

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

MODELACIÓN DE LA MOROSIDAD DE CRÉDITO COMERCIAL Y DE CONSUMO PARA LA BANCA PRIVADA ECUATORIANA USANDO MODELOS OCULTOS DE MARKOV

**TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO MATEMÁTICO**

PROYECTO DE INVESTIGACIÓN

EDWIN DAVID SIMBAÑA FARINANGO

edwinsimb@gmail.com

DIRECTORA: ADRIANA UQUILLAS ANDRADE, PhD.

adriana.uquillas@epn.edu.ec

Quito, Junio 2019

DECLARACIÓN

Yo, EDWIN DAVID SIMBAÑA FARINANGO, declaro bajo juramento que el trabajo aquí escrito es de mi autoría, que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

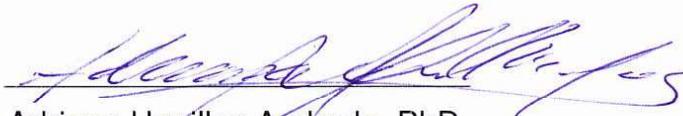
A través de la presente declaración cedo mis derechos de propiedad intelectual, correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por normatividad institucional vigente.



Edwin David Simbaña Farinango

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por EDWIN DAVID SIMBAÑA FARI-
NANGO, bajo mi supervisión.

A handwritten signature in blue ink, appearing to read 'Adriana Uquillas Andrade', written over a horizontal line.

Adriana Uquillas Andrade, PhD.
DIRECTORA DEL PROYECTO

AGRADECIMIENTOS

En primer lugar, agradezco a Dios, por brindarme salud y bienestar, por permitirme alcanzar mis objetivos, cumplir mis metas y sobre todo por la culminación de esta etapa de mi vida.

Agradezco también a mis padres, por todo su amor, sacrificio, paciencia, y apoyo incondicional; ya que fueron un pilar fundamental en mi vida estudiantil y mi formación profesional.

Agradezco también a mi directora Dra. Adriana Uquillas, por haberme brindado la oportunidad de recurrir a su capacidad, conocimiento y experiencia, así como también, por haberme tenido toda la paciencia del mundo para guiarme durante todo el desarrollo de este trabajo.

Agradezco también a la Escuela Politécnica Nacional y la Facultad de Ciencias, por permitirme estudiar mi carrera, así como también, a los diferentes docentes que la conforman, por brindarme sus conocimientos y apoyo para seguir adelante en la formación de mi carrera universitaria.

Y para finalizar, también agradezco a mis amigos y compañeros de clase, pues gracias a su compañerismo, amistad y apoyo aportaron un alto porcentaje a mis ganas de seguir adelante en mi vida estudiantil.

Edwin Simbaña

DEDICATORIA

A mis padres, quienes me brindaron su apoyo y amor incondicional; a mis hermanas quienes fueron mi motivación; a mis abuelos quienes fueron las personas más amorosas que he conocido.

¡Muchas gracias!

Edwin Simbaña

Índice general

DECLARACIÓN	I
CERTIFICACIÓN	II
AGRADECIMIENTOS	III
DEDICATORIA	IV
Índice de general	V
Índice de figuras	VII
Índice de cuadros	X
Resumen	XII
Abstract	XIII
1. Introducción.	1
1.1. Sistema financiero nacional ecuatoriano	2
1.2. Revisión de literatura	7
2. Marco Teórico.	11
2.1. Modelos de mezclas	11
2.2. Cadenas de Markov	15
2.2.1. Probabilidades de transición	16

2.2.2. Probabilidades de transición en n pasos	18
2.2.3. Distribución inicial	19
2.2.4. Evolución de las distribuciones	20
2.2.5. Distribuciones estacionarias	20
2.2.6. Estructura Gráfica de las cadenas de Markov	21
2.2.7. Cadenas de Markov de orden superior	21
2.3. Modelos ocultos de Markov	23
2.3.1. Estimación de los parámetros por el algoritmo (EM)	27
2.3.2. Pronósticos, decodificación y predicción de estados	37
2.3.3. Selección y validación del modelo	41
2.3.4. Variaciones de los modelos ocultos de Markov	44
3. Metodología y Resultados.	47
3.1. Descripción del modelo	47
3.2. Modelamiento	52
3.2.1. Modelamiento de la cartera de crédito comercial	52
3.2.2. Modelamiento de la cartera de crédito de consumo	69
4. Conclusiones y Recomendaciones.	89
Bibliografía	93
A. Modelos ARIMA - SARIMA	99
B. Bases de datos	101
C. Funciones de R y esquema del procedimiento	103

Índice de figuras

1.1. Series de activos, pasivos y patrimonio de la banca privada.	3
1.2. Índices de morosidad de las carteras de crédito de la banca privada. . . .	4
1.3. Participación de las carteras de crédito de la banca privada.	5
2.1. Ejemplos de modelos de mezclas gaussianas con 3 componentes.	14
2.2. Arquitectura de una cadena de Markov de primer orden	16
2.3. Representación gráfica de una matriz de transición asociada a la cadena de Markov.	22
2.4. Arquitectura de una cadena de Markov de segundo orden	22
2.5. Arquitectura de un modelo oculto de Markov de primer orden	24
2.6. Construcción de los pseudo-residuos normales.	43
2.7. Dependencias adicionales en los modelos ocultos de Markov.	45
2.8. Dependencia adicional de las observaciones con observaciones previas.	45
2.9. Dependencia adicional de las observaciones con estados previos.	46
3.1. Procedimiento del modelo.	48
3.2. Serie de morosidad de la cartera de crédito comercial.	53
3.3. Histograma de la cartera de crédito comercial.	53
3.4. Grupos generados con los modelos de mezclas para la morosidad de la cartera de crédito comercial.	54
3.5. Densidades de mezcla formados por los modelos de mezcla para la mo- rosidad de la cartera de crédito comercial.	56

3.6. Gráfico de los pseudo-residuos de los modelos ocultos de Markov estimados para la morosidad de la cartera de crédito comercial.	59
3.7. Histogramas de los pseudo-residuos de los modelos ocultos de Markov estimados para la morosidad de la cartera de crédito comerciala.	60
3.8. Gráficos qqplot de los pseudo-residuos de los modelos ocultos de Markov estimados para la morosidad de la cartera de crédito comercial. . . .	61
3.9. Decodificación de los estados mediante el uso del algoritmo de Viterbi para la morosidad de la cartera de crédito comercial.	64
3.10. Probabilidades de pertenencia a los estados para la morosidad de la cartera de crédito comercial.	64
3.11. Predicción de los estados para el año 2017 para la morosidad de la cartera de crédito comercial.	65
3.12. Densidades y distribuciones de pronóstico para la morosidad de la cartera de crédito comercial.	66
3.13. Ajustes y pronósticos con el modelo oculto de Markov y el modelo ARIMA con las observaciones reales para la morosidad de la cartera de crédito comercial.	68
3.14. Comparación de pronósticos e intervalos de confianza de los modelos con las observaciones reales de la morosidad de la cartera de crédito comercial.	70
3.15. Serie de morosidad de la cartera de crédito de consumo.	70
3.16. Histograma de la cartera de crédito de consumo.	71
3.17. Grupos generados con los modelos de mezclas para la morosidad de la cartera de crédito de consumo.	72
3.18. Densidades de mezcla formados por los modelos de mezcla para la morosidad de la cartera de crédito de consumo.	74
3.19. Gráfico de los pseudo-residuos de los modelos ocultos de Markov estimados para la morosidad de la cartera de crédito de consumo.	77
3.20. Histogramas de los pseudo-residuos de los modelos ocultos de Markov estimados para la morosidad de la cartera de crédito de consumo.	78
3.21. Gráficos qqplot de los pseudo-residuos de los modelos ocultos de Markov estimados para la morosidad de la cartera de crédito de consumo. . .	79

3.22. Decodificación de los estados mediante el algoritmo de Viterbi para la morosidad de la cartera de crédito de consumo.	82
3.23. Probabilidades de pertenencia a los estados para la morosidad de la cartera de crédito de consumo.	82
3.24. Predicción de los estados para el año 2017 para la morosidad de la cartera de crédito de consumo.	83
3.25. Densidades y distribuciones de pronóstico para la morosidad de la cartera de crédito de consumo.	85
3.26. Ajustes y pronósticos con el modelo oculto de Markov y el modelo SARIMA con las observaciones reales para la morosidad de la cartera de crédito de consumo.	87
3.27. Comparación de pronósticos e intervalos de confianza de los modelos con las observaciones reales de la morosidad de la cartera de crédito de consumo.	88
A.1. Resultados del modelo ARIMA	99
A.2. Resultados del modelo SARIMA	100
C.1. Flujograma y funciones del algoritmo.	106

Índice de cuadros

1.1. Composición de la cartera improductiva y la cartera bruta.	6
2.1. Parámetros de las figuras de los modelos de mezclas 2.1	15
3.1. Cuadro de estadísticos de la serie de morosidad comercial.	52
3.2. Parámetros para el modelo de mezcla con 2 grupos para la morosidad de la cartera de crédito comercial.	55
3.3. Parámetros para el modelo de mezcla con 3 grupos para la morosidad de la cartera de crédito comercial.	55
3.4. Parámetros para el modelo de mezcla con 4 grupos para la morosidad de la cartera de crédito comercial.	55
3.5. Parámetros para el modelo de mezcla con 5 grupos para la morosidad de la cartera de crédito comercial.	55
3.6. Criterios para seleccionar el mejor modelo oculto de Markov para la morosidad de la cartera de crédito comercial.	58
3.7. Estimaciones e intervalos de confianza de los parámetros del modelo oculto de Markov seleccionado para la morosidad de la cartera comercial.	63
3.8. Probabilidades de predicción del estado para el año 2017 para la morosidad de la cartera de crédito comercial.	65
3.9. Estadísticos de las distribuciones de pronósticos para el año 2017 para la morosidad de la cartera de crédito comercial.	67
3.10. Medidas de exactitud para la morosidad de la cartera de crédito comercial.	69
3.11. Cuadro de estadísticos de la serie de morosidad de consumo.	71

3.12. Parámetros para el modelo de mezcla con 2 grupos para la morosidad de la cartera de crédito de consumo.	73
3.13. Parámetros para el modelo de mezcla con 3 grupos para la morosidad de la cartera de crédito de consumo.	73
3.14. Parámetros para el modelo de mezcla con 4 grupos para la morosidad de la cartera de crédito de consumo.	73
3.15. Parámetros para el modelo de mezcla con 5 grupos para la morosidad de la cartera de crédito de consumo.	73
3.16. Criterios para seleccionar el mejor modelo oculto de Markov para la morosidad de la cartera de crédito de consumo.	76
3.17. Estimaciones e intervalos de confianza de los parámetros del modelo oculto de Markov seleccionado para la morosidad de la cartera de consumo.	81
3.18. Probabilidades de predicción del estado para el año 2017 para la cartera de crédito de consumo.	84
3.19. Estadísticos de las distribuciones de pronósticos para el año 2017 para la morosidad de la cartera de consumo.	86
3.20. Medidas de exactitud para la morosidad de la cartera de crédito de consumo.	86
B.1. Participación de la cartera de crédito comercial	101
B.2. Participación de la cartera de crédito de consumo	101
B.3. Serie de morosidad de la cartera de crédito comercial	102
B.4. Serie de morosidad de la cartera de crédito de consumo	102

Resumen

El propósito de este trabajo de investigación es crear un modelo para la morosidad de las carteras de crédito comercial y de consumo del sistema financiero nacional, centrándose en el sistema de bancos privados, para lo cual se usará como herramienta de modelación una extensión de las cadenas de Markov, como lo son los modelos ocultos de Markov, cuya particularidad son sus estados no visibles.

Se consideran únicamente las series de morosidad de las carteras de crédito comercial y de consumo, debido a su gran representatividad en el sistema financiero ecuatoriano, para lo cual se toman datos con frecuencia mensual a partir de julio 2002 hasta diciembre 2017.

En esta investigación se describe la teoría y metodología de los modelos ocultos de Markov, se ajustan modelos de mezcla para cada cartera con el fin de generar los estados de la cadena de Markov subyacente que representan niveles de riesgo de morosidad, se estiman valores iniciales para el algoritmo de Baum-Welch para el cálculo de los parámetros del modelo oculto de Markov, se realiza la decodificación de los estados ocultos, se realiza la predicción de estados y pronósticos. Finalmente, se analizan los resultados obtenidos con los modelos clásicos ARIMA/SARIMA. Adicionalmente, se hace uso del software estadístico R para obtener los resultados del modelo.

Palabras claves: Procesos estocásticos, Modelos ocultos de Markov, Matrices de transición, Morosidad de crédito, Programación en R.

Abstract

The purpose of this research is to create a model for commercial and consumer default credit portfolio for the national financial system, by focusing on the private banking system, it will be used an extension of Markov chains as modeling tool called hidden Markov models, whose particularity is its non-visible states.

Default credit series of commercial and consumer credit portfolios are considered, due to their great representation in the national financial system, then monthly information is taken from July 2002 to December 2017.

In this research, theory and methodology are described, mixed models are fitted for each portfolio in order to generate states for underlying Markov chain which represent default risk levels, initial values for the Baum-Welch algorithm are estimated in order to compute the hidden Markov model parameters, hidden states are decoded and states predictions and forecasts are computed. Finally, results are analyzed with classic ARI-MA/SARIMA models. Furthermore, statistical software R will be used to get results of the model.

Keywords: Stochastic processes, Hidden Markov models, Transition matrices, Default credit, R programming.

Capítulo 1

Introducción.

En las últimas décadas varios países no sólo los emergentes sino también países como Estados Unidos, Japón, entre otros, han sufrido crisis financieras ocasionados por variaciones bruscas en el sistema financiero. Éste a su vez, es un eje fundamental para la economía de un país, ya que por su intermedio se ejecutan todas las tracciones personales, empresariales y gubernamentales tanto a nivel nacional e internacional. Además, contar con un sistema saludable y sólido permite que los recursos financieros circulen eficientemente desde las personas que tienen recursos monetario excedentes hacia las personas que necesiten dichos recursos, lo que permite el desarrollo de la economía de un país.

Muchos son los factores que influyen para que se produzca una crisis financiera, pero entre las más comunes se mencionan: *“las malas prácticas bancarias, malos negocios, inadecuada supervisión bancaria, violación de las normas, fraude, ajustes estructurales en la economía, inestabilidad macroeconómica, errores en la política macroeconómica, sistemas de control inadecuados, fallas operacionales, entre otras.”* (Cuevas, 2013)

Ecuador no está exento de crisis. De hecho, sufrió una crisis bancaria en el año 1999, marcando un punto crítico en la historia del país. Ocasionando varios eventos desfavorables como: *“alza de tasas de interés, devaluación de la moneda, quiebra de empresas, problemas de liquidez entre otras. Afectado principalmente a los sectores más vulnerables.”* (Bonilla Toapanta and Gálvez Castillo, 2007). Por tal motivo, es importante realizar un seguimiento y evaluación de las entidades financieras; es decir, monitorear la situación de la entidad financiera periódicamente y prevenir eventos desfavorables, para asegurar la estabilidad del sistema financiero y así evitar situaciones como la disminución de crédito, problemas de liquidez, desempleo, etc.

Por su parte, el Comité de Supervisión Bancaria de Basilea¹, en su segundo pilar establece que *“El proceso del examen supervisor establecido en este marco no tiene por objetivo únicamente garantizar que los bancos posean el capital necesario para cubrir los riesgos de sus actividades, sino que también insta a los bancos a que desarrollen y utilicen mejores técnicas de gestión de riesgos en el seguimiento y control de los mismos.”* (Basilea, 2006).

Por tanto, de lo mencionado anteriormente, se motiva el desarrollo de nuevos métodos y técnicas que permitan controlar y gestionar el riesgo que pueda tener el sistema financiero, y de ser necesario, tomar acciones correctivas. Un indicador de riesgo es el índice de morosidad, es decir, el porcentaje de la cartera improductiva frente al total de la cartera de la entidad. Este indicador es de mucha importancia, pues una tasa elevada de morosidad indica un alto riesgo para que una entidad financiera entre en una crisis, lo que ocasionaría un fuerte golpe a la economía del país.

1.1. Sistema financiero nacional ecuatoriano

El sistema financiero es un eje fundamental para la economía ecuatoriana, ya que por su intermedio se ejecutan todas las transacciones personales, empresariales y gubernamentales, tanto a nivel nacional e internacional.

La Superintendencia de Bancos del Ecuador² define el sistema financiero como *“un servicio de orden público, regulado y controlado por el Estado, que pueden ser prestadas por las entidades que conforman el sistema financiero nacional, previa autorización de los organismos de control, en el marco de la normativa que expida la Junta de Política y Regulación Monetaria y Financiera.”* (SBS, 2016c)

El sistema financiero está integrado por tres sectores: el sector financiero público (*compuesto por bancos y corporaciones*), el sector financiero privado (*compuesto por bancos múltiples y bancos especializados, servicios financieros, servicios auxiliares del sistema financiero y por subsidiarias o afiliadas de las entidades financieras domiciliadas en el Ecuador*) y el sector financiero popular y solidario (*compuesto por cooperativas y mutualistas de ahorro y crédito para la vivienda*).

De acuerdo a cifras de la Superintendencia de Bancos del Ecuador, los bancos

¹Organización mundial cuya función es fortalecer la solidez de los sistemas financieros constituida en 1975.

²Ente supervisor y controlador de las actividades que ejercen las entidades financieras y de seguridad social, públicas y privadas.

constituyen el mayor y más importante participante del mercado con más del 90% de las operaciones del total del sistema³. Además, canaliza gran parte de la riqueza del país. Esto se puede apreciar en la figura 1.1 donde a través del tiempo la administración de la riqueza, por parte de la banca privada, es cada vez mayor.



Figura 1.1: Series de activos, pasivos y patrimonio de la banca privada.

Además, la Superintendencia de Bancos del Ecuador menciona que, “*el sector de bancos privados está conformado por entidades bancarias que canalizan recursos del público en forma de depósitos y los coloca en forma de créditos e inversiones a fin de financiar actividades productivas y rentables, capaces de generar empleo y crecimiento económico, usando para ello recursos humanos, operativos y financieros y por ende, asumiendo riesgos.*” (SBS, 2016c)

Uno de los riesgos que asume los bancos es el riesgo de crédito representado por el índice de morosidad de la cartera de crédito⁴, donde la Superintendencia de Bancos en su papel de supervisor lo monitorea y calcula, por línea de negocio, de la siguiente manera:

$$\text{Morosidad de la cartera} = \frac{\text{Cartera improductiva}}{\text{Cartera total bruta}} \quad (1.1)$$

Dentro del sistema financiero nacional, “*la Cartera de créditos a partir de agosto de*

³Romero, Bertha (2015). Estructura del sistema financiero ecuatoriano, Educación Financiera

⁴Debido a que la principal actividad de los bancos es la concesión de créditos, pues contribuye gran parte de sus ingresos, pero a la vez es un detonante de riesgo.

2012 está clasificada en 6 líneas de negocio: Comercial, Consumo, Vivienda, Micro-empresa, Crédito Educativo e Inversión Pública⁵”.

En la figura 1.2 se presenta la evolución de los índices de morosidad de las carteras de crédito de las 6 líneas de negocio para la banca privada. En ésta, se aprecia que la cartera comercial, en períodos iniciales, presenta valores altos de morosidad, especialmente en noviembre 2002 donde alcanza un valor máximo de 11,49 %, mientras que la morosidad de la cartera de consumo inicia con un valor de 7,93 % para luego alcanzar su valor más alto de 11,62 % en febrero 2004. Además, se puede apreciar que el índice de morosidad de la cartera de crédito comercial tiende a decrecer hasta que en los últimos períodos se muestra constante hasta llegar a un valor de 1,01 % en diciembre 2017. Mientras, el índice de mora de la cartera de crédito de consumo tiene un comportamiento oscilante hasta llegar a un valor de 5,21 % en diciembre 2017. Por otro lado, la tasa de mora de la cartera de crédito educativa muestra valores a partir de agosto 2015, debido a que a partir de dicha fecha se comenzaron a realizar préstamos al sector educativo por parte de la banca privada. Finalmente, se muestra un comportamiento nulo del índice de mora de la cartera de crédito de inversión pública debido a que la banca privada no destina recursos a dicho sector.

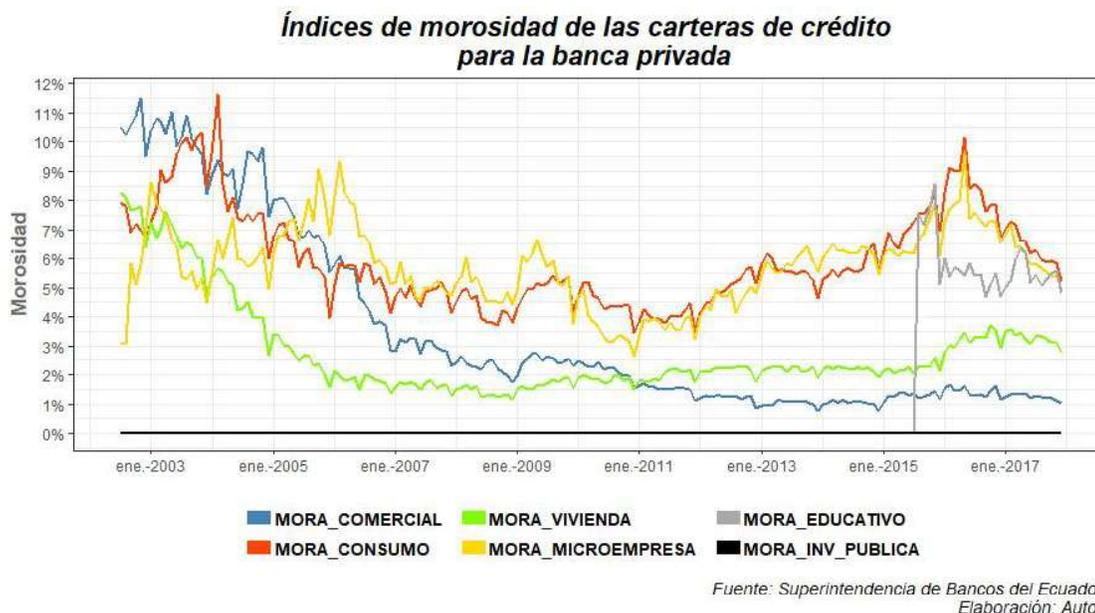


Figura 1.2: Índices de morosidad de las carteras de crédito de la banca privada.

De las carteras de crédito descritas, las carteras de crédito comercial y de consumo presentan una gran representatividad en el sistema financiero ecuatoriano, como se

⁵Superintendencia de Bancos, *Nota técnica 6*, Boletines Mensuales Bancos Privados.

aprecia en la figura 1.3 donde se muestra un diagrama de áreas a través del período de análisis. En ésta, se muestra que en cada período la cartera comercial ocupa gran parte del sistema financiero (*alrededor del 50 %*), seguida de la cartera de consumo, la cartera de vivienda, la cartera de microempresa y finalmente, la cartera educativa en menor proporción en los períodos finales. Por tanto, en este trabajo de investigación se presta mayor atención a las carteras: comercial y de consumo.

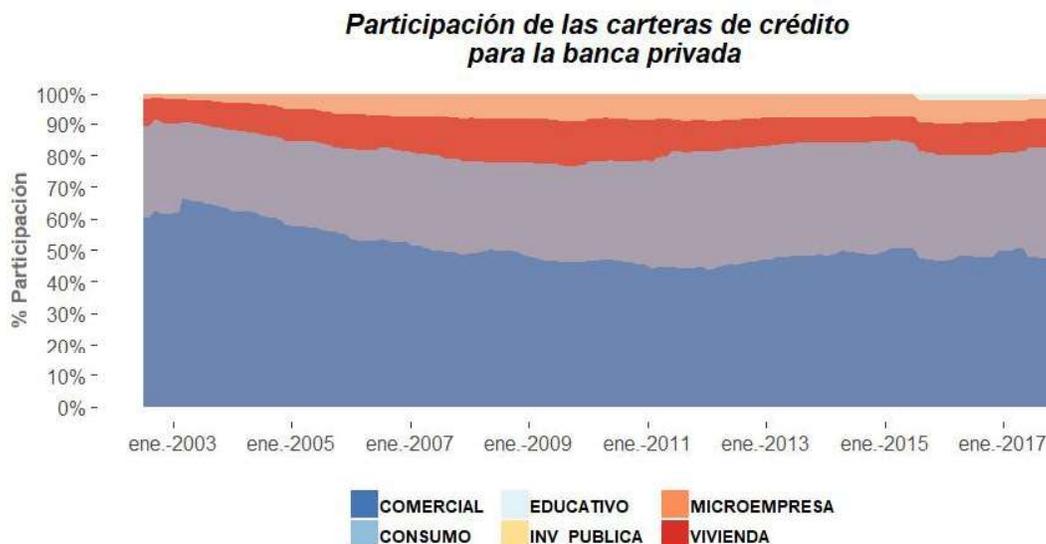


Figura 1.3: Participación de las carteras de crédito de la banca privada.

Por tal motivo, el objetivo principal de este proyecto de investigación es crear un modelo para la morosidad de las carteras de crédito comercial y de consumo para el sistema financiero nacional, específicamente para el sistema de bancos privados. Utilizando como herramienta de modelación los modelos ocultos de Markov.

De acuerdo a la Regulación No.JB-2011-1897 del 15 de Marzo del 2011 del Banco Central del Ecuador, en su Capítulo II, Sección II, Artículo 5, se tienen las siguientes definiciones:

Definición 1.1 (Crédito comercial). “Se entiende por créditos comerciales todos aquellos otorgados a personas naturales o jurídicas destinados al financiamiento de actividades de producción y comercialización de bienes y servicios en sus diferentes fases, cuya fuente de pago constituyen los ingresos por ventas u otros conceptos redituables, directamente relacionados con la actividad financiada.” (BCE, 2011)

Definición 1.2 (Crédito de consumo). “Son créditos otorgados a personas naturales destinados al pago de bienes, servicios o gastos no relacionados con una actividad

productiva, cuya fuente de pago es el ingreso neto mensual promedio del deudor, entendiéndose por éste el promedio de los ingresos brutos mensuales del núcleo familiar menos los gastos familiares estimados mensuales obtenidos de fuentes estables como: sueldos, salarios, honorarios, remesas.” (BCE, 2011)

La composición de la cartera improductiva y la cartera bruta se muestra en el cuadro 1.1.

Cartera improductiva	<ul style="list-style-type: none"> ■ Cartera de crédito que no devenga intereses ■ Cartera de crédito refinanciada que no devenga intereses ■ Cartera de crédito reestructurada que no devenga intereses ■ Cartera de crédito vencida ■ Cartera de crédito refinanciada vencida ■ Cartera de crédito reestructurada vencida
Cartera bruta	<ul style="list-style-type: none"> ■ Cartera de crédito por vencer ■ Cartera de crédito refinanciada por vencer ■ Cartera de crédito reestructurada por vencer ■ Cartera de crédito improductiva

Cuadro 1.1: Composición de la cartera improductiva y la cartera bruta.

Definición 1.3 (Cartera por Vencer). *“Se entiende por cartera por vencer aquella cuyo plazo no ha vencido y aquella vencida hasta 60, 30 o 90 días posteriores a la fecha de vencimiento de la operación, según sea el caso conforme a la Normativa vigente, o de acuerdo a lo establecido en el manual crediticio de la institución y en el respectivo contrato, sin perjuicio del cálculo de los intereses por mora a que hubiere lugar desde el día siguiente del vencimiento de la operación. También se la conoce como Cartera Vigente.” (BCE, 2011)*

Definición 1.4 (Cartera que no devenga intereses). *“Es aquella cartera que ha dejado de generar ingresos por intereses y con riesgos potenciales en su recuperación.” (BCE, 2011)*

Definición 1.5 (Cartera Vencida). *“Registra los créditos, en todas sus modalidades que no hubieren sido cancelados dentro de los plazos establecidos en la Codificación*

de Resoluciones de la Superintendencia de Bancos y de la Junta Bancaria. Adicionalmente registra las operaciones contingentes pagadas por la institución por cuenta de clientes que no se hayan recaudado.” (BCE, 2011)

1.2. Revisión de literatura

Un modelo oculto de Markov es un modelo estadístico que *“permite modelar procesos estocásticos, donde la ocurrencia de los estados está asociada con una distribución de probabilidad y donde las transiciones entre los estados (contables) están gobernadas por un conjunto de probabilidades llamadas probabilidades de transición de estados. En un estado particular, una observación se genera también de acuerdo a una distribución de probabilidad. Los estados no son visibles (ocultos) y su ocurrencia depende del estado en el instante anterior”*. (Maldonado, 2012). El objetivo es determinar los parámetros desconocidos (ocultos) de dicha cadena a partir de los parámetros observables.

Un modelo oculto de Markov consiste en un proceso estocástico doble, en el cual las realizaciones del primer proceso (proceso oculto - cadena de Markov), dan origen a un segundo proceso (proceso observado). Los dos procesos se logran caracterizar usando solo el proceso observable.

Los modelos ocultos de Markov tienen una amplia gama de aplicaciones en distintos campos. Para tipos de reconocimiento se tiene, Chediak and Maldonado (2014) quienes lo usan para el reconocimiento automático de fonemas; Martinez et al. (2003) y Ramos (2011) lo usan para el reconocimiento de voz, Macas et al. (2012) lo usan para el reconocimiento automático del habla y Jurafsky and Martin (2017) lo usan para el procesamiento del habla y lenguaje. En el área de la medicina, Evin et al. (2011) lo usan para la predicción de estados de hipotensión y Cooper and Lipsitch (2004) lo usan para analizar datos de infección hospitalaria. En el área de la economía, Bhar and Hamori (2006) detalla algunas aplicaciones económicas financieras. En otras áreas, pueden ser aplicados para modelar la preferencia de los consumidores como lo hace Patiño Antonioli (2011), donde son aplicados en el sector de consumo masivo para modelar las preferencias de marca en el mercado cervecero. En este estudio se usa como variables de mucha importancia al volumen de ventas de las empresas y las transacciones entre las marcas más representativas para finalmente realizar una comparación entre el modelo markoviano y el oculto. Por otro lado, se aplica para el modelamiento

de series financieras como lo hace Lihn (2017). Incluso pueden ser usados para modelar la selección de acciones como lo hace Nguyen and Nguyen (2015), donde ajustan estos modelo a series de índices como: la inflación, IPC, entre otros, para luego, realizar predicciones y conjuntamente con sus históricos identificar períodos de similitud y determinar las características de las acciones que han sido bien recompensadas. Así, se asignan puntuaciones y ponderaciones correspondientes a cada una de dichas características. Sobre estos puntajes se realiza la selección de acciones.

Existen varios estudios que usan diferentes métodos para la gestión del riesgo. Uno de las más usuales es la utilización de modelos de stress donde a través de diferentes variables microeconómicas y macroeconómicas explican la variable de interés. En Ecuador, Uquillas and González (2017), lo realizan a través de un modelo SARIMAX mientras que Guzmán Garzón (2017) usa indicadores financieros. Otros estudios hacen uso de matrices de transición, como es el caso de Montoya Sánchez et al. (2014) quienes modelan el riesgo de crédito de vivienda para el sector de bancos privados del Ecuador. Malik and Thomas (2012), los usan para modelar la calificación de crédito del consumidor, Vázquez and Vaquero (2013), crean matrices de probabilidad de transición de microcréditos. La Superintendencia de Bancos del Ecuador usa mucho las matrices de transición para el sistema de bancos como se muestra en SBS (2017), para análisis de cosechas del sistema bancario como se muestra en SBS (2016a) y para analizar el comportamiento crediticio sectorial como se muestra en SBS (2016b). Un enfoque más complejo se muestra en Kipkoech (2016) que usa cadenas de Markov multivariantes para medir y manejar el riesgo de crédito.

Existen varios estudios que usan los modelos ocultos de Markov relacionados con la gestión de riesgo. En América Latina, existe muy poca bibliografía, por ejemplo: en Ecuador, Guzmán Garzón (2017) lo usa como un modelo estocástico para modelar la quiebra de empresas ecuatorianas donde además usa los modelos de Altman para definir los estados de quiebre (*verde, gris, rojo*) para dos instantes de tiempo consecutivos para diferentes bancos del Ecuador, para en base a ellos, calcular las probabilidades de transición de la cadena de Markov. Luego, define las señales mediante la morosidad de crédito (*mora y no mora*). Finalmente, simula diferentes secuencias de señales para realizar una decodificación de los estados en dichas secuencias. En Colombia, Mateus (2008) presenta algunas de sus aplicaciones que mayoritariamente se enfoca a usarlos para automatización de procesos. En Venezuela, Jabbour and Maldonado (2009) lo usan conjuntamente con redes neuronales para predecir índices bursátiles como un sistema híbrido.

En otras regiones, la bibliografía es mayor, por ejemplo: Giampieri et al. (2005) analizan los datos de incumplimiento; Ching et al. (2009), tienen el mismo enfoque, pero lo hacen de manera interactiva; Korolkiewicz (2012) analiza la calidad de crédito; Benyacoub et al. (2014) lo usan para clasificación, donde se evalúa la probabilidad de pertenecer a un grupo basados en los modelos ocultos de Markov como clasificador lineal. Así, se definen un conjunto de ecuaciones donde cada uno de ellos representan probabilidades de pertenencia a los distintos estados (*o grupos*). Cada una de estas ecuaciones se expresan como una combinación lineal de las probabilidades del proceso observable. Ahuja et al. (2011), Oguz and Gurgun (2008) analizan el riesgo de crédito usando las probabilidades de incumplimiento; Teng et al. (2013) lo utilizan para medir el puntaje de crédito del consumidor; Ntwiga et al. (2016) quienes lo usan para modelar la calificación de crédito; Ntwiga (2016) quien analiza las redes sociales para modelar el riesgo de crédito. Otra aplicación es para la detección de fraude bancario por internet como lo hacen Mhamane and Lobo (2012).

Recopilando todas estas referencias, en este trabajo se usará las distribuciones de mezcla para clasificar en grupos de riesgo (señales ocultas de la cadena de Markov). Posteriormente, se usará los modelos ocultos de Markov para decodificar los estados ocultos en cada período y realizar pronósticos de morosidad de las carteras de crédito comercial y de consumo del sistema financiero nacional.

Finalmente, en los trabajos anteriormente mencionados se centran en modelar bancos. Este trabajo de investigación es innovador, pues se pretende modelar el sistema financiero nacional, a través de las series mensuales⁶ de la morosidad de las carteras de crédito comercial y de consumo, centrándose en los bancos privados, pues presentan gran participación en el sistema financiero nacional. Por otro lado, los modelos ocultos de Markov contienen inherente una matriz de transición que, de acuerdo a Pérez Delgado (2014), es de suma importancia para tomar decisiones, pues permite conocer la probabilidad de que un nivel de riesgo cambie otro, ya sea mejor o peor, en un período de tiempo. Lo que permite un adecuado control y gestión del riesgo y permitiría implementar políticas económicas que permitan mantener o mejorar la calidad de las diferentes carteras de crédito.

La estructura del presente trabajo de investigación es la siguiente, en el capítulo 2, se presenta el marco teórico donde describen los aspectos teóricos para comprender la metodología de los modelos de mezclas y los modelos ocultos de Markov. En el capítulo 3, se aplica la metodología analítica para el modelamiento de la morosidad

⁶La información fue obtenida la página de la Superintendencia de Bancos, Estadísticas

de las carteras de crédito comercial y de consumo, usando los modelos ocultos de Markov. Finalmente, en el capítulo 4, se presenta las principales conclusiones que se obtuvieron en el desarrollo de este trabajo.

Capítulo 2

Marco Teórico.

En este capítulo se presentan las nociones y definiciones teóricas que son necesarias para entender la metodología utilizada para modelar las series de morosidad de crédito de la banca privada usando los modelos ocultos de Markov. Se presenta la teoría de los modelos de mezclas, las cadenas de Markov y finalmente los modelos ocultos de Markov; en el cual se fundamenta este trabajo.

2.1. Modelos de mezclas

En esta sección se toman como principales referencias a Aggarwal and Reddy (2013, cap. 3), Salas González (2008, cap. 1) y Zucchini et al. (2016, cap. 1).

Los modelos de mezclas han sido una herramienta de modelado estadístico de gran flexibilidad, capaz de modelar muchos fenómenos en varias disciplinas como en genética, bioinformática, ingeniería entre otras. De acuerdo a Zucchini et al. (2016), los modelos de mezcla son modelos diseñados para acomodar la heterogeneidad no observada en la población; es decir, la población puede consistir en grupos no observados, cada uno con una distribución diferente. De acuerdo a Aggarwal and Reddy (2013), los modelos mezcla son modelos probabilísticos que se usan para encontrar clústers para un conjunto de datos.

En los modelos de mezcla para el análisis de agrupamiento, se suponen que las observaciones a agrupar, se obtienen de uno o varios grupos de distribuciones de probabilidad, y el objetivo es estimar dichos parámetros para que mejor se ajusten a los datos. Inferir los parámetros de estas distribuciones e identificar cuál generó cada observación conduce a un agrupamiento de las observaciones.

