=1}^r q_{i,j} & i > 0. \end{aligned}$$

Intuitivamente, significa que si la industria i recibe un peso para para su uso, lo debe gastar en comprar $p_{i,j}$ a la industria j . El resto del peso, si queda, es decir, la cantidad p_{i0} es la ganancia, y se debe pensar que está siendo depositada en un banco. El hecho es que el estado banca es un estado absorbente, lo cual significa que el banco obtiene dinero pero no lo gasta. Si Q

satisface las condiciones 1) y 2), y la cadena de Markov tiene como unico estado absorbente a s_0 , entonces $(I-Q)^{-1} = N$ existe y es no-negativa, por lo tanto $\pi = \gamma N$ es la solución deseada. En otro caso $(I-Q)^{-1}$ que da el número promedio de tiempo en varios estados antes de que alcance s_0 , podria no existir.

Una sencilla interpretación económica de este resultado indica que dado que debe ser posible alcanzar el estado banca a partir de cualquier estado, y que solamente una industria productiva puede alcanzar dicho estado directamente, una industria infructuosa debe alcanzar el estado banca a través de una productiva. Por lo tanto la condición establece que cada una de las industrias debe ser productiva o depender de una que lo sea. Si esta condición no es satisfecha, entonces la economía no puede abastecer todas las demandas posibles.

Considérese el caso general en que la condición es violada. La cadena de Markov asociada no es una cadena absorbente con un solo estado absorbente. Esto implica que existe un conjunto cerrado de estados, distinto de $\langle s_0 \rangle$, de industrias que no son productivas y que no dependen de ninguna industria fuera de dicho grupo. Sea \bar{Q} la submatriz de dichas industrias, entonces $Q\bar{x} = \xi$ (\bar{x} es un vector columna con todas sus entradas 1), por lo tanto no puede satisfacer demandas externas, y entonces la economía no puede cumplir los requerimientos de bienes producidos por esas industrias, y aquellos cuya producción requiera materias primas de estas industrias tampoco podrán ser suministrados, pues estas actuarían como demandas externas al grupo cerrado de industrias. Así, si se retiran de la economía este grupo cerrado de industrias improductivas y todas las que dependen de ellas, las restantes (si las hay) podrán cumplir los requerimientos, y satisfacer demandas arbitrarias. Para detectar que industrias pueden soportar una demanda, se utiliza el siguiente algoritmo de clasificación de estados:

1) Revise que todos los renglones de la matriz Q correspondan a industrias que sean productivas, es decir que cada renglón sume uno, y márquelos.

2) Revise las columnas que tienen los mismos índices que los renglones marcados y cheque, en esas columnas, los renglones que tiene entradas positivas.

3) Repita (2) hasta que no produzca nuevos renglones. Entonces pueden ocurrir cualquiera de las siguientes posibilidades:

a) Todos los renglones están marcados, en cuyo caso la cadena de Markov asociada tiene a s_0 como único estado absorbente, y entonces cualquier demanda no-negativa puede ser recibida.

b) No todos los renglones están marcados. En este caso dichos renglones corresponden al grupo cerrado maximal de empresas improductivas. Todos los estados que dependen de estas industrias se pueden encontrar quitando las marcas puestas previamente y aplicando (2) repetidamente.

Las industrias que no puedan recibir demanda exterior forman un segmento completamente inútil de la economía, por ello, se supondrá que han sido borradas, y por lo tanto $(I-Q)^{-1}$ existe.

Ejemplo: Suponga que los coeficientes tecnológicos para seis industrias están dados por

$$Q = \begin{pmatrix} 1/2 & 0 & 1/4 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/4 & 3/4 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1/4 & 0 & 1/4 & 0 & 1/4 \end{pmatrix}$$

y por tanto

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/2 & 0 & 1/4 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/4 & 3/4 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 \end{pmatrix} \begin{matrix} s_0 \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \end{matrix}$$

De la matriz P se puede ver directamente que existen únicamente tres industrias productivas que son s_1 , s_2 , y s_6 ; s_4 y s_5 son improductivas y no dependen de ninguna industria productiva, mientras que s_3 depende de ellas, por lo que estas tres industrias deben ser borradas. Así, la matriz de transición es

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/4 & 1/2 & 0 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix} \begin{matrix} s_0 \\ s_1 \\ s_2 \\ s_3 \end{matrix}$$

y

$$N = \begin{pmatrix} 4 & 0 & 2 \\ 8/3 & 4/3 & 2 \\ 4 & 0 & 4 \end{pmatrix}.$$

Entonces, si se supone $\gamma = (1, 3, 2)$, entonces se deberán producir $\gamma N = (20, 4, 16)$ unidades por cada una de las industrias, teniéndose una producción total de 40 pesos.

