



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

**ESTIMACIÓN DE PROCESOS DE SALTOS DE MARKOV VÍA
ALGORITMO ESPERANZA MAXIMIZACIÓN CON APLICACIÓN A
RIESGO DE CRÉDITO**

TESIS

QUE PARA OBTENER EL TÍTULO DE:

LICENCIATURA EN ACTUARÍA

PRESENTA:

CHAVARRÍA GARCÍA, MARIANA

ASESOR: BALTAZAR LARIOS, FERNANDO

Ciudad Universitaria, México, D. F.

2013



Universidad Nacional
Autónoma de México



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 esta 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.

Estimación de procesos de saltos de Markov vía
algoritmo Esperanza Maximización con
aplicación a riesgo de crédito.

Chavarría García Mariana

13 de marzo de 2013

Facultad de Ciencias, UNAM.

Hoja de Datos del Jurado

1. Datos del alumno

Apellido Paterno
 Apellido Materno
 Nombre(s)
 Teléfono
 Universidad
 Facultad de Ciencias
 Carrera
 Número de Cuenta

2. Datos del tutor

Grado
 Nombre(s)
 Apellido Paterno
 Apellido Materno

3. Datos del sinodal 1

Grado
 Nombre(s)
 Apellido Paterno
 Apellido Materno

4. Datos del sinodal 2

Grado
 Nombre(s)
 Apellido Paterno
 Apellido Materno

5. Datos del sinodal 3

Grado
 Nombre(s)
 Apellido Paterno
 Apellido Materno

6. Datos del sinodal 4

Grado
 Nombre(s)
 Apellido Paterno
 Apellido Materno

7. Datos del trabajo escrito

Título
 Número de páginas
 Año

1. Datos del alumno

Chavarría
 García
 Mariana
 58458479
 Universidad Nacional Autónoma de México
 Facultad de Ciencias
 Actuaría
 304114047

2. Datos del tutor

Dr.
 Fernando
 Baltazar
 Larios

3. Datos del sinodal 1

Dra.
 María Asunción Begoña
 Fernández
 Fernández

4. Datos del sinodal 2

Dra.
 Ana
 Meda
 Guardiola

5. Datos del sinodal 3

M. en C.
 David Josafat
 Santana
 Cobian

6. Datos del sinodal 4

M. en C.
 Jorge Humberto
 del Castillo
 Spíndola

7. Datos del trabajo escrito

Estimación de procesos de saltos de Markov vía
 algoritmo Esperanza Maximización con aplicación
 a riesgo de crédito.
 85
 2013

Índice general

1. Introducción.	7
2. Sobre el Riesgo de Crédito.	9
2.1. Riesgo de crédito.	9
2.2. Marco histórico de los Modelos de Riesgo de Crédito.	11
2.3. Calificadoras y calificación de Riesgo.	13
2.4. Marco normativo del Riesgo de Crédito.	15
3. Proceso de saltos de Markov.	19
3.1. Definiciones.	20
3.2. Matriz de intensidad.	26
3.3. Propiedades del proceso de saltos de Markov.	29
3.3.1. Procesos de saltos de Markov regulares.	29
3.3.2. Relación con Cadenas de Markov a tiempo discreto.	30
3.3.3. Ejemplos.	34
3.3.4. Simulación del proceso de saltos de Markov.	36
3.4. Inferencia estadística para procesos de saltos de Markov.	38
3.4.1. Introducción.	39
3.4.2. Función de verosimilitud.	41
3.4.3. Métodos para el cálculo del estimador.	47
4. Algoritmo Esperanza-Maximización.	51
4.1. Antecedentes.	52
4.2. Descripción del Algoritmo EM.	52
4.3. Aplicación del Algoritmo EM a Procesos de Saltos de Markov.	60

5. Aplicación a riesgo de crédito.	67
5.1. Descripción del modelo.	68
5.2. Algoritmo EM.	71
5.3. Implementación a riesgo de crédito.	72
6. Conclusiones.	75
I Apéndice	77

Capítulo 1

Introducción.

Todos los días, los seres humanos, nos enfrentamos a la incertidumbre, que muy frecuentemente no podemos prevenir, cuantificar o evitar, lo cual nos indica que siempre estamos expuestos a eventos adversos que tengan consecuencias negativas. Dicha incertidumbre ha llevado a numerosos estudios para intentar predecir lo que ocurrirá del modo más preciso con la finalidad de minimizar el riesgo, y la manera más efectiva de hacer predicciones es mediante probabilidades.

En el contexto que nos interesa estudiarlo, el riesgo se entiende como la posible pérdida financiera por los distintos factores aleatorios presentes en el mercado. Dentro de los tipos de riesgos financieros que existen nos enfocaremos en el riesgo de crédito, el cual está asociado a la probabilidad de que la contraparte de un contrato de crédito, no cumpla con sus obligaciones y cause a sus acreedores una pérdida financiera. Dichas probabilidades de incumplimiento son calculadas mediante diferentes modelos de riesgo de crédito.

Una de las metodologías utilizadas para hacer el cálculo de la probabilidad de incumplimiento, es mediante los sistemas de calificación crediticias. Las agencias calificadoras determinan la calidad crediticia de los acreditados y con base a ello les asignan calificaciones, que son de las principales métricas dentro del riesgo de crédito, pues representan la solvencia del acreditado, y por lo tanto se utilizan para analizar, admitir y dar seguimiento continuo y oportuno de los acreditados.

Las agencias calificadoras, asignan las calificaciones periódicamente, y de este modo se va formando un historial del acreditado a lo largo del tiempo, esto quiere decir que la información respecto de los acreditados es reportada de manera

discreta en el tiempo. Nos interesa entonces, conocer la posibilidad de que un acreditado que tiene asignada cierta calificación, se mueva a otra diferente en un horizonte de tiempo determinado y la manera de realizar la estimación es mediante matrices de transición que representen éstas probabilidades.

El propósito de este trabajo es presentar un modelo para el cálculo de dichas matrices, estudiado por Bladt y Sorensen en [4], que consiste en suponer que conocemos información histórica de los acreditados, es decir observaciones discretas en el tiempo, de tal modo que la información faltante entre dichas observaciones completa un modelo de la evolución de las diferentes calificaciones crediticias mediante un proceso de Markov a tiempo continuo.

En su estudio, Bladt y Sorensen, se basan en este modelo Markoviano a tiempo continuo para realizar la estimación de matrices de intensidad, que posteriormente nos permiten conocer la matriz de transición de calificaciones. La estimación se realiza mediante dos métodos: el Algoritmo Esperanza-Maximización (algoritmo EM) y el método Markov Chain Monte Carlo (MCMC). Debido a la naturaleza del problema, que es de información faltante, nos enfocaremos en el algoritmo EM que fue propuesto por Dempster en [8].

Derivado de este estudio, este trabajo se divide en cuatro partes; en el segundo capítulo, se presenta un panorama general del riesgo de crédito, por lo que se mencionan definiciones importantes, se describe el marco histórico y normativo del riesgo de crédito y también se definen los conceptos de calificadora y calificación de riesgo.

El tercer capítulo está dedicado a la teoría de procesos de saltos de Markov, que es el proceso estocástico que modela la transición de calificaciones, por lo que se presentan definiciones importantes y ciertas propiedades de procesos de saltos de Markov, dentro de las cuales se describe la manera de simular un proceso de saltos de Markov, además se introducen conceptos de verosimilitud.

El cuarto capítulo trata de la herramienta que utilizaremos para realizar el cálculo del estimador que nos interesa, es decir el algoritmo EM. En este capítulo se describen los antecedentes del algoritmo, se presentan ejemplos, y se explica como es aplicado particularmente en los procesos de saltos de Markov, donde también se presenta un ejemplo numérico y su simulación.

En el quinto capítulo se presenta el modelo de estimación máximo verosímil mediante el algoritmo EM, el cual se aplica a una cartera de acreditados cuyo comportamiento es modelado por un proceso de saltos de Markov y de la cual se cuenta con información real (a tiempo discreto). Por último se presentan conclusiones y resultados numéricos.

Capítulo 2

Sobre el Riesgo de Crédito.

Para comenzar, se analizarán algunas definiciones, conceptos y también el marco histórico y legal del Riesgo de Crédito.

2.1. Riesgo de crédito.

En general, el riesgo, es la posibilidad de ocurrencia de un evento adverso y sus consecuencias. El riesgo es el producto de la incertidumbre que existe sobre el valor de los activos financieros, ante movimientos adversos de los factores que determinan su precio; a mayor incertidumbre mayor riesgo.

Particularmente el riesgo financiero, está relacionado con la posibilidad de que ocurra un evento que se traduzca en pérdidas para inversionistas, deudores o entidades financieras.

La rentabilidad debe ser proporcional al riesgo: a mayor rentabilidad, mayor riesgo y viceversa. Si no se quiere correr ningún riesgo, se debe invertir en instrumentos financieros libres de riesgo y se obtendrá la rentabilidad libre de riesgo.

Si se invierte en un activo con riesgo, se espera obtener la rentabilidad libre de riesgo más una prima de rentabilidad o prima de riesgo. Existen varios tipos de riesgo financiero: de mercado, de crédito, de liquidez, operativo, país y sistémico.

El tipo de riesgo financiero en el que nos enfocaremos es el riesgo de crédito; el cual existe cuando se da la posibilidad de que una de las partes de un contrato financiero¹ sea incapaz de cumplir con las obligaciones financieras contraídas,

¹Contrato en el cual existe una transferencia de dinero que debe ser devuelto después de un tiempo determinado.

haciendo que la otra parte del contrato incurra en una pérdida.

El riesgo de crédito se puede analizar mediante tres dimensiones:

- **Riesgo de incumplimiento:** es la posibilidad de que se presente un incumplimiento en el pago de un crédito, depende de aspectos internos al acreditado, como su situación crediticia, y de aspectos externos, como la situación económica del país o el comportamiento de los mercados financieros nacionales e internacionales.
- **Riesgo de exposición:** es el generado por la incertidumbre respecto a los montos futuros en riesgo. Cuando los créditos pueden pagarse total o parcialmente de manera anticipada, se presenta el riesgo de exposición, ya que al no cumplirse con exactitud el plazo de liquidación se dificulta la estimación de los montos en riesgo.
- **Riesgo de recuperación:** está presente en el evento de un incumplimiento, ya que la recuperación es impredecible por depender del tipo de incumplimiento y de factores relacionados con las garantías que se hayan recibido, el tipo de garantía y su situación al momento de incumplimiento.

Históricamente, el riesgo de crédito es el más antiguo y el que mayor importancia tiene en términos de las pérdidas potenciales que su inadecuado manejo puede implicar para una institución de crédito. Sus orígenes datan, por lo menos desde 1800 a.C., en esencia ha permanecido sin variantes desde aquella época.

Una de las principales métricas dentro del riesgo de crédito, además de las pérdidas esperada, que se pueden estimar mediante modelos matemáticos, y las no esperadas, que no es posible estimar debido a que están determinadas por variables exógenas; corresponde a la calificación de la institución deudora, también conocida como calificación del riesgo deudor, que es un mecanismo que determina la calidad crediticia del deudor y que se utiliza para analizar, admitir y dar seguimiento continuo y oportuno al riesgo de crédito.

La problemática que implicaba el conocer el riesgo de incumplimiento de las instituciones, motivó el surgimiento de los **Modelos de Riesgo de Crédito**, que calculan la probabilidad de incumplimiento en el pago de los créditos otorgados y cuantifican el riesgo al enfrentar este incumplimiento.

Para diseñar un modelo se necesita partir del establecimiento de las relaciones entre las diversas variables que afectan al riesgo y con ellas construir el

2.2. MARCO HISTÓRICO DE LOS MODELOS DE RIESGO DE CRÉDITO.11

modelo. Entre las metodologías existentes aplicadas para estimar la probabilidad de incumplimiento de las empresas, se encuentran: los modelos de clasificación, árboles de decisión, modelos de elección cualitativa, y el uso de estadística y procesos estocásticos (matrices de transición). En la siguiente sección se describe la evolución de los modelos de Riesgo de crédito a través del tiempo.

2.2. Marco histórico de los Modelos de Riesgo de Crédito.

El desarrollo de los Modelos de Riesgo de Crédito surge de manera formal a final de los años 60 y durante la década de los 70. Sin embargo, desde los años 30 se inician estudios basados en el análisis tradicional de razones financieras.

En 1849, una compañía llamada *Dun & Bradstreet* (Cincinnati, Ohio), proporcionaba servicios a las empresas relacionadas con el análisis de crédito con base en el análisis de la información financiera. Asimismo, en 1967 aparece un trabajo de W. Beaver (1967), que puede ser considerado como uno de los trabajos clásicos en el área de análisis de razones y clasificación de quiebras.

El estudio de Beaver, titulado “*Financial ratios predictors of failure*”; realiza un análisis univariado de una serie de indicadores de bancarrota, sentando las bases para la realización de análisis multivariado para la estimación de la probabilidad de incumplimiento posteriores, dichas estimaciones fueron estudiadas después por E. Deakin (1972) en su trabajo “*A discriminant analysis of predictors of business failure*”, quien aplicó a los indicadores una serie de modelos discriminantes multivariados ².

Posteriormente surgieron las modificaciones a estos modelos con el Modelo Zeta (1977) de Altman, Haldeman y Narayanan, y el Modelo *EMS*³ (1995) de Altman, Heine, Hartzell y Peck.

Los estudios antes mencionados, proporcionaron una definición tentativa útil acerca de las principales razones financieras con poder predictivo acerca del evento de bancarrota, entre las cuales estaban: razones que miden rentabilidad, liquidez y solvencia de una institución, prevaleciendo como los más significativos.

Existen varias técnicas utilizadas para la construcción de Modelos de Riesgo de crédito, dentro de las que se encuentran: técnicas econométricas, redes

²Modelo conocido como *Z-Score*

³Emerging Markets Corporate Bond Scoring System

neuronales, modelos de optimización, sistemas de expertos y sistemas híbridos. Además los modelos pueden clasificarse en cuatro: Modelos Tradicionales, Modelos de Valor de Firma (estructurales), Modelos de Primer Tiempo de Paso y Modelos Contingentes (de intensidad de incumplimiento).

Los Modelos Tradicionales son de fácil cálculo y por lo mismo son los más usados, dentro de los principales se encuentran: Sistemas de calificación de crédito, el Modelo VaR ⁴, Modelo Basado en datos históricos y los Basados en datos de precios de bonos.

Es importante mencionar que el modelo VaR es el que cuenta con mayor avance en el desarrollo de metodologías, ya que su utilidad radica en que no sólo mide el riesgo de una única inversión sino también para un portafolio de inversiones, y además se puede determinar el valor máximo que podría ganar con la inversión.

Las dos metodologías mas importantes para el cálculo del VaR que fueron creados en J.P. Morgan & Co. son: *RiskMetrics*, (1989), también conocido como el método de exponencial suave, y posteriormente *Creditmetrics* (1997) creado por J.P. Morgan en su escrito “*Introduction to CreditMetrics, The benchmark for understanding credit risk*”, donde utiliza información histórica para determinar la máxima pérdida (o ganancia).

Desde los estudios publicados por Altman en los años sesenta hasta mediados de la década de los noventa sólo se habían implementado los llamados modelos tradicionales de riesgo de crédito, que predicen la quiebra de las empresas a partir de variables independientes (razones financieras, indicadores micro y macroeconómicos) más relevantes a nivel estadístico.

Los Modelos estructurales, toman en cuenta el valor de la firma, y el incumplimiento se presenta cuando el valor alcanza un límite en una barrera determinada que depende del tiempo. Entre los más importantes se encuentran: El Modelo de Merton (1974), El Modelo de Lenard y Toft (1996), Modelo KMV ⁵.

El objetivo de los Modelos de Primer Tiempo de Paso, es solucionar la carencia principal del modelo de Merton, respecto a la posibilidad de incumplimiento previo al vencimiento del instrumento; se asume que la quiebra se produce cuando el valor de la firma cruza cierto límite predeterminado, que depende del tiempo.

⁴Valor en Riesgo, mide la pérdida máxima de valor de un activo durante un periodo de tiempo, con un nivel de confianza específico.

⁵Por las siglas de sus autores Kecholfer, McQuown y Vasicek; también conocido como *Credit Monitor (CM)*

Dentro de estos modelos encontramos los siguientes: *Black-Cox* (1976), *Box y Cox* (1976), *Brenan y Schwartz* (1980), *Mason y Bhattacharya* (1981), *Frank y Torous* (1989), *Nielsen, Saa Requejo y Santa Clara*, *Kim Ramaswamy y Sundarson* (1993), *Longstaff-Schwartz* (1995), *Anderson y Sundaresan* (1996), *Mella-Barral y Perraudin* (1997), *Brys y Varene* (1997).

Los Modelos Contingentes, representan un enfoque alternativo de los estructurales. Aquí se define un proceso estocástico que gobierna la eventualidad del incumplimiento. Entre los principales tenemos: *Jarrow y Turnbull* (1995), *Jarrow, Landon y Turnbull* (1997), *Jarrow y Turnbull* (1998), *Duffie y Singleton* y el *Modelo de Bonos*. Se puede revisar información más completa sobre los modelos en [10].

Dentro de los modelos estructurales del riesgo de crédito, se encuentra el *rating* o calificación crediticia, que es del que hablaremos en la siguiente sección y en el que nos basaremos en este trabajo.

2.3. Calificadoras y calificación de Riesgo.

Las llamadas, “calificadoras de riesgo” o “agencias calificadoras”, son sociedades cuyo objeto social exclusivo, es la calificación de valores o riesgos relacionados con la actividad financiera, aseguradora, bursátil y cualquier otra relativa al manejo, aprovechamiento e inversión de recursos captados del público.

Mediante los sistemas de calificación, las agencias calificadoras determinan la calidad crediticia de los acreditados y permiten cuantificar la probabilidad de incumplimiento de los deudores con sus obligaciones y la severidad de las pérdidas en caso de incumplimiento.

Los sistemas de de calificación determinan una **calificación**, que es otorgada por una calificadora, y que representa una valoración profesional, acerca de la capacidad de un emisor de pagar el capital y los intereses en los términos y plazos pactados en cada emisión. Es decir, una opinión sobre la aptitud de pago de una empresa y la estimación razonable sobre la probabilidad de que el calificado cumpla con sus obligaciones contractuales, sobre el impacto de los riesgos que está asumiendo, o sobre la habilidad para administrar inversiones o portafolios de terceros, según sea el caso.

Debido a que la calificación se relaciona con la probabilidad y riesgo relativo de un instrumento de inversión, se basa en el análisis de información cuantitativa; como estados financieros, flujos de efectivo, razones financieras, proyecciones, entre otras; y también cualitativa, como la calidad de la administración, planes

y estrategias, oportunidades de mercado, investigación y desarrollo de nuevos productos, recursos humanos, políticas de control y auditoría, aspectos fiscales, entre otras. Asimismo, se incorpora dentro de la evaluación el ambiente económico de cada industria o sector y la situación social y política del país.

La calificación de riesgo tiene su origen en Estados Unidos de América hacia finales del siglo XIX, en donde se había desarrollado un sistema de información crediticia que era utilizado por inversionistas e instituciones financieras. A principios del siglo XX, con el desarrollo de la industria ferroviaria, las empresas de ese sector se convirtieron en las principales emisoras de bonos, hecho que dio pie a la creación de otras empresas dedicadas a estudiar la calidad de dichos instrumentos.

Algunas de las firmas más importantes de Estados Unidos de América comenzaron sus primeros pasos en el campo de la evaluación de riesgo, por ejemplo, *Poor's Publishing Co.*, publicó su primera calificación en 1916 y *Standard Statistics Bureau*, inicia actividades en 1922. Es hasta 1972 que se funda la primera calificadora fuera de Estados Unidos de América, en Canadá específicamente; en Europa la primera de estas firmas surgió en 1985 en España, con el nombre *Renta S.A.*

Mientras que en América Latina, la primera calificadora de riesgo se autorizó en Chile en 1988 y el segundo país de la región en donde se constituyó fue México, en 1990.

Actualmente, alrededor del mundo existen más de 74 calificadoras, pero el mercado se encuentra controlado básicamente por 3 principales compañías neoyorquinas que dominan aproximadamente el 90% del mercado, y aunque cada una tiene su propio sistema de calificaciones, éstos son muy similares. Dichas compañías son: *Standard & Poor's (S&P)*, *Moody's Investors Service (Moody's)* y *Fitch Ratings (Fitch)*, las cuales tienen la misma presencia en México.