En este trabajo, se usará el enfoque de agrupamiento de los modelos de mezcla para generar grupos, que para el caso de la tasa de morosidad de la cartera de crédito representará niveles de riesgo (*valores altos de morosidad denotarían la presencia de crisis (mayor riesgo) y para valores bajos, su ausencia (menor riesgo)*), los cuales servirán como conjunto de estados (*finito*) para la cadena de Markov subyacente de los modelos ocultos de Markov.

Matemáticamente, si la variable aleatoria (*discreta o continua*) X es una mezcla de m componentes (*clústers*), entonces la función de probabilidad de X , $P_X(x)$, es una suma de distribuciones con distintos pesos ω_i , $i = 1, 2, \dots, m$. Este modelo se representa de la siguiente manera:

Definición 2.1. Sean m componentes del modelo de mezcla, donde cada componente es representado por una distribución (*discreta o continua*) $p(x|\theta_i)$ con parámetros θ_i y ω_i su respectivos pesos, con $i = 1, 2, \dots, m$. Entonces, la función de probabilidad de X está dada por:

$$p_X(x) = \sum_{i=1}^m \omega_i p(x|\theta_i) \quad (2.1)$$

donde $0 \leq \omega_i \leq 1$, $i = 1, 2, \dots, m$ y $\sum_{i=1}^m \omega_i = 1$.

En general, este modelo ha sido estudiado para diversas distribuciones $p(x|\theta_i)$ (*discreta o continua*). En particular, es muy estudiado con distribuciones gaussianas. En este trabajo, las observaciones de morosidad toman valores de 0 a 1, por lo que, se opta por usar distribuciones beta para los modelos de mezclas.

Ejemplo 2.1. Un modelo de mezcla gaussiana, donde cada componente tiene una distribución gaussiana $p(x|\theta_i) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left\{ -\frac{(x-\mu_i)^2}{2\sigma_i^2} \right\}$, quedaría expresado de la siguiente forma:

$$p_X(x) = \sum_{i=1}^m \omega_i \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left\{ -\frac{(x-\mu_i)^2}{2\sigma_i^2} \right\} \quad (2.2)$$

La esperanza y el momento k -ésimo de una distribución de mezcla es calculado en términos de las esperanzas y momentos de las distribuciones de sus componentes, respectivamente. Es decir, si X representa la variable aleatoria de la distribución de mezcla y X_i representa la variable aleatoria de la distribución del componente i ($i = 1, 2, \dots, m$). Entonces, la esperanza de la distribución de mezcla se expresa de la siguiente manera:

$$E(X) = \sum_{i=1}^m \omega_i E(X_i) \quad (2.3)$$

Y el momento k -ésimo se expresa de la siguiente manera:

$$E(X^k) = \sum_{i=1}^m \omega_i E(X_i^k), \quad k = 1, 2, \dots \quad (2.4)$$

La estimación de los parámetros (ω_i y los parámetros θ_i de la distribución de los componentes.) del modelo de mezcla se lo realiza por máxima verosimilitud. Se asume que el parámetro general del modelo de mezcla es un conjunto $\Theta = \{\omega_1, \omega_2, \dots, \omega_m, \theta_1, \theta_2, \dots, \theta_m\}$ y que se tiene de una muestra $X = \{x_1, x_2, \dots, x_n\}$ de n observaciones independientes extraída de la distribución de mezcla desconocida. De esta manera, la función de verosimilitud del modelo de mezcla está dado por:

$$p(X|\Theta) = \prod_{i=1}^n \sum_{j=1}^m \omega_j p(x_i|\theta_j) \quad (2.5)$$

O en forma de logaritmo de la verosimilitud,

$$\log p(X|\Theta) = \sum_{i=1}^n \log \sum_{j=1}^m \omega_j p(x_i|\theta_j) \quad (2.6)$$

En estadística, la estimación por máxima verosimilitud (*MLE*) es un enfoque estadístico muy importante para estimar los parámetros, que considera como mejor estimación a aquella que maximiza el logaritmo de la verosimilitud, es decir,

$$\Theta_{MLE} = \underset{\Theta}{arg \max} \{ \log p(X|\Theta) \} \quad (2.7)$$

Resolver 2.7 puede resultar muy complejo, pues la sumatoria dentro del logaritmo de 2.6 impide que el logaritmo actúe directamente sobre la distribución conjunta, dando como resultado expresiones complicadas para hallar su solución. Un método eficiente para encontrar las soluciones que maximiza la verosimilitud se conoce como algoritmo de esperanza-maximización (*EM*).

El algoritmo EM es un método iterativo popular para calcular estimaciones de los parámetros por máxima verosimilitud para problemas probabilísticos con variables latentes (*datos no observables o perdidos*). El algoritmo EM es un método de optimización iterativo que empieza inicializando los valores de los parámetros del modelo, Luego, sigue dos pasos, el paso de esperanza (*E*) y el paso de maximización (*M*).

1. **Paso E.** Se calcula la esperanza de la verosimilitud mediante la inclusión de las variables latentes como si fueran observables.
2. **Paso M.** Se calculan los estimadores de máxima verosimilitud de los parámetros mediante la verosimilitud esperada del paso E

Estos dos pasos se repiten hasta que algún criterio de convergencia sea satisfecha. En cada iteración, se garantiza que el algoritmo EM aumenta el valor de la verosimilitud. Así lo aseguran Aggarwal and Reddy (2013).

El algoritmo EM es muy popular debido a su implementación simple con el aumento monótono garantizado de la verosimilitud durante la optimización. Sin embargo, existen varias limitaciones, como: la existencia de múltiples máximos locales; función de verosimilitud no acotada; convergencia lenta; dependencia de los valores iniciales para el algoritmo iterativo, lo que consecuentemente produce estimaciones de máxima verosimilitud subóptimas y además, el número de grupos es desconocido inicialmente. Salas González (2008), Peñalver Benavent (2007) y Gómez Losada (2014) mencionan algunas alternativas para solucionar dicho problema, entre ellas están: determinación aleatoria de valores iniciales de los parámetros de entrada e incluso se podría generar múltiples conjuntos de valores iniciales, para posteriormente, seleccionar la estimación más probable; y generar previamente clústers con alguna técnica de segmentación.

Ejemplo 2.2. En el conjunto de figuras 2.1 se presenta ejemplos de modelos de mezclas que se obtienen al variar sus parámetros. Estas figuras fueron simuladas usando 3 componentes con distribuciones gaussianas, donde los parámetros se detallan en el cuadro 2.1.

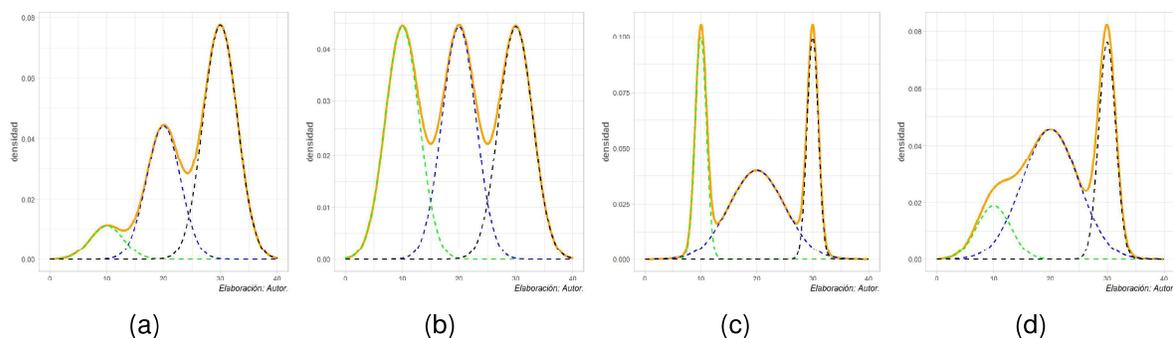


Figura 2.1: Ejemplos de modelos de mezclas gaussianas con 3 componentes.

	$(\omega_1, \omega_2, \omega_3)$	(μ_1, μ_2, μ_3)	$(\sigma_1, \sigma_2, \sigma_3)$
figura 2.1(a)	(1/12, 1/3, 7/12)	(10.0, 20.0, 30.0)	(3.0, 3.0, 3.0)
figura 2.1(b)	(1/3, 1/3, 1/3)	(10.0, 20.0, 30.0)	(3.0, 3.0, 3.0)
figura 2.1(c)	(1/4, 1/2, 1/4)	(10.0, 20.0, 30.0)	(1.0, 5.0, 1.0)
figura 2.1(d)	(1/7, 4/7, 2/7)	(10.0, 20.0, 30.0)	(3.0, 5.0, 1.5)

Cuadro 2.1: Parámetros de las figuras de los modelos de mezclas 2.1

2.2. Cadenas de Markov

En esta sección se toman como principales referencias a Rojo and Miranda (2009, cap. 1), Rincón (2012, cap. 1), Ross (2010, cap. 4) y Zucchini et al. (2016, cap. 1).

“Un proceso estocástico es un modelo matemático que describe el comportamiento de un sistema dinámico, sometido a un fenómeno de naturaleza aleatoria. La presencia de un fenómeno aleatorio hace que el sistema evolucione según un parámetro, que normalmente es el tiempo cambiando probabilísticamente de estado. En otras palabras, al realizar una serie de observaciones del proceso, en diferentes ocasiones y bajo idénticas condiciones, los resultados de las observaciones serán, en general, diferentes” (Rojo and Miranda, 2009).

Definición 2.2 (Proceso estocástico). *Un proceso estocástico es una colección de variables aleatorias $\{X_t : t \in T\}$ parametrizada por un conjunto T , llamado espacio parametral, en donde las variables toman valores en un conjunto C llamado espacio de estados. (Rincón, 2012)*

De acuerdo con la naturaleza del espacio de estados C y del espacio parametral T , los procesos estocásticos tienen diferentes clasificaciones. En este trabajo se considera que el espacio de estados es contable, pues representa los niveles de riesgo de la cartera de crédito (*niveles altos de riesgo denotarían que el sistema financiero se encuentra cerca de una crisis. Recíprocamente, niveles bajos denotarían que el sistema financiero es saludable y por tanto, está lejos de una crisis*). Se considera además que espacio parametral es discreto, pues los registros de la morosidad de la cartera de crédito son de frecuencia mensual.

Se dice que un proceso es a tiempo discreto si el espacio parametral es discreto, y se denota por $\{X_n : n = 0, 1, 2, \dots\}$. De manera similar, un proceso es a tiempo continuo si el espacio parametral es continuo, y se denota por $\{X_t : t \geq 0\}$.

Un tipo de proceso estocástico muy común es el proceso de Markov donde suponiendo conocido el estado presente se puede determinar el estado futuro del sistema, mientras que los estados pasados no tienen influencia en los estados futuros. Esta condición se conoce como propiedad de Markov y puede expresarse de la siguiente forma: Para cualesquiera estados x_1, \dots, x_{n-1} (pasado), x_n (presente), x_{n+1} (futuro), se cumple la igualdad:

$$P(X_{n+1} = x_{n+1} | X_1 = x_1, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n) \quad (2.8)$$

De esta forma “la probabilidad del evento futuro ($X_{n+1} = x_{n+1}$) sólo depende del evento presente ($X_n = x_n$), mientras que la información correspondiente al evento pasado ($X_1 = x_1, \dots, X_{n-1} = x_{n-1}$) es irrelevante.” (Rincón, 2012)

Definición 2.3 (Cadena de Markov). Una cadena de Markov a tiempo discreto $\{X_n : n = 1, 2, \dots\}$ es un proceso estocástico con espacio de estados contable que satisface la propiedad de Markov 2.8.

En la figura 2.2 se muestra la arquitectura que tiene una cadena de Markov, donde el pasado y el futuro dependen únicamente del presente.

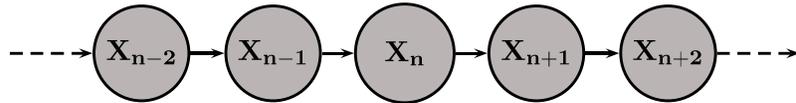


Figura 2.2: Arquitectura de una cadena de Markov de primer orden

La condición 2.8 es equivalente a poder calcular la distribución conjunta de las variables X_1, \dots, X_n de la siguiente manera:

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_1) \cdot P(X_2 = x_2 | X_1 = x_1) \cdot \dots \cdot P(X_n = x_n | X_{n-1} = x_{n-1}) \quad (2.9)$$

2.2.1. Probabilidades de transición

Sean i y j dos estados de la cadena de Markov. La probabilidad

$$\gamma_{ij}(n, n+1) = P(X_{n+1} = j | X_n = i) \quad (2.10)$$

representa la probabilidad de transición del estado i en el tiempo n al estado j al tiempo $n + 1$. Estas probabilidades se conocen como probabilidades de transición a un paso. En general, las probabilidades de transición no solo dependen de los estados, sino también del instante en el cual se efectúan. Cuando dichas probabilidades son independientes del tiempo n ; es decir, pueden tomar cualquier valor, se dice que la cadena de Markov es homogénea en el tiempo, y se denota por:

$$\gamma_{ij} = P(X_{n+1} = j | X_n = i) \quad (2.11)$$

Al variar los índices i y j sobre el conjunto de estados C se obtiene la matriz de probabilidades de transición en un paso, denotada por $\Gamma = (\gamma_{ij})$. Donde la entrada (i, j) de esta matriz es la probabilidad de pasar del estado i al estado j en una unidad de tiempo.

$$\Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \dots \\ \gamma_{21} & \gamma_{22} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (2.12)$$

Proposición 2.1. *La matriz de probabilidades de transición $\Gamma = (\gamma_{ij})$ cumple las siguientes propiedades:*

$$\gamma_{ij} \geq 0, \quad \forall i, j \in C \quad y \quad \sum_j \gamma_{ij} = 1, \quad \forall i \in C$$

Rincón (2012), menciona que “en general, toda matriz cuadrada que cumpla estas dos condiciones se dice que es una matriz estocástica. Si además, la matriz satisface $\sum_i \gamma_{ij} = 1, \forall j \in C$, entonces se dice que es doblemente estocástica.”

Estimación de las probabilidades de transición

Las estimaciones de las probabilidades de la matriz de transición se obtienen a partir de la función de verosimilitud donde se define la variable f_{ij} que denota el número de transiciones observadas del estado i al estado j , para una cadena de Markov con m estados $C = \{1, 2, \dots, m\}$. De acuerdo a Küchler et al. (1997), la función de verosimilitud está dada por:

$$L = \prod_{i=1}^m \prod_{j=1}^m \gamma_{ij}^{f_{ij}} \quad (2.13)$$

o de manera equivalente, aplicando el logaritmo queda como:

$$\mathcal{L} = \sum_{i=1}^m \left(\sum_{j=1}^m f_{ij} \log \gamma_{ij} \right) \quad (2.14)$$

La maximización se la realiza bajo la restricción de la propiedad de Markov $\sum_{j=1}^m \gamma_{ij} = 1, \forall i = 1, 2, \dots, m$, y cuya solución está dada por:

$$\hat{\gamma}_{ij} = f_{ij} / \sum_{j=1}^m f_{ij} \quad \forall i, j = 1, 2, \dots, m \quad (2.15)$$

2.2.2. Probabilidades de transición en n pasos

La probabilidad $P(X_{m+n} = j | X_m = i)$ corresponde a la probabilidad de pasar del estado i al tiempo m al estado j al tiempo $m+n$. Dada la suposición de homogeneidad en el tiempo, esta probabilidad no depende de m , por lo que coincide con $P(X_n = j | X_1 = i)$, y se denota por $\gamma_{ij}(n)$ o también $\gamma_{ij}^{(n)}$, y se la denomina probabilidad de transición en n pasos.

Al variar los índices i y j se obtiene la matriz de probabilidad de transición en n pasos que se denota por $\Gamma(n) = (\gamma_{ij}(n))$ o $\Gamma^{(n)} = (\gamma_{ij}^{(n)})$:

$$\Gamma(n) = \begin{pmatrix} \gamma_{11}(n) & \gamma_{12}(n) & \dots \\ \gamma_{21}(n) & \gamma_{22}(n) & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (2.16)$$

Cuando $n = 0$, se define $\gamma_{ij}(0) = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases}$ (función delta de Kronecker). Es decir, después de realizar cero pasos, la cadena no puede estar en otro lugar mas que es su estado de partida.

Ecuación de Chapman-Kolmogorov

Rincón (2012), la define como “una fórmula sencilla y muy útil que permite descomponer la probabilidad de pasar del estado i al estado j en n pasos, en la suma de probabilidades de las trayectorias que van de i a j , y que atraviesan por un estado k ”

cualquiera, en el tiempo intermedio r .”

Proposición 2.2 (Ecuación de Chapman-Kolmogorov). *Para cualesquiera par de números enteros r y n tales que $0 \leq r \leq n$, y para cualesquiera estados i y j se cumple*

$$\gamma_{ij}(n) = \sum_k \gamma_{ik}(r) \gamma_{kj}(n-r) \quad (2.17)$$

Una consecuencia importante de la ecuación de Chapman-Kolmogorov es que la probabilidad de transición en n pasos, $\gamma_{ij}(n)$, está dado por la entrada (i, j) de la n -ésima potencia de la matriz Γ ; es decir, $\gamma_{ij}(n) = (\Gamma^n)_{ij}$.

En palabras, este resultado establece que el difícil problema de calcular las probabilidades de transición en n pasos se transforma en obtener la n -ésima potencia de la matriz de probabilidades de transición en un paso; es decir,

$$\Gamma(n) = \begin{pmatrix} \gamma_{11}(n) & \gamma_{12}(n) & \cdots \\ \gamma_{21}(n) & \gamma_{22}(n) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots \\ \gamma_{21} & \gamma_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}^n = \Gamma^n \quad (2.18)$$

Si en la ecuación 2.17 se toma $n = r + s$, se tiene

$$\gamma_{ij}(r+s) = \sum_k \gamma_{ik}(r) \gamma_{kj}(s) = \sum_k \gamma_{kj}(r) \gamma_{ik}(s)$$

Lo que expresado de forma matricial queda de la siguiente manera:

$$\Gamma^{(r+s)} = \Gamma^{(r)} \Gamma^{(s)} \quad (2.19)$$

Cuando una matriz estocástica Γ es diagonalizable; es decir, cuando puede ser descrita en la forma QDQ^{-1} donde D es una matriz diagonal; las potencias de Γ se calculan fácilmente, pues $\Gamma^n = QD^nQ^{-1}$. Como D es diagonal, D^n es la matriz con cada elemento de la diagonal elevado a la n -ésima potencia.

2.2.3. Distribución inicial

“En general, una cadena de Markov inicia su evolución partiendo de un estado i cualquiera, de manera general, considerando una distribución de probabilidad inicial sobre el espacio de estados. Una distribución inicial para una cadena de Markov con

espacio de estados $\{1, 2, \dots, m\}$ es simplemente una distribución de probabilidad sobre este conjunto; es decir, una colección de números (u_1, u_2, \dots, u_m) no negativos que suman 1. $u_i = P(X_1 = i)$ corresponde a la probabilidad de que la cadena inicie en el estado i ." (Rincón, 2012)

2.2.4. Evolución de las distribuciones

Consideremos un espacio de estados finito $C = \{1, 2, \dots, m\}$, y denotemos a la distribución de probabilidad inicial $u(1) = (u_1(1), u_2(1), \dots, u_m(1))$. Luego de transcurrido la primera unidad de tiempo, la cadena se encuentra en cualquiera de sus posibles estados de acuerdo a la distribución $u(2) = (u_1(2), u_2(2), \dots, u_m(2))$, donde la j -ésima entrada de ese vector es:

$$u_j(2) = P(X_2 = j) = \sum_{i=1}^m P(X_2 = j | X_1 = i) P(X_1 = i) = \sum_{i=1}^m u_i(1) \gamma_{ij}$$

Es decir, el vector $u(2)$ se obtiene a partir del vector $u(1)$ y de la matriz de probabilidades de transición Γ a través del producto $u(2) = u(1)\Gamma$; es decir,

$$(u_1(2), u_2(2), \dots, u_m(2)) = (u_1(1), u_2(1), \dots, u_m(1)) \begin{pmatrix} \gamma_{11} & \dots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \dots & \gamma_{mm} \end{pmatrix}$$

A su vez el vector $u(2)$ se transforma en el vector $u(3)$ a través de la ecuación $u(3) = u(2)\Gamma = u(1)\Gamma^2$, y así sucesivamente. En general, para $n \geq 1$,

$$u(n) = u(n-1)\Gamma = u(1)\Gamma^{n-1} \quad (2.20)$$

De esta forma se obtiene una sucesión de distribuciones de probabilidad $u(1), u(2), \dots, u(m)$, donde cada una de ellas, excepto la primera, se obtiene de la anterior multiplicada por la matriz de probabilidades de transición a un paso.

2.2.5. Distribuciones estacionarias

Definición 2.4. (Rincón, 2012) Una distribución de probabilidad $\delta = (\delta_1, \dots, \delta_m)$ es estacionaria o invariante para una cadena de Markov con matriz de probabilidades de

transición $\Gamma = (\gamma_{ij})$ si

$$\delta_j = \sum_i \delta_i \gamma_{ij}$$

En términos matriciales, la distribución de probabilidad δ es estacionaria si $\delta = \delta\Gamma$ y además $\sum_{j=1}^m \delta_j = 1$. Esta identidad tiene como consecuencia, que para cualquier número natural n se cumple que $\delta = \delta\Gamma^n$; es decir, δ es también una distribución estacionaria para la matriz Γ^n . Esto significa que si la variable aleatoria inicial X_1 tiene distribución δ , entonces la distribución de X_n también es δ ; es decir, esta distribución no cambia con el paso del tiempo.

Definición 2.5 (Accesibilidad y Comunicación). (Rincón, 2012) *Se dice que un estado j es accesible desde el estado i si existe un número entero $n \geq 0$ tal que $\gamma_{ij}(n) > 0$, ésto se describe como $i \rightarrow j$. Se dice además que los estados i y j se comunican, y se escribe como $i \leftrightarrow j$, si se cumple que $i \rightarrow j$ y $j \rightarrow i$.*

Definición 2.6 (Cadena irreductible). (Rincón, 2012) *Se dice que una cadena de Markov es irreductible si todos los estados se comunican entre sí.*

Proposición 2.3. (Rincón, 2012) *Una cadena de Markov irreductible, homogénea, a tiempo discreto con espacio de estados finito tiene una única y estrictamente positiva distribución estacionaria.*

2.2.6. Estructura Gráfica de las cadenas de Markov

Una cadena de Markov está totalmente determinada por la distribución inicial $u(1)$ y la matriz de probabilidades Γ . De esta manera, la matriz de probabilidades puede ser descrita como un grafo dirigido, donde cada nodo representa un estado y la probabilidad de transición determina la existencia de un arco (si la probabilidad de transición es cero, el arco no existe). Esto permite observar el comportamiento de la cadena de Markov.

Ejemplo 2.3. *Se considera una cadena de Markov con 3 estados, con matriz de probabilidades de transición Γ . Su representación gráfica está dada por la figura 2.3.*

2.2.7. Cadenas de Markov de orden superior

“En los casos donde las observaciones en un proceso con espacio de estados contable no satisfacen la propiedad de Markov 2.8, se sugiere utilizar una cadena de

$$\mathbf{\Gamma} = \begin{matrix} & \text{estado 1} & \text{estado 2} & \text{estado 3} \\ \text{estado 1} & \begin{pmatrix} 0,80 & 0,20 & 0,00 \\ 0,16 & 0,70 & 0,14 \\ 0,02 & 0,05 & 0,84 \end{pmatrix} \\ \text{estado 2} & \\ \text{estado 3} & \end{matrix}$$

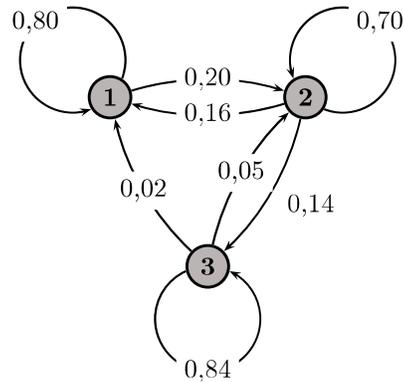


Figura 2.3: Representación gráfica de una matriz de transición asociada a la cadena de Markov.

Markov de orden superior; es decir, un proceso $\{X_n : 1, 2, \dots\}$ que satisfaga la siguiente generalización de la propiedad de Markov para algún $l \geq 2$: (Zucchini et al., 2016)

$$P(X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_1 = x_1) = P(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_{n-l+1} = x_{n-l+1}) \quad (2.21)$$

De este modo para $l = 2$, se tiene una cadena de Markov de segundo orden donde las probabilidades de transición satisfacen la siguiente ecuación:

$$\gamma_{ijk} = P(X_{n+1} = k | X_n = j, X_{n-1} = i) \quad (2.22)$$

Es decir, la expresión 2.22 establece que la distribución de probabilidad del estado del proceso al tiempo futuro $n + 1$ depende únicamente de los estados del proceso al tiempo n y $n - 1$, y no de los estados en los tiempos pasados $1, \dots, n - 2$. Una secuencia de observaciones que obedece a este modelo se observa en la figura 2.4.

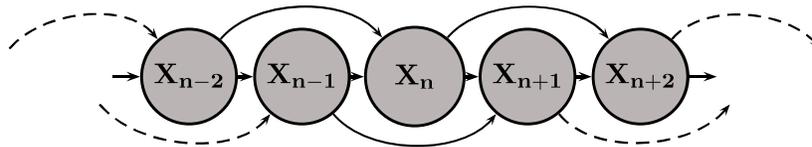


Figura 2.4: Arquitectura de una cadena de Markov de segundo orden

2.3. Modelos ocultos de Markov

En esta sección se toman como principales referencias a Zucchini et al. (2016), Dymarski (2011) y Rabiner and Juang (1986).

Los modelos ocultos de Markov han sido ampliamente usados en aplicaciones de procesamiento de señales, especialmente en el contexto de reconocimiento automático de la voz, pero su teoría ha sido expandido a otros campos, como por ejemplo: bioinformática, finanzas, ecología, medicina, etc.

De acuerdo a Dymarski (2011), un modelo oculto de Markov es resumido como un tipo de proceso estocástico doble con los siguientes aspectos:

1. El primer proceso estocástico es una cadena de Markov (*se denotará por $\{C_n\}$*) con espacio contable de estados (*no visibles*), donde cada uno de ellos está generalmente asociado con una distribución de probabilidad.
2. El segundo proceso observable (*se denotará por $\{X_n\}$*), donde cada una de las observaciones (*visibles*) es generada por la distribución de algún estado (*oculto*) de la cadena de Markov asociada en cada instante de tiempo.

Definición 2.7 (Modelo oculto de Markov). *Sea $\{X_n : n \in \mathbb{N}\}$ un modelo oculto de Markov y $\{C_n : n \in \mathbb{N}\}$ una cadena de Markov asociada. Se define $\mathbf{X}^{(n)} = \{X_1, X_2, \dots, X_n\}$ y $\mathbf{C}^{(n)} = \{C_1, C_2, \dots, C_n\}$ como sus historias del tiempo 1 al tiempo n . Se resume el modelo por:*

$$P(C_n = c_n | \mathbf{C}^{(n-1)}) = P(C_n = c_n | C_{n-1} = c_{n-1}), \quad n = 2, 3, \dots \quad (2.23)$$

$$P(X_n = x_n | \mathbf{X}^{(n-1)}, \mathbf{C}^{(n)}) = P(X_n = x_n | C_n = c_n), \quad n \in \mathbb{N} \quad (2.24)$$

El modelo consiste de dos partes: primero, un proceso de parámetros no observados $\{C_n : n = 1, 2, \dots\}$ que satisface la propiedad de Markov; y segundo, un proceso dependiente del estado $\{X_n : n = 1, 2, \dots\}$, donde la distribución de X_n depende del estado actual C_n y no de estados u observaciones previas.

El gráfico 2.5 presenta la arquitectura de un modelo oculto de Markov mediante un grafo dirigido, donde C_n representa los estados (*ocultos*) de la cadena de Markov y X_n las observaciones generadas por las distribuciones dependientes del estado.

En lo que sigue, se asume que la cadena de Markov $\{C_n\}$ tiene m estados. De esta manera, se define p_i con $i = 1, 2, \dots, m$ como la función de probabilidad de X_n si la

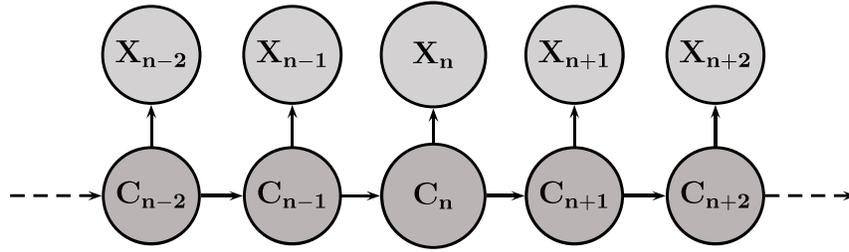


Figura 2.5: Arquitectura de un modelo oculto de Markov de primer orden

cadena de Markov se encuentra en el estado i al tiempo n ; es decir,

$$p_i(x) = P(X_n = x | C_n = i) \quad (2.25)$$

Para el caso continuo, p_i denota la densidad de probabilidad de X_n asociado al estado i . De manera general, se refiere a p_i como la distribución que depende del estado.

Distribuciones marginales

Distribuciones univariantes. Para X_n se define $u_i(n) = P(C_n = i)$ con $n = 1, 2, \dots, N$ e $i = 1, 2, \dots, m$ y se tiene:

$$P(X_n = x) = \sum_{i=1}^m P(C_n = i)P(X_n = x | C_n = i) = \sum_{i=1}^m u_i(n)p_i(x) \quad (2.26)$$

Expresada en forma matricial queda de la siguiente manera:

$$P(X_n = x) = (u_1(n), \dots, u_m(n)) \cdot \begin{pmatrix} p_1(x) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & p_m(x) \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \mathbf{u}(n)\mathbf{P}(x)\mathbf{1}' \quad (2.27)$$

donde $\mathbf{P}(x)$ es una matriz diagonal cuyo i -ésimo elemento es $p_i(x)$. Además, por la ecuación 2.20 se tiene que $\mathbf{u}(n) = \mathbf{u}(1)\mathbf{\Gamma}^{n-1}$. Por lo tanto:

$$P(X_n = x) = \mathbf{u}(1)\mathbf{\Gamma}^{n-1}\mathbf{P}(x)\mathbf{1}' \quad (2.28)$$

Distribuciones bivariantes. La distribución de un conjunto de variables aleatorias V_i está dada por:

$$P(V_1, V_2, \dots, V_n) = \prod_{i=1}^n P(V_i | pa(V_i)) \quad (2.29)$$

donde $pa(V_i)$ denota todos los padres o antecesores de V_i en el conjunto V_1, V_2, \dots, V_n . Por ejemplo si tenemos las variables $X_n, X_{n+k}, C_n, C_{n+k}$ para k un entero positivo, donde C_t no tienen padres o antecesores, $pa(X_n) = \{C_n\}$, $pa(C_{n+k}) = C_n$ y $pa(X_{n+k}) = \{C_{n+k}\}$, tenemos:

$$P(X_n, X_{n+k}, C_n, C_{n+k}) = P(C_n)P(X_n|C_n)P(C_{n+k}|C_n)P(X_{n+k}|C_{n+k})$$

Por lo que, se tiene que,

$$\begin{aligned} P(X_n = v, X_{n+k} = w) &= \sum_{i=1}^m \sum_{j=1}^m P(X_n = v, X_{n+k} = w, C_n = i, C_{n+k} = j) \\ &= \sum_{i=1}^m \sum_{j=1}^m P(C_n = i) p_i(v) P(C_{n+k} = j | C_n = i) p_j(w) \\ &= \sum_{i=1}^m \sum_{j=1}^m u_i(t) p_i(v) \gamma_{ij}(k) p_j(w) \end{aligned}$$

Expresada en forma matricial queda de la siguiente manera:

$$P(X_n = v, X_{n+k} = w) = \mathbf{u}(n) \mathbf{P}(v) \mathbf{\Gamma}^k \mathbf{P}(w) \mathbf{1}' \quad (2.30)$$

La verosimilitud

Zucchini et al. (2016) muestra que la expresión para la verosimilitud L_N de N observaciones consecutivas x_1, x_2, \dots, x_N que se asumen fueron generadas por un modelo de oculto de Markov de m estados, que tiene una distribución inicial δ y una matriz de probabilidades de transición $\mathbf{\Gamma}$ para la cadena de Markov, y una función de probabilidad dependiente del estado p_i ; está dada por la siguiente proposición.

Proposición 2.4. *La verosimilitud de un modelo oculto de Markov está dada por*

$$L_N = \delta \mathbf{P}(x_1) \mathbf{\Gamma} \mathbf{P}(x_2) \mathbf{\Gamma} \mathbf{P}(x_3) \cdots \mathbf{\Gamma} \mathbf{P}(x_N) \mathbf{1}' \quad (2.31)$$

Demostración. Nótese que

$$L_N = P(\mathbf{X}^{(N)} = \mathbf{x}^{(N)}) = \sum_{c_1, c_2, \dots, c_N=1}^m P(\mathbf{X}^{(N)} = \mathbf{x}^{(N)}, \mathbf{C}^{(N)} = \mathbf{c}^{(N)})$$

y por la ecuación 2.29,

$$P(\mathbf{X}^{(N)}, \mathbf{C}^{(N)}) = P(C_1) \prod_{k=2}^N P(C_k | C_{k-1}) \prod_{k=1}^N P(X_k | C_k) \quad (2.32)$$

De donde se sigue,

$$\begin{aligned} L_N &= \sum_{c_1, c_2, \dots, c_N=1}^m (\delta_{c_1} \gamma_{c_1, c_2} \gamma_{c_2, c_3} \cdots \gamma_{c_{N-1}, c_N}) (p_{c_1}(x_1) p_{c_2}(x_2) \cdots p_{c_N}(x_N)) \\ &= \sum_{c_1, c_2, \dots, c_N=1}^m \delta_{c_1} p_{c_1}(x_1) \gamma_{c_1, c_2} p_{c_2}(x_2) \gamma_{c_2, c_3} \cdots \gamma_{c_{N-1}, c_N} \cdots p_{c_N}(x_N) \\ &= \delta \mathbf{P}(x_1) \Gamma \mathbf{P}(x_2) \Gamma \mathbf{P}(x_3) \cdots \Gamma \mathbf{P}(x_N) \mathbf{1}' \end{aligned}$$

■

De las expresiones anteriores se deduce que el cálculo de L_N , por medio de la ecuación 2.31, involucra la suma de m^N términos donde cada uno es un producto de $2N$ factores; es decir, se requiere de $O(Nm^N)$ operaciones para obtener L_N , lo cual es ineficiente excepto para valores de N pequeños.

Una solución al problema de obtener la verosimilitud por medio de la ecuación 2.31 es el algoritmo de avance, que consiste en cálculos recursivos, lo que resulta computacionalmente más eficiente.

Para indicar el algoritmo de avance, se define el vector α_n , para $n = 1, 2, \dots, N$, por

$$\alpha_n = \delta \mathbf{P}(x_1) \Gamma \mathbf{P}(x_2) \Gamma \mathbf{P}(x_3) \cdots \Gamma \mathbf{P}(x_n) = \delta \mathbf{P}(x_1) \prod_{s=2}^n \Gamma \mathbf{P}(x_s) \quad (2.33)$$

A los elementos del vector α_n se denominan probabilidades de avance. Con la convención de que el producto vacío es la matriz identidad, se sigue inmediatamente que:

$$L_N = \alpha_N \mathbf{1}' \quad y \quad \alpha_n = \alpha_{n-1} \Gamma \mathbf{P}(x_n), \quad para \quad n \geq 2$$

Consecuentemente, se establece el algoritmo para calcular la verosimilitud L_N como:

Algoritmo 2.1: Cálculo de la verosimilitud L_N

Data: x_1, x_2, \dots, x_N

Result: L_N

- 1 Inicializar $\alpha_1 = \delta \mathbf{P}(x_1)$;
 - 2 **for** $n = 2, 3, \dots, N$ **do**
 - 3 $\alpha_n = \alpha_{n-1} \delta \mathbf{P}(x_n)$;
 - 4 $L_N = \alpha_N \mathbf{1}'$
-

De esta forma, el cálculo de L_N involucra $O(Nm^2)$ operaciones, que es eficiente cuando N es muy grande; es decir, para cada valor de n dentro del lazo, hay m elementos de α_n por calcular y cada uno es la suma de m productos de 3 cantidades (α_{n-1} , γ_{ij} y $p_j(x_n)$).

2.3.1. Estimación de los parámetros por el algoritmo (EM)

Una forma para estimar los parámetros de un modelo oculto de Markov (δ , Γ y los parámetros de la distribución que dependen del estado) es por medio de la maximización directa de la verosimilitud con respecto a los parámetros; es decir, se podría derivar la función de verosimilitud considerando las restricciones correspondientes de los parámetros, pero este enfoque es demasiado complejo debido al gran número de cálculos que involucra. Además, la función de verosimilitud presenta otros problemas como poseer múltiples máximos locales o no estar acotada.

Un método comúnmente usado para encontrar los estimadores de máxima verosimilitud de los modelos ocultos de Markov es el algoritmo (EM). En el contexto de los modelos ocultos de Markov el algoritmo (EM) es conocido como el algoritmo de Baum-Welch.