La teoría de colas es un área de la Investigación de Operaciones que se ha desarrollado para analizar y planear el comportamiento de las líneas de espera en cualquier industria, buscando siempre optimizar la utilización de los recursos y los resultados de la misma. Como es de esperarse, aun cuando se cuenta con distintos modelos de colas, todos tienen una estructura común, dentro de la cual están las variables de decisión que considera siendo las principales las siguientes: el número (s) de servidores en una oficina de servicio; la eficiencia de cada servidor (μ); el número de oficinas de servicio y su localización; el valor esperado del costo de proveer el servicio en un sistema de cola ($EC(S)$), que es una tasa de costo con dimensiones pesos/tiempo; el valor esperado del costo asociado con los clientes que están en el sistema de cola ($EC(WC)$); y $E(TC)$ el valor esperado del costo total, en donde este es la suma de los valores esperados de costo de servicio y costo de espera, es decir, $E(TC) = E(S) + E(WC)$.

Si se supone que el costo de espera en cualquier momento particular depende de cuantos clientes están en el sistema, entonces

$$E(WC) = \sum_{n=1}^{\infty} g(n)P_n$$

donde $g(n)$ es la tasa de costo cuando hay n clientes en el sistema, y P_n son las probabilidades de recurrencia. Es decir, que el costo promedio de espera es proporcional al número promedio de clientes en el sistema.

Sea C_v la tasa de costo por individuo en el sistema, entonces se tiene que

$$g(n) = C_v n,$$

y

$$ECWC) = C_v L.$$

Si se supone ahora que el costo de espera para un individuo es una función del tiempo que permanece en el sistema, y se definen
 T = la variable aleatoria que denota el tiempo en el sistema,
 $h(T)$ = función de costo para un cliente en el sistema,
 $f_T(w)$ = función de densidad de probabilidad del tiempo.

entonces,

$$E(h(T)) = \int_0^{\infty} h(w) f_T(w) dw,$$

pero como este valor esperado es un costo más que una tasa de costo, la tasa $ECWC)$ se obtiene como

$$ECWC) = \lambda E(h(T))$$

donde λ es la tasa de llegadas de clientes al sistema cuando se supone que los tiempos entre llegadas son variables aleatorias positivas, independientes e idénticamente distribuidas.

Es importante hacer notar que la longitud de la cola al momento t forma una cadena de Markov irreducible y aperiódica, y por tanto su distribución límite es independiente de la distribución inicial y si además es estacionaria, con $\lambda < \mu$ se tiene que el tamaño promedio de la cola es $\lambda / (\mu - \lambda)$ donde μ es el parámetro de la distribución del tiempo de servicio.

El tiempo virtual de espera en el instante t , es el tiempo que un cliente deberá esperar si él se incorpora a la cola en el instante t . Si el proceso es estacionario, es decir, si $\lambda < \mu$ existe su distribución límite y es

$$\lim_{t \rightarrow \infty} W(t, x) = W^*(x) = \begin{cases} 1 - \frac{\lambda}{\mu} e^{-(\lambda - \mu)x} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$$

y si $\lambda \geq \mu$ entonces $W^*(x) = 0$ para toda x .

Por lo tanto si el proceso es estacionario, el tiempo promedio de espera de un cliente es $\lambda / \mu(\lambda + \mu)$.

Cuando el costo de espera es una función lineal del tiempo de permanencia en el sistema, el costo promedio de espera es, otra vez, proporcional al número promedio en el sistema. Esto es,

$$h(CT) = C_v T$$

y

$$E(WC) = \lambda E(h(CT)) = \lambda C_v W = C_v L$$

donde C_v es la tasa del costo de espera para un individuo en el sistema.

Para determinar el número de servidores que debe haber en el sistema si se supone que cada servidor tiene un costo de C_s y que los tiempos de servicio son variables aleatorias independientes e idénticamente distribuidas e independientes de los tiempos de llegadas, el modelo indica que, si s es entera

$$E(TC) = sC_s + E(WC)$$

Suponiendo un sistema con tasas de llegada y servicio por canal constantes. Además si el número de servidores crece, el número promedio y el tiempo en el sistema decrecen.