Las calificadoras, asignan la calificación a las empresas según su capacidad de cumplimiento en un contrato de crédito, dependiendo del tiempo de vencimiento de los mismos; ya sea a corto plazo (que vencen en un periodo menor a un año) y a largo plazo (cuyo vencimiento se presenta después de un año o más).

Cada calificadora tiene su propia escala, pero todas se corresponden con la "escala homogénea" convenida entre ellas para permitir a los inversionistas establecer comparaciones (tabla de calificación homogénea).

En el cuadro 2.1 y 2.2 se presentan las calificaciones proporcionadas por las principales calificadoras para corto y largo plazo respectivamente.

Calificaciones para la deuda a corto plazo (menor a un año)			
Standard & Poor's	Moody's	Fitch	Descripción
A-1	1	F1	Fuerte grado de seguridad respecto al pago. Máxima capacidad para amortizar obligaciones. Alta calidad crediticia.
A-2	2	F2	Capacidad de pago satisfactoria. Capacidad de pago superior al promedio. Buena calidad crediticia.
A-3	3	F3	Adecuada capacidad de pago oportuno. Capacidad de pago promedio. Adecuada calidad crediticia.
B	4	B	Mayor incertidumbre. Capacidad de pago inferior al promedio. Especulativa capacidad incierta.
C		C	Dudosa capacidad oportuna de pago. Alto riesgo de incumplimiento.
D		D	Ha incurrido en incumplimiento de pago. Incumplimiento inminente.
		E	Calificación suspendida, (no se dispone de información adecuada).

Cuadro 2.1: Calificaciones otorgadas por principales calificadoras, para deudas a corto plazo.

Calificaciones para la deuda a largo plazo (mayor a un año)			
Standard & Poor's	Moody's	Fitch	Descripción
AAA	Aaa	AAA	Sustancialmente fuerte. Más fuerte y menor probabilidad de pérdida. La más alta calidad crediticia.
AA	Aa	AA	Muy fuerte. Muy fuerte y poca probabilidad de pérdida. Muy alta calidad crediticia.
A	A	A	Fuerte. Arriba del promedio. Alta calidad crediticia.
BBB	Baa	BBB	Adecuada. Promedio. Adecuada calidad crediticia.
BB	Ba	BB	Menor vulnerabilidad de incumplimiento. Debajo del promedio. Especulativa.
B	B	B	Mayor vulnerabilidad de incumplimiento. Capacidad crediticia débil. Altamente especulativa.
CCC	Caa	CCC/CC/C	Posibilidad de incumplimiento. Capacidad crediticia muy débil. Alto riesgo de incumplimiento.
CC	Ca	D	Alta susceptibilidad de no pagar. Capacidad crediticia extremadamente débil. Incumplimiento.
D	C	E	Incurrido en incumplimiento. Extremadamente especulativas. Calificación suspendida.

Cuadro 2.2: Calificaciones otorgadas por principales calificadoras, para deudas a largo plazo.

2.4. Marco normativo del Riesgo de Crédito.

Debido a la complejidad y dinamismo del mundo económico actual, donde la tecnología afecta a corporaciones y entidades financieras; es necesario incluir supervisión al momento de proporcionar estabilidad en los mercados, tanto micro como macroeconómica.

El Comité de Supervisión Bancaria (Comité de Basilea), que hasta 1990 se llamaba Comité de Regulación y Prácticas Supervisoras Bancarias, fue creado

a finales de 1974 por los países del Grupo de los Seis⁶ para coordinar la supervisión de los bancos internacionales. Surgió ante la aparición de unos mercados financieros más interrelacionados y, en principio, más volátiles como se explica en [20].

El primer Acuerdo de Basilea (Basilea I) surge en 1988, éste establecía una definición de "capital regulatorio" compuesto por elementos que se agrupan en dos categorías si cumplen ciertos requisitos de permanencia, de capacidad de absorción de pérdidas y de protección ante quiebra. Dicho capital debía ser igual o superior al 8% del valor nacional de los activos agrupados según su grado de riesgo.

Este capital debía ser suficiente para hacer frente a los riesgos de crédito, mercado y tipo de cambio. Cada uno de estos riesgos se medía con unos criterios aproximados y sencillos, el capital regulatorio se obtenía simplemente como una media ponderada del riesgo de crédito.

Sin embargo, éste primer acuerdo tenía como principal limitación la insensibilidad ante las variaciones de riesgo, además de ignorar la calidad crediticia, pues consideraba que los créditos tenían la misma probabilidad de incumplir.

Debido a lo anterior el Comité de Basilea en 2004 publicó una nueva propuesta (Basilea II) que se apoya en tres pilares:

- **Pilar I: Cálculo de los requisitos mínimos de capital.** Se establece que el riesgo de crédito se calcula a través de tres componentes fundamentales: PD (Probability of Default) Probabilidad de incumplimiento, LGD (Loss Given Default) Pérdida en el momento de incumplimiento también conocida como severidad, y (EAD Exposure At Default) exposición en el momento de incumplimiento.
- **Pilar II: Proceso de Supervisión de la gestión de los fondos propios.** Los organismos supervisores deben validar tanto los métodos estadísticos empleados para calcular los parámetros exigidos en el primer pilar, como la suficiencia de los niveles de fondos propios para hacer frente a una crisis económica.
- **Pilar III: Disciplina de mercado.** Se establecen normas de transparencia y se exige la publicación periódica de información acerca de su exposición a

⁶Países industrializados cuyo peso político, económico y militar es importante globalmente, en 1974 estaba formado por Estados Unidos, Japón, Alemania, Italia, Francia y el Reino Unido, actualmente se conoce como el grupo de los ocho o G8.

los diferentes riesgos y la suficiencia de sus fondos propios.

Para una descripción más detallada sobre la participación del riesgo de crédito bajo los fundamentos de Basilea II se puede revisar en [20].

Debido a la crisis financiera internacional que comenzó en 2007, en Europa, en 2010, el Comité de Basilea desarrolló una tercera propuesta, que tiene como base los fundamentos de Basilea II, y pretende mejorar el marco de regulación bancaria.

Con esta nueva propuesta se persigue:

- Mejorar la capacidad del sector bancario para afrontar perturbaciones ocasionadas por tensiones financieras o económicas de cualquier tipo.
- Mejorar la gestión de riesgos y el buen gobierno en los bancos
- Reforzar transparencia y divulgación de información de los bancos.

El marco de Basilea III se resume en una tabla que recoge las distintas medidas adoptadas por el Comité, se puede revisar la tabla en [22]. El Acuerdo de Basilea III pretende entrar en vigor a partir de 2013, ante esta situación algunos bancos han solicitado prórrogas, puesto que el desequilibrio financiero actual no permite la transición inmediata a la nueva reglamentación. En América Latina se pretende que México comience a implementar la reglamentación de Basilea III en 2013.

Actualmente en México, opera una variante del acuerdo de Basilea para la medición del riesgo de crédito. En síntesis, el acuerdo establece los porcentajes de riesgo de crédito para cada tipo de contraparte. Haciendo uso de estos porcentajes se determina una suma ponderada denominada Activos Sujetos a Riesgo de Crédito (ASRC).

Por otro lado, la circular 1423 expedida por la Comisión Nacional Bancaria y de Valores (CNBV) relativa a las disposiciones que deben observarse para una adecuada administración integral de riesgos, señala en su disposición vigésima, sección II, inciso d, que las instituciones de crédito deberán calcular la probabilidad de incumplimiento de los deudores en lo que respecta a su cartera crediticia.

Así mismo en nuestro país la información disponible para estimar los modelos de riesgo es más escasa, siendo esto un factor adicional a tomar en cuenta en el desarrollo de la medición del riesgo de crédito.

En México, para ayudar al inversionista a medir el riesgo de las emisiones de deuda y de las sociedades de inversión de deuda, a partir de enero de 1998, las instituciones Calificadoras de Valores evalúan la calidad de los instrumentos de deuda, así como los activos que integran los portafolios de las sociedades de inversión de deuda, la capacidad de pago del emisor y la sensibilidad ante cambios en la economía.

Capítulo 3

Proceso de saltos de Markov.

Un proceso estocástico es una familia de variables aleatorias, que sirve como modelo para representar la evolución aleatoria de un sistema a lo largo del tiempo. Dichos procesos pueden analizarse tanto en tiempo discreto, como en tiempo continuo.

Los procesos estocásticos tienen muchas aplicaciones en la vida cotidiana, por ejemplo son utilizados para modelar: redes de telecomunicación, índice de la bolsa por segundo, señales biomédicas como electrocardiogramas y encefalogramas, el tiempo de espera en cola de cada usuario que llega a una ventanilla, sistemas de producción e inventario o sistemas de entrada y salida.

Uno de los principales procesos estocásticos son las cadenas de Markov. Las cadenas de Markov a tiempo continuo, al igual que las cadenas a tiempo discreto se caracterizan por la propiedad Markoviana: dado el presente, el futuro es independiente del pasado; dichos procesos suelen ser candidatos naturales para modelar sistemas en tiempo real.

Un caso particular de las cadenas de Markov a tiempo continuo, es el proceso de saltos de Markov; con el que trabajaremos en este capítulo bajo las siguientes características: espacio de estados numerable, trayectorias continuas por la derecha y constantes a tramos de duración finita, y además, el número de saltos en un tiempo finito, también es finito.

En este capítulo se define el concepto de proceso de saltos de Markov, se estudian sus propiedades y se define la matriz de intensidad, se presentan algunos ejemplos; además se introducen conceptos sobre inferencia estadística, que serán una herramienta importante en el desarrollo del trabajo en capítulos posteriores.

3.1. Definiciones.

La mayoría del siguiente contenido está basado en [18, 14, 15].

Definición 1. Tomando como base un espacio de probabilidad (Ω, \mathcal{F}, P) ¹, un **proceso estocástico**, es una familia de variables aleatorias $\{X_t\}_{t \in \mathcal{T}}$ parametrizadas por un conjunto \mathcal{T} , llamado espacio parametral, en donde las variables toman valores en un conjunto E llamado espacio de estados.

Un proceso estocástico puede considerarse como una función de dos variables $X : \mathcal{T} \times \Omega \rightarrow E$, tal que a la pareja (t, ω) se le asocia el valor o estado $X_t(\omega)$. Para cada valor de $t \in \mathcal{T}$, el mapeo $\omega \rightarrow X_t(\omega)$ es una variable aleatoria, mientras que para cada $\omega \in \Omega$ fijo, la función $t \rightarrow X_t(\omega)$ es llamada una realización o trayectoria del proceso. Es decir que a cada ω del espacio muestral le corresponde una trayectoria del proceso.

Dependiendo de la característica del espacio parametral, es decir si \mathcal{T} es un conjunto discreto o continuo, distinguiremos a un proceso estocástico a tiempo discreto $X = \{X_n\}_{n \in \mathbb{N}}$ o a tiempo continuo $X = \{X_t\}_{t \geq 0}$ respectivamente; en general, en situaciones reales \mathcal{T} se interpreta como el tiempo. Se pueden ver más detalles de los procesos estocásticos en [18].

Nos interesan particularmente dos tipos de procesos estocásticos:

1. Procesos con incrementos independientes. Se dice que un proceso estocástico $\{X_t\}_{t \geq 0}$, tiene incrementos independientes, si para cualesquiera tiempos $0 \leq t_1 < t_2 < \dots < t_n$, las variables $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ son independientes. Es decir que los desplazamientos que tiene el proceso en intervalos mutuamente disjuntos de tiempo son independientes unos a otros.
2. Procesos con incrementos estacionarios. Diremos que un proceso estocástico $\{X_t\}_{t \geq 0}$, tiene incrementos estacionarios si, para cualesquiera tiempos $t < s$, y para cualquier $h > 0$, las variables $X_{t+h} - X_{s+h}$ y $X_t - X_s$ tienen la misma distribución de probabilidad. Es decir, el incremento que tiene el proceso entre los tiempos s y t sólo depende de estos tiempos a través del tiempo transcurrido, y no de los valores específicos de s y t .

¹Terna formada por un espacio muestral Ω , \mathcal{F} una σ -álgebra de subconjuntos de Ω y una medida de probabilidad P .

Definición 2. Una **cadena de Markov a tiempo continuo**, es un proceso estocástico $\{X_t\}_{t \geq 0}$, con espacio de estados E , a lo más numerable², que cumple con la siguiente propiedad: para todo $s, t \geq 0$, y para $i, j, k \in E$, $0 \leq u \leq s$:

$$P(X_{t+s} = j \mid X_s = i, X_u = k, 0 \leq u < s) = P(X_{t+s} = j \mid X_s = i).$$

Es decir, una cadena de Markov a tiempo continuo, es un proceso estocástico que cumple con la propiedad Markoviana, la distribución condicional del futuro X_{t+s} dado el presente X_s , y el pasado X_u , $0 \leq u \leq s$, sólo depende del presente, y es independiente del pasado.

Si además $P(X_{t+s} = j \mid X_s = i)$, es independiente de s , es decir si $P(X_{t+s} = j \mid X_s = i) = P(X_t = j \mid X_0 = i)$, se dice que la cadena de Markov a tiempo continuo es homogénea en el tiempo. En este trabajo consideraremos cadenas homogéneas.

A continuación se describen conceptos relevantes relacionados con las cadenas de Markov.

Sean i, j elementos del espacio de estados E y $s < t$, denotaremos por $p_{ij}(s, t)$ a la probabilidad

$$P(X_t = j \mid X_s = i),$$

que representa la **probabilidad de transición** del estado i en el tiempo s , al estado j en el tiempo t . La función definida por $p : \mathbb{R}_+^2 \times E \times E \rightarrow [0, 1]$, es llamada también función de transición.

La distribución de probabilidad μ_0 que satisface: $\mu_0(i) = P(X_0 = i), \forall i \in E$, es llamada distribución inicial. Si existe un estado $i \in E$, tal que $\mu_0(i) = 1$, entonces decimos que i es el estado inicial.

Definiremos el concepto de semigrupo, que nos servirá en resultados importantes que veremos más adelante. Una familia $\{G(t)\}_{t \geq 0}$ contenida en un conjunto de operadores de K en K , denotado por $\mathcal{B}(K, K)$, donde K es un espacio de Banach³; es llamado **semigrupo** si:

1. $G(0) = 1$

²Por numerable entendemos un conjunto con cardinalidad igual a los naturales o un subconjunto de él.

³Espacio vectorial normado y completo

2. $G(s+t) = G(t) + G(s) \forall s, t \geq 0$. Y además satisface que:
3. $\lim_{t \rightarrow 0} \|G(t) - I\| = 0$, decimos que es un semigrupo uniformemente continuo, donde $\|G\| = \sup_{f \in K, \|f\|=1} \|Gf\|$ y $\|\cdot\|$ es la norma usual.

Definición 3. Un proceso de saltos de Markov, es un proceso de Markov $X = \{X_t\}_{t \geq 0}$ a tiempo continuo, con espacio de estados un conjunto numerable E , el cual comienza en el tiempo cero $t_0 = 0$ en un estado inicial x_0 , en el cual permanece hasta el tiempo t_1 , tiempo en el que se traslada al estado x_1 y permanece ahí hasta un tiempo posterior $t_2 > t_1$ en donde salta a un estado diferente x_2 , y así sucesivamente. Entonces, si t_1, t_2, \dots es el conjunto de los tiempos de salto, $X_t = x_0$ para $t \in [0, t_1)$, $X_t = x_1$ para $t \in [t_1, t_2)$, y así sucesivamente. Además el $\lim_{n \rightarrow \infty} t_n = \infty$, esto garantiza que el proceso está bien definido para toda $t \geq 0$, evita que el proceso efectúe un número infinito de saltos durante un intervalo de tiempo acotado o que el proceso “explote”.

Considerando que el proceso es homogéneo en el tiempo y un espacio de estados $E = \{1, 2, \dots, m\}$, definimos la matriz de transición asociada al proceso de saltos de Markov $X = \{X_t\}_{t \geq 0}$ (matriz de probabilidades de transición al tiempo t), como P_t , matriz de $m \times m$ con entradas $p_{ij}(t)$ $i, j \in E$, tales que:

$$p_{ij}(t) = P[X_{t+s} = j \mid X_s = i], \quad (3.1.1)$$

que denotan la probabilidad de que el proceso que se encuentra actualmente en el estado i , esté en el estado j después de un cierto tiempo t .

Supondremos también, que las probabilidades de transición son estacionarias en el tiempo, es decir para cualquier $h > 0$, la distribución de X_t es la misma que la de X_{t+h} .

Las probabilidades de transición satisfacen una versión continua de la ecuación de Chapman-Kolmogorov:

Proposición 3.1.1. Para cualesquiera estados $i, j \in E$ y para todo $t \geq 0$ y $s \geq 0$,

$$p_{ij}(t+s) = \sum_{k \in E} p_{ik}(t)p_{kj}(s).$$

Demostración:

$$\begin{aligned}
p_{ij}(t+s) &= P[X_{t+s} = j | X_0 = i] \\
&= \sum_{k \in E} P[X_{t+s} = j, X_t = k | X_0 = i] \\
&= \sum_{k \in E} P[X_{t+s} = j | X_t = k] P[X_t = k | X_0 = i] \\
&= \sum_{k \in E} p_{ik}(t) p_{kj}(s) \blacksquare
\end{aligned} \tag{3.1.2}$$

El resultado de la proposición puede escribirse en notación matricial de este modo:

$$P_{t+s} = P_t P_s. \tag{3.1.3}$$

El conjunto de matrices que componen la ecuación (3.1.3) pertenecen a la colección $\{P_t\}_{t \geq 0}$ que constituye un semigrupo de matrices estocásticas, o semigrupo estocástico, esto quiere decir que cumple con las siguientes propiedades:

1. $P_0 = I$ donde I es la matriz identidad.
2. P_t es una matriz estocástica, es decir que cumple:
 - a) $p_{ij}(t) \geq 0 \forall i, j \in E = \{1, \dots, m\}, t \geq 0$ y
 - b) $\sum_{j=1}^m p_{ij}(t) = 1 \forall i, j \in E$
3. $P_{t+s} = P_t P_s$ para cualesquiera $t, s \geq 0$.

Asumiremos también que $p_{ij}(t)$ es continua para $t > 0$ y en particular $p_{ij}(0) = \delta_{ij}$, donde δ_{ij} es conocida como la función delta de Kronecker:

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j. \end{cases} \tag{3.1.4}$$

Denotaremos por

$$\mathcal{P} = \{P_t = [p_{ij}(t)]_{i,j \in E} \mid 0 \leq p_{ij}(t) \leq 1 \text{ y } \sum_{j \in E} p_{ij}(t) = 1, t \geq 0\},$$

al conjunto de todas las matrices estocásticas respecto a un espacio de estados de dimensión m .

Definición 4. El semigrupo $\{P_t\}_{t \geq 0}$ es llamado **estándar** si $P(t) \rightarrow 0$ cuando $t \downarrow 0$, es decir:

$$\lim_{t \rightarrow 0^+} p_{ij}(t) = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

para todo $i, j \in E$.

Es importante notar que el semigrupo es estándar si y sólo si sus elementos $p_{ij}(t)$ son funciones continuas de t .

La información sobre el conjunto $\{P_t\}_{t \geq 0}$ y la distribución inicial, determinan el comportamiento probabilístico de un proceso de saltos de Markov.

También será importante analizar el comportamiento del tiempo en cada momento del proceso, y su relación con el estado donde se encuentra, para lo cual necesitaremos las siguientes definiciones.

Definición 5. Denotaremos por T_i a la variable aleatoria que representa el **tiempo de permanencia** o de estancia del proceso en el estado i como sigue:

$$T_i = \inf\{t > 0 | X_t \neq i\}.$$

Si el proceso está en el estado i , el tiempo que permanece en ese estado es una variable aleatoria T_i con función de distribución $F_i(t)$. Cada estado i , puede tener una función de distribución diferente $F_i(t)$.

La siguiente proposición nos muestra que la distribución de T_i , es exponencial, con parámetro que depende del estado en el que se encuentra.