Probabilidades de avance y retroceso

Se define el vector de probabilidades de avance α_n , para $n = 1, 2, \dots, N$ como:

$$\alpha_n = \delta\mathbf{P}(x_1)\Gamma\mathbf{P}(x_2) \dots \Gamma\mathbf{P}(x_n) = \delta\mathbf{P}(x_1) \prod_{s=2}^n \Gamma\mathbf{P}(x_s) \quad (2.34)$$

Y el vector de probabilidades de retroceso β_n , para $n = 1, 2, \dots, N$ como:

$$\beta'_n = \Gamma\mathbf{P}(x_{n+1})\Gamma\mathbf{P}(x_{n+2}) \dots \Gamma\mathbf{P}(x_N)\mathbf{1}' = \left(\prod_{s=n+1}^N \Gamma\mathbf{P}(x_s) \right) \mathbf{1}' \quad (2.35)$$

Donde δ denota la distribución inicial de la cadena de Markov. Con la convención de que el producto vacío es la matriz identidad tenemos que para $n = 1$, $\alpha_1 = \delta\mathbf{P}(x_1)$ y para $n = N$, $\beta_N = \mathbf{1}$

Probabilidades de avance. Se sigue inmediatamente de la definición de α_n que, para $n = 1, 2, \dots, N - 1$, $\alpha_{n+1} = \alpha_n\Gamma\mathbf{P}(x_{n+1})$ o, en su forma escalar,

$$\alpha_{n+1}(j) = \sum_{i=1}^m \alpha_n(i)\gamma_{ij}p_j(x_{n+1}) \quad (2.36)$$

Proposición 2.5. Para $n = 1, 2, \dots, N$ y $j = 1, 2, \dots, m$,

$$\alpha_n(j) = P(\mathbf{X}^{(n)} = \mathbf{x}^{(n)}, C_n = j) \quad (2.37)$$

Demostración. Para la demostración usaremos el siguiente resultado,

$$P(\mathbf{X}^{(n+1)} = \mathbf{x}^{(n+1)}, C_n = i, C_{n+1} = j) = P(\mathbf{X}^{(n)} = \mathbf{x}^{(n)}, C_n = i) \\ P(C_{n+1} = j | C_n = i)P(X_{n+1} = x_{n+1} | C_{n+1} = j) \quad (2.38)$$

Por inducción, para $n = 1$ tenemos que $\alpha_1 = \delta\mathbf{P}(x_1)$. Por tanto,

$$\alpha_1(j) = \delta_j p(x_1) = P(C_1 = j)P(X_1 = x_1 | C_1 = j) = P(X_1 = x_1, C_1 = j),$$

Suponemos que la proposición se cumple para $n \in \mathbb{N}$, se probará que se cumple para $n + 1$:

$$\begin{aligned}
\alpha_{n+1}(j) &= \sum_{i=i}^m \alpha_n(i) \gamma_{ij} p_j(x_{n+1}) \\
&= \sum_{i=i}^m P(\mathbf{X}^{(n)} = \mathbf{x}^{(n)}, C_n = i) P(C_{n+1} = j | C_n = i) P(X_{n+1} = x_{n+1} | C_{n+1} = j) \\
&= \sum_{i=i}^m P(\mathbf{X}^{(n+1)} = \mathbf{x}^{(n+1)}, C_n = i, C_{n+1} = j) \\
&= P(\mathbf{X}^{(n+1)} = \mathbf{x}^{(n+1)}, C_{n+1} = j)
\end{aligned}$$

■

Probabilidades de retroceso. Se sigue inmediatamente de la definición de β_n que, para $n = 1, 2, \dots, N - 1$, $\beta'_n = \mathbf{\Gamma P}(x_{n+1})\beta'_{n+1}$ o, en su forma escalar,

$$\beta_n(j) = \sum_{i=1}^m \gamma_{ij} p_i(x_{n+1}) \beta_{n+1}(i) \quad (2.39)$$

Proposición 2.6. Para $n = 1, 2, \dots, N - 1$ e $i = 1, 2, \dots, m$,

$$\beta_n(i) = P(X_{n+1} = x_{n+1}, X_{n+2} = x_{n+2}, \dots, X_N = x_N | C_n = i), \quad (2.40)$$

En una notación más compacta,

$$\beta_n(i) = P(\mathbf{X}_{n+1}^N = \mathbf{x}_{n+1}^N | C_n = i) \quad (2.41)$$

donde X_a^b denota el vector $(X_a, X_{a+1}, \dots, X_b)$

Demostración. Para la demostración usaremos los siguientes resultados,

$$P(X_{n+1} = x_{n+1} | C_{n+1} = i) P(\mathbf{X}_{n+2}^N = \mathbf{x}_{n+2}^N | C_{n+1} = i) = P(\mathbf{X}_{n+1}^T = \mathbf{x}_{n+1}^N | C_{n+1} = i) \quad (2.42)$$

y

$$P(\mathbf{X}_{n+1}^N = \mathbf{x}_{n+1}^N | C_{n+1} = i) = P(\mathbf{X}_{n+1}^N = \mathbf{x}_{n+1}^N | C_n = j, C_{n+1} = i) \quad (2.43)$$

Por inducción, para $n = N - 1$ se tiene que, $\beta'_{N-1} = \Gamma P(x_N) \mathbf{1}'$. Por tanto,

$$\begin{aligned}
 \beta_{N-1}(i) &= \sum_j P(C_N = j | C_{N-1} = i) P(X_N = x_N | C_N = j) \\
 &= \sum_j P(C_N = j | C_{N-1} = i) P(X_N = x_N | C_{N-1} = i, C_N = j) \\
 &= \sum_j \frac{P(X_N = x_N, C_{N-1} = i, C_N = j)}{P(C_{N-1} = i)} \\
 &= \frac{1}{P(C_{N-1} = i)} \sum_j P(X_N = x_N, C_{N-1} = i, C_N = j) \\
 &= \frac{P(X_N = x_N, C_{N-1} = i)}{P(C_{N-1} = i)} \\
 &= P(X_N = x_N | C_{N-1} = i)
 \end{aligned}$$

Suponemos que la proposición se cumple para $n + 1$, se probará que se cumple para n :

$$\begin{aligned}
 \beta_n(i) &= \sum_j \gamma_{ij} p_j(x_{n+1}) \beta_{n+1}(j) \\
 &= \sum_j P(C_{n+1} = j | C_n = i) P(X_{n+1} = x_{n+1} | C_{n+1} = j) P(\mathbf{X}_{n+2}^N = \mathbf{x}_{n+2}^N | C_{n+1} = j) \\
 &= \frac{1}{P(C_n = i)} \sum_j P(\mathbf{X}_{n+1}^N = \mathbf{x}_{n+1}^N, C_n = i, C_{n+1} = j) \\
 &= \frac{P(\mathbf{X}_{n+1}^N = \mathbf{x}_{n+1}^N, C_n = i)}{P(C_n = i)} \\
 &= P(\mathbf{X}_{n+1}^N = \mathbf{x}_{n+1}^N | C_n = i)
 \end{aligned}$$

■

Propiedades de las probabilidades de avance y retroceso. Se establecen resultados que relacionan las probabilidades $\alpha_n(i)$ y $\beta_n(i)$, a las probabilidades $P(\mathbf{X}^{(N)} = \mathbf{x}^{(N)}, C_n = i)$ que son usados en el algoritmo EM para los modelos ocultos de Markov.

Proposición 2.7. Para $n = 1, 2, \dots, N$ e $i = 1, 2, \dots, m$,

$$\alpha_n(i)\beta_n(i) = P(\mathbf{X}^{(N)} = \mathbf{x}^{(N)}, C_n = i) \quad (2.44)$$

y consecuentemente $\alpha_n\beta'_n = P(\mathbf{X}^{(N)} = \mathbf{x}^{(N)}) = L_N$, para cada n .

Demostración. Por las proposiciones anteriores y el siguiente resultado,

$$P(\mathbf{X}_1^N = \mathbf{x}_1^N | C_n = i) = P(\mathbf{X}_1^n = \mathbf{x}_1^n | C_n = i)P(\mathbf{X}_{n+1}^N = \mathbf{x}_{n+1}^N | C_n = i) \quad (2.45)$$

Tenemos que,

$$\begin{aligned} \alpha_n(i)\beta_n(i) &= P(\mathbf{X}_1^n = \mathbf{x}_1^n, C_n = i)P(\mathbf{X}_{n+1}^N = \mathbf{x}_{n+1}^N | C_n = i) \\ &= P(C_n = i)P(\mathbf{X}_1^n = \mathbf{x}_1^n | C_n = i)P(\mathbf{X}_{n+1}^N = \mathbf{x}_{n+1}^N | C_n = i) \\ &= P(C_n = i)P(\mathbf{X}_1^N = \mathbf{x}_1^N | C_n = i) \\ &= P(\mathbf{X}^N = \mathbf{x}^N, C_n = i) \end{aligned}$$

La segunda conclusión se sigue inmediatamente de la expresión matricial para la verosimilitud y la definiciones de α_n y β_n :

$$L_N = \left(\delta \mathbf{P}(x_1) \Gamma \mathbf{P}(x_2) \cdots \Gamma \mathbf{P}(x_n) \right) \cdot \left(\Gamma \mathbf{P}(x_{n+1}) \cdots \Gamma \mathbf{P}(x_N) \mathbf{1}' \right) = \alpha_n \beta'_n$$

■

Proposición 2.8. Tenemos que,

(a) Para $n = 1, 2, \dots, N$,

$$P(C_n = j | \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) = \frac{\alpha_n(j)\beta_n(j)}{L_N} \quad (2.46)$$

(b) Para $n = 2, 3, \dots, N$,

$$P(C_{n-1} = j, C_n = k | \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) = \frac{\alpha_{n-1}(j)\gamma_{jk}p_k(x_n)\beta_n(k)}{L_N} \quad (2.47)$$

Demostración. Se tiene que,

De la ecuación 2.44, tenemos

$$\begin{aligned}
 \text{(b)} \quad P(C_n = j | \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) &= \frac{P(C_n = j, \mathbf{X}^{(N)} = \mathbf{x}^{(N)})}{P(\mathbf{X}^{(N)} = \mathbf{x}^{(N)})} = \frac{\alpha_n(j)\beta_n(j)}{L_N} \\
 P(C_{n-1} = j, C_n = k | \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) &= \frac{1}{L_n} P(\mathbf{X}^{(N)} = \mathbf{x}^{(N)}, C_{n-1} = j, C_n = k) \\
 &= \frac{1}{L_t} P(\mathbf{X}^{(n-1)} = \mathbf{x}^{(n-1)}, C_{n-1} = j) \\
 &\quad P(C_n = k | C_{n-1} = j) P(\mathbf{X}_t^N = \mathbf{x}_t^N | C_n = k) \\
 &= \frac{1}{L_N} \alpha_{n-1}(j) \gamma_{jk} P(X_n = x_n | C_n = k) \\
 &\quad P(\mathbf{X}_{n+1}^N = \mathbf{x}_{n+1}^N | C_n = k) \\
 &= \frac{\alpha_{n-1}(j) \gamma_{jk} p_k(x_n) \beta_n(k)}{L_N}
 \end{aligned}$$

■

El algoritmo EM para los modelos ocultos de Markov

El algoritmo EM es un método iterativo para realizar estimaciones por máxima verosimilitud, cuando los datos son faltantes o perdidos, y explota el hecho que el logaritmo de la verosimilitud de los datos completos (*logaritmo de la verosimilitud de los datos observados y faltantes*) puede ser sencillo de maximizar, incluso si la verosimilitud de los datos observados no lo son. En el caso de los modelos ocultos de Markov, los estados son considerados como valores faltantes (*pues son no observados*). De acuerdo a Zucchini et al. (2016), es conveniente representar la secuencia de estados c_1, c_2, \dots, c_N generadas por la cadena de Markov por variables aleatorias binarias $\{0, 1\}$ definidas como:

$$u_j(n) = 1 \text{ si y solo si } c_n = j \quad (n = 1, 2, \dots, N)$$

y

$$v_{jk}(n) = 1 \text{ si y solo si } c_{n-1} = j \text{ y } c_n = k \quad (n = 2, 3, \dots, N)$$

Con esta notación, el logaritmo de la verosimilitud de datos completos de un modelo oculto de Markov (*el logaritmo de la verosimilitud de las observaciones* x_1, x_2, \dots, x_N

más los valores faltantes c_1, c_2, \dots, c_N) está dado por

$$\begin{aligned} \log (P(\mathbf{x}^{(N)}, \mathbf{c}^{(N)})) &= \log \left(\delta_{c_1} \prod_{n=2}^N \gamma_{c_{n-1}, c_n} \prod_{n=1}^N p_{c_n}(x_n) \right) \\ &= \log \delta_{c_1} + \sum_{n=2}^N \log \gamma_{c_{n-1}, c_n} + \sum_{n=1}^N \log p_{c_n}(x_n) \end{aligned}$$

Por tanto, el logaritmo de la verosimilitud de los datos completos es

$$\begin{aligned} \log (P(\mathbf{x}^{(N)}, \mathbf{c}^{(N)})) &= \underbrace{\sum_{j=1}^m u_j(1) \log \delta_j}_{\text{término 1}} + \underbrace{\sum_{j=1}^m \sum_{k=1}^m \left(\sum_{n=2}^N v_{jk}(n) \right) \log \gamma_{jk}}_{\text{término 2}} \\ &\quad + \underbrace{\sum_{j=1}^m \sum_{n=1}^N u_j(n) \log p_j(x_n)}_{\text{término 3}} \quad (2.48) \end{aligned}$$

donde δ es la distribución inicial de la cadena de Markov, que no necesariamente es una distribución estacionaria.

El algoritmo EM para los modelos ocultos de Markov procede de la siguiente manera:

Algoritmo 2.2: Estimación de los parámetros del modelo oculto de Markov (Baum-Welch)

Data: Valores iniciales de los parámetros δ_0 , Γ_0 y los parámetros de la distribución dependientes del estado θ_0

Result: Estimaciones de los parámetros $\hat{\delta}$, $\hat{\Gamma}$ y los parámetros de la distribución dependientes del estado $\hat{\theta}$

1 **repeat**

2 **PASO E:** Estimamos $v_{jk}(n)$ y $u_j(n)$ por // Dado los parámetros actuales

$$\hat{u}_j(n) = P(C_n = j | \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) = \frac{\alpha_n(j)\beta_n(j)}{L_N} \quad (2.49)$$

$$\hat{v}_{jk}(n) = P(C_{n-1} = j, C_n = k | \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) = \frac{\alpha_{n-1}(j)\gamma_{jk}p_k(x_n)\beta_n(j)}{L_N} \quad (2.50)$$

PASO M: Maximizar el logaritmo de la verosimilitud de datos completos 2.48, con respecto a los tres conjuntos de parámetros; la distribución inicial δ , la matriz de probabilidades de transición Γ y los parámetros de las distribuciones dependientes del estado.;

3 **until** *Hasta que algún criterio de convergencia sea satisfecha;*

La ecuación 2.48 muestra que el **paso M** se divide en tres maximizaciones separadas, ya que los parámetros del término 1 depende solo de la distribución inicial δ , el término 2 de la matriz de probabilidades de transición Γ y el término 3 de los parámetros dependientes del estado. Por tanto se debe maximizar:

1. $\sum_{j=1}^m \hat{u}_j(1) \log \delta_j$ con respecto a δ ;
2. $\sum_{j=1}^m \sum_{k=1}^m \left(\sum_{n=2}^N \hat{v}_{jk}(n) \right) \log \gamma_{jk}$ con respecto a Γ ; y
3. $\sum_{j=1}^m \sum_{n=1}^N \hat{u}_j(n) \log p_j(x_n)$ con respecto a los parámetros que dependen del estado. Notar que los únicos parámetros que dependen el término $\sum_{n=1}^N \hat{u}_j(n) \log p_j(x_n)$ son los de la *j-ésima* distribución p_j dependiente del estado.

La solución queda de la siguiente manera:

1. $\delta_j = \hat{u}_j(1) / \sum_{j=1}^m \hat{u}_j(1) = \hat{u}_j(1)$.

$$2. \gamma_{jk} = \sum_{n=2}^N \hat{v}_{jk}(n) / \sum_{k=1}^m \sum_{n=2}^N \hat{v}_{jk}(n).$$

3. La maximización del tercer término puede resultar fácil o no, de acuerdo a la distribución dependiente del estado asumida. Esencialmente es el problema estándar de estimación por máxima verosimilitud para las distribuciones.

Si se considera un modelo oculto de Markov cuya distribución dependiente del estado es una Poisson; es decir, $p_j(x) = \frac{e^{-\lambda_j} \lambda_j^x}{x!}$; la maximización del tercer término con respecto a λ_j es:

$$\hat{\lambda}_j = \sum_{n=1}^N \hat{u}_j(n)x(n) / \sum_{n=1}^N \hat{u}_j(n) \quad (2.51)$$

Si se considera un modelo oculto de Markov cuya distribución dependiente del estado es una Normal; es decir, $p_j(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu_j)^2}{2\sigma_j^2}\right\}$; la maximización del tercer término con respecto a μ_j y σ_j es:

$$\hat{\mu}_j = \sum_{n=1}^N \hat{u}_j(n)x(n) / \sum_{n=1}^N \hat{u}_j(n) \quad (2.52)$$

y

$$\hat{\sigma}_j^2 = \sum_{n=1}^N \hat{u}_j(n)(x_n - \hat{\mu})^2 / \sum_{n=1}^N \hat{u}_j(n) \quad (2.53)$$

Errores estándar e intervalos de confianza

Las propiedades de los estimadores de máxima verosimilitud de los modelos ocultos de Markov son resultados asintóticos y difíciles de calcular. Para explotar estos resultados se requiere estimar la matriz de varianzas-covarianza de los estimadores de los parámetros. Una alternativa es usar el método de bootstrap paramétrico. La idea de bootstrap paramétrico es evaluar las propiedades del modelo con parámetros Θ usando los del modelo con parámetros $\hat{\Theta}$. Para luego obtener las propiedades de distribución de $\hat{\Theta}$. El algoritmo 2.3 permite estimar la matriz de varianzas-covarianza de $\hat{\Theta}$.

Algoritmo 2.3: Bootstrap paramétrico

Data: Modelo ajustado con parámetros $\hat{\Theta}$

Result: Matriz de varianza-covarianza $\widehat{Var-Cov}(\hat{\Theta})$

- 1 **for** $b = 1, 2, \dots, B$ **do** // B suficientemente grande
 - 2 | Generamos una muestra, (*muestra de bootstrap*), de observaciones del
| modelo ajustado; es decir, del modelo con parámetros $\hat{\Theta}$.; // De
| igual longitud que las observaciones originales
 - 3 | Estimamos los parámetros Θ por $\hat{\Theta}^*$ para cada muestra de bootstrap.;
 - 4 Calculamos la matriz de varianza-covarianza $\widehat{Var-Cov}(\hat{\Theta})$, con 2.54
-

La matriz de varianza-covarianza de $\hat{\Theta}$ es entonces estimada por la matriz de varianza-covarianza muestral de las estimaciones de bootstrap $\hat{\Theta}^*(b)$, $b = 1, 2, \dots, B$.

$$\widehat{Var-Cov}(\hat{\Theta}) = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{\Theta}^*(b) - \hat{\Theta}^*(\cdot) \right)' \left(\hat{\Theta}^*(b) - \hat{\Theta}^*(\cdot) \right) \quad (2.54)$$

donde $\hat{\Theta}^*(\cdot) = \frac{1}{B} \sum_{b=1}^B \hat{\Theta}^*(b)$.

El método de bootstrap permite estimar intervalos de confianza de los parámetros. Una alternativa es usar el método del percentil; el cual, luego de obtener los registros de $\hat{\Theta}^*(b)$ de las muestras de bootstrap, el intervalo de confianza de tipo percentil al $(1 - \alpha)100\%$ de confianza para cada parámetro está determinado por los cuantiles:

$$[q(\alpha/2), q(1 - \alpha/2)] \quad (2.55)$$

Otra alternativa es utilizar la siguiente ecuación:

$$[\hat{\theta} - 2SE(\hat{\theta}), \hat{\theta} + 2SE(\hat{\theta})] \quad (2.56)$$

donde $SE(\hat{\theta})$ representa el error estándar de $\hat{\theta}$, que corresponde a la raíz cuadrada del elemento de la diagonal de $\widehat{Var-Cov}(\hat{\Theta})$ del parámetro respectivo.

2.3.2. Pronósticos, decodificación y predicción de estados

Distribuciones Condicionales

Usando la verosimilitud de un modelo oculto de Markov y las probabilidades de avance y retroceso, se sigue inmediatamente que para $n = 2, 3, \dots, N$, la distribución de X_n condicionada al resto de observaciones está dada por:

$$P(X_n = x | \mathbf{X}^{(-n)} = \mathbf{x}^{(-n)}) = \frac{\delta \mathbf{P}(x_1) \mathbf{B}_2 \cdots \mathbf{B}_{n-1} \Gamma \mathbf{P}(x) \mathbf{B}_{n+1} \cdots \mathbf{B}_N \mathbf{1}'}{\delta \mathbf{P}(x_1) \mathbf{B}_2 \cdots \mathbf{B}_{n-1} \Gamma \mathbf{B}_{n+1} \cdots \mathbf{B}_N \mathbf{1}'} \propto \alpha_{n-1} \Gamma \mathbf{P}(x) \beta_n' \quad (2.57)$$

Con $\mathbf{x}^{(-n)} = (x_1, \dots, x_{n-1}, x_{n+1}, \dots, x_N)$, $\mathbf{B}_n = \Gamma \mathbf{P}(x_n)$, $\alpha_n = \delta \mathbf{P}(x_1) \mathbf{B}_2 \cdots \mathbf{B}_n$ y $\beta_n' = \mathbf{B}_{n+1} \cdots \mathbf{B}_N \mathbf{1}'$.

En la distribución condicional anterior, el numerador es la verosimilitud de las observaciones excepto que la observación x_n es reemplazada por x y el denominador es la verosimilitud de las observaciones excepto que x_n es tratado como un dato faltante.

En la ecuación 2.57, la probabilidad tiene la siguiente forma: una vector fila multiplicado por una matriz diagonal de dimensión $m \times m$, $\mathbf{P}(x) = \text{diag}(p_1(x), \dots, p_m(x))$, multiplicada por un vector columna. Se sigue que para $n = 1, 2, \dots, N$, se tiene que

$$P(X_n = x | \mathbf{X}^{(-n)} = \mathbf{x}^{(-n)}) \propto \sum_{i=1}^m d_i(n) p_i(x)$$

donde, $d_i(n)$ es el producto de la i -ésima entrada del vector $\alpha_{n-1} \Gamma$ y la i -ésima entrada del vector β_n . Por tanto,

$$P(X_n = x | \mathbf{X}^{(-n)} = \mathbf{x}^{(-n)}) = \sum_{i=1}^m w_i(n) p_i(x) \quad (2.58)$$

donde $w_i(n) = d_i(n) / \sum_{j=1}^m d_j(n)$ son funciones de las observaciones $\mathbf{x}^{(-n)}$ y de los parámetros del modelo.

Distribuciones de pronóstico

La distribución condicional de X_{N+h} dado $\mathbf{X}^{(N)} = \mathbf{x}^{(N)}$, donde h representa el horizonte de tiempo, para observaciones de valores discretos es:

$$\begin{aligned} P(X_{N+h} = x | \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) &= \frac{P(\mathbf{X}^{(N)} = \mathbf{x}^{(N)}, X_{N+h} = x)}{P(\mathbf{X}^{(N)} = \mathbf{x}^{(N)})} \\ &= \frac{\delta \mathbf{P}(x_1) \mathbf{B}_2 \mathbf{B}_3 \cdots \mathbf{B}_N \mathbf{\Gamma}^h \mathbf{P}(x) \mathbf{1}'}{\delta \mathbf{P}(x_1) \mathbf{B}_2 \mathbf{B}_3 \cdots \mathbf{B}_N \mathbf{1}'} \\ &= \frac{\alpha_N \mathbf{\Gamma}^h \mathbf{P}(x) \mathbf{1}'}{\alpha_N \mathbf{1}'} \end{aligned}$$

Si se denota $\phi_N = \alpha_N / \alpha_N \mathbf{1}'$, se tiene

$$P(X_{N+h} = x | \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) = \phi_N \mathbf{\Gamma}^h \mathbf{P}(x) \mathbf{1}' \quad (2.59)$$

La distribución de pronóstico puede ser escrita como mezcla de las distribuciones de probabilidad dependientes del estado:

$$P(X_{N+h} = x | \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) = \sum_{i=1}^m \xi_i(h) p_i(x) \quad (2.60)$$

donde $\xi_i(h)$ es la i -ésima entrada del vector $\phi_N \mathbf{\Gamma}^h$. En el caso de observaciones de valores continuos 2.60 representa la densidad de pronóstico.

Decodificación

En muchas aplicaciones es de interés determinar los estados de la cadena de Markov más probables (*de acuerdo al modelo ajustado*) que haya generado la secuencia de observaciones.

La decodificación local del estado al tiempo n se refiere a la determinación del estado más probable a ese momento. Por el contrario, decodificación global se refiere a la determinación de una secuencia de estados más probables.

Probabilidades de estado y decodificación local. Para la derivación del estado más probable de la cadena de Markov al tiempo $n \in \{1, \dots, N\}$. Usando la ecuación 2.44, la distribución condicional de C_n dadas las observaciones para $i = 1, 2, \dots, m$, se

define:

$$\begin{aligned} P(C_n = i | \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) &= \frac{P(C_n = i, \mathbf{X}^{(N)} = \mathbf{x}^{(N)})}{P(\mathbf{X}^{(N)} = \mathbf{x}^{(N)})} \\ &= \frac{\alpha_n(i)\beta_n(i)}{L_N} \end{aligned} \quad (2.61)$$

Para cada tiempo $n \in \{1, \dots, N\}$ se puede determinar la distribución del estado C_n , dadas las observaciones $\mathbf{x}^{(N)}$. Para los m estados es una distribución de probabilidad discreta con soporte en $\{1, \dots, m\}$.

Para cada $n \in \{1, \dots, N\}$ el estado más probable i_n^* , dadas las observaciones, es definida como

$$i_n^* = \underset{i=1, \dots, m}{\operatorname{argmax}} P(C_n = i | \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) \quad (2.62)$$

Este enfoque determina el estado más probable para cada n (*por separado*) por la maximización de la probabilidad condicional $P(C_n = i | \mathbf{X}^{(N)} = \mathbf{x}^{(N)})$, por lo que es llamada decodificación local.

Decodificación Global. En muchas de las aplicaciones, no es de mucho interés el estado más probable para un tiempo n separado, sino la secuencia más probable de estados (*ocultos*). En lugar de maximizar $P(C_n = i | \mathbf{X}^{(N)} = \mathbf{x}^{(N)})$ sobre i para cada n , se busca la secuencia de estados c_1, c_2, \dots, c_N que maximiza la probabilidad condicional

$$P(\mathbf{C}^{(N)} = \mathbf{c}^{(N)} | \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) \quad (2.63)$$

o equivalentemente, y más conveniente, la probabilidad conjunta

$$P(\mathbf{C}^{(N)} = \mathbf{c}^{(N)}, \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) = \delta_{c_1} \prod_{n=2}^N \gamma_{c_{n-1}, c_n} \prod_{n=1}^N p_{c_n}(x_n)$$

Este es un problema de maximización sutilmente diferente a la decodificación local, y se denomina decodificación global. Los resultados de la decodificación local y global a menudo son muy similares, pero no idénticos.

Al maximizar 2.63 sobre todas las posibles secuencias de estados c_1, c_2, \dots, c_N por fuerza bruta involucra m^N operaciones; lo que no es factible excepto para N pequeño.

Afortunadamente, existe un algoritmo eficiente de programación dinámica para determinar la secuencia de estados más probable, llamada algoritmo de Viterbi.

Se empieza definiendo

$$\xi_{1i} = P(C_1 = i, X_1 = x_1) = \delta_i p_i(x_1),$$

y, para $n = 2, 3, \dots, N$

$$\xi_{ni} = \max_{c_1, c_2, \dots, c_{n-1}} P(\mathbf{C}^{(n-1)} = \mathbf{c}^{(n-1)}, C_n = i, \mathbf{X}^{(n)} = \mathbf{x}^{(n)})$$

Las probabilidades ξ_{ni} satisfacen la siguiente recursión, para $n = 2, 3, \dots, N$ y $j = 1, 2, \dots, m$:

$$\xi_{nj} = \left(\max_i (\xi_{n-1,i} \gamma_{ij}) \right) p_j(x_n) \quad (2.64)$$

Esto proporciona un medio eficiente para calcular la matriz de dimensión $N \times m$ de valores ξ_{nj} , ya que, el esfuerzo computacional es lineal en N . La secuencia de maximización de los estados i_1, i_2, \dots, i_N puede entonces ser determinada recursivamente desde:

$$i_N = \underset{i=1, \dots, m}{\operatorname{argmax}} \xi_{Ni} \quad (2.65)$$

y, para $n = N - 1, N - 2, \dots, 1$, de

$$i_n = \underset{i=1, \dots, m}{\operatorname{argmax}} (\xi_{ni} \gamma_{i, i_{n+1}}) \quad (2.66)$$

Para medir el rendimiento de la decodificación, se puede identificar los estados correctos, por lo que, es usual simular una serie de observaciones del modelo ajustado y luego aplicar el algoritmo de Viterbi para decodificar sus estados, y luego comparar los estados de la serie (*conocida*) con los estados de la serie simulada.

El algoritmo de Viterbi para decodificar los estados ocultos procede de la siguiente manera:

Algoritmo 2.4: Decodificación de estados (*Viterbi*)

Data: Modelo ajustado

Result: Secuencia de estados i_1, i_2, \dots, i_N

- 1 $\xi_{1i} = \delta_i p_i(x_1)$;
 - 2 **for** $n = 2, 3, \dots, N$ **do**
 - 3 $\xi_{nj} = \left(\max_i (\xi_{n-1,i} \gamma_{ij}) \right) p_j(x_n)$;
 - 4 $i_N = \underset{i=1, \dots, m}{\operatorname{argmax}} \xi_{Ni}$;
 - 5 **for** $n = N - 1, N - 2, \dots, 2, 1$ **do**
 - 6 $i_n = \underset{i=1, \dots, m}{\operatorname{argmax}} (\xi_{ni} \gamma_{i, i_{n+1}})$;
-

Predicción de estados

La fórmula para la distribución condicional del estado C_n , para $n = 1, 2, \dots, N$, dadas las observaciones $\mathbf{x}^{(N)}$, consideraba solo los estados pasados y presentes. Para realizar la predicción del estado se considera $n > N$.

Dadas las observaciones x_1, \dots, x_N y la ecuación 2.61, se tiene que la probabilidad condicional de predicción del estado puede expresarse para $i = 1, 2, \dots, m$, como:

$$P(C_{N+h} = i | \mathbf{X}^{(N)} = \mathbf{x}^{(N)}) = \frac{\alpha_N \Gamma^h \mathbf{e}'_i}{L_N} = \phi_N \Gamma^h \mathbf{e}'_i \quad (2.67)$$

con $\phi_N = \alpha_N / \alpha_N \mathbf{1}'$, h el horizonte de tiempo y $\mathbf{e}_i = (0, \dots, 0, \underbrace{1}_{i\text{-ésima posición}}, 0, \dots, 0)$.

2.3.3. Selección y validación del modelo

Selección del modelo por AIC y BIC

“Un problema que surge cuando se usa los modelos ocultos de Markov (o cualquier otro) es seleccionar un modelo apropiado, por ejemplo, elegir el número de estados m o elegir la distribución dependiente del estado (binomial, Poisson, normal, beta, etc). Por lo que es necesario algún criterio para comparar los diferentes modelos posibles.” (Zucchini et al., 2016)

Se asume que las observaciones x_1, x_2, \dots, x_N fueron generadas por un modelo verdadero, pero desconocido f y que a estas observaciones se ajustan modelos de dos familias diferentes aproximadas $\{g_1 \in G_1\}$ y $\{g_2 \in G_2\}$. El objetivo de la selección del modelo es identificar el modelo, que en algún sentido, es el mejor.

En el enfoque frecuentista, se selecciona la familia estimada más cercana al modelo verdadero; que se simplifica en el “*Criterio de Información de Akaike*” (*AIC*), que está definido por:

$$AIC = -2\log(L) + 2p, \quad (2.68)$$

En el enfoque bayesiano se selecciona la familia estimada que es más probable al modelo verdadero; que se simplifica en el “*Criterio de Información Bayesiano*” (*BIC*), que está definido por:

$$BIC = -2\log(L) + p \log(N), \quad (2.69)$$

Donde $\log(L)$ es el logaritmo de la verosimilitud del modelo ajustado, p denota en número de parámetros del modelo y N es el número de observaciones.

Finalmente, en ambos criterios, se tiene que el mejor modelo para el conjunto de datos dado, es aquel que presenta el menor valor *AIC* y *BIC*, respectivamente.

Validación del modelo con pseudo-residuos

“A pesar de haber seleccionado el mejor modelo con alguno de los criterios anteriores, existe aún el problema de decidir si el modelo es necesariamente el adecuado, por lo que es necesario una herramienta que evalúe la bondad de ajuste general del modelo, e identificar valores atípicos relativos al modelo.” (Zucchini et al., 2016)

En el contexto de los modelos de regresión, los residuos es una herramienta para validar el modelo. En el caso de los modelos ocultos de Markov, la herramienta usada son los pseudo-residuos.

Introducción a pseudo-residuos. Para motivar los pseudo-residuos es necesario el siguiente resultado: “Sea X una variable aleatoria con función de distribución continua F . Entonces $U \equiv F(X)$ está uniformemente distribuido en el intervalo unitario, que se escribe de la siguiente manera:”

$$U \sim \mathbf{U}(0, 1)$$

El pseudo-residuo uniforme de una observación x_n de una variable aleatoria X_n es-

tá definida como la probabilidad, bajo el modelo ajustado, de obtener una observación menor o igual a x_n :

$$u_n = P(X_n \leq x_n) = F_{X_n}(x_n)$$

Es decir, u_n es la observación x_n transformada por su función distribución bajo el modelo. Si el modelo es correcto, este tipo de pseudo-residuo esta distribuido $U(0, 1)$, con residuos para observaciones extremas cercanas a 0 o 1. Si el histograma o gráfico qq-plot de u_n muestra dudas de que sea $U(0, 1)$, se deduce que el modelo no es válido.

Los pseudo-residuos uniformes son útiles, pero tienen el problema de no poder identificar valores atípicos, pues no son distinguibles valores como 0,97 o 0,999 gráficamente.

Esta deficiencia de los pseudo-residuos uniformes puede ser solucionada si se utiliza el siguiente resultado. “Sea Φ la función de distribución normal estándar y X una variable con función de distribución F . Entonces $Z = \Phi^{-1}(F(X))$ está distribuida como una normal estándar.” Por lo que, se define el pseudo-residuo normal como

$$z_n = \Phi^{-1}(u_n) = \Phi^{-1}(F_{X_n}(x_n))$$

Si las observaciones x_1, \dots, x_N fueron generadas por el modelo $X_n \sim F_n$, los pseudo-residuos normales z_n seguirían una distribución normal estándar. Lo que es fácil de verificar, usando un histograma, un gráfico qq-plot o un test de normalidad.

En la figura 2.6 se muestra el procedimiento para la construcción de los pseudo-residuales normales. Donde, para $n = 1, 2, \dots, N$, el primer cuadro muestra la distribución de la observación x_n , el segundo cuadro muestra la distribución del pseudo-residual uniforme u_n y por último, el tercer cuadro muestra la distribución del pseudo-residual normal z_n .

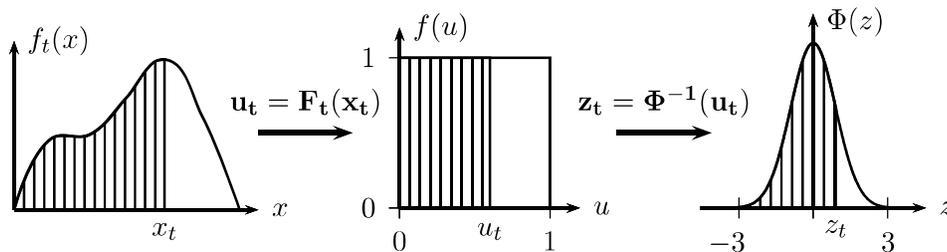


Figura 2.6: Construcción de los pseudo-residuos normales.

El análisis de pseudo-residuos de un modelo oculto de Markov tiene dos objetivos: evaluar el ajuste general del modelo seleccionado y detectar valores atípicos.

Pseudo-residuos para modelos ocultos de Markov. Considera las observaciones una a la vez y busca aquellas que, en relación con el modelo y el resto de observaciones de la serie, son lo suficientemente extremas como para sugerir que difieren en naturaleza u origen de las otras. Esto significa que el cálculo del pseudo-residuo z_n de la distribución condicional de X_n dado $\mathbf{X}^{(-n)}$ está dada por:

$$z_n = \Phi^{-1} (P(X_n \leq x_n | \mathbf{X}^{(-n)} = \mathbf{x}^{(-n)}))$$

Si el modelo es correcto, z_n sigue una distribución normal estándar. El cálculo de la probabilidad condicional está dada por la ecuación 2.58.

