

UNIVERSIDAD NACIONAL DE CÓRDOBA



**Análisis de la Dinámica de la Trayectoria Estudiantil
mediante Cadenas de Markov Absorbentes.
Caso: Profesorado para la Educación Secundaria en
Matemática, Instituto de Enseñanza Superior Simón
Bolívar.**

Para optar al grado de:
Magíster en Estadística Aplicada

Autor:
Lic.: Víctor Omar Palazzesi

2023



Análisis de la Dinámica de la Trayectoria Estudiantil mediante Cadenas de Markov Absorbentes. Caso: Profesorado para la Educación Secundaria en Matemática, Instituto de Enseñanza Superior Simón Bolívar por Víctor Palazzesi se distribuye bajo una licencia [Creative Commons Reconocimiento 4.0 Internacional License](https://creativecommons.org/licenses/by/4.0/).

Índice general

1. Introducción	1
1.1. Problema de Investigación	1
1.1.1. Percepción de la Institución sobre el problema planteado	4
1.2. Antecedentes	6
1.3. Objetivos	9
1.3.1. Objetivo General	9
1.3.2. Objetivos Específicos	9
1.4. Organización del Trabajo	10
2. Marco teórico	12
2.1. Cadenas de Markov a tiempo discreto	12
2.1.1. Definiciones Preliminares	13
2.1.2. Función de transición y distribución inicial	14
2.1.3. Tiempos de llegada	19
2.1.4. Matriz de transición	21
2.1.5. Ejemplo: jugador ambicioso	22
2.1.6. Ejemplo: cadena de Ehrenfest	35
2.1.7. Clasificación de los estados	51
2.1.8. Descomposición del espacio de estados	59
2.1.9. Representación canónica de la matriz de transición	79
2.2. Distribuciones Estacionarias	81
2.2.1. Existencia y Unicidad de Distribuciones Estacionarias	85
2.3. Cadenas absorbentes	97
2.3.1. Probabilidades de absorción	98
2.3.2. Distribución estacionaria de cadenas absorbentes	100
2.3.3. Probabilidades de transición en cadenas absorbentes	105
2.3.4. Tiempo medio de absorción	109
3. Aplicación de las cadenas de Markov al problema planteado	116
3.1. Introducción	116
3.2. Descripción del problema	118

3.3.	Metodología de trabajo	120
3.4.	Ajustes, resultados y análisis	123
3.4.1.	Transición por el campo de la Formación General	123
3.4.2.	Transición por el campo de la Formación Específica	146
3.4.3.	Transición por el campo de la Práctica Docente	166
4.	Un modelo alternativo: el Análisis de Supervivencia	189
4.1.	Generalidades del análisis de supervivencia	189
4.2.	Notación y terminología	190
4.3.	Curvas de supervivencia de Kaplan-Meier	192
4.4.	Una solución competitiva empleando Análisis de Supervivencia	193
4.4.1.	Metodología de Trabajo	194
4.4.2.	Ajuste del modelo de Kaplan-Meier	194
4.5.	Comparación entre Cadenas de Markov y Análisis de Supervivencia	198
5.	Conclusiones y Recomendaciones	201
5.1.	Importancia de las Cadenas de Markov	201
5.2.	Modelo empleando cadenas de Markov	203
5.3.	Cadenas de Markov vs. Análisis de Supervivencia	205
5.4.	Comparación de los resultados obtenidos con la percepción Institucional	206
5.5.	Recomendaciones y futuras líneas de investigación	207

Agradecimientos

En primer lugar, agradecer a todas aquellas personas y compañeros de trabajo que me alentaron, apoyaron y brindaron toda su confianza, para empezar y poder terminar con esta nueva aventura: la Maestría en Estadística Aplicada. Gracias a todos aquellos que escucharon mis relatos de cursada y me aconsejaban, o simplemente prestaban su oído.

Agradezco a mi Directora de Tesis, Dra. Silvia María Ojeda, por todo su tiempo dedicado, asesoría y ayuda para cumplir con este objetivo.

Seguidamente, agradezco a las autoridades del Instituto de Enseñanza Superior Simón Bolívar, quienes desde un primer momento se mostraron interesados en este trabajo, facilitándome toda la información necesaria para poder llevarlo a cabo.

Mi gratitud para con las autoridades de la Escuela Superior de Comercio Manuel Belgrano, quienes hicieron lo posible para organizar mis horarios para poder cursar la Maestría.

Gracias a todos los docentes, no docentes y compañeros de la Maestría en Estadística Aplicada; por sus valiosas enseñanzas, dedicación y compañerismo; respectivamente.

Me gustaría expresar mi sincera gratitud hacia los docentes que me brindaron una formación inicial excepcional. No puedo dejar de reconocer el profundo impacto que han tenido en mi vida y en mi camino hasta este momento. Ellos fueron mucho más que simples transmisores de conocimiento; a través de su dedicación y enseñanzas, lograron cultivar en mí una pasión por el aprendizaje que aún perdura hoy en día.

Resumen

Palabras clave: Cadenas de Markov a Tiempo Discreto, Estados Absorbentes, Probabilidad de Absorción, Matriz de Transición, Matriz Fundamental, Análisis de Supervivencia, Trayectorias Estudiantiles.

En esta tesis, se busca estimar las probabilidades de transición entre campos de formación por año de la cursada, sobre los cuales se estructura el plan de estudios del Profesorado de Educación Secundaria en Matemática de la Provincia de Córdoba. Para ello, se toma como caso al Instituto de Enseñanza Superior Simón Bolívar, el cual enfrenta todos los años la problemática de la deserción en la mencionada carrera, así como la de estudiantes que deciden pedir un pase para continuar sus estudios en otras instituciones educativas que dictan la misma carrera.

Es aquí donde cobran relevancia las cadenas de Markov a tiempo discreto, las cuales permiten estimar las probabilidades de transición entre diferentes estados. En particular, las cadenas absorbentes son aquellas que permiten obtener la probabilidad de acceder a estados absorbentes, es decir, aquellos de los cuales no se puede acceder a otro.

En primer lugar, se analiza en detalle el marco teórico de referencia para plantear estos modelos, así como también los supuestos que deben cumplirse para su planteo. Se analizan las propiedades y teoremas fundamentales que permitirán arribar a conclusiones relacionadas con la problemática descripta.

Luego de este exhaustivo análisis teórico, se plantean los modelos de las diferentes cadenas de Markov que permiten describir y realizar predicciones sobre la trayectoria académica de los estudiantes por un plan de estudios, así como analizar cuáles son las transiciones más críticas y favorecer la toma de decisiones en pos de fortalecerlas y disminuir la deserción estudiantil. Para llevar a cabo estos ajustes, se consideran los datos provistos por la mencionada institución, correspondientes a los registros académicos de estudiantes inscriptos en el plan de estudios desde 2011 hasta 2019.

Posteriormente se analiza otra alternativa de trabajo estadístico, el Análisis de Supervivencia, con el objetivo de abordar las ventajas y desventajas de utilizar una Cadena de Markov, para describir y tomar decisiones frente al problema

considerado.

Finalmente, se cotejan los resultados obtenidos mediante la percepción que la institución educativa tiene sobre estas trayectorias académicas y se proponen diferentes líneas de acción para mejorar estos modelos.

Summary

Keywords: Discret-Time Markov Chains, Absorbing States, Absorption Probability, Transition Matrix, Fundamental Matrix, Survival Analysis, Students Academic Progress.

This thesis aims to estimate the transition probabilities between fields of study by year, which form the structure of the curriculum of the Secondary Education Mathematics Teaching program in the Province of Córdoba. To do this, the case of the Simón Bolívar Higher Education Institute is taken, which faces the problem of student dropout in the aforementioned program every year, as well as students who decide to transfer to other educational institutions that offer the same program.

This is where discrete-time Markov chains become relevant, which allow for the estimation of transition probabilities between different states. In particular, absorbing chains allow to estimate for the probability of accessing absorbing states, which are those from which no other state can be accessed.

First, the theoretical framework is analyzed in detail to formulate these models, as well as the assumptions that must be met for their formulation. The fundamental properties and theorems are analyzed to arrive at conclusions related to the described problem.

After this exhaustive theoretical analysis, models of different Markov chains are formulated to describe and make predictions about the academic trajectory of students in a curriculum, as well as to analyze which transitions are most critical and facilitate decision-making in order to strengthen them and reduce student dropout. To carry out these adjustments, data provided by the aforementioned institution, corresponding to the academic records of students enrolled in the curriculum from 2011 to 2019, are considered.

Subsequently, another alternative statistical approach, Survival Analysis, is analyzed with the aim of addressing the advantages and disadvantages of using a Markov Chain to describe and make decisions in the face of the considered problem.

Finally, the results obtained are compared through the perception that the educational institution has of these academic trajectories and different lines of action are proposed to improve these models.

Índice de figuras

2.1.	Valores de X_n estimados con simulación en función del número de jugadas n . Caso: apuestas a 7.	26
2.2.	Valores de X_n estimados con simulación en función del número de jugadas n . Caso apuestas a menor que 7.	27
2.3.	Tiempos de espera para lograr una racha de longitud 2.	28
2.4.	Tiempos de espera para obtener longitudes de racha $z=2,3,4,5$	29
2.5.	Gráfico Cadena de Markov. Caso: 10000 apuestas a 7.	34
2.6.	Diagrama de la cadena de Ehrenfest con 3 cajas y 10 bolas.	48
2.7.	Valores de X_n estimados con simulación en función del número de transiciones n . Caso: cadena de Ehrenfest con 3 cajas.	50
2.8.	Partición del espacio de estados	64
3.1.	Probabilidades de transición en un año.	132
3.2.	Probabilidades de transición por año lectivo, dado que actualmente es Ingresante.	135
3.3.	Probabilidades de transición por año lectivo, dado que actualmente es Ingresante.	135
3.4.	Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el primer año del campo de la Formación General.	136
3.5.	Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el primer año del campo de la Formación General.	137
3.6.	Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el segundo año del campo de la Formación General.	138
3.7.	Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el segundo año del campo de la Formación General.	138
3.8.	Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el tercer año del campo de la Formación General.	139
3.9.	Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el tercer año del campo de la Formación General.	140
3.10.	Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el cuarto año del campo de la Formación General.	141

3.11. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el cuarto año del campo de la Formación General.	141
3.12. Probabilidades de transición en un año - Campo de la Formación Específica.	152
3.13. Probabilidades de transición por año lectivo, dado que actualmente es Ingresante.	155
3.14. Probabilidades de transición por año lectivo, dado que actualmente es Ingresante.	156
3.15. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el primer año del campo de la Formación Específica.	157
3.16. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el primer año del campo de la Formación Específica.	157
3.17. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el segundo año del campo de la Formación Específica.	158
3.18. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el segundo año del campo de la Formación Específica.	159
3.19. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el tercer año del campo de la Formación Específica.	160
3.20. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el tercer año del campo de la Formación Específica.	160
3.21. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el cuarto año del campo de la Formación Específica.	161
3.22. Probabilidades de transición en un año.	173
3.23. Probabilidades de transición por año lectivo, dado que actualmente es Ingresante.	176
3.24. Probabilidades de transición por año lectivo, dado que actualmente es Ingresante.	177
3.25. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente I.	178
3.26. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente I.	178
3.27. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente II.	180
3.28. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente II.	180
3.29. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente III.	182
3.30. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente III.	182

3.31. Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente IV.	183
4.1. Función de supervivencia. Estimación: Kaplan-Meier.	195
4.2. Curvas de supervivencia por aprobación de FE1.	198

Índice de cuadros

3.1. Espacios curriculares por campo de la formación y formato.	121
3.2. Transiciones entre estados por año - Campo de la Formación General.	127
3.3. Frecuencia de transiciones entre el estado <i>I</i> y todos los estados para cada periodo.	127
3.4. Frecuencia de transiciones entre el estado <i>FG1</i> y todos los estados para cada periodo.	128
3.5. Frecuencia de transiciones entre el estado <i>FG2</i> y todos los estados para cada periodo.	128
3.6. Frecuencia de transiciones entre el estado <i>FG3</i> y todos los estados para cada periodo.	128
3.7. Frecuencia de transiciones entre el estado <i>FG4</i> y todos los estados para cada periodo.	129
3.8. Resultados de la prueba de independencia entre periodos y transiciones (2014-2019).	130
3.9. Matriz de transición - Campo de la Formación General.	131
3.10. Matriz de transición en 4 pasos - Transición por el campo de la Formación General.	132
3.11. Matrices de transición en más de 4 pasos - Campo de la Formación General.	134
3.12. Conteo de transiciones entre estados del campo de la Formación General para el caso de estudiantes que desertaron.	142
3.13. Conteo de transiciones entre estados del campo de la Formación General para el caso de estudiantes que solicitaron un pase.	143
3.14. Conteo de transiciones entre estados del campo de la Formación General para el caso de estudiantes que egresaron.	143
3.15. Matriz Fundamental de la cadena del campo de la Formación General para el caso de estudiantes que desertaron.	144
3.16. Matriz Fundamental de la cadena del campo de la Formación General para el caso de estudiantes que solicitaron un pase.	144
3.17. Matriz Fundamental de la cadena del campo de la Formación General para el caso de estudiantes que egresaron.	144

3.18. Matriz de acceso eventual a los estados absorbentes desde uno transitorio, para el campo de la Formación General.	145
3.19. Transiciones entre estados por año - Campo de la Formación Específica.	149
3.20. Frecuencia de transiciones entre el estado <i>I</i> y todos los estados para cada periodo.	149
3.21. Frecuencia de transiciones entre el estado <i>FE1</i> y todos los estados para cada periodo.	150
3.22. Frecuencia de transiciones entre el estado <i>FE2</i> y todos los estados para cada periodo.	150
3.23. Frecuencia de transiciones entre el estado <i>FE3</i> y todos los estados para cada periodo.	150
3.24. Frecuencia de transiciones entre el estado <i>FE4</i> y todos los estados para cada periodo.	151
3.25. Resultados de la prueba de independencia entre periodos y transiciones (2014-2019).	151
3.26. Matriz de transición - Campo de la Formación Específica.	152
3.27. Matriz de transición en 4 pasos - Transición por el campo de la Formación Específica.	153
3.28. Matrices de transición en más de 4 pasos - Campo de la Formación Específica.	155
3.29. Conteo de transiciones entre estados del campo de la Formación Específica para el caso de estudiantes que desertaron.	163
3.30. Conteo de transiciones entre estados del campo de la Formación Específica para el caso de estudiantes que solicitaron un pase.	163
3.31. Conteo de transiciones entre estados del campo de la Formación Específica para el caso de estudiantes que egresaron.	163
3.32. Matriz Fundamental de la cadena del campo de la Formación Específica para el caso de estudiantes que desertaron.	164
3.33. Matriz Fundamental de la cadena del campo de la Formación Específica para el caso de estudiantes que solicitaron un pase.	164
3.34. Matriz Fundamental de la cadena del campo de la Formación Específica para el caso de estudiantes que egresaron.	164
3.35. Matriz de acceso eventual a los estados absorbentes desde uno transitorio, para el campo de la Formación Específica.	166
3.36. Transiciones entre estados por año - Campo de la Práctica Docente.	170
3.37. Frecuencia de transiciones entre el estado <i>I</i> y todos los estados para cada periodo.	170
3.38. Frecuencia de transiciones entre el estado <i>PD1</i> y todos los estados para cada periodo.	170

3.39. Frecuencia de transiciones entre el estado $PD2$ y todos los estados para cada periodo.	171
3.40. Frecuencia de transiciones entre el estado $PD3$ y todos los estados para cada periodo.	171
3.41. Frecuencia de transiciones entre el estado $PD4$ y todos los estados para cada periodo.	171
3.42. Resultados de la prueba de independencia entre periodos y transiciones (2014-2019).	172
3.43. Matriz de transición - Campo de la Práctica Docente.	172
3.44. Matriz de transición en 4 pasos - Transición por el campo de la Práctica Docente.	174
3.45. Matrices de transición en más de 4 pasos - Campo de la Práctica Docente.	176
3.46. Conteo de transiciones entre estados del campo de la Práctica Docente para el caso de estudiantes que desertaron.	185
3.47. Conteo de transiciones entre estados del campo de la Práctica Docente para el caso de estudiantes que solicitaron un pase.	185
3.48. Conteo de transiciones entre estados del campo de la Práctica Docente para el caso de estudiantes que egresaron.	185
3.49. Matriz Fundamental de la cadena del campo de la Práctica Docente para el caso de estudiantes que desertaron.	186
3.50. Matriz Fundamental de la cadena del campo de la Práctica Docente para el caso de estudiantes que solicitaron un pase.	186
3.51. Matriz Fundamental de la cadena del campo de la Práctica Docente para el caso de estudiantes que egresaron.	187
3.52. Matriz de acceso eventual a los estados absorbentes desde uno transitorio, para el campo de la Práctica Docente.	188
4.1. Ajuste de la función de supervivencia.	195
4.2. Media y mediana de los tiempos de supervivencia.	196
4.3. Media y mediana de los tiempos de supervivencia habiendo aprobado o no $FE1$	196
4.4. Ajuste de la función de supervivencia sin aprobar $FE1$	196
4.5. Ajuste de la función de supervivencia aprobando $FE1$	197
4.6. Prueba de comparación Log-Rank entre el grupo que no aprobó $FE1$ y el que sí lo aprobó.	197

Capítulo 1

Introducción

1.1. Problema de Investigación

Toda institución educativa plantea en su Proyecto Educativo Institucional (PEI) el objetivo de formar a sus estudiantes con un perfil de egresado específico. Por lo tanto, dichas instituciones esperan que la gran mayoría de sus estudiantes ingresantes puedan completar los trayectos formativos que en ellas se ofrecen.

Para ello, cada institución educativa implementa diversas medidas para fomentar la continuidad en la trayectoria estudiantil y evitar el abandono o interrupción de los estudios que ofrece. Sin embargo, a veces estas intervenciones no tienen el efecto deseado y no logran mantener a los estudiantes en los trayectos académicos. En consecuencia, es importante identificar las causas de la deserción estudiantil para poder abordar adecuadamente este problema.

El abandono de un trayecto académico, especialmente en carreras de formación docente, tiene un impacto negativo tanto en las instituciones como en el sistema educativo en general. Además, los estudiantes desertores también se ven afectados por las consecuencias de abandonar sus estudios. Las altas tasas de abandono escolar en la educación superior, así como el retraso en el egreso de los estudiantes, son factores que impiden la cobertura de vacantes docentes en los diferentes niveles del sistema educativo, lo cual produce efectos perjudiciales sobre los individuos afectados por estas deficiencias. Es por eso que resulta fundamental abordar este problema para reducir su impacto negativo en la sociedad.

Todos estos aspectos negativos reflejan la importancia que reviste el estudio y la prevención de la deserción, no sólo para los individuos afectados por estas problemáticas y las instituciones superiores, sino también para la sociedad en su conjunto.

Una forma de abordar el problema de la deserción académica es conocer la dinámica de transición de los estudiantes dentro del trayecto formativo, identi-

1.1. PROBLEMA DE INVESTIGACIÓN

ficando los momentos en que se produce mayor deserción y aquellos en que el avance en el trayecto se ralentiza. De esta manera, las instituciones educativas podrán indagar sobre los factores que influyen en la deserción de sus estudiantes y desarrollar acciones efectivas para abordarlos. Por lo tanto, resulta fundamental realizar un análisis detallado de la dinámica de transición de los estudiantes para tomar medidas preventivas y fomentar la continuidad de los mismos en sus estudios.

En Argentina, existen pocas investigaciones sobre deserción estudiantil en el nivel superior no universitario; más aún, son mínimas las que utilizan técnicas estadísticas avanzadas para mostrar esta problemática. La mayoría de los estudios se limitan a proporcionar cantidades de matriculados, por trayectos y por provincias. A modo de ejemplo, pueden citarse a los Anuarios Estadísticos del Ministerio de Educación (2018), los cuales se restringen a realizar una estadística descriptiva.

Otros estudios aprovechan esta perspectiva para desarrollar acciones que buscan garantizar la permanencia de los estudiantes en los trayectos formativos. Entre ellos, se encuentra el de Murtaugh, Burns y Schuster (1999), quienes emplean un modelo de supervivencia que utiliza las características del individuo para predecir el tiempo que tardará un estudiante en abandonar sus estudios. Awhadi y Ahmed (2002), por otro lado, ajustan un modelo de regresión logística para identificar los factores más significativos externos a la institución (socioeconómicos, familiares, demográficos, psicológicos, entre otros) y el grado de asociación con la decisión de interrumpir sus estudios.