Para un sistema con población y cola infinitas, el valor mínimo de s es igual al menor entero mayor que λ/μ , mientras que para uno finito el mínimo de s es 1.

Para determinar el número de servidores y la eficiencia del servicio, supóngase que hay varios tipos alternativos de servidores con diferentes eficiencias medidas por la tasa de servicio μ . El problema es escoger el tipo de servidores y el número s de ellos. El costo para cada servidor es una función de μ , $f(\mu)$. El modelo de costo es entonces

$$E(TC) = s f(\mu) + EC(WC)$$

y la solución se encuentra determinando el valor óptimo de s para cada alternativa y seleccionando la alternativa con menor costo total.

Si lo que se desea determinar ahora es el número de oficinas de servicio que debe haber en el sistema, se utiliza una tasa de llegadas al sistema que se considera dividiendo el número de clientes que llegan al sistema entre el número de oficinas de servicio. Si λ_p es la tasa de llegadas para la población y n es el número de oficinas, la tasa de llegadas en cualquiera de ellas es

$$\lambda = \lambda_p / n .$$

Supóngase que hay un costo fijo C_f de establecer una oficina y un costo variable C_v por servidor, entonces el costo total del sistema de oficinas es

$$E(TC) = n (C_f + s C_v + EC(WC))$$

donde $EC(WC)$ es el costo promedio de espera en cada oficina. La solución a este modelo siempre resultará en una sola oficina, por ello muchos sistemas de cola reales tienen una sola cola.

Ejemplo: Se han diseñado dos alternativas para una oficina de servicio de un sistema de cola finita. La alternativa A tiene un solo servidor y la B dos. El costo de la oficina de servicio tiene dos partes: un costo fijo de \$ 50,000 por año y que es

independiente del número de servidores; y uno variable de \$ 30.000 por servidor por año. Una simulación mostró que, con la alternativa A el 10% de los clientes abandona el sistema de cola por encontrarlo lleno, y con la B solamente lo hacen un 5% de los mismos. La tasa de llegadas al sistema (incluyendo a los que se van) es 1500 clientes por año.

La administración ha determinado que, el que un cliente abandone el sistema, le cuesta a la compañía \$ 100 en pérdida de utilidades, y que el costo de que una persona espere en el sistema es \$ 20 por hora, por cliente en el sistema.

El problema es determinar cual de los dos sistemas alternativos minimiza el costo total. Supóngase para ello que un año tiene 2000 horas.

Para resolver el problema se realiza una simulación del comportamiento de cada uno de los sistemas A y B, y los resultados se presentan en las figuras 4.1 y 4.2 respectivamente. Se calculan los costos para cada alternativa y se comparan en busca del menor de ellos.

Así el costo para el caso de un solo servidor (alternativa A) se obtiene como sigue:

A partir de la simulación para la alternativa A se tiene

$$L = 1.878915$$

$$C_v = 20$$

y como $ECWC = C_v L$, entonces $ECWC = 37.5383$ y la tasa promedio de llegadas es .7021064.

Si el 10% de los clientes abandonan el sistema y cada uno tiene un costo de \$ 100, entonces el costo de los que abandonan el sistema es $(.07021064)(100) = 7.021064$

GENERAL MODEL

STATE PROBABILITIES

STATE	ARRIVAL RATE	DEPT. RATE	PROBABILITY
0	1	0	.1176527
1	.75	.48	.2451097
2	.75	.48	.382984
3	.75	1.9	.1511779
4	0	1.1	.1030758

AVERAGE QUANTITIES

AVERAGE ARRIVAL RATE = .7021064 PER HOUR
 AVERAGE NUMBER IN SYSTEM = 1.876915
 AVERAGE NUMBER IN QUEUE = .9945672
 AVERAGE TIME IN SYSTEM = 2.673263 HOURS
 AVERAGE TIME IN QUEUE = 1.416548 HOURS

QUEUEING SYSTEM RESULTS

PROPORTION OF THE TIME THE SYSTEM IS EMPTY = .1176527
 PROPORTION OF THE TIME THE SYSTEM IS FULL = .1030758
 PROPORTION OF THE TIME ALL CHANNELS ARE BUSY = .8823474
 PERCENT UTILIZATION = 88.23473
 EXPECTED NUMBER OF BUSY CHANNELS = .8823474
 EXPECTED NUMBER OF IDLE CHANNELS = .1176527