2.3.4. Variaciones de los modelos ocultos de Markov

Se pueden obtener varios modelos al considerar diferentes distribuciones de probabilidad dependientes del estado. Por ejemplo, para series de conteos, una alternativa es considerar una distribución de Poisson o binomial; para series de datos binarios, una alternativa es considerar una distribución de Bernoulli. Para series de valores continuos se consideran densidades de probabilidad como: normal, exponencial, gamma, beta (*para valores entre $[0, 1]$*), etc. Incluso, es posible considerar una distribución multinomial (*generalización de una distribución binomial*). Si se considera varias series de tiempo a modelar a la vez, una alternativa es usar modelos ocultos de Markov multivariantes.

Los modelos ocultos de Markov se pueden modificar para permitir la influencia de covariables, permitiendo que algunos de sus parámetros dependan de covariables, como en los parámetros de las distribuciones dependientes del estado o en las probabilidades de transición de la cadena de Markov. En la primera, se establece que las distribuciones dependientes del estado, que generan las observaciones, dependen de dichas covariables, como se observa en la figura 2.7(a). En la segunda, se abandona la suposición de homogeneidad de la cadena de Markov y se establece que las probabilidades de transición son función del tiempo o de dichas covariables.

Otra modificación que se puede realizar en los modelos ocultos de Markov es considerar que la cadena de Markov subyacente es de segundo orden, como se observa

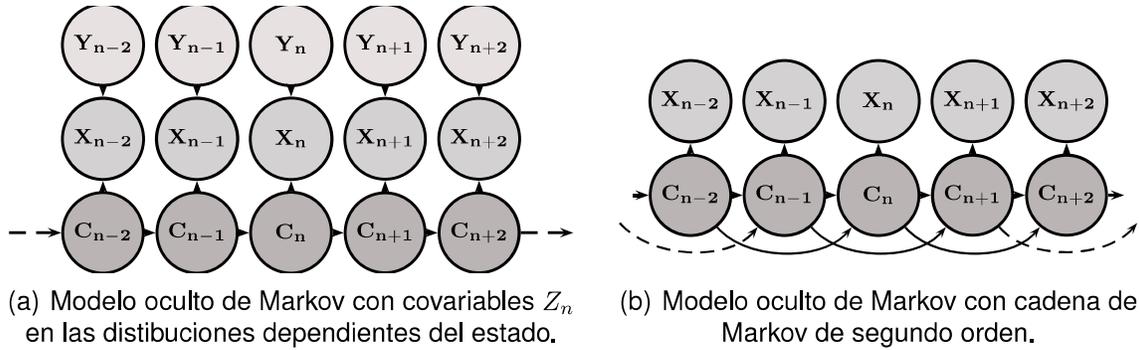


Figura 2.7: Dependencias adicionales en los modelos ocultos de Markov.

en la figura 2.7(b); es decir, que las probabilidades de transición están caracterizadas por:

$$\gamma(i, j, k) = P(C_{n+1} = k | C_n = j, C_{n-1} = i)$$

Se puede considerar otras dependencias adicionales en los modelos ocultos de Markov. Por ejemplo, la suposición de que la observación X_n depende únicamente del estado actual C_n se puede eliminar. Permitiendo dependencia de observaciones previas, así como de C_n , como se observa en las figuras 2.8. Estos modelos son llamados “*Markov-switching models*”.

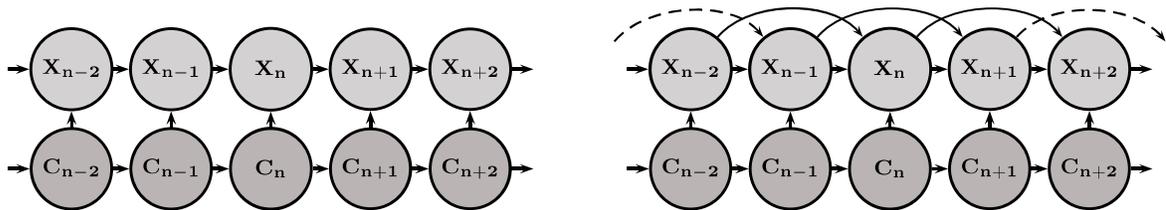


Figura 2.8: Dependencia adicional de las observaciones con observaciones previas.

Otro tipo de dependencia adicional permite que la distribución de X_n dependa de estados previos, así como del estado actual C_n . Es decir, en lugar de la suposición usual 2.24, se cambia por $P(X_n | \mathbf{X}^{(n-1)}, \mathbf{C}^{(n)}) = P(X_n | C_{n-1}, C_n)$, como se observa en la figura 2.9.

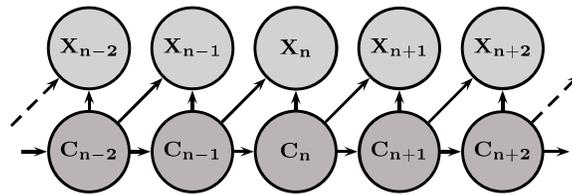


Figura 2.9: Dependencia adicional de las observaciones con estados previos.

Capítulo 3

Metodología y Resultados.

En este capítulo se detalla la metodología utilizada para el desarrollo de la modelación de las carteras de crédito comercial y de consumo del sector bancario privado ecuatoriano, para lo cual se inicia con la descripción de las variables a modelar, se ajustan los modelos a cada variable, para finalmente presentar los resultados y validar los modelos obtenidos.

La información considerada para la construcción de los modelos consiste en datos de series temporales con frecuencia mensual a partir julio 2002 hasta diciembre 2017. Estos datos fueron tomados del portal web de la Superintendencia de Bancos del Ecuador, que son libre acceso y pueden ser descargados de la sección sistema financiero: estadísticas de bancos privados¹.

3.1. Descripción del modelo

Basado en la bibliografía recopilada en la revisión de literatura del capítulo 1, se establece una metodología en común. De manera general, se crean grupos que representan los estados de la cadena de Markov, se establece la distribución dependiente del estado que generan las observaciones, en base a los grupos creados se calculan valores iniciales que posteriormente son usados en el algoritmo EM (*Baum-Welch*) para estimar los parámetros del modelo oculto de Markov, luego se decodifica los estados ocultos y finalmente, en pocos casos, se realizan pronósticos.

En este trabajo de investigación se usa una metodología similar con algunas modificaciones. El procedimiento a seguir se presenta en la figura 3.1 y se detalla a conti-

¹Los datos pueden ser descargados del siguiente enlace: Boletín de series mensual

nuación:

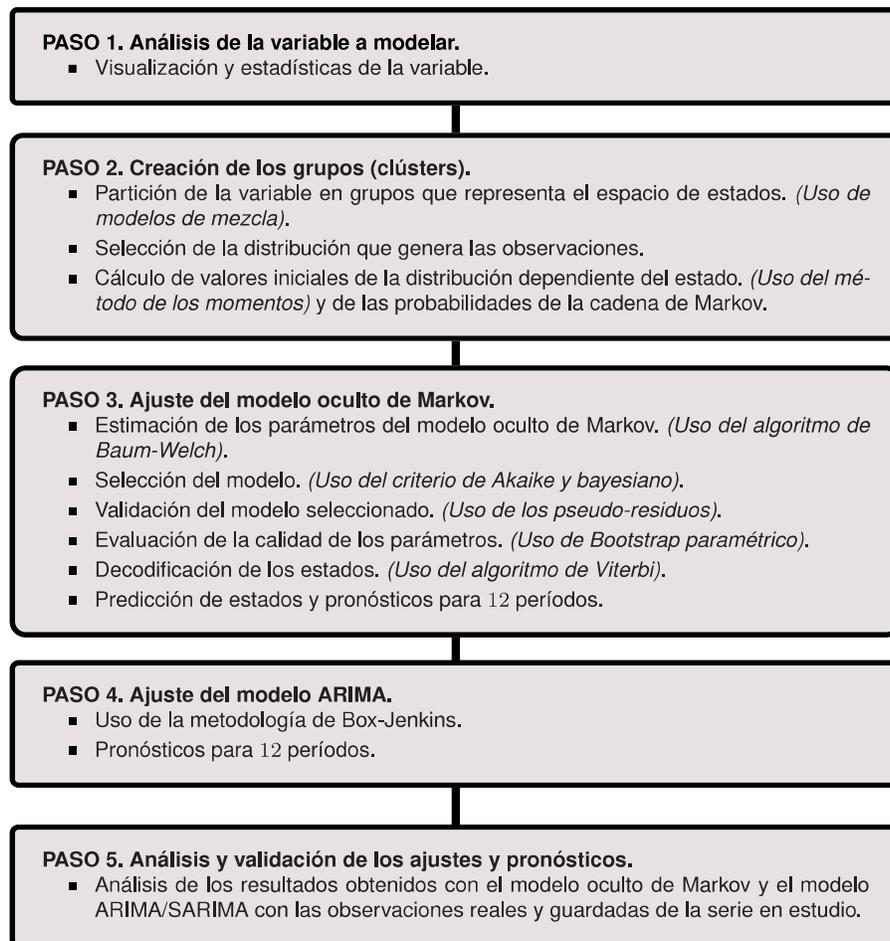


Figura 3.1: Procedimiento del modelo.

PASO 1. Análisis de la variable.

- Se dispone de información de las series de morosidad de las carteras de crédito comercial y de consumo a partir de julio 2002 hasta diciembre 2017. Para lo cual se usará la información de julio 2002 a diciembre 2016 como datos de entrenamiento para ajustar el modelo, mientras que la información del período 2017 se guardará para ser usado como datos de validación para compararse con los resultados obtenidos del modelo.

PASO 2. Creación de los grupos (clústers).

- Para cada variable se ajusta un modelo de mezcla para particionarla en 2, 3, 4 y 5 grupos (*clústers*) que representarán el conjunto de estados de la cadena de Markov. Además, se presentará el gráfico de los grupos formados y la densidad de mezcla resultante.
- Luego, con los grupos formados se calcula los valores de inicio de los parámetros de la cadena oculta de Markov que servirán para algoritmo de Baum-Welch. Para la distribución inicial δ , se considera valores iguales, es decir, si $m = 3$ entonces $\delta = (1/3, 1/3, 1/3)$. Para la matriz de probabilidades Γ , se usará la expresión 2.15. Para los parámetros de la distribución dependiente del estado, se considera una distribución $beta(\alpha, \beta)$ (*pues los datos están entre [0, 1]*), para lo cual se usa del método de los momentos para estimarlos. Con el método de los momentos, las estimaciones de los parámetros de la distribución beta, para un conjunto de datos, se calcula de la siguiente manera:

$$\hat{\alpha} = \bar{x} \left[\frac{\bar{x}(1 - \bar{x})}{s^2} - 1 \right] \quad (3.1)$$

$$\hat{\beta} = (1 - \bar{x}) \left[\frac{\bar{x}(1 - \bar{x})}{s^2} - 1 \right] \quad (3.2)$$

donde \bar{x} representa la media y s la desviación estándar del conjunto de datos.

PASO 3. Ajuste del modelo oculto de Markov.

- Con los valores iniciales obtenidos en el paso anterior, se realiza el entrenamiento del modelo oculto de Markov, para obtener sus parámetros (δ , Γ , α y β) para $m = 2, 3, 4, 5$ estados.
- Con los modelos obtenidos previamente se procederá a realizar la selección del mejor modelo con ayuda del criterio de información de Akaike (*ecuación 2.68*) y el criterio de información bayesiano (*ecuación 2.69*) y se validará con ayuda de los pseudo-residuos normales.
- Con el modelo seleccionado y validado, se procederá a realizar la evaluación de la calidad de los parámetros estimados, obteniendo sus intervalos de confianza respectivos con ayuda de Bootstrap paramétrico (*algoritmo 2.3*).
- Con ayuda del algoritmo de Viterbi se decodificará los estados ocultos de la cadena de Markov (*algoritmo 2.4*).

- Se realiza la predicción de los estados para 12 períodos futuros (*ecuación 2.67*).
- Finalmente, se presentará las distribuciones de pronóstico (*ecuación 2.60*) para los 12 períodos futuros para obtener los pronósticos en cada período correspondiente. En este caso, al usar distribuciones continuas, la ecuación 2.60 da a lugar a una densidad de probabilidad, por lo que se usará como valores de pronóstico a la media, mediana y moda de cada densidad. Para el cálculo de las sumas, se toma una partición $\{x_0, x_1, \dots, x_i, \dots, x_{n-1}, x_n\}$ de $[0, 1]$, donde $x_0 = 0$ y $x_1 = 1$ con n suficientemente grande. De esta manera se tiene:

Media (\bar{x}):

$$\bar{x} = \int_0^1 x f(x) dx \approx \sum_{i=0}^n x_i f(x_i) (x_i - x_{i-1}) \quad (3.3)$$

Mediana (Md): Valor para el cual se cumple

$$\int_0^{Md} f(x) dx = 0,5 \iff \sum_{i=0}^k f(x_i) (x_i - x_{i-1}) = 0,5 \quad (3.4)$$

donde $Md = X_k$, con $0 \leq k \leq n$.

Moda (Mo): Punto de máxima densidad.

De manera similar, es posible calcular otros estadísticos, tales como:

Varianza ($Var(x)$):

$$\begin{aligned} Var(x) &= E(x^2) - (E(x))^2 = \int_0^1 x^2 f(x) dx - \left(\int_0^1 x f(x) dx \right)^2 \\ &\approx \sum_{i=0}^n x_i^2 f(x_i) (x_i - x_{i-1}) - \left(\sum_{i=0}^n x_i f(x_i) (x_i - x_{i-1}) \right)^2 \end{aligned} \quad (3.5)$$

Desviación estándar (S):

$$S = \sqrt{Var(x)} \quad (3.6)$$

Primer cuartil (Q_1): Valor para el cual se cumple

$$\int_0^{Q_1} f(x) dx = 0,25 \iff \sum_{i=0}^{k'} f(x_i) (x_i - x_{i-1}) = 0,25 \quad (3.7)$$

donde $Q_1 = X_{k'}$, con $0 \leq k' \leq n$.

Tercer cuartil (Q_3): Valor para el cual se cumple

$$\int_0^{Q_3} f(x)dx = 0,75 \iff \sum_{i=0}^{k''} f(x_i)(x_i - x_{i-1}) = 0,75 \quad (3.8)$$

donde $Q_3 = X_{k''}$, con $0 \leq k'' \leq n$.

En las expresiones anteriores, al no disponer de una función que represente a la densidad, los estadísticos serán estimados mediante la definición de la integral de Riemann. Además, antes de calcular estos estadísticos es necesario que la densidad obtenida cumpla con las condiciones de positividad y que la integral de $-\infty$ a ∞ (de 0 a 1 para la distribución beta), sea igual a 1.

PASO 4. Ajuste del modelo ARIMA/SARIMA.

- Adicional a los modelos ocultos de Markov se aplicará un modelo ARIMA/SARIMA con las series de morosidad aplicando la metodología de Box-Jenkins que se detallada en Capa Santos (2016).
- Se realizará pronósticos para los 12 períodos futuros, los cuales se usarán para el siguiente paso.

PASO 5. Análisis y validación de los pronósticos.

- Para finalizar, es este paso se realizará un análisis de los ajustes y pronósticos obtenidos con el modelo oculto de Markov y el modelo ARIMA/SARIMA, según sea el caso, con las observaciones reales de entrenamiento y las observaciones de validación guardadas del año 2017, para analizar el modelo oculto de Markov obtenido.

Para analizar sus pronósticos se hace uso del cuadrado medio debido al error y del error porcentual absoluto medio, que de acuerdo a Anderson et al. (2011) se calculan por:

Cuadrado medio debido al error (CME):

$$CME = \frac{1}{K} \sum_{i=1}^K (Valor\ real_i - Pronóstico_i)^2 \quad (3.9)$$

Error porcentual absoluto medio (EPAM):

$$EPAM = \frac{1}{K} \sum_{i=1}^K \frac{|Valor\ real_i - Pronóstico_i|}{Valor\ real_i} \times 100 \quad (3.10)$$

donde K representa el número de pronósticos que se realice. De acuerdo a estas medidas, un modelo es adecuado cuando reporta valores pequeños.

3.2. Modelamiento

Para el modelamiento de cada una de las series de morosidad de las carteras de crédito se considera un conjunto de datos de entrenamiento y un conjunto de datos de validación. En este caso, para el primer conjunto se toman las observaciones a partir de julio 2002 hasta diciembre 2016 y para el segundo conjunto se consideran las observaciones correspondientes al año 2017.

3.2.1. Modelamiento de la cartera de crédito comercial

En la figura 3.2 se muestra la evolución de la morosidad de la cartera de crédito comercial donde la sección sombreada representa al conjunto de observaciones de validación y la sección no sombreada representa el conjunto de observaciones de entrenamiento. En esta figura se observa que al inicio del período de estudio se tienen valores de morosidad altos, alrededor del 12%. Posteriormente, la morosidad disminuye hasta que en los últimos períodos tiende a estabilizarse, alrededor del 1%.

Para modelar la serie de morosidad comercial se usan 174 observaciones que conforman el conjunto de valores de entrenamiento donde el valor de la media es de 0,0369 con un valor mínimo de 0,0075 y un valor máximo de 0,1149, como se aprecia en el cuadro 3.1.

Nro. Obs.	Media	Desv. Est.	Mínimo	1er cuartil	2do cuartil	3er cuartil	Máximo
174	0.0369	0.0324	0.0075	0.0128	0.0227	0.0568	0.1149

Cuadro 3.1: Cuadro de estadísticos de la serie de morosidad comercial.

En la figura 3.3 se muestra la distribución de los valores de morosidad de la cartera comercial junto a su media (*recta segmentada vertical*). En ésta se observa que existe

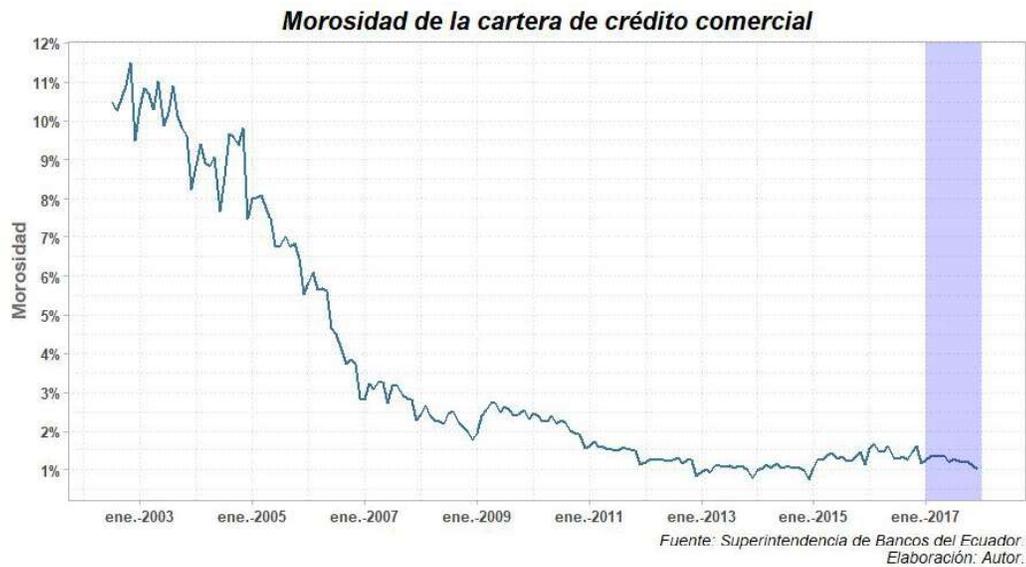


Figura 3.2: Serie de morosidad de la cartera de crédito comercial.

una gran cantidad de observaciones entre 0 y 0,03, el resto de valores se distribuye con poca frecuencia entre 0,03 y 0,12. Esto también se corrobora en la figura 3.2.

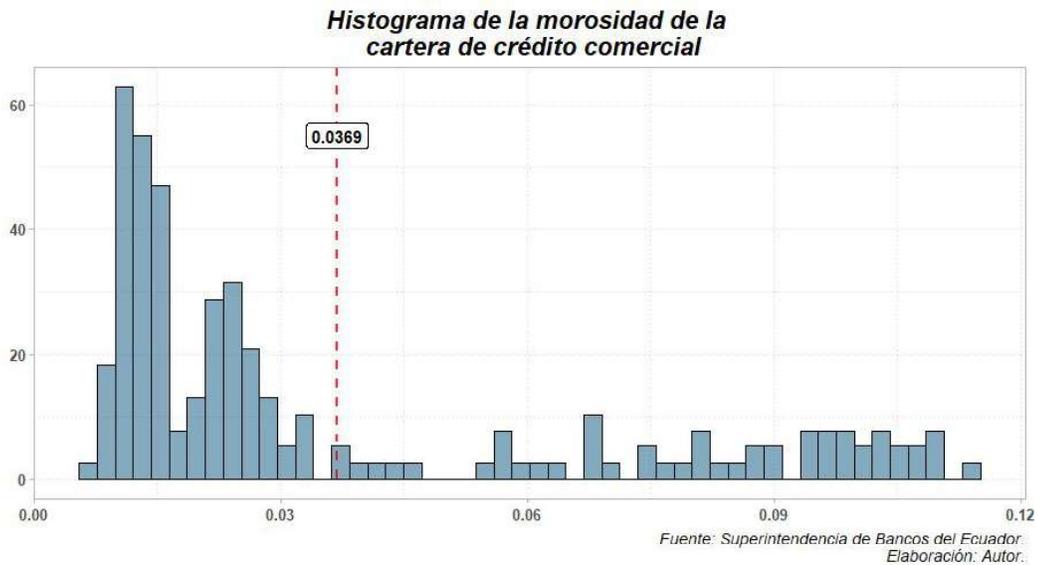


Figura 3.3: Histograma de la cartera de crédito comercial.

Para comenzar a modelar la serie de morosidad se empieza creando los grupos o clústers, que representarán niveles de riesgo y que a la vez formarán el conjunto finito de estados de la cadena de Markov. En este caso, cuando los valores son cercanos a 1, representaría un estado de *crisis* indicando que un quiebre del sistema financiero está cerca. Por el contrario, valores cercanos a 0 indicarían que un quiebre del sistema

financiero está lejos de producirse; es decir, se definiría como un estado de *no crisis*. Para esto, se utilizan los modelos de mezcla con los cuales se generan los grupos en los cuales las observaciones de morosidad siguen una distribución beta con parámetros determinados por las observaciones que conforman cada grupo y que pueden ser estimados mediante el método de los momentos, dadas por las ecuaciones 3.1 y 3.2.

De esta manera, se considera crear modelos de mezcla con 2, 3, 4 y 5 grupos, los cuales se pueden apreciar en la figura 3.4, donde la recta segmentada representa la media del grupo generado por el modelo de mezcla respectivo en cada caso.

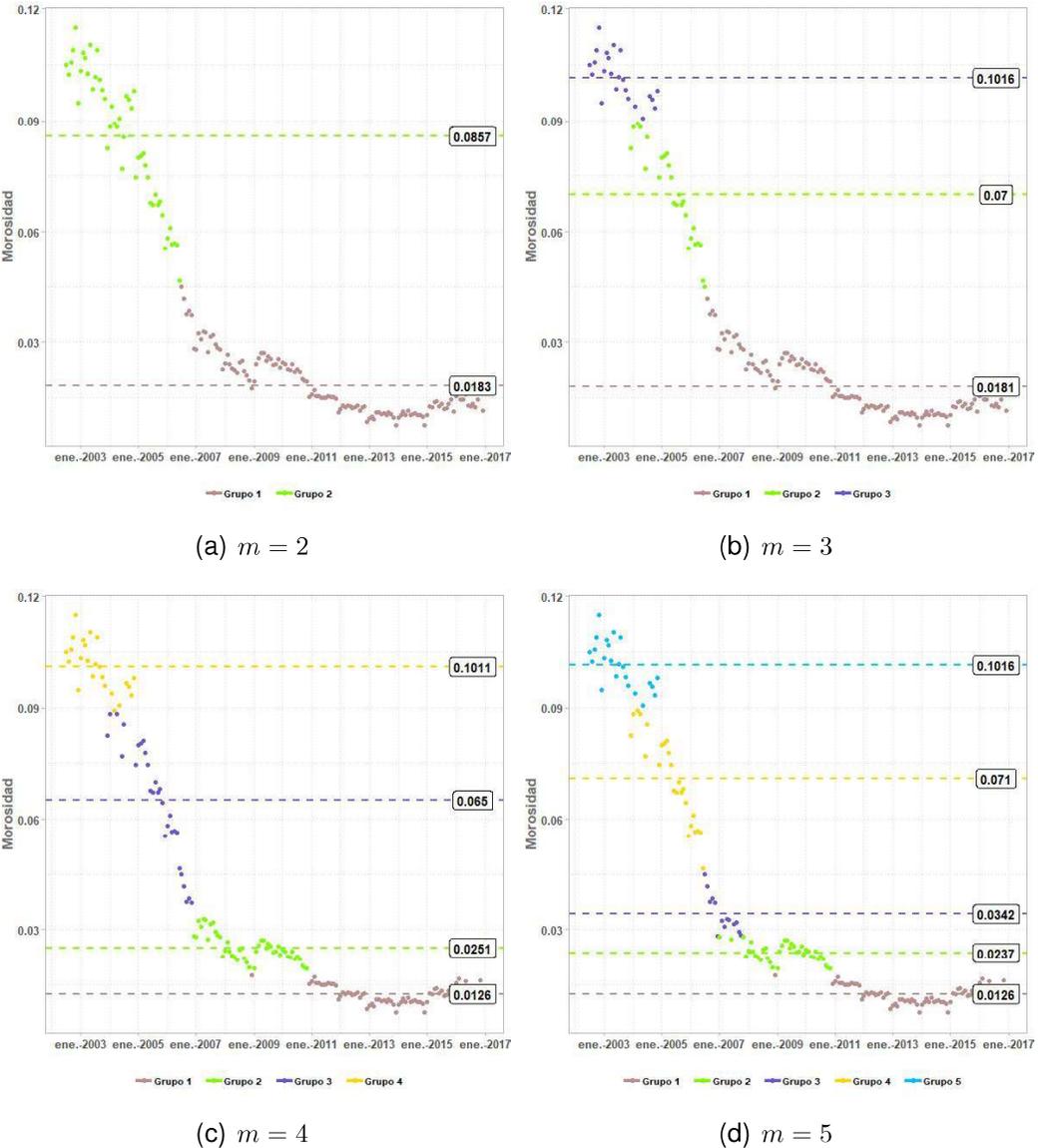


Figura 3.4: Grupos generados con los modelos de mezclas para la morosidad de la cartera de crédito comercial.

Los clústers formados con los diferentes modelos de mezcla estimados, son usados para estimar los parámetros de la distribución beta de cada clúster para cada modelo de mezcla usando las ecuaciones 3.1 y 3.2. Los resultados de los parámetros estimados se encuentran tabulados en los cuadros 3.2, 3.3, 3.4 y 3.5.

	ω	α	β
Grupo 1	0.2768063	20.15	215.07
Grupo 2	0.7231937	5.13	274.78

Cuadro 3.2: Parámetros para el modelo de mezcla con 2 grupos para la morosidad de la cartera de crédito comercial.

	ω	α	β
Grupo 1	0.121892	231.33	2045.56
Grupo 2	0.718129	5.46	295.93
Grupo 3	0.159979	27.40	364.05

Cuadro 3.3: Parámetros para el modelo de mezcla con 3 grupos para la morosidad de la cartera de crédito comercial.

	ω	α	β
Grupo 1	0.4195065	29.55	2322.43
Grupo 2	0.2818972	33.89	1289.73
Grupo 3	0.1776235	19.90	273.31
Grupo 4	0.1209728	231.33	2045.56

Cuadro 3.4: Parámetros para el modelo de mezcla con 4 grupos para la morosidad de la cartera de crédito comercial.

	ω	α	β
Grupo 1	0.1238681	231.33	2045.56
Grupo 2	0.1801895	102.51	4241.06
Grupo 3	0.1520404	32.12	420.22
Grupo 4	0.4287925	29.55	2322.43
Grupo 5	0.1151095	42.10	1204.48

Cuadro 3.5: Parámetros para el modelo de mezcla con 5 grupos para la morosidad de la cartera de crédito comercial.

Los parámetros estimados son utilizados para calcular la densidad de mezcla de acuerdo a la ecuación 2.1. Las figuras de las densidades de mezclas resultantes se presentan en 3.5. En las cuales se aprecia la mejora del ajuste de la densidad de mezcla a medida que se aumenta el número de grupos.

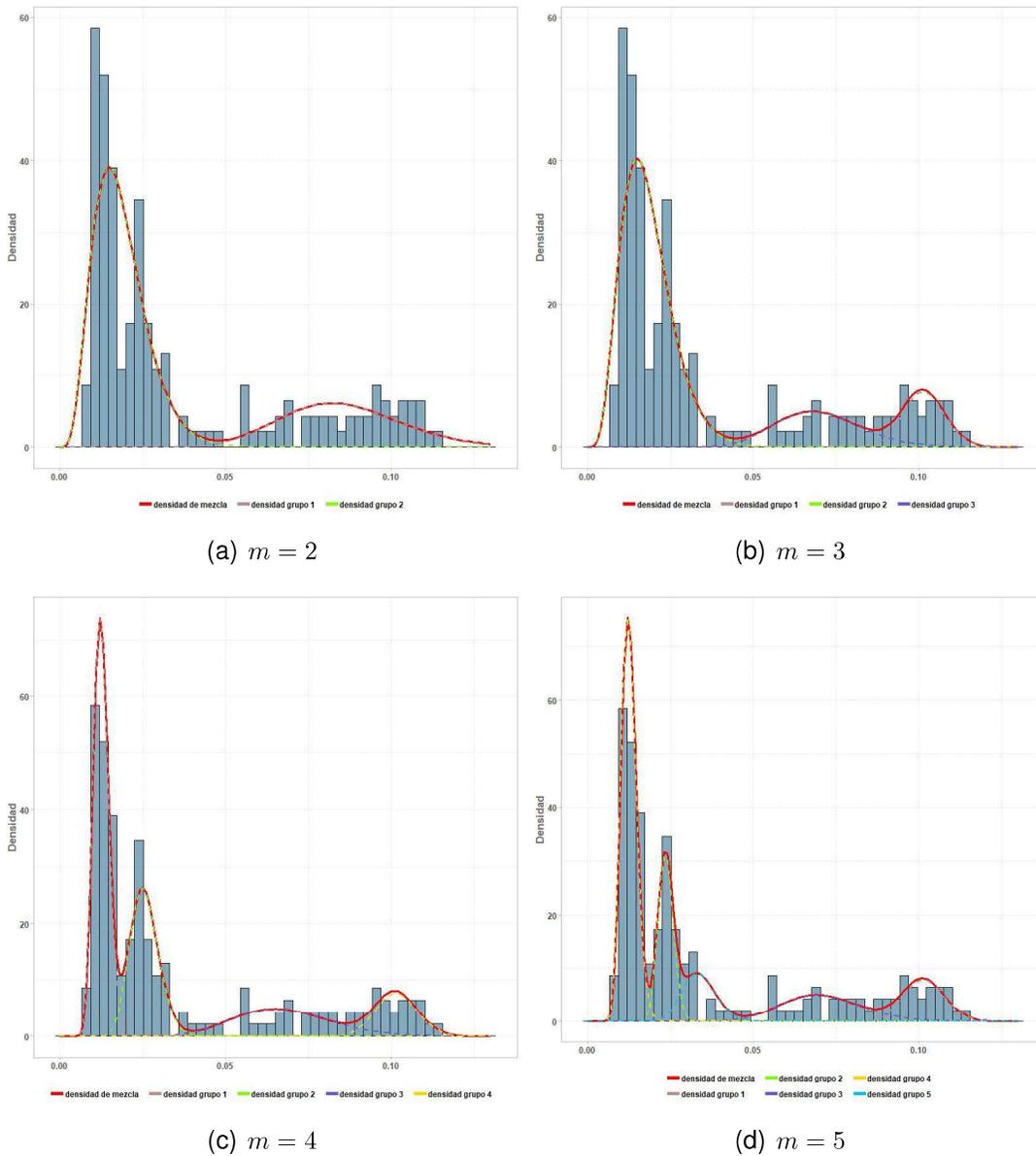


Figura 3.5: Densidades de mezcla formados por los modelos de mezcla para la morosidad de la cartera de crédito comercial.

Con ayuda de los grupos generados con los modelos de mezcla, se generan los valores iniciales que ayudarán a la estimación de los parámetros del modelo oculto de Markov usando el algoritmo de Baum-Welch. Los parámetros a estimarse son: la distribución inicial de la cadena de Markov (δ), la matriz de transición de probabilidades (Γ) y los parámetros de la distribución beta que dependen del estado. Para la distribución inicial se asignarán valores equiprobables; para la matriz de transición se estimará de acuerdo a los grupos formados con los modelos de mezcla estimados, con ayuda de

la ecuación 2.15; y por último, para los parámetros de la distribuciones dependientes del estado (α, β) se usarán los valores obtenidos de los modelos de mezcla que se encuentran tabulados en los cuadros 3.2, 3.3, 3.4 y 3.5.

De esta manera, los valores iniciales para el algoritmo de Baum-Welch con $m = 2, 3, 4$ y 5 estados, son los siguientes:

Para $m = 2$:

$$\delta = (1/2, 1/2)$$

$$\Gamma = \begin{pmatrix} 0,98 & 0,02 \\ 0,00 & 1,00 \end{pmatrix}$$

Para $m = 3$:

$$\delta = (1/3, 1/3, 1/3)$$

$$\Gamma = \begin{pmatrix} 0,83 & 0,00 & 0,17 \\ 0,00 & 1,00 & 0,00 \\ 0,12 & 0,04 & 0,85 \end{pmatrix}$$

Para $m = 4$:

$$\delta = (1/4, 1/4, 1/4, 1/4)$$

$$\Gamma = \begin{pmatrix} 0,99 & 0,01 & 0,00 & 0,00 \\ 0,04 & 0,94 & 0,02 & 0,00 \\ 0,00 & 0,07 & 0,82 & 0,11 \\ 0,00 & 0,00 & 0,17 & 0,83 \end{pmatrix}$$

Para $m = 5$:

$$\delta = (1/5, 1/5, 1/5, 1/5, 1/5)$$

$$\Gamma = \begin{pmatrix} 0,83 & 0,00 & 0,17 & 0,00 & 0,00 \\ 0,00 & 0,89 & 0,00 & 0,05 & 0,05 \\ 0,12 & 0,00 & 0,84 & 0,00 & 0,04 \\ 0,00 & 0,01 & 0,00 & 0,99 & 0,00 \\ 0,00 & 0,20 & 0,00 & 0,00 & 0,80 \end{pmatrix}$$

Con los valores iniciales anteriores se estiman los modelos ocultos de Markov con $m = 2, 3, 4$ y 5 estados, respectivamente. De los cuales se seleccionó como mejor modelo a aquel que tiene 4 estados, debido a que el criterio bayesiano es el menor de los 4 modelos calculados. Además, se acepta la hipótesis nula de normalidad de los

pesudo-residuos, pues el valor-p es mayor a 0,05. Por otro lado, se tiene un modelo parsimonioso, pues no es complejo, al no tener muchos parámetros por estimar (*bajo coste computacional*), satisface los criterios de normalidad y se previene un posible sobre ajuste.

El número de parámetros estimados dependerá del número de estados seleccionado; es decir, para m estados, en la distribución inicial δ se deben calcular $m - 1$ parámetros, pues su suma debe ser 1; en la matriz de transición Γ se deben calcular $m^2 - m$ estados, pues su suma por fila debe ser 1; por último, en los parámetros que dependen del estado (α, β) se deben calcular $2m$ parámetros. Lo que da un total de $m^2 + 2m$ parámetros por calcular. Así por ejemplo, si $m = 2$ se tendrá que calcular 7 parámetros. De manera similar para $m = 3, 4, 5$, respectivamente.

Estos resultados se encuentran resumidos en el cuadro 3.6.

Nro. de estados	Nro. de parámetros estimados	log. verosimilitud	AIC	BIC	p valor
2	7	564.3190	-1114.6381	-1092.5247	0.0000
3	14	597.6279	-1167.2558	-1123.0290	0.0006
4	23	694.0363	-1342.0726	-1269.4143	0.3294
5	34	714.1651	-1360.3302	-1252.9223	0.1653

Cuadro 3.6: Criterios para seleccionar el mejor modelo oculto de Markov para la morosidad de la cartera de crédito comercial.

Se presenta además, el gráfico de los pseudo-residuos para cada uno de los modelos ocultos de Markov, en los cuales se aprecian que con $m = 2$ y $m = 3$ estados, los pseudo-residuos se comportan con cierta tendencia, mientras que cuando $m = 4$ y $m = 5$ estados, éstos se muestran dispersos. Lo que se aprecia en la figura 3.6, donde además se presentan sus límites al 95% y 99% de confianza (*franjas azules y rojas, respectivamente*).

En la figura 3.7, se muestra la distribución de los pseudo-residuos con la respectiva gráfica de la densidad de la distribución normal estándar. En estas figuras se observan que los histogramas $m = 2$ y $m = 3$ estados no tienen un comportamiento igual al de una normal estándar, lo que si sucede en los pseudo-residuos cuando $m = 4$ y $m = 5$ estados.