Proposición 3.1. Consideremos un proceso de saltos de Markov $X = \{X_t\}_{t \geq 0}$ que se encuentra en el estado $i \in E$ en un tiempo $t \geq 0$. Entonces existe $\lambda_i \geq 0$ independiente del tiempo t , tal que

$$P(T_i > s | X_t = i) = e^{-\lambda_i s}$$

Para toda $s > 0$ y donde $\lambda_i \in [0, +\infty)$ es llamada la tasa de salto (intensidad de salto) asociada al estado $i \in E$, (si $\lambda_i = +\infty$, entonces $e^{-\lambda_i s} = 0, \forall s \geq 0$). Esto quiere decir que los tiempos de estancia tienen distribución exponencial.

Demostración:

Siguiendo la definición 5, $P(T_i > s \mid X_0 = i) = P(T_i > s)$. Entonces la función $g_i(t) = 1 - F_i(t)$, es la probabilidad de cola de T_i entonces:

$g_i(t + s) = P(T_i > t + s) = P(X_u = i, \forall u \in (0, t + s] \mid X_0 = i)$, por propiedades de probabilidad condicional:

$= P(X_u = i, \forall u \in (0, t + s) \mid X_0 = i, X_w = i, \forall w \in (0, t)) \cdot P(X_w = i, \forall w \in (0, t) \mid X_0 = i)$, ahora por propiedad de Markov:

$= P(X_u = i, \forall u \in (t, t + s) \mid X_t = i) \cdot g_i(t)$, por ser homogéneo el proceso:

$$\begin{aligned} &= P(X_u = i, \forall u \in (0, s) \mid X_0 = i) \cdot g_i(t) \\ &= g_i(s)g_i(t) \end{aligned}$$

Dado que las únicas funciones con valores reales que satisfacen la propiedad $g_i(t + s) = g_i(t)g_i(s)$ son de la forma exponencial $g(t) = e^{-\alpha t}$ (ver [15]), y asumiendo que la función g es continua en cero, entonces la solución para $g_i(s)$ es:

$$g_i(s) = e^{-\lambda_i s}.$$

Por lo tanto la función de distribución $F_i(t)$ es la distribución exponencial $\forall i \in E$ es decir, $F_i(t) = 1 - e^{-\lambda_i t}$, donde λ_i es el parámetro de la distribución exponencial en el estado i . ■

Es posible clasificar a los estados en función de su intensidad, como lo describe la siguiente definición.

Definición 6. Si $i \in E$ tiene tasa de intensidad asociada λ_i , entonces decimos que es:

1. **Absorbente o permanente**, si $\lambda_i = 0$. Y si $X_t = i$ en el tiempo t , entonces: $P(X_s = i \mid X_t = i) = 1$, para toda $s > t$, por tanto el proceso permanece en i por siempre.
2. **Estable**, si $P(T_i < \infty) = 1$.
3. **Instantáneo**, si $\lambda_i = \infty$ (no aplica para procesos de saltos de Markov con trayectorias continuas por la derecha). Si $P(T_i = 0) = 1$, o sea la estancia en el tiempo i dura cero casi seguramente.

3.2. Matriz de intensidad.

Nos interesa estudiar las probabilidades de transición para ciertos intervalos de tiempo, particularmente cuando son intervalos pequeños (infinitesimales), es decir, si el proceso de saltos de Markov se encuentra en el estado i en el tiempo t , analizaremos lo que sucede en el intervalo de tiempo $(t, t + h)$ cuando h es pequeño.

Como observamos anteriormente los elementos del conjunto $\{P_t\}_{t \geq 0}$, son funciones continuas por la derecha, por lo que nos interesa analizar el comportamiento del límite de estas funciones.

Karlin & Taylor [14] realizan un desarrollo detallado sobre las propiedades de diferenciabilidad de las probabilidades de transición, de las que se desprenden los siguientes resultados importantes:

Proposición 3.3. Considerando el conjunto $\{P_t\}_{t \geq 0}$, de un proceso de saltos de Markov, entonces:

$$\lim_{h \rightarrow 0^+} \frac{P_{ij}(h) - I}{h}$$

existe, es decir:

$$\lim_{h \rightarrow 0^+} \frac{1 - p_{ii}(h)}{h} = q_i \quad (3.2.1)$$

y

$$\lim_{h \rightarrow 0^+} \frac{p_{ij}(h)}{h} = q_{ij} \quad (3.2.2)$$

con $i \neq j$ existen y definen el generador infinitesimal del proceso de saltos de Markov, donde $0 \leq q_{ij} \leq \infty$ y $0 \leq q_i \leq \infty$.

Demostración. Ver [13].

La ecuación (3.2.2) es llamada la tasa de transición del estado i al estado j , mientras que la ecuación (3.2.1) es la tasa de permanencia en el estado i .

Considerando la relación $1 = p_{ii}(h) + \sum_{i \neq j} p_{ij}(h)$, dividiendo entre h y haciéndola tender a cero tenemos que:

$$q_i = \sum_{i \neq j} q_{ij}$$

De este modo se define la siguiente matriz infinitesimal, también llamada la matriz de intensidad o generador infinitesimal:

$$Q = \begin{pmatrix} -q_1 & q_{12} & \cdots & q_{1m} \\ q_{21} & -q_2 & \cdots & q_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ q_{m1} & q_{m2} & \cdots & -q_m \end{pmatrix}$$

De las ecuaciones (3.2.1) y (3.2.2) tenemos que:

$$\lim_{h \rightarrow 0^+} \frac{P_h - I}{h} = Q \quad (3.2.3)$$

Utilizando las expresiones (3.2.3) y (3.1.3), se sigue que:

$$\frac{P_{t+h} - P_t}{h} = \frac{P_t(P_h - I)}{h} = \frac{P_h - I}{h} P_t \quad (3.2.4)$$

entonces, tomando el límite cuando h tiende a cero, considerando la derivada de P_t dada por:

$$P_t' = P_t Q = Q P_t \quad (3.2.5)$$

donde P_t' denota una matriz cuyos elementos son $p'_{ij}(t)$.

La ecuación (3.2.5) resulta ser una ecuación diferencial ordinaria, que se puede resolver mediante métodos para resolver sistemas de ecuaciones diferenciales ordinarias y con la condición inicial $P_0 = I$, de donde resulta:

$$P_t = \exp(tQ) = I + \sum_{n=1}^{\infty} \frac{Q^n t^n}{n!} \quad (3.2.6)$$

Donde $\exp(*)$ denota la función matriz exponencial. Q es el generador infinitesimal del proceso de saltos de Markov $X = \{X_t\}_{t \geq 0}$, que tiene como principales características que los elementos fuera de la diagonal son no negativos, y la suma de cada renglón es igual a cero.

La matriz Q también es conocida como matriz de tasas y es de suma importancia, pues como vimos anteriormente, es auxiliar para conocer las probabilidades de transición del proceso de saltos de Markov.

El siguiente teorema es importante para la construcción de la densidad del proceso de saltos de Markov, dada una muestra de realizaciones del mismo.

Teorema 3.1. Sea Q el generador infinitesimal del proceso de saltos de Markov $X = \{X_t\}_{t \geq 0}$ entonces:

1. Condicionado a que el proceso de saltos de Markov se encuentre en el estado i al tiempo t , la variable aleatoria T_i (tiempo de estancia) se distribuye exponencialmente con parámetro q_i , es decir

$$P(T_i > s) = \exp(-q_i s) = e^{-q_i s}.$$

2. Si $X_t = i$ y $q_i > 0$, para alguna $t > 0$ existe, casi seguramente, una discontinuidad de la función, que es un salto. Si $0 < t < h \leq \infty$, la probabilidad condicional de que sea un salto a j , dado que $X_t = i$ en $[t, t+h)$, es $\frac{q_{ij}}{q_i}$.

Demostración.

1. Por la proposición 3.1.1 tenemos que la función de distribución $F_{T_i}(t)$ es exponencial, entonces $1 - F_{T_i}(t) = e^{-\lambda_i t}$, donde $\lambda_i = F'_{T_i}(0) = q_i$.

2. Suponemos que $h, t > 0$ y que el proceso salta sólo una vez en $(t, t+h]$, entonces

$$P(\text{saltar a } j \text{ en algún tiempo dentro de } (t, t+h) | X_t = i) \simeq \frac{p_{ij}(h)}{1 - p_{ii}} \rightarrow \frac{q_{ij}}{q_i} \text{ cuando } h \downarrow 0. \blacksquare$$

Sea Q el generador infinitesimal del proceso $X = \{X_t\}_{t \geq 0}$ con espacio de estados $E = \{1, \dots, m\}$ y $P_t = e^{tQ}$. El semigrupo $\{P_t\}_{t \geq 0}$ es la única solución a las ecuaciones retrospectivas (o ecuaciones hacia adelante y hacia atrás) de Kolmogorov:

$$\begin{aligned} \frac{d}{dt} P_t &= P_t Q, \\ \frac{d}{dt} P_t &= Q P_t, \end{aligned}$$

con $P_0 = I$; es decir $p_{ij}(t)$ satisface respectivamente:

$$\begin{aligned} p'_{ij}(t) &= \sum_{k \in E} p_{ik}(t) q_{kj}, \\ p'_{ij}(t) &= \sum_{k \in E} q_{ik} p_{kj}(t). \end{aligned}$$

Seguido de lo anterior obtenemos que las tasas de transición del proceso de

saltos de Markov para $i \neq j$ son:

$$q_{ij} = q_i p_{ij}$$

3.3. Propiedades del proceso de saltos de Markov.

En esta sección describiremos algunas propiedades del proceso de saltos de Markov que son de suma importancia en el desarrollo del trabajo.

3.3.1. Procesos de saltos de Markov regulares.

Definición 7. Sea $X = \{X_t\}_{t \geq 0}$ un proceso de saltos de Markov y t_n el tiempo en el cual el n -ésimo salto ocurre, entonces decimos que el proceso es regular si

$$t_\infty := \sup_{n \in \mathbb{N}} t_n = \infty.$$

Lo anterior quiere decir que, si el proceso salta infinitamente, no lo hace en un tiempo finito, es decir no explota.

Considerando el tiempo en el cual sucede el n -ésimo salto, podemos escribir el tiempo de permanencia o estancia en un estado visitado en el intervalo de tiempo $[t_n, t_{n+1})$, del siguiente modo :

$$T(n) = t_{n+1} - t_n.$$

La siguiente proposición muestra que, el tiempo en el que se realizará el siguiente salto y el estado al cual saltará, son independientes entre si.

Proposición 3.1. Sea $X = \{X_t\}_{t \geq 0}$, un proceso de saltos de Markov regular, con espacio de estados E y $t_{n+1} < \infty$ casi seguramente. Entonces, condicionado a $X_t = i$, las variables aleatorias $T(n+1)$ y $X_{t_{n+1}}$, son independientes, es decir

$$P(T(n+1) > t, X_{t_{n+1}} = j | X_{t_n} = i) = P(T(n+1) > t | X_{t_n} = i) P(X_{t_{n+1}} = j | X_{t_n} = i).$$

Demostración.

Por propiedades de probabilidad condicional se tiene:

$$\begin{aligned} P(T(n+1) > t, X_{t_{n+1}} = j | X_{t_n} = i) &= \\ &= P(T(n+1) > t | X_{t_n} = i) P(X_{t_{n+1}} = j | X_{t_n} = i, T(n+1) > t), \end{aligned}$$

y por propiedad de Markov:

$$\begin{aligned} P(X_{t_{k+1}} = j | X_{t_k} = i, T(n+1) > t) &= \\ &= P(X_{t_{n+1}} = j, X_s = i, t_n \leq s < t_{n+1} | X_{t_{n+t}} = i) \\ &= P(X_{t_{n+1}+T(t_n+t)} = j, X_s = i, t_n \leq s < t_{n+1} + T(t_{n+1} + t) | X_{t_n} = i) \\ &= P(X_{t_{n+1}} = j | X_{t_n} = i). \end{aligned}$$

Por lo tanto

$$P(T(n+1) > t, X_{t_{n+1}} = j | X_{t_n} = i) = P(T(n+1) > t | X_{t_n} = i) P(X_{t_{n+1}} = j | X_{t_n} = i). \blacksquare$$

3.3.2. Relación con Cadenas de Markov a tiempo discreto.

Todo proceso de saltos de Markov $X = \{X_t\}_{t \geq 0}$ tiene asociada una cadena a tiempo discreto a la cual llamaremos $Y = \{Y_n\}_{n=0,1,\dots}$, y que está dada por la primera, pero observada en los tiempos en donde se efectúan los saltos, regularmente este proceso es llamado “proceso incrustado”.

Sea $\omega \in \Omega$ y considerando las definiciones de la sección anterior, $T(n)$ es el tiempo de permanencia en el momento n , mientras que $t_{n+1} = t_n + T(n+1)$ es el tiempo en el momento $n+1$ en el cual hay un cambio de estado, podemos describir la realización $t \rightarrow X_t(\omega)$ del proceso de saltos $X = \{X_t\}_{t \geq 0}$, del siguiente modo: inicialmente, en el tiempo $t_0 = 0$ el proceso está en algún estado que denotaremos por $Y_0(\omega) = X_0(\omega)$, y permanece ahí un cierto tiempo positivo, para después, en un tiempo t_1 saltar a un nuevo estado $Y_1(\omega) = X_{t_1}(\omega)$, donde permanecerá hasta un tiempo t_2 donde saltará a un estado diferente $Y_2(\omega) = X_{t_2}(\omega)$, y así sucesivamente.

De este modo los tiempos t_0, t_1, t_2, \dots son los momentos de transición del proceso de saltos de Markov, y Y_0, Y_1, Y_2, \dots es la sucesión de estados visitados por $X = \{X_t\}_{t \geq 0}$, por lo que si $Y_n = i$, el intervalo $[t_n, t_{n+1})$, es el intervalo de permanencia en el estado i . Además por la propiedad de no explosión $0 < T(t_n) < \infty$ y también $Y_n \in E$ para cada $n \in \mathbb{N}$.

Definición 8. Denotamos por $\{Y_n\}_{n \in \mathbb{N}}$, al proceso incrustado en el proceso de saltos de Markov $X = \{X_t\}_{t \geq 0}$, el cual es una cadena de Markov homogénea en tiempo discreto con espacio de estados E , y con matriz de probabilidades de transición $\mathbf{P} = [p_{ij}]_{ij \in E}$, para la cual

$$p_{ij} = P(X_{t_n} = j \mid X_0 = i),$$

y $p_{ii} = 0$.

Se pueden ver a detalle todas las propiedades de las cadenas de Markov discretas en [18, 19, 15]. Como veremos en la siguiente sección, el proceso incrustado nos ayudará a simular un proceso de saltos de Markov, pero primero debemos estudiar los siguientes propiedades que nos servirán como herramienta para justificar resultados más adelante.

El siguiente teorema demuestra que $\{Y_n\}_{n \in \mathbb{N}}$ es una cadena de Markov discreta y que $t_{n+1} - t_n$, tiene distribución exponencial con parámetro que depende de el proceso Y_n .

Teorema 3.2. Para cualquier $n \in \mathbb{N}, j \in E$ y $u \in \mathbb{R}^+$ tenemos que

$$P[Y_{n+1} = j, t_{n+1} - t_n > u \mid Y_0, \dots, Y_n; t_0, \dots, t_n] = p_{ij} e^{-\lambda_i u} \quad (3.3.1)$$

si sucede que $Y_n = i$, y donde $p_{ij} \geq 0$, $p_{ii} = 0$ y $\sum_{j \in E} p_{ij} = 1$.

Demostración.

Fijamos j, n y u . Dado que t_n representa el momento en el que cambia de estado el proceso, $t_n \leq t$ si y solo si existen $0 < s_1 < s_2 < \dots < s_n \leq t$, tales que $X_0 \neq X_{s_1} \neq \dots \neq X_{s_n}$, y la existencia de cada s_1, \dots, s_n , puede determinarse si se conoce $X_s \forall s \leq t$.

Escribiremos $T(n) = t_{n+1} - t_n$ y denotaremos $Y_{n+1} = X_{t_n + T(n)}$, esto nos indica que si conocemos Y_0, \dots, Y_n y los tiempos t_0, \dots, t_n entonces conocemos X_t para toda $t \leq t_n$. Entonces el lado izquierdo de la ecuación (3.3.1) se puede escribir como

$$P[X_{t_n+T(n)} = j, T(n) > u | X_t; t \leq t_n],$$

por la propiedad de Markov se puede escribir como $g(X_{t_n}) = g(Y_n)$, donde

$$g(i) = P[X_{T(0)} = j, T(0) > u | X_0 = i] \quad (3.3.2)$$

por lo que debemos demostrar que $g(i)$ es de la forma $g(i) = p_{ij}e^{-\lambda_i u}$. Entonces podemos escribir (3.3.2) como:

$$g(i) = P[T(0) > u | X_0 = i]P[X_{T(0)} = j | X_0 = i, T(0) > u].$$

Por la proposición 3.1 $P[T(0) > u | X_0 = i] = e^{-\lambda_i u}$. Por otro lado el evento $\{X_0 = i, T(0) > u\}$ es el mismo que $\{X_s = i, s \leq u\}$, y dado que $T(0) > u$ tenemos que $T(0) = u + T(u)$, de esta forma tenemos

$$\begin{aligned} P[X_{T(0)} = j | X_0 = i, T(0) > u] &= P[X_{u+T(u)} = j | X_s = i; s \leq u] \\ &= P[X_{u+T(u)} = j | X_u = i] \\ &= P[X_{T(0)} = j | X_0 = i] = p_{ij} \end{aligned}$$

y que es independiente de u , por lo que con el resultado anterior y el resultado de la proposición 3.1 podemos reescribir $g(i) = p_{ij}e^{-\lambda_i u}$ ■.

Si en el resultado del Teorema 3.2 tomamos $u = 0$, se tiene la siguiente consecuencia:

Corolario. La sucesión de los estados visitados $\{Y_n\}_{n \in \mathbb{N}}$, forma una cadena de Markov discreta con matriz de transición P_{ij} . Del teorema 3.2 se sigue que

$$P[t_{n+1} - t_n > u | Y_n = i, Y_{n+1} = j] = e^{-\lambda_i u}$$

independiente de u . Es decir la distribución del tiempo de estancia solo depende del estado en el que se encuentra durante ese tiempo y no del estado al que salta después. En general los tiempos entre transiciones son condicionalmente independientes unos de otros, dada la sucesión de estados visitada, y cada tiempo de permanencia tiene distribución exponencial con parámetro que depende del estado visitado.

Observación. Si consideramos un proceso de saltos de Markov $X = \{X_t\}_{t \geq 0}$

con espacio de estados E y con un proceso incrustado $\{Y_n\}_{n \in \mathbb{N}}$, entonces el proceso incrustado $\{Y_n\}_{n \in \mathbb{N}}$, es una cadena de Markov en $E \times \mathbb{R}$, con probabilidades de transición

$$P(Y_{n+1} = j, t_{n+1} \leq t | Y_n = i, t_n) = p_{ij}(1 - e^{-q_i t}) \quad (3.3.3)$$

con $i, j \in E$, $t \geq 0$, donde q_i son constantes positivas y $P = \{p_{ij}\}$ es la matriz de transición con $p_{ii} = 0$.

Es importante mencionar que las probabilidades de salto p_{ij} son diferentes a las probabilidades de transición $p_{ij}(t)$, pues las primeras denotan la probabilidad de cambiar al estado j cuando el proceso se encuentra en el estado i y tiene un salto, mientras que las segundas denotan la probabilidad de encontrar al proceso en el estado j , partiendo del estado i , al finalizar un intervalo de tiempo de longitud t .

Proposición 3.4. Si el proceso de saltos de Markov $X = \{X_t\}_{t \geq 0}$ en E , es tal que su proceso incrustado $\{Y_n\}_{n \in \mathbb{N}}$ satisface las condiciones del Teorema 3.2, entonces sus probabilidades de transición satisfacen

$$p_{ij}(t) = e^{-q_i t} \delta_{ij} + \int_0^t \sum_{k \neq i} p_{kj}(t-s) q_{ik} e^{-q_i s} ds.$$

donde δ_{ij} es la delta de Kronecker

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j. \end{cases}$$

Demostración.