Figura 4.1

GENERAL MODEL

STATE PROBABILITIES

STATE	ARRIVAL RATE	DEPT. RATE	PROBABILITY
0	1	0	.379628
1	.75	1.25	.3087303
2	.75	1.1	.2070888
3	.75	2.6	5.973716E-02
4	0	.9	4.978097E-02

AVERAGE QUANTITIES

AVERAGE ARRIVAL RATE	= .8075799	PER HOUR
AVERAGE NUMBER IN SYSTEM	= 1.096243	
AVERAGE NUMBER IN QUEUE	= .1592991	
AVERAGE TIME IN SYSTEM	= 1.357442	HOURS
AVERAGE TIME IN QUEUE	= .1972549	HOURS

QUEUEING SYSTEM RESULTS

PROPORTION OF THE TIME THE SYSTEM IS EMPTY	= .379628
PROPORTION OF THE TIME THE SYSTEM IS FULL	= 4.97809E-02
PROPORTION OF THE TIME ALL CHANNELS ARE BUSY	= .3158069
PERCENT UTILIZATION	= 46.84721
EXPECTED NUMBER OF BUSY CHANNELS	= .9369441
EXPECTED NUMBER OF IDLE CHANNELS	= 1.063056

Figura 4.2

El hecho de que el costo fijo sea \$ 50,000 por año es equivalente a un costo de \$ 25 por hora, y el costo variable de \$ 30,000 por año equivale a \$15 por hora. Por lo tanto

$$E(TC) = 25 + 7.021064 + 15 + 37.5383 = \underline{94.559364}$$

Cuando se consideran dos servidores el resultado es el siguiente:

$$L = 1.096243$$

$$C_v = 20$$

$$\text{entonces } E(WC) = 21.92486$$

y la tasa promedio de llegadas es .8075799

Si el 5% de los clientes abandonan el sistema con un costo de \$100 por cada uno, el costo de los que abandonan el sistema es

$$(0.040378995)(100) = 4.0378995$$

Por lo tanto

$$E(TC) = 25 + 4.0378995 + 15(2) + 21.92486 = \underline{90.9627525}$$

Comparando los costos de cada alternativa el resultado que se obtiene para el problema es que el sistema que minimiza el costo total corresponde a la alternativa B con dos servidores.

Existen también dentro de la teoría de colas otro tipo de sistemas, por ejemplo los sistemas de colas con un solo canal y con distribuciones generales de tiempo de servicio, también conocidos como sistemas (M/G/1), cuyas fórmulas generales fueron obtenidas por Pollaczeh-Khintchine utilizando para ello la teoría de los procesos de ramificación, por resultar estos una aproximación bastante adecuada a los sistemas que se desean analizar. Dichas fórmulas cubren también el caso de los sistemas M/D/1/∞ que fueron estudiados por Fry.

Un modelo de colas M/G/1 se puede describir como un sistema de cola con un solo servidor y proceso de llegadas Poisson. Los tiempos de servicio se suponen independientes de cada uno de los otros y de las llegadas. Si se denota por $N_1(\omega)$ al número de llegadas durante el intervalo de tiempo $[0, t]$; $\Sigma_1(\omega)$, $\Sigma_2(\omega)$, ... a los tiempos de servicio de los clientes; y $Y_1(\omega)$ al número de clientes esperando servicio en el tiempo t ; y si se supone además que el proceso de llegadas es Poisson con parámetro λ , para conocer el comportamiento del proceso "longitud de cola" $Y = \{Y_t; t \geq 0\}$, se define X_n como el número de clientes en el sistema en el instante siguiente de la n -ésima salida. Los clientes que están inicialmente en la cola forman la 0-ésima generación, los que llegan durante el tiempo en que dicha generación es atendida, forman la primera generación, y así sucesivamente. Sea $N(n, i)$ el número de clientes que llegan mientras es atendido el i -ésimo individuo de la n -ésima generación. Estas son independientes e idénticamente distribuidas y por lo tanto describen un proceso de ramificación.