Autores como Alhawadi y Konsowa (2010) han realizado análisis a nivel institucional, en los cuales, a través de cadenas de Markov, se analiza el comportamiento de los estudiantes respecto a la permanencia en las asignaturas. Caicedo, Atuesta y Caicedo (2016), emplean también cadenas de Markov, analizando la retención de estudiantes de Ingeniería Industrial de la Universidad de Carabobo, Venezuela, basado en la pérdida consecutiva de una misma asignatura.

Si bien los primeros estudios mencionados emplean modelos que identifican y analizan los factores más significativos que influyen en la deserción estudiantil, en muchas ocasiones esta información no está disponible para las instituciones de formación docente. En estos casos, es necesario realizar una instancia de recolección de datos, la cual no sería fácil de obtener una vez producida la interrupción del cursado. Otros estudios se enfocan en el tiempo que transcurre desde el ingreso hasta que se produce el abandono, identificando cuándo es más probable que ocurra la deserción.

Lo que sí es posible desarrollar, son acciones antes de que se produzca el abandono estudiantil, conociendo esta dinámica en la trayectoria académica, que puede modelarse con la información con la que cuentan las instituciones educativas en sus registros, sobre la historia académica de cada estudiante.

1.1. PROBLEMA DE INVESTIGACIÓN

Por esta razón, las cadenas de Markov a tiempo discreto son relevantes, ya que permiten modelar las transiciones en la trayectoria estudiantil. Al definir como estados los diferentes espacios curriculares que cada estudiante debe transitar y aprobar, es posible identificar dónde se produce el mayor retraso o a partir de qué instancia es más probable que se produzca la deserción. A diferencia de otros métodos estadísticos, las cadenas de Markov permiten obtener información detallada no solo sobre el momento en que se produce el abandono o retraso, sino también sobre en qué parte del trayecto académico ocurre.

Considerando los párrafos anteriores y la importancia de contar con indicadores que aborden la problemática de la deserción en las instituciones de educación superior, este trabajo propone la estimación de las probabilidades de transición entre campos de formación por año de la cursada a través del uso de cadenas de Markov. En particular, se utiliza como caso de estudio el Profesorado para la Educación Secundaria en Matemática de la Provincia de Córdoba, una carrera que enfrenta cada año el problema de la deserción, así como el de estudiantes que deciden solicitar un pase para continuar sus estudios en otras instituciones educativas que ofrecen la misma carrera.

Las cadenas de Markov a tiempo discreto son una familia de procesos estocásticos que se definen como una sucesión de variables aleatorias, donde la dependencia entre los eventos sucesivos está determinada únicamente por una unidad de tiempo. Esto significa que la probabilidad futura del proceso depende exclusivamente del estado actual del mismo y no está influenciada por su historia pasada (propiedad de Markov). Tanto la deserción estudiantil como el avance en un trayecto académico pueden ser modelados a través de variables aleatorias que poseen características estocásticas importantes, lo que permite su análisis mediante cadenas de Markov (Brezavšček, Bach & Baggia, 2017).

Con este trabajo se busca ajustar las probabilidades de desertar, pedir un pase o egresar del Profesorado para la Educación Secundaria en Matemática de la Provincia de Córdoba, a partir de diferentes estados de la formación y el tiempo medio que le lleva a los estudiantes llegar a dichos estados. Además, se analizan y comparan las probabilidades de transición entre los campos de formación propuestos en el diseño curricular, para identificar qué año de cada campo resulta más crítico a la hora de que los estudiantes decidan abandonar la carrera. Para plantear estos modelos, se utilizarán las denominadas cadenas absorbentes, que permiten estimar las probabilidades de acceder a estados absorbentes, de los cuales no hay retorno. Este tipo de cadenas es especialmente útil para modelar situaciones como éstas, ya que permiten estimar las probabilidades de transición entre estados y la posibilidad de llegar a un estado absorbente.

Los resultados obtenidos a través de esta modelización pueden ser muy útiles para la institución educativa, ya que permiten identificar los momentos y áreas de

la formación en los que los estudiantes tienen mayor riesgo de abandonar la carrera o cambiarse de institución. Con esta información, se pueden implementar medidas de apoyo y seguimiento específicas para los estudiantes que enfrentan dificultades en estos momentos críticos. De esta manera, se podrían reducir los índices de deserción y mejorar la retención de estudiantes, lo que, a su vez, contribuiría a mejorar la calidad y la eficiencia de la educación superior.

Si bien esta propuesta se realiza sobre una carrera dictada en un instituto de formación docente tomado como caso de estudio, es un trabajo que puede replicarse en otras carreras. Es necesario adaptar el ajuste de los modelos de acuerdo con la información que se tenga disponible en cada caso sobre las trayectorias académicas del estudiantado.

El resultado de este trabajo permitirá brindar un protocolo para que las instituciones de educación superior puedan analizar la deserción del estudiantado. Además, con su implementación, se podrá incorporar información nueva a la metodología y establecer comparaciones entre los resultados obtenidos, con el objetivo de elaborar conclusiones más claras y así fortalecer la permanencia y una transición más fluida en los trayectos educativos.

1.1.1. Percepción de la Institución sobre el problema planteado

En esta sección se analiza cómo se percibe la problemática desde el equipo directivo del Instituto de Enseñanza Superior Simón Bolívar. Para ello, se consultó al personal directivo de la mencionada institución si cuenta con informes de las trayectorias estudiantiles sobre la carrera de Profesorado para la Educación Secundaria en Matemática. En cuanto a la respuesta obtenida, se comunicó que los únicos datos que manejan son los correspondientes al sistema de gestión de alumnos, los cuales no aparecen resumidos o procesados. Por este motivo, recomendaron que se entrevistase a la Coordinadora de Curso designada para la carrera, quien está más cerca de percibir la dinámica en la trayectoria de estudiantes, así como también las problemáticas que en la misma se detectan.

De la entrevista realizada, puede señalarse que la Coordinadora de Curso observa los siguientes aspectos en cuanto a la trayectoria académica de los estudiantes por el Plan de Estudios de la carrera considerada:

- En cuanto a la deserción estudiantil en esta carrera se produce, en gran medida, durante el primer año de cursado. Pueden mencionarse dos momentos durante el primer año de cursada: el primero es cuando los estudiantes se preinscriben y realizan los seminarios de ingreso; posteriormente se inscriben a los espacios curriculares de primer año y entre los primeros días se observa

1.1. PROBLEMA DE INVESTIGACIÓN

un descenso en la cantidad de estudiantes que continúan el cursado. En segundo lugar, luego del receso invernal, se identifica otro momento en donde se produce un alto porcentaje de deserción.

- Respecto a la solicitud de pases para cursar la misma carrera en otra institución, puede decirse que desde que el estudiante ingresa hasta que solicita el pase, transcurren aproximadamente 5 años. En general, se produce cuando el estudiante necesita recibirse con urgencia, ya superó el tiempo estipulado de cursada y no puede lograr egresar. Específicamente, lo solicita el estudiantado que está cursando el tercer y cuarto año.
- Desde que el estudiante ingresa, hasta que logra egresar, transcurren aproximadamente entre 5 y 7 años. Los estudiantes culminan con la carrera aprobando, en última instancia, espacios curriculares del campo de la Formación Específica de tercer y cuarto año.
- En relación al campo de la Formación General, se observa que los estudiantes pueden aprobar los diferentes espacios curriculares, según los tiempos previstos de cursada en el plan de estudios. No se detectan retrasos en el avance por este campo, debido a que no guardan correlatividad con los espacios de otros campos de la formación adeudados de años anteriores. En algunas ocasiones, se registran estudiantes que aprueban los espacios de la Formación General de cuarto año sin haber aprobado los espacios específicos de primer o segundo año. Por ello, se vislumbra un retraso en el egreso desde que el estudiante aprobó este campo.
- Observando las trayectorias por el campo de la Formación Específica, se pueden identificar serias dificultades para lograr la acreditación de los diferentes espacios. Puede verse que un gran número de estudiantes debe recursar entre 3 y 4 veces los espacios de este campo para lograr su aprobación, especialmente los de primer año. En mayor proporción, se destaca la dificultad para aprobar los espacios específicos de primer año, seguidos por los de segundo año.
- En el campo de la Práctica Docente, los estudiantes pueden acreditar la Práctica Docente I y II sin mayores inconvenientes. Sin embargo, se produce una imposibilidad de continuar con el avance hacia la Práctica Docente III ya que requiere la regularidad de espacios curriculares específicos de primer y segundo año. Una vez que el estudiantado acredita la Práctica Docente III, avanza sin problemas hacia la Práctica Docente IV.
- No se podría señalar un momento en particular en la trayectoria del estudiante en el que se pueda asegurar el egreso y descartar la posibilidad de deserción

o cambio de institución. Se considera que los estudiantes tienen mayores posibilidades de terminar la carrera en esta institución y de no desertar, una vez que han logrado acreditar los espacios del campo de la Formación Específica de segundo y tercer año. Sin embargo, en algunos casos particulares, se percibe dificultad en la acreditación de espacios específicos, lo que puede llevar a solicitar el pase a otra institución.

- Poniendo el foco de atención en el estudiante que logra egresar, puede señalarse que permanece en el trayecto académico entre 5 y 7 años. En general, es un cursante que ha podido incorporar el oficio de estudiante de Nivel Superior, es decir que, puede pensar en su trayectoria formativa pudiendo comprender la necesidad de cursar y acreditar cada año propuesto por el diseño curricular, para poder avanzar hacia el año siguiente, entendiendo la importancia de cómo está pensada la cursada en esta formación inicial.

Estas apreciaciones serán retomadas nuevamente cuando se obtengan los resultados de los modelos planteados, con el objetivo de establecer comparaciones entre ambos.

1.2. Antecedentes

Dentro de los artículos de investigación publicados se encuentra diversa información y resultados inherentes al ingreso, permanencia, deserción y egreso estudiantil dentro de un trayecto educativo. En particular, pueden citarse, entre otras publicaciones similares, los boletines de Síntesis de Información de Estadísticas Universitarias del Ministerio de Educación de la Presidencia de la Nación (2017). La mayoría de ellos se centran en resultados obtenidos mediante estadística descriptiva analizando matrículas por área de estudio.

Otros trabajos son de índole pedagógico poniendo foco en problemáticas socio-antropológicas en educación y en la articulación entre niveles del sistema educativo, basados en entrevistas y encuestas realizadas a estudiantes de los primeros años de la educación superior (Ramírez Arballo & Denazis, 2004).

Dentro de las metodologías estadísticas alternativas, se encuentra el trabajo propuesto por Casini, González, Gertel y Giuliadori (2012); en el cual se analizan los determinantes del rendimiento académico y del fracaso escolar en el nivel universitario. Los autores trabajan sobre la hipótesis de que estos determinantes están relacionados y que pueden sintetizarse en los resultados académicos de los estudiantes. Además, realizaron una comparación de los hallazgos reportados en distintos estudios previos a su investigación, donde abordaron el rendimiento académico y la deserción escolar sobre una cohorte de ingresantes a cinco carreras

dictadas en tres facultades de la Universidad Nacional de Córdoba (UNC), Argentina, utilizando análisis de regresión lineal multivariada, técnicas de regresión cuantílica y diferentes especificaciones de funciones de riesgo, respectivamente. El estudio de comparación realizado por los autores, plantea como su principal objetivo el de destacar semejanzas y diferencias en el peso que, dentro de cada modelo, asumen los factores considerados como determinantes de los resultados académicos. Los últimos aparecen circunscriptos a dos medidas sencillas y específicas: un índice de rendimiento o “avance en la carrera” y un indicador de deserción como medida de “fracaso”. Los distintos modelos comparados se apoyan principalmente en las técnicas de OLS (ordinary least squares), regresión logística (LR), quantile regression (QR), e historias de vida o modelos de duración (MD).

Por su parte, se encuentran estudios en donde se logra predecir la deserción universitaria en un periodo de tiempo determinado, mediante un modelo de supervivencia. En algunos de estos trabajos se incorpora la estimación de probabilidades de deserción mediante modelos lineales generalizados a partir de considerar factores que son determinantes en la decisión de que un estudiante abandone sus estudios. Dentro de estos factores pueden mencionarse: la condición de vida, hogar de origen, entorno social, la caracterización socio-económica, aptitudes, entre otros. Los mismos son analizados mediante encuestas previas que se les realiza a los estudiantes al momento de ingresar a la carrera. (Escanés, Herrero y Merlino, 2013).

Entre otros estudios que aprovechan el análisis de factores sociodemográficos y socioeconómicos de los estudiantes, para desarrollar acciones en pos de asegurar la permanencia de los estudiantes en los trayectos formativos, se encuentra la propuesta de Murtaugh, Burns y Schuster (1999), quienes emplean un modelo de supervivencia que utiliza las características del individuo para predecir el tiempo hasta que deserta. Además, Awhadi y Ahmed (2002) emplean un modelo de regresión logística para identificar los factores más significativos y el grado de asociación con la decisión de desertar.

Existen otros trabajos que emplean cadenas de Markov para abordar esta problemática de deserción estudiantil. Caicedo, Atuesta y Caicedo (2016), plantean cadenas de Markov a tiempo discreto para predecir la probabilidad de que un estudiante abandone sus estudios, basado en el número de veces que pierde una misma asignatura de su plan de estudios. Este enfoque le permite a la Universidad, no sólo predecir el momento en que es probable que se produzca la deserción, sino también identificar aquellos espacios curriculares que impactan en esta decisión estudiantil.

Entre los trabajos que abordan la problemática utilizando cadenas de Markov a tiempo discreto, se encuentra el de Ñaupá Ñaupá (2017), quien no considera a los espacios curriculares del plan de estudios como en el trabajo mencionado en

el párrafo anterior, sino que trabaja sobre la condición del estudiante durante la cursada (ingresante - estudiante regular - abandono transitorio - deserción - egreso). Si bien este ajuste permite predecir el tiempo en que se producirá la deserción o el egreso, no permite puntualizar en qué espacios de la cursada es más probable que se produzca el abandono estudiantil.

González Campos, Carvajal Muquillaza y Aspeé Chacón (2020) utilizan también el modelo de caminata aleatoria que constituye una cadena de Markov a tiempo discreto, donde se considera que el estudiante de ser regular, puede seguir en el mismo estado o desertar. En este trabajo también se estiman las probabilidades a partir de modelos de regresión logística que permiten estimar las probabilidades de transición. Estos modelos de regresión incorporan factores que se consideran determinantes en la decisión de desertar.

Si bien entre los estudios mencionados se emplean modelos en los que se identifican y analizan los factores más significativos que influyen en la deserción estudiantil, en muchas ocasiones esta información no está disponible para las instituciones sino que requerirían de una instancia de recolección de datos, la cual no sería de fácil acceso una vez producida la deserción. Sí es posible desarrollar acciones antes de que se produzca el abandono, conociendo esta dinámica en la trayectoria estudiantil, que puede modelarse con la información con la que cuenta cada organismo educativo en sus registros sobre la historia académica de cada estudiante.

Álvarez y Orrego (2000) siguen la misma línea de trabajo, pero no consideran los factores que pueden influir en la decisión de desertar, sino que trabajan en el análisis de explorar cuáles espacios curriculares son más difíciles de aprobar y, por lo tanto, que puedan ser los causantes de la deserción estudiantil. Para ello, en este trabajo, se analizan las llamadas líneas de correlatividades y los autores definen todas las combinaciones posibles de acuerdo al plan de estudios. De esta manera, se estudia la probabilidad de transición por los distintos espacios de esas líneas mediante cadenas de Markov a tiempo discreto. Definen, en base a los resultados, la velocidad de transición en cada cadena ajustada, así como un “índice de fricción” cuyo objetivo es analizar en qué espacios curriculares se produce mayor retardo.

Otros autores como Alhawadi y Konsowa (2010) han realizado análisis a nivel institucional, en los cuales, a través de cadenas de Markov, se analiza el comportamiento de los estudiantes respecto a la permanencia en las asignaturas. Para ello, estos autores definen diferentes niveles (estados) que pueden alcanzar los estudiantes, incluyendo el número de créditos obtenidos a lo largo de la transición por la carrera, así como también los estados de *graduación*, *deserción* y *transferencia* a otra Universidad.

En síntesis, dentro de la literatura disponible, existen algunos estudios de índole descriptivo general y otros más específicos que emplean métodos estadísticos más

avanzados para puntualizar la problemática en diferentes unidades académicas.

Dentro de los trabajos que emplean cadenas de Markov para el abordaje de esta problemática se encuentran diferentes ajustes, entre los cuales se aprecian los que incorporan información de los estudiantes relacionada con factores sociodemográficos y socioeconómicos, y otros que sólo emplean la información académica disponible en cada institución. Puede verse que cada uno de estos ajustes explorados dependen fuertemente de las características del plan de estudios, del propósito en el que se enmarca la problemática de deserción y en la información que es plausible de recolectar en cada una de ellas.

1.3. Objetivos

1.3.1. Objetivo General

Contribuir con el desarrollo de un protocolo para proveer información sobre la dinámica de la deserción estudiantil en carreras de nivel superior a partir del estudio de cadenas de Markov a tiempo discreto absorbentes.

1.3.2. Objetivos Específicos

- Realizar una búsqueda bibliográfica exhaustiva sobre el uso de cadenas de Markov en trabajos sobre deserción estudiantil.
- Estudiar cuáles son las condiciones suficientes para la existencia y unicidad de distribuciones estacionarias para estimar probabilidades de absorción en una cadena de Markov absorbente.
- Evaluar el cumplimiento de los supuestos que permiten ajustar cadenas de Markov a tiempo discreto para modelar la trayectoria estudiantil sobre el Profesorado para la Educación Secundaria en Matemática, dictado en el Instituto de Enseñanza Superior Simón Bolívar.
- Comparar el modelo de análisis de supervivencia frente a las cadenas de Markov, para explorar ventajas y desventajas de cada uno de éstos métodos, a la hora de describir y tomar decisiones frente a la problemática de la deserción estudiantil.
- Analizar y comparar las probabilidades de transición por cada campo de la formación superior, hasta que se produce la absorción, para identificar en qué momento de la trayectoria estudiantil se llega a estos estados.

- Estudiar los tiempos medios de absorción de la trayectoria modelada, con el propósito de brindar información al Instituto de Enseñanza Superior Simón Bolívar, para que el mismo pueda tender líneas de acción focalizadas en pos de mejorar estas trayectorias.
- Brindar a las instituciones educativas de nivel superior, un protocolo para replicar este estudio en las trayectorias académicas que las mismas ofrecen.

1.4. Organización del Trabajo

En el capítulo 1 se presenta el planteamiento del problema, una breve discusión sobre los principales trabajos y antecedentes donde se abordó la problemática en estudio, además de señalar el objetivo general y los específicos de este trabajo.

En el capítulo 2 se lleva a cabo un estudio exhaustivo y riguroso de las cadenas de Markov a tiempo discreto, donde se detallan las definiciones, propiedades y teoremas que sustentan la construcción de modelos para obtener conclusiones acerca de la problemática en cuestión. Si bien muchas de las demostraciones que aparecen en este capítulo pueden encontrarse en bibliografía especializada, se han completado detalles que pueden resultar útiles para el lector y que no son tan evidentes en estos textos. Además, se elaboran ejemplos novedosos tomando como referencia algunos abordados en la literatura, los cuales servirán para ilustrar el marco teórico.

En el capítulo 3 se realiza un análisis detallado del ajuste de cadenas de Markov a tiempo discreto absorbentes para modelizar la trayectoria académica. Se explica la metodología empleada para construir cada uno de los modelos propuestos y se evalúa su adecuación a los supuestos necesarios para su aplicación. Además, se presentan los resultados obtenidos y se procede con su interpretación y discusión. En particular, se describe paso a paso cómo obtener y analizar los resultados para permitir al lector comprender y replicar el proceso.

En el capítulo 4 se presenta un modelo competitivo a las cadenas de Markov: el análisis de supervivencia. Se detallan las condiciones estudiadas en la investigación bibliográfica para construir un ajuste de Kaplan-Meier, que es una herramienta estadística útil para estudiar el tiempo hasta que ocurre un evento de interés. Se presentan los resultados obtenidos al aplicar este método en la problemática abordada y se comparan con los obtenidos en el capítulo anterior. Se exploran similitudes y diferencias entre las dos metodologías, considerando tanto la calidad de los ajustes como la interpretación de los resultados.