Para cualesquiera $i, j \in E$ y condicionando con $N(t)$ que representa al mayor entero n para el cual $t_n < t$, se tiene:

$$P(X_{t+u} = j | X_t = i, X_s, s < t) = \sum_{n=0}^{\infty} (\alpha_n(t, u) + \beta_n(t, u)) P(N(t) = n | X_t = i, X_s, s < t)$$

donde

$$\begin{aligned}
\alpha_n(t, u) &= P(X_{t+u} = j, t_{n+1} > t + u | F_n(i, t)) \\
\beta_n(t, u) &= P(X_{t+u} = j, t_{n+1} \leq t + u | F_n(i, t)) \\
F_n(i, t) &= \{N(t) = n, X_t = i, X_s, s < t\}
\end{aligned}$$

Dado que el tiempo restante $t_{n+1} - t$, se distribuye exponencial con parámetro q_i , entonces

$$\beta_n(t, u) = \delta_{ij} P(t_{n+1} - t > u | F_n(i, t)) = \delta_{ij} e^{-q_i u}$$

Ahora, condicionando a X_{n+1} y t_{n+1} y utilizando el resultado de la observación, o sea la ecuación (3.3.3),

$$\begin{aligned}
\beta_n(t, u) &= \sum_{k \neq i} \int_0^u P(X_{t+u} = j | t_{n+1} = t + v, X_{n+1} = k, F_n(i, t)) \\
&\quad \times P(X_{n+1} = k, t_{n+1} - t \in dv | F_n(i, t)) \\
&= \sum_{k \neq i} \int_0^u p_{kj}(u - v) q_i p_{ik} e^{-q_i v}
\end{aligned}$$

usando las tasas de transición de nuestro generador infinitesimal tenemos,

$$\beta_n(t, u) = \sum_{k \neq i} \int_0^u p_{kj}(u - v) q_{ik} e^{-q_i v} dv$$

Sabiendo que lo anterior es independiente de n y t podemos sustituir la expresión del siguiente modo

$$P(X_{t+u} = j | X_t = i, X_s, s < t) = e^{-q_i t} \delta_{ij} + \int_0^u \sum_{k \neq i} p_{kj}(u - v) q_{ik} e^{-q_i v} dv \blacksquare.$$

En la siguiente sección se presentan un par de ejemplos específicos para comprender mejor lo que se ha estudiado hasta este momento.

3.3.3. Ejemplos.

Ejemplo 1. Proceso de nacimiento y muerte. Consideremos un sistema

el cual, un estado a cualquier tiempo, es representado por el número de individuos que hay en el sistema en ese tiempo. Supongamos que siempre que el sistema existan n individuos, entonces: i) nuevos arribos llegan al sistema con una tasa exponencial λ_n , y ii) los individuos abandonan el sistema con una tasa μ_n ; esto es, siempre que en el sistema existan n individuos, entonces el tiempo hasta el siguiente arribo se distribuye exponencial con media $1/\lambda_n$ y es independiente del tiempo hasta el siguiente abandono, el cual también se distribuye exponencialmente pero con media $1/\mu_n$.

Un sistema de este tipo es llamado proceso de nacimiento y muerte y los parámetros $\{\lambda_n\}_{n=0}^{\infty}$ y $\{\mu_n\}_{n=0}^{\infty}$, son llamados, tasas instantáneas de llegada (nacimiento) y abandono (muerte), respectivamente.

De este modo, un proceso de nacimiento y muerte es un proceso de saltos de Markov con espacio de estados $\{0, 1, \dots\}$, para el cual las transiciones del estado n , pueden ir únicamente al estado $n - 1$, o al estado $n + 1$.

El generador infinitesimal para un proceso de nacimiento y muerte con las características mencionadas es:

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \dots \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

A partir de esta matriz podemos concluir que el tiempo de estancia en cualquier estado $i \geq 0$ tiene distribución exponencial con parámetro $(\lambda_i + \mu_i)$, donde definimos $\mu_0 = 0$ y la probabilidad de transición del estado 0 al 1 es $p_{01} = 1$, las demás probabilidades de transición, tienen una relación con las instantáneas de nacimiento y muerte, del siguiente modo:

$$p_{ij} = \begin{cases} \frac{\lambda_i}{\lambda_i + \mu_i} & \text{si } j = i + 1 \\ \frac{\mu_i}{\lambda_i + \mu_i} & \text{si } j = i - 1 \\ 0 & \text{otro caso} \end{cases}$$

Existen algunas variantes del proceso de nacimiento y muerte, como son el proceso de nacimiento puro, proceso de muerte puro, o el proceso de Yule ver [13, 18]. El proceso de Poisson con tasa λ es un proceso de nacimiento puro con tasas de nacimiento y muerte $\lambda_n = \lambda$ y $\mu_n = 0$, detalles en [15].

Ejemplo 2. Sistemas de Markov de colas. Supongamos que clientes llegan a un solo servidor en una estación de servicio de acuerdo a un proceso Poisson con tasa λ . Esto es, los tiempos sucesivos entre arribos son variables aleatorias independientes exponenciales con media $1/\lambda$. A su llegada, cada cliente se dirige directamente al servidor si se encuentra vacío, si no el cliente entra en la cola (es decir, espera su turno). Cuando el servidor termina con el cliente, éste deja el sistema y el siguiente cliente en espera (si existe) entra al servidor. Los tiempos sucesivos de servicio son variables aleatorias independientes exponenciales de media $1/\mu$, diferentes a los tiempos de arribos.

El modelo anterior se conoce como sistema de colas M/M/1 (la primera M se refiere a que el proceso entre llegadas, por ser un proceso de Poisson, es Markoviano, la segunda M es debido a que la distribución del servidor es exponencial y por lo tanto Markoviano, y el 1 es referente a un solo servidor).

Este sistema de colas es también un proceso de nacimiento y muerte con tasas: $\mu_n = \mu, n \geq 1$ y $\lambda_n = \lambda, n \geq 0$.

La matriz de intensidad para este ejemplo es:

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -(\lambda + \mu) & \lambda & 0 & \cdots \\ 0 & \mu & -(\lambda + \mu) & \lambda & \cdots \\ 0 & 0 & \mu & -(\lambda + \mu) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Existen algunas variantes de los sistemas de Markov de colas, como el sistema de colas para k servidores o sistema M/M/ k , e incluso el sistema M/M/ ∞ con servidores infinitos, los ejemplos y detalles de sistemas de Markov de colas pueden ser estudiados en [13, 14].

3.3.4. Simulación del proceso de saltos de Markov.

A continuación describiremos, utilizando todo lo visto hasta ahora, un algoritmo que nos ayudará a simular un proceso de saltos de Markov.

Considerando un proceso de saltos de Markov $X = \{X_t\}_{t \geq 0}$ con tasas de intensidad $\{q_{ij}\}_{i,j \in E}$. Sea $Y = \{Y_n\}_{n \geq 0}$ la cadena de Markov asociada al proceso, o el proceso incrustado, el tiempo de permanencia en cada estado i se distribuye exponencialmente con parámetro que depende del estado. Llamamos K a la matriz de transición de primer paso del proceso Y , recordando las tasas de intensidad entonces: $q_i = \sum_{j \neq i} q_{ij}$, entonces $K_{ij} = \frac{q_{ij}}{q_i}$ para $i \neq j$, esto im-

plica que $K_{ii} = 0$. Si llamamos T_1, T_2, \dots a los tiempos de estancia del proceso de saltos y t_0, t_1, \dots a los tiempos en los que ocurre un salto del proceso en $1, 2, \dots$; el algoritmo para generar un proceso de saltos de Markov es:

Algoritmo 1. Simulación de un proceso de saltos de Markov.

1. Inicializar t_0 . Generar Y_0 de la distribución inicial de Y . Hacemos $X_0 = Y_0$ y $n = 0$.
2. Hacemos $Y_n = i$
3. Generamos T_{n+1} de la exponencial que depende de Y , $Exp(q_i)$.
4. Hacemos $t_{n+1} = t_n + T_{n+1}$.
5. Ahora $X_n = Y_n$ para $t_n \leq t < t_{n+1}$.
6. Generamos Y_{n+1} de la distribución correspondiente de K . Ahora $n = n + 1$, regresamos al paso 2.

Ejemplo 3. Simulación de un proceso de saltos de Markov.

Implementando el algoritmo anterior con un espacio de estados finito $E = \{1, 2, 3\}$

Suponiendo que nuestro generador infinitesimal es:

$$Q = \begin{pmatrix} -2.51 & 1.6 & 0.91 \\ 2.5 & -3 & 0.5 \\ 0.97 & 3.02 & -3.99 \end{pmatrix} \quad (3.3.4)$$

Se implementó el algoritmo 1 utilizando R (ver código en el Apéndice), para obtener una matriz de transición calculada con la ecuación (3.2.6), para un tiempo $t = 5$ es:

$$P_5 = \exp(5Q) = \begin{pmatrix} 0.46 & 0.39 & 0.15 \\ 0.45 & 0.40 & 0.15 \\ 0.44 & 0.40 & 0.16 \end{pmatrix} \quad (3.3.5)$$

Utilizando los resultados anteriores, a continuación se ilustra la realización de una trayectoria de un Proceso X en el intervalo de tiempo $[0,5]$, comenzando en el estado 1, que también se obtuvo con el algoritmo 1.

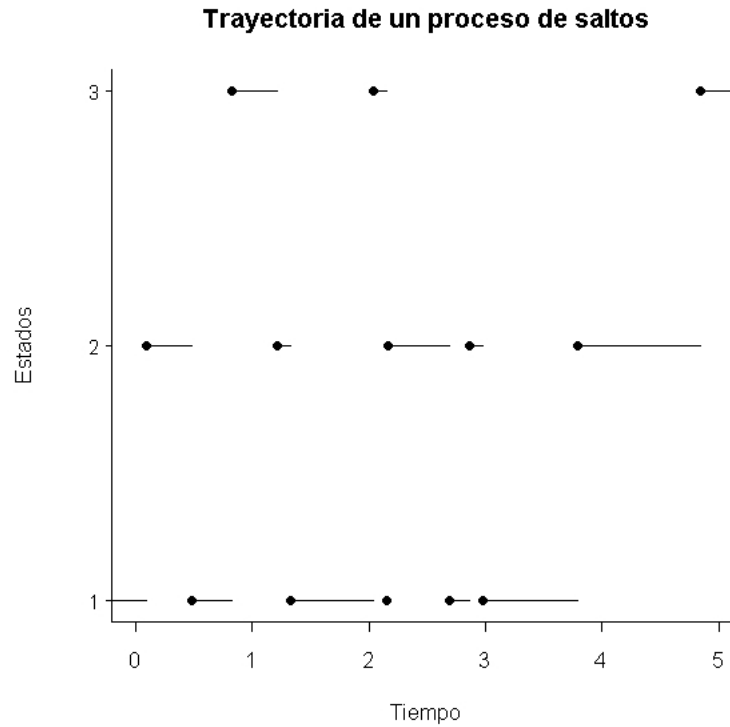


Figura 3.3.1: Realización típica de un proceso de saltos de Markov

3.4. Inferencia estadística para procesos de saltos de Markov.

La inferencia estadística, se define como la colección de técnicas que permiten formular inferencias inductivas. Los parámetros o sus funciones se estiman con base en estadísticas, que a su vez se obtienen a partir de la información contenida en una muestra aleatoria. Consideraremos como estadística a cualquier función de las variables aleatorias que se observan en la muestra, tal que, dicha función no contiene cantidades desconocidas.

En el caso de procesos de saltos de Markov, como veremos en esta sección, nos interesará estimar el generador infinitesimal Q , cuando se cuenta únicamen-

te con una muestra finita del proceso de saltos en tiempos discretos, sabiendo y justificando que el número de saltos de un estado a otro y el tiempo de permanencia en cada estado son estadísticas suficientes.

La principal herramienta que se necesita para el cálculo del estimador máximo verosímil es la función de verosimilitud, la cual explicaremos a continuación.

3.4.1. Introducción.

Sea Y una variable aleatoria, definida de tal modo que sus valores representan las posibles opciones en que puede ocurrir un fenómeno aleatorio antes de llevar a cabo un experimento, entonces su función de probabilidad (función de densidad) es $f_Y(y)$. Sea $f(x|y)$ la función de densidad de probabilidad condicional de cualquier variable aleatoria X^* , que representa evidencia muestral en función de una alternativa fija y de Y , entonces la función $f(x|y)$ recibe el nombre de función de verosimilitud, pues representa el grado de concordancia del resultado muestral x , dado el valor y de Y .

Consideraremos que nuestra población está constituida por un número infinito de posibles resultados para alguna característica medible de interés. Sea X una característica medible y $f(x; \theta)$ la función de densidad de probabilidad de la distribución de la población que depende de un parámetro θ , nos interesará entonces estimar este parámetro.

Definimos a continuación una muestra aleatoria.

Definición 9. Una muestra aleatoria de una población, es un conjunto de n variables aleatorias X_1, X_2, \dots, X_n independientes, que tienen la misma función de probabilidad que la de la distribución de la población.

Entonces la función conjunta de probabilidad de X_1, X_2, \dots, X_n es la función de verosimilitud de la muestra dada por:

$$L(x^*; \theta) = \prod_{i=1}^n f(x_i; \theta),$$

donde $x^* = \{x_i\}_{i=1, \dots, n}$ denota los datos muestreados, y cuando éstos se conocen, $L(x^*; \theta)$ es una función de parámetro desconocido θ . Dicha función nos es útil para estimar parámetros.

Para el caso del proceso de saltos de Markov, es necesario observar algunos resultados para poder construir la función de verosimilitud que nos interesa, y

el parámetro que nos interesará estimar más adelante.

Supongamos que el proceso de saltos de Markov con el que hemos trabajado hasta el momento, $X = \{X_t\}_{t \geq 0}$, con espacio de estados finito $E = \{1, 2, \dots, m\}$ y matriz de intensidad Q ; ha sido observado continuamente durante cierto intervalo de tiempo $[0, \tau]$.

Podemos representar la trayectoria de $X = \{X_t\}_{0 \leq t \leq \tau}$, como una secuencia ordenada de variables aleatorias:

$$\{X_t\}_{0 \leq t \leq \tau} = \{(x_1, \Delta t_1), (x_2, \Delta t_2), \dots, (x_n, \Delta t_n), (x_{n+1}, \Delta t_{n+1})\}.$$

Donde, la trayectoria comienza en el estado x_1 en el tiempo cero y permanece ahí durante Δt_1 unidades de tiempo, después salta a x_2 y permanece en ese estado durante Δt_2 unidades de tiempo, y así sucesivamente salta al estado x_n y permanece ahí durante Δt_n unidades de tiempo, finalmente salta al estado x_{n+1} y permanece ahí por lo menos hasta el tiempo τ , entonces n saltos han sido realizados en la trayectoria.

Definimos como $N(\tau)$, como el mayor entero n , para el cual $t_n < \tau$, es decir el número de saltos realizados antes del tiempo τ ; entonces podemos escribir la distribución de probabilidad del siguiente modo:

Teorema 3.3. Sea

$$r_{ij} = \begin{cases} 0 & \text{si } i = j \\ q_{ij} & \text{si } i \neq j \end{cases}$$

entonces,

$$P(N(\tau) = n, X_{t_0} = x_1, \Delta t_1 \leq s_1, \dots, X_{t_{n-1}} = x_n, \Delta t_n \leq s_n, X_{t_n} = x_{n+1}) = P(X_{t_0} = x_1) e^{-q_{x_{n+1}} \tau} \int_{t^*} \prod_{i=1}^n d\Delta t_i r_{x_i x_{i+1}} e^{-(q_{x_i} - q_{x_{n+1}}) \Delta t_i}$$

donde:

$$t^* = \{(\Delta t_1, \Delta t_2, \dots, \Delta t_n) : \sum_{i=1}^n \Delta t_i < \tau, 0 \leq \Delta t_i \leq s_i\},$$

si $n > 0$ y $P(N(\tau) = 0, X_{t_0} = x_1) = P(X_{t_0} = x_1) e^{-q_{x_1} \tau}$.

Demostración. Ver [3].

Este teorema nos permite ver la densidad conjunta del proceso de saltos de

Markov como el producto de las densidades de todo el proceso, este resultado nos permitirá justificar la construcción de la función de verosimilitud, pero además necesitamos el resultado del teorema que describiremos a continuación.

Definimos a B , como una posible trayectoria dentro del intervalo $[0, \tau]$, y llamamos μ a la medida sobre el espacio de todas las trayectorias del proceso en el intervalo $[0, \tau]$, podemos ahora escribir la densidad del proceso de saltos de Markov mediante el siguiente teorema:

Teorema 3.4. Si B es una trayectoria dentro del intervalo $[0, \tau]$, la cual es medible respecto a μ , entonces

$$P(B) = \int_B f_Q(x^*) d\mu(x^*),$$

donde x^* denota la realización de $X = \{X_t\}_{0 \leq t \leq \tau}$ y

$$f_Q(x^*) = \begin{cases} P(X_{t_0} = x_1) e^{-q_{x_1} \tau} & \text{si } x^* = x_1 \\ P(X_{t_0} = x_1) e^{-q_{x_{n+1}} \tau} \prod_{i=1}^n r_{x_i x_{i+1}} e^{-(q_{x_i} - q_{x_{n-1}}) \Delta t_i} & \text{si } x^* = \{(x_1, \Delta t_1), (x_2, \Delta t_2), \dots, (x_1, \Delta t_1), x_{n+1}\} \\ 0 & \text{en otro caso.} \end{cases}$$

En el segundo caso cuando $n > 0$, $x_i \in E$, $\Delta t_i \geq 0$, $i = 1, 2, \dots, n$ y $\sum_{i=1}^n \Delta t_i < t$.

Demostración. Ver [9].

Ahora, suponiendo k realizaciones independientes $x_1^*, x_2^*, \dots, x_k^*$ de $X = \{X_t\}_{0 \leq t \leq \tau}$, que han sido observadas de manera continua, la función de verosimilitud en caso continuo denotada por $L_\tau^{(c)}(Q)$ es tradicionalmente definida mediante la siguiente ecuación:

$$L_{\tau,k}^{(c)}(Q) = \prod_{i=1}^k f_Q(x_i^*). \quad (3.4.1)$$

3.4.2. Función de verosimilitud.

Trabajaremos con los conceptos de función de verosimilitud en los casos cuando se tienen observaciones en tiempo continuas y discretas y también se define el estimador máximo verosímil en dichos casos.

El desarrollo de la estimación del generador infinitesimal, fue trabajada por Bladt & Sorensen [4, 5], quienes muestran importantes resultados respecto a la dificultad que presenta el cálculo de la verosimilitud en el caso cuando el proceso de saltos es observado únicamente en puntos discretos del tiempo, como veremos a continuación.

Caso continuo.

Suponemos un proceso de saltos de Markov $X = \{X_t\}_{t \geq 0}$ con espacio de estados finito $E = \{1, 2, \dots, m\}$ y matriz de intensidad $Q = \{q_{ij}\}_{i,j \in E}$. El proceso X ha sido observado continuamente durante un intervalo de tiempo $[0, \tau]$, es decir $\{X(t) | 0 \leq t \leq \tau\}$.

Definimos la variable aleatoria $R_i(\tau)$ como:

$$R_i(\tau) = \int_0^\tau I_i(X_s) ds \quad (3.4.2)$$

la cual representa el tiempo de permanencia en el estado i antes del tiempo τ , con $I_i(X_s)$ la función indicadora; y denotaremos como $N_{ij}(\tau)$ al número de transiciones del estado i al estado j , $\forall i, j \in E$ en el intervalo de tiempo $[0, \tau]$. Utilizando los resultados de los teoremas 3.3 y 3.4 podemos escribir la función de verosimilitud $L_\tau^{(c)}$ cuando se tienen observaciones continuas de las trayectorias observadas de $\{X_t : 0 \leq t \leq \tau\}$ del siguiente modo:

$$L_\tau^{(c)}(Q) = \prod_{i=1}^m \prod_{j \neq i} q_{ij}^{N_{ij}(\tau)} e^{-q_{ij} R_i(\tau)} \quad (3.4.3)$$

Aplicando la función logaritmo a la verosimilitud anterior tenemos:

$$\log L_\tau^{(c)}(Q) = \sum_{i=1}^m \sum_{j \neq i} [N_{ij}(\tau) \log(q_{ij}) - q_{ij} R_i(\tau)], \quad (3.4.4)$$

mediante la factorización de Halmos-Savage, que consiste en factorizar la función de verosimilitud de tal forma que uno de los elementos de la factorización es una función que no depende del parámetro, concluimos que $N_{ij}(\tau)$ y $R_i(\tau)$, son estadísticas suficientes para Q , es decir que contienen toda la información respecto a Q .