Para corroborar que los pseudo-residuos del modelo oculto de Markov con 4 estados siguen una distribución normal estándar. Se realiza un gráfico cuantil-cuantil en donde, mientras más cercanos a la recta identidad se encuentren los pseudo-residuos, se

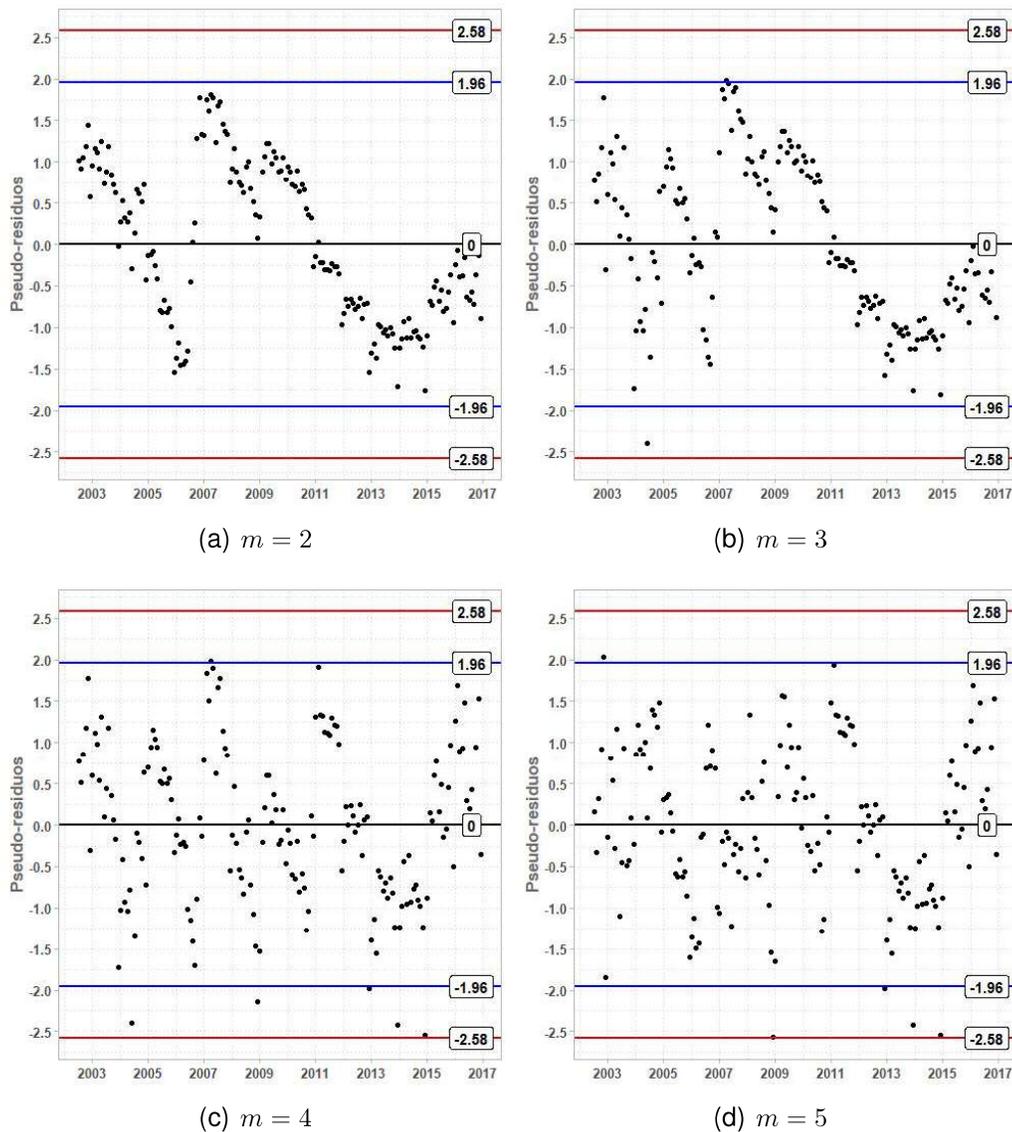


Figura 3.6: Gráfico de los pseudo-residuos de los modelos ocultos de Markov estimados para la morosidad de la cartera de crédito comercial.

concluye que siguen una distribución normal estándar. Esto sucede cuando $m = 4$ y $m = 5$, como se puede apreciar en la figura 3.8.

EL modelo oculto de Markov con 5 estados no fue seleccionado a pesar que cumple el criterio de normalidad de los pseudo-residuos, debido a que involucra el cálculo de un gran número de estados y su criterio de Akaike es alto comparado con los demás.

Para el modelo seleccionado se estiman los parámetros del modelo oculto de Markov con sus respectivos errores estándar e intervalos de confianza de acuerdo al método del percentil (ecuación 2.55) y al método de Bootstrap (ecuación 2.56) al 95 %

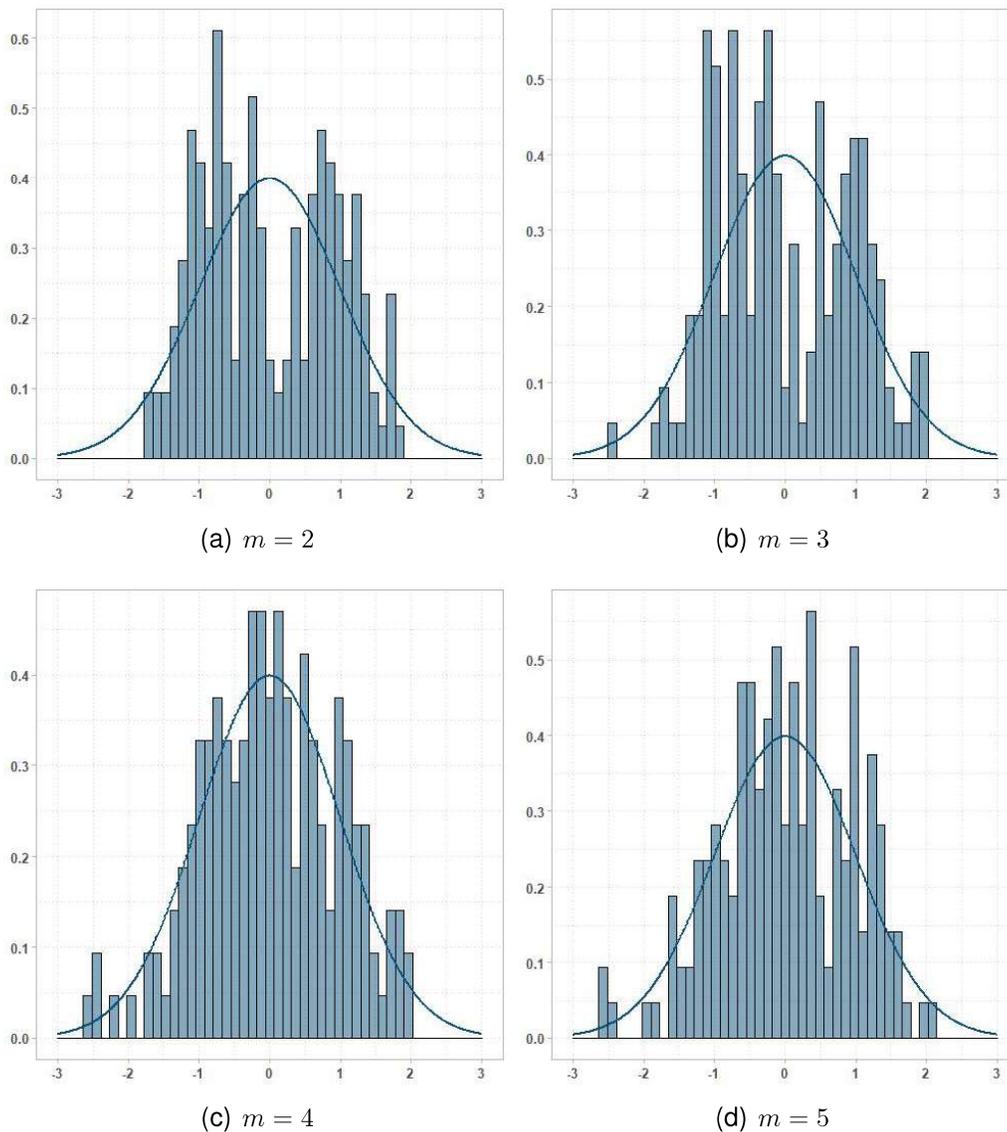


Figura 3.7: Histogramas de los pseudo-residuos de los modelos ocultos de Markov estimados para la morosidad de la cartera de crédito comercial.

de confianza. Estos resultados se encuentran tabulados en el cuadro 3.7, donde θ_0 denota los valores iniciales para el algoritmo de Baum-Welch; θ denota los valores resultantes de los parámetros luego de aplicar el algoritmo; $SE(\theta)$ denota el error estándar de bootstrap para cada parámetro estimado; y finalmente, $[q(\alpha/2), q(1 - \alpha/2)]$ y $[\theta - 2 * SE(\theta), \theta + 2 * SE(\theta)]$ denotan los intervalos de confianza con el método del percentil y de bootstrap, respectivamente.

Respecto a las estimaciones del modelo oculto de Markov con 4 estados, se tiene

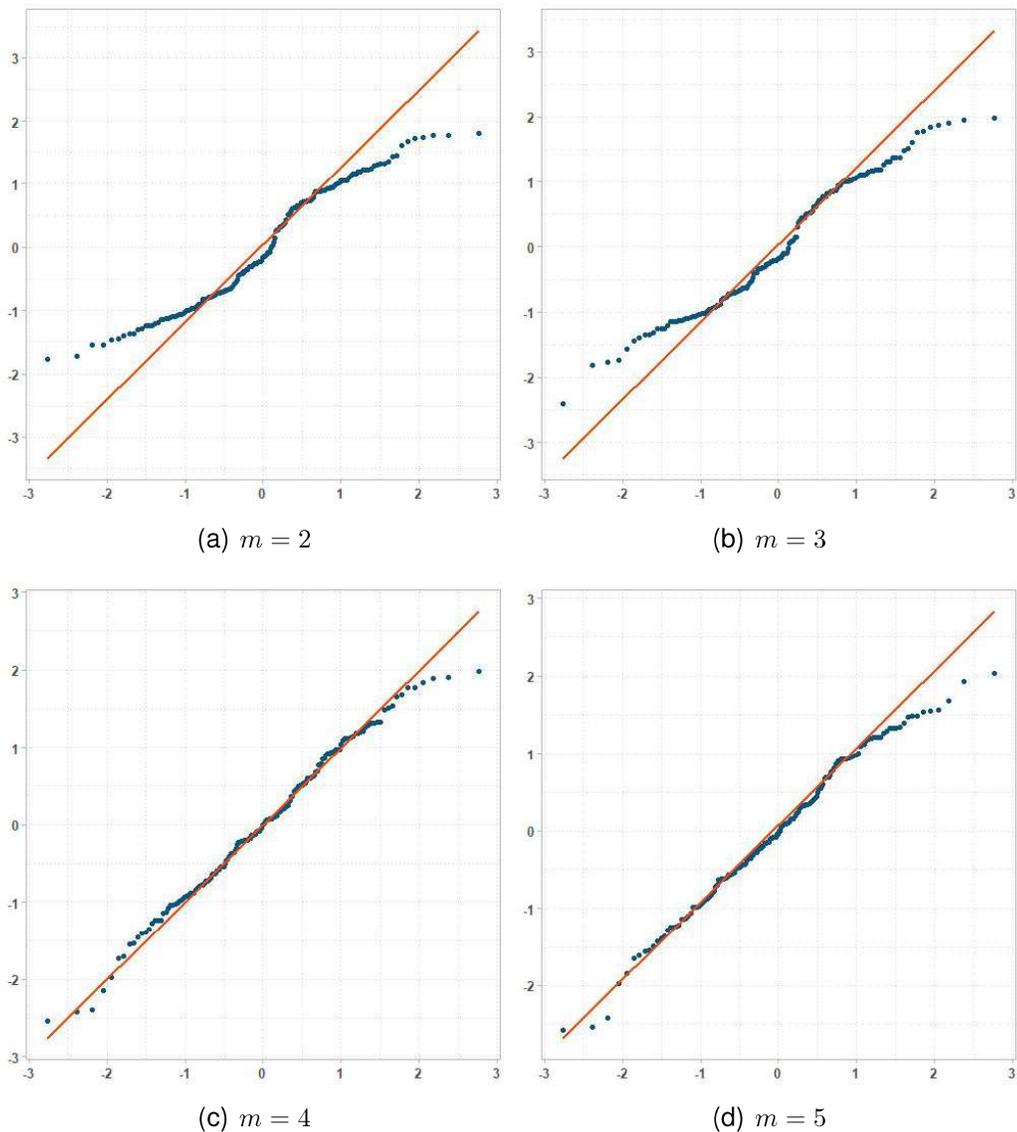


Figura 3.8: Gráficos qqplot de los pseudo-residuos de los modelos ocultos de Markov estimados para la morosidad de la cartera de crédito comercial.

que la distribución inicial está dada por:

$$\delta = \begin{pmatrix} 0,00000 & 0,00000 & 0,00000 & 1,00000 \end{pmatrix}$$

De donde se concluye que siempre inicia en el estado 4. Para la matriz de transición de probabilidades, se tiene:

$$\Gamma = \begin{pmatrix} 1,00000 & 0,00000 & 0,00000 & 0,00000 \\ 0,02080 & 0,97920 & 0,00000 & 0,00000 \\ 0,00000 & 0,04261 & 0,95739 & 0,00000 \\ 0,00000 & 0,00000 & 0,03405 & 0,96595 \end{pmatrix}$$

De donde se puede concluir que existe una alta probabilidad, cercana a 1, de que en un período parta de un estado y al siguiente vuelva al mismo; la probabilidad de partir del estado 1 a cualquiera de los demás estados es nula; la probabilidad de partir del estado 2 al estado 1 es de 0,02080 y 0 para el resto de estados excepto para sí mismo; la probabilidad de partir del estado 3 al estado 2 es de 0,04261 y 0 para el resto de estados excepto para sí mismo; y finalmente, la probabilidad de partir del estado 4 al estado 3 es de 0,03405 y 0 para el resto de estados excepto para sí mismo.

Por otro lado, las distribuciones que dependen del estado son distribuciones beta (*al tomar valores entre* $[0, 1]$), donde los parámetros para la distribución del primer estado son $\alpha = 30,27561$ y $\beta = 2391,1450$; para el segundo estado son $\alpha = 43,01097$ y $\beta = 1677,6301$; para el tercer estado son $\alpha = 16,87218$ y $\beta = 260,1111$; y finalmente, para el cuarto estado son $\alpha = 100,69080$ y $\beta = 927,5959$.

En el cuadro 3.7 se aprecian que las estimaciones realizadas por el algoritmo Baum-Welch (θ) varían de los valores inicialmente considerados (θ_0). Se observa además que los intervalos de confianza, para los parámetros del modelo oculto de Markov seleccionado, con ambos métodos son buenos, en especial para los de la distribución inicial, seguidos de las probabilidades de transición y por último los parámetros de las distribuciones dependientes de los estados, los cuales son regulares.

Luego de las estimaciones de los parámetros del modelo oculto de Markov seleccionado se procede con la decodificación de los estados, es decir, descubrir qué estado generó la observación en cada uno de los períodos. Para esto, se hace uso del algoritmo de Viterbi (*algoritmo 2.4*).

Los resultados de la decodificación se muestran en las figuras 3.9 y 3.10. En la primera figura se muestra la pertenencia de cada observación a cada uno de los 4 estados distinguidos por colores. En esta figura se puede diferenciar claramente los estados a los que pertenecen, además de una recta segmentada en cada grupo que indica su promedio respectivo. En este caso, estos estados representan niveles de riesgo, valores altos mostrarían una tendencia a una *crisis* y una mayor exposición a una quiebra, mientras que valores bajos representarían una tendencia a *no crisis* y una

Parámetro	θ_0	θ	$SE(\theta)$	$[q(\alpha/2), q(1 - \alpha/2)]$	$[\theta - 2 * SE(\theta), \theta + 2 * SE(\theta)]$
δ_1	0.2500	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
δ_2	0.2500	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
δ_3	0.2500	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
δ_4	0.2500	1.0000	0.0000	[1.0000, 1.0000]	[1.0000, 1.0000]
γ_{11}	0.9863	1.0000	0.0000	[1.0000, 1.0000]	[1.0000, 1.0000]
γ_{12}	0.0137	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{13}	0.0000	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{14}	0.0000	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{21}	0.0426	0.0208	0.0920	[0.0130, 0.2506]	[0.0000, 0.2048]
γ_{22}	0.9574	0.9792	0.0920	[0.7494, 0.9870]	[0.7952, 1.1632]
γ_{23}	0.0000	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{24}	0.0000	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{31}	0.0000	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{32}	0.0345	0.0426	0.0561	[0.0156, 0.1617]	[0.0000, 0.1548]
γ_{33}	0.8621	0.9574	0.0561	[0.8383, 0.9844]	[0.8452, 1.0696]
γ_{34}	0.1034	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{41}	0.0000	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{42}	0.0000	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{43}	0.1667	0.0340	0.0769	[0.0141, 0.2360]	[0.0000, 0.1878]
γ_{44}	0.8333	0.9660	0.0769	[0.7640, 0.9859]	[0.8122, 1.1197]
α_1	29.550	30.275	6.1131	[23.540, 40.421]	[18.049, 42.501]
α_2	45.767	43.011	10.556	[30.232, 62.386]	[21.898, 64.123]
α_3	15.802	16.872	4.8680	[11.577, 28.337]	[7.1363, 26.608]
α_4	205.81	100.67	31.786	[68.361, 168.73]	[37.104, 164.25]
β_1	2322.4	2391.1	482.70	[1860.7, 3247.9]	[1425.7, 3356.5]
β_2	1289.7	1677.6	414.91	[1173.7, 2489.2]	[847.80, 2507.4]
β_3	273.31	260.11	76.016	[178.34, 442.57]	[108.08, 412.15]
β_4	2045.6	927.47	289.15	[627.18, 1518.5]	[349.16, 1505.7]

Cuadro 3.7: Estimaciones e intervalos de confianza de los parámetros del modelo oculto de Markov seleccionado para la morosidad de la cartera comercial.

menor exposición a una quiebra.

En la segunda figura, se muestra la probabilidad de pertenencia de cada observación a cada uno de los 4 estados en todo el período de estudio. En esta figura se observa que las probabilidades están bien definidas en cada estado, es decir, las probabilidades de pertenencia al estado son muy cercanas a 1.

Posteriormente, se realiza la predicción de los estados para el año 2017, para lo cual, se hace uso de la ecuación 2.67 donde la predicción del estado se lo hace de acuerdo al mayor valor de probabilidad que se obtenga en cada uno de los estados para cada período. Estos resultados se los pueden observar en la figura 3.11 en la

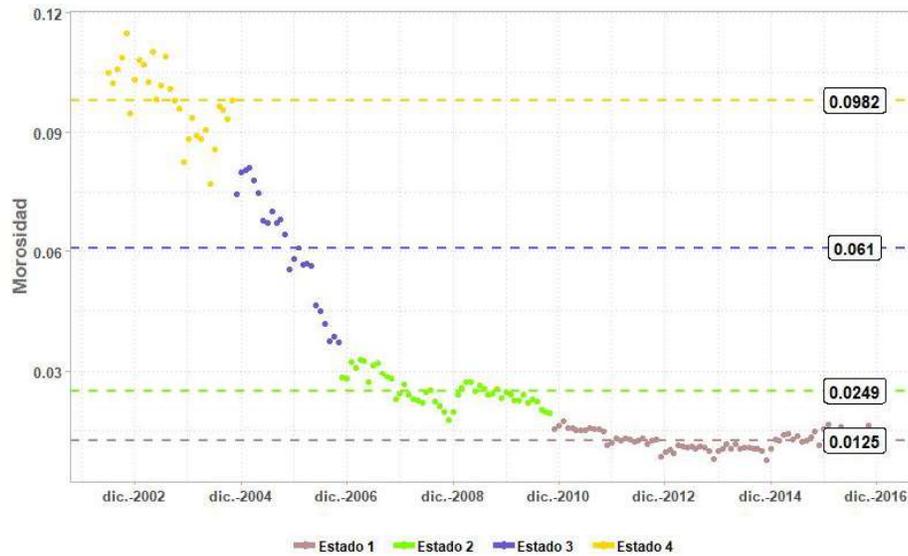


Figura 3.9: Decodificación de los estados mediante el uso del algoritmo de Viterbi para la morosidad de la cartera de crédito comercial.

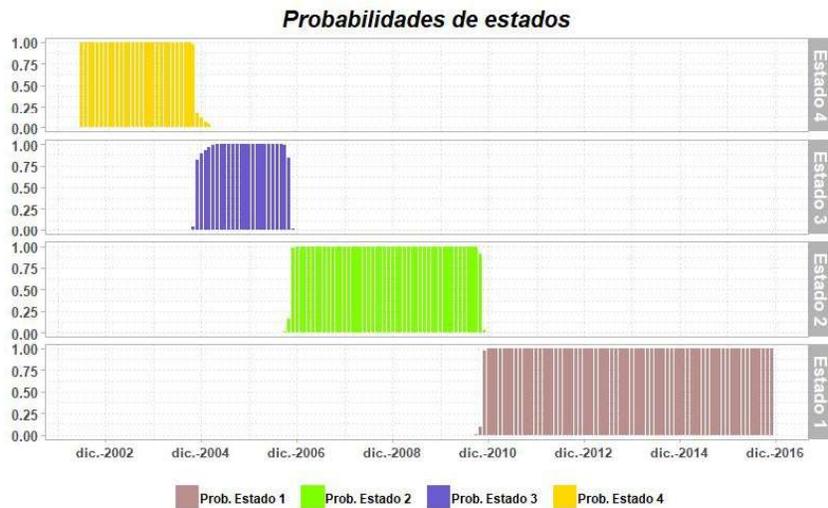


Figura 3.10: Probabilidades de pertenencia a los estados para la morosidad de la cartera de crédito comercial.

cual se aprecia que existe una alta probabilidad de pertenecer al estado 1 en los 12 períodos de predicción.

Las probabilidades de predicción del estado se presentan en el cuadro 3.8 en la cual se muestra que la probabilidad de pertenecer al estado 1 es del 100% en los 12 períodos de predicción. Estos resultados son coherentes, pues de acuerdo a la figura 3.2, las observaciones reales correspondientes al período de validación mantienen un comportamiento constante; es decir, se hallan en el mismo estado 1 (*no crisis*).



Figura 3.11: Predicción de los estados para el año 2017 para la morosidad de la cartera de crédito comercial.

Período	Probabilidad de pertenecer al estado 1	Probabilidad de pertenecer al estado 2	Probabilidad de pertenecer al estado 3	Probabilidad de pertenecer al estado 4
Enero 2017	1.0000	0.0000	0.0000	0.0000
Febrero 2017	1.0000	0.0000	0.0000	0.0000
Marzo 2017	1.0000	0.0000	0.0000	0.0000
Abril 2017	1.0000	0.0000	0.0000	0.0000
Mayo 2017	1.0000	0.0000	0.0000	0.0000
Junio 2017	1.0000	0.0000	0.0000	0.0000
Julio 2017	1.0000	0.0000	0.0000	0.0000
Agosto 2017	1.0000	0.0000	0.0000	0.0000
Septiembre 2017	1.0000	0.0000	0.0000	0.0000
Octubre 2017	1.0000	0.0000	0.0000	0.0000
Noviembre 2017	1.0000	0.0000	0.0000	0.0000
Diciembre 2017	1.0000	0.0000	0.0000	0.0000

Cuadro 3.8: Probabilidades de predicción del estado para el año 2017 para la morosidad de la cartera de crédito comercial.

Como siguiente paso se procede a calcular las distribuciones de pronóstico, para lo cual se hace uso de la ecuación 2.60 para cada uno de los 12 períodos de validación.

Dichas distribuciones se encuentran en la figura 3.12, en la cual se muestran las distribuciones de pronóstico para el año 2017. En esta figura se muestra la densidad de pronóstico (*en azul*) con su respectiva distribución de probabilidad (*en rojo*).

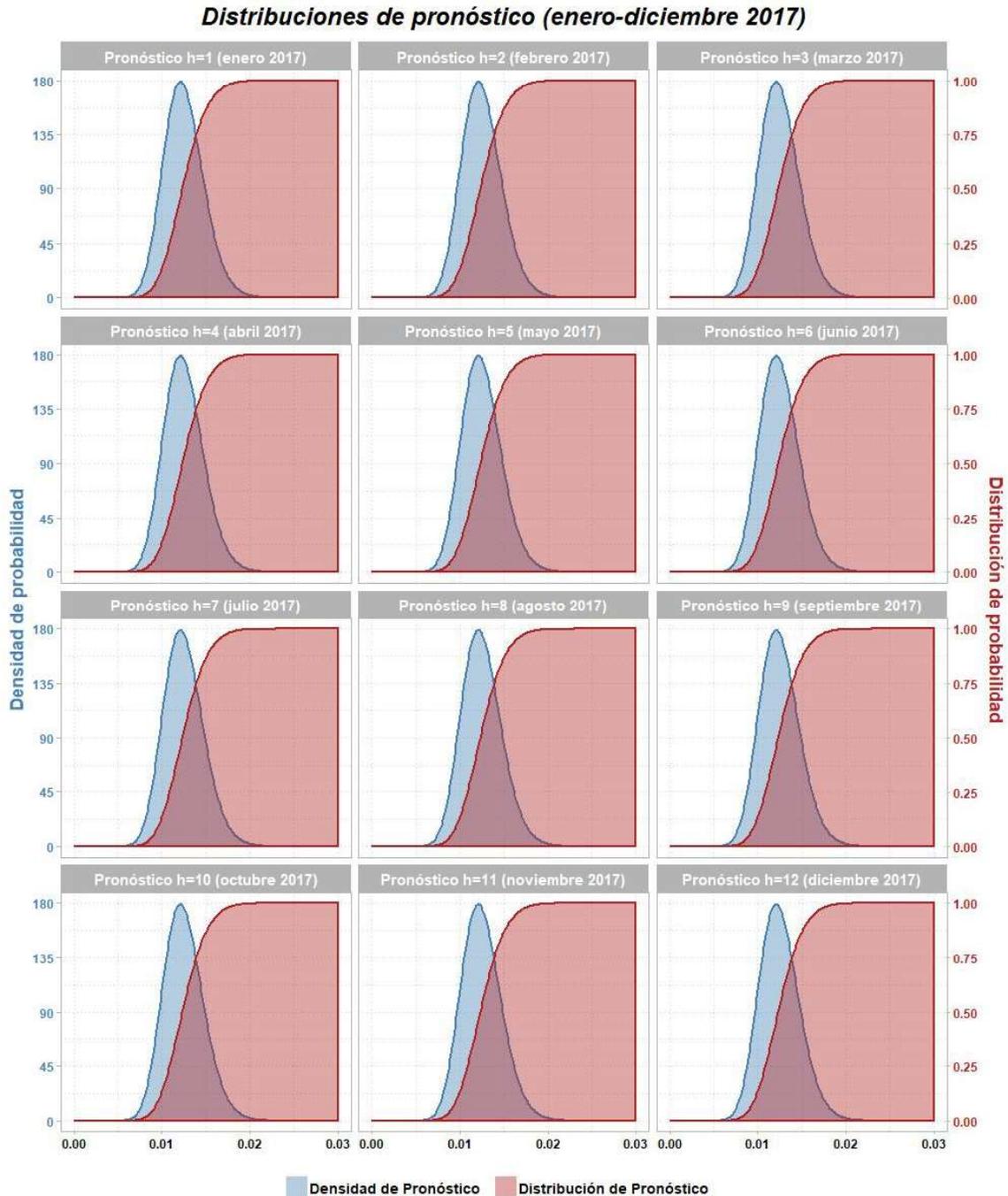


Figura 3.12: Densidades y distribuciones de pronóstico para la morosidad de la cartera de crédito comercial.

En esta figura se puede observar que en los 12 períodos de validación se tienen

densidades de pronóstico muy similares. Esto muestra que en los próximos períodos se tendrán valores de pronósticos muy parecidos. Además, estos resultados son coherentes, pues las observaciones reales tienen un comportamiento parecido; es decir, se mantienen una tendencia constante con una ligera variabilidad.

Basados en dichas distribuciones se muestran algunos estadísticos descriptivos para cada período de validación, de acuerdo con las ecuaciones 3.3, 3.4, 3.5, 3.6, 3.7 y 3.8. Estos resultados se visualizan en el cuadro 3.9.

Estadísticos	Enero 2017	Febrero 2017	Marzo 2017	Abril 2017	Mayo 2017	Junio 2017	Julio 2017	Agosto 2017	Septiembre 2017	Octubre 2017	Noviembre 2017	Diciembre 2017
Horizonte	1	2	3	4	5	6	7	8	9	10	11	12
Media	0.01250	0.01250	0.01250	0.01250	0.01250	0.01250	0.01250	0.01250	0.01250	0.01250	0.01250	0.01250
Varianza	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001
Desv.Est	0.00226	0.00226	0.00226	0.00226	0.00226	0.00226	0.00226	0.00226	0.00226	0.00226	0.00226	0.00226
Moda	0.01210	0.01210	0.01210	0.01210	0.01210	0.01210	0.01210	0.01210	0.01210	0.01210	0.01210	0.01210
Cuartil 1	0.01080	0.01080	0.01080	0.01080	0.01080	0.01080	0.01080	0.01080	0.01080	0.01080	0.01080	0.01080
Cuartil 2	0.01230	0.01230	0.01230	0.01230	0.01230	0.01230	0.01230	0.01230	0.01230	0.01230	0.01230	0.01230
Cuartil 3	0.01380	0.01380	0.01380	0.01380	0.01380	0.01380	0.01380	0.01380	0.01380	0.01380	0.01380	0.01380

Cuadro 3.9: Estadísticos de las distribuciones de pronósticos para el año 2017 para la morosidad de la cartera de crédito comercial.

En el cuadro 3.9 se puede apreciar que todos los estadísticos en cada uno de los 12 períodos de validación son idénticos, lo que también se puede visualizar en la figura 3.12, pues su comportamiento es similar en cada período.

Para observar el poder predictivo del modelo oculto de Markov, se presenta el ajuste al conjunto de entrenamiento, usando la ecuación 2.58 y los pronósticos obtenidos con dicho modelo al conjunto de validación (2017). Puesto que, en cada período se dispone de una distribución, para obtener un valor de ajuste y pronóstico, respectivamente, se pretende usar las medidas de tendencia central, tales como: media, mediana y moda; para cada período. Además, se realiza un ajuste con el modelo ARIMA para la cartera en estudio y se realizan pronósticos para la validación. Los resultados del modelo ARIMA se encuentra en el anexo A.

En la figura 3.13 se presenta una comparación gráfica de los 2 modelos con las observaciones reales, donde se observa que el modelo ARIMA proporciona un mejor ajuste a las observaciones reales que el modelo oculto de Markov, aunque éste modela muy bien la tendencia de la serie usando como ajuste y pronóstico la media, mediana y moda. En ambos modelos, los pronósticos son muy buenos, pues visualmente son muy cercanos.

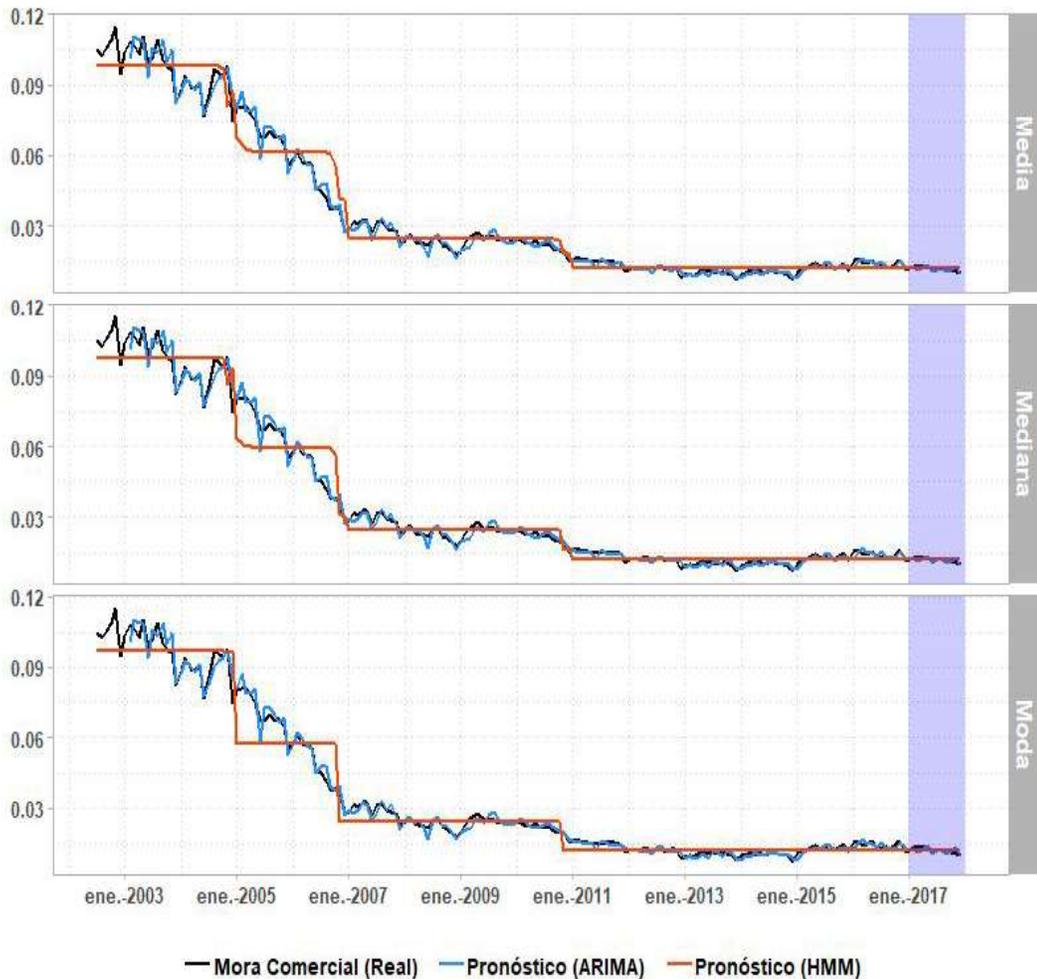


Figura 3.13: Ajustes y pronósticos con el modelo oculto de Markov y el modelo ARIMA con las observaciones reales para la morosidad de la cartera de crédito comercial.

En un modelo ARIMA se fija un estado del espacio de estados y se varía respecto al espacio parametral, por lo que al pronosticar se lo realiza período a período. Mientras que en un modelo oculto de Markov varía respecto al espacio de estados y al espacio parametral, por lo que en cada período se obtiene una distribución de pronóstico y para obtener un valor se estima la media, mediana o moda como pronóstico. De esta manera, al comparar los resultados, se tiene que el del modelo oculto de Markov no son tan buenos, como el del modelo ARIMA, pues sus medidas de exactitud son mejores (*tabla 3.10*), pero si se encuentran en el dentro del estado predicho (*estado 1*). Esto hacen que sean mejores sobre los modelo ARIMA pues predicen estados y pronósticos dentro de ellos.

	Pronóstico ARIMA	Pronóstico HMM (Media)	Pronóstico HMM (Mediana)	Pronóstico HMM (Moda)
EAM	0.00072814	0.00086586	0.00085320	0.00086468
CME	0.00000100	0.00000113	0.00000108	0.00000111
EPAM	6.02 %	7.39 %	7.17 %	7.15 %

Cuadro 3.10: Medidas de exactitud para la morosidad de la cartera de crédito comercial.

En la figura 3.13 se presentan los ajustes (*conjunto de entrenamiento*) y los pronósticos (*conjunto de validación*) del modelo ARIMA y del modelo oculto de Markov, y se los compara con las observaciones reales. En esta figura se observan buenos resultados del modelo ARIMA sobre el modelo oculto de Markov al realizar pronósticos punto a punto. Por otro lado, los resultados del modelo oculto de Markov también son buenos pero para estimar estados en cada período. Esta figura puede ser de gran interés y utilidad para un ente regulador, pues permite conocer dónde se encontró y se encontrará la situación de la cartera comercial y sus fluctuaciones en el tiempo.

Por otro lado, no se debe olvidar que los modelos ocultos de Markov no realizan pronósticos mediante una expresión, como lo hacen los modelos ARIMA, sino que proporcionan probabilidades que se expresan en densidades (caso continuo) y distribuciones de probabilidad.

Además, en la figura 3.14 se muestra los pronósticos de los modelos aplicados con sus respectivos intervalos de confianza y las observaciones reales en donde se observa que las observaciones reales están contenidas en el intervalos de confianza del modelo oculto de Markov y éstas a su vez están contenidas en los intervalos de confianza del modelo ARIMA. Lo que cual indica que el modelo oculto de Markov reporta pronósticos más certeros.

Finalmente, los pronósticos del modelo oculto de Markov generan buenos resultados, pues las medidas de exactitud 3.9 y 3.10 reportan valores pequeños de aproximadamente 0,00000103 para el caso del cuadrado medio debido al error y de 7,17 % para el caso del error porcentual absoluto medio.