Finalmente, en el capítulo 5, se aborda la conclusión acerca de la relevancia de las cadenas de Markov como modelo matemático para abordar la problemática de la deserción estudiantil. Se examinan las conclusiones derivadas de los resultados

1.4. ORGANIZACIÓN DEL TRABAJO

y se comparan con los del método alternativo del análisis de supervivencia, así como con las percepciones previas de la institución sobre la cuestión. Asimismo, se presentan posibles líneas de investigación futura para profundizar en el tema y mejorar la comprensión del abandono de trayectos académicos en la educación superior.

Capítulo 2

Marco teórico

En este capítulo se lleva a cabo un estudio riguroso y exhaustivo de las cadenas de Markov a tiempo discreto, a partir de la investigación bibliográfica realizada. Se incluyen definiciones, propiedades y teoremas fundamentales para la construcción de modelos que permitan explicar la problemática de la deserción estudiantil. Aunque muchas de las demostraciones presentadas en este capítulo se pueden encontrar en diversas fuentes, se han agregado detalles específicos que pueden resultar útiles para el lector y que no son tan evidentes en estas fuentes. Además, se presentan ejemplos novedosos que ilustran el marco teórico, elaborados a partir de los que se abordan en la literatura.

Es importante destacar que el marco general de este tema es amplio y complejo, por lo que se sugiere al lector consultar bibliografía especializada para obtener un tratamiento más detallado sobre las distintas clasificaciones de cadenas de Markov existentes.

En síntesis, este capítulo se enfoca en brindar una perspectiva completa y detallada sobre las cadenas de Markov a tiempo discreto absorbentes y enriquece el conocimiento existente sobre este tema en la bibliografía especializada.

2.1. Cadenas de Markov a tiempo discreto

Las cadenas de Markov llevan su nombre en honor al profesor matemático ruso Andrei A. Markov (1856-1922), quien publicó su primer resultado sobre el tema en 1906. Markov se matriculó en la Universidad de San Petersburgo, donde obtuvo una maestría y un doctorado. Fue discípulo de Tchebysheff, quien al retirarse de la universidad fue sustituido por el propio Markov. Desde allí se desempeñó como profesor en San Petersburgo y también fue miembro de la Academia de Ciencias de Rusia. Se retiró en 1905, pero continuó su enseñanza en la universidad hasta su muerte, dictando diferentes cursos de teoría de probabilidades.

Markov es particularmente recordado por su estudio de las cadenas que llevan su nombre. Sus trabajos de investigación sobre este tema, dieron inicio al estudio de los procesos estocásticos junto con sus diversas aplicaciones. Su intención era crear un modelo probabilístico para analizar la frecuencia con la que aparecen las vocales en poemas y textos literarios. El éxito del modelo propuesto por Markov radica en que es lo suficientemente complejo como para describir ciertas características no triviales de algunos sistemas, pero al mismo tiempo es lo suficientemente sencillo para ser analizado matemáticamente.

Las cadenas de Markov pueden aplicarse a una amplia gama de fenómenos científicos y sociales, y se cuenta con una teoría matemática extensa al respecto.

¿Pero qué es un proceso estocástico? A continuación se realiza una breve introducción para luego abordar la definición formal de cadena de Markov y mostrar algunos resultados importantes.

Sea T un subconjunto de \mathbb{R} y (Ω, \mathcal{A}, P) un espacio de probabilidad. Un *proceso aleatorio* es una función

$$X : T \times \Omega \rightarrow \mathbb{R}$$

tal que para cada $t \in T$, $X(t, \cdot)$ es una variable aleatoria $X : \Omega \rightarrow \mathbb{R}$.

Si se fija $\omega \in \Omega$ se obtiene una función $X(\cdot, \omega) : T \rightarrow \mathbb{R}$ que se conoce como una *trayectoria* del proceso: $X : T \rightarrow \mathbb{R}$.

En general, se interpreta el parámetro t como el tiempo aunque también pueden considerarse procesos con índices en espacios más generales. En lo que se desarrolla en esta tesis, como se trabajará con cadenas de Markov a tiempo discreto, T será un subconjunto de \mathbb{N}_0 .

2.1.1. Definiciones Preliminares

Se considera un sistema conformado por un número finito o infinito numerable de elementos. Sea \mathcal{S} este conjunto de elementos al que puede llamarse *espacio de estados del sistema* y a dichos elementos se los llamará *estados* del sistema.

Con X_n se denota el estado del sistema observado en el momento n . Como aquí interesan los sistemas no determinísticos, se piensa en X_n con $n \in \mathbb{N}_0$, como variables aleatorias definidas en el mismo espacio de probabilidad.

En muchos sistemas de la práctica, estados pasados y presentes ejercen influencia en los estados futuros. Pero otros sistemas tienen la propiedad que dado el estado presente, el estado pasado no tiene influencia sobre estados futuros. Esta propiedad se conoce como *propiedad de Markov* y los sistemas que poseen esta propiedad se llaman *cadenas de Markov*.

Definición 2.1. Una *cadena de Markov* a tiempo discreto es una sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}_0}$ que toman valores en un conjunto finito o infinito numerable \mathcal{S} , conocido como espacio de estados, y que satisface la siguiente propiedad:

$$P(X_{n+1} = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} \mid X_n = x_n) \quad (2.1)$$

donde $x_0, x_1, \dots, x_n \in \mathcal{S}$.

La propiedad (2.1) se conoce como *propiedad de Markov*.

Entonces, una cadena de Markov, es una sucesión de variables aleatorias X_0, X_1, \dots , tal que para cada instante n , la distribución del “futuro” X_{n+1} es independiente de los estados “pasados” $X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}$, dado el estado “presente” $X_n = x_n$.

Definición 2.2. La probabilidad condicional

$$P(X_{n+1} = y \mid X_n = x) \quad (2.2)$$

se llama *probabilidad de transición de la cadena* y se denotará en adelante con $P(x, y)$ e indica la probabilidad de transición de la cadena en un paso desde el estado x al estado y .

De aquí en lo que sigue $\{X_n\}_{n \in \mathbb{N}_0}$ conforma una cadena de Markov. Esa sucesión de variables aleatorias satisfacen (2.1) y tiene probabilidad de transición.

Cabe mencionar, que se estudiarán aquí cadenas de Markov tomando probabilidades de transición *estacionarias* u *homogéneas*, que son aquellas cadenas en las cuales $P(X_{n+1} = y \mid X_n = x)$ es independiente de n .

2.1.2. Función de transición y distribución inicial

Definición 2.3. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} . La función $P(x, y)$, con $x, y \in \mathcal{S}$, definida por

$$P(x, y) = P(X_1 = y \mid X_0 = x) \quad (2.3)$$

se llama *función de transición de la cadena* y es tal que:

- i) Por tratarse de una función de probabilidad se cumple el axioma de no negatividad:

$$\forall x, y \in \mathcal{S} : P(x, y) \geq 0 \quad (2.4)$$

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

- ii) Por tratarse de la función de probabilidad condicional de una variable aleatoria, resulta:

$$\forall x \in \mathcal{S} : \sum_{y \in \mathcal{S}} P(x, y) = 1 \quad (2.5)$$

- iii) Puesto que la cadena de Markov tiene probabilidades de transición estacionarias.

$$\forall n \in \mathbb{N} : P(X_{n+1} = y \mid X_n = x) = P(x, y) \quad (2.6)$$

- iv) Como consecuencia de (2.1), surge:

$$P(X_{n+1} = x_{n+1} \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = P(x, y) \quad (2.7)$$

Si la cadena de Markov está en el estado x en el momento n , entonces no importa cómo llegó a x , tiene probabilidad $P(x, y)$ de estar en el estado y en el siguiente paso.

En otros casos, interesa calcular la probabilidad de que un proceso pase del estado x al estado y en m pasos, en vez de en uno, como se abordó con anterioridad.

Definición 2.4. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Sean $x, y \in \mathcal{S}$, y $m, n \in \mathbb{N}_0$. Se define la probabilidad de que la cadena de Markov pase del estado x al estado y en m pasos, y se denota con $P^{(m)}(x, y)$, como:

$$P^{(m)}(x, y) = P(X_{n+m} = y \mid X_n = x) \quad (2.8)$$

En el caso en que $m = 1$ se tiene:

$$P^{(1)}(x, y) = P(x, y) \quad (2.9)$$

Y en caso que $m = 0$:

$$P^{(0)}(x, y) = \begin{cases} 1 & \text{si } x = y, \\ 0 & \text{en caso contrario.} \end{cases} \quad (2.10)$$

El siguiente resultado, refuerza la idea de que si se conoce el presente, el pasado no tiene influencia en el comportamiento futuro de la cadena de Markov.

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

Proposición 2.1. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Sean $x, y, z \in \mathcal{S}$ y $0 \leq m \leq n - 1$ con $m, n \in \mathbb{N}_0$. Se tiene que:

$$\begin{aligned} P(X_{n+1} = y \mid X_n = x, X_m = z) &= P(X_{n+1} = y \mid X_n = x) \\ &= P(x, y) \end{aligned} \quad (2.11)$$

Demostración. Para simplificar la notación se realiza la prueba en el caso en que $m = 0$, la demostración del caso general es completamente análoga. En primer lugar, se considera que el espacio muestral puede escribirse como la siguiente unión disjunta:

$$\Omega = \bigcup_{x_1, \dots, x_{n-1} \in \mathcal{S}} \{X_1 = x_1, \dots, X_{n-1} = x_{n-1}\} \quad (2.12)$$

En particular, se considera $\Omega = \{X_i \in \mathcal{S}\}$, para calcular la probabilidad que aparece en la tesis del enunciado de la proposición.

De esta manera, aplicando propiedades de la probabilidad condicional:

$$\begin{aligned} &P(X_{n+1} = y \mid X_n = x, X_0 = z) \\ &= \frac{P(X_{n+1} = y, X_n = x, X_{n-1} \in \mathcal{S}, \dots, X_1 \in \mathcal{S}, X_0 = z)}{P(X_n = x, X_0 = z)} \\ &= \sum_{x_1, \dots, x_{n-1} \in \mathcal{S}} \frac{P(X_{n+1} = y, X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = z)}{P(X_n = x, X_0 = z)} \\ &= \sum_{x_1, \dots, x_{n-1} \in \mathcal{S}} P(X_{n+1} = y \mid X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = z) \cdot \\ &\quad \frac{P(X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = z)}{P(X_n = x, X_0 = z)} \end{aligned} \quad (2.13)$$

Ahora, aplicando la propiedad de Markov en el primer factor de cada término de la sumatoria del último miembro de la serie de igualdades, se obtiene:

$$\begin{aligned} P(X_{n+1} = y \mid X_n = x, X_0 = z)(*) &= \sum_{x_1, \dots, x_{n-1} \in \mathcal{S}} P(X_{n+1} = y \mid X_n = x) \cdot \\ &\quad \frac{P(X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = z)}{P(X_n = x, X_0 = z)} \end{aligned} \quad (2.14)$$

Usando la notación de función de probabilidad de transición en un solo paso y poniéndola en evidencia como factor por no depender del índice de la sumatoria,

se cumple:

$$\begin{aligned}
 (*) &= P(x, y) \cdot \sum_{x_1, \dots, x_{n-1} \in \mathcal{S}} \frac{P(X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = z)}{P(X_n = x, X_0 = z)} \\
 &= P(x, y) \cdot \frac{P(X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = z)}{P(X_n = x, X_0 = z)} \\
 &= P(x, y)
 \end{aligned} \tag{2.15}$$

■

En el siguiente teorema se plantea una ecuación que permite obtener las probabilidades de transición entre dos estados en un número finito de pasos dado, a través de las probabilidades de transición de los posibles estados intermedios. Dicha ecuación recibe el nombre de sus creadores: Sidney Chapman y Andrey Kolmogorov.

Teorema 1. Ecuación de Chapman-Kolmogorov

Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Se cumple:

$$P^{(m)}(x, y) = \sum_{z \in \mathcal{S}} P^{(n)}(x, z) P^{(m-n)}(z, y) \tag{2.16}$$

para cualquier $m, n \in \mathbb{N}_0$ tal que $n \leq m$.

Demostración. Para calcular $P^{(m)}(x, y)$ se aplica la definición de probabilidad de transición en m pasos, se realiza una partición del espacio de estados según los valores posibles de la cadena en el instante n , se aplica la definición de probabilidad

condicional y, por último, se utiliza la relación (2.11)

$$\begin{aligned}
 P^{(m)}(x, y) &= P(X_m = y \mid X_0 = x) \\
 &= \sum_{z \in \mathcal{S}} P(X_m = y, X_n = z \mid X_0 = x) \\
 &= \sum_{z \in \mathcal{S}} \frac{P(X_m = y, X_n = z, X_0 = x)}{P(X_0 = x)} \\
 &= \sum_{z \in \mathcal{S}} \frac{P(X_m = y, X_n = z, X_0 = x)}{P(X_0 = x)} \cdot \frac{P(X_n = z, X_0 = x)}{P(X_n = z, X_0 = x)} \\
 &= \sum_{z \in \mathcal{S}} \frac{P(X_m = y, X_n = z, X_0 = x)}{P(X_n = z, X_0 = x)} \cdot \frac{P(X_n = z, X_0 = x)}{P(X_0 = x)} \quad (2.17) \\
 &= \sum_{z \in \mathcal{S}} P(X_m = y \mid X_n = z, X_0 = x) \cdot P(X_n = z \mid X_0 = x) \\
 &= \sum_{z \in \mathcal{S}} P(X_m = y \mid X_n = z) \cdot P(X_n = z \mid X_0 = x) \\
 &= \sum_{z \in \mathcal{S}} P^{(n)}(x, z) P^{(m-n)}(z, y)
 \end{aligned}$$

■

Corolario 1. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Si la probabilidad de que el proceso esté inicialmente en x es $\pi_0(x)$:

$$P(X_0 = x) = \pi_0(x)$$

entonces la probabilidad de que el proceso esté en el estado y en el instante n es:

$$P(X_n = y) = \sum_{x \in \mathcal{S}} \pi_0(x) P^{(n)}(x, y) \quad (2.18)$$

Demostración. Si π_0 es la distribución inicial de la cadena de Markov, se cumple:

$$\begin{aligned}
 P(X_n = y) &= \sum_{x \in \mathcal{S}} P(X_0 = x, X_n = y) \\
 &= \sum_{x \in \mathcal{S}} P(X_0 = x) P(X_n = y \mid X_0 = x) \quad (2.19) \\
 &= \sum_{x \in \mathcal{S}} \pi_0(x) P^{(n)}(x, y)
 \end{aligned}$$

■

Nota. Se denotará con $P_x(\cdot)$ a la probabilidad de aquellos eventos definidos en términos de una cadena de Markov que comienza en x .

2.1.3. Tiempos de llegada

En muchas situaciones interesa conocer el tiempo que tarda una cadena de Markov en pasar de un estado a otro determinado.

Definición 2.5. Sea $A \subset \mathcal{S}$. Se define el tiempo de llegada T_A de A como la variable aleatoria:

$$T_A = \min\{n > 0 / X_n \in A\}$$

si $X_n \in A$ para algún $n \in \mathbb{N}$.

Por su parte, se denotará con $T_A = \infty$ si $X_n \notin A$, cualquiera sea $n \in \mathbb{N}$.

En otras palabras, T_A es el menor número de pasos necesarios desde el inicio de la cadena hasta que ésta visita a A . También, $T_A = \infty$ se interpreta como que desde el inicio de la cadena, ésta nunca visita A .

En el caso en que A sea un conjunto unitario $A = \{a\}$ se utilizará la notación T_a en lugar de $T_{\{a\}}$.

La siguiente proposición brinda una relación entre la probabilidad de transición en n pasos y los tiempos de llegada.

Proposición 2.2. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} , donde $x, y \in \mathcal{S}$. Luego, se cumple:

$$P^{(n)}(x, y) = \sum_{m=1}^n P_x(T_y = m) P^{(n-m)}(y, y), \quad \forall n \geq 1 \quad (2.20)$$

Demostración. Para demostrar esta proposición se descompone al evento de interés $\{T_y = m, X_n = y\}$, con $1 \leq m \leq n$, de acuerdo al instante a la primera visita al

estado y . Como dichos eventos son disjuntos se obtiene:

$$\begin{aligned}
 P^{(n)}(x, y) &= P_x \left(\bigcup_{m=1}^n \{T_y = m, X_n = y\} \right) \\
 &= \sum_{m=1}^n P_x(T_y = m, X_n = y) \\
 &= \sum_{m=1}^n P_x(X_n = y \mid T_y = m) P_x(T_y = m) \\
 &= \sum_{m=1}^n P(X_n = y \mid X_m = y, X_{m-1} \neq y, \dots, X_1 \neq y, X_0 = x) P_x(T_y = m) \\
 &= \sum_{m=1}^n P(X_n = y \mid X_m = y) P_x(T_y = m) \\
 &= \sum_{m=1}^n P_x(T_y = m) P^{(n-m)}(y, y)
 \end{aligned} \tag{2.21}$$

En la serie de igualdades que se empleó en la demostración, se aplicó la probabilidad de una unión disjunta de eventos que es igual a la suma de las probabilidades de dichos eventos, la definición de probabilidad condicional, la definición de tiempo de llegada, la propiedad de Markov y la definición de probabilidad de transición en n pasos. ■

Puede observarse que:

$$P_x(T_y = 1) = P(x, y)$$

es decir, que la probabilidad de que la cadena de Markov que comienza en x visite por primera vez al estado y en un paso, es igual a la probabilidad de transición del estado x al y en un paso.

Además, como consecuencia del Teorema 1:

$$\begin{aligned}
 P_x(T_y = 2) &= \sum_{z \neq y} P_x(X_1 = z, X_2 = y) \\
 &= \sum_{z \neq y} P(x, z) P(z, y)
 \end{aligned} \tag{2.22}$$

Para valores mayores de n se tiene:

$$P_x(T_y = n + 1) = \sum_{z \neq y} P(x, z) P_z(T_y = n), \quad n \geq 1.$$

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

Es decir que, dada una cadena de Markov, para llegar a y desde x por primera vez en $n + 1$ pasos, es necesario visitar algún estado z distinto de y en el 1° paso y luego ir de z a y por primera vez al final de n pasos adicionales.

Definición 2.6. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Se define:

$$\rho_{xy} = P_x(T_y < \infty)$$

como la probabilidad de que la cadena de Markov que comienza en x visite al estado y en algún momento, cualesquiera sean los estados $x, y \in \mathcal{S}$.

En particular ρ_{yy} denota la probabilidad de que la cadena de Markov comience en y y regrese a y .

2.1.4. Matriz de transición

Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados finito $\mathcal{S} = \{0, 1, \dots, d\}$. Con las probabilidades de transición $P(x, y)$, $\forall x, y \in \mathcal{S}$, puede construirse una matriz P cuyas entradas sean dichas probabilidades de transición en un paso.

En este caso puede pensarse en P como la matriz de transición con $d + 1$ filas y $d + 1$ columnas definida por:

$$P = \begin{pmatrix} P(0,0) & P(0,1) & \dots & P(0,d) \\ P(1,0) & P(1,1) & \dots & P(1,d) \\ \vdots & \vdots & \ddots & \vdots \\ P(d,0) & P(d,1) & \dots & P(d,d) \end{pmatrix}$$

La siguiente proposición establece una relación importante entre la probabilidad de transición en n pasos y la matriz de transición, facilitando el cálculo de dichas probabilidades de transición conociendo previamente la matriz de transición.

Proposición 2.3. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P , con espacio de estados finito $\mathcal{S} = \{0, 1, \dots, d\}$ y matriz de transición $P = (P(x, y))$, $\forall x, y \in \mathcal{S}$. Se cumple que la matriz de transición en n pasos, cuyas entradas son las probabilidades de transición en n pasos, es $P^n = (P^{(n)}(x, y))$.

Demostración. Para demostrar esta proposición se realizará inducción sobre n .