De lo anterior se sigue que:

$$\frac{\partial \log L_\tau^{(c)}(Q)}{\partial q_{ij}} = \sum_{i=1}^m \sum_{j \neq i} \left[\frac{N_{ij}(\tau)}{q_{ij}} - R_i(\tau) \right]$$

3.4. INFERENCIA ESTADÍSTICA PARA PROCESOS DE SALTOS DE MARKOV.43

por lo que

$$\frac{\partial \log L_\tau^{(c)}(\mathbf{Q})}{\partial q_{ij}} = 0 \Leftrightarrow q_{ij} = \frac{N_{ij}(\tau)}{R_i(\tau)}$$

y

$$\frac{\partial \log L_\tau^{(c)}(\mathbf{Q})}{\partial q_{ij} \partial q_{kl}} = \frac{-N_{ij}(\tau)}{q_{ij}^2} I_k(i) I_l(j) \leq 0,$$

con $i, j, k, l \in E$.

Los resultados anteriores nos muestran cual es el valor del parámetro que maximiza la función de verosimilitud por lo que- obtenemos que el Estimador Máximo Verosímil de \mathbf{Q} es:

$$\hat{q}_{ij}(\tau) = \frac{N_{ij}(\tau)}{R_i(\tau)} \quad (3.4.5)$$

seguido del hecho de que $R_i(\tau) > 0$.

Es importante mencionar que si el proceso no ha estado en el estado i , no existe información sobre q_{ij} y el estimador máximo verosímil de q_{ij} no existe.

Ejemplo 4. Estimador máximo verosímil caso continuo.

Utilizando los datos del ejemplo de la simulación del proceso de saltos de la sección anterior (Ejemplo 3) y utilizando un intervalo de tiempo $[0, 1000]$ obtenemos que nuestra matriz de saltos es:

$$N_{ij} = \begin{pmatrix} 0 & 757 & 413 \\ 1011 & 0 & 200 \\ 158 & 455 & 0 \end{pmatrix}$$

y los tiempos de permanencia por estado son $R_1 = 472.55, R_2 = 379.39$ y $R_3 = 148.04$.

Estimando \mathbf{Q} utilizando el estimador de la ecuación (3.4.5) obtenemos:

$$\hat{\mathbf{Q}} = \begin{pmatrix} -2.47 & 1.60 & 0.87 \\ 2.66 & -3.19 & 0.53 \\ 1.06 & 3.07 & -4.14 \end{pmatrix}$$

El cuadro 3.1 nos muestra , la convergencia hacia los valores reales, variando el tiempo, mientras que para el cuadro 3.2. se fijó el horizonte de tiempo en 1000

Transición de estados		Tiempo						Real
De	A	10	100	250	500	750	1000	
1	2	1.2903	1.6618	1.4948	1.4375	1.6359	1.6019	1.60
1	3	0.6451	0.8429	1.0092	0.8829	0.9390	0.8739	0.91
2	1	4.2582	2.0250	2.1659	2.6517	2.6447	2.6647	2.50
2	3	2.1291	0.3877	0.4594	0.6544	0.5649	0.5271	0.50
3	1	1.3777	0.8290	0.8089	0.8377	0.9597	1.0672	0.97
3	2	4.8222	3.5649	3.2359	2.8190	2.7701	3.0733	3.02

Cuadro 3.1: Estimación de intensidades de transición en diferentes momentos en el tiempo.

y se realizaron 5 trayectorias diferentes.

Transición de estados		Trayectorias con horizonte de tiempo t=1000					Real
De	A	1	2	3	4	5	
1	2	1.5042	1.6535	1.6482	1.6358	1.5758	1.60
1	3	0.9102	0.9410	0.9084	0.9874	0.8911	0.91
2	1	2.4276	2.5856	2.5420	2.4728	2.4750	2.50
2	3	0.5702	0.5654	0.6091	0.5270	0.4687	0.50
3	1	0.8177	0.9234	0.8843	0.9437	1.0617	0.97
3	2	2.9520	2.9646	2.9717	2.9905	3.1013	3.02

Cuadro 3.2: Estimación de intensidades con horizonte fijo y diferentes trayectorias.

Caso Discreto.

Ahora consideramos el caso en el que el proceso es observado únicamente en puntos de tiempo discretos, $0 = t_0 < t_1 < \dots < t_n = \tau$, es decir, $\{X_{t_0}, \dots, X_{t_n}\}$; entonces el proceso $Y = \{Y_i\}_{i=0}^n = \{X_{t_i}\}_{i=0}^n$ es una Cadena de Markov discreta, generalmente en un tiempo no homogéneo, para el cual la matriz de transición en el tiempo i es $P^{\Delta_i}(Q) = \exp(\Delta_i Q) = e^{\Delta_i Q}$, donde $\Delta_i = t_{i+1} - t_i$. Además definimos a \mathcal{Q} como el conjunto de todas las matrices de intensidad (aquellas cuya diagonal está compuesta por entradas negativas y la suma por renglones es igual a 0), es decir

$$\mathcal{Q} = \{Q = (q_{ij})_{i,j \in E} | q_{ij} \geq 0 \forall i \neq j, q_{ii} = -\sum_{j \neq i} q_{ij}\}.$$

Así pues, la función de verosimilitud para tiempo discreto está dada en términos de la matriz de transición $P^t(Q) = \exp(tQ)$ $t > 0$ de la siguiente manera:

$$L_n^{(d)}(Q) = \prod_{i=1}^{n-1} P_{x_i, x_{i+1}}^{\Delta_i}(Q) \quad Q \in \mathcal{Q} \quad (3.4.6)$$

donde x_i, \dots, x_n denota los valores observados de X .

En el caso en el que las observaciones son equidistantes, es decir, $\Delta_i = \Delta$ para alguna $\Delta > 0$, la Cadena de Markov discreta Y es homogénea en el tiempo y tiene matriz de transición $P^\Delta(Q)$, por lo cual la función de verosimilitud en tiempo discreto esta dada por:

$$L_n^{(d)}(Q) = \prod_{i=1}^m \prod_{j=1}^m (P_{ij}^\Delta(Q))^{c_{ij}(n)} \quad (3.4.7)$$

donde $c_{ij}(n)$ denota el número de transiciones de i a j antes de punto de tiempo discreto (n) de la Cadena de Markov Y . En este caso la función de log-verosimilitud es:

$$\log L(Q) = \sum_{i=1}^m \sum_{j=1}^m (\log(P_{ij}^\Delta(Q))^{c_{ij}(n)}) \quad (3.4.8)$$

Derivando respecto a las entradas de Q :

$$\frac{\partial}{\partial Q} \log L_n^{(d)}(Q) = \sum_{k=1}^{\infty} \sum_{l=1}^k \frac{\Delta^k}{k!} (Q^T)^{l-1} Z (Q^T)^{k-l} \quad (3.4.9)$$

donde $Z = (Z_{ij})_{ij \in E}$ y $z_{ij} = c_{ij}(n)/e^{(\Delta Q)_{ij}}$, esta derivada es su caso simplificado el cual, tiene una forma complicada que no puede ser encontrada analíticamente. Por lo cual, no existe una expresión analítica para el estimador máximo verosímil respecto a Q .

Por otro lado la log-verosimilitud respecto a la matriz de transición P es

$$L_n(P) = \prod_{i=1}^m \prod_{j=1}^m P_{ij}^{c_{ij}(n)}, \quad P \in \mathcal{P},$$

donde $\mathcal{P} = \{P_t = p_{ij}(t)_{ij \in E} | 0 \leq p_{ij}(t) \leq 1 \text{ y } \sum_{j \in E} p_{ij}(t) = 1, t \geq 0\}$. Se puede observar que esa función de verosimilitud es idéntica a una verosimilitud de m distribuciones multinomiales independientes por lo que su derivada puede ser obtenida analíticamente y el estimador máximo verosímil para el parámetro P es:

$$\hat{P}_{ij} = \frac{c_{ij}(n)}{c_i(n)} \quad (3.4.10)$$

$$\text{Donde } c_i(n) = \sum_{j=1}^m c_{ij}(n).$$

Debido a que no es posible calcular analíticamente el estimador máximo verosímil de Q en el caso discreto, se considerara el problema de como determinar el generador si solo se cuenta con observaciones en puntos de tiempo discretos $0 = t_0, t_1, \dots, t_n = \tau$, es decir se tendrá una observación incompleta y por lo tanto la información entre dos observaciones consecutivas está escondida y $R_i(\tau)$ y $N_{ij}(\tau)$ son desconocidos.

Si definimos a $\mathcal{P}_0 = \{e^Q | Q \in \mathcal{Q}\}$ como el conjunto de las matrices de transición correspondientes a la observación de tiempo discreto de un proceso de saltos de Markov en tiempo continuo y suponemos que calcularemos \hat{P} con la ecuación (3.4.10), basados en las observaciones de tiempo discreto del proceso de saltos de Markov a tiempo continuo, entonces si $\hat{P} \in \mathcal{P}_0$, existe $\hat{Q} \in \mathcal{Q}$ tal que $P^\Delta(\hat{Q}) = \hat{P}$, y la función de verosimilitud (3.4.7) alcanza su máximo en \hat{Q} , que es el estimador máximo verosímil.

Sin embargo, existen algunas observaciones que mencionar: primero, el conjunto \mathcal{P}_0 suele ser complicado excepto para el caso en el que $m = 2$, y segundo hecho es que la matriz de exponenciales no es inyectiva en todo el dominio (si los eigenvalores son complejos) y por lo tanto \hat{Q} no es necesariamente única; éste problema de identificar el conjunto \mathcal{P}_0 , es el llamado “*Problema de acoplamiento*”⁴ para cadenas de Markov finitas.

Cuando $\hat{P} \notin \mathcal{P}_0$, la situación no es clara, debido a la complicada estructura de \mathcal{P}_0 , pero no parece ser poco común que el estimador máximo verosímil no exista, particularmente cuando el tiempo Δ entre observaciones es muy largo.

Los problemas antes mencionados han sido ampliamente estudiados por Bladt y Sorensen en [4], y resumen en el siguiente teorema, los principales resultados sobre la existencia y unicidad del estimador máximo verosímil, en particular, la probabilidad de que $\hat{P} \in \mathcal{P}_0$ tiende a 1 cuando $n \rightarrow \infty$.

Teorema 3.4. Si \hat{P} dada por la ecuación (3.4.10) pertenece a \mathcal{P}_0 , entonces, el estimador máximo verosímil de la matriz de intensidad \hat{Q} existe, y es la solución a $\hat{P} = e^{\Delta \hat{Q}}$. Si $\hat{P} \notin \mathcal{P}_0$, entonces o el estimador máximo verosímil existe y satisface la condición $e^{\Delta \hat{Q}} \in \delta \mathcal{P}_0$, o la función de verosimilitud dada

⁴“Embedding Problem”

3.4. INFERENCIA ESTADÍSTICA PARA PROCESOS DE SALTOS DE MARKOV.47

por (3.4.7), no tiene máximo en \mathcal{Q} . Si la verdadera matriz de transición Q_0 satisface la condición $e^{\Delta Q_0} \in \text{int}(\mathcal{P}_0)$ y si el proceso de Markov es ergódico, entonces la probabilidad de que el estimador máximo verosímil exista tiende a 1 cuando $n \rightarrow \infty$, y $e^{\Delta \hat{Q}} \rightarrow e^{\Delta Q_0}$ casi seguramente. Además si Q_0 satisface la condición $e^{\Delta Q_0} \in \text{int}(\mathcal{P}_{00})$, entonces la probabilidad de que el estimador máximo verosímil sea único tiende a 1 y $\hat{Q} \rightarrow Q_0$ casi seguramente cuando $n \rightarrow \infty$. La condición $e^{\Delta Q_0} \in \text{int}(\mathcal{P}_{00})$ se satisface cuando Δ es suficientemente pequeña.

Donde $\delta\mathcal{P}_0$ denota la frontera de \mathcal{P}_0 relativa a \mathcal{P}_+ definido como $\mathcal{P}_+ = \{P \in \mathcal{P} \mid \det(P) > 0\}$, entonces: $\delta\mathcal{P}_0 = (\bigcup_{i \neq j} E_{ij}) \cup \mathcal{E}$ donde E_{ij} es el subconjunto no vacío del conjunto de las matrices de intensidad exponenciales con $q_{ij} = 0$, y \mathcal{E} es un subconjunto no vacío de las matrices de transición de $m \times m$, con menos de m distintos eigenvalores, y \mathcal{P}_{00} denota el subconjunto de \mathcal{P}_0 de las matrices de transición $P \in \mathcal{P}_0$ para el cual Q está únicamente determinada por $P = \exp(Q)$. Para detalles del teorema y de la demostración revisar [4].

Existen varios métodos para el cálculo del estimador máximo verosímil, lo que nos interesa es seleccionar un método que nos facilite el cálculo de la verosimilitud y que sea eficiente respecto a las dificultades antes mencionadas. En la siguiente sección se enlistan algunos de estos métodos.

3.4.3. Métodos para el cálculo del estimador.

En la última década se han trabajado varias maneras de realizar estimaciones con información incompleta, entre las cuales se encuentran: el *Método Resolvente* (2001), el *Método de Optimización Cuadrática* (2006), y el *Método de Máxima Verosimilitud* (2005). A continuación se describe brevemente en que consiste cada uno de estos métodos, pero se pueden encontrar detalles específicos en [17].

Método Resolvente.

Consiste en aproximar un parámetro de la matriz Q llamado el *resolvente de Q* , $R(\alpha) = (\alpha I - Q)^{-1}$ $\alpha > 0$, escribiéndolo en términos de la matriz de transición P_t , esto es:

$$R(\alpha) = \int_0^{\infty} \exp(-\alpha t) P_t dt \quad \alpha > 0 \quad (3.4.11)$$

y usando la expresión (3.4.11) estimar el generador infinitesimal subyacente mediante la identidad $Q = \alpha I - R^{-1}(\alpha)$. Donde el valor óptimo de α se puede

aproximar mediante máxima verosimilitud.

Método de Optimización Cuadrática.

Este método consiste, en realizar la eigendescomposición de la matriz de transición $P = UAU^{-1}$ donde A es la matriz diagonal que contiene los eigenvalores $\lambda_1, \lambda_2, \dots, \lambda_m$, de Q , y U contiene los eigenvectores de P .

Con la ecuación anterior modificada haciendo $z_k = \frac{\lambda_k}{t}$ los elementos de la matriz diagonal Z se tiene $Q = UZU^{-1}$ y entonces, se puede expresar en términos de la matriz exponencial, una aproximación de la matriz de transición: $e^{tQ} = e^{U \log(A) U^{-1}} = UAU^{-1} = P$

Posteriormente se calcula el generador infinitesimal estimado \hat{Q} resolviendo un problema de minimización cuadrática.

Los dos métodos anteriores han sido trabajados con más detalle en [16].

Método de Máxima Verosimilitud.

En general, la estimación por máxima verosimilitud, es un método para ajustar un modelo y encontrar sus parámetros. Si se tiene una muestra de observaciones independientes, extraídas de una función de distribución desconocida pero con una función de probabilidad o densidad $f_m(\cdot)$, la cual pertenece a una familia de distribuciones que dependen de un parámetro θ , $\{f(\cdot|\theta) : \theta \in \Theta\}$ y que llamaremos a esta familia modelo paramétrico, de tal modo que la densidad de la muestra f_m , corresponde al parámetro $\theta = \theta_m$.

El objetivo principal de la estimación máximo verosímil es encontrar el estimador o valor $\hat{\theta}$, que se aproxime de manera más exacta al valor verdadero de θ_m . Dicha estimación se realiza encontrando la función de densidad conjunta de todas las observaciones de tal manera que los valores observados sean fijos pero el parámetro varíe libremente, es decir buscar la función de verosimilitud, para posteriormente buscar el valor de θ que la maximice.

En el caso de procesos de saltos de Markov, la idea de este método es encontrar el generador infinitesimal \hat{Q} tal que maximice la función de verosimilitud de las observaciones dadas.

Lo anterior puede ser calculado mediante diferentes técnicas, por ejemplo, las Cadenas de Markov de Montecarlo; cuyo procedimiento consiste en generar cadenas de Markov estacionarias tales que las transiciones sean entre los posibles valores de los parámetros de interés, con lo que se obtiene su distribución, a partir de la cual se obtendrán estadísticos informativos sobre los actuales.

Otra técnica utilizada para realizar estos cálculos es el Algoritmo Esperanza-Maximización, que es una técnica para calcular el estimador máximo verosímil de los parámetros de una distribución subyacente de cierta información dada

3.4. INFERENCIA ESTADÍSTICA PARA PROCESOS DE SALTOS DE MARKOV.49

previamente, la cual se encuentra incompleta o con algunos valores perdidos.

Podemos ver el caso de la verosimilitud en el caso discreto, como un problema de información incompleta, pues conocemos cierta información, pero el resto de ella se encuentra escondida, es por eso resulta razonable utilizar el método del Algoritmo Esperanza-Maximización para estimar el generador infinitesimal, el cual se desarrollará a detalle en el siguiente capítulo.

Capítulo 4

Algoritmo

Esperanza-Maximización.

Como se comentó en el capítulo anterior, el algoritmo Esperanza-Maximización, en adelante algoritmo EM, es una de las maneras que se utilizan en estadística para realizar el cálculo del estimador máximo verosímil, cuando la información con la que se cuenta es incompleta.

En cualquier tipo de análisis estadístico se desea hacer inferencias válidas sobre una población de interés, donde muchas veces nos encontramos con información faltante dentro de las series de datos obtenidas, lo cual implica inconvenientes como la pérdida de eficiencia, complicación de análisis de datos faltantes y estimadores sesgados que ponen en duda la validez del proceso.

El algoritmo EM es un algoritmo iterativo que nos permite estimar parámetros mediante máxima verosimilitud, cuando algunas de las variables aleatorias involucradas no han sido observadas, es decir cuando existe información faltante o incompleta.

Este capítulo está dirigido a la comprensión del algoritmo EM; por lo que a continuación mencionaremos antecedentes del algoritmo, para posteriormente dar una descripción y ejemplificarlo. Para finalizar se explica su aplicación al proceso de saltos de Markov y la matriz de intensidad Q .

4.1. Antecedentes.

En 1933 Yates propuso, que si los datos faltantes fueran reemplazados por sus estimadores mínimos cuadrados aplicados a datos completos, se produciría un estimador de mínimos cuadrados correcto.

Bartlett en 1937, propuso en cambio, un método de estimación de datos faltantes basado en indicadores faltantes, tomados a través de covariables, que tienen la ventaja de obtener estimadores y errores estándares correctos.

Posteriormente, varios autores, entre ellos Healy & Wesmacott en 1965, proponen la estimación iterativa de información faltante, teoría profundizada más adelante por Dempster (1977), con el algoritmo EM[2].

El algoritmo EM se ha convertido en una herramienta muy popular para la estimación estadística que involucran información incompleta (por ejemplo, la información histórica de alguna empresa observada únicamente cada cierta temporalidad (anual, semestral, etc.), los datos escondidos serían los faltantes en el tiempo no observado).

El algoritmo EM, se usa en muchos campos de investigación científica para tratar muestras de todo tipo, en medicina para estimar frecuencias de genotipo, y algunas de sus variantes se han utilizado en métodos de restauración de múltiples fotogramas de superresolución que combinan la estimación de movimiento a lo largo de líneas en física, etc.

4.2. Descripción del Algoritmo EM.

El algoritmo EM es un procedimiento iterativo eficiente para calcular el estimador máximo verosímil, cuando hay ausencia de información en los datos. La idea general de los pasos a seguir en el algoritmo es:

1. reemplazar valores faltantes por valores estimados,
2. estimar parámetros.
3. repetir, del siguiente modo: el paso 1, utilizando los valores de los parámetros estimados como valores reales y; el paso 2, utilizando valores estimados como “valores observados”, iterando hasta la convergencia.

La idea anterior fue utilizada en el año 1972 por Orchard y Woodbury, en su documento *Principios sobre información faltante*, donde se proporcionó la base

teórica de la idea fundamental. El término Esperanza-Maximización la introdujo Dempster y Rubin en 1977, donde se dió primero, una prueba de los resultados generales sobre el comportamiento, así como un gran número de aplicaciones, ver [2].

De modo general, supongamos que se tiene un vector aleatorio y cuya densidad conjunta $f(y; \theta)$ está indexada por el parámetro p -dimensional $\theta \in \Theta \subseteq \mathbb{R}^p$.