3.2.2. Modelamiento de la cartera de crédito de consumo

En la figura 3.15 se muestra la evolución de la morosidad de la cartera de crédito de consumo donde la sección sombreada representa al conjunto de observaciones de validación y la sección no sombreada representa el conjunto de observaciones de

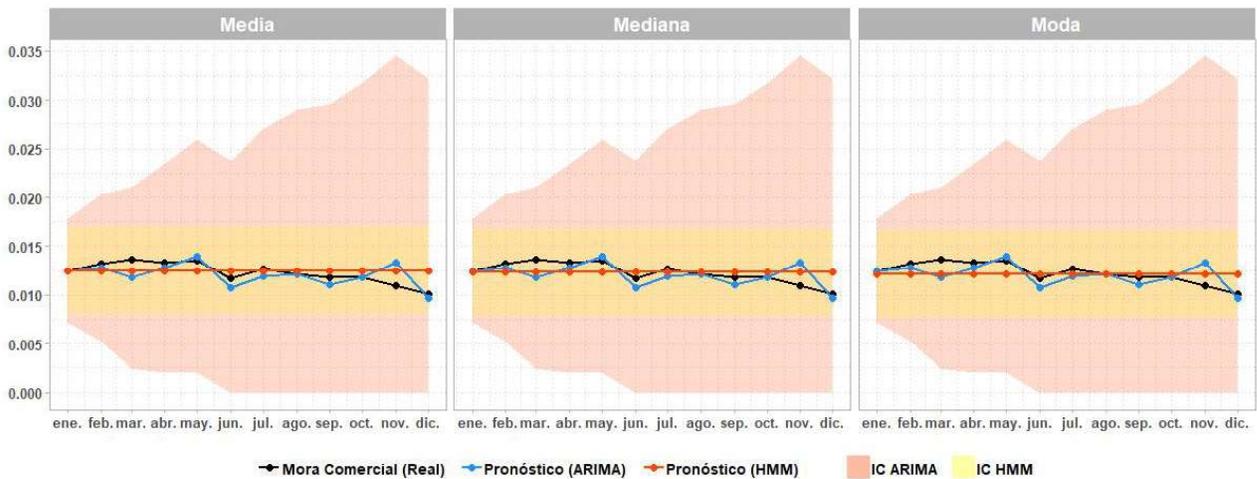


Figura 3.14: Comparación de pronósticos e intervalos de confianza de los modelos con las observaciones reales de la morosidad de la cartera de crédito comercial.

entrenamiento. En esta figura se observa que la morosidad se comporta de manera volátil, es decir, existe momento en que la morosidad aumenta hasta un valor máximo aproximado de 12% y en otros períodos baja considerablemente hasta un valor mínimo aproximado de 3%. Además, en todo el período es muy variable, para finalmente, en el período de validación se comporta de manera decreciente hasta llegar a valores cercanos al 5%.

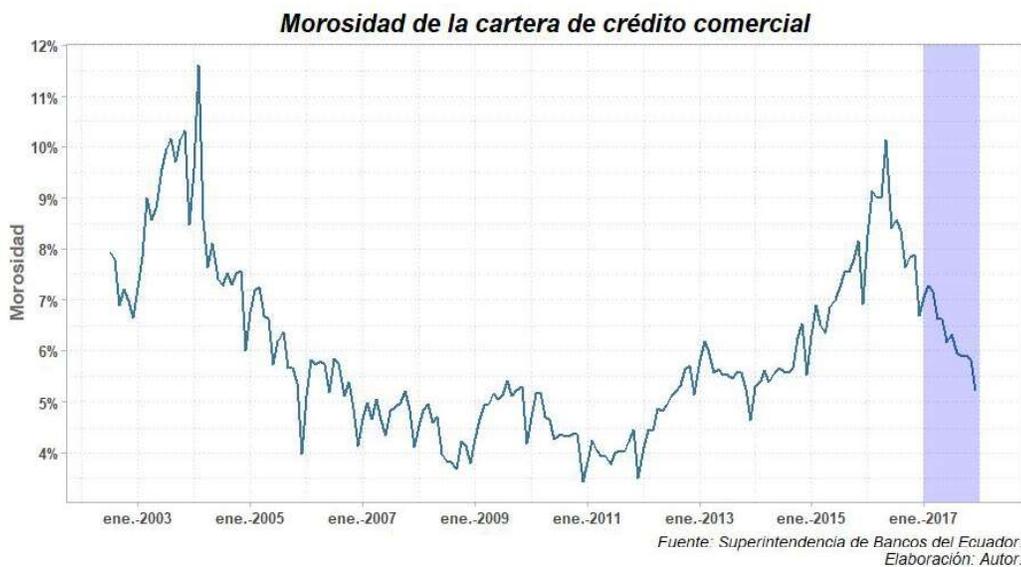


Figura 3.15: Serie de morosidad de la cartera de crédito de consumo.

Para modelar la serie de morosidad de consumo se usan 174 observaciones que

conforman el conjunto de valores de entrenamiento donde el valor de la media es de 0,0597 con un valor mínimo de 0,0344 y un valor máximo de 0,1162, como se aprecia en el cuadro 3.11.

Nro. Obs.	Media	Desv. Est.	Mínimo	1er cuartil	2do cuartil	3er cuartil	Máximo
174	0.0597	0.0169	0.0344	0.0469	0.0554	0.0714	0.1162

Cuadro 3.11: Cuadro de estadísticos de la serie de morosidad de consumo.

En la figura 3.16 se muestra la distribución de los valores de morosidad de la cartera de consumo junto a su media (*recta segmentada vertical*). En ésta se observa que existe una gran cantidad de observaciones entre 0 y 0,062, otra parte de valores se distribuye con menor frecuencia entre 0,062 y 0,087; y el restante con frecuencia muy pequeña entre 0,087 y 0,12. Esto también se corrobora en la figura 3.15.

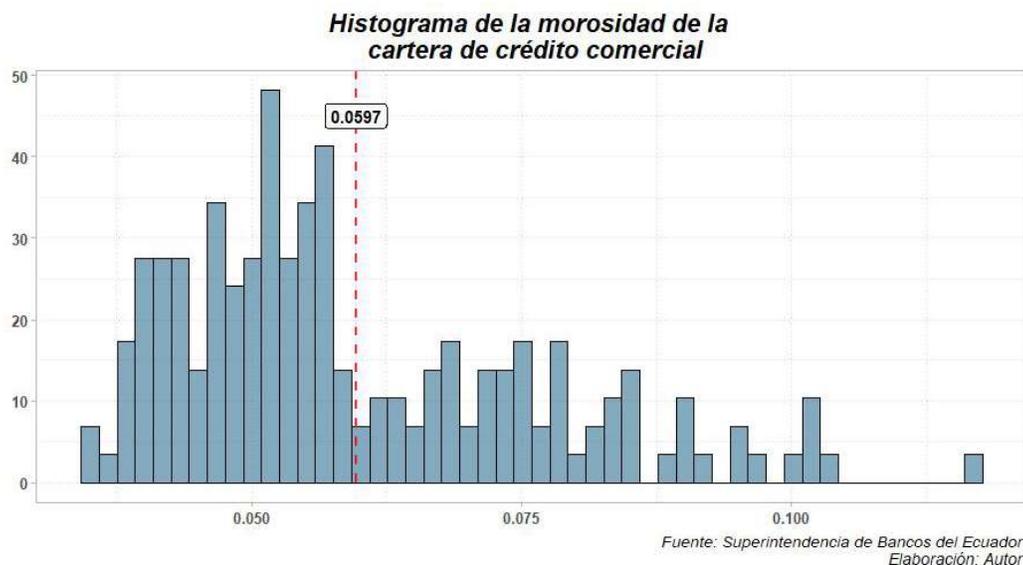


Figura 3.16: Histograma de la cartera de crédito de consumo.

Para comenzar a modelar la serie de morosidad se empieza creando los grupos o clústers, que representarán niveles de riesgo y que a la vez formarán el conjunto finito de estados de la cadena de Markov. En este caso, cuando los valores son cercanos a 1, representaría un estado de *crisis* indicando que un quiebre del sistema financiero está cerca. Por el contrario, valores cercanos a 0 indicarían que un quiebre del sistema financiero está lejos de producirse; es decir, se definiría como un estado de *no crisis*. Para esto, se utilizan los modelos de mezcla con los cuales se generan los grupos en los cuales las observaciones de morosidad siguen una distribución beta con parámetros determinados por las observaciones que conforman cada grupo y que pueden ser

estimados mediante el método de los momentos, dadas por las ecuaciones 3.1 y 3.2.

De esta manera, se considera crear modelos de mezcla con 2, 3, 4 y 5 grupos, los cuales se pueden apreciar en la figura 3.17, donde la recta segmentada representa la media del grupo generado por el modelo de mezcla respectivo en cada caso.

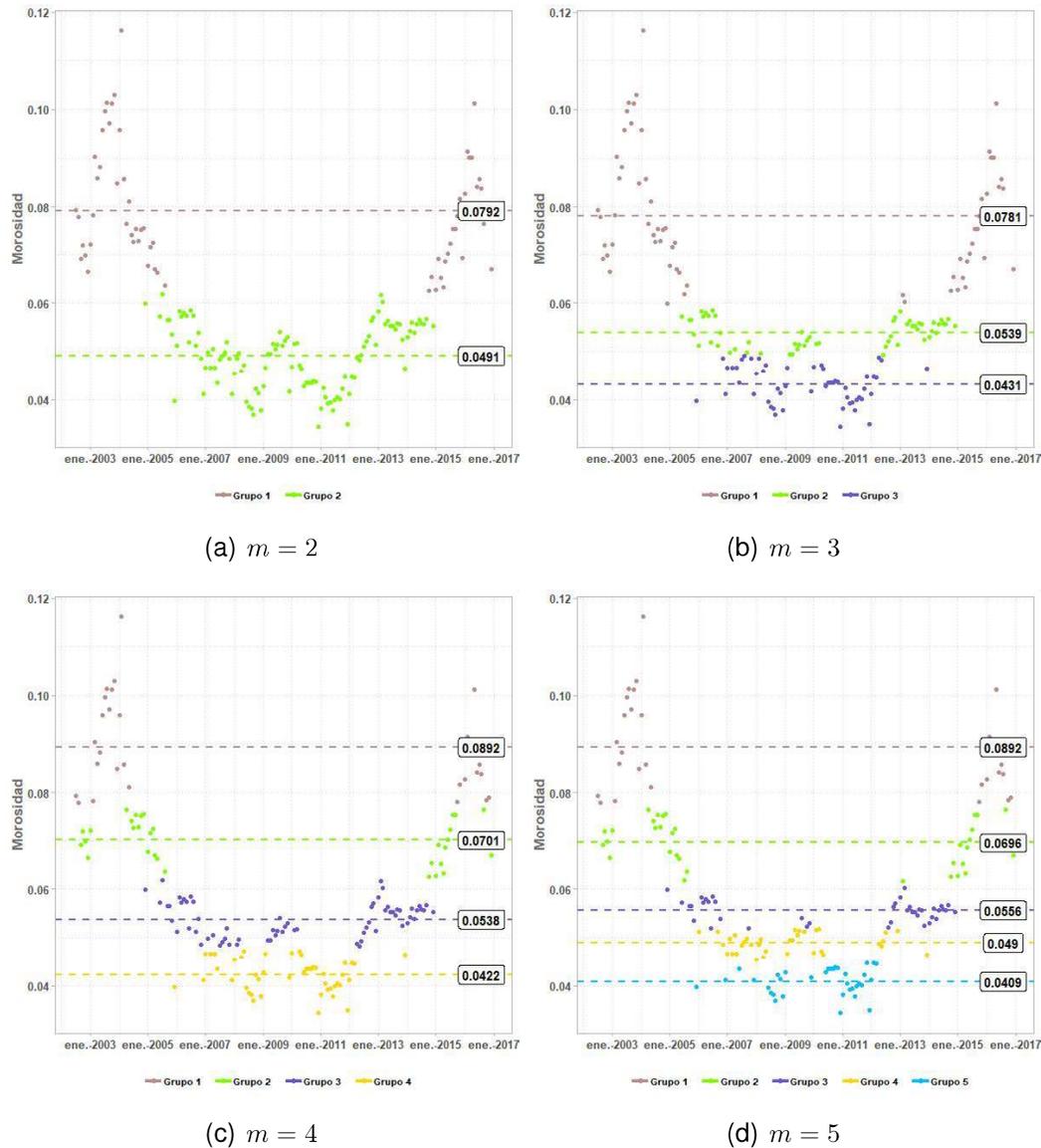


Figura 3.17: Grupos generados con los modelos de mezclas para la morosidad de la cartera de crédito de consumo.

Los clústers formados con los diferentes modelos de mezcla estimados, son usados para estimar los parámetros de la distribución beta de cada clúster para cada modelo de mezcla usando las ecuaciones 3.1 y 3.2. Los resultados de los parámetros estimados se encuentran tabulados en los cuadros 3.12, 3.13, 3.14 y 3.15.

	ω	α	β
Grupo 1	0.6232966	51.49076	997.446
Grupo 2	0.3767034	39.67194	461.131

Cuadro 3.12: Parámetros para el modelo de mezcla con 2 grupos para la morosidad de la cartera de crédito de consumo.

	ω	α	β
Grupo 1	0.4160523	35.97781	424.7355
Grupo 2	0.3103734	284.83788	5054.2709
Grupo 3	0.2735743	148.56021	3369.8969

Cuadro 3.13: Parámetros para el modelo de mezcla con 3 grupos para la morosidad de la cartera de crédito de consumo.

	ω	α	β
Grupo 1	0.1975562	78.35106	799.6985
Grupo 2	0.1623688	241.68565	3204.1084
Grupo 3	0.3339503	259.78383	4530.5827
Grupo 4	0.3061247	126.44338	2825.2528

Cuadro 3.14: Parámetros para el modelo de mezcla con 4 grupos para la morosidad de la cartera de crédito de consumo.

	ω	α	β
Grupo 1	0.2476922	159.72714	3653.6923
Grupo 2	0.2240059	614.68032	11500.1421
Grupo 3	0.1425048	2370.77346	39647.2276
Grupo 4	0.1883416	174.64856	2352.2272
Grupo 5	0.1974555	78.35106	799.6985

Cuadro 3.15: Parámetros para el modelo de mezcla con 5 grupos para la morosidad de la cartera de crédito de consumo.

Los parámetros estimados son utilizados para calcular la densidad de mezcla de acuerdo a la ecuación 2.1. Las figuras de las densidades de mezclas resultantes se presentan en 3.18. En las cuales se aprecia la mejora del ajuste de la densidad de mezcla a medida que se aumenta el número de grupos.

Con ayuda de los grupos generados con los modelos de mezcla, se generan los valores iniciales que ayudarán a la estimación de los parámetros del modelo oculto de Markov usando el algoritmo de Baum-Welch. Los parámetros a estimarse son: la distribución inicial de la cadena de Markov (δ), la matriz de transición de probabilidades (Γ) y los parámetros de la distribución beta que dependen del estado. Para la distribución

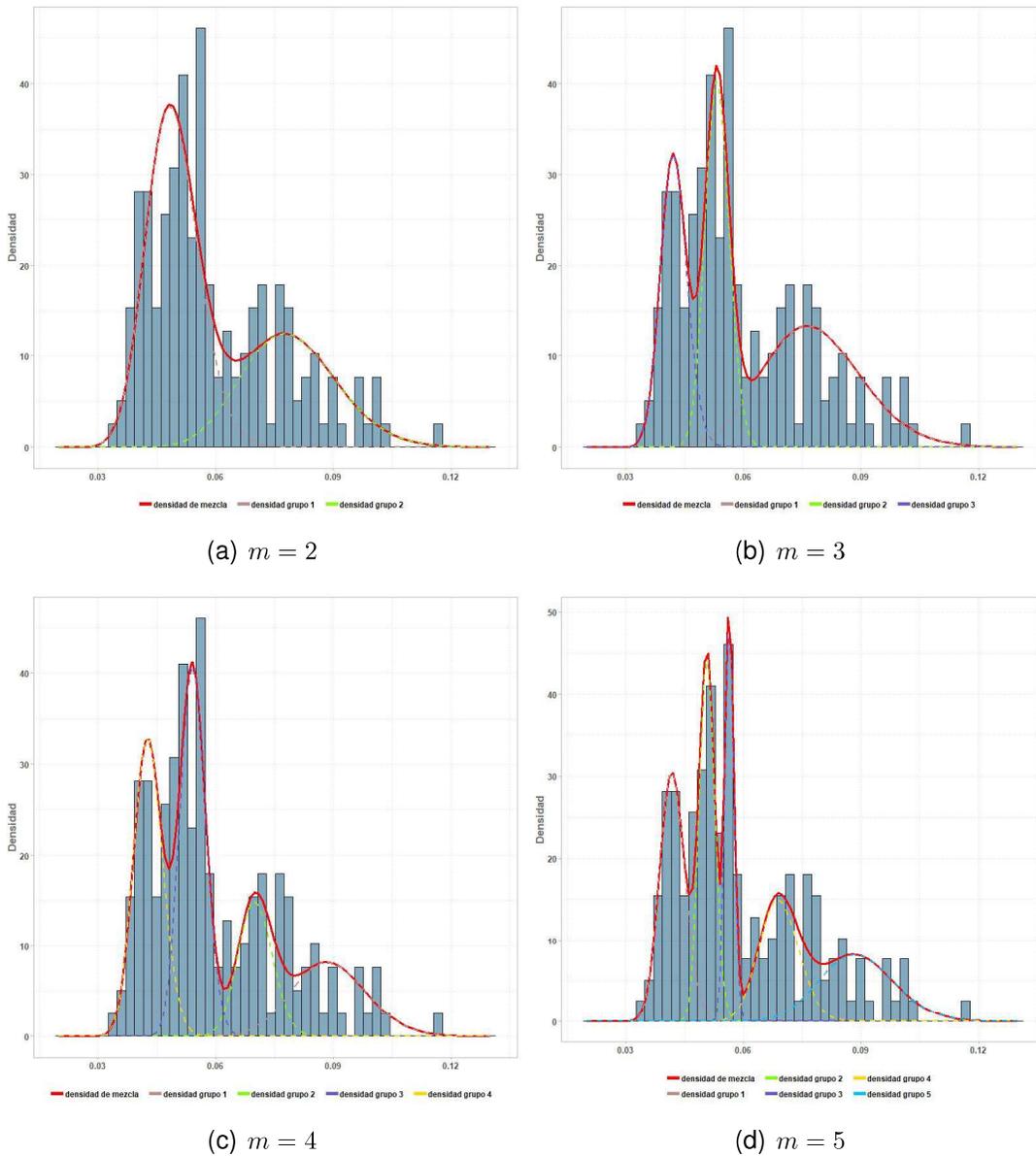


Figura 3.18: Densidades de mezcla formados por los modelos de mezcla para la morosidad de la cartera de crédito de consumo.

inicial se asignarán valores equiprobables; para la matriz de transición se estimará de acuerdo a los grupos formados con los modelos de mezcla estimados, con ayuda de la ecuación 2.15; y por último, para los parámetros de la distribuciones dependientes del estado (α, β) se usarán los valores obtenidos de los modelos de mezcla que se encuentran tabulados en los cuadros 3.12, 3.13, 3.14 y 3.15.

De esta manera, los valores iniciales para el algoritmo de Baum-Welch con $m = 2, 3, 4$ y 5 estados, son los siguientes:

Para $m = 2$:

$$\delta = (1/2, 1/2)$$
$$\Gamma = \begin{pmatrix} 0,96 & 0,04 \\ 0,07 & 0,93 \end{pmatrix}$$

Para $m = 3$:

$$\delta = (1/3, 1/3, 1/3)$$
$$\Gamma = \begin{pmatrix} 0,94 & 0,06 & 0,00 \\ 0,06 & 0,80 & 0,14 \\ 0,00 & 0,20 & 0,80 \end{pmatrix}$$

Para $m = 4$:

$$\delta = (1/4, 1/4, 1/4, 1/4)$$
$$\Gamma = \begin{pmatrix} 0,79 & 0,21 & 0,00 & 0,00 \\ 0,16 & 0,71 & 0,13 & 0,00 \\ 0,00 & 0,07 & 0,77 & 0,16 \\ 0,00 & 0,00 & 0,20 & 0,80 \end{pmatrix}$$

Para $m = 5$:

$$\delta = (1/5, 1/5, 1/5, 1/5, 1/5)$$
$$\Gamma = \begin{pmatrix} 0,77 & 0,23 & 0,00 & 0,00 & 0,00 \\ 0,26 & 0,59 & 0,15 & 0,00 & 0,00 \\ 0,00 & 0,22 & 0,63 & 0,15 & 0,00 \\ 0,00 & 0,00 & 0,12 & 0,74 & 0,14 \\ 0,00 & 0,00 & 0,00 & 0,21 & 0,79 \end{pmatrix}$$

Con los valores iniciales anteriores se estiman los modelos ocultos de Markov con $m = 2, 3, 4$ y 5 estados, respectivamente. De los cuales se seleccionó como mejor modelo a aquel que tiene 4 estados, debido a que el criterio bayesiano es el menor de los 4 modelos calculados. Además, se acepta la hipótesis nula de normalidad de los pseudo-residuos, pues el valor-p es mayor a 0,05. Por otro lado, se tiene un modelo parsimonioso, pues no es complejo, al no tener muchos parámetros por estimar (*bajo coste computacional*), satisface los criterios de normalidad y se previene un posible sobre ajuste.

El número de parámetros estimados dependerá del número de estados seleccionado; es decir, para m estados, en la distribución inicial δ se deben calcular $m - 1$

parámetros, pues su suma debe ser 1; en la matriz de transición Γ se deben calcular $m^2 - m$ estados, pues su suma por fila debe ser 1; por último, en los parámetros que dependen del estado (α, β) se deben calcular $2m$ parámetros. Lo que da un total de $m^2 + 2m$ parámetros por calcular. Así por ejemplo, si $m = 2$ se tendrá que calcular 7 parámetros. De manera similar para $m = 3, 4, 5$, respectivamente.

Estos resultados se encuentran resumidos en el cuadro 3.16.

Nro. de estados	Nro. de parámetros estimados	log. verosimilitud	AIC	BIC	p valor
2	7.0000	578.6421	-1143.2842	-1121.1708	0.5506
3	14.0000	622.3443	-1216.6886	-1172.4618	0.6144
4	23.0000	646.1245	-1246.2490	-1173.5907	0.2101
5	34.0000	664.1772	-1260.3544	-1152.9465	0.0058

Cuadro 3.16: Criterios para seleccionar el mejor modelo oculto de Markov para la morosidad de la cartera de crédito de consumo.

Se presenta además, el gráfico de los pseudo-residuos para cada uno de los modelos ocultos de Markov, en los cuales se aprecian que con $m = 2$ estados, presenta variabilidad con ligeros patrones; con $m = 3$ estados, los pseudo-residuos se muestran dispersos en comparación al anterior; con $m = 4$ y $m = 5$ estados se muestra una mayor dispersión. Lo que se aprecia en la figura 3.19, donde además se presentan sus límites al 95% y 99% de confianza (*franjaz azules y rojas, respectivamente*).

En la figura 3.20, se muestra la distribución de los pseudo-residuos con la respectiva gráfica de la densidad de la distribución normal estándar. En estas figuras se observan que los histogramas con $m = 1$ y $m = 5$ estados no tienen un comportamiento similar al de una normal estándar, lo que si sucede en los pseudo-residuos cuando $m = 3$ y $m = 4$ estados.

Para corroborar que los pseudo-residuos del modelo oculto de Markov con 4 estados siguen una distribución normal estándar. Se realiza un gráfico cuantil-cuantil en donde, mientras más cercanos a la recta identidad se encuentren los pseudo-residuos, se concluye que siguen una distribución normal estándar. Esto sucede cuando $m = 3$ y $m = 4$; para $m = 2$ se muestra una ligera desviación; y para $m = 5$ se muestra una gran diferencia, como se puede apreciar en la figura 3.21.

Se seleccionó como mejor modelo oculto de Markov para la morosidad de la cartera de crédito de consumo a aquel con 4 estados, debido a que su valor del criterio bayesiano es menor y sus pseudo-residuos son normales. El modelo oculto de Markov

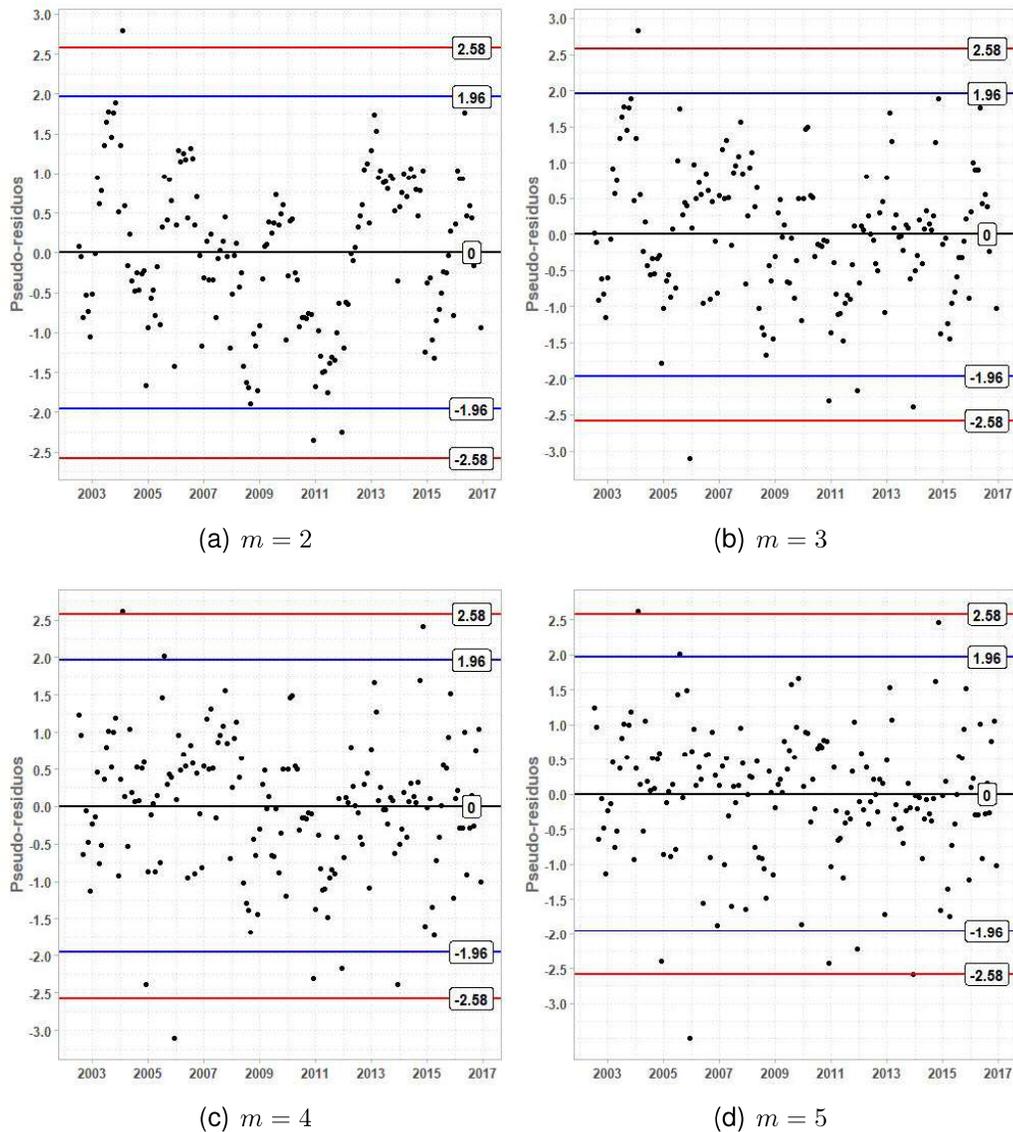


Figura 3.19: Gráfico de los pseudo-residuos de los modelos ocultos de Markov estimados para la morosidad de la cartera de crédito de consumo.

con 2 y 3 estados también pudieron haber sido seleccionados, pues su valor-p es mayor a 0.05, es decir, sus pseudo-residuos son normales pero su criterio de Akaike y bayesiano son altos. Además, no se consideró al modelo con 5 estados, ya que a pesar de tener el valor más pequeño en el criterio Akaike, su valor-p es menor a 0.05, es decir, sus pseudo-residuos no son normales. Además de ser muy complejo, pues requiere muchos cálculos para obtener sus parámetros.

Para el modelo seleccionado se estiman los parámetros del modelo oculto de Markov con sus respectivos errores estándar e intervalos de confianza de acuerdo al mé-

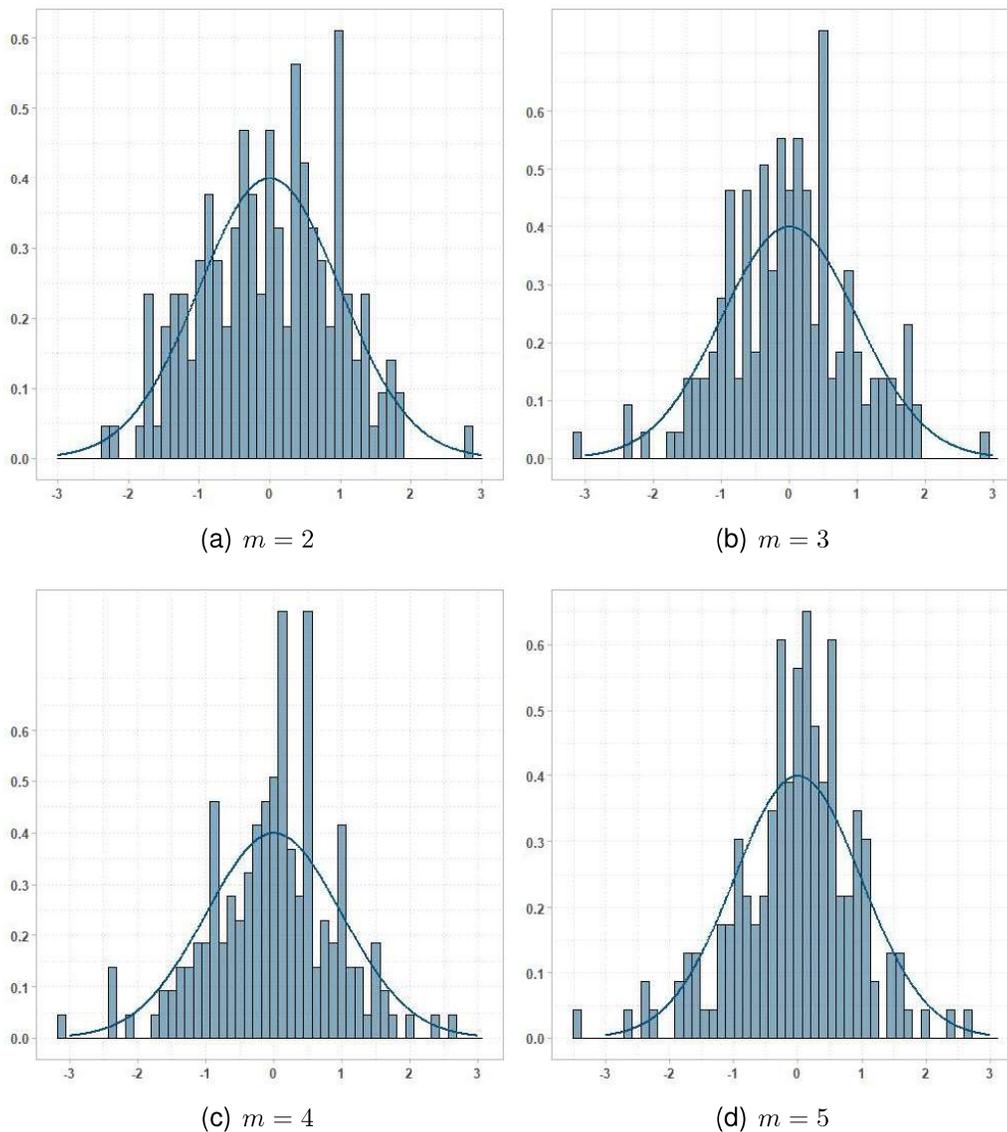


Figura 3.20: Histogramas de los pseudo-residuos de los modelos ocultos de Markov estimados para la morosidad de la cartera de crédito de consumo.

todo del percentil (ecuación 2.55) y al método de Bootstrap (ecuación 2.56) al 95% de confianza. Estos resultados se encuentran tabulados en el cuadro 3.17, donde θ_0 denota los valores iniciales para el algoritmo de Baum-Welch; θ denota los valores resultantes de los parámetros luego de aplicar el algoritmo; $SE(\theta)$ denota el error estándar de bootstrap para cada parámetro estimado; y finalmente, $[q(\alpha/2), q(1 - \alpha/2)]$ y $[\theta - 2 * SE(\theta), \theta + 2 * SE(\theta)]$ denotan los intervalos de confianza con el método del percentil y de bootstrap, respectivamente.

Respecto a las estimaciones del modelo oculto de Markov con 4 estados, se tiene

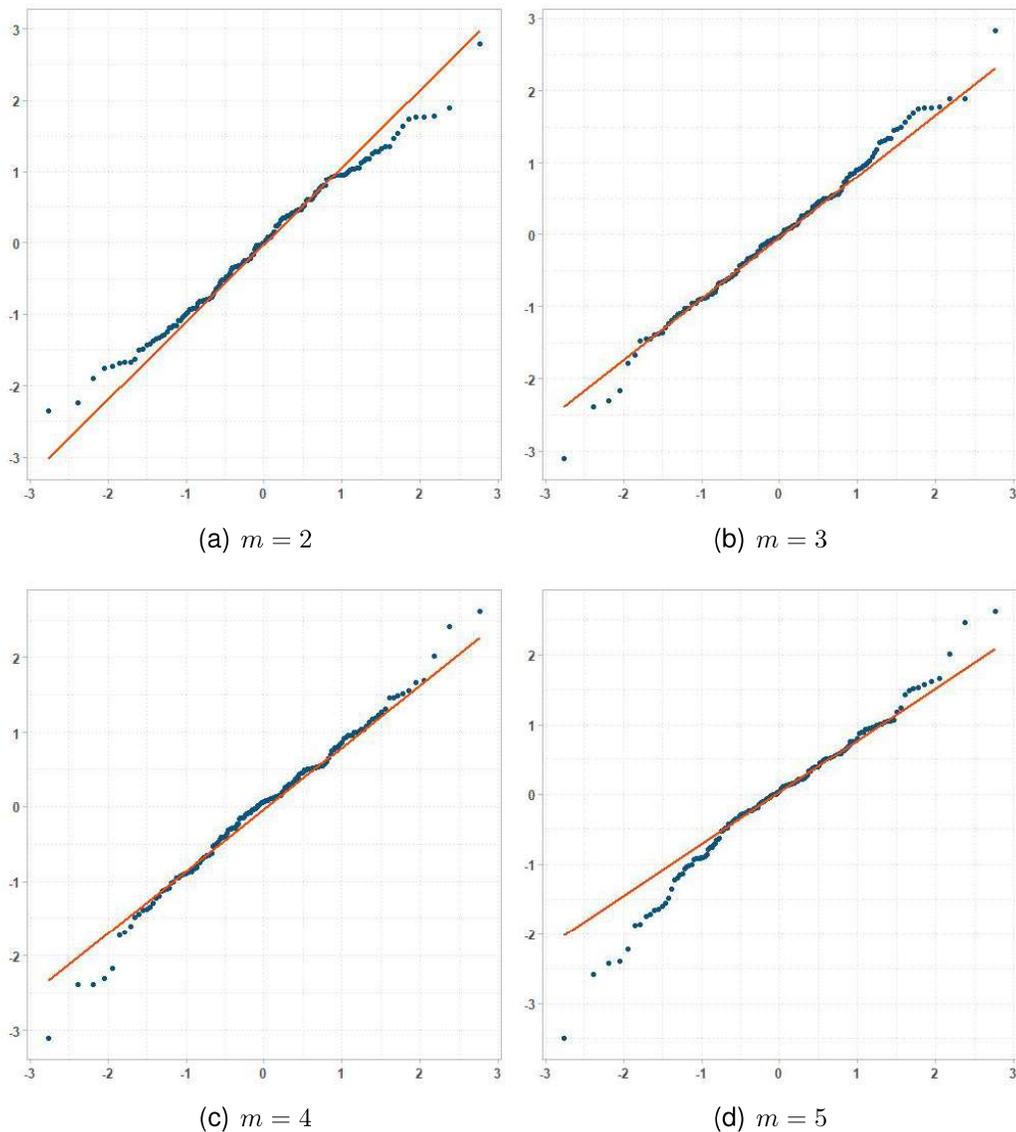


Figura 3.21: Gráficos qqplot de los pseudo-residuos de los modelos ocultos de Markov estimados para la morosidad de la cartera de crédito de consumo.

que la distribución inicial está dada por:

$$\delta = \begin{pmatrix} 0,00000 & 1,00000 & 0,00000 & 0,00000 \end{pmatrix}$$

De donde se concluye que siempre inicia en el estado 2. Para la matriz de transición de probabilidades, se tiene:

$$\Gamma = \begin{pmatrix} 0,90298 & 0,09702 & 0,00000 & 0,00000 \\ 0,05556 & 0,91655 & 0,02789 & 0,00000 \\ 0,00000 & 0,01926 & 0,92031 & 0,06043 \\ 0,00000 & 0,00000 & 0,05121 & 0,94879 \end{pmatrix}$$

De donde se puede concluir que existe una alta probabilidad, cercana a 1, de que en un período parta de un estado y al siguiente vuelva al mismo; la probabilidad de partir del estado 1 al estado 2 es de 0,09702 y 0 para el resto de estados excepto para sí mismo; la probabilidad de partir del estado 2 al estado 1 es de 0,05556, al estado 3 es de 0,02789 y 0 para el estado 4; la probabilidad de partir del estado 3 al estado 1 es 0, al estado 2 de 0,01926 y al estado 4 es de 0,06043; y finalmente, la probabilidad de partir del estado 4 al estado 3 es de 0,05121 y 0 para el resto de estados excepto para sí mismo.