Como resultado de sus estudios Pollaczeh y Khintchine llegaron a que

- La probabilidad de que no haya ningún cliente en el sistema es

$$P_0 = 1 - \rho$$

- La longitud promedio de la cola se puede calcular como

$$\bar{q} = \rho^2 \frac{(\lambda + \mu^2 \sigma^2)}{2(\lambda - \rho)}$$

- El número promedio en el sistema es

$$\bar{n} = \bar{q} + \rho$$

- El tiempo promedio de espera que se obtiene es

$$\bar{v} = \frac{\bar{q}}{\rho\mu}$$

- Por último el tiempo promedio en el sistema es

$$\bar{d} = \bar{v} + \frac{1}{\mu}$$

Existen varios ejemplos de aplicación de este tipo de sistemas, y entre ellos se encuentran los siguientes:

Ejemplo 1: El tiempo para procesar una reclamación en una oficina de Seguros es exactamente 25 minutos por cada reclamación propuesta. Si los reclamantes llegan aleatoriamente con una tasa promedio de 1 cada 40 minutos, ¿ que tiempo en promedio debe esperar un reclamante para ser atendido, dado que hay un solo escritorio de seguros procesando las reclamaciones ?

$$\sigma_a^2 = 0, \quad \lambda = 1.5 / \text{hr.}, \quad \mu = 2.4 / \text{hr.} \quad \rho = 1.5 / 2.4 = 0.625$$

entonces

$$\text{Tiempo promedio de espera } \bar{v} = \frac{0.625}{4.900.375} = 0.3472 \text{ hrs.} = \underline{20.83 \text{ min.}}$$

Ejemplo 2: En una tienda de maquinaria pesada, se utiliza la grúa general en un 75%. El estudio de observaciones en el tiempo dió el tiempo promedio de levantamiento como 10.5 min. con una desviación estándar de 8.8 min. ¿ Cual es la tasa promedio de llamadas para los servicios de la grúa y cual es el retraso promedio en obtener el servicio?. Si el tiempo promedio de servicio es cortado a 8 min. con desviación estándar de 6, ¿ Cuanta reducción ocurrirá en promedio en el retraso de ser servido ?. La situación que se tiene inicialmente es

$\rho = 0.75$, $\mu = 60 / 10.5 = 5.71 / \text{hr.}$, y $\lambda = \rho \cdot \mu = 4.29 / \text{hr.}$, entonces el tiempo promedio de espera es

$$\bar{v} = \frac{0.75}{0.5} (1.70) \left[\frac{50}{5.71} \right] = 26.8 \text{ min}$$

Si el tiempo de servicio es cortado a 8 minutos, los valores de ρ y μ son

$$\mu = 60 / 8 = 7.5 / \text{hr.} \quad \text{y} \quad \rho = 4.29 / 7.5 = 0.571,$$

es decir que la utilización de la grúa se reduce al 57.1%, y por lo tanto

$$\bar{v} = \frac{0.571}{2(0.429)} (1.562) \text{ h} = 8.3 \text{ min.}$$

lo que significa una reducción de 18.5 minutos, o aproximadamente un 70%.

Ejemplo 3: Una firma emplea un equipo de ajustadores experimentados para reparar fallas que ocurren a sus máquinas. Para ello considérese que existen los siguientes métodos alternativos de operación:

a) Ninguna especialización, todos los proveedores de equipo son reparadores. Este método da una desviación estándar de tiempo de reparación igual a la tasa media de reparación (coeficiente de variación igual a uno).

b) Especialización parcial, la cual reduce la desviación estándar del tiempo de reparación a la mitad de la tasa media de reparación.

c) Especialización total, lo cual asegura que todas las reparaciones son completadas en el mismo tiempo.

Dado que la tasa media de reparación es la misma para todos los métodos de operación y que las reparaciones llegan aleatoriamente con respecto al tiempo, muestre que

i) El método b) da una reducción del 37.5% en la esperanza del tiempo de espera para reparación sobre el método a).

ii) El método c) da una reducción del 50% en el tiempo promedio de espera sobre el método a).

Nótese que todos los procesos propuestos son M/G/1 con las mismas tasas de servicio μ y de llegadas λ . El tiempo promedio de espera para cada uno está dado por \bar{v} .