Si el vector de información completa y fuera observado, sería interesante calcular el estimador máximo verosímil de θ , basado en la distribución de y . Es decir la log-verosimilitud de y $\log L(\theta; y) = \ell(\theta; y) = \log f(y; \theta)$ debe ser maximizada. En el caso de la información faltante, sólo se observa una función del vector de datos y .

Por lo tanto denotaremos a y como (y_o, y_m) donde y_o es la parte observada pero “incompleta” y y_m es la parte no observada o faltante; se asume que la información faltante (y_m) , es faltante de manera aleatoria, entonces:

$$\begin{aligned} f(y; \theta) &= f(y_o, y_m; \theta) \\ &= f_1(y_o; \theta) \cdot f_2(y_m | y_o; \theta), \end{aligned}$$

donde f_1 es la distribución conjunta de y_o y f_2 es la distribución conjunta de y_m dada la información observada, respectivamente. Entonces la log-verosimilitud de la información observada $\ell_o(\theta, y_o)$ es:

$$\ell_o(\theta, y_o) = \ell(\theta; y) - \log(f_2(y_m | y_o; \theta)) \quad (4.2.1)$$

El algoritmo EM es útil cuando es difícil maximizar $\ell_o(\theta, y_o)$, pero maximizar la log-verosimilitud de la información completa $\ell(\theta, y)$ es simple. El algoritmo EM intenta maximizar $\ell(\theta; y)$ iterativamente, reemplazándolo por la esperanza condicional dada la información observada y_o .

Esta esperanza se calcula respecto a la distribución de la información completa evaluada en una estimación actual de θ . Más específicamente, si $\theta^{(0)}$ es un valor inicial para θ , entonces para la primera iteración se requiere calcular:

$$Q(\theta; \theta^{(0)}) = \mathbb{E}_{\theta^{(0)}}[\ell(\theta; y) | y_o].$$

La función $Q(\theta; \theta^{(0)})$ se maximiza ahora respecto a θ , esto quiere decir que $\theta^{(1)}$ es hallada tal que:

$$Q(\theta^{(1)}; \theta^{(0)}) \geq Q(\theta; \theta^{(0)}) \quad \forall \theta \in \Theta$$

Con esto podemos definir el algoritmo EM con los pasos E y M del siguiente modo:

Algoritmo 2. Algoritmo EM.

Paso E: Calcular $Q(\theta; \theta^{(n)})$ donde

$$Q(\theta; \theta^{(n)}) = \mathbb{E}_{\theta^{(n)}}[\ell(\theta; y)|y_o],$$

Paso M: Encontrar $\theta^{(n+1)} \in \Theta$ tal que

$$Q(\theta^{(n+1)}; \theta^{(n)}) \geq Q(\theta; \theta^{(n)}) \quad \forall \theta \in \Theta.$$

Los pasos anteriores se repiten alternadamente hasta que la diferencia $L(\theta^{(n+1)}) - L(\theta^{(n)}) < \delta$ para una cierta $\delta > 0$ suficientemente pequeña, como se explica en [8].

Cuando la distribución del vector de información completa y pertenece a la familia exponencial, el paso E se reduce a calcular la esperanza del estadístico suficiente de la información completa dada la información observada. También el paso M se simplifica, pues implica maximizar la esperanza de la log-verosimilitud calculada en el paso E.

Particularmente en el caso de la familia exponencial, maximizar para obtener la siguiente iteración, puede reemplazarse por sustituir directamente las esperanzas condicionales de los estadísticos suficientes calculados en el paso E, por los estadísticos suficientes resultantes de las expresiones obtenidas para los estimadores máximo verosímiles de la información completa de θ , como veremos a continuación.

Consideraremos que una función es de la familia exponencial si:

$$f(y_o|\theta) = \frac{b(y) \exp(\theta t(y_o)^T)}{a(\theta)} \quad (4.2.2)$$

donde $t(y_o)^T$ es un vector transpuesto de estadísticas suficientes.

Dado que el objetivo es encontrar un valor θ^* que maximice la ecuación (4.2.1), maximizaremos el logaritmo de la función $f(y_o|\theta)$ del siguiente modo,

$$\log(y_o|\theta) = -\log a(\theta) + \log b(y_o) + \theta t(y_o)^T,$$

esto equivale a maximizar sólo $-\log a(\theta) + \theta t(y_o)^T$, entonces para la familia exponencial nuestra función f_2 debe cumplir

$$f_2 = \frac{b(y_m) \exp(\theta t(y_m)^T)}{a(\theta|y_o)}$$

y dado que $a(\theta)$ y $a(\theta|y_o)$ son constantes de parametrización se tiene

$$a(\theta|y_o) = \int_y b(y_m) \exp(\theta t(y_m)^T) dy_m$$

y

$$a(\theta) = \int b(y_m) \exp(\theta t(x)^T) dy_m$$

por lo que se observa que la función f_2 , y $f(y_o|\theta)$ son familias con el mismo parámetro θ y la misma estadística suficiente $t(y_m)$, aunque definidas sobre diferentes espacios muestrales, por lo que de la ecuación (4.2.1) se tiene,

$$\begin{aligned} \ell_o(\theta, y_o) &= \log f(y_m|\theta) - f_2 \\ &= -\log a(\theta) + \log a(\theta|y_o) \end{aligned}$$

Derivando la ecuación anterior se obtiene que:

$$\begin{aligned} \frac{\partial \log a(\theta)}{\partial \theta} &= \int t(y_m)^T f(y_m|\theta) dy_m \\ &= \mathbb{E}[t|\theta] \end{aligned}$$

y

$$\begin{aligned} \frac{\partial \log a(\theta|y_o)}{\partial \theta|y_o} &= \int t(y_m)^T f_2 dy_m \\ &= \mathbb{E}[t|y_o, \theta] \end{aligned}$$

Igualando la derivada a cero se tiene que $\mathbb{E}[t|\theta] = \mathbb{E}[t|y_o, \theta]$ y por lo tanto $\theta^{(n)} = \theta^{(n+1)} = \theta^*$. Esto demuestra que existe convergencia para las familias exponenciales. Una exposición detallada se puede encontrar en [8].

Existen tres situaciones en los casos de información faltante, que han sido estudiadas ampliamente, dichos casos son: el modelo multinomial, el modelo normal lineal y el modelo normal multivariado.

En los primeros dos casos, los estadísticos suficientes para el problema de

información completa son lineales, eso implica que el paso de la estimación en el algoritmo EM equivale a primero estimar o “rellenar” los puntos individuales de la información y entonces calcular las estadísticas suficientes usando esos valores rellenados. En el tercer caso, el “rellenado” de forma directa no es apropiado, debido a que algunas estadísticas suficientes en los valores de la información son cuadráticas.

Para tener más claros los conceptos anteriormente mencionados, a continuación se presentan ejemplos numéricos de los casos descritos arriba.

Ejemplo 5. Muestreo de una Multinomial.

Sea $y = (y_1, y_2, y_3, y_4)^T$ con probabilidades multinomiales: $(\frac{1}{2} - \frac{1}{2}\theta, \frac{1}{4}\theta, \frac{1}{4}\theta, \frac{1}{2}) \equiv (p_1, p_2, p_3, p_4)$, y es considerado el vector de información completa. Definiremos entonces a nuestro vector de datos observados como $y_o = (y_1, y_2, y_3 + y_4)^T$, la cual es una función de nuestro vector de información completa. De este vector tenemos que, como sólo $y_3 + y_4$ ha sido observado pero y_3 y y_4 no, la información observada se considera incompleta. En este ejemplo, la información observada puede ser vista como una función del espacio de la información completa. Asignando valores numéricos $y_o = (38, 34, 125)^T$ es el conteo observado con probabilidades: $(\frac{1}{2} - \frac{1}{2}\theta, \frac{1}{4}\theta, \frac{1}{2} + \frac{1}{4}\theta)$. El objetivo es obtener el estimador máximo verosímil de θ .

La log-verosimilitud de la información completa es:

$$\ell(\theta, y) = y_1 \log p_1 + y_2 \log p_2 + y_3 \log p_3 + y_4 \log p_4 + C$$

La cual es lineal en y_1, y_2, y_3 y y_4 que son también estadísticas suficientes. Entonces el paso E requiere calcular $\mathbb{E}_\theta(y|y_o)$, esto es:

$$\begin{aligned} \mathbb{E}_\theta(y_1|y_o) &= y_1 &= 38 \\ \mathbb{E}_\theta(y_2|y_o) &= y_2 &= 34 \\ \mathbb{E}_\theta(y_3|y_o) &= \mathbb{E}_\theta(y_3|y_3 + y_4) &= 125(\frac{1}{4}\theta)/(\frac{1}{2} + \frac{1}{4}\theta). \end{aligned}$$

Del último resultado podemos observar que, condicionado a $y_3 + y_4$, y_3 se distribuye Binomial(125, p) con $p = \frac{\frac{1}{4}\theta}{\frac{1}{2} + \frac{1}{4}\theta}$. De manera similar:

$$\mathbb{E}_\theta(y_4|y_o) = \mathbb{E}_\theta(y_4|y_3 + y_4) = 125(\frac{1}{2})/(\frac{1}{2} + \frac{1}{4}\theta)$$

Que se calcula de manera parecida a $\mathbb{E}_\theta(y_3|y_o)$ pero tenemos que sólo

$$y_3^{(n)} = \mathbb{E}_{\theta^{(n)}}(y_3|y_o) = \frac{125(\frac{1}{4})\theta^{(n)}}{(\frac{1}{2} + \frac{1}{4}\theta^{(n)})} \quad (4.2.3)$$

necesita calcularse en la t-ésima iteración del paso-E.

Para el paso-M, tenemos que maximizar:

$$\ell(\theta, y) = y_1 \log\left(\frac{1}{2} - \frac{1}{2}\theta\right) + y_2 \log \frac{1}{4}\theta + y_3 \log \frac{1}{4}\theta + y_4 \log \frac{1}{2},$$

de donde resulta que el estimador máximo verosímil de θ es: $\frac{y_2+y_3}{y_1+y_2+y_3}$.

Así podemos sustituir las esperanzas del paso-E por las estadísticas suficientes de la expresión obtenida para el estimador máximo verosímil de θ antes mencionado y obtener así:

$$\theta^{(n+1)} = \frac{34 + y_3^{(n)}}{72 + y_3^{(n)}} \quad (4.2.4)$$

Las iteraciones de las ecuaciones (4.1.1) y (4.1.2) definen los pasos del algoritmo EM para este problema. La siguiente tabla muestra la convergencia obtenida después de implementar el algoritmo EM al problema comenzando con $\theta^{(0)} = 0.50$.

n	$\theta^{(n)}$	$\theta^{(n)} - \hat{\theta}$	$(\theta^{(n+1)} - \hat{\theta})/(\theta^{(n)} - \hat{\theta})$
0	0.50000000	0.108247423	0.14848
1	0.60824742	0.016073628	0.13487
2	0.62432105	0.002167829	0.13306
3	0.62648888	0.000288443	0.13282
4	0.62677732	0.000038309	0.13278
5	0.62681563	0.000005086	0.13278
6	0.62682072	0.000000675	0.13278
7	0.62682139	0.000000089	⋮
8	0.62682148	0.000000001	⋮

Ejemplo 6. Muestreo normal univariado.

Sea $y = (y_1, \dots, y_n)^T$ el vector de información completa de una muestra aleatoria de una distribución normal $N(\mu, \sigma^2)$. Entonces

$$\begin{aligned}
f(y; \mu, \sigma^2) &= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma^2}\right\} \\
&= -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n y_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n y_i - \frac{n\mu^2}{\sigma^2} + C
\end{aligned}$$

La logverosimilitud basada en la información completa es lineal para los estadísticos suficientes de la información completa. Supongamos que $y_i, i = 1, \dots, m$ es observado y $y_i, i = m + 1, \dots, n$ es la información faltante aleatoriamente y donde se asume que las y_i son independientes e idénticamente distribuidas $N(\mu, \sigma^2)$. Denotamos al vector de información observada por $y_o = (y_1, \dots, y_m)^T$. Como el vector y de información completa es de la familia exponencial, el paso E consistirá en el cálculo de

$$E_{\theta}\left(\sum_{i=1}^n y_i | y_o\right) \text{ y } E_{\theta}\left(\sum_{i=1}^n y_i^2 | y_o\right),$$

en vez de calcular la esperanza de la logverosimilitud de la información completa. Así pues, en la t -ésima iteración del paso E, se calcula

$$\begin{aligned}
s_1^{(t)} &= E_{\mu^{(t)}, \sigma^{2(t)}}\left(\sum_{i=1}^n y_i | y_o\right) \\
&= \sum_{i=1}^m y_i + (n - m)\mu^{(t)}
\end{aligned} \tag{4.2.5}$$

como $E_{\mu^{(t)}, \sigma^{2(t)}}(y_i) = \mu^{(t)}$ donde $\mu^{(t)}$ y $\sigma^{2(t)}$ son las estimaciones actuales de μ y σ^2 , y

$$\begin{aligned}
s_2^{(t)} &= E_{\mu^{(t)}, \sigma^{2(t)}}\left(\sum_{i=1}^n y_i^2 | y_o\right) \\
&= \sum_{i=1}^m y_i^2 + (n - m)[\sigma^{(t)^2} + \mu^{(t)^2}]
\end{aligned} \tag{4.2.6}$$

pues $E_{\mu^{(t)}, \sigma^{2(t)}}(y_i^2) = \sigma^{(t)^2} + \mu^{(t)^2}$.

Para el paso M, notemos primero que los estimadores máximo verosímiles de la información completa para μ y σ^2 son:

$$\hat{\mu} = \frac{\sum_{i=1}^n y_i}{n} \text{ y } \hat{\sigma}^2 = \frac{\sum_{i=1}^n y_i^2}{n} - \left(\frac{\sum_{i=1}^n y_i}{n}\right)^2.$$

El paso M, está definido mediante la sustitución de las esperanzas calculadas en el paso E por las estadísticas suficientes de la información completa en el lado derecho de las expresiones anteriores y de este modo obtener expresiones para las nuevas iteraciones de μ y σ^2 . Es importante notar que las estadísticas suficientes de la información completa no pueden ser calculadas directamente pues y_{m+1}, \dots, y_n no han sido observadas.

Así pues obtenemos las siguientes expresiones:

$$\mu^{(t+1)} = \frac{s_1^{(t)}}{n} \quad (4.2.7)$$

y

$$\sigma^{2(t+1)} = \frac{s_2^{(t)}}{n} - \mu^{(t+1)^2}. \quad (4.2.8)$$

De este modo, el paso E para este ejemplo implica evaluar las ecuaciones (4.2.5) y (4.2.6) comenzando con los valores iniciales $\mu^{(0)}$ y $\sigma^{2(0)}$. Mientras que el paso M implica sustituir los valores del paso E en las ecuaciones (4.2.7) y (4.2.8) para poder calcular los nuevos valores de $\mu^{(1)}$ y $\sigma^{2(1)}$ y así sucesivamente.

La mayor ventaja de este algoritmo es que, es numéricamente estable, esto debido a que cada iteración es necesaria para aumentar la log-verosimilitud $\ell(\theta; y_o)$, y si es acotada, entonces la sucesión $\ell(\theta^{(t)}; y_o)$ converge a un valor estacionario. Si la sucesión $\theta^{(t)}$ converge, lo hace hacia un máximo local o punto silla de $\ell(\theta; y_o)$ y por lo tanto al único estimador máximo verosímil de $\ell(\theta; y_o)$.

Lo anterior queda demostrado con el siguiente Teorema:

Teorema 4.1. Supongamos que $\theta^{(t)}$, para $t = 0, 1, 2, \dots$ ha sido obtenido mediante el algoritmo EM, y es tal que:

1. $\ell(\theta^{(t)})$ es acotada, y
2. $Q(\theta^{(n+1)}; \theta^{(n)}) - Q(\theta^{(n)}; \theta^{(n)}) \geq \lambda(\theta^{(n+1)} - \theta^{(n)})(\theta^{(n+1)} - \theta^{(n)})^T$, para algún escalar $\lambda > 0$ y $\forall n$

Entonces, la sucesión $\theta^{(n)}$ converge a algún $\hat{\theta}$ en la cerradura de Θ .

Demostración. Ver [8].

Por otro lado, una desventaja del algoritmo EM, es que la tasa de convergencia puede ser demasiado lenta si el número de datos faltantes es muy grande; Dempster, Laird y Rubin (1977) demostraron que la convergencia es lineal con tasa proporcional a la fracción de la información sobre θ en $\ell(\theta; y)$ que es observada.

4.3. Aplicación del Algoritmo EM a Procesos de Saltos de Markov.

Como hemos visto en el capítulo anterior, el estimador máximo verosímil en el caso continuo, puede ser obtenido analíticamente, pero no así para el caso discreto, eso implica que la información entre dos observaciones discretas consecutivas se encuentra escondida, y por lo tanto $R_i(\tau)$ y $N_{ij}(\tau)$ son desconocidos.

El algoritmo EM, es un buena opción para aproximar iterativamente el máximo local de la verosimilitud discreta $L_n^{(d)}$ de la ecuación (3.4.7), calculando los valores esperados de $R_i(\tau)$ y $N_{ij}(\tau)$ dada la información observada, y un generador inicial supuesto Q_0 . El desarrollo de lo anterior lo trabajaron Bladt y Sorensen en [4], y es la idea principal con la que trabajaremos a partir de este momento.

Consideraremos el proceso de saltos de Markov $X = \{X_t\}_{0 \leq t \leq \tau}$, pero sólo conocemos la información parcial sobre el proceso $Y = \{Y_i\}_{i=1, \dots, n}$ donde $Y_i = X_{t_i}$ y para el cual $t_i = 0$ y $t_n = \tau$, es decir el proceso Y_i , son las observaciones discretas (cadena de Markov) y la información disponible con la que contamos.

Para este caso en particular, el paso E consistirá en calcular la función de log-verosimilitud condicional, la cual puede representarse, para la función (3.4.3) del siguiente modo:

$$\mathbb{E}_{Q_0}[\log L_\tau^{(c)}(Q)|Y = y] \quad (4.3.1)$$

donde $Y = \{Y_i | i = 1, \dots, n\}$ y Q_0 es una matriz de intensidad dada. Por lo tanto la expresión anterior puede verse como:

$$\mathbb{E}_{Q_0}[\log L_\tau^{(c)}(Q)|Y = y] = \sum_{i=1}^m \sum_{j \neq i} [\log(q_{ij}) \mathbb{E}_{Q_0}\{N_{ij}(\tau)|Y = y\} - q_{ij} \mathbb{E}_{Q_0}\{R_i(\tau)|Y = y\}] \quad (4.3.2)$$

La log-verosimilitud de tiempo continuo dada por (4.3.2) tiene estadísticas $\mathbb{E}_{Q_0}[N_{ij}(\tau)|Y]$ y $\mathbb{E}_{Q_0}[R_i(\tau)|Y]$, y es maximizada como función de Q mediante (3.4.5) por:

$$\hat{q}_{ij} = \begin{cases} \frac{\mathbb{E}_{Q_0}[N_{ij}(\tau)|Y]}{\mathbb{E}_{Q_0}[R_i(\tau)|Y]} & i \neq j \\ -\sum_{i \neq j} \hat{q}_{ij} & i = j \end{cases},$$

esta parte es el paso M de nuestro algoritmo.

La tarea no trivial del algoritmo, implica evaluar las esperanzas condicionales $\mathbb{E}_{Q_0}[N_{ij}(\tau)|Y]$ y $\mathbb{E}_{Q_0}[R_i(\tau)|Y]$ respectivamente.

Debido a la propiedad de Markov y a la homogeneidad del proceso de saltos de Markov, durante un tiempo constante τ , será suficiente encontrar:

$$\hat{M}_{ij}^k(t) = \mathbb{E}_{Q_0}[R_k(t)|X(t) = j, X(0) = i] \quad (4.3.3)$$

y

$$\hat{f}_{ij}^{kl}(t) = \mathbb{E}_{Q_0}[N_{kl}(t)|X(t) = j, X(0) = i] \quad (4.3.4)$$

esto debido a que

$$\mathbb{E}_{Q_0}[N_{ij}(\tau)|Y = y] = \sum_{k=1}^{n-1} \hat{f}_{y_k, y_{k+1}}^{ij}(t_{k+1} - t_k), \quad (4.3.5)$$

$$\mathbb{E}_{Q_0}[R_l(\tau)|Y = y] = \sum_{k=1}^{n-1} \hat{M}_{y_k, y_{k+1}}^l(t_{k+1} - t_k). \quad (4.3.6)$$

Para calcular (4.3.3) es conveniente estudiar la relación (se omitirá el subíndice Q_0 a partir de este momento):

$$M_{ij}^k(t) = \mathbb{E}[R_k(t)I\{X(t) = j\}|X(0) = i].$$

Los siguientes resultados y detalles de los mismos pueden ser encontrados en

[6].