- Para $n = 1$ la propiedad se cumple, ya que P es la matriz de transición cuyas entradas son las probabilidades de transición en un solo paso $P(x, y)$.
- Se supone válida la proposición para $n = k$, es decir que P^k es la matriz cuyas entradas son las probabilidades de transición en k pasos.
- Para $n = k + 1$, retomando el Teorema 1 de la ecuación de Chapman-Kolmogorov, se tiene que:

$$P^{(k+1)}(x, y) = \sum_{z \in \mathcal{S}} P^{(k)}(x, z)P(z, y)$$

y, por definición de multiplicación de matrices, se tiene que:

$$\begin{aligned} P^{(k+1)}(x, y) &= [(P^k \cdot P)(x, y)] \\ &= [(P^{k+1})(x, y)] \end{aligned} \tag{2.23}$$

Luego, la proposición es válida $\forall n \in \mathbb{N}$. ■

2.1.5. Ejemplo: jugador ambicioso

La longitud de rachas es un ejemplo clásico que se encuentra con frecuencia en la literatura sobre cadenas de Markov. En esta subsección, se explorará en detalle esta cadena y se presentarán ejemplos novedosos que ilustran el marco teórico estudiado hasta el momento.

Se realiza una sucesión de juegos en idénticas condiciones con probabilidad de éxito p y de fracaso $1 - p$. Se dice que ocurre una racha de longitud z en el n -ésimo juego si han ocurrido z éxitos sucesivos en el instante n luego de un fracaso en la jugada $n - z$.

$$\underbrace{F}_{n-z} \quad \underbrace{E}_{n-z+1} \quad \underbrace{E}_{n-z+2} \quad \underbrace{E}_{n-z+3} \quad \dots \quad \underbrace{E}_n$$

Para estudiar este proceso como una cadena de Markov, se definen los siguientes estados: si el ensayo resulta en fracaso, el estado es 0, mientras que si resulta en éxito, el estado es el número de éxitos que han ocurrido en sucesión. Por lo tanto, desde cualquier estado x puede haber una transición al estado 0 (si hay un fracaso en el próximo juego) con probabilidad $1 - p$, mientras que si hay un éxito, la racha continúa.

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

Suponga que se quiere simular la suerte de un jugador que apuesta toda su fortuna cada vez que juega en un juego que le permite ganar \$1 con probabilidad p o perder todo el dinero apostado con probabilidad $1 - p$. Si la casa de juegos le permite seguir jugando aunque ya no tenga dinero, se define la variable aleatoria:

X_n : “Saldo del jugador en el instante n ”

Entonces:

$$X_{n+1} = \begin{cases} X_n + 1 & \text{con probabilidad } p, \\ 0 & \text{con probabilidad } 1 - p. \end{cases} \quad (2.24)$$

Mientras que en n pasos, se tienen dos posibilidades, que el jugador no pierda ni una sola vez, o bien pierde alguna vez. Entonces, cuando el jugador no pierde:

$$P^{(n)}(x, z) = p^n, \quad \text{si } z = x + n$$

Si el jugador pierde en alguno de los n juegos:

$$P^{(n)}(x, z) = (1 - p)p^z, \quad 0 \leq z \leq n - 1$$

El caso en el que el jugador no pierde, puede demostrarse por inducción:

- Para $n = 1$, se tiene que el jugador gana una jugada con probabilidad p , entonces, si $z = x + 1$:

$$\begin{aligned} P(x, z) &= p \\ &= p^1 \end{aligned} \quad (2.25)$$

Luego, la propiedad es válida para $n = 1$.

- Se supone válida la propiedad para $n = k$, es decir:

$$P^{(k)}(x, z) = p^k, \quad \text{si } z = x + k$$

Para $n = k + 1$ se tiene:

$$\begin{aligned} P^{(k+1)}(x, z) &= \sum_{y \in \mathcal{S}} P^{(k)}(x, y)P(y, z) \\ &= p^k p \\ &= p^{k+1} \end{aligned} \quad (2.26)$$

Se aplicó la ecuación de Chapman-Kolmogorov y el hecho de que los términos de la sumatoria son todos nulos excepto el término en donde $y = z - 1$.

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

Luego, la propiedad es válida $\forall n \in \mathbb{N}$.

Para demostrar la segunda afirmación se analiza en primer lugar cómo calcular la probabilidad en el caso en que $z = 0$. Por la ecuación de Chapman-Kolmogorov: $\forall n \geq 2, n \in \mathbb{N}$

$$\begin{aligned} P^{(n)}(x, 0) &= \sum_{z \in \mathcal{S}} P^{(n-1)}(x, z)P(z, 0) \\ &= (1 - p) \sum_{z \in \mathcal{S}} P^{(n-1)}(x, z) \\ &= 1 - p \end{aligned} \tag{2.27}$$

Del segundo al tercer paso de la serie de igualdades, se aplicó el hecho de que la sumatoria de las probabilidades $P^{(n-1)}(x, z)$ sobre todos los z del espacio de estados, es igual a 1, por ser una distribución de probabilidad.

Ahora, el evento $\{X_n = z\}$, dado que $X_0 = x$, ocurre cuando se gana sin parar z jugadas sucesivas, antecidos de una sucesión de $n - z$ juegos que terminan por un juego perdido:

$$\underbrace{(x, \dots, \dots, 0)}_{n - z \text{ juegos}}$$

Esto puede verse con la ecuación de Chapman-Kolmogorov, para $n \geq 2$ y $0 \leq z \leq n - 1$:

$$\begin{aligned} P^{(n)}(x, z) &= \sum_{y \in \mathcal{S}} P^{(n-z)}(x, y)P^{(z)}(y, z) \\ &= P^{(n-z)}(x, 0)P^{(z)}(0, z) \\ &= (1 - p)p^z \end{aligned} \tag{2.28}$$

En la segunda igualdad, se utiliza el hecho de que la única manera de acceder al estado z , en exactamente z pasos, es a partir del estado 0 (el jugador no tiene dinero) y no pierde ningún juego, lo cual ocurre con probabilidad p^z .

Como puede verse, el espacio de estados de esta cadena de Markov es infinito, ya que el número de éxitos depende del número de jugadas que el jugador desee realizar.

Se quiere simular un modelo de una cadena de Markov para el caso de un apostador que juega al “7” con los dados. Este es un juego de casino en el que se utilizan dos dados y las apuestas se determinan con un solo tiro de los dados. El objetivo del juego es apostar a que saldrá un número mayor, menor o igual a 7.

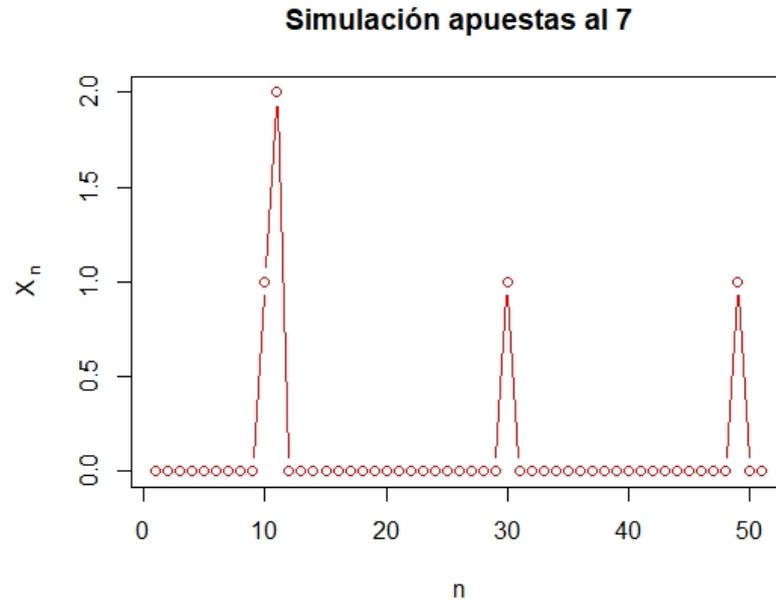


Figura 2.1: Valores de X_n estimados con simulación en función del número de jugadas n . Caso: apuestas a 7.

```
+ X[i+1]<-sample(x=c(X[i]+1,0),size=1,prob=c(p,1-p))
+ }
> X
[1] 0 1 2 0 0 1 2 3 0 1 0 1 0 1 0 0 0 0 1 2 0 0 0 1 0 1 2 0 1 2
[31] 0 0 0 1 2 0 1 2 3 4 0 1 0 1 0 1 0 0 1 0 1
```

En el gráfico (Figura 2.2) se muestran los valores de X_n estimados con esta simulación en función del número de jugadas n .

Aquí, al tener más probabilidad de ganar, aumenta el número de rachas en estas 50 jugadas y la longitud de las mismas asciende hasta 4 en una oportunidad.

Puede analizarse también, la variable aleatoria: tiempo que le toma a la cadena de Markov alcanzar una racha de una longitud z . Para construir esta simulación en R, se procede de manera análoga con la diferencia de que el proceso se simulará hasta que X_n sea igual a z , guardando dichos n .

Se simula la distribución de la variable aleatoria T_z : tiempo que le toma a la cadena alcanzar una racha de longitud $z = 2$:

```
> N=1000 #Numero de simulaciones
> z=2 #Numero de rachas
```

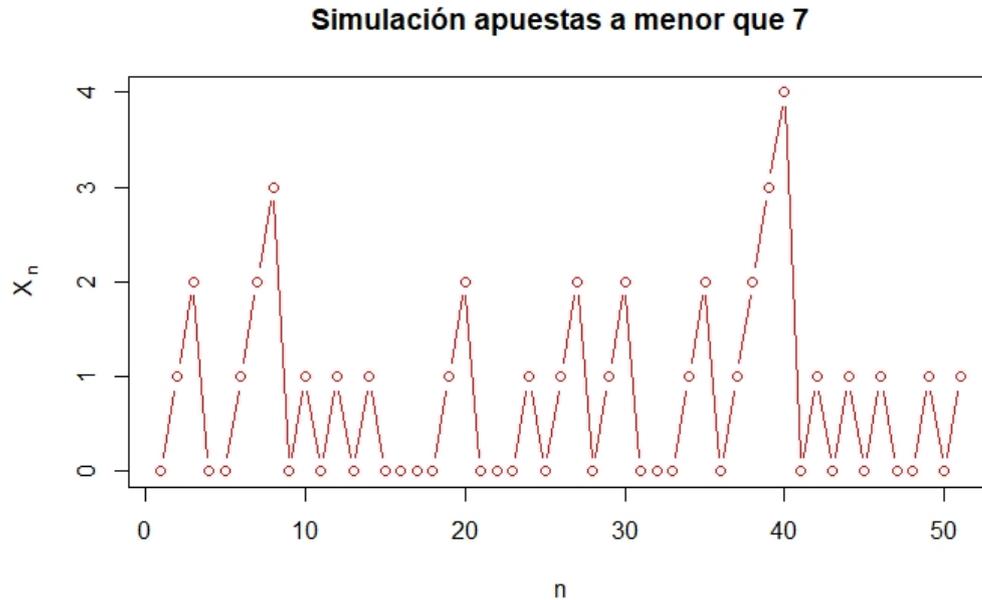


Figura 2.2: Valores de X_n estimados con simulación en función del número de jugadas n . Caso apuestas a menor que 7.

```

> I=0 #Punto Inicial
> p=1/6 #Probabilidad de Éxito
> T_z=rep(0,N)
> for(i in 1:N){
+   Contar=0
+   Racha=I
+   while(Racha<z){
+     Contar=Contar+1
+     Racha=sample(x=c(Racha+1,0),size=1,prob=c(p,1-p))}
+   T_z[i]=Contar}
> barplot(prop.table(table(T_z)),main="Probabilidades Empiricas de T_z",
+         col="blue",xlab="n",ylab="P(T_n=z)")

```

En esta salida puede observarse que el tiempo mínimo que debe esperarse para obtener una racha de longitud 2 es de dos jugadas, el cual coincide con el tiempo de llegada teórico para una cadena que comienza en el estado 0. De todos los tiempos estimados con esta simulación, la mayoría ocurre entre los menores números de jugadas.

En la Figura 2.3, puede verse que la distribución de la variable aleatoria T_2 es

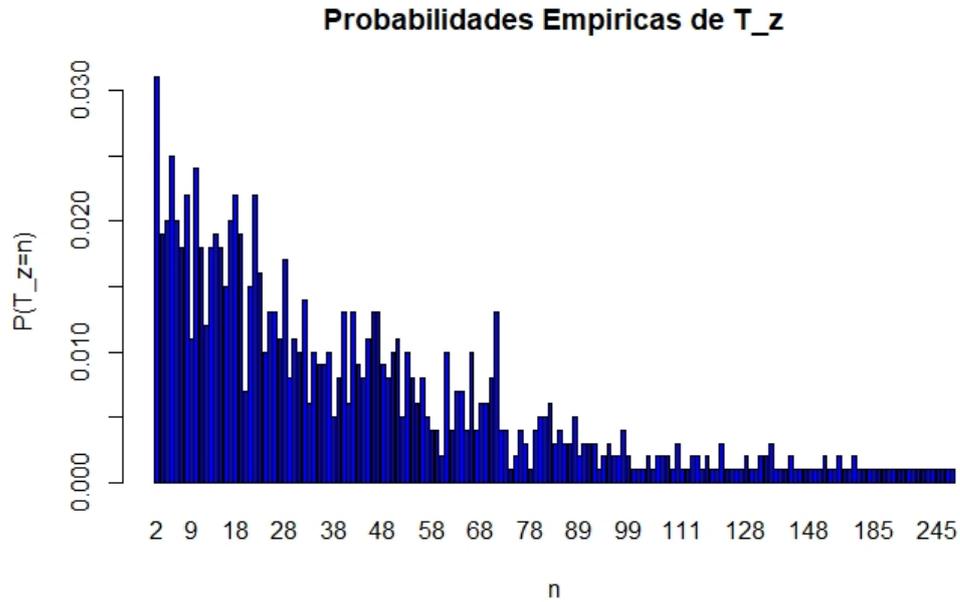


Figura 2.3: Tiempos de espera para lograr una racha de longitud 2.

asimétrica hacia la derecha, indicando que las longitudes de racha 2 ocurren con mayores probabilidades en tiempos de espera más cortos y que es poco probable tener que esperar un tiempo largo hasta que ocurra dicha racha.

Si se repite el procedimiento para longitudes de rachas que varían desde $z = 2$ hasta $z = 5$ en el juego de apostar al 7, se obtienen las siguientes salidas:

z	mín	Q_1	Mediana	Media	Q_3	Máximo
2	2	14	31	43,52	60	293
3	3	77	191,5	266,2	370	1893
4	9	435,5	1013,5	1481	2120,2	11352
5	7	2888	6766	9516	13325	79673

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

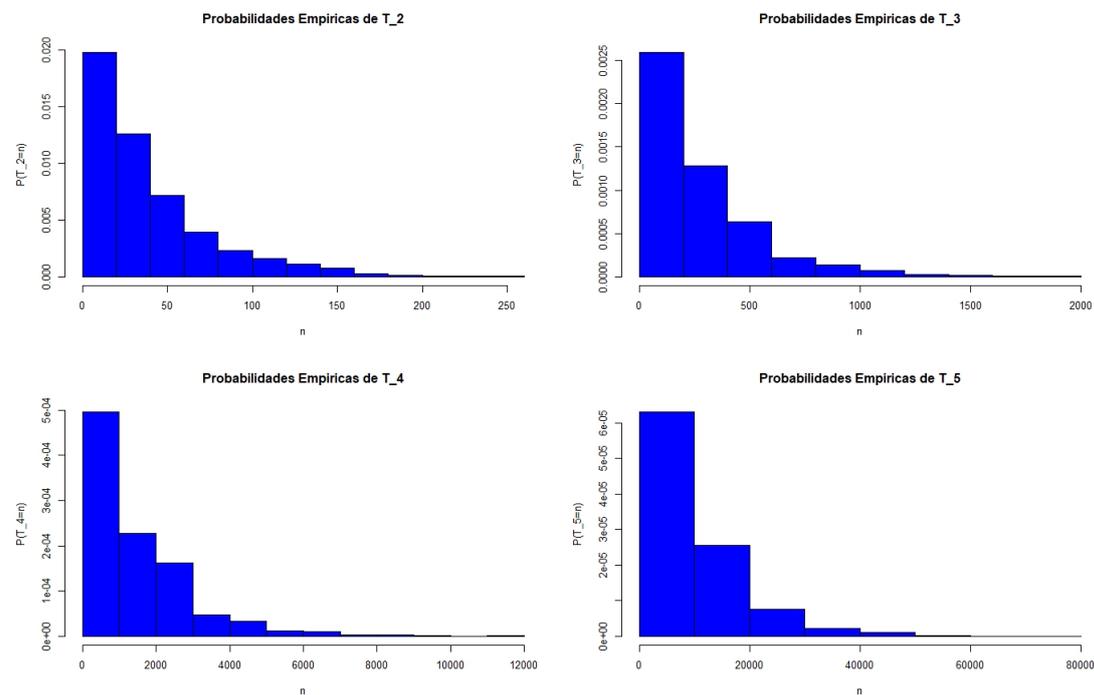


Figura 2.4: Tiempos de espera para obtener longitudes de racha $z=2,3,4,5$.

En la Figura 2.4 puede verse que, conforme crece el número de rachas, el jugador debe esperar un tiempo mayor para obtener dicha racha y que las probabilidades de dicho evento disminuyen. El rango de la variable T_z se amplía considerablemente conforme crece z dado que la probabilidad de obtener una racha cada vez más grande, va disminuyendo.

Ahora, si se desean obtener las probabilidades de transición, puede construirse la matriz de transición estimando dichas probabilidades para la simulación realizada en el caso del jugador que apuesta a 7.

Teniendo en cuenta que el apostador realiza 50 jugadas idénticas en esta simulación, y que cada X_n tiene dos resultados posibles en el siguiente paso: que su valor aumente una unidad (éxito) o ser igual a 0 (fracaso); que cada resultado es independiente uno de otro; y que, la probabilidad de aumentar en una unidad en cada paso es constante (p), puede decirse entonces que se trata de un experimento binomial y que la variable aleatoria Y que cuenta el número de veces que determinado valor de X_n aumentó en una unidad en la próxima jugada, dentro de las n jugadas, tiene distribución binomial de parámetros n y p :

$$Y \sim \mathcal{B}(n, p)$$

Se trata de buscar quién es el estimador de máxima verosimilitud de p para construir las probabilidades de transición empíricas en esta simulación.

Teorema 2. Sea Y_1, Y_2, \dots, Y_n una muestra aleatoria de una distribución Bernoulli de parámetro p . El estimador de máxima verosimilitud de p es:

$$\hat{p}_{MV} = \frac{\sum_{i=1}^n Y_i}{n}$$

Demostración. Sea Y_1, Y_2, \dots, Y_n una muestra aleatoria de una distribución Bernoulli de parámetro p . Luego, cada variable Y_i , con $i = 1, \dots, n$, tiene función de probabilidad:

$$P(y_i, p) = p^{y_i}(1 - p)^{n - y_i}$$

Luego, la función de probabilidad conjunta, por ser Y_1, \dots, Y_n una muestra aleatoria, es:

$$P(\mathbf{y}, \mathbf{p}) = \prod_{i=1}^n p^{y_i}(1 - p)^{n - y_i}$$

Como el espacio paramétrico $\Theta = (0, 1)$ es un conjunto abierto, el soporte de $p(\mathbf{y}, \mathbf{p})$ no depende de \mathbf{p} y $p(\mathbf{y}, \mathbf{p})$ tiene derivadas parciales respecto de p_i , puede utilizarse el método de máxima verosimilitud para obtener el estimador máximo verosímil de p .