Entonces:

$$a) \sigma_s^2 = 1 / \mu^2, \quad y \quad \bar{v} = \rho / [\mu(1 - \rho)]$$

$$b) \sigma_s^2 = 1 / 4\mu^2, \quad y \quad \bar{v} = 5\rho / [8\mu(1 - \rho)]$$

$$c) \sigma_s^2 = 0, \quad y \quad \bar{v} = \rho / [2\mu(1 - \rho)].$$

El costo de la espera en la cola por trabajos es proporcional a $\lambda \bar{v}$. Los porcentajes de ahorro alcanzados son:

i) El ahorro en el tiempo de espera es

$$\frac{w_a - w_b}{w_a} \cdot 100\% = 300 / 8 = 37.5 \%$$

ii) El porcentaje de ahorro en el tiempo de espera se calcula como

$$\frac{w_a - w_c}{w_a} \cdot 100\% = 50 \%$$

Dado que no es el objetivo de este trabajo el presentar de manera exhaustiva la Teoría de Colas, probablemente no se alcance a visualizar completamente la aplicación de los resultados de los Procesos Markovianos. Sin embargo, existen muchas propiedades dentro de los modelos de colas que tienen su base en las distribuciones del tiempo de espera y de la longitud de la cola, así como en sus estructuras características de Procesos de Markov¹.

¹Puede revisarse Takács, Lajos, "Introduction to the Theory of Queues", New York Oxford University Press, 1962.

4.6 CONTROL DE INVENTARIOS

Considérese un sistema de inventarios como un tipo de sistema de colas M/G/1 en el cual se tiene al tiempo t un inventario que se puede identificar como el tiempo virtual de espera. Se supone que el sistema tiene una capacidad finita k . También se supone que en cada momento el sistema está en una de los dos posibles fases 1 y 2 y también en cualquier momento puede ser cambiado de una a otra sin pérdida de tiempo. Si el sistema está en la fase i , las épocas en las cuales los clientes llegan son generadas por un proceso Poisson con tasa λ_i , $i = 1, 2$. Sea Y_i una variable aleatoria positiva con función de distribución de probabilidad $F_i(x) = P(Y_i \leq x)$, $i = 1, 2$. Cualquier cliente que llegue mientras el sistema está en la fase i y el inventario del sistema es x , agranda el inventario en una cantidad distribuida como $\min\{k-x, Y_i\}$ y causa un excedente el cual está distribuido como $\max\{0, Y_i - k + x\}$, $i = 1, 2$. Si el sistema está en el estado i y el inventario es positivo, entonces entre épocas de llegadas el inventario decrece linealmente a una tasa $\alpha_i > 0$, $i = 1, 2$.

Se impone al modelo una estructura de costos en la cual cuando se tiene un inventario x y en la fase i el costo de inventario es $h_i(x)$ donde $h_1(x)$ y $h_2(x)$ se suponen acotadas y con solo un número finito de discontinuidades en $0 \leq x \leq k$. Cuando se tiene un excedente " y " se incurre en un costo $p_i(y)$ si el sistema está en la fase i , donde $p_i(y)$ es una función no-decreciente de $y \geq 0$, con

$$\int_0^{\infty} p_i(y) dF_i(y) < \infty$$

para $i = 1, 2$.

Finalmente se tiene un costo fijo γ cuando el sistema es cambiado del estado 2 al 1.

La forma de controlar el inventario está especificada por dos niveles alternativos y_1 e y_2 con $0 \leq y_2 \leq y_1 < k$. Esta política (y_1, y_2) prescribe cambiar el sistema de la fase 1 a la 2 solo cuando el inventario excede el valor y_1 y cambiar el sistema de la fase 2 a la 1 solo cuando el inventario ha decrecido al valor y_2 .

Para $i = 1, 2$, sea $H_i(x) = 0$ para $x < 0$ y sea

$$H_i(x) = \frac{\lambda}{\sigma_i} \int_0^x (1 - F_i(y)) dy \quad \text{para } x \geq 0$$

y se define δ_i como el unico número que satisface

$$\int_0^{\infty} e^{-\delta_i y} dH_i(y) = 1 \quad i = 1, 2$$

También para $i = 1, 2$ definase la función de distribución G_i como

$$G_i(x) = \begin{cases} 0 & \text{si } x < 0 \\ \int_0^{\infty} e^{-\delta_i y} dH_i(y) & \text{si } x \geq 0 \end{cases}$$

Finalmente se define la función de renovación M_i $i = 1, 2$, como

$$M_i = \begin{cases} 0 & \text{si } x < 0 \\ \sum_{n=1}^{\infty} G_i^n(x) & \text{para } x \geq 0, \end{cases}$$

donde G_i^n es la n -ésima convolución de G_i consigo misma.