Por otro lado, las distribuciones que dependen del estado son distribuciones beta (*al ser datos entre* $[0, 1]$), donde los parámetros para la distribución del primer estado son $\alpha = 107,26117$ y $\beta = 1047,848$; para el segundo estado son $\alpha = 163,30459$ y $\beta = 2088,441$; para el tercer estado son $\alpha = 217,59265$ y $\beta = 3703,780$; y finalmente, para el cuarto estado son $\alpha = 89,51251$ y $\beta = 1931,034$.

En el cuadro 3.7 se aprecian que las estimaciones realizadas por el algoritmo Baum-Welch (θ) varían de los valores inicialmente considerados (θ_0). Se observa además que los intervalos de confianza, para los parámetros del modelo oculto de Markov, con ambos métodos son buenos, en especial para los de la distribución inicial, seguidos de las probabilidades de transición y por último los parámetros de las distribuciones dependientes de los estados, los cuales son regulares.

Luego de las estimaciones de los parámetros del modelo oculto de Markov seleccionado se procede con la decodificación de los estados, es decir, descubrir qué estado generó la observación en cada uno de los períodos. Para esto, se hace uso del algoritmo de Viterbi (*algoritmo 2.4*).

Los resultados de la decodificación se muestran en las figuras 3.22 y 3.23. En la primera figura se muestra la pertenencia de cada observación a cada uno de los 4 estados distinguidos por colores. En esta figura se puede diferenciar claramente los estados a los que pertenecen, además de una recta segmentada en cada grupo que indica su promedio respectivo. En este caso, estos estados representan niveles de riesgo, valores altos mostrarían una tendencia a una *crisis* y una mayor exposición a

Parámetro	θ_0	θ	$SE(\theta)$	$[q(\alpha/2), q(1 - \alpha/2)]$	$[\theta - 2 * SE(\theta), \theta + 2 * SE(\theta)]$
δ_1	0.2500	0.0000	0.1666	[0.0000, 0.0000]	[0.0000, 0.3332]
δ_2	0.2500	1.0000	0.1666	[1.0000, 1.0000]	[0.6668, 1.3332]
δ_3	0.2500	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
δ_4	0.2500	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{11}	0.7931	0.9030	0.1228	[0.7091, 0.9747]	[0.6574, 1.1485]
γ_{12}	0.2069	0.0970	0.1228	[0.0253, 0.2909]	[0.0000, 0.3426]
γ_{13}	0.0000	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{14}	0.0000	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{21}	0.1613	0.0556	0.0500	[0.0224, 0.1616]	[0.0000, 0.1555]
γ_{22}	0.7097	0.9166	0.0691	[0.7618, 0.9504]	[0.7783, 1.0548]
γ_{23}	0.1290	0.0279	0.0325	[0.0131, 0.0994]	[0.0000, 0.0928]
γ_{24}	0.0000	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{31}	0.0000	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{32}	0.0645	0.0193	0.0335	[0.0000, 0.1000]	[0.0000, 0.0862]
γ_{33}	0.7742	0.9203	0.0654	[0.7729, 0.9658]	[0.7895, 1.0511]
γ_{34}	0.1613	0.0604	0.0458	[0.0176, 0.1575]	[-0.031, 0.1520]
γ_{41}	0.0000	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{42}	0.0000	0.0000	0.0000	[0.0000, 0.0000]	[0.0000, 0.0000]
γ_{43}	0.1961	0.0512	0.0511	[0.000, 0.1521]	[0.0000, 0.1533]
γ_{44}	0.8039	0.9488	0.0511	[0.8479, 1.0000]	[0.8467, 1.0509]
α_1	78.351	107.26	37.913	[66.266, 185.83]	[31.451, 183.15]
α_2	241.68	163.30	46.149	[120.93, 272.16]	[70.996, 255.59]
α_3	259.78	217.59	60.531	[147.35, 341.10]	[96.527, 338.65]
α_4	126.44	89.512	25.501	[60.422, 141.49]	[38.509, 140.52]
β_1	799.69	1047.8	366.69	[648.06, 1884.6]	[314.61, 1781.4]
β_2	3204.1	2088.4	596.02	[1541.3, 3508.3]	[896.27, 3280.3]
β_3	4530.5	3703.7	1029.0	[2499.3, 1645.7]	[1645.7, 5761.8]
β_4	2825.2	1931.0	553.45	[1300.5, 3108.3]	[824.13, 3037.9]

Cuadro 3.17: Estimaciones e intervalos de confianza de los parámetros del modelo oculto de Markov seleccionado para la morosidad de la cartera de consumo.

una quiebra, mientras que valores bajos representarían una tendencia a *no crisis* y una menor exposición a una quiebra.

En la segunda figura, se muestra la probabilidad de pertenencia de cada observación a cada uno de los 4 estados en todo el período de estudio. En esta figura se observa que las probabilidades de pertenencia están bien definidas, pues son muy altas, es decir, es muy clara la pertenencia de las observaciones a cada uno de los estados.

Posteriormente, se realiza la predicción de los estados para el año 2017, para lo cual, se hace uso de la ecuación 2.67 donde la predicción del estado se lo hace de

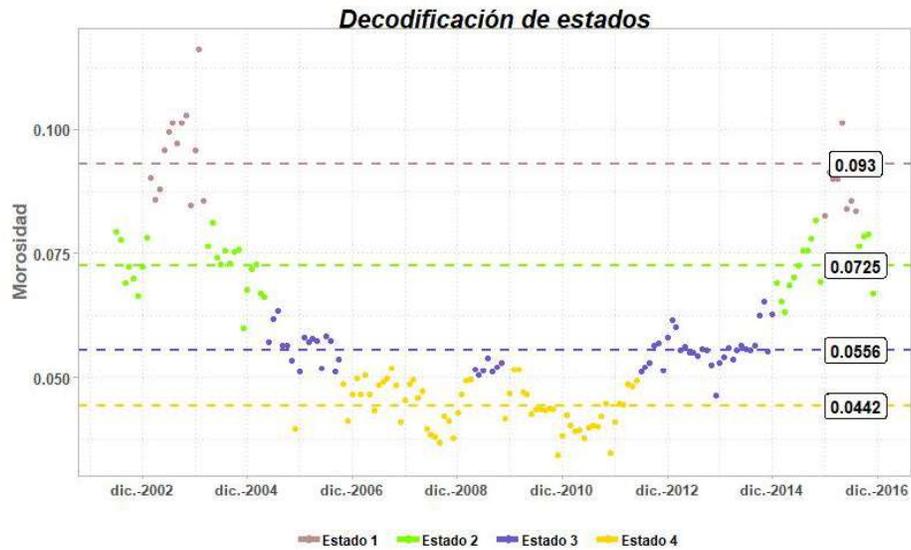


Figura 3.22: Decodificación de los estados mediante el algoritmo de Viterbi para la morosidad de la cartera de crédito de consumo.

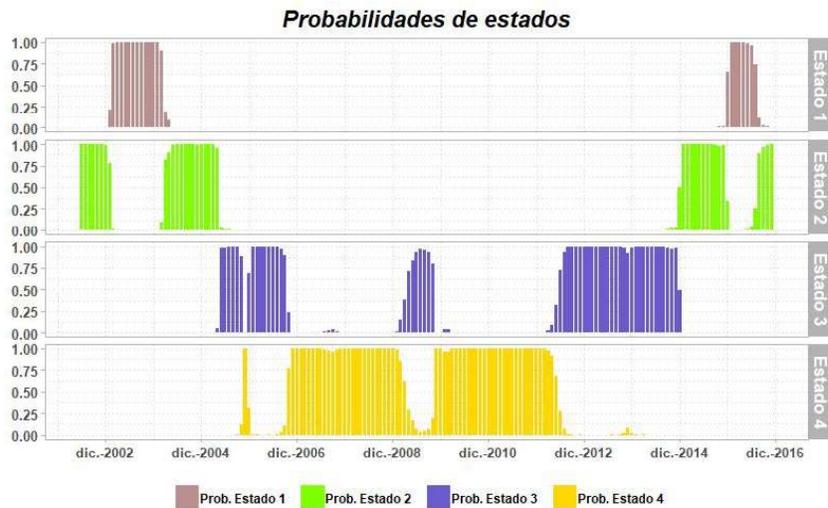


Figura 3.23: Probabilidades de pertenencia a los estados para la morosidad de la cartera de crédito de consumo.

acuerdo al mayor valor de probabilidad que se obtenga en cada uno de los estados para cada período. Estos resultados se los pueden observar en la figura 3.24 en la cual se aprecia que para enero-2017 existe una alta probabilidad de que se encuentre en el estado 2 y muy bajas probabilidades en el resto de estados; paulatinamente en los siguientes meses se siguen mostrando que la probabilidad de encontrarse en el estado 1 siguen siendo mayor que el resto de estados, pero va disminuyendo poco a poco mientras que la probabilidad de encontrarse en el resto de estados aumenta en

la misma proporción. Además, se observa que la probabilidad de encontrarse en el estado 1 va aumentando, lo que muestra que para un horizonte de tiempo más amplio habrá una gran probabilidad de encontrarse en el estado 1. En este caso, el estado 1 (con niveles altos de morosidad) representaría crisis y el estado 4 (con niveles bajos de morosidad) representaría no crisis. Lo que indicaría que para un horizonte de tiempo muy grande se tendrá posiblemente una quiebra del sistema financiero.



Figura 3.24: Predicción de los estados para el año 2017 para la morosidad de la cartera de crédito de consumo.

Las probabilidades de predicción del estado se presentan en el cuadro 3.18 en la cual se muestra numéricamente lo dicho anteriormente, es decir, que la probabilidad de encontrarse en el estado 1 es muy alta y que tiende a decrecer en el período de validación, mientras que la probabilidad de encontrarse en el estado 1 tiende a aumentar lentamente; es decir, tiende a una crisis.

Como siguiente paso se procede a calcular las distribuciones de pronóstico, para lo cual se hace uso de la ecuación 2.60 para cada uno de los 12 períodos de validación. Dichas distribuciones se encuentran en la figura 3.25, en la cual se muestran las distribuciones de pronóstico para el año 2017. En esta figura se muestra la densidad de pronóstico (en azul) con su respectiva distribución de probabilidad (en rojo).

Período	Probabilidad de pertenecer al estado 1	Probabilidad de pertenecer al estado 2	Probabilidad de pertenecer al estado 3	Probabilidad de pertenecer al estado 4
Enero 2017	0.0563	0.9152	0.0285	0.0000
Febrero 2017	0.1016	0.8448	0.0518	0.0018
Marzo 2017	0.1387	0.7852	0.0713	0.0048
Abril 2017	0.1689	0.7345	0.0877	0.0089
Mayo 2017	0.1933	0.6913	0.1017	0.0137
Junio 2017	0.2130	0.6543	0.1136	0.0192
Julo 2017	0.2287	0.6226	0.1237	0.0250
Agosto 2017	0.2411	0.5952	0.1325	0.0312
Septiembre 2017	0.2507	0.5715	0.1402	0.0376
Octubre 2017	0.2582	0.5508	0.1469	0.0442
Noviembre 2017	0.2637	0.5327	0.1528	0.0508
Diciembre 2017	0.2677	0.5168	0.1581	0.0574

Cuadro 3.18: Probabilidades de predicción del estado para el año 2017 para la cartera de crédito de consumo.

En esta figura se puede observar que en los 12 períodos de validación las densidades van cambiando en el tiempo, es decir, se muestra inicialmente que en enero-2017 existe pico que posteriormente va disminuyendo mientras su cola izquierda aumenta. Luego su cola derecha sigue el mismo patrón, es decir, también va incrementando su tamaño. Estas densidades en comparación a las de la figura 3.12 son cambiantes, lo cual es coherente, pues en la figura 3.15 se observa que en la zona de validación las observaciones reales son muy variables mostrando un ligero decrecimiento.

Basados en dichas distribuciones se muestran algunos estadísticos descriptivos para cada período de validación, de acuerdo con las ecuaciones 3.3, 3.4, 3.5, 3.6, 3.7 y 3.8. Estos resultados se visualizan en el cuadro 3.19.

En el cuadro 3.19 se puede apreciar que los estadísticos son variables en cada uno de los 12 períodos de validación, lo que no sucedía con los estadísticos del modelo anterior. Lo cual es coherente, pues las observaciones reales tienen un comportamiento similar. Esto también se lo puede visualizar en la figura 3.25, pues la forma de las distribuciones se van modificando en cada período.

Para observar el poder predictivo del modelo oculto de Markov, se presenta el ajuste al conjunto de entrenamiento, usando la ecuación 2.58 y los pronósticos obtenidos con dicho modelo al conjunto de validación (2017). Puesto que, en cada período se dispone de una distribución, para obtener un valor de ajuste y pronóstico, respectivamente, se pretende usar las medidas de tendencia central, tales como: media, mediana y

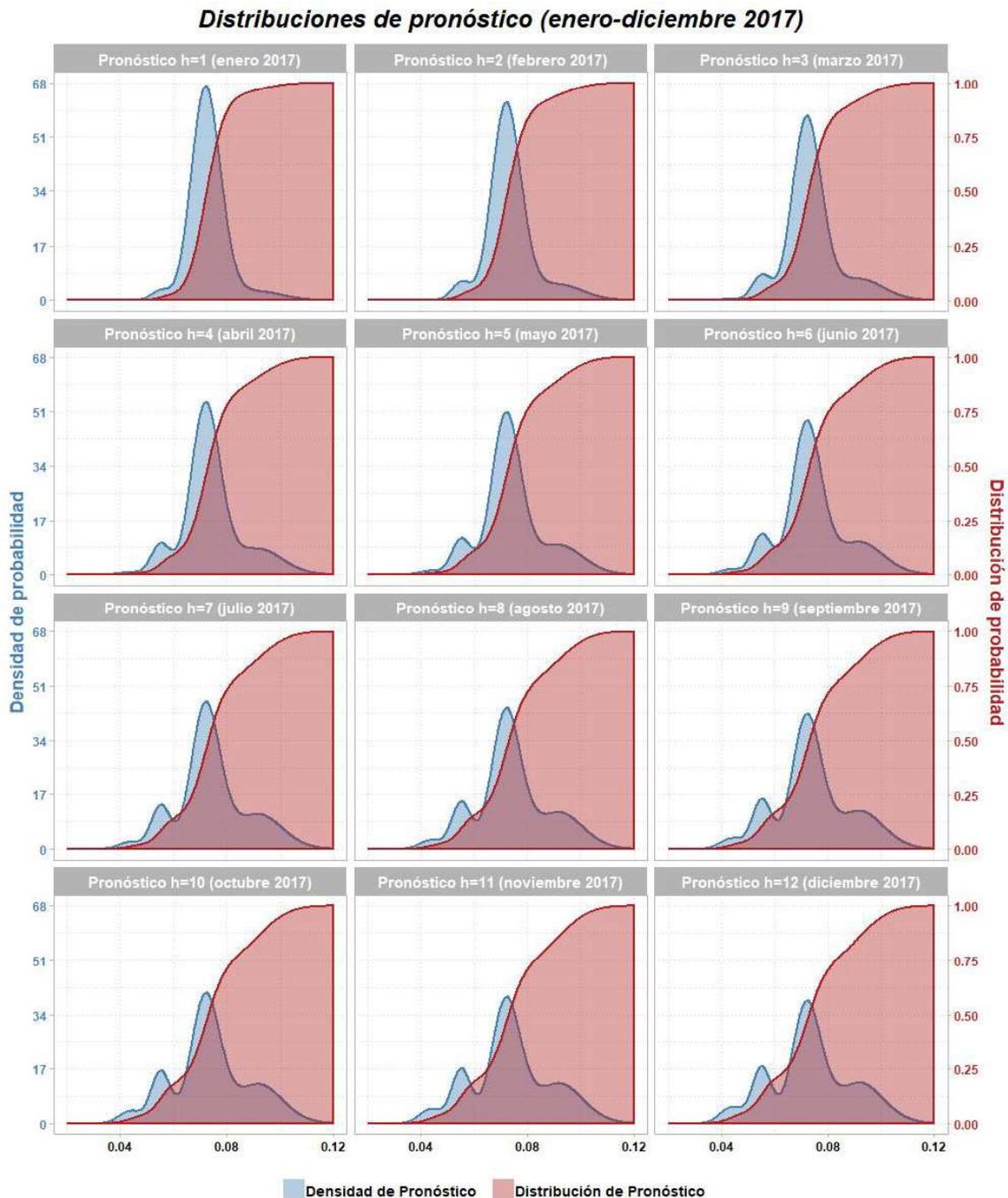


Figura 3.25: Densidades y distribuciones de pronóstico para la morosidad de la cartera de crédito de consumo.

moda; para cada período. Además, se realiza un ajuste con el modelo SARIMA para la cartera en estudio y se realizan pronósticos para la validación. Los resultados del modelo SARIMA se encuentra en el anexo A.

En la figura 3.26 se presenta una comparación gráfica de los 2 modelos con las

Estadísticos	Enero 2017	Febrero 2017	Marzo 2017	Abril 2017	Mayo 2017	Junio 2017	Julio 2017	Agosto 2017	Septiembre 2017	Octubre 2017	Noviembre 2017	Diciembre 2017
Horizonte	1	2	3	4	5	6	7	8	9	10	11	12
Media	0.07317	0.07364	0.07397	0.07418	0.07430	0.07434	0.07432	0.07424	0.07413	0.07398	0.07380	0.07361
Varianza	0.00006	0.00009	0.00011	0.00014	0.00015	0.00017	0.00019	0.00020	0.00021	0.00022	0.00023	0.00024
Desv.Est	0.00795	0.00954	0.01072	0.01166	0.01242	0.01307	0.01362	0.01410	0.01452	0.01489	0.01521	0.01551
Moda	0.07220	0.07220	0.07220	0.07220	0.07220	0.07220	0.07220	0.07230	0.07230	0.07230	0.07230	0.07230
Cuartil 1	0.06850	0.06840	0.06820	0.06800	0.06770	0.06740	0.06700	0.06670	0.06620	0.06570	0.06520	0.06450
Cuartil 2	0.07250	0.07270	0.07280	0.07300	0.07300	0.07310	0.07310	0.07320	0.07310	0.07310	0.07300	0.07290
Cuartil 3	0.07670	0.07740	0.07810	0.07890	0.07960	0.08030	0.08090	0.08150	0.08200	0.08240	0.08280	0.08300

Cuadro 3.19: Estadísticos de las distribuciones de pronósticos para el año 2017 para la morosidad de la cartera de consumo.

observaciones reales, donde se observa que el modelo SARIMA proporciona un mejor ajuste a las observaciones reales que el modelo oculto de Markov, aunque éste modela muy bien la tendencia de la serie usando como ajuste y pronóstico la media, mediana y moda. En ambos modelos, los pronósticos son muy buenos, pues visualmente son muy cercanos.

En un modelo SARIMA se fija un estado del espacio de estados y se varía respecto al espacio parametral, por lo que al pronosticar se lo realiza período a período. Mientras que en un modelo oculto de Markov varía respecto al espacio de estados y al espacio parametral, por lo que en cada período se obtiene una distribución de pronóstico y para obtener un valor se estima la media, mediana o moda como pronóstico. De esta manera, al comparar los resultados, se tiene que el del modelo oculto de Markov no son tan buenos, como el del modelo SARIMA, pues sus medidas de exactitud son mejores (*tabla 3.20*), pero si se encuentran en el dentro del estado predicho (*estado 1*). Esto hacen que sean mejores sobre los modelo SARIMA pues predicen estados y pronósticos dentro de ellos.

	Pronóstico ARIMA	Pronóstico HMM (Media)	Pronóstico HMM (Mediana)	Pronóstico HMM (Moda)
EAM	0.01201	0.01064	0.00964	0.00901
CME	0.00016	0.00015	0.00013	0.00012
EPAM	19.70 %	17.87 %	16.29 %	15.25 %

Cuadro 3.20: Medidas de exactitud para la morosidad de la cartera de crédito de consumo.

En la figura 3.26 se presentan los ajustes (*conjunto de entrenamiento*) y los pronósticos (*conjunto de validación*) del modelo SARIMA y del modelo oculto de Markov,

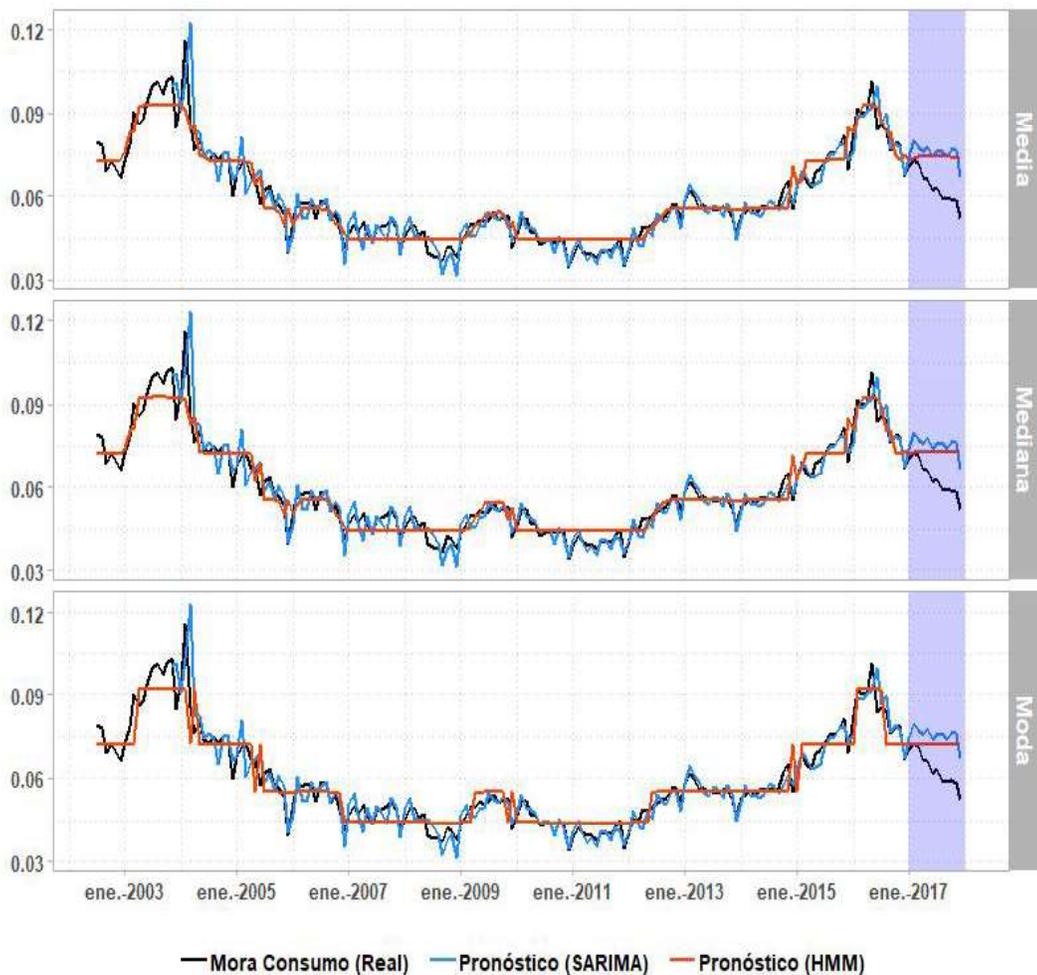


Figura 3.26: Ajustes y pronósticos con el modelo oculto de Markov y el modelo SARIMA con las observaciones reales para la morosidad de la cartera de crédito de consumo.

y se los compara con las observaciones reales. En esta figura se observan buenos resultados del modelo SARIMA sobre el modelo oculto de Markov al realizar pronósticos punto a punto. Por otro lado, los resultados del modelo oculto de Markov también son buenos pero para estimar estados en cada período. Esta figura puede ser de gran interés y utilidad para un ente regulador, pues permite conocer dónde se encontró y se encontrará la situación de la cartera comercial y sus fluctuaciones en el tiempo.

Por otro lado, no se debe olvidar que los modelos ocultos de Markov no realizan pronósticos mediante una expresión, como lo hacen los modelos SARIMA, sino que proporcionan probabilidades que se expresan en densidades (caso continuo) y distribuciones de probabilidad.

Además, en la figura 3.27 se muestra los pronósticos de los modelos aplicados con sus respectivos intervalos de confianza y las observaciones reales en donde se observa que las observaciones reales están contenidas en el intervalos de confianza del modelo oculto de Markov y éstas a su vez están contenidas en los intervalos de confianza del modelo SARIMA. Lo que cual indica que el modelo oculto de Markov reporta pronósticos más certeros.

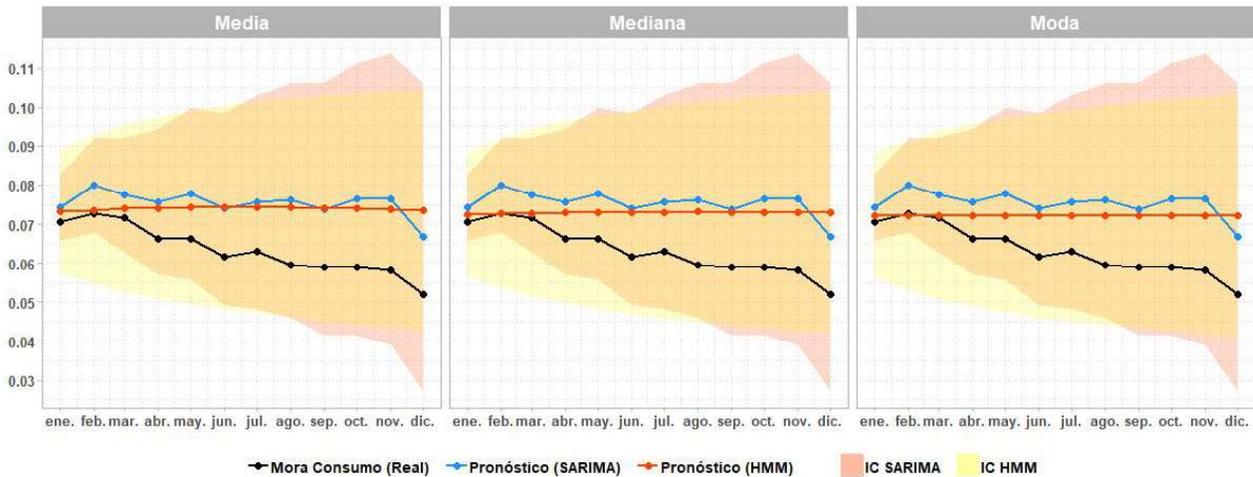


Figura 3.27: Comparación de pronósticos e intervalos de confianza de los modelos con las observaciones reales de la morosidad de la cartera de crédito de consumo.

Finalmente, los pronósticos del modelo oculto de Markov generan buenos resultados, pues las medidas de exactitud 3.9 y 3.10 reportan valores pequeños de aproximadamente 0,00013 para el caso del cuadrado medio debido al error y de 16,29% para el caso del error porcentual absoluto medio.

Capítulo 4

Conclusiones y Recomendaciones.

Las instituciones financieras y entes reguladores hacen especial énfasis al estudio del riesgo, pues una inadecuada gestión provocaría problemas económicos, que incluso, pueden desembocar en la quiebre de dichas instituciones, y en el peor de los casos, la caída de la economía de un país. Uno de los riesgos más importante para las instituciones financieras es el riesgo de crédito, representado por el índice de morosidad, ya que la actividad crediticia es la que genera mayores ingresos a la institución.

Por tal motivo, el principal objetivo de este trabajo de investigación es crear un modelo para la morosidad de la cartera de crédito, enfocándose en las carteras comercial y de consumo al ser las más representativas del sistema financiero nacional y estudiando al sistema de bancos privados, pues tienen un gran tamaño en la misma. Para lo cual se utiliza una extensión de las cadenas de Markov, como lo son los modelos ocultos de Markov como herramienta de modelación.

Dentro de los principales aportes de este estudio, se tiene: la descripción detallada de cada paso del modelo, su innovación al modelar el sistema financiero nacional, su aplicación en otras áreas y la obtención de resultados mediante el software libre R, lo que reduce el tiempo de obtención de resultados de manera eficiente y permite tomar decisiones oportunas y basados en argumentos matemáticos y no bajo supuestos.

Tomando en consideración lo anterior se puede concluir que:

- La metodología de los modelos ocultos de Markov para modelar series de tiempo es diferente a las tradicionales como los modelos ARIMA/SARIMA, ya que dichos modelos intentan modelar una de las posibles trayectorias, mientras que en el caso de los modelos ocultos de Markov se reconoce que ésta es una de todas las posibles y por tal motivo se trabaja con matrices de cambio de estado.

- Esta metodología tiene la ventaja de obtener una matriz de transición con la cual se puede obtener probabilidades de salto entre los diferentes niveles de riesgos establecidos. Además, dispone de distribuciones de probabilidad que generan las observaciones de la serie en cada período de tiempo con lo cual se puede inferir su comportamiento.
- Otra interpretación de estos estados o niveles de riesgo es la de considerarlos como estados de crisis, en donde los niveles altos de morosidad indicaría una crisis del sistema financiero. Por el contrario, niveles bajos de morosidad indicarían que el sistema está saludable y lejos de una crisis. De esta manera, al pronosticarse estados de crisis, el modelo revelaría que existe una gran exposición al quiebre del sistema financiero de una institución o de un país. Además, da lugar a tomar acciones para encontrar las fallas que están ocasionando dicha crisis, como: malas prácticas, malos negocios, fallas operacionales, corrupción, etc.
- Otra ventaja de esta metodología, es que la predicción de los estados y los pronósticos están basados en probabilidades, lo que da lugar a formar distribuciones de probabilidad, lo que brinda un valor agregado, pues permite inferir su comportamiento en los periodos futuros y permite varias alternativas para gestionar el riesgo.
- Los resultados obtenidos con esta metodología son muy buenos, a pesar de usar únicamente información histórica de la misma serie. Y esto se puede constatar en los ajustes obtenidos tanto en la distribución de la serie con los modelos de mezcla, como en su trayectoria con los modelos ocultos de Markov. Además, se cumplen los criterios de normalidad de los pseudo-residuos.
- En este estudio se usaron los modelos de mezcla para generar los niveles de riesgo que posteriormente formaron los estados de la cadena de Markov, limitándose a 2, 3, 4 y 5. Al aumentar el número de estados se generaría una mejor explicación del comportamiento de la serie, pero al hacerlo, se pierde grados de libertad sin mencionar el alto costo computacional que implica la estimación de los parámetros del modelo oculto de Markov.
- Las matrices de transición de probabilidades de cada cartera presentan una mayor concentración en la diagonal, lo que muestra que una transición de un estado a otro diferente es poco probable, es decir, que las series no presentan alta variabilidad.

- Esta metodología proporciona elementos importantes para la gestión del riesgo, pues permite estudiar su comportamiento pasado y realizar inferencias acerca de su futuro, con lo cual, de pronosticarse eventos desfavorables, se podría realizar las acciones pertinentes para mitigar el riesgo y evitar la caída del sistema financiero nacional.
- Por otro lado, este modelo permite al ente regulador saber en qué rangos se encontrará la morosidad, sus fluctuaciones y la frecuencia con la que lo hace. En lugar de pronósticos puntuales.
- Se elaboró una herramienta de fácil aplicación e interpretación, robusta y parsimoniosa, basada en procesos estocásticos que modela el riesgo de crédito, reporta buenos resultados y se automatiza cada paso en el software estadístico R.

Con el fin de facilitar trabajos futuros y mejorar los resultados obtenidos con este modelo es necesario realizar las siguientes recomendaciones:

- Tomar en consideración información externa, covariables, que puedan expresar la morosidad de la cartera, ya que este modelo es univariante, una mejora sería considerarlo multivariante. La inclusión de otras variables permitiría obtener resultados más certeros, como: PIB, IPC, Precio del petróleo, etc. Así como lo hacen Uquillas and González (2017) quienes usan un modelo SARIMAX usando variables microeconómicas y macroeconómicas que generan mejores pronósticos de que modelo SARIMA por si solo. Una posibilidad es considerar las extensiones de los modelos ocultos de Markov, como los detallados en la sección final del capítulo 2.
- Debido al alto costo computacional que requiere la estimación de los parámetros del modelo se recomienda el uso de herramientas más robustas o complementarias al software R para agilizar la obtención de los resultados, como el uso de Python.
- En este trabajo se consideró los modelos de mezcla para generar los estados de la cadena de Markov, es posible obtener mejoras del modelo si se considera otros modelos de agrupamiento. Así, como otros criterios para la selección del número óptimo de grupos, como quiebre de grupos, para determinar el número de grupos de manera estadística.

- Extender este modelo a otras carteras de crédito, así, como para otros tipos de riesgos, no solo de crédito, y aplicarlos en instituciones financieras y entidades de supervisión bancaria.
- Usarlo como herramienta complementaria a las metodologías ya existentes y que hacen uso las entidades de supervisión bancaria para gestionar el riesgo. Y que sirva como un sistema de alerta temprana que permita implementar políticas económicas para mantener o mejorar la calidad de la cartera.

Bibliografía

- Aggarwal, C. C. and Reddy, C. K. (2013). *Data Clustering: Algorithms and Applications*. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series.
- Ahuja, S., Bunsupha, S., Tan, K., Tong, X., and Wen, V. (2011). Hidden markov model of portfolio credit risk. *Stanford University*.
- Anderson, D., Sweeney, D., and Williams, T. (2011). *Estadística para negocios y economía*. Estadística. CENGAGE Learning, 11va edition.
- Basilea, C. (2006). Convergencia internacional de medidas y normas de capital. *Working Paper*.
- BCE (2011). Resolución no. jb-2011-1897. http://oidprd.sbs.gob.ec/medios/PORTALDOCS/downloads/normativa/2011/resol_JB-2011-1897.pdf.
- Benyacoub, B., ElBernoussi, S., Zoglat, A., and Ismail, E. M. (2014). Classification with hidden markov model. *Applied Mathematical Sciences*, 8(50):2483–2496.
- Bhar, R. and Hamori, S. (2006). *Hidden Markov models: applications to financial economics*, volume 40. Springer Science & Business Media.
- Bonilla Toapanta, J. G. and Gálvez Castillo, J. A. (2007). Análisis de la crisis bancaria de 1999 y sus efectos en la industria ecuatoriana en el período 1999-2004. B.S. thesis, Escuela Politécnica Nacional.
- Capa Santos, H. (2016). *Series temporales: la ciencia y el arte de la modelación y los pronósticos*. Monografías de estadística y matemática. Escuela Politécnica Nacional, 1ra edition.
- Chediak, G. J. and Maldonado, J. L. (2014). Reconocimiento automático de fonemas en habla continua venezolana por medio de sistemas híbridos basados en modelos ocultos de markov y redes neuronales artificiales. *Ciencia e Ingeniería*, 35(1):29–38.

- Ching, W.-K., Siu, T. K., Li, L.-m., Li, T. L., and Li, W.-K. (2009). Modeling default data via an interactive hidden markov model. *Springer Series*.
- Cooper, B. and Lipsitch, M. (2004). The analysis of hospital infection data using hidden markov models. *Biostatistics*, 5(2):223–237.
- Cribari-Neto, F. and Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, 34(2):1–24.
- Cuevas, D. C. (2013). Cuando un banco quiebra. <https://www.listindiario.com/puntos-de-vista/2013/10/03/294427/cuando-un-banco-quiebra>.
- Dymarski, P. (2011). *Hidden Markov Models, Theory and Applications*. InTech, Chapters.
- Evin, D., Hadad, A., Martina, M., and Drozdowicz, B. (2011). Predicción de estados de hipotensión empleando modelos ocultos de markov. *Facultad de Ingeniería*, 20:55–63.
- Giampieri, G., Davis, M., and Crowder, M. (2005). Analysis of default data using hidden markov models. *Quantitative Finance*, 5(1):27–34.
- Gómez Losada, Á. (2014). Modelos de mixturas finitas para la caracterización y mejora de las redes de monitorización de la calidad del aire. Master's thesis, Universidad de Granada.
- Guzmán Garzón, R. O. (2017). Modelo estocástico de cadenas de markov ocultas para el problema de quiebra de las empresas ecuatorianas, en un sector específico de la economía ecuatoriana. B.S. thesis, Escuela Politécnica Nacional.
- Harte, D. (2017). *HiddenMarkov: Hidden Markov Models*. Statistics Research Associates, Wellington. R package version 1.8-11.
- Jabbour, G. and Maldonado, J. L. (2009). Predicción de índices bursátiles mediante un sistema híbrido basado en modelos ocultos de markov y redes neuronales artificiales. *Ciencia e Ingeniería*, 30(2):127–136.
- Jurafsky, D. and Martin, J. H. (2017). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, chapter 9, pages 122–141. Prentice Hall, 3era edition.