Para obtener la función de log-verosimilitud, se considera el logaritmo natural de la función de probabilidad y aplicando propiedades de los logaritmos en el segundo miembro de la igualdad anterior se obtiene:

$$\begin{aligned} \ln P(\mathbf{y}, \mathbf{p}) &= \ln \left(\prod_{i=1}^n p^{y_i}(1 - p)^{n - y_i} \right) \\ &= \sum_{i=1}^n (\ln(p^{y_i}(1 - p)^{n - y_i})) \\ &= \sum_{i=1}^n (\ln(p^{y_i}) + \ln((1 - p)^{n - y_i})) \\ &= \sum_{i=1}^n (y_i \ln(p) + (n - y_i) \ln(1 - p)) \end{aligned} \tag{2.29}$$

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

Para obtener el máximo de la función de log-verosimilitud se procede a obtener su derivada respecto a p , para analizar la existencia de valores críticos:

$$\frac{\delta \ln P(\mathbf{y}, \mathbf{p})}{\delta p} = \sum_{i=1}^n \left(\frac{y_i}{p} - \frac{n - y_i}{1 - p} \right) \quad (2.30)$$

Entonces:

$$\begin{aligned} \frac{\delta \ln P(\mathbf{y}, \mathbf{p})}{\delta p} = 0 &\Rightarrow \sum_{i=1}^n \left(\frac{y_i}{p} - \frac{n - y_i}{1 - p} \right) = 0 \\ &\Rightarrow \sum_{i=1}^n \frac{y_i(1 - p) - (n - y_i)p}{p(1 - p)} = 0 \\ &\Rightarrow \sum_{i=1}^n \frac{y_i - y_i p - np + y_i p}{p(1 - p)} = 0 \\ &\Rightarrow \sum_{i=1}^n y_i - \sum_{i=1}^n np = 0 \\ &\Rightarrow p = \frac{\sum_{i=1}^n y_i}{n} \end{aligned} \quad (2.31)$$

Si se calcula la derivada segunda respecto a p_i de la función de log-verosimilitud se obtiene:

$$\frac{\delta^2 \ln P(\mathbf{y}, \mathbf{p})}{\delta p^2} = \sum_{i=1}^n \left(-\frac{y_i}{p^2} - \frac{n - y_i}{(1 - p)^2} \right)$$

que resulta ser negativa para cualquier valor de $p \in (0, 1)$.

Por lo tanto, por el criterio de existencia de extremos relativos de una función:

$$\hat{p}_{MV} = \frac{\sum_{i=1}^n Y_i}{n}$$

■

El resultado de este teorema afirma que, el estimador de máxima verosimilitud de la probabilidad de transición en un paso, de un estado a otro, es la frecuencia relativa de la variable que cuenta el número de veces que aparece esa transición en la sucesión de transiciones de un proceso.

Si se denota con n_{xy} al número de transiciones en un paso en que la cadena pasa del estado x al y , y con n_{xz} al número de transiciones en un paso en que

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

la cadena pasa del estado x al z , entonces en virtud del teorema anterior, puede escribirse:

$$\hat{P}(x, y)_{MV} = \frac{n_{xy}}{\sum_{z \in \mathcal{S}} n_{xz}}$$

Es decir que, el estimador de máxima verosimilitud de $P(x, y)$ es el cociente entre: el número de veces que desde x se accede a y , en un paso, y el número de veces que desde x se accede a cualquier estado $z \in \mathcal{S}$, en un paso.

En virtud de este resultado, puede armarse la matriz de conteos de veces que la cadena de Markov pasa de un estado a otro, para la simulación del jugador que apuesta 50 veces al 7:

```
> P1=createSequenceMatrix(X);P1
  0 1 2
0 43 3 0
1  2 0 1
2  1 0 0
```

Y mediante el paquete “markovchain” puede generarse la matriz de transición la cual estima las probabilidades de transición como se mencionó anteriormente, con la frecuencia relativa y un nivel de confianza específico:

```
> fit=markovchainFit(X);fit
$estimate
MLE Fit
A 3 - dimensional discrete Markov Chain defined by the following states:
0, 1, 2
The transition matrix (by rows) is defined as follows:
      0      1      2
0 0.9347826 0.06521739 0.0000000
1 0.6666667 0.00000000 0.3333333
2 1.0000000 0.00000000 0.0000000
```

De esta manera puede verse que, las estimaciones de: la probabilidad de empezar a jugar sin dinero y permanecer en dicho estado es $P(0, 0) = 0,93$; la probabilidad de empezar sin dinero y ganar la primera jugada es $P(0, 1) = 0,065$; la de perder, habiendo ganado una jugada, es $P(1, 0) = 0,667$; la de ganar habiendo ganado una jugada, es $P(1, 2) = 0,333$ y, la de perder habiendo ganado dos jugadas,

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

$P(2, 0) = 1$. Se observa que las probabilidades nulas son aquellas correspondientes a eventos imposibles. Los resultados de este experimento se alejan bastante de las probabilidades teóricas del modelo.

Si el número de jugadas aumenta a $n = 10000$:

```
> N<-10000 #Numero de jugadas
> p<-1/6 #Probabilidad de Éxito
> I<-0
> X<-rep(I,N+1)
> for(i in 1:N){
+   X[i+1]<-sample(x=c(X[i]+1,0),size=1,prob=c(p,1-p))
+ }
> P1=createSequenceMatrix(X);P1
      0   1   2   3  4  5
0 6900 1406   0   0  0  0
1 1165   0 241   0  0  0
2  204   0   0 37  0  0
3  28    0   0  0 9  0
4   8    0   0  0  0  1
5   1    0   0  0  0  0
> fit=markovchainFit(X);fit
$estimate
MLE Fit
A 6 - dimensional discrete Markov Chain defined by the following states:
0, 1, 2, 3, 4, 5
The transition matrix (by rows) is defined as follows:
      0   1   2   3   4   5
0 0.8307248 0.1692752 0.0000000 0.000000 0.0000000 0.0000000
1 0.8285917 0.0000000 0.1714083 0.000000 0.0000000 0.0000000
2 0.8464730 0.0000000 0.0000000 0.153527 0.0000000 0.0000000
3 0.7567568 0.0000000 0.0000000 0.000000 0.2432432 0.0000000
4 0.8888889 0.0000000 0.0000000 0.000000 0.0000000 0.1111111
5 1.0000000 0.0000000 0.0000000 0.000000 0.0000000 0.0000000
```

Puede verse que, conforme crece el valor de n , las probabilidades empíricas se aproximan cada vez más a las teóricas.

La matriz de transición de las probabilidades teóricas, suponiendo que 5 es la máxima longitud de rachas que puede alcanzarse, es:

Cadena de Markov, 10000 apuestas al 7

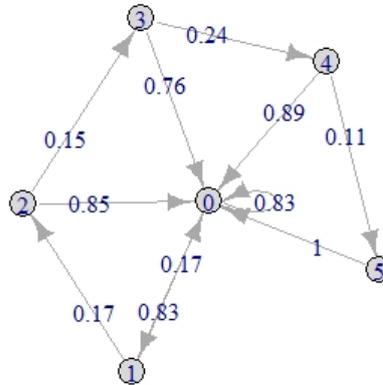


Figura 2.5: Gráfico Cadena de Markov. Caso: 10000 apuestas a 7.

$$P_T = \begin{pmatrix} 5/6 & 1/6 & 0 & 0 & 0 & 0 \\ 5/6 & 0 & 1/6 & 0 & 0 & 0 \\ 5/6 & 0 & 0 & 1/6 & 0 & 0 \\ 5/6 & 0 & 0 & 0 & 1/6 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

También puede representarse a una cadena de Markov mediante un grafo. En los vértices de dicho grafo se representan los estados del espacio de estados de la cadena y en las aristas “orientadas” se señala la probabilidad de transición de un estado a otro.

Para la cadena de Markov simulada para 10000 jugadas, se obtiene el gráfico que se muestra en la figura 2.5.

```
> plot(mc, main="Cadena de Markov, 10000 apuestas al 7")
```

Si se quieren obtener las estimaciones de las probabilidades de transición en más pasos, con el software es sencillo calcular las potencias n -ésimas de la matriz de transición para la simulación:

$$P^3 = \begin{pmatrix} 0,8309 & 0,1406 & 0,0241 & 0,0045 & 0 & 0 \\ 0,8285 & 0,1411 & 0,0240 & 0,0064 & 0 & 0 \\ 0,8277 & 0,1387 & 0,0246 & 0,0091 & 0 & 0 \\ 0,8261 & 0,1376 & 0,0220 & 0,0144 & 0 & 0 \\ 0,8311 & 0,1417 & 0,0258 & 0 & 0,0014 & 0 \\ 0,8304 & 0,1406 & 0,0290 & 0 & 0 & 0 \end{pmatrix}$$

De esta manera, por ejemplo, se estima que la probabilidad de que el jugador no tenga dinero y que siga sin dinero luego de tres jugadas es $P^{(3)}(0,0) = 0,5733$. La probabilidad de que el jugador pase de tener \$2 a tener \$3 en tres jugadas se estima que es $P^{(3)}(2,3) = 0,0091$.

Si se compara con las probabilidades teóricas, en el primer caso debe calcularse: $P_T^{(3)}(0,0)$. Para ello puede ocurrir que comenzando en el estado 0 permanezca en 0 en las tres jugadas siguientes; o que avance al estado 1 en un paso, regrese a 0 en el siguiente y permanezca allí en la tercera jugada; que llegue a 1 en dos jugadas y en la tercera regrese a 0; o que llegue a 2 en dos jugadas y en la tercera regrese a 0, es decir:

$$\begin{aligned} P_T^{(3)}(0,0) &= \left(\frac{5}{6}\right)^3 + \frac{1}{6} \cdot \left(\frac{5}{6}\right)^2 \cdot 2 + \left(\frac{1}{6}\right)^2 \cdot \frac{5}{6} \\ &= 0,8333 \end{aligned} \tag{2.32}$$

En el segundo caso debe calcularse: $P_T^{(3)}(2,3)$. Pero es imposible que el jugador pase del estado 2 al 3 en tres jugadas, puesto que del estado 2 puede avanzar al 3 o al 0, en la próxima jugada. Si llega al estado 0 en la primer jugada, es imposible que en las dos jugadas siguientes llegue al estado 3, puesto que requeriría un mínimo de tres jugadas. Si llega al estado 3 en la primer jugada, en la segunda puede pasar al estado 4 o regresar a 0, lo que implicaría el caso anterior. Luego:

$$P_T^{(3)}(2,3) = 0$$

2.1.6. Ejemplo: cadena de Ehrenfest

Este modelo, propuesto inicialmente por Paul y Tatiana Ehrenfest, representa una descripción matemática simplificada del proceso de difusión de calor, gases o líquidos a través de una membrana.

El modelo consiste en dos cajas etiquetadas A y B , y d bolas etiquetadas $1, 2, \dots, d$. Inicialmente, algunas de estas bolas están en la caja A y el resto en la caja B . Se selecciona al azar un número entero entre $1, 2, \dots, d$ y la bola etiquetada

por ese número entero se retira de su caja y se coloca en su caja opuesta. El procedimiento se repite indefinidamente y las selecciones son independientes de un ensayo a otro.

Sea X_n : “Número de bolas en la caja A después de la n -ésima transición”. Entonces $\{X_n\}_{n \in \mathbb{N}_0}$ es una cadena de Markov con espacio de estados:

$$\mathcal{S} = \{0, 1, 2, \dots, d\}$$

La función de transición P para esta cadena puede determinarse observando que si en la caja A hay x bolas en el instante n , entonces con probabilidad x/d la bola será extraída en el instante $(n + 1)$ de la caja A para ser transferida a la B , quedando en la caja A , $x - 1$ bolas en dicho instante.

Análogamente, con probabilidad $(d - x)/d$, la bola extraída en el instante $n + 1$ será de la caja B y será transferida a la caja A en el instante $n + 1$.

Así, la función de transición de esta cadena de Markov está dada por:

$$P(x, y) = \begin{cases} \frac{x}{d} & \text{si } y = x - 1, \\ \frac{d-x}{d} & \text{si } y = x + 1. \end{cases} \quad (2.33)$$

A continuación, se trabajará con un proceso de los diferentes que pueden plantearse, adaptando el modelo de Ehrenfest utilizando tres cajas. Este planteo es de elaboración propia y, en lo que sigue, será utilizado para ejemplificar el marco teórico trabajado.

Se consideran tres cajas etiquetadas con A , B y C , una al lado de la otra en el orden indicado. Se cuenta con d bolas enumeradas con $1, 2, \dots, d$ repartidas en las tres cajas. En cada transición se sortea un número entre $1, 2, \dots, d$, se extrae la bola de la caja en donde esté y se considera el siguiente criterio para su reubicación:

- Si la bola es extraída de la caja A y en la caja B hay un número menor de bolas que en A , la bola se pasa a la caja B .
- Si la bola es extraída de la caja A y en la caja B hay un número mayor o igual de bolas que en A , la bola se queda en la caja A .
- Si la bola es extraída de la caja C y en la caja B hay un número menor de bolas que en C , la bola se pasa a la caja B .
- Si la bola es extraída de la caja C y en la caja B hay un número mayor o igual de bolas que en C , la bola se queda en la caja C .
- Si la bola es extraída de la caja B y en la caja A o C hay menor número de bolas que en la caja B , la bola pasa a la que tenga menor número de bolas. En caso de que A y C tengan el mismo número de bolas, la bola se queda en B .

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

- Si la bola es extraída de B y en las cajas A y C hay un número mayor o igual de bolas que en la caja B , la bola se queda en la caja B .

Sea X_n : “Número de bolas en la caja B después de la n -ésima transición”. Entonces $\{X_n\}_{n \in \mathbb{N}_0}$ es una cadena de Markov cuyo espacio de estados es el conjunto $\mathcal{S} = \{0, 1, 2, \dots, d\}$.

En una transición, el número de bolas en la caja B , sólo puede disminuir en una unidad, mantener el número de bolas o incrementarse en una unidad. Es por esto, que se trata de averiguar $P(b, b-1)$, $P(b, b)$, y $P(b, b+1)$. En los restantes casos, dicha probabilidad será 0.

Para calcular $P(b, b-1)$, debe tenerse en cuenta que para que de B se quite una bola, la misma debe pertenecer a la caja B y, además, en A o C debe haber un número menor de bolas que en B , excepto el caso en que A y C tengan el mismo número de bolas.

- Para que la bola a extraer esté en la caja B , se tiene una probabilidad de $\frac{b}{d}$.
- Sabiendo que en B hay b bolas, puede ocurrir que:
 - en A haya $0, 1, \dots, b-1$ bolas, mientras que b sea menor o igual al entero más próximo del cociente entre d y 3.
 - Análogamente sucede con la cantidad de bolas en C .
 - Es decir que, dada esta condición, se tienen $2b$ posibilidades de que en A o C el número de bolas sea menor al de B y, que A y C no tengan el mismo número de bolas.
 - Si b es mayor que el entero más próximo al cociente entre d y 3, se tendrán $d-b+1$ casos en que b sea mayor al número de bolas de A o de C , excepto en el caso en que $d-b$ sea par, el cual se corresponde a la situación en que A y C tengan el mismo número de bolas.
 - Por su parte, si B tiene b bolas, los casos posibles de distribución de las d bolas en las tres urnas, es de $d-b+1$.

De esta manera:

$$P(b, b-1) = \begin{cases} \frac{2b}{d-b+1} \cdot \frac{b}{d} & \text{si } b \leq q_3(d), \\ \frac{d-b+I_{\mathbb{N}_0}(d-b+1)}{d-b+1} \cdot \frac{b}{d} & \text{si } b > q_3(d). \end{cases} \quad (2.34)$$

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

siendo $q_3(d)$ la función que redondea al entero más próximo, el cociente de la división de d por 3, e $I_{\mathbb{N}_0}(d - b + 1)$ una función indicadora que vale 1 si $d - b + 1$ es impar y 0 si $d - b + 1$ es par.

Para calcular $P(b, b)$ deben tenerse en cuenta los siguientes casos, mutuamente excluyentes:

- i) que la bola sorteada esté en la caja B y que en A y C el número de bolas sea mayor o igual que el de B . O bien, que salga sorteada de B y que en A y C haya el mismo número de bolas.
- ii) que la bola sorteada esté en A o en C y el número de bolas en B sea mayor o igual que el de A o C , respectivamente.

Para el caso i), la probabilidad de que la bola salga sorteada de B , es $\frac{b}{d}$. Y para que se de la segunda condición, se tienen en cuenta dos casos:

- Si b es menor o igual que la parte entera de $d/3$: de los $d + 1 - b$ casos posibles, deben restarse los $2b$ casos en que el número de bolas en A o C sea menor que el de B .
- Si b es mayor que la parte entera de $d/3$, no habrá casos en que el número de bolas de A y C sea mayor que B , excepto en el que la bola sorteada salga de B y, en A y C haya el mismo número de bolas, situación que se da cuando $d - b$ es par.

Para el caso ii) deben tenerse en cuenta los siguientes casos:

- b es menor o igual que la parte entera de $d/2$ y la bola sorteada sale de la caja A . Como en A puede haber $1, 2, \dots, b$ bolas, se tendrá la siguiente probabilidad para el total de los casos entre los $(d - b + 1)$ posibles:

$$\frac{1}{d - b + 1} \sum_{i=1}^b \frac{i}{d}$$

cuyo desarrollo es:

$$\frac{b(b + 1)}{2d(d - b + 1)}$$

- b es mayor que la parte entera de $d/2$ y la bola sorteada sale de la caja A . Como en A puede haber $1, 2, \dots, d - b$ bolas, se tendrá la siguiente probabilidad para el total de los casos entre los $d - b + 1$ posibles:

$$\frac{1}{d - b + 1} \sum_{i=1}^{d-b} \frac{i}{d}$$

cuyo desarrollo es:

$$\frac{d-b}{2d}$$

- Análogamente a los dos casos anteriores, en el caso en que la bola salga sorteada de C .

Teniendo en cuenta estos análisis, puede concluirse que:

$$P(b, b) = \begin{cases} \frac{b(d-2b+2)}{d(d-b+1)} & \text{si } b \leq \left\lceil \frac{d}{3} \right\rceil, \\ \frac{b \cdot (b+1 + I_{\mathbb{N}_0}(d-b))}{d(d-b+1)} & \text{si } \left\lceil \frac{d}{3} \right\rceil < b \leq \left\lfloor \frac{d}{2} \right\rfloor, \\ \frac{b \cdot I_{\mathbb{N}_0}(d-b) + (d-b)(d-b+1)}{d(d-b+1)} & \text{si } b > \left\lfloor \frac{d}{2} \right\rfloor \end{cases} \quad (2.35)$$

Para calcular $P(b, b+1)$, debe tenerse en cuenta que la bola no puede salir sorteada de la caja B , puesto que en ella se incrementa el número de bolas en una unidad. Es por esto, que dicha bola, debe salir de la caja A o la C , para luego poder ser colocada en la B .

Para que esto ocurra, en la caja A o en la C debe haber mayor número de bolas que en la caja B .

Se considera el caso de que la bola sorteada salga de la caja A , puesto que si sale de C , el caso es análogo. La caja A tendrá más bolas que la caja B , mientras b sea menor o igual que la parte entera de $\frac{d-1}{2}$, pues de lo contrario, puede ocurrir que tenga el mismo o menor número de bolas que en B , lo cual no permitiría el incremento del número de bolas en B .

En este caso, en la caja A puede haber $b+1, b+2, \dots, d-b$ bolas y, por consiguiente, la probabilidad para el total de los casos entre los $(d-b+1)$ posibles será:

$$\frac{1}{d-b+1} \sum_{i=b+1}^{d-b} \frac{i}{d}$$

cuyo desarrollo es:

$$\frac{(d-b)(d-b+1) - b(b+1)}{2d(d-b+1)}$$

Teniendo en cuenta que, el caso es análogo si la bola es extraída de la caja C , puede concluirse que:

$$P(b, b+1) = \begin{cases} \frac{(d-b)(d-b+1) - b(b+1)}{d(d-b+1)} & \text{si } b \leq \left\lfloor \frac{d-1}{2} \right\rfloor, \\ 0 & \text{si } b > \left\lfloor \frac{d-1}{2} \right\rfloor. \end{cases} \quad (2.36)$$

En conclusión, la función de transición de esta cadena de Markov está dada por:

$$P(x, y) = \begin{cases} P(b, b-1) & \text{si } y = x - 1, \\ P(b, b) & \text{si } y = x, \\ P(b, b+1) & \text{si } y = x + 1, \\ 0 & \text{en caso contrario.} \end{cases} \quad (2.37)$$

Se quiere buscar la matriz de transición de probabilidades para el caso en que se tengan 10 bolas enumeradas con $1, 2, \dots, 10$, repartidas aleatoriamente en tres cajas según el modelo planteado anteriormente. Para ello, se analizan las probabilidades de transición.

Para la primera fila de la matriz de transición P deben determinarse las probabilidades $P(0, y)$ con $y = 0, 1, \dots, 10$:

Casos en que B no tenga bolas											
Caja A	10	9	8	7	6	5	4	3	2	1	0
Caja B	0	0	0	0	0	0	0	0	0	0	0
Caja C	0	1	2	3	4	5	6	7	8	9	10

- Es imposible que la bola haya sido extraída de la caja B que en este caso tiene 0 bolas. Y si saliera de A o de C , pasaría a B puesto que tiene 0 bolas. Luego

$$P(0, 0) = 0$$

- Como se mencionó en el ítem anterior, al estar vacía la caja B , la bola sorteada en el ensayo será de la caja A o C en cuyo caso pasará a la caja B . Luego,

$$P(0, 1) = 1$$

- $P(0, y) = 0$ cualquiera sea $y \geq 2$, ya que en una transición, el número de bolas en B puede disminuir o aumentar en una unidad o permanecer igual.