Definase el estado del sistema como x cuando el nivel de inventario es x y el sistema está en la fase 1, y como x' si el nivel de inventario es x y el sistema está en la fase 2. Sean también $X(t)$ y $S(t)$ el nivel de inventario y el estado del sistema al tiempo t respectivamente donde se toman los procesos $\{X(t), t \geq 0\}$ y $\{S(t), t \geq 0\}$ continuos por la derecha. Para derivar la fórmula

para el costo promedio, se estudiará la cadena de Markov del proceso $\langle S(t) \rangle$. Considerese que el sistema de inventario está controlado por una política fija (y_1, y_2) donde $y_2 > 0$. Se supondrá que en la época cero, el sistema está vacío, y sean: $T_0 = 0$ y T_n , $n \geq 1$ la n -ésima época en la cual el nivel de inventario excede y_1 mientras que el sistema está en la fase 1 o el nivel de inventario decrece a y_2 mientras que el sistema está en la fase 2, o bien, el inventario se hace cero. Para cualquier $n \geq 0$, se define S_n como el estado del sistema en la época T_n . El proceso a tiempo-discreto $\langle S_n, n = 0, 1, \dots \rangle$ es una cadena de Markov con espacio de estados $S = \{0\} \cup \{x' \mid y_1 < x \leq k\}$.

Como la cadena $\langle S_n \rangle$ tiene tiempo medio de recurrencia del estado s al estado 0, finito para todo $s \in S$, se tiene que $\langle S_n \rangle$ tiene una única medida de probabilidad π tal que para todo subconjunto de Borel A de S ,

$$\pi(A) = \int_S P(s, A) \pi(ds)$$

donde $P(\dots)$ denota la función de distribución de probabilidad de transición en un paso de $\langle S_n \rangle$.

Para determinar la distribución estacionaria π , se define para toda $0 \leq u \leq y_1$ y $y_1 \leq v \leq k$,

$p(u, v)$ = probabilidad de que el primer valor del proceso $\langle X(t), t \geq 0 \rangle$ tomado en el conjunto $\{0\} \cup \{x \mid y_1 < x \leq k\}$ pertenezca al conjunto $\{x \mid v \leq x \leq k\}$ dado que $X(0) = u$.

y sea $p_0(u) = 1 - p(u, y_1)$ para $0 \leq u \leq y_1$, por facilidad de notación, se escribirá

$$\pi_0 = \pi(\{0\}), \quad \pi_2 = \pi(\{y_2\}), \quad \pi(v) = \pi(\{x' \mid v \leq x \leq k, y_1 \leq v \leq k\}).$$

También, sea $F_1^-(y) = 1 - \lim_{x \rightarrow y} F_1^-(x) = P\{Y_1 \geq y\}$. Entonces por la

definición de $\pi(\cdot)$,

$$\pi(v) = \pi_0 \left\{ \bar{F}_1(v) + \int_0^v p(y, v) dF_1(y) \right\} + \pi_2 p(y_2, v), \quad y_1 \leq v \leq k \quad \dots (1)$$

$$\pi_0 = \pi_0 \int_0^{y_1} p_0(y) dF_1(y) + \pi_2 p_0(y_2) \quad \text{y} \quad \pi_2 = \pi(y_2) \quad \dots (2)$$

De (1) y (2) y la relación $\pi_0 + \pi_2 + \pi(y_1) = 1$ se puede determinar la distribución estacionaria π una vez que se han calculado las probabilidades $p(u, v)$.

Si se fija v con $y_1 \leq v \leq k$, utilizando la continuidad de $p(u, v)$ en $0 \leq u \leq y_1$, se tiene que $p(0, v) = 0$, y

$$p(u, v) = \phi(u, v) + \int_0^{y_1-u} e^{-\delta y} \phi(u+y, v) dM_1(y) \quad \text{para} \quad 0 \leq u \leq y_1$$

con lo cual se completa la determinación de la distribución estacionaria π donde ϕ es una medida finita en los conjuntos de Borel S tal que $\phi(A) > 0$ si y solo si $0 \in A$.