En adelante asumiremos que δ_{ij} es igual a 1 si y sólo si $i = j$ y 0 en otro caso, es decir la delta de Kronecker.

Teorema 4.2. La función M_{ij}^k resuelve la ecuación diferencial:

$$\frac{d}{dt}M_{ij}^k(t) = \sum_l M_{il}^k(t)q_{lj} + \exp(tQ)ij\delta_{jk} \quad (4.3.7)$$

con condición inicial $M_{ij}^k(0) = 0$.

Definimos $M_{i\cdot}^k(t) = (M_{il}^k(t), \dots, M_{in}^k(t))$ (vector renglón). Entonces (4.3.7) se puede escribir como:

$$\frac{d}{dt}M_{i\cdot}^k(t) = M_{i\cdot}^k(t)Q + A_i^k(t),$$

donde $A_i^k(t) = e_i' \exp(Qt) e_k e_k'$ con e_i que denota el vector unitario cuya coordenada i -ésima es igual a 1 y e_i' es su vector transpuesto. Este es un sistema de ecuaciones diferenciales lineales no homogéneas con condición inicial $M_{i\cdot}^k(0) = 0$, y que puede resolverse mediante métodos numéricos, por ejemplo el método Runge-Kutta de cuarto orden.

Alternativamente, notamos que la solución al sistema de ecuaciones diferenciales está dado por:

$$\begin{aligned} M_{i\cdot}^k(t) &= \int_0^t A_i^k(s) \exp\{(t-s)Q\} ds \\ &= e_i' \int_0^t \exp(sQ) (e_k e_k') \exp\{(t-s)Q\} ds, \end{aligned}$$

el cual puede ser evaluado numéricamente mediante la adecuada expansión de matrices exponenciales utilizando el método de uniformización. Específicamente, elegir $\lambda \geq \max_{i=1, \dots, m} (-Q_{ii})$ y definir

$$B = I + \frac{1}{\lambda}Q = \frac{1}{\lambda}(\lambda I + Q).$$

Entonces $M^k = \{M_{ij}^k\}_{ij \in E}$ esta dada por:

$$M^k(t) = \exp(-\lambda t) \lambda^{-1} \sum_{n=0}^{\infty} \frac{(\lambda t)^{n+1}}{(n+1)!} \sum_{l=0}^n B^l (e_k e_k') B^{n-l}.$$

Ambos métodos son eficientes para dimensiones menores que 30, para dimen-

4.3. APLICACIÓN DEL ALGORITMO EN A PROCESOS DE SALTOS DE MARKOV. 63

siones mayores, el método de uniformización es más efectivo.

Ahora podemos resolver la ecuación (4.3.7) mediante:

$$\hat{M}_{ij}(t) = \frac{M_{ij}^k(t)}{e_i' \exp(Qt) e_j}. \quad (4.3.8)$$

Para calcular la ecuación (4.3.4), que es el número esperado de transiciones del estado k al estado l en un intervalo de longitud t dado que el proceso comienza en el estado i y finaliza en el estado j , primero debemos considerar

$$f_{ij}^{kl}(t) = \mathbb{E}[N_{kl}(t) I\{X(t) = j\} | X(0) = i] \quad (4.3.9)$$

para k y l fijos.

Teorema 4.3. La función f_{ij}^{kl} resuelve la ecuación diferencial

$$\frac{d}{dt} f_{ij}^{kl}(t) = \sum_{h=1}^m f_{ih}^{kl}(t) q_{hj} + q_{kl} \exp(Qt)_{ik} \delta_{jl}, \quad (4.3.10)$$

con condición de frontera $f_{ij}^{kl}(0) = 0$ para toda i y j .

$f_{ij}^{kl}(t)$ puede resolverse mediante los métodos de cuarto orden de Runge-Kutta o mediante uniformización. Utilizando uniformización se obtiene:

$$f_{ij}^{kl}(t) = q_{kl} \exp(-\lambda t) \lambda^{-1} \sum_{n=0}^{\infty} \frac{(\lambda t)^{n+1}}{(n+1)!} \sum_{j=0}^n B^j (e_k e_l') B^{n-j}, \quad (4.3.11)$$

donde λ y B son como en el caso de $M_{ij}^k(t)$.

Ahora podemos calcular $\hat{f}_{ij}^{kl}(t)$ definida en (4.3.4):

$$\hat{f}_{ij}^{kl}(t) = \frac{f_{ij}^{kl}(t)}{e_i' \exp\{Q(t)\} e_j}. \quad (4.3.12)$$

Se pueden revisar detalles de los teoremas 4.2 y 4.3 en [6].

Con toda la información anterior podemos describir el algoritmo EM para la estimación máximo verosímil de \hat{Q} , como sigue:

Algoritmo 3. Algoritmo EM para procesos de saltos de Markov.

1. Calcular $\hat{M}_{y_i, y_{i+1}}^k(t_{i+1} - t_i)$ y $\hat{f}_{y_i, y_{i+1}}^{kl}(t_{i+1} - t_i)$ para toda k y l bajo el modelo con matriz de intensidad Q con las ecuaciones (4.3.7) y (4.3.10).
2. Calcular $\mathbb{E}_Q[R_i(\tau)|Y = y]$ y $\mathbb{E}_Q[N_{ij}(\tau)|Y = y]$ con las ecuaciones (5.2.1) y (5.2.2).
3. Calcular \hat{Q} con $\hat{Q}_{ij} = \mathbb{E}_Q[N_{ij}|Y = y]/\mathbb{E}_Q[R_i(\tau)|Y = y]$ para toda $i \neq j$.
4. Hacemos $Q := \hat{Q}$ y regresar a 1.

Sea Q_0, Q_1, Q_2, \dots la sucesión de matrices de intensidad obtenidas mediante el Algoritmo EM, entonces $L_n(Q_{n+1}) \geq L_n(Q_n)$ para $n = 0, 1, 2, \dots$ donde L_n es la función de verosimilitud discreta. Como vimos en la sección anterior se converge al máximo de la función de verosimilitud. Sin embargo, como se explica en [4], debemos asegurar que el determinante de $\{\exp(Q)\}$ no converja a cero, para lo cual debemos restringir el espacio de nuestro conjunto de generadores infinitesimales para alguna $\epsilon > 0$ suficientemente pequeña, del siguiente modo

$$\mathcal{Q}_\epsilon = \{Q \in \mathcal{Q} \mid \det\{\exp(Q)\} \geq \epsilon\}.$$

Esta condición asegura que para cualquier generador infinitesimal de este conjunto, los tiempos de estancia esperados y el número de saltos esperados, que calculamos mediante el algoritmo EM son estrictamente positivos para todos los estados posibles, y por lo tanto es posible obtener el estimador máximo verosímil usando estas esperanzas como estadísticas en la función de verosimilitud discreta y por lo tanto en cada iteración el generador infinitesimal resultante pertenece también a este conjunto. Lo anterior se puede resumir en el siguiente teorema descrito en [4].

Teorema 4.4. Supongamos que la matriz inicial Q_0 pertenece al interior del espacio parametral \mathcal{Q} , es decir que, $(Q_0)_{ij} > 0$ para toda $i \neq j$. Entonces la sucesión $\{Q_n\}$ puede ya sea converger a un punto estacionario de la función de verosimilitud L_n , o $\det\{\exp(Q_n)\} \rightarrow 0$.

Es por esta razón que es necesario elegir Q_0 de tal modo que $\det\{\exp(Q_n)\}$, se encuentre lejos de 0.

4.3. APLICACIÓN DEL ALGORITMO EN A PROCESOS DE SALTOS DE MARKOV. 65

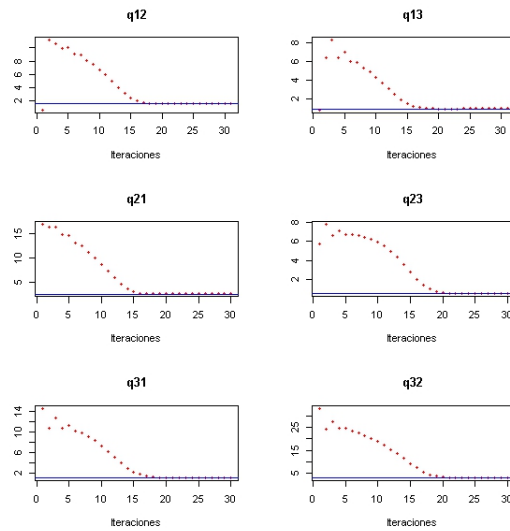
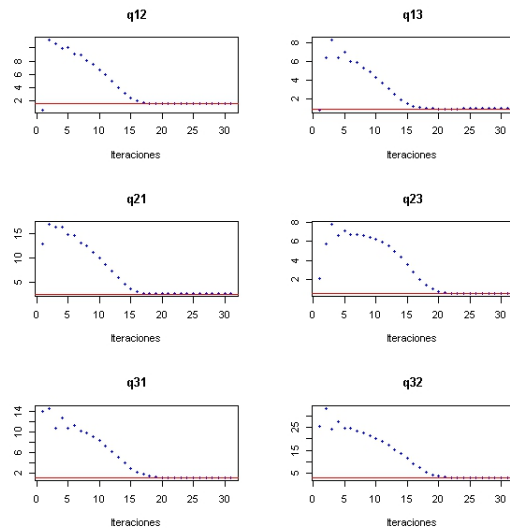
A continuación se presenta un ejemplo aplicado, en el cual, de nuestro ejemplo 9 donde simulamos un proceso de saltos de Markov tomamos observaciones discretas del mismo y consideramos tres tipos de tamaños de los intervalos de nuestras observaciones, esto es tenemos diferentes $\Delta = t_{i+1} - t_i$, en cada caso se estima la matriz de intensidad conocida de dicho ejemplo.

Ejemplo 7. Para este ejemplo estimaremos la matriz de intensidad Q de nuestro ejemplo número 3. La simulación del proceso fue realizada en el programa Maple con un intervalo de tiempo $[0, 100]$ para $\Delta = 1$ con un total de 101 observaciones discretas, $\Delta = 0.1$ con 1001 observaciones discretas y $\Delta = 0.01$ con 10001 observaciones discretas. Se realizaron iteraciones hasta obtener una convergencia con $\delta = 0.000001$.

Transición de estados		Diferentes tamaños de intervalos para las muestras			
De	A	$\Delta = 1$	$\Delta = 0.1$	$\Delta = 0.01$	Real
1	2	1.5568	1.6779	1.5568	1.6
1	3	0.9315	0.9036	0.9315	0.91
2	1	2.5797	2.5057	2.5797	2.5
2	3	0.5660	0.3289	0.5060	0.5
3	1	1.0418	0.8245	1.0418	0.97
3	2	2.9734	2.0403	2.9734	3.02
# de iteraciones		32	21	30	

Cuadro 4.1: Estimación de intensidades.

En las siguientes gráficas se muestra la convergencia de cada intensidad, dependiendo el tamaño de Δ .

Figura 4.3.2: Convergencia para cada intensidad con $\Delta = 0.01$ Figura 4.3.1: Convergencia para cada intensidad con $\Delta = 1$

Capítulo 5

Aplicación a riesgo de crédito.

Como vimos al principio de este trabajo, el estudio de los modelos de riesgo de crédito ha evolucionado dramáticamente, especialmente durante los últimos 30 años. Existen básicamente tres objetivos que se ha buscado estudiar en el desarrollo de los modelos de riesgo de crédito, el primero es trabajar con intervalos de tiempo que nos muestren un comportamiento más aproximado de las transiciones entre calificaciones de crédito de una firma, y dado que el comportamiento del mundo real no es lineal se ha buscado como segundo objetivo pronosticar con mayor precisión los modelos probabilísticos; finalmente se busca tener un sólido modelo teórico que sea base para el estudio especialmente de calificación crediticia para pronosticar ruinas[1].

La metodología markoviana ha sido utilizada para estimar las probabilidades de transición, desde 1992 por Altman y Kao. En 1995 Jarrow y Turnbull desarrollan una nueva metodología para derivados que involucran dos tipos de riesgo de crédito, cuando el activo subyacente puede incumplir y el segundo cuando el tenedor del activo subyacente incumple [12].

Más recientemente, durante los últimos 6 años, se han estudiado modelos de riesgo de crédito mediante procesos de difusión (distribuciones tipo fase), procesos semi-Markovianos, procesos de Markov mezclados y modelos con Procesos Poisson, algunos de los cuales contemplan la estimación de parámetros.

En la actualidad, el análisis de la evolución de información sobre calificaciones crediticias mediante modelos de Markov de tiempo continuo ha resultado ser una mejor herramienta, que con con el modelo a tiempo discreto. Esto es debido a que el modelo en tiempo continuo, permite una estimación significativa de la

probabilidad de transiciones no comunes, incluso cuando dichas transiciones no han sido observadas realmente en el conjunto de información disponible, un ejemplo podría ser la probabilidad de impago de una categoría de calificación alta. Dichos resultados los explican Jarrow, Lando y Turnbull en [11].

Utilizando los resultados de los dos capítulos anteriores, podremos observar como es que el modelo de Markov a tiempo continuo puede utilizarse para analizar observaciones en tiempos discretos, por ejemplo dónde las calificaciones han sido observadas sólo en puntos discretos del tiempo (mensual o anualmente).

La tarea principal, es estimar las intensidades de transición, es decir el generador infinitesimal del proceso de saltos de Markov, con un número finito de estados, que representarán las calificaciones crediticias asignadas. Las intensidades estimadas pueden utilizarse para calcular estimadores de las probabilidades de transición entre cualesquiera dos calificaciones sobre cualquier horizonte de tiempo. En particular, la probabilidad de incumplimiento de cualquier clase de calificación puede ser estimado, como podremos observar enseguida.

En este capítulo estudiaremos cómo los temas que hemos visto hasta el momento, se relacionan entre sí para poder realizar la aplicación a riesgo de crédito.

5.1. Descripción del modelo.

Nuestros supuestos para desarrollar el modelo son los siguientes:

- Se estudiarán N firmas, empresas o bonos que serán representadas mediante $X^k = \{X_t^k\}_{t \geq 0}$, $k = 1, \dots, N$ el conjunto de N procesos de saltos de Markov independientes, que comparten las mismas tasas de transición.
- El conjunto de las calificaciones crediticias: {AAA, AA, A, BBB, BB, B, C,D} representará nuestro espacio de estados finito $E = \{1, \dots, m\}$, con la característica principal de que el último elemento del conjunto (el último estado) será la peor calificación, la de endeudamiento total, y a dicho estado sólo se puede acceder una sola vez y nunca salir, es decir es un estado absorbente.

Nuestros procesos de saltos independientes mencionados anteriormente comparten el mismo espacio de estados y la misma matriz de intensidad $Q = \{q_{ij}\}$. En particular q_{ij} es la tasa de transición del estado i al estado j , y $q_{ii} = -\sum_{j \neq i} q_{ij}$.

Si los X^k han sido observados continuamente en el intervalo de tiempo $[0, \tau]$, la función de verosimilitud esta dada por:

$$L^{(c)}(Q) = \prod_{k=1}^N \prod_{i=1}^m \prod_{j \neq i} q_{ij}^{N_{ij}^k(\tau)} e^{-q_{ij} R_i^k(\tau)} = \prod_{i=1}^m \prod_{j \neq i} q_{ij}^{N_{ij}(\tau)} e^{-q_{ij} R_i(\tau)}, \quad (5.1.1)$$

donde $N_{ij}^k(t)$ es el número de transiciones del estado i al estado j en el intervalo de tiempo $[0, t]$ del proceso X^k , es decir, $N_{ij}^k(t)$, denotarán el número de veces que la empresa k , que se encuentra en la calificación i , salta a la calificación j durante un periodo de tiempo determinado t , que podría ser un año, un mes, una semana, etc.; además

$$R_i^k(t) = \int_0^t I\{X_s^k = i\} ds \quad (5.1.2)$$

es el tiempo de estancia en el estado i antes del tiempo t en el proceso X^k y por lo tanto:

$$N_{ij}(t) = \sum_{k=1}^N N_{ij}^k(t), \quad R_i(t) = \sum_{k=1}^N R_i^k(t).$$

Mientras que $R_i^k(t)$, nos indica cuanto tiempo permaneció la empresa k , en la calificación i durante el periodo de tiempo determinado.

Con la notación anterior, el estimador máximo verosímil de Q , basado en el registro continuo de historiales de calificación crediticia está dado para toda $i \neq j$ por:

$$\hat{q}_{ij}^{(c)}(\tau) = \frac{N_{ij}(\tau)}{R_i(\tau)}, \quad (5.1.3)$$

seguido del hecho $R_i(\tau) > 0$. Evidentemente $\hat{q}_{ii}(\tau) = -\sum_{j \neq i} \hat{q}_{ij}(\tau)$ para $i = 1, \dots, m$.

Las observaciones de calificación crediticia generalmente se encuentran disponibles únicamente en ciertos puntos discretos del tiempo, por ejemplo anualmente. Sean $0 \leq t_1^i < \dots < t_{n_i}^i \leq \tau$ los n_i puntos en el tiempo en los cuales el proceso X_i ha sido observado. El historial observado $Y_s^k = X_{t_s^k}^k$ es la cadena de Markov discreta, para la cual la matriz de transición en el intervalo de tiempo s es $P^{\Delta_s^k}(Q)$, donde $\Delta_s^k = t_{s+1}^k - t_s^k$ y

$$P^t(Q) = \exp(tQ), t > 0. \quad (5.1.4)$$

Es importante notar que el proceso en tiempo discreto será homogéneo cuando los puntos en el tiempo t_s^k sean equidistantes, que es nuestro caso.

Para nuestro propósito, el tiempo continuo de nuestras observaciones, será representado por un “registro continuo”, es decir la transición de las empresas, entre las diferentes calificaciones crediticias son registradas en las fechas en que ocurrieron, eso significa que el registro es completo, en el sentido de que todos los cambios han sido registrados. Por lo tanto un “registro continuo” significará observaciones diarias.

Nos interesa estimar las intensidades dependiendo de la frecuencia con la que se observan las calificaciones crediticias, por lo cual utilizaremos observaciones discretas en el tiempo, éstas a diferencia del registro continuo definido anteriormente, son tiempos observados cada cierto número de días: 1,7,15,30,90,180,365, que van desde información diaria (“registro continuo”) hasta información anual.

En este caso la función de verosimilitud en tiempo discreto esta dada por:

$$L^{(d)}(Q) = \prod_{k=1}^N \prod_{i=1}^{n_k-1} P^{\Delta_i^k}(Q)_{x_i^k x_{i+1}^k}, \quad (5.1.5)$$

donde $x_1^k, \dots, x_{n_k}^k$ denotan los valores observados de X^k , y donde $P^t(Q)_{ij}$ denota la ij -ésima entrada de la matriz de transición $P^t(Q)$.

En el caso especial donde $\Delta_i^k = \Delta$, la función de verosimilitud se escribe como:

$$L^{(d)}(Q) = \prod_{i=1}^m \prod_{j=1}^m P^{\Delta}(Q)_{ij}^{K_{ij}}, \quad (5.1.6)$$

donde K_{ij} denota el número de transiciones del estado i al j observados en todos los procesos $Y_\ell = \{Y_k^\ell\}_{k=1, \dots, n_\ell}$, $\ell = 1, \dots, N$. El estimador máximo verosímil de $P = \exp(\Delta Q)$ está dado entonces por:

$$\hat{P}_{ij} = \frac{K_{ij}}{K_i}, \quad (5.1.7)$$

donde $K_i = \sum_{j=1}^m K_{ij}$.

Específicamente nos interesa estimar la matriz de intensidad basada en las calificaciones crediticias de todas las empresas cuando estuvieron un determinado momento en el tiempo (algún día del año).