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

Para la segunda fila de la matriz de transición P deben determinarse las probabilidades $P(1, y)$ con $y = 0, 1, \dots, 10$:

Casos en que B tenga 1 bola										
Caja A	9	8	7	6	5	4	3	2	1	0
Caja B	1	1	1	1	1	1	1	1	1	1
Caja C	0	1	2	3	4	5	6	7	8	9

- Para determinar la probabilidad $P(1, 0)$, la bola extraída debe ser de la caja B para pasar a alguna de las cajas A o C , solo si en alguna de estas no hay bolas. Luego:

$$P(1, 0) = \frac{1}{10} \cdot \frac{2}{10} = \frac{2}{100}$$

- Para determinar $P(1, 1)$ se consideran los casos en que la bola sorteada esté en la caja B y, en A o en C haya 1 bola (lo cual implica que en la otra de las dos el número de bolas supera al de B), que la cantidad de bolas de A y C superen a la de B , o que la bola sorteada salga de A o de C , siendo esta la única bola en dichas cajas. Entonces:

$$P(1, 1) = \frac{1}{10} \cdot \frac{2}{10} + \frac{1}{10} \cdot \frac{6}{10} + \frac{1}{10} \cdot \frac{2}{10} = \frac{10}{100}$$

- Para el caso en que B tenga 1 bola, y se quiera obtener la probabilidad de que en la próxima transición tenga 2 bolas, se consideran los casos en que la bola sale sorteada de A o C y en alguna de ellas la cantidad de bolas supere a la de B .

$$P(1, 2) = \left(\frac{9}{10} + \frac{8}{10} + \frac{7}{10} + \frac{6}{10} + \frac{5}{10} + \frac{4}{10} + \frac{3}{10} + \frac{2}{10} \right) \cdot \frac{1}{10} \cdot 2 = \frac{88}{100}$$

- $P(1, y) = 0$ cualquiera sea $y \geq 3$, ya que en una transición, el número de bolas en B puede disminuir o aumentar en una unidad o permanecer igual.

Para la tercera fila de la matriz de transición P deben determinarse las probabilidades $P(2, y)$ con $y = 0, 1, \dots, 10$:

Casos en que B tenga 2 bolas									
Caja A	8	7	6	5	4	3	2	1	0
Caja B	2	2	2	2	2	2	2	2	2
Caja C	0	1	2	3	4	5	6	7	8

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

- Para calcular la probabilidad $P(2, 1)$, la bola extraída debe ser de la caja B para pasar a alguna de las cajas A o C , solo si en alguna de estas hay 1 bola o ninguna. Luego:

$$P(2, 1) = \frac{2}{10} \cdot \frac{4}{9} = \frac{8}{90}$$

- Para determinar $P(2, 2)$, se consideran los casos en que la bola sorteada esté en la caja B y, en A y en C haya 2 bolas o más; o que salga sorteada de A o de C y, que en dicha caja, haya 2 bolas o menos. Entonces:

$$P(2, 2) = \frac{2}{10} \cdot \frac{5}{9} + \frac{1}{10} \cdot \frac{2}{9} + \frac{2}{10} \cdot \frac{2}{9} = \frac{16}{90}$$

- Para el caso en que B tenga 2 bolas, y se quiere obtener la probabilidad de que en la próxima transición tenga 3 bolas, se consideran los casos en que la bola sale sorteada de A o C y, en alguna de ellas, la cantidad de bolas supere a la de B .

$$P(2, 3) = \left(\frac{8}{10} + \frac{7}{10} + \frac{6}{10} + \frac{5}{10} + \frac{4}{10} + \frac{3}{10} \right) \cdot \frac{1}{9} \cdot 2 = \frac{66}{90}$$

- $P(2, y) = 0$ cualquiera sea $y \geq 4$ o $y = 0$, ya que en una transición, el número de bolas en B puede disminuir o aumentar en una unidad o permanecer igual.

Para la cuarta fila de la matriz de transición P deben determinarse las probabilidades $P(3, y)$ con $y = 0, 1, \dots, 10$:

Casos en que B tenga 3 bolas								
Caja A	7	6	5	4	3	2	1	0
Caja B	3	3	3	3	3	3	3	3
Caja C	0	1	2	3	4	5	6	7

- Para calcular la probabilidad $P(3, 2)$, la bola extraída debe ser de la caja B para pasar a alguna de las cajas A o C , solo si en alguna de estas hay 2 bolas o menos. Luego:

$$P(3, 2) = \frac{3}{10} \cdot \frac{6}{8} = \frac{18}{80}$$

- Para determinar $P(3, 3)$ se consideran los casos en que la bola sorteada esté en la caja B y, en A y en C haya 3 bolas o más; o que salga sorteada de A o de C y, que en dicha caja, haya 3 bolas o menos. Entonces:

$$P(3, 3) = \frac{3}{10} \cdot \frac{2}{8} + \left(\frac{1}{10} + \frac{2}{10} + \frac{3}{10} \right) \cdot \frac{1}{8} \cdot 2 = \frac{18}{80}$$

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

- Para el caso en que B tenga 3 bolas, y se quiera obtener la probabilidad de que en la próxima transición tenga 4 bolas, se consideran los casos en que la bola sale sorteada de A o C y, en alguna de ellas, la cantidad de bolas supere a la de B .

$$P(3, 4) = \left(\frac{7}{10} + \frac{6}{10} + \frac{5}{10} + \frac{4}{10} \right) \cdot \frac{1}{8} \cdot 2 = \frac{44}{80}$$

- $P(3, y) = 0$ cualquiera sea $y \neq 2, 3, 4$, ya que en una transición, el número de bolas en B puede disminuir o aumentar en una unidad o permanecer igual.

Para la quinta fila de la matriz de transición P deben determinarse las probabilidades $P(4, y)$ con $y = 0, 1, \dots, 10$:

Casos en que B tenga 4 bolas							
Caja A	6	5	4	3	2	1	0
Caja B	4	4	4	4	4	4	4
Caja C	0	1	2	3	4	5	6

- Para calcular la probabilidad $P(4, 3)$, la bola extraída debe ser de la caja B para pasar a alguna de las cajas A o C , solo si en alguna de estas hay 2 bolas o menos (porque en caso de que alguna de estas tenga 3 bolas, la otra también). Luego:

$$P(4, 3) = \frac{4}{10} \cdot \frac{6}{7} = \frac{24}{70}$$

- Para determinar $P(4, 4)$, se consideran los casos en que la bola sorteada esté en la caja B y, en A y en C haya 3 bolas; o que salga sorteada de A o de C y, que en dicha caja haya 4 bolas o menos. Entonces:

$$P(4, 4) = \frac{4}{10} \cdot \frac{1}{7} + \left(\frac{4}{10} + \frac{3}{10} + \frac{2}{10} + \frac{1}{10} \right) \cdot \frac{1}{7} \cdot 2 = \frac{24}{70}$$

- Para el caso en que B tenga 4 bolas, y se quiera obtener la probabilidad de que en la próxima transición tenga 5 bolas, se consideran los casos en que la bola sale sorteada de A o C y, en alguna de ellas la cantidad de bolas supere a la de B .

$$P(4, 5) = \left(\frac{6}{10} + \frac{5}{10} \right) \cdot \frac{1}{7} \cdot 2 = \frac{22}{70}$$

- $P(4, y) = 0$ cualquiera sea $y \neq 3, 4, 5$, ya que en una transición, el número de bolas en B puede disminuir o aumentar en una unidad o permanecer igual.

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

Para la sexta fila de la matriz de transición P deben determinarse las probabilidades $P(5, y)$ con $y = 0, 1, \dots, 10$:

Casos en que B tenga 5 bolas						
Caja A	5	4	3	2	1	0
Caja B	5	5	5	5	5	5
Caja C	0	1	2	3	4	5

- Para calcular la probabilidad $P(5, 4)$, la bola extraída debe ser de la caja B para pasar a alguna de las cajas A o C , solo si en alguna de estas hay 4 bolas o menos. Luego:

$$P(5, 4) = \frac{5}{10} \cdot \frac{6}{6} = \frac{30}{60}$$

- Para determinar $P(5, 5)$, se consideran los casos en que la bola sorteada esté en la caja B y, en A o en C haya 5 bolas o más (lo cual es imposible, puesto que tiene salida a alguna de esas dos cajas); o que salga sorteada de A o de C y, que en dicha caja, haya 5 bolas o menos. Entonces:

$$P(5, 5) = \frac{5}{10} \cdot \frac{0}{6} + \left(\frac{5}{10} + \frac{4}{10} + \frac{3}{10} + \frac{2}{10} + \frac{1}{10} \right) \cdot \frac{1}{6} \cdot 2 = \frac{30}{60}$$

- Para el caso en que B tenga 5 bolas, y se quiera obtener la probabilidad de que en la próxima transición tenga 6 bolas, se consideran los casos en que la bola sale sorteada de A o C y, en alguna de ellas, la cantidad de bolas supere a la de B , lo cual no es posible según los casos analizados en el cuadro. Luego, $P(5, 6) = 0$.
- $P(5, y) = 0$ cualquiera sea $y \neq 4, 5, 6$, ya que en una transición, el número de bolas en B puede disminuir o aumentar en una unidad o permanecer igual.

Para la séptima fila de la matriz de transición P deben determinarse las probabilidades $P(6, y)$ con $y = 0, 1, \dots, 10$:

Casos en que B tenga 6 bolas						
Caja A	4	3	2	1	0	
Caja B	6	6	6	6	6	
Caja C	0	1	2	3	4	

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

- Para calcular la probabilidad $P(6, 5)$, la bola extraída debe ser de la caja B para pasar a alguna de las cajas A o C (excepto el caso en que A y C tengan dos bolas cada una. Luego:

$$P(6, 5) = \frac{6}{10} \cdot \frac{4}{5} = \frac{24}{50}$$

- Para determinar $P(6, 6)$ se consideran los casos en que la bola sorteada esté en la caja B y, en A y en C haya 2 bolas en cada una; o que salga sorteada de A o de C . Entonces:

$$P(6, 6) = \frac{6}{10} \cdot \frac{1}{5} + \left(\frac{4}{10} + \frac{3}{10} + \frac{2}{10} + \frac{1}{10} \right) \cdot \frac{1}{5} \cdot 2 = \frac{26}{50}$$

- Para el caso en que B tenga 6 bolas, no es posible que en una transición pase a tener 7, puesto que la bola debería salir sorteada de A o de C y el número de bolas en estas cajas es menor que el de B . Luego, $P(6, 7) = 0$.
- $P(6, y) = 0$ cualquiera sea $y \neq 5, 6, 7$, ya que en una transición, el número de bolas en B puede disminuir o aumentar en una unidad o permanecer igual.

Para la octava fila de la matriz de transición P deben determinarse las probabilidades $P(7, y)$ con $y = 0, 1, \dots, 10$:

Casos en que B tenga 7 bolas				
Caja A	3	2	1	0
Caja B	7	7	7	7
Caja C	0	1	2	3

- Para calcular la probabilidad $P(7, 6)$, la bola extraída debe ser de la caja B para pasar a alguna de las cajas A o C . Luego:

$$P(7, 6) = \frac{7}{10} = \frac{28}{40}$$

- Para determinar $P(7, 7)$, se consideran los casos en que la bola sorteada salga de A o de C . Entonces:

$$P(7, 7) = \left(\frac{3}{10} + \frac{2}{10} + \frac{1}{10} \right) \cdot \frac{1}{4} \cdot 2 = \frac{12}{40}$$

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

- Para el caso en que B tenga 7 bolas, no es posible que en una transición pase a tener 8, puesto que la bola debería salir sorteada de A o de C y, el número de bolas en estas cajas, es menor que el de B . Luego, $P(7, 8) = 0$.
- $P(7, y) = 0$ cualquiera sea $y \neq 6, 7, 8$, ya que en una transición, el número de bolas en B puede disminuir o aumentar en una unidad o permanecer igual.

Para la novena fila de la matriz de transición P deben determinarse las probabilidades $P(8, y)$ con $y = 0, 1, \dots, 10$:

Casos en que B tenga 8 bolas			
Caja A	2	1	0
Caja B	8	8	8
Caja C	0	1	2

- Para calcular la probabilidad $P(8, 7)$, la bola extraída debe ser de la caja B para pasar a alguna de las cajas A o C (excepto que A y C tengan una bola cada una. Luego:

$$P(8, 7) = \frac{8}{10} \cdot \frac{2}{3} = \frac{16}{30}$$

- Para determinar $P(8, 8)$ se consideran los casos en que la bola sorteada salga de B y en A haya 1 bola al igual que en C ; o que salga sorteada de A o de C . Entonces:

$$P(8, 8) = \frac{8}{10} \cdot \frac{1}{3} + \left(\frac{2}{10} + \frac{1}{10} \right) \cdot \frac{1}{3} \cdot 2 = \frac{14}{30}$$

- Para el caso en que B tenga 8 bolas, no es posible que en una transición pase a tener 9, puesto que la bola debería salir sorteada de A o de C y, el número de bolas en estas cajas, es menor que el de B . Luego, $P(8, 9) = 0$.
- $P(8, y) = 0$ cualquiera sea $y \neq 7, 8, 9$, ya que en una transición, el número de bolas en B puede disminuir o aumentar en una unidad o permanecer igual.

Para la décima fila de la matriz de transición P deben determinarse las probabilidades $P(9, y)$ con $y = 0, 1, \dots, 10$:

Casos en que B tenga 9 bolas		
Caja A	1	0
Caja B	9	9
Caja C	0	1

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

- Para calcular la probabilidad $P(9, 8)$, la bola extraída debe ser de la caja B para pasar a alguna de las cajas A o C . Luego:

$$P(9, 8) = \frac{9}{10} = \frac{18}{20}$$

- Para determinar $P(9, 9)$, se consideran los casos en que la bola sorteada salga de A o de C . Entonces:

$$P(9, 9) = \frac{1}{10} \cdot \frac{1}{2} \cdot 2 = \frac{2}{20}$$

- Para el caso en que B tenga 9 bolas, no es posible que en una transición pase a tener 10, puesto que la bola debería salir sorteada de A o de C y, el número de bolas en estas cajas, es menor que el de B . Luego, $P(9, 10) = 0$.
- $P(8, y) = 0$ cualquiera sea $y \neq 7, 8, 9$, ya que en una transición, el número de bolas en B puede disminuir o aumentar en una unidad o permanecer igual.

Para la undécima fila de la matriz de transición P deben determinarse las probabilidades $P(10, y)$ con $y = 0, 1, \dots, 10$. Aquí todas las bolas están en B , por lo que saldrá sorteada de esta caja y la bola permanecerá allí indefinidamente puesto que, al tener igual número de bolas (ninguna) A y C , no podrá salir de la caja B . Luego, $P(10, 10) = 1$ y $P(10, y) = 0$ cualquiera sea $y < 10$.

Con estos resultados, los cuales coinciden con las probabilidades determinadas para el caso de n bolas, la matriz de transición para esta cadena de Markov queda:

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{100} & \frac{10}{100} & \frac{88}{100} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{8}{90} & \frac{16}{90} & \frac{66}{90} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{18}{80} & \frac{90}{80} & \frac{44}{80} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{24}{70} & \frac{80}{70} & \frac{24}{70} & \frac{22}{70} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{30}{60} & \frac{30}{60} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{24}{50} & \frac{26}{50} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{38}{40} & \frac{12}{40} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{40}{30} & \frac{14}{30} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{30}{20} & \frac{18}{20} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

En la siguiente figura se puede observar el diagrama correspondiente a esta cadena:

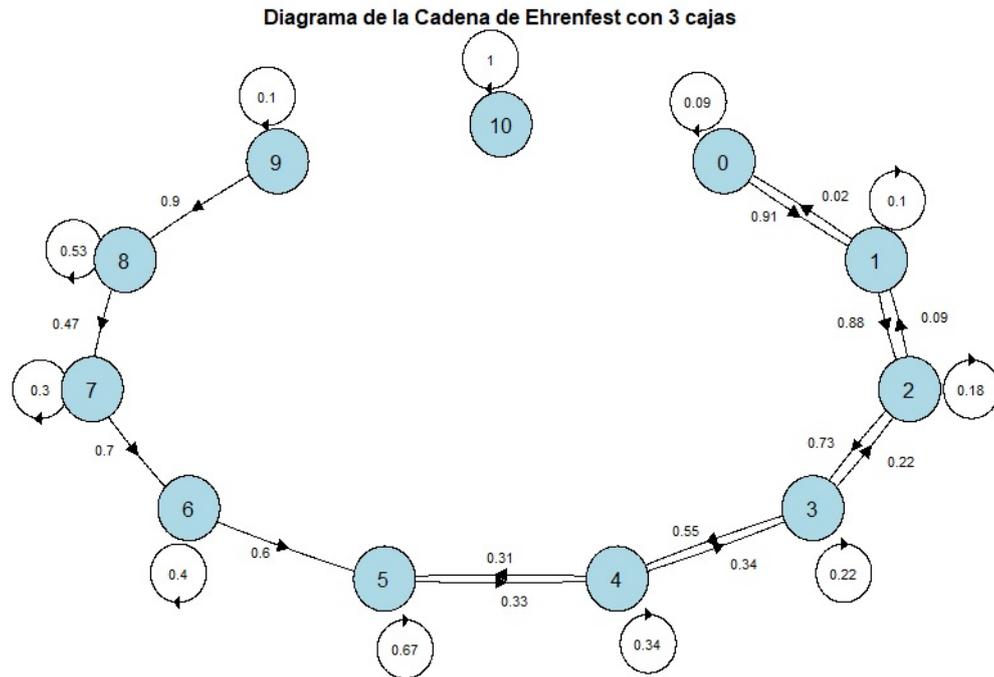


Figura 2.6: Diagrama de la cadena de Ehrenfest con 3 cajas y 10 bolas.

Como puede verse en el gráfico, si inicialmente la caja B tiene 0 bolas, en una transición, con seguridad tendrá 1 bola. En los siguientes pasos, se aprecia mayor probabilidad de que aumente en una unidad el número de bolas en cada transición, hasta llegar a 4 bolas, instante en el cual la probabilidad de permanecer en ese estado y de disminuir una bola es la misma.

Cuando la caja B tiene 5 bolas, tiene la misma probabilidad de permanecer en ese estado que la de perder una bola. Sin embargo, a partir de este estado, la probabilidad es nula de que la caja aumente su número de bolas.

En caso de que la caja B contenga 6, 7, 8 o 9 bolas, se aprecia que es más probable que en un paso, pierda una bola. Se destaca aún mayor probabilidad de que la caja tienda a contener un número par de bolas. Este hecho se debe a que si el número de bolas es par, y la cantidad de bolas en la caja B es par, puede ocurrir que el número de bolas en A y en C coincidan, con lo cual de B ya no podrían extraerse bolas.

También se observa el caso en que B tiene 10 bolas, estado al cual no puede accederse desde otro estado, y del cual no puede salirse.