- Kassambara, A. (2017). *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 0.1.6.
- Kipkoech, K. G. (2016). A multivariate markov chain model for credit risk measurement and management. Master's thesis, University Of Nairobi.
- Korolkiewicz, M. W. (2012). A dependent hidden markov model of credit quality. *International Journal of Stochastic Analysis*.
- Küchler, U., Sorensen, U., Sorensen, M., and ensen, M. (1997). *Exponential families of stochastic processes*. Springer Science & Business Media, 1era edition.
- Leisch, F. (2004). FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software*, 11(8):1–18.
- Lihn, S. H. (2017). Hidden markov model for financial time series and its application to s&p 500 index. *Quantitative Finance*.
- Macas, M., Xismena, D., and Padilla Pineda, W. A. (2012). Estudio de los modelos ocultos de markov y desarrollo de un prototipo para el reconocimiento automático del habla. B.S. thesis, Universidad Politécnica Salesiana.
- Maldonado, L. (2012). Los modelos ocultos de markov, mom. *Telos*, 14:433–438.
- Malik, M. and Thomas, L. C. (2012). Transition matrix models of consumer credit ratings. *International Journal of Forecasting*, 28(1):261–272.
- Martinez, F., Portale, G., Klein, H., and Olmos, O. (2003). Reconocimiento de voz, apuntes de cátedra para introducción a la inteligencia artificial. http://www.secyt.frba.utn.edu.ar/giar/IA1_IntroReconocimientoVoz.pdf.
- Mateus, E. O. (2008). Hidden markov models (hmm's) y aplicaciones. B.S. thesis, Universidad Tecnológica de Bolívar.
- Mhamane, S. S. and Lobo, L. J. (2012). Use of hidden markov model as internet banking fraud detection. *International Journal of Computer Applications*.
- Montoya Sánchez, L. A., Lapo, A., and Vladimir, E. (2014). Modelo para el análisis de riesgo crediticio de la cartera de vivienda basado en matrices de transición de calificación para el sector de bancos privados nacionales. B.S. thesis, Universidad de las Fuerzas Armadas ESPE. Maestría en Evaluación y Auditoría de Sistemas Tecnológicos.

- Nguyen, N. and Nguyen, D. (2015). Hidden markov model for stock selection. *Risks*, 3(4):455–473.
- Ntwiga, D., Bundi, W., and Weke, P. (2016). Credit scoring for m-shwari using hidden markov model. *European Scientific Journal*, 12:1857–7881.
- Ntwiga, D. B. (2016). *Social network analysis for credit risk modeling*. PhD thesis, University of Nairobi.
- Oguz, H. T. and Gurgun, F. S. (2008). Credit risk analysis using hidden markov model. In *2008 23rd International Symposium on Computer and Information Sciences*, pages 1–5.
- Patiño Antonioli, M. Á. (2011). Aplicación de las cadenas ocultas de markov para la preferencia de los consumidores en el mercado cervecero. Master's thesis, Pontificia Universidad Católica del Perú.
- Peñalver Benavent, A. (2007). *Clustering EBEM: modelos de mezclas gaussianas basados en maximización de entropía*. PhD thesis, Universidad de Alicante.
- Pérez Delgado, J. (2014). Matrices de transición y su papel en la crisis económica. Master's thesis, Universidad Pública de Navarra.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rabiner, L. and Juang, B. (1986). An introduction to hidden markov models. *IEEE ASSP Magazine*, 3(1):4–16.
- Ramos, M. (2011). Cadenas ocultas de markov aplicadas al reconocimiento de voz. *Universidad de la República - Uruguay*.
- Rincón, L. (2012). *Introducción a los procesos estocásticos*. Facultad de Ciencias UNAM.
- Rojo, H. and Miranda, M. (2009). Cadenas de markov. http://campus.fi.uba.ar/pluginfile.php/146938/mod_resource/content/3/CM_VERSION_BETA.pdf, Universidad de Buenos Aires.
- Ross, S. M. (2010). *Introduction to probability models*. Academic Press, 10ma edition.

- Salas González, D. (2008). *Modelos de mezcla de distribuciones alpha-estables*. PhD thesis, Universidad de Granada.
- SBS (2016a). Análisis de cosechas del sistema de bancos. http://www.superbancos.gob.ec/medios/PORTALDOCS/downloads/articulos_financieros/Estudios_Especializados/EC1_dic_16.pdf.
- SBS (2016b). Comportamiento crediticio sectorial. http://www.superbancos.gob.ec/medios/PORTALDOCS/downloads/articulos_financieros/Estudios%20Sectoriales/2016/ES1_2016.pdf.
- SBS (2016c). Memorias 2016. http://www.superbancos.gob.ec/medios/PORTALDOCS/downloads/Memoria_2016.pdf.
- SBS (2017). Matrices de transición del sistema de bancos. http://www.superbancos.gob.ec/medios/PORTALDOCS/downloads/articulos_financieros/Estudios_Especializados/MT1_jun_17.pdf.
- Teng, G.-E., He, C.-Z., Xiao, J., and Jiang, X.-Y. (2013). Customer credit scoring based on hmm/gmdh hybrid model. *Knowledge and information systems*, 36(3):731–747.
- Uquillas, A. and González, C. (2017). Determinantes macro y microeconómicos para pruebas de tensión de riesgo de crédito: un estudio comparativo entre Ecuador y Colombia basado en la tasa de morosidad. *Ensayos sobre Política Económica*, 35(84):245 – 259.
- Vázquez, V. P. R. and Vaquero, J. H. (2013). Matriz de probabilidad de transición de microcréditos: el caso de una microfinanciera mexicana. *Estudios Económicos*, pages 39–77.
- Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12):1–20.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H. (2017a). *scales: Scale Functions for Visualization*. R package version 0.5.0.
- Wickham, H. (2017b). *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1.

Wickham, H. and Bryan, J. (2018). *readxl: Read Excel Files*. R package version 1.1.0.

Wickham, H., François, R., Henry, L., and Müller, K. (2018). *dplyr: A Grammar of Data Manipulation*. R package version 0.7.5.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov models for time series: an introduction using R*, volume 150 of *Chapman & Hall/CRC Monographs on Statistics & Applied Probability*. CRC press, segunda edition.

Anexo A: Modelos ARIMA - SARIMA

Modelo ARIMA para la serie de morosidad de la cartera comercial

Null Hypothesis: D(MORA_COMERCIAL) has a unit root
Exogenous: Constant
Lag Length: 4 (Fixed)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-9.273553	0.0000
Test critical values:		
1% level	-3.469451	
5% level	-2.878618	
10% level	-2.575954	

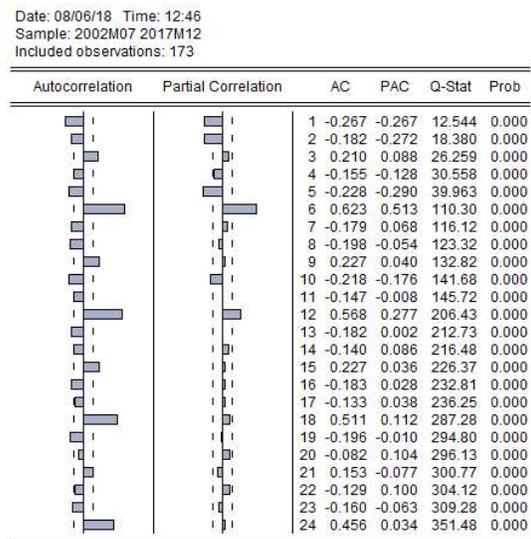
*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(MORA_COMERCIAL_2)
Method: Least Squares
Date: 08/06/18 Time: 12:45
Sample (adjusted): 2003M01 2016M12
Included observations: 168 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(MORA_COMERCIAL(-1))	-2.003365	0.216030	-9.273553	0.0000
D(MORA_COMERCIAL(-1),2)	0.704878	0.187393	3.761492	0.0002
D(MORA_COMERCIAL(-2),2)	0.482186	0.156267	3.085653	0.0024
D(MORA_COMERCIAL(-3),2)	0.500441	0.112443	4.450619	0.0000
D(MORA_COMERCIAL(-4),2)	0.284486	0.068607	4.146607	0.0001
C	-0.001057	0.000289	-3.660475	0.0003

R-squared	0.722511	Mean dependent var	9.26E-05
Adjusted R-squared	0.713947	S.D. dependent var	0.006359
S.E. of regression	0.003401	Akaike info criterion	-8.494285
Sum squared resid	0.001874	Schwarz criterion	-8.382715
Log likelihood	719.5200	Hannan-Quinn criter.	-8.449005
F-statistic	84.36156	Durbin-Watson stat	1.532173
Prob(F-statistic)	0.000000		

(a) Prueba de raíz unitaria.



(b) Correlograma de la morosidad diferenciada estacionariamente.

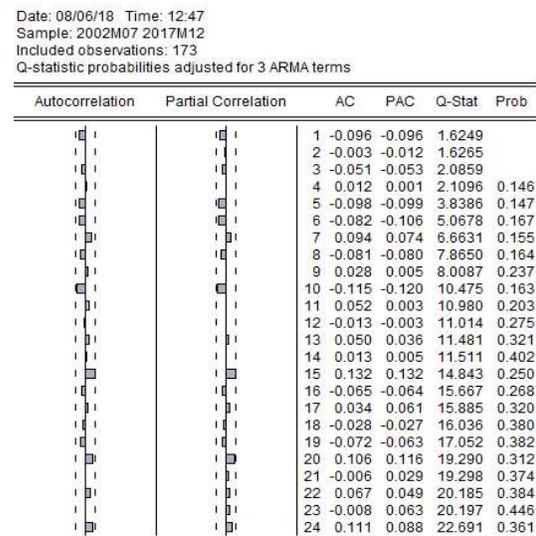
Dependent Variable: D(MORA_COMERCIAL)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 08/06/18 Time: 12:47
Sample: 2002M08 2016M12
Included observations: 173
Convergence achieved after 146 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(6)	0.962948	0.012316	78.18442	0.0000
MA(6)	-0.559971	0.063601	-8.804473	0.0000
MA(23)	-0.182492	0.071200	-2.563083	0.0112
SIGMASQ	6.96E-06	5.37E-07	12.95467	0.0000

R-squared	0.593346	Mean dependent var	-0.000540
Adjusted R-squared	0.586128	S.D. dependent var	0.004148
S.E. of regression	0.002669	Akaike info criterion	-8.933938
Sum squared resid	0.001203	Schwarz criterion	-8.861029
Log likelihood	776.7856	Hannan-Quinn criter.	-8.904359
Durbin-Watson stat	2.187328		

Inverted AR Roots	.99	.50+ .86i	.50- .86i	-.50+ .86i
		-.99		
Inverted MA Roots	.97	.89- .22i	.89+ .22i	.77- .47i
	.77+ .47i	.61+ .70i	.61- .70i	.46+ .85i
	.46- .85i	.21+ .89i	.21- .89i	-.07+ .90i
	-.07- .90i	-.35+ .87i	-.35- .87i	-.53- .80i
	-.53+ .80i	-.69- .59i	-.69+ .59i	-.84- .35i
	-.84+ .35i	-.95+ .09i	-.95- .09i	

(c) Coeficientes del modelo.



(d) Correlograma de los residuos del modelo.

Figura A.1: Resultados del modelo ARIMA

Modelo SARIMA para la serie de morosidad de la cartera de consumo

Null Hypothesis: D(MORA_CONSUMO,1,12) has a unit root
 Exogenous: None
 Lag Length: 0 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-14.69989	0.0000
Test critical values:		
1% level	-2.579495	
5% level	-1.942830	
10% level	-1.615384	

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(MORA_CONSUMO,2,12)
 Method: Least Squares
 Date: 08/06/18 Time: 12:52
 Sample (adjusted): 2003M09 2016M12
 Included observations: 160 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(MORA_CONSUMO(-1),1,12)	-1.151245	0.078317	-14.69989	0.0000

R-squared	0.576096	Mean dependent var	-1.78E-05
Adjusted R-squared	0.576096	S.D. dependent var	0.009041
S.E. of regression	0.005887	Akaike info criterion	-7.426010
Sum squared resid	0.005510	Schwarz criterion	-7.406790
Log likelihood	595.0808	Hannan-Quinn criter.	-7.418206
Durbin-Watson stat	2.046325		

(a) Prueba de raíz unitaria.

Date: 08/06/18 Time: 12:53
 Sample: 2002M07 2017M12
 Included observations: 161

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.152	-0.152	3.7882	0.052
		2	-0.123	-0.149	6.2690	0.044
		3	0.301	0.269	21.274	0.000
		4	0.003	0.081	21.275	0.000
		5	-0.092	-0.019	22.693	0.000
		6	0.007	-0.095	22.700	0.001
		7	0.003	-0.051	22.701	0.002
		8	-0.003	0.019	22.703	0.004
		9	-0.104	-0.082	24.573	0.003
		10	0.024	0.008	24.669	0.006
		11	0.095	0.085	26.245	0.006
		12	-0.375	-0.344	51.060	0.000
		13	0.036	-0.066	51.286	0.000
		14	0.097	-0.019	52.976	0.000
		15	-0.022	0.235	53.063	0.000
		16	-0.066	0.009	53.847	0.000
		17	0.049	-0.025	54.291	0.000
		18	0.063	-0.050	55.010	0.000
		19	-0.053	-0.063	55.537	0.000
		20	0.013	0.045	55.566	0.000
		21	0.015	-0.028	55.806	0.000
		22	-0.118	-0.124	58.252	0.000
		23	0.003	-0.001	58.253	0.000
		24	-0.021	-0.219	58.337	0.000

(b) Correlograma de la morosidad diferenciada estacionaria y estacionalmente.

Dependent Variable: D(MORA_CONSUMO,1,12)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 08/06/18 Time: 12:56
 Sample: 2003M08 2016M12
 Included observations: 161
 Convergence achieved after 35 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(3)	0.336828	0.072160	4.667769	0.0000
MA(12)	-0.899410	0.142969	-6.290953	0.0000
SIGMASQ	1.75E-05	2.71E-06	6.471572	0.0000

R-squared	0.499723	Mean dependent var	-0.000139
Adjusted R-squared	0.493390	S.D. dependent var	0.005941
S.E. of regression	0.004228	Akaike info criterion	-7.953926
Sum squared resid	0.002825	Schwarz criterion	-7.896509
Log likelihood	643.2911	Hannan-Quinn criter.	-7.930612
Durbin-Watson stat	2.251982		

Inverted AR Roots	.70	-.35+.60i	-.35-.60i	
Inverted MA Roots	.99	.86+.50i	.86-.50i	.50+.86i
		.50-.86i	-.00-.99i	-.50-.86i
		-.50+.86i	-.86+.50i	-.86-.50i
				-.99

(c) Coeficientes del modelo.

Date: 08/06/18 Time: 12:56
 Sample: 2002M07 2017M12
 Included observations: 161
 Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.127	-0.127	2.6549	
		2	-0.102	-0.120	4.3784	
		3	-0.020	-0.051	4.4445	0.055
		4	0.046	0.025	4.8024	0.091
		5	-0.071	-0.071	5.6588	0.129
		6	0.032	0.020	5.8360	0.212
		7	-0.013	-0.020	5.8638	0.320
		8	0.045	0.042	6.2088	0.400
		9	-0.011	0.004	6.2312	0.513
		10	-0.132	-0.136	9.2519	0.321
		11	0.053	0.024	9.7486	0.371
		12	0.034	0.009	9.9582	0.444
		13	0.000	0.014	9.9582	0.534
		14	-0.023	-0.009	10.054	0.611
		15	0.042	0.023	10.377	0.663
		16	-0.083	-0.070	11.617	0.637
		17	0.031	0.014	11.792	0.695
		18	0.100	0.110	13.622	0.627
		19	0.025	0.047	13.738	0.686
		20	0.047	0.080	14.153	0.719
		21	-0.053	-0.028	14.685	0.742
		22	-0.173	-0.175	20.326	0.438
		23	0.058	0.012	20.964	0.461
		24	0.037	0.005	21.220	0.507

(d) Correlograma de los residuos del modelo.

Figura A.2: Resultados del modelo SARIMA

Anexo B: Bases de datos

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
2002							1756874.14	1766842.13	1853225.14	1879946.83	1861821.34	1879375.37
2003	1882933.76	1871135.57	1996692.15	2003279.11	2019307.23	2013771.28	2033655.47	2038517.78	2095660.82	2105232.70	2118126.27	2118002.76
2004	2092029.02	2085264.77	2108082.32	2140156.70	2184434.28	2232797.12	2261013.89	2318490.29	2362393.90	2441388.69	2428639.10	2472518.16
2005	2497299.50	2512232.15	2562905.82	2617008.81	2668013.08	2748886.10	2766949.30	2794007.59	2848854.03	2871337.06	2898194.93	2959364.55
2006	2911284.23	2947440.74	2976129.53	3016211.06	3068139.63	3125243.43	3194232.37	3256947.56	3347483.47	3364994.53	3445280.45	3566359.21
2007	3572754.67	3546634.47	3598274.04	3584273.48	3574605.62	3591325.13	3588889.30	3608809.70	3594480.85	3635236.94	3649428.34	3681338.13
2008	3720661.48	3755831.79	3879014.85	3985345.76	4101585.56	4217687.31	4281278.88	4434458.41	4541694.49	4667215.59	4722946.02	4676773.77
2009	4559780.61	4528550.38	4340344.62	4230927.92	4155466.08	4166473.87	4169788.27	4115368.27	4149532.48	4191522.57	4316010.09	4305777.46
2010	4343894.34	4316360.65	4372491.88	4561836.30	4606880.13	4704868.97	4751151.48	4806238.44	4881363.50	4979855.33	5023098.35	5175209.62
2011	5140821.76	5118270.98	5248667.67	5394609.25	5462142.71	5501891.40	5555675.86	5606737.54	5667223.81	5804055.64	6031492.05	6116610.87
2012	6017413.16	6001190.83	6233586.59	6502802.61	6590915.12	6719682.85	6800324.93	6987475.37	7126963.63	7130326.78	7272176.19	7399107.23
2013	7346606.77	7364957.63	7506735.75	7717308.19	7691746.11	7807888.64	7900723.41	8030846.07	8071001.21	8076649.40	8288700.63	8403771.17
2014	8268875.73	8445117.21	8642092.97	8970001.89	8965608.06	8967145.68	9006186.93	8997628.42	9114840.93	9218579.60	9274913.89	9676594.87
2015	9847838.37	10234415.79	10235167.66	10366822.50	10167805.57	10157739.22	10002609.28	9415457.07	9211553.19	9107537.18	8936896.05	8734009.01
2016	8634320.79	8676364.70	8782843.67	8952265.59	8906606.92	8994513.82	9070274.61	9018714.06	9077535.46	9255187.46	9455376.27	10099559.34
2017	10043455.73	10183605.56	10398123.19	10736312.78	10732099.19	10874379.47	10978982.31	11042690.54	11208492.84	11292544.88	11560088.41	11767772.45

Cuadro B.1: Participación de la cartera de crédito comercial

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
2002							837888.57	853436.76	865803.58	863406.46	882362.08	864846.34
2003	852336.54	852650.96	744233.18	733049.62	757013.87	749530.43	759511.41	773824.76	788715.18	799650.54	819286.70	830305.99
2004	849095.94	850622.46	856393.27	867190.53	888803.99	906510.01	943142.36	973627.55	1006547.77	1027572.98	1053201.22	1131276.69
2005	1141061.63	1160177.73	1193355.45	1231091.34	1280672.56	1296659.30	1341940.48	1356604.22	1389789.91	1361802.62	1426414.84	1441653.29
2006	1553785.20	1577908.26	1618075.90	1625889.30	1675981.82	1715258.47	1750352.16	1796119.45	1853608.09	1893965.93	1920091.52	1908454.24
2007	2037185.23	2060481.25	2041739.34	2133333.37	2105718.69	2184959.08	2207328.43	2184023.95	2195966.15	2188209.26	2209580.91	2259835.20
2008	2260766.77	2271853.21	2264820.04	2253792.03	2291270.98	2325900.96	2394406.07	2478410.01	2557744.29	2612599.66	2712706.42	2818043.97
2009	2826618.61	2819586.41	2776473.22	2729555.67	2778633.65	2729855.86	2730713.79	2740684.39	2767947.54	2779554.03	2810610.40	2872243.12
2010	2893091.69	2906028.97	2940677.65	3003682.75	3085824.27	3174274.44	3246169.33	3287870.33	3402654.50	3461843.27	3597244.41	3721432.19
2011	3817557.92	3969201.26	4092834.85	4208712.35	4278142.65	4414264.64	4552432.18	4675698.22	4781268.19	4877392.23	4964084.47	5015874.61
2012	5067894.33	5104455.84	5195336.57	5282134.45	5381018.52	5431719.15	5492209.54	5570121.26	5622211.35	5661291.70	5689761.86	5689244.01
2013	5676144.00	5634945.01	5588939.15	5661220.85	5731867.88	5811924.61	5913190.00	5967955.07	5994767.96	6030709.39	6075103.82	6111643.65
2014	6144496.37	6183238.54	6192539.63	6208141.59	6259626.10	6355843.62	6417175.82	6491306.72	6559556.73	6825787.84	6874812.20	6924281.15
2015	6926436.61	6911010.60	6926903.40	6844401.57	6796328.38	6738757.79	6687705.00	6664692.75	6643086.51	6575575.55	6471401.16	6331052.99
2016	6225450.76	6113876.17	6040029.68	5967549.13	5975830.37	5990626.49	6020790.73	6175240.43	6223990.18	6273137.98	6357543.89	6381920.04
2017	6366376.22	6358986.74	6411568.72	6435974.75	6533758.77	7869068.35	7953297.11	8115900.58	8270541.34	8352349.53	8557690.54	8604908.21

Cuadro B.2: Participación de la cartera de crédito de consumo

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
2002							0.1049	0.1024	0.1058	0.1089	0.1149	0.0948
2003	0.1034	0.1083	0.1070	0.1027	0.1103	0.0986	0.1018	0.1090	0.1010	0.0982	0.0961	0.0824
2004	0.0884	0.0939	0.0893	0.0883	0.0906	0.0770	0.0856	0.0967	0.0958	0.0935	0.0981	0.0745
2005	0.0801	0.0805	0.0811	0.0779	0.0746	0.0677	0.0673	0.0700	0.0673	0.0681	0.0644	0.0553
2006	0.0581	0.0610	0.0566	0.0569	0.0562	0.0465	0.0449	0.0418	0.0374	0.0385	0.0372	0.0284
2007	0.0280	0.0323	0.0309	0.0330	0.0326	0.0272	0.0315	0.0321	0.0293	0.0285	0.0281	0.0228
2008	0.0243	0.0266	0.0240	0.0228	0.0225	0.0218	0.0245	0.0251	0.0222	0.0210	0.0197	0.0176
2009	0.0195	0.0241	0.0256	0.0271	0.0271	0.0249	0.0262	0.0255	0.0240	0.0241	0.0255	0.0231
2010	0.0246	0.0240	0.0226	0.0225	0.0241	0.0219	0.0227	0.0221	0.0202	0.0197	0.0194	0.0153
2011	0.0161	0.0172	0.0156	0.0156	0.0151	0.0150	0.0150	0.0155	0.0153	0.0153	0.0147	0.0112
2012	0.0119	0.0129	0.0124	0.0129	0.0126	0.0122	0.0124	0.0129	0.0116	0.0125	0.0126	0.0084
2013	0.0095	0.0100	0.0092	0.0112	0.0110	0.0107	0.0109	0.0105	0.0110	0.0106	0.0098	0.0077
2014	0.0098	0.0103	0.0114	0.0104	0.0116	0.0104	0.0107	0.0108	0.0104	0.0103	0.0098	0.0075
2015	0.0105	0.0127	0.0125	0.0138	0.0142	0.0127	0.0135	0.0121	0.0123	0.0134	0.0147	0.0113
2016	0.0154	0.0165	0.0145	0.0146	0.0160	0.0130	0.0128	0.0134	0.0125	0.0146	0.0161	0.0116
2017	0.0124	0.0132	0.0137	0.0133	0.0135	0.0117	0.0127	0.0122	0.0118	0.0118	0.0110	0.0101

Cuadro B.3: Serie de morosidad de la cartera de crédito comercial

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
2002							0.0793	0.0779	0.0690	0.0721	0.0699	0.0665
2003	0.0723	0.0783	0.0902	0.0859	0.0881	0.0957	0.0996	0.1015	0.0971	0.1013	0.1031	0.0847
2004	0.0957	0.1162	0.0857	0.0765	0.0812	0.0742	0.0728	0.0754	0.0729	0.0753	0.0757	0.0600
2005	0.0676	0.0718	0.0726	0.0670	0.0662	0.0573	0.0618	0.0635	0.0566	0.0566	0.0536	0.0397
2006	0.0513	0.0583	0.0572	0.0580	0.0574	0.0520	0.0585	0.0575	0.0513	0.0539	0.0487	0.0413
2007	0.0467	0.0499	0.0466	0.0505	0.0466	0.0435	0.0484	0.0491	0.0499	0.0520	0.0486	0.0411
2008	0.0454	0.0486	0.0497	0.0460	0.0472	0.0397	0.0385	0.0381	0.0369	0.0422	0.0413	0.0379
2009	0.0428	0.0466	0.0494	0.0496	0.0516	0.0506	0.0515	0.0540	0.0513	0.0523	0.0531	0.0417
2010	0.0469	0.0516	0.0518	0.0471	0.0466	0.0428	0.0435	0.0435	0.0434	0.0438	0.0437	0.0344
2011	0.0382	0.0424	0.0405	0.0393	0.0393	0.0377	0.0400	0.0404	0.0402	0.0423	0.0447	0.0349
2012	0.0411	0.0447	0.0446	0.0488	0.0482	0.0493	0.0511	0.0521	0.0531	0.0564	0.0570	0.0515
2013	0.0583	0.0617	0.0602	0.0557	0.0563	0.0552	0.0553	0.0546	0.0558	0.0556	0.0525	0.0465
2014	0.0530	0.0542	0.0561	0.0539	0.0556	0.0565	0.0559	0.0556	0.0566	0.0625	0.0653	0.0553
2015	0.0628	0.0691	0.0652	0.0633	0.0685	0.0702	0.0725	0.0755	0.0754	0.0780	0.0816	0.0693
2016	0.0828	0.0913	0.0900	0.0901	0.1013	0.0841	0.0857	0.0837	0.0764	0.0785	0.0790	0.0670
2017	0.0705	0.0728	0.0716	0.0663	0.0662	0.0616	0.0631	0.0596	0.0589	0.0590	0.0583	0.0521

Cuadro B.4: Serie de morosidad de la cartera de crédito de consumo

Anexo C: Funciones de R y esquema del procedimiento

Para este proyecto de investigación se utilizó el software estadístico R (R Core Team, 2018). Se utilizaron los siguientes paquetes:

```
library{readxl}      # Para leer archivos excel. (Wickham and Bryan, 2018)
library{dplyr}       # Para manejo y análisis de datos. (Wickham et al., 2018)
library{tidyverse}   # Para manejo de datos. (Wickham, 2017b)
library{reshape2}    # Para reestructurar data.frame. (Wickham, 2007)
library{flexmix}     # Para modelos de mezcla. (Leisch, 2004)
library{betareg}     # Para modelos de mezcla. (Cribari-Neto and Zeileis, 2010)
library{HiddenMarkov} # Para modelos ocultos de Markov (Harte, 2017)
library{ggplot2}     # Para gráficos especializados. (Wickham, 2016)
library{scales}      # Para escalar gráficos de ggplot2. (Wickham, 2017a)
library{ggpubr}      # Para personalizar gráficos. (Kassambara, 2017)
```

Las funciones usadas para los modelos de mezcla (*estimación y clúster*) y los ocultos de Markov (*estimación, pseudo-residuos, simulación, probabilidades de avance y retroceso y decodificación de estados*) fueron:

```
betamix(formula, data, k, subset, cluster = NULL, ...) # Modelos de mezcla
dthmm(x, Pi, delta, distn, pm, pn = NULL, discrete = NULL) # Define un modelo oculto de Markov
BaumWelch(object, control = bwcontrol(), ...) # Algoritmo de Baum-Welch
residuals(object, ...) # Pseudo-residuos
simulate(object, nsim = 1, seed = NULL, ...) # Simula un modelo oculto de Markov
backward(x, Pi, distn, pm, pn = NULL) # Probabilidades de avance
forward(x, Pi, delta, distn, pm, pn = NULL) # Probabilidades de retroceso
Viterbi(object, ...) # Algoritmo de Viterbi
```

Las funciones de predicción de estados, distribución condicional y distribución de pronósticos no se encuentran en los paquetes anteriores. Por lo que fueron creados modificando los existentes en Zucchini et al. (2016).

```
# FUNCION PARA PREDICCION DE ESTADOS -----
HMM.Pred.Estado <- function(data.mora, modelo, h=1){
  n <- length(data.mora)
```

```

m      <- length(modelo$delta)
delta  <- modelo$delta
Pi     <- modelo$Pi
alfa   <- modelo$pm$shape1
beta   <- modelo$pm$shape2
la     <- t(forward(data.mora, Pi, delta, "beta",list(shape1=alfa, shape2=beta)))
c      <- max(la[,n])
llk    <- c+log(sum(exp(la[,n]-c)))
statepreds <- matrix(NA,ncol=h,nrow=m)
foo    <- exp(la[,n]-llk)
for (i in 1:h){
  foo <- foo%*%Pi
  statepreds[,i] <- foo
}
return(statepreds)
}

# FUNCION PARA DISTRIBUCION DE PROBABILIDAD CONDICIONAL -----
HMM.Prob.conditional <- function(data.mora, rango, modelo){
  n      <- length(data.mora)
  m      <- length(modelo$delta)
  nxc    <- length(rango)
  dxc    <- matrix(NA,nrow=nxc,ncol=n)
  delta  <- modelo$delta
  Pi     <- modelo$Pi
  alfa   <- modelo$pm$shape1
  beta   <- modelo$pm$shape2
  Px     <- t(sapply(c(1:m), function(i) dbeta(rango, shape1=alfa[i], shape2=beta[i])))
  la     <- t(forward(x=data.mora, Pi=Pi, delta=delta, distn="beta",
    pm=list(shape1=alfa, shape2=beta)))
  lb     <- t(backward(x=data.mora, Pi=Pi, distn="beta",
    pm=list(shape1=alfa, shape2=beta)))
  la     <- cbind(log(delta),la)
  lafact <- apply(la,2,max)
  lbfact <- apply(lb,2,max)
  for (i in 1:n){
    foo    <- (exp(la[,i]-lafact[i])%*%Pi)*exp(lb[,i]-lbfact[i])
    foo    <- foo/sum(foo)
    dxc[,i] <- foo%*%Px
  }
  return(dxc)
}

```

```

# FUNCION PARA DISTRIBUCION DE PRONOSTICOS -----
HMM.Dist.Pronostico <- function(data.mora, rango, modelo, h=1){
  n      <- length(data.mora)
  nxf    <- length(rango)
  m      <- length(modelo$delta)
  delta  <- modelo$delta
  Pi     <- modelo$Pi
  alfa   <- modelo$pm$shape1
  beta   <- modelo$pm$shape2
  probs  <- sapply(c(1:m), function(i) dbeta(data.mora, shape1=alfa[i], shape2=beta[i]))
  dxf    <- matrix(0,nrow=h,ncol=nxf)
  foo    <- delta*probs[1,]
  sumfoo <- sum(foo)
  lscale <- log(sumfoo)
  foo    <- foo/sumfoo
  for (i in 2:n){
    foo    <- foo**Pi*probs[i,]
    sumfoo <- sum(foo)
    lscale <- lscale+log(sumfoo)
    foo    <- foo/sumfoo
  }
  for (i in 1:h){
    foo    <- foo**Pi
    for (j in 1:m) dxf[i,] <- dxf[i,] + foo[j]*dbeta(rango,shape1=alfa[j],shape2=beta[j])
  }
  return(dxf)
}

# ESTIMACION DE LA DISTRIBUCION INICIAL -----
dist.ini <- function(n){return(rep(1,n)/n)}
# dist.ini <- function(secuencia){return(as.numeric(prop.table(table(secuencia))))}

# ESTIMACION DE LAS PROBABILIDADES DE LA MATRIZ DE TRANSICION -----
matrix.trans <- function(secuencia, prob=T){
  MT <- table(secuencia[-length(secuencia)], secuencia[-1])
  if(prob) MT <- prop.table(MT, 1)
  return(as.matrix.data.frame(MT))
}

# ESTIMACION DE LOS PARAMETROS DE LA DISTRIBUCION BETA -----
alfa.moment <- function(x){return(mean(x)*(mean(x)*(1-mean(x))/sd(x)^2-1))}
beta.moment <- function(x){return((1-mean(x))*(mean(x)*(1-mean(x))/sd(x)^2-1))}

```

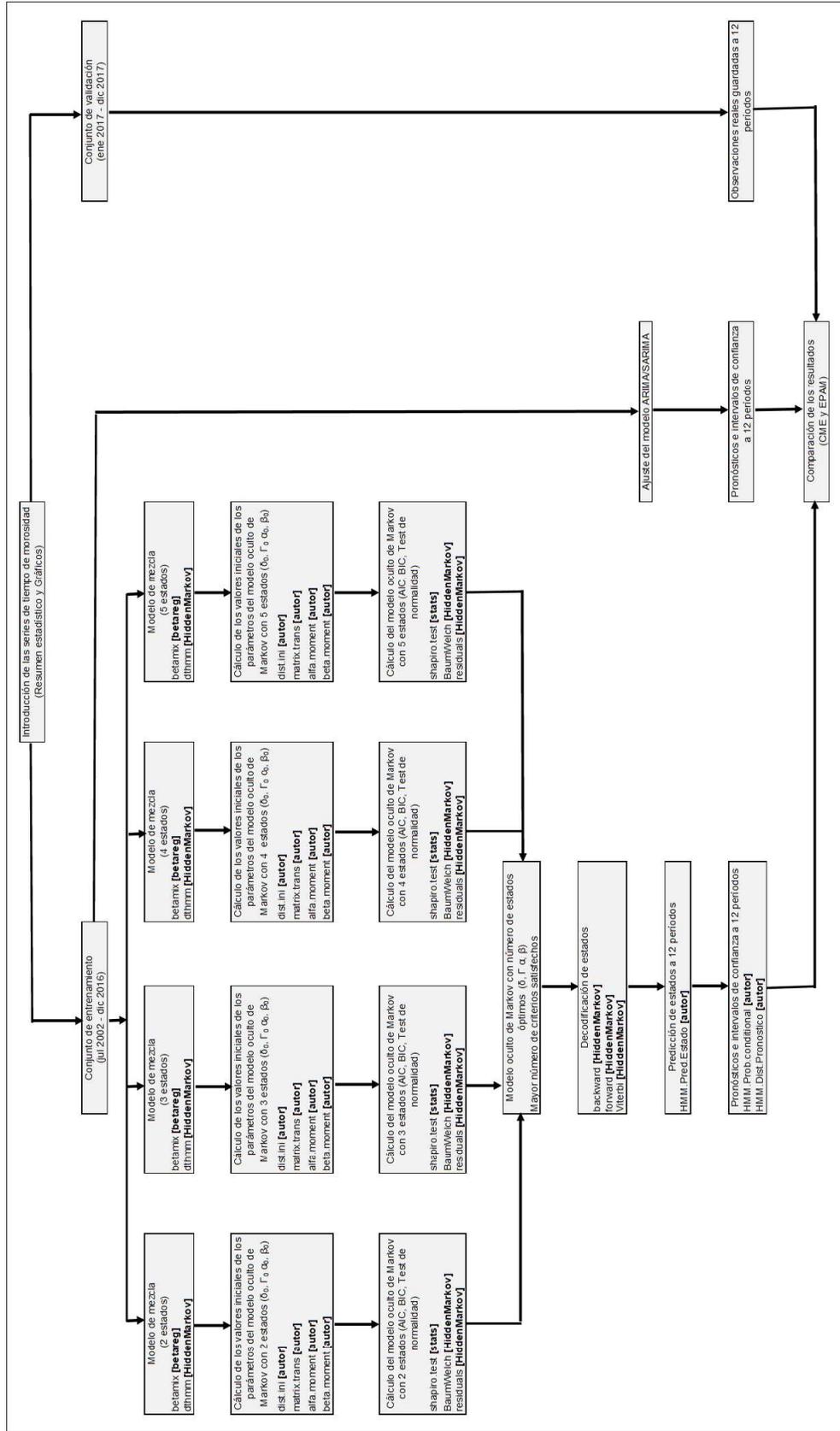


Figura C.1: Flujograma y funciones del algoritmo.



ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

CARRERA DE INGENIERÍA MATEMÁTICA

ORDEN DE EMPASTADO

De acuerdo a lo estipulado en el artículo 27 del Instructivo para la Implementación de la Unidad de Titulación en las Carreras y Programas vigentes de la Escuela Politécnica Nacional, aprobado por Consejo de Docencia en sesión extraordinaria del 29 de abril de 2015 y una vez verificado el cumplimiento del formato de presentación establecido, se autoriza la impresión y encuadernación final del Trabajo de Titulación presentado por el señor **EDWIN DAVID SIMBAÑA FARINANGO**.

Fecha de autorización: Quito, D.M., 3 de junio de 2019.



Dr. Luis Lascano
DECANO
FACULTAD DE CIENCIAS