5.2. Algoritmo EM.

En esta parte, modificaremos lo visto en la sección 4.2 para el caso de varios procesos de saltos de Markov, de la siguiente manera. Sabemos que la parte principal del algoritmo EM es encontrar:

$$\begin{aligned} \mathbb{E}_{Q_0}(\log L^{(c)}(Q)|Y_1 = y_1, \dots, Y_N = y_N) \\ = \sum_{i=1}^m \sum_{j \neq i} \log(q_{ij}) \mathbb{E}_{Q_0}(N_{ij}(\tau)|Y_1 = y_1, \dots, Y_N = y_N) \\ - \sum_{i=1}^m \sum_{j \neq i} q_{ij} \mathbb{E}_{Q_0}(R_i(\tau)|Y_1 = y_1, \dots, Y_N = y_N), \end{aligned}$$

donde Q y Q_0 son típicamente distintos. Para una matriz de intensidad dada Q , se define:

$$\begin{aligned} \hat{M}_{ij}^k(\Delta, Q) &= \mathbb{E}_Q(R_k^1(\Delta)|X_\Delta^1 = j, X_0^1 = i) \\ \hat{f}_{ij}^{k\ell}(\Delta, Q) &= \mathbb{E}_Q(N_{k\ell}^1(\Delta)|X_\Delta^1 = j, X_0^1 = i). \end{aligned}$$

Como los N procesos X_i con $i = 1, \dots, N$ son independientes e idénticamente distribuidos entonces,

$$\mathbb{E}_Q(N_{ij}(\tau)|Y_1 = y_1, \dots, Y_N = y_N) = \sum_{\ell=1}^N \sum_{k=1}^{n_\ell-1} \hat{f}_{y_k^\ell, y_{k+1}^\ell}^{ij}(t_{k+1}^\ell - t_k^\ell, Q) \quad (5.2.1)$$

$$\mathbb{E}_Q(R_i(\tau)|Y_1 = y_1, \dots, Y_N = y_N) = \sum_{\ell=1}^N \sum_{k=1}^{n_\ell-1} \hat{M}_{y_k^\ell, y_{k+1}^\ell}^i(t_{k+1}^\ell - t_k^\ell, Q) \quad (5.2.2)$$

Las cantidades \hat{M}_{ij}^k y $\hat{f}_{ij}^{k\ell}$, se calculan con los teoremas 4.2 y 4.3 del capítulo anterior. Ahora podemos resumir el algoritmo EM del siguiente modo:

Algoritmo 4. Algoritmo EM aplicado a riesgo de crédito.

Sea Q_0 cualquier matriz de intensidad para un proceso de saltos de Markov con espacio de estados E , inicializamos $Q = Q_0$.

1. Calcular $\hat{M}_{y_k^\ell, y_{k+1}^\ell}^i(t_{k+1}^\ell - t_k^\ell, Q)$ y $\hat{f}_{y_k^\ell, y_{k+1}^\ell}^{ij}(t_{k+1}^\ell - t_k^\ell, Q)$ para toda i, j, k, ℓ .

2. Calcular $\mathbb{E}_Q(R_i(\tau)|Y_1 = y_1, \dots, Y_N = y_N)$ y $\mathbb{E}_Q(N_{ij}(\tau)|Y_1 = y_1, \dots, Y_N = y_N)$ mediante (5.2.1) y (5.2.2).
3. Calcular \hat{Q} con:

$$\hat{Q}_{ij} = \mathbb{E}_Q(N_{ij}(\tau)|Y_1 = y_1, \dots, Y_N = y_N) / \mathbb{E}_Q(R_i(\tau)|Y_1 = y_1, \dots, Y_N = y_N)$$

para toda $i \neq j$ con el resultado (5.1.3).

4. Hacer $Q := \hat{Q}$ y regresar al paso 1.

En la siguiente sección veremos la implementación de este algoritmo con información real para intentar estimar una matriz de transición de probabilidad de calificaciones crediticias.

5.3. Implementación a riesgo de crédito.

Para esta sección utilizamos el Historial completo de Calificación crediticia de *Fitch Ratings*, que puede descargarse de [23], el cual muestra el historial crediticio de 112 países, durante los años 1994 a 2012. Debido a que no todos los países tienen el mismo número de calificaciones, y no en los mismos años, se trabajó la información para obtener un historial más homogéneo con 44 países cada uno de ellos con 8 calificaciones anuales, esto quiere decir que la información de cada país es equidistante.

Para el óptimo manejo de la información, crearemos una situación hipotética en la cual nuestro espacio de estados estará reducido a cuatro elementos:

- El espacio de estados de cada proceso de saltos de Markov es $E = \{A, B, C, D\}$, por lo que la matriz de transición de calificaciones se representará con una matriz de 4×4 . Donde la clasificación A , representa el conjunto de las calificaciones AAA, AA y A de *Fitch Ratings*, la clasificación B , representa el conjunto de las calificaciones BBB, BB y B de *Fitch Ratings*, y las clasificaciones C y D son como las asignadas por *Fitch*.
- $N = 44$ el número de países observados, cada uno de ellos representa un proceso de saltos de Markov y son independientes entre si. Además suponemos que todos comparten el mismo generador infinitesimal Q . Veremos

	A	B	C	D
A	86	10	0	0
B	13	176	6	1
C	0	4	4	2
D	0	3	0	3

Cuadro 5.1: Saltos entre calificaciones crediticias.

a este conjunto de procesos de saltos de Markov independientes como como $X^k = \{X_t^k\}_{t \geq 0}$, $k = 1, \dots, 44$.

Obtenemos entonces un total de 44 procesos con 8 observaciones cada uno, un total de 352 observaciones que representan 308 saltos entre calificaciones del siguiente modo:

Esta matriz está cargada a la diagonal, y nos sirve para calcular $\hat{f}_{y_k^\ell, y_{k+1}^\ell}^{ij}$ y $\hat{M}_{y_k^\ell, y_{k+1}^\ell}^i$, para todas las calificaciones posibles.

Por lo tanto se realizó la estimación de Q obteniendo:

	A	B	C	D
A	-0.022902	0.022902	0	0
B	0.000004	-0.000007	0.000003	0.000001
C	0	18230.42	-18230.42	0.000000
D	0	0.007059	0	-0.007059

Cuadro 5.2: Matriz de intensidad estimada

con la que podemos calcular la matriz de transición de probabilidades P :

	A	B	C	D
A	0.97735841	0.02264158	0.00000000	0.00000001
B	0.00000358	0.99999557	0.00000000	0.00000084
C	0.00000000	0.00000000	0.99999558	0.00000084
D	0.00000001	0.00703420	0.00000000	0.99296579

Cuadro 5.3: Matriz de probabilidades de transición

Como podemos observar en nuestros resultados, las calificaciones con mayor estabilidad son B y C. Lo que quiere decir que las entidades calificadas con estas categorías tienen una mayor probabilidad de finalizar el año con la misma calificación. También podemos observar que si la entidad se encuentra en A,

tiene una probabilidad positiva de terminar el año en calificación B y también si la entidad inicia con calificación B puede terminar con una probabilidad pequeña en A. En nuestro ejemplo no estamos considerando al estado de Default como absorbente aunque también observamos que tiene una probabilidad grande de permanecer en esa calificación. Los modelos reales consideran este estado como absorbente.

Las calificaciones C y D resultan tener menor estabilidad debido a la poca concentración de datos en estas calificaciones. La ventaja que tenemos utilizando el algoritmo EM es que dado que tenemos que $\Delta = t_{k+1} - t_k$ es la misma para toda k , dado que se tomaron observaciones de manera homogénea, el algoritmo es esencialmente insensible al número de puntos observados, mientras sean estos sean representativos. Por lo que aunque parezca que tenemos poca información para realizar la estimación de Q , ésta es suficiente.

Por ejemplo, en el reporte especial realizado por *Fitch Ratings* en 2011 (que puede revisarse en [24]), presentan las matrices de transición por año y se mencionan el número de observaciones por año. Sin embargo debe tenerse cautela al momento de interpretar el análisis de aquellas calificaciones en las que el número de observaciones es limitado.

Capítulo 6

Conclusiones.

La importancia de los procesos estocásticos a tiempo continuo para modelar el comportamiento de situaciones de la vida real a través del tiempo, radica en las propiedades que éstos tienen. En este trabajo la homogeneidad y la independencia entre los procesos de saltos de Markov, nos permitió desarrollar un modelo para estimar su generador infinitesimal, en los casos cuando se cuenta con información completa y cuando sólo se cuenta con observaciones discretas del proceso. En el segundo caso utilizamos el algoritmo EM, que funciona para el cálculo del estimador máximo verosímil, cuando se cuenta con información incompleta.

En el último capítulo, se realizó la implementación de la estimación del generador infinitesimal para poder conocer la matriz de transición de probabilidades de calificación crediticia.

La única desventaja que se encontró al trabajar en la implementación con datos reales, fué que dado que los sistemas de ecuaciones lineales que nos ayudan a calcular \hat{f}_{kl}^{ij} y \hat{M}_{kl}^i aumentan exponencialmente mientras el espacio de estados es más grande, el cálculo de los mismos se vuelve mucho más complicado.

Dado que existen otros métodos para realizar el cálculo del generador infinitesimal, como son las Cadenas de Markov vía Monte Carlo (MCMC por sus siglas en inglés) y aleatorización; resultaría interesante realizar una comparación entre las matrices de intensidad estimadas con cada uno de los métodos .

Parte I
Apéndice

Código en R Project: Simulación de un proceso de saltos de Markov y obtener observaciones discretas del proceso.

```

est<-3 # Número de estados
Q<-matrix(0,est , est )

#Llenamos la matriz Q únicamente con las intradas i dif
  de j
Q[1,2]<-1.6
Q[1,3]<-.91
Q[2,1]<-2.5
Q[2,3]<-.5
Q[3,1]<-.97
Q[3,2]<-3.02
b<-0
c<-array(data=NA, dim=est )
for (i in 1:est){
  for (j in 1:est){
    b<-b+Q[i , j ]
  }
  c [ i ]<-(-b)
  b<-0
}
c
for (i in 1:est){
  for (j in 1:est){
    if (i==j)
      Q[i , j ]<-c [ i ]
    else
      Q[i , j ]<-Q[i , j ]
  }
}
Q

#Creamos la matriz K (matriz de transición de primer paso
  del proceso discreto)
K<-matrix(0,est , est )

```

```

for (i in 1:est){
  for (j in 1:est){
    if(i==j)
      K[i,j]<0
    else
      K[i,j]<-(-Q[i,j]/Q[i,i])
  }
}

```

K

#####Algoritmo que genera el proceso de saltos de Markov

```

k=1 #estado inicial
T<-0 #tiempos de salto
N<-100 #tiempo de duración total del proceso
t<-0 #tiempo de permanencia en cada estado
S<-matrix(0,est,est) #matriz de saltos

i=1 #inicializas en 1 pues R no acepta 0
while(T[i]<=N)
  {
    j=k[i] #se define j referenciando al estado
    L=(-Q[j,j])
    T[i+1]=T[i]+rexp(1, rate=L)
    r[i]=runif(1)
    for (z in 1:est)
      {
        if (r[i]<cumsum(K[j,])[z])
          { #suma acumulada de los elementos del renglon j
            k[i+1]=z #el nuevo estado
            break
          }
      }
    i=i+1 #aumento de i para que continúe
  }

```

```

                                el proceso
}

#tiempo no acotado
T
#generamos el vector t de los tiempos de estancia en cada
  intervalo de tiempo
if(T[i]>N){
T[i]=N}
for(i in 1:length(T)-1){
  t[i]=T[i+1]-T[i]
}
T
k
t

kp<-k[1*1:i] #nos muestra los saltos, excepto el último (
  nos sirve para sacar el tiempo de estancia por estado)
kp

Tp<-T[-1]
Tp
TP<-Tp[-(length(Tp))] #sirve para graficar:nos muestra
  los tiempos de salto excepto el primero (no nos
  interesa porque es 0)
TP

st<-array(0,dim=est) #vector de tiempo total de
  permanencia por estado
for(z in 1:length(t)){
for(j in 1:est){
if(j==kp[z]){
st[j]=st[j]+t[z]
}
}
}
}

```

```

st
sum(st) #tiempo total del proceso

#generamos una matriz que muestre los saltos entre
    estados
for (d in 1:length(kp)-1){
S[kp[d],kp[d+1]]<-S[kp[d],kp[d+1]]+1
}
S
sum(S) #total de saltos = kp-1

salto<-stepfun(TP,kp, f =0, right = FALSE)#define la
    función de salto para graficar
#dos maneras de graficar el proceso
#grafica la función de salto sin el último salto
plot.stepfun(salto ,main="Trayectoria de un proceso de
    saltos ",xlab="Tiempo",ylab="Estados",verticals=FALSE,
    do.points=TRUE, pch=16,col.points="black",ylim=c(0,est
    ),xlim=c(0,N))
#salto2<-stepfun(Tp,k, f =0, right = FALSE)#define la
    función de salto con el último salto

#grafica la función de salto con el último salto
#plot.stepfun(salto2 ,main="Trayectoria de un proceso de
    saltos ",xlab="Tiempo",ylab="Estados",verticals=FALSE,
    do.points=TRUE, pch=16,col.points="black",ylim=c(0,est
    ),xlim=c(0,N))
abline(v=N, col="red", lty=4)

#Esta parte nos sirve para saber que nuestra simulación
    del proceso funciona correctamente.
#se estima la matriz Q mediante el número de saltos
    obtenidos de un estado a otro en el proceso.

QE<-matrix(0,est,est)
for (i in 1:est){

```

```

for (j in 1:est){
QE[i , j]<-S [ i , j] / st [ i ]
}
}
f<-0
cc<-array (data=NA, dim=est )
for ( i in 1:est){
    for ( j in 1:est){
        f<-f+QE[ i , j ]
    }
    cc [ i ]<-(-f)
    f<-0
}
for ( i in 1:est){
    for ( j in 1:est){
        if (i==j)
QE[ i , j ]<-cc [ i ]
        else
QE[ i , j ]<-QE[ i , j ]
    }
}
QE
Q

# Se generan las matrices de transición de probabilidades
en el tiempo N para la matriz real Q y el estimador
QE

Pq<-expm(Q)
Pqe<-expm(QE)
Pq
Pqe

##### En la siguiente parte del código se obtienen
puntos discretos del proceso de saltos de Markov
generado###

```

```

h<-.01 #tamaño del intervalo de la observación, mientras
      más pequeño sea más observaciones tendremos.

#se divide el horizonte de tiempo en intervalos de tamaño
  h

y<-seq(0,N,h)
y1<-length(y)
Y<-rep(0,y1)
for (i in 2:y1-1){
  for (j in 1:length(T)-1){
    if ((y[i]>=T[j])&&(y[i]<T[j+1]))
      Y[i]<-kp[j]
    }
  }
Y1<-Y[-length(Y)]
Y2<-c(kp[1],Y1,kp[length(kp)])
Y2
abline(v=y, lty=2) #en cada "punto discreto" se dibuja una
                  línea vertical

#####Generamos una matriz que nos indica el número de
          veces que el saltó a otro estado o permaneció en él.

O<-matrix(0,est,est)
for(j in 1:length(Y2)-1){
  O[Y2[j],Y2[j+1]]<-O[Y2[j],Y2[j+1]]+1
}
O
sum(O) #número de observaciones discretas totales

```

Código en R Project: Método Runge Kutta de cuarto orden para calcular un sistema de dos ecuaciones diferenciales.

```
ff <- function(x,y){ return( y*(-.007*y+6) ) } #
  ecuación diferencial original

x0 <- 0
x1 <- .5
h <- .01

x <- seq(x0,x1,h) #hace la partición de los intervalos en
  los que se harán las iteraciones
m <- length(x)
y <- rep(0,m)
y[1] <- 10 # condiciones iniciales

for( i in 2:m ){
f0 <- ff(x[i-1],y[i-1])
k1 <- h*f0
k2 <- h*ff(x[i-1]+(h/2),y[i-1]+(k1/2))
k3<-h*ff(x[i-1]+(h/2),y[i-1]+(k2/2))
k4<-h*ff(x[i-1]+h,y[i-1]+k3)
y[i] <- y[i-1]+(k1+(2*k2)+(2*k3)+k4)/6 } # Runge
  -Kutta-4

yrk4 <- y[m]

yrk4 #solución del sistema de ecuaciones
```


Bibliografía

- [1] Altman, E., Saunders, A. (1998). "Credit risk measurement: Developments over the last 20 years". *Journal of Banking & Finance* 21: 1721-1742. Salomon Brothers Center, Leonard Stern School of Business, New York University. New York.
- [2] Ayala, Y., Melo, O. (2007) "Estimación de datos faltantes en medidas repetidas con respuesta binaria". *Revista Colombiana de Estadística*, Departamento de Matemáticas y Estadística Universidad Pedagógica y Tecnológica de Colombia, Tunja Colombia. Departamento de Estadística, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá Colombia.
- [3] Baltazar Larios, F. (2011). "Estimation of discretely observed Markov Processes". Tesis para obtener grado de Doctor en ciencias. UNAM, México D.F.
- [4] Bladt, M. & Sørensen, M. (2006). "Statistical inference for discretely observed Markov jump processes". *Statist. Soc.*, 77:395–410.
- [5] Bladt, M. & Sørensen, M. (2009). "Efficient estimation of transition rates between credit ratings from observations at discrete time points". *Quantitative Finance*, 9:147–160.
- [6] Bladt, M., Neuts, M.F., Meini, B. and Sericola, B. (2002). "Distributions of reward functions on continuous-time, Markov chains. In *Matrix-analytic Methods: Theory and Applications* (ed. G. Latouche). Singapore: World Scientific Publishing Company.
- [7] Cinlar, E. (1941). "Introduction to Stochastic Processes". Englewood cliffs : Prentice Hall.

- [8] Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society, Series B, Methodological*.
- [9] Doob, J. (1953). "Stochastic Processes" Wiley, New York.
- [10] Elizondo, Alan. (2004). "Medición Integral del Riesgo de Crédito". Editorial Limusa, Grupo Noriega Editores. México, D.F.
- [11] Jarrow, R., Lando D., Turnbull, S. (1997). "A Markov Model for the Term Structure of Credit Risk Spreads". *Rev. Financial Stud.*, 10:481-523.
- [12] Jarrow, R., Turnbull, S. (1995). "Pricing Derivatives on Financial Securities Subject to Credit Risk". *The Journal of Finance*, Vol. 50 No.1 53-85.
- [13] Karlin, S. & Taylor, H.M. (1975). "A First Course in Stochastic Processes". Academic Press, Orlando.
- [14] Karlin, S. & Taylor, H.M. (1981). "A Second Course in Stochastic Processes". Academic Press, Orlando.
- [15] Lawler, G.F. (1995). "Introduction to Stochastic Processes". Chapman & Hall Probability Series, New York.
- [16] Metzner, P., Dittmer, E., Jahnke, T., Schütte, Ch. (2007). "Generator Estimation of Markov Jump Processes". *Journal of Computational Physics* 227:353-375 ScienceDirect, Institute of Mathematics II, Free University of Berlin, Berlin Germany.
- [17] Metzner, P., Horenko, I., Schütte, Ch. (2007). "Generator Estimation of Markov Jump Processes based on incomplete observations non-equidistant in time". Institute of Mathematics II, Free University of Berlin, Berlin Germany.
- [18] Rincón, Luis. (2012). "Introducción a los Procesos Estocásticos". Departamento de Matemáticas, Facultad de Ciencias UNAM, México D.F.
- [19] Ross, S. M. (1997). "Introduction to Probability Models". Department of Industrial Engineering and Operations Research, University of California, Berkeley California.

- [20] Samaniego Medina, Reyes. (2008). "El riesgo de crédito en el marco del Acuerdo Basilea II". Delta publicaciones universitarias. Madrid, España.
- [21] Serfozo, R. (2009). "Basics of Applied Stochastic Processes". Georgia Institute of Technology, School of Industrial & Systems Engineering, Atlanta.
- [22] "Reformas del Comité de Supervisión Bancaria de Basilea-Basilea III", año 2012. http://www.bis.org/bcbs/basel3/b3summarytable_es.pdf
- [23] "Fitch-Complete Sovereign Rating History", Fitch Ratings. 24 de agosto de 2012. http://www.fitchratings.com/web_content/ratings/sovereign_ratings_history.xls
- [24] "Global Covered Bonds 2011 Rating Transition Study. Covered Bonds' Safe Product Status Threatened by the Global Financial Crisis. Special Report." Fitch Ratings. http://www.fitchratings.com/creditdesk/reports/report_frame?rpt_id=660011