Como $\{S_n\}$ es uniformemente recurrente, se tiene que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} E(f(S_k)) = \int_S f(s) \pi(ds)$$

y si $Z(t)$ es el costo total en el que se incurre en $(0, t)$ y Z_n el costo total incurrido en (T_n, T_{n+1}) , sean $c(s) = E(Z_n | S_n = s)$ y $r(s) = E(T_{n+1} - T_n | S_n = s)$ para $s \in S$, y como $\{S(t)\}$ es regenerativo, se tiene que el costo promedio de la política $\{y_1, y_2\}$ está dado por

$$\lim_{t \rightarrow \infty} \frac{1}{t} E(Z(t)) = \int_S c(s) \pi(ds) / \int_S r(s) \pi(ds) \quad \dots (3)$$

Para obtener la distribución estacionario del inventario, se define, para cualquier $t \geq 0$, una variable aleatoria $A(t) = 1$

cuando el sistema está en la fase i al tiempo t , $i = 1, 2$, donde el proceso $\{A(t)\}$ es continuo por la derecha, obteniéndose que dicha distribución estacionaria está determinada por el lado derecho de la ecuación (3).

Se puede también calcular el número promedio de cambios de fase por unidad de tiempo como

$$(1 - \pi_0 - \pi_2) / \int_S r(s) \pi(ds)$$

*NOTA: El material presentado en esta sección corresponde solamente a un extracto de un artículo publicado en 1978, con la siguiente referencia:

H. C. TIJMS y F. A. van der DUYN SCHOUTEN, "Inventory control with two switch-over levels for a class of M/G/1 queueing systems with variable arrival and service rate", Stochastic Processes and Their Applications, Vol. 6, No. 2, January 1978.

CONCLUSIONES

A lo largo de este trabajo se ha presentado parte de la teoría desarrollada para las Cadenas de Markov en general, y los Procesos de Ramificación Univariados como una particularización de ellos. Asimismo y mediante la presentación de algunos ejemplos en los que estos modelos han sido utilizados, se muestran la importancia y gran oportunidad de aplicabilidad de ellos mismos.

De ninguna manera se pretende que este trabajo cubra toda la teoría desarrollada alrededor de dichos temas, pero sí lo suficiente para satisfacer los requerimientos de las aplicaciones presentadas.

Como se puede observar, definitivamente el campo de utilización de estos procesos es vasto y parece, aún, inagotable. Sin embargo, pensamos que la teoría desarrollada hasta ahora no ha sido completamente explotada, todavía, dentro de las aplicaciones, y que probablemente con ello podrían aumentar las posibilidades de uso de estos modelos en el futuro.

BIBLIOGRAFIA

- 1) ATHREYA, K. B., NEY, P. E., "Branching Processes", Ed. Springer-Verlag.
- 2) DOOB, J. L. " Stochastic Processes ", Ed. John Wiley & Sons, Inc., New York.
- 3) FELLER, W., " Introducción a la Teoría de Probabilidades y sus Aplicaciones ", Ed. Limusa.
- 4) GNEDENKO, B. V., " Theory of Probability ", Ed. Chelsea Publishing Co.
- 5) HARRIS, THEODORE E., " The Theory of Branching Processes ", Ed. Springer-Verlag, Berlin, 1963.
- 6) HOWARD, RONALD A., " Dynamic Probabilistic Systems ", Vol. I, Ed. John Wiley & Sons, Inc.
- 7) HUNTER, JEFFREY J., " Mathematical Techniques of Applied Probability ", Vol. II, Ed. Academic Press.
- 8) HILLIER, FREDERICK E., LIEBERMAN, GERALD J., " Introduction to the Operations Research " Ed. Holden-Day.
- 9) ISSACSON & MADSEN, "Markov Chains: Theory and Applications ", Ed. Wiley
- 10) KARLIN, SAMUEL, "A first Course in Stochastic Processes ", Ed. New York Academic.
- 11) KEMENY, JOHN G., SNELL J. LAURIE, " Finite Markov Chains ", Ed. D. Van Nostrand Company, Inc.

- 12) KOLMOGOROV, A. N., " Foundations of the Theory of Probability"
Ed. Chelsea, New York, 1966.
- 13) NEUTS, MARCEL F., " Probability ", Ed. Allyn & Bacon, Inc.
- 14) RIORDAN, JOHN, " Stochastic Service Systems ", Ed. John Wiley
& Sons, Inc., 1962.
- 15) SAATY, THOMAS L., " Elements of Queuing Theory With
Applications ", Ed. McGraw-Hill, 1961.
- 16) TAKACS, LAJOS, " Introduction to the Theory of Queues ", Ed.
New York Oxford University Press, 1962.
- 17) TIJMS, H. C., VAN DER DUYN SCHOUTEN, F. A., " Inventory
Control with two Switch-Over levels for a class of M/G/1
queuing systems with variable arrival and Service rate ",
Stochastic Processes and Their Applications, Vol. 8, No. 2,
January 1978.