El siguiente código en R, muestra una simulación para esta cadena:

```
> N = 10 #Número de Bolas
> N_B = sample(0:N,1);N_B #Número de bolas en la caja B
```

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

```
[1] 0
> N_A = sample(0:(N-N_B),1);N_A #Número de bolas en la caja A
[1] 1
> N_C = N-N_B-N_A; N_C #Número de Bolas en la Caja C
[1] 9

# Función de Transición
> Ehrenfest = function(x) {
+   if (x == "A" && N_B<N_A) {
+     N_B <<- N_B+1
+     N_A <<- N_A-1
+   } else {
+     if (x == "C" && N_B<N_C) {
+       N_B <<- N_B+1
+       N_C <<- N_C-1
+     } else {
+       if (x == "B" && N_A<N_B && N_A<N_C) {
+         N_A <<- N_A+1
+         N_B <<- N_B-1
+       } else {
+         if ( x == "B" && N_C<N_B && N_C<N_A) {
+           N_C <<- N_C+1
+           N_B <<- N_B-1
+         } else {
+           N_A <<- N_A
+           N_B <<- N_B
+           N_C <<- N_C
+         }
+       }
+     }
+   }
+   return(N_B)
+ }

> numrep = 30 #Número de Repeticiones

> X_n = rep(1,numrep) #Transiciones

> for (i in 1:numrep){
+   X_n[i] = Ehrenfest(sample(c("A","B","C"),
```

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

```
+           size = 1,  
+           prob = c(N_A/N,N_B/N,N_C/N))  
+ }  
  
X_n  
[1] 2 3 4 5 5 4 3 2 3 2 2 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4  
  
# Gráfico  
> plot(X_n,main = "Simulación Cadena de Ehrenfest",  
+       xlab = "n",  
+       ylab = expression("X"[n]),  
+       type = "b",  
+       col = "darkviolet",  
+       ylim = c(0,10)  
+       )
```

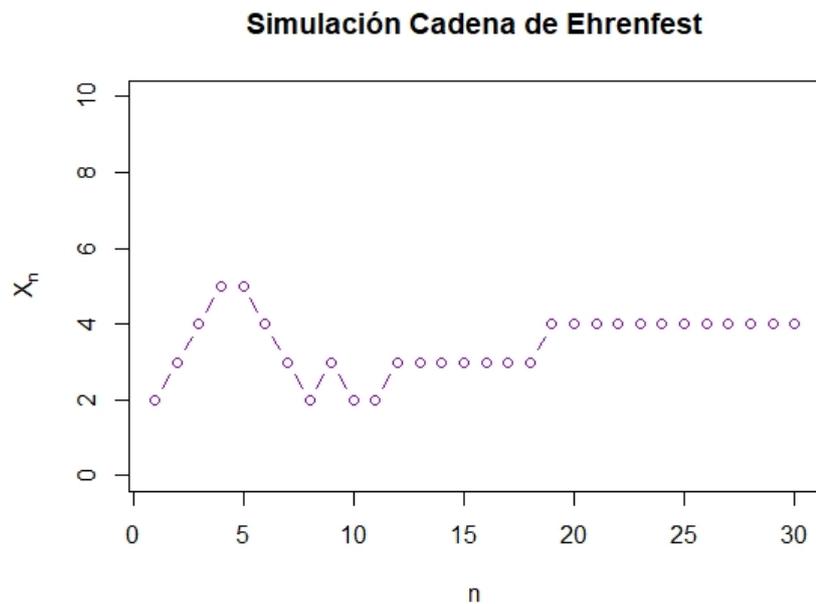


Figura 2.7: Valores de X_n estimados con simulación en función del número de transiciones n . Caso: cadena de Ehrenfest con 3 cajas.

2.1.7. Clasificación de los estados

En este apartado se examina la clasificación de los estados en una cadena de Markov genérica. En primer lugar, se distinguen dos tipos de estados según si la cadena puede o no regresar con certeza a un estado específico partiendo del mismo. Luego, se realiza una subclasificación de los estados en los que es posible regresar a sí mismos, considerando si el tiempo de llegada es finito o “infinito”.

Además, se analiza otro tipo de estados y su relación con los tipos de estados de la clasificación anterior, los cuales son aquellos en los que la cadena de Markov entra en ellos y permanecen allí indefinidamente.

Definición 2.7. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Un estado $y \in \mathcal{S}$ es llamado *recurrente* si $\rho_{yy} = 1$ y *transitorio* si $\rho_{yy} < 1$.

Esta clasificación puede interpretarse de la siguiente manera: un estado es considerado recurrente si la cadena tiene la capacidad de regresar eventualmente a ese estado con probabilidad uno y, cuando esto ocurre en algún momento finito, gracias a la propiedad de Markov, es seguro que regresa a él de manera repetitiva. Debido a este comportamiento, se le llama recurrente. Por otro lado, un estado se clasifica como transitorio si hay una probabilidad positiva de que la cadena, comenzando en ese estado, nunca regrese a él.

Definición 2.8. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Un estado $y \in \mathcal{S}$ es llamado *estado absorbente* si $P(y, y) = 1$, o de manera equivalente si $P(y, x) = 0, \forall y \neq x$.

Proposición 2.4. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Si un estado $y \in \mathcal{S}$ es absorbente entonces es recurrente.

Demostración. Como y es un estado absorbente entonces se cumple, por definición, que $P(y, y) = 1$. El primer miembro de la última igualdad puede expresarse como la probabilidad de que la cadena que comienza en y siga en y en el próximo paso, es decir: $P_y(T_y = 1) = 1$; y en consecuencia, $\rho_{yy} = 1$. Luego, por definición, resulta que y es un estado recurrente. ■

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

La implicación recíproca de la Proposición 2.4 no es cierta. En efecto, sea una cadena de Markov con espacio de estados $\mathcal{S} = \{x, y, z\}$ y matriz de transición:

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ \alpha & 1 - \alpha & 0 \end{pmatrix}$$

donde $\alpha < 1$.

El estado z es recurrente ya que $\rho_{zz} = 1$, es decir: $P_z(T_z < \infty) = 1$. Esto significa que, si la cadena comienza en z puede regresar a z en un número finito de pasos. Lo que se quiere probar es que comenzando en z regresa a z pero no en un paso. En este ejemplo, si $T_z = 2$, por la Proposición 2.2:

$$\begin{aligned} P_z(T_z = 2) &= P(z, x)P(x, z) + P(z, y)P(y, z) \\ &= \alpha \cdot 1 + (1 - \alpha) \cdot 1 \\ &= \alpha + 1 - \alpha \\ &= 1 \end{aligned} \tag{2.38}$$

Y el estado z no es absorbente porque $P(z, z) \neq 1$.

Definición 2.9. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} . Sean $y, z \in \mathcal{S}$ y sea $I_y(z)$ una función indicadora del conjunto $\{y\}$ definida por:

$$I_y(z) = \begin{cases} 1 & \text{si } z = y, \\ 0 & \text{si } z \neq y. \end{cases} \tag{2.39}$$

Se define el número de momentos en que la cadena de Markov visita el estado y como:

$$N(y) = \sum_{n=1}^{\infty} I_y(X_n)$$

Esta definición es sencilla de ver, ya que se están sumando tantos unos como momentos en que la cadena visita al estado y , puesto que $I_y(X_n) = 1$ si la cadena está en el estado y en el momento n y que $I_y(X_n) = 0$ en caso contrario.

Proposición 2.5. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Sean $x, y \in \mathcal{S}$ y $m, n \in \mathbb{N}$, la probabilidad de que la cadena de Markov que comienza en x visite el estado y por primera vez luego de m pasos y que la próxima visita ocurra n pasos después, es:

$$P_x(T_y = m)P_y(T_y = n)$$

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

Demostración. Se comienza escribiendo la probabilidad de que la cadena visite al estado y por primera vez luego de m pasos, comenzando en el estado x , y que la visite por segunda vez luego de n pasos:

$$\begin{aligned}
 & P(X_{n+m} = y, X_{n+m-1} \neq y, \dots, X_{m+1} \neq y, X_m = y, \dots, X_1 \neq y \mid X_0 = x) \\
 &= P(X_{n+m} = y, X_{n+m-1} \neq y, \dots, X_{m+1} \neq y \mid X_m = y, \dots, X_1 \neq y, X_0 = x) \\
 &\quad \cdot P(X_m = y, \dots, X_1 \neq y \mid X_0 = x) \\
 &= P(X_{n+m} = y, X_{n+m-1} \neq y, \dots, X_{m+1} \neq y \mid X_m = y) \\
 &\quad \cdot P(X_m = y, \dots, X_1 \neq y \mid X_0 = x) \\
 &= P_y(T_y = n)P_x(T_y = m) \\
 &= P_x(T_y = m)P_y(T_y = n)
 \end{aligned} \tag{2.40}$$

En las igualdades planteadas se utilizó la definición de probabilidad condicional, la propiedad de Markov y la definición de tiempos de llegada. ■

Proposición 2.6. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Sean $x, y \in \mathcal{S}$ y $m \in \mathbb{N}$, entonces se cumplen:

- i) $P_x(N(y) \geq m) = \rho_{xy}\rho_{yy}^{m-1}$
- ii) $P_x(N(y) = m) = \rho_{xy}\rho_{yy}^{m-1}(1 - \rho_{yy})$
- iii) $P_x(N(y) = 0) = 1 - \rho_{xy}$

Demostración. En primer lugar, se demuestra el ítem *i)* de la proposición y luego se lo utilizará para demostrar *ii)* y *iii)*.

i) Para demostrar esta igualdad se aplicará inducción sobre m .

- Para $m = 1$ se considera el hecho de que el evento $\{N(y) \geq 1\}$ es igual a $\{T_y < \infty\}$. De esta manera:

$$\begin{aligned}
 P_x(N(y) \geq 1) &= P_x(T_y < \infty) \\
 &= \rho_{xy} \\
 &= \rho_{xy}\rho_{yy}^{1-1}
 \end{aligned} \tag{2.41}$$

- Suponiendo válida la propiedad para $m = k$ se tiene que:

$$P_x(N(y) \geq k) = \rho_{xy}\rho_{yy}^{k-1}$$

- Para $m = k + 1$, considerando $n \in \mathbb{N}$ y la Proposición 2.5, se obtiene:

$$\begin{aligned}
 P_x(N(y) \geq k + 1) &= \sum_{m=k}^{\infty} \sum_{n=1}^{\infty} P_x(T_y = m) P_y(T_y = n) \\
 &= \left(\sum_{m=k}^{\infty} P_x(T_y = m) \right) \left(\sum_{n=1}^{\infty} P_y(T_y = n) \right) \quad (2.42) \\
 &= \rho_{xy} \rho_{yy}^{k-1} \rho_{yy} \\
 &= \rho_{xy} \rho_{yy}^k
 \end{aligned}$$

Luego, la propiedad es válida $\forall m \in \mathbb{N}$.

- ii) Si se descompone el evento de interés de tal modo que pueda aplicarse lo demostrado en *i*), se obtiene:

$$\begin{aligned}
 P_x(N(y) = m) &= P_x(N(y) \geq m) - P_x(N(y) \geq m + 1) \\
 &= \rho_{xy} \rho_{yy}^{m-1} - \rho_{xy} \rho_{yy}^m \quad (2.43) \\
 &= \rho_{xy} \rho_{yy}^{m-1} (1 - \rho_{yy})
 \end{aligned}$$

- iii) Calculando la probabilidad del complemento del evento en interés y aplicando lo demostrado en *i*):

$$\begin{aligned}
 P_x(N(y) = 0) &= 1 - P_x(N(y) \geq 1) \\
 &= 1 - \rho_{xy} \quad (2.44)
 \end{aligned}$$

■

Nota. Se denotará con $E_x(\cdot)$ a la esperanza de una variable aleatoria dado que la cadena de Markov comienza en x .

La siguiente proposición relaciona el valor esperado del número de visitas al estado y para una cadena de Markov que comienza en x , con la función de transición en n pasos de x a y .

Proposición 2.7. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Sean $x, y \in \mathcal{S}$ y $n \in \mathbb{N}$, entonces se cumple:

$$E_x(N(y)) = \sum_{n=1}^{\infty} P^{(n)}(x, y)$$

Se denotará a $E_x(N(y))$ con $G(x, y)$ e indica el valor esperado del número de visitas a y de una cadena de Markov que comienza en x .

Demostración. Utilizando la definición del número de visitas a un estado, el hecho de que la esperanza de una suma de variables aleatorias es igual a la suma de las esperanzas de dichas variables aleatorias, el cálculo de la esperanza planteada y la definición de probabilidad de transición en n pasos:

$$\begin{aligned}
 E_x(N(y)) &= E_x\left(\sum_{n=1}^{\infty} I_y(X_n)\right) \\
 &= \sum_{n=1}^{\infty} E_x(I_y(X_n)) \\
 &= \sum_{n=1}^{\infty} (1 \cdot P_x(X_n = y) + 0 \cdot P_x(X_n \neq y)) \\
 &= \sum_{n=1}^{\infty} P^{(n)}(x, y)
 \end{aligned} \tag{2.45}$$

■

El siguiente teorema brinda una relación entre la clasificación de los estados y , la probabilidad y esperanza del número de visitas a un estado.

Teorema 3. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Sean $x, y \in \mathcal{S}$, entonces se cumple:

i) Si y es un estado transitorio, entonces:

$$P_x(N(y) < \infty) = 1$$

y

$$G(x, y) = \frac{\rho_{xy}}{1 - \rho_{yy}}, \quad \forall x \in \mathcal{S}$$

donde G es finita.

ii) Si y es un estado recurrente entonces:

$$P_y(N(y) = \infty) = 1$$

y

$$G(y, y) = \infty$$

Además.

$$\begin{aligned}
 \forall x \in \mathcal{S} : P_x(N(y) = \infty) &= P_x(T_y < \infty) \\
 &= \rho_{xy}
 \end{aligned} \tag{2.46}$$

Si $\rho_{xy} = 0$ entonces $G(x, y) = 0$, mientras que si $\rho_{xy} > 0$ entonces se cumple que $G(x, y) = \infty$.

Demostración.

- i) Sea y un estado transitorio. Como $0 \leq \rho_{yy} < 1$ por definición de estado transitorio y axioma de no negatividad de la probabilidad y, como $P_x(N(y) \geq m) = \rho_{xy}\rho_{yy}^{m-1}$, $\forall m \geq 1$, se sigue que:

$$\begin{aligned} P_x(N(y) = \infty) &= \lim_{m \rightarrow \infty} P_x(N(y) \geq m) \\ &= \lim_{m \rightarrow \infty} \rho_{xy}\rho_{yy}^{m-1} \\ &= 0 \end{aligned} \tag{2.47}$$

Ahora, como ya se demostró anteriormente que

$$\forall m \geq 1 : P_x(N(y) = m) = \rho_{xy}\rho_{yy}^{m-1}(1 - \rho_{yy})$$

:

$$\begin{aligned} G(x, y) &= E_x(N(y)) \\ &= \sum_{m=1}^{\infty} m P_x(N(y) = m) \\ &= \sum_{m=1}^{\infty} m \rho_{xy}\rho_{yy}^{m-1}(1 - \rho_{yy}) \end{aligned} \tag{2.48}$$

Por otro lado, se tiene el siguiente resultado para series de potencias:

$$\sum_{m=1}^{\infty} m t^{m-1} = \frac{1}{(1-t)^2}, \quad \text{para } |t| < 1$$

y usando este resultado se obtiene:

$$\begin{aligned} G(x, y) &= \frac{\rho_{xy}(1 - \rho_{yy})}{(1 - \rho_{yy})^2} \\ &= \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty \end{aligned} \tag{2.49}$$

- ii) Si y es un estado recurrente entonces, por definición, se cumple que $\rho_{yy} = 1$ y como $\forall m \geq 1 : P_x(N(y) \geq m) = \rho_{xy}\rho_{yy}^{m-1}$ resulta:

$$\begin{aligned} P_x(N(y) = \infty) &= \lim_{m \rightarrow \infty} P_x(N(y) \geq m) \\ &= \lim_{m \rightarrow \infty} \rho_{xy}\rho_{yy}^{m-1} \\ &= \rho_{xy} \end{aligned} \tag{2.50}$$

y en particular:

$$\begin{aligned}
 P_y(N(y) = \infty) &= \lim_{m \rightarrow \infty} P_y(N(y) \geq m) \\
 &= \lim_{m \rightarrow \infty} \rho_{xy} \rho_{yy}^{m-1} \\
 &= 1
 \end{aligned} \tag{2.51}$$

Si una variable aleatoria no negativa tiene probabilidad positiva de no ser finita, su valor esperado no es finito. Luego:

$$\begin{aligned}
 G(y, y) &= E_y(N(y)) \\
 &= \infty
 \end{aligned} \tag{2.52}$$

Si $\rho_{xy} = 0$, entonces $\forall m \in \mathbb{N} : P_x(T_y = m) = 0$ y entonces, por la Proposición 2.2, se cumple que,

$$\forall n \in \mathbb{N} : P^{(n)}(x, y) = \sum_{m=1}^{\infty} \underbrace{P_x(T_y = m)}_{=0} P^{(n-m)}(y, y)$$

de donde resulta que,

$$\forall n \in \mathbb{N} : P^{(n)}(x, y) = 0$$

Por lo tanto, $G(x, y) = 0$, en este caso.

Si $\rho_{xy} > 0$, entonces $P_x(N(y) = \infty) = \rho_{xy} > 0$ y en consecuencia:

$$\begin{aligned}
 G(x, y) &= E_x(N(y)) \\
 &= \infty
 \end{aligned} \tag{2.53}$$

■

El teorema anterior describe la diferencia fundamental entre un estado transitorio y un estado recurrente. Si y es un estado transitorio, no importa donde comience la cadena de Markov, sólo realiza un número finito de visitas a y y el número esperado de visitas a y es finito.

Si y es un estado recurrente, entonces la cadena de Markov que comienza en y , regresa a y infinitas veces. Si la cadena de Markov comienza en algún otro estado x , puede resultar imposible que llegue al estado y ; sin embargo, si es posible y lo visita al menos una vez, entonces lo hace infinitamente.

Definición 2.10. Una cadena de Markov se llama *cadena transitoria* si todos los estados son transitorios y *cadena recurrente* si todos los estados son recurrentes.

Definición 2.11. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Se define

$$N_n(y) = \sum_{m=1}^n I_y(X_m)$$

como el número de visitas de la cadena de Markov al estado y durante $m = 1, \dots, n$.

El valor esperado del número de visitas para la cadena que comienza en x está dado por

$$\begin{aligned} E_x [N_n(y)] &= G_n(x, y) \\ &= \sum_{m=1}^n P^{(m)}(x, y) \end{aligned} \tag{2.54}$$

Teorema 4. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Si $y \in \mathcal{S}$ es un estado transitorio, entonces:

- i) $P \left(\lim_{n \rightarrow \infty} \frac{N_n(y)}{n} = 0 \right) = 1$;
- ii) $\lim_{n \rightarrow \infty} \frac{G_n(x, y)}{n} = 0, \quad \forall x \in \mathcal{S}$

Demostración. Sea y un estado transitorio, entonces por el Teorema 3 se ha visto que, con probabilidad 1:

$$\lim_{n \rightarrow \infty} N_n(y) < \infty$$

y en consecuencia:

$$\lim_{n \rightarrow \infty} \frac{N_n(y)}{n} = 0 \quad \text{con probabilidad 1}$$

En el mismo teorema, se estableció que cualquiera sea $x \in \mathcal{S}$:

$$\lim_{n \rightarrow \infty} G_n(x, y) = G(x, y) < \infty$$

y, por lo tanto:

$$\lim_{n \rightarrow \infty} \frac{G_n(x, y)}{n} = 0 \quad \forall x \in \mathcal{S}$$

■

Se observa que $N_n(y)/n$ es la proporción de las primeras n unidades de tiempo que la cadena está en el estado y y que $G_n(x, y)/n$ es el valor esperado de esta proporción para una cadena que comienza en x .

Sea ahora y un estado recurrente y se denota con $m_y = E_y [T_y]$ al tiempo medio de regreso a y para una cadena que comienza en y , si este tiempo de regreso tiene esperanza finita y, en caso contrario, se denotará con $m_y = \infty$.

2.1.8. Descomposición del espacio de estados

En esta sección se estudia la comunicación entre estados de una cadena de Markov. Dos estados se comunican si existe una probabilidad no nula de transitar desde uno hacia el otro en un número finito de pasos. Se analizará cómo esta relación de comunicación produce una descomposición del espacio de estados y cómo esto se relaciona con las clasificaciones de estados previamente abordadas. Para ello, se introducirá una definición rigurosa de la comunicación entre estados y se demostrarán algunas de sus propiedades clave.

Definición 2.12. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Sean $x, y \in \mathcal{S}$, se dice que desde el estado x se llega al estado y si se cumple $P^{(n)}(x, y) > 0$ para algún $n \in \mathbb{N}_0$.

En símbolos: $x \rightarrow y$.

El caso en el que desde x no se pueda llegar a y se denotará con $x \not\rightarrow y$.

Es inmediato ver que, si $x = y$ entonces $x \rightarrow y$. A este resultado se lo demostrará un poco más adelante.

En el caso en que $x \neq y$ se tiene el siguiente resultado:

Proposición 2.8. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Sean $x, y \in \mathcal{S}$, tal que $x \neq y$, entonces desde x se llega a y si y sólo si $\rho_{xy} > 0$.

Demostración. Para demostrar la implicación directa se probará su contrarrecíproca, es decir que, para cualesquiera estados diferentes x y y , si la probabilidad de que comenzando en x la cadena visite a y es nula, entonces desde x no se puede llegar a y .

$$\begin{aligned} \forall x \neq y : \rho_{xy} = 0 &\Rightarrow P_x(T_y < \infty) = 0 \\ &\Rightarrow \forall n \in \mathbb{N}_0 : P^{(n)}(x, y) = 0 \\ &\Rightarrow x \not\rightarrow y \end{aligned} \tag{2.55}$$

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

Para demostrar la implicación recíproca, se toman dos estados cualesquiera diferentes x y y , y se quiere probar que si $\rho_{xy} > 0$ entonces desde x se puede acceder a y :

$$\begin{aligned} \forall x \neq y : \rho_{xy} > 0 &\Rightarrow P_x(T_y < \infty) > 0 \\ &\Rightarrow \exists n \in \mathbb{N}_0 : P^{(n)}(x, y) > 0 \\ &\Rightarrow x \rightarrow y \end{aligned} \tag{2.56}$$

■

Definición 2.13. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Si $x, y \in \mathcal{S}$, tales que $x \rightarrow y$ y $y \rightarrow x$, se dice que x y y se comunican. En símbolos $x \leftrightarrow y$.

En consecuencia, si dos estados x y y no se comunican, es porque o la probabilidad de transición en n pasos del estado x al y es nula o bien, la probabilidad de transición en n pasos del estado y al x es nula, cualquiera sea $n \in \mathbb{N}_0$.

En símbolos:

$$x \not\leftrightarrow y \quad \Leftrightarrow \quad \forall n \in \mathbb{N}_0 : P^{(n)}(x, y) = 0 \vee P^{(n)}(y, x) = 0$$

Proposición 2.9. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . La relación de comunicación entre estados es una relación de equivalencia.

Demostración. Para demostrar que la relación de comunicación entre estados es de equivalencia debe probarse que es reflexiva, simétrica y transitiva.

i) Reflexividad:

¿Todo estado del espacio de estados de la cadena se comunica con si mismo?

$$\begin{aligned} \forall x \in \mathcal{S} : \exists n = 0 / P^{(n)}(x, x) = 1 > 0 &\Rightarrow x \rightarrow x \\ &\Rightarrow x \rightarrow x \wedge x \rightarrow x \\ &\Rightarrow x \leftrightarrow x \end{aligned} \tag{2.57}$$

Luego, “ \leftrightarrow ” es reflexiva.

ii) Simetría

¿Si un estado se comunica con otro, el segundo se comunica con el primero?

$$\begin{aligned}
 \forall x, y \in \mathcal{S} : x \leftrightarrow y &\Rightarrow x \rightarrow y \wedge y \rightarrow x \\
 &\Rightarrow \exists n \in \mathbb{N}_0 / P^{(n)}(x, y) > 0 \wedge \exists m \in \mathbb{N}_0 / P^{(m)}(y, x) > 0 \\
 &\Rightarrow \exists m \in \mathbb{N}_0 / P^{(m)}(y, x) > 0 \wedge \exists n \in \mathbb{N}_0 / P^{(n)}(x, y) > 0 \\
 &\Rightarrow y \rightarrow x \wedge x \rightarrow y \\
 &\Rightarrow y \leftrightarrow x
 \end{aligned} \tag{2.58}$$

Luego, “ \leftrightarrow ” es simétrica.

iii) Transitividad

¿Si un estado se comunica con otro y este último se comunica con un tercero, entonces el primero se comunica con el tercero?

$$\begin{aligned}
 \forall x, y, z \in \mathcal{S} : x \leftrightarrow y \wedge y \leftrightarrow z \\
 \Rightarrow \exists n \in \mathbb{N}_0 / P^{(n)}(x, y) > 0 \wedge \exists m \in \mathbb{N}_0 / P^{(m)}(y, z) > 0
 \end{aligned} \tag{2.59}$$

Entonces:

$$\begin{aligned}
 \exists n + m \in \mathbb{N}_0 / P^{(n+m)}(x, z) &= \sum_{w \in \mathcal{S}} P^{(n)}(x, w) P^{(m)}(w, z) \\
 &\geq P^{(n)}(x, y) P^{(m)}(y, z) \\
 &> 0
 \end{aligned} \tag{2.60}$$

De aquí resulta que $x \rightarrow y$ (*).

Análogamente, puede probarse que $\exists m_2 \in \mathbb{N}_0 / P^{(m_2)}(z, x) > 0$, con lo que $y \rightarrow x$ (**).

De (*) y (**) resulta que $x \leftrightarrow z$.

Luego, “ \leftrightarrow ” es transitiva.

De i), ii) y iii), resulta que “ \leftrightarrow ” es una relación de equivalencia. ■

Una consecuencia de la propiedad anterior es que por el Teorema Fundamental de las Relaciones de Equivalencia, al ser “ \leftrightarrow ” una relación de este tipo, se induce una partición en el conjunto donde dicha relación está definida. En este caso, la relación de comunicación entre estados, particiona al espacio de estados \mathcal{S} en clases de equivalencia, las cuales reciben el nombre de *clases de comunicación*.

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

Por definición de clase de equivalencia, los estados que pertenecen a una clase de comunicación son aquellos estados del espacio que se comunican entre sí.

Definición 2.14. Se dice que una cadena de Markov es *irreducible* si hay una sola clase de equivalencia, es decir si todos los elementos del espacio de estados se comunican entre sí.

Definición 2.15. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Un conjunto no vacío $C \subset \mathcal{S}$ se llama *cerrado* si y sólo si no se puede salir de él; es decir, desde ningún estado de C se puede llegar a estados que no pertenecen a C .

En símbolos:

$$\forall x \in C, y \notin C : \rho_{xy} = 0$$

O equivalentemente:

$$\forall x \in C, y \notin C, n \in \mathbb{N}_0 : P^{(n)}(x, y) = 0$$

Proposición 2.10. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} . Se cumple:

- i) Si $C \subset \mathcal{S}$ es un conjunto cerrado y la cadena de Markov comienza en C entonces, con probabilidad 1, se queda en C .
- ii) Si x es un estado absorbente entonces $\{x\}$ es un conjunto cerrado.

Demostración. Como por definición $C \neq \emptyset$, entonces existe al menos un elemento en dicho conjunto.

- i) Sea $x \in C$ y se considera la partición de \mathcal{S} en los conjuntos disjuntos de elementos que pertenecen a C y el conjunto de elementos que no pertenecen a C . Entonces como C es cerrado, se cumple:

$$\forall x \in C, \forall y \in C, \forall n \in \mathbb{N}_0 : P^{(n)}(x, y) = 0$$

Y entonces, la probabilidad del complemento será:

$$\forall x \in C, \forall y \notin C, \forall n \in \mathbb{N}_0 : P^{(n)}(x, y) = 1$$

Por lo tanto, si la cadena de Markov comienza en C , con probabilidad 1, se queda en C .

- ii) Como x es un estado absorbente entonces, por definición, se cumple que $P(x, x) = 1$ o, equivalentemente, si $x \neq y$ se tiene que $P(x, y) = 0$. Luego, $\{x\}$ es cerrado.

Puede demostrarse por inducción, que es suficiente conocer la probabilidad de transición en un solo paso para decidir si el conjunto es cerrado o no. Ya se vio que, para el caso de $n = 1$, la definición de conjunto cerrado se mantiene. Si se supone válido para $n = k$ pasos, es decir que

$$P^{(k)}(x, y) = 0, \quad \forall x \in C, \forall y \notin C$$

entonces para $n = k + 1$ pasos:

$$\begin{aligned} P^{(k+1)}(x, y) &= \sum_{z \in \mathcal{S}} P^{(k)}(x, z)P(z, y) \\ &= \sum_{z \in C} P^{(k)}(x, z)P(z, y) + \sum_{z \notin C} P^{(k)}(x, z)P(z, y) \\ &= \sum_{z \in C} 1 \cdot 0 + \sum_{z \notin C} 0 \cdot P(z, y) \\ &= 0 \end{aligned} \tag{2.61}$$

■

Definición 2.16. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} . Sea $C \subset \mathcal{S}$ un conjunto cerrado. Se dice que C es irreducible si todos sus estados se comunican.

Proposición 2.11. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Sea $C \subset \mathcal{S}$ un conjunto cerrado e irreducible de estados recurrentes. Entonces para cualesquiera $x, y \in C$ se cumple:

- i) $\rho_{xy} = 1$
- ii) $P_x(N(y) = \infty) = 1$
- iii) $G(x, y) = \infty$

Demostración. Sea $C \subset \mathcal{S}$ un conjunto cerrado e irreducible de estados recurrentes.

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

- i) Sean $x, y \in C$. Como C es cerrado e irreducible entonces, por definición, x se comunica con y . Por definición de la relación de comunicación se tiene:

$$\exists n \in \mathbb{N}_0 / P^{(n)}(x, y) = 1 \wedge P^{(n)}(y, x) = 1$$

Luego, como existe un número entero no negativo n tal que $P^{(n)}(x, y) = 1$, resulta que $\rho_{xy} = 1$.

- ii) Ya se analizó que el evento $\{N(y) = \infty\}$ es igual a $\{T_y < \infty\}$. Entonces, calculando la probabilidad de dichos eventos comenzando en x , utilizando lo demostrado en el ítem i) y el hecho de que $y \in C$ es recurrente:

$$\begin{aligned} P_x(N(y) = \infty) &= P_x(T_y < \infty) \\ &= \rho_{xy} \\ &= 1 \end{aligned} \tag{2.62}$$

- iii) Como $P_x(N(y) = \infty) = 1$ entonces el valor esperado de que la cadena visite y infinitas veces comenzando en x no es finito, es decir $G(x, y) = \infty$.

■

Teorema 5. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Sea \mathcal{S}_R el conjunto formado por todos los estados recurrentes de \mathcal{S} . Si $\mathcal{S}_R \neq \emptyset$ entonces \mathcal{S}_R es una unión finita o infinita numerable de conjuntos cerrados, disjuntos e irreducibles.

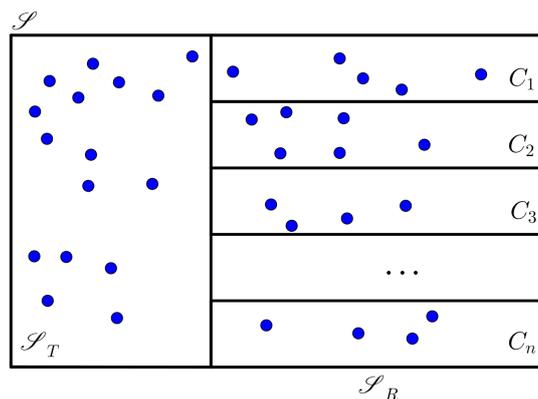


Figura 2.8: Partición del espacio de estados

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

Demostración. Sea $x \in \mathcal{S}_R$ y se define

$$C = \{y \in \mathcal{S}_R / x \longrightarrow y\}$$

Como x es recurrente, por definición se cumple que, $\rho_{xx} = 1$, entonces $x \longrightarrow x$ y luego $x \in C$

¿ C es cerrado e irreducible?

Sea $y \in C$ y $y \longrightarrow z$. Como y es recurrente por pertenecer a C , ya se probó anteriormente que si un estado es recurrente y desde él se puede acceder a otro estado, éste último también es recurrente. Luego z es recurrente.

Como $x \longrightarrow y$ (ya que $y \in C$) y $y \longrightarrow z$, entonces, por transitividad de “ \longrightarrow ”, se cumple que $x \longrightarrow z$. En consecuencia, $z \in C$, por definición de C .

Esto demuestra que C es cerrado ya que, tomando cualquier estado al que se puede acceder desde un estado de C , el primero también pertenece a C .

Sean ahora $y, z \in C$. Por ser x recurrente y $x \longrightarrow y$, entonces $y \longrightarrow x$ puesto que, $\rho_{xx} = 1$. Como $y \longrightarrow x$ y $x \longrightarrow z$, entonces $y \longrightarrow z$ y por lo tanto, C es irreducible.

Se consideran ahora C y D dos conjuntos cerrados e irreducibles de \mathcal{S}_R . Dichos conjuntos pueden ser disjuntos o no. Si son disjuntos el teorema queda demostrado. Si $C \cap D \neq \emptyset$, entonces $\exists x / x \in C \cap D$ lo cual implica, por definición de intersección de conjuntos, que $x \in C$ y $x \in D$. Si $y \in C$, entonces $x \longrightarrow y$ porque $x \in C$ y C es irreducible. Como D es cerrado, $x \in D$ y $x \longrightarrow y$, entonces $y \in D$. Por lo tanto, todo estado de C pertenece también a D , es decir $C \subset D$. Análogamente puede probarse que $D \subset C$ y, por definición de igualdad de conjuntos resulta, $C = D$.

Luego, \mathcal{S}_R es una unión finita o infinita numerable de conjuntos cerrados, disjuntos e irreducibles. ■

A continuación, se enuncia la Ley Fuerte de los Grandes Números y el Teorema de Convergencia Acotada, los cuales se necesitarán para demostrar un teorema.

Teorema 6. Ley Fuerte de los Grandes Números. Sea ξ_1, \dots, ξ_n una sucesión de variables aleatorias independientes e idénticamente distribuidas. Si estas variables tienen media finita μ , entonces

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \xi_i = \mu$$

con probabilidad 1.

Teorema 7. Teorema de Convergencia Acotada. Sea ξ_1, \dots, ξ_n una sucesión de variables aleatorias. Si existe una constante K tal que $\forall i = 1, \dots, n : |\xi_i| < K$ y si $\xi_i \rightarrow \xi$ cuando $i \rightarrow \infty$, entonces $E(\xi_i) \rightarrow E(\xi)$

Ahora sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con probabilidades de transición estacionarias que comienza en un estado recurrente y . Con probabilidad 1, la cadena regresa a y infinitas veces. Para $r \in \mathbb{N}$, sea T_y^r el instante de la r -ésima visita a y :

$$T_y^r = \min\{n \in \mathbb{N} / N_n(y) = r\}$$

Sea $W_y^1 = T_y^1 = T_y$ y para $r \geq 2$, sea $W_y^r = T_y^r - T_y^{r-1}$, el tiempo de espera entre la $(r-1)$ -ésima visita a y y la r -ésima visita. Puede observarse que:

$$T_y^r = W_y^1 + \dots + W_y^r$$

Lema 1. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con probabilidades de transición estacionarias, que comienza en un estado recurrente y . Las variables aleatorias W_y^r , $r \in \mathbb{N}$, son independientes e idénticamente distribuidas.

Demostración. En primer lugar, se define $t_0 = 0$ y para $1 \leq i \leq r+1$ se define $t_i = t_{i-1} + z_i = \sum_{q=1}^i z_q$. Sea, además,

$$A_r = \{t_1, t_2, \dots, t_r\}$$

el conjunto de los instantes en los cuales el proceso realiza las primeras r visitas al estado y . Por lo tanto, para $1 \leq s \leq t_r$, $X_s = y$, si $s \in A_r$ y, $X_s \neq y$, si $s \notin A_r$

Usando esta notación, puede verse que el evento

$$\{W_y^1 = z_1, W_y^2 = z_2, \dots, W_y^r = z_r\}$$

se puede escribir como:

$$\{X_s = y \text{ para } s \in A_r, X_s \neq y \text{ para } s \notin A_r, 1 \leq s \leq t_r\}$$

Por lo tanto,

$$P_y(W_y^{r+1} = z_{r+1} \mid W_y^1 = z_1, \dots, W_y^r = z_r)$$

puede escribirse como:

$$\begin{aligned} P_y(X_{t_{r+1}} = y, X_s \neq y \text{ para } t_r + 1 \leq s < t_{r+1} \mid X_s = y \text{ para } s \in A_r, \\ X_s \neq y \text{ para } s \notin A_r, 1 \leq s \leq t_r) \end{aligned} \quad (2.63)$$

Por la propiedad de Markov se obtiene:

$$P_y(X_{t_{r+1}} = y, X_s \neq y \text{ para } t_r + 1 \leq s < t_{r+1} \mid X_{t_r} = y)$$

y, teniendo en cuenta la definición de los t_i y la homogeneidad de la cadena, esto es igual a

$$P_y(W_j^1 = z_{r+1})$$

De manera que, se demostró:

$$P_y(W_j^{r+1} = z_{r+1} \mid W_y^1 = z_1, \dots, W_y^r = z_r) = P_y(W_j^1 = z_{r+1})$$

Para ver que todas las variables aleatorias tienen la misma distribución, se analiza la distribución de W_y^2 :

$$\begin{aligned} P_y(W_y^2 = z) &= \sum_{s=1}^{\infty} P_y(W_y^2 = z \mid W_y^1 = s) P_y(W_y^1 = s) \\ &= \sum_{s=1}^{\infty} P_y(W_y^1 = z) P_y(W_y^1 = s) \\ &= P_y(W_y^1 = z) \end{aligned} \tag{2.64}$$

En esta deducción se utilizó el teorema de la probabilidad total y el resultado que se encontró al comienzo de la demostración. Luego, procediendo análogamente para el resto de las variables W_y^r , se deduce que tienen la misma distribución.

Para analizar si las variables aleatorias son independientes, se realiza inducción sobre r .

- Para $r = 1$ es trivial.
- Suponiendo válida la independencia para $r = k$ se tiene que:

$$P_y(W_y^1, \dots, W_y^k) = P_y(W_y^1) \dots P_y(W_y^k)$$

Para $r = k + 1$ puede aplicarse la definición de probabilidad condicional, la hipótesis inductiva y el hecho de que las variables W_y^r tienen la misma distribución para todo $r \in \mathbb{N}$. De esta manera:

$$\begin{aligned} P_y(W_y^1, \dots, W_y^k, W_y^{k+1}) &= P_y(W_y^{k+1} \mid W_y^1, \dots, W_y^k) P_y(W_y^1, \dots, W_y^k) \\ &= P(W_y^1) \cdot P_y(W_y^1) \dots P_y(W_y^k) \\ &= P(W_y^{k+1}) \cdot P_y(W_y^1) \dots P_y(W_y^k) \\ &= P_y(W_y^1) \dots P_y(W_y^k) P(W_y^{k+1}) \end{aligned} \tag{2.65}$$

Por lo tanto, W_y^r , $r \in \mathbb{N}$ son independientes. ■

Teorema 8. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Sea $y \in \mathcal{S}$ un estado recurrente, entonces:

i) $P \left(\lim_{n \rightarrow \infty} \frac{N_n(y)}{n} = \frac{I_{\{T_y < \infty\}}}{m_y} \right) = 1$

ii) $\forall x \in \mathcal{S}$:

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{G_n(x, y)}{n} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P^{(m)}(x, y) \\ &= \frac{\rho_{xy}}{m_y} \end{aligned} \tag{2.66}$$

Demostración. Si se supone que la cadena de Markov comienza en el estado recurrente y entonces, por definición, dicha cadena regresa a y infinitas veces. Por el Lema 1, las variables W_y^1, W_y^2, \dots son independientes e idénticamente distribuidas. Entonces, tienen la misma media $E_y[W_y^1] = E_y[T_y] = m_y$. Ahora, por la Ley Fuerte de los Grandes Números enunciada con anterioridad:

$$P \left(\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k W_y^i = m_y \right) = 1$$

Y, como se había definido a la variable aleatoria $T_y^k = \sum_{i=1}^k W_y^i$.

$$P \left(\lim_{k \rightarrow \infty} \frac{T_y^k}{k} = m_y \right) = 1$$

Si $N_n(y) = k$, entonces al instante n la cadena ha realizado exactamente k visitas al estado y . Por lo tanto, la visita k -ésima ocurre en el instante n o antes, mientras que la visita $(k+1)$ -ésima ocurre después de n . Es decir:

$$T_y^{N_n(y)} \leq n < T_y^{N_n(y)+1}$$

y multiplicando en todos los miembros de la desigualdad por el inverso multiplicativo de $N_n(y) \geq 1$:

$$\frac{T_y^{N_n(y)}}{N_n(y)} \leq \frac{n}{N_n(y)} < \frac{T_y^{N_n(y)+1}}{N_n(y)}$$

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

Tomando límite para $n \rightarrow \infty$, se sabe que $N_n(y) \rightarrow \infty$ con probabilidad 1, entonces, por lo demostrado anteriormente y aplicando el Teorema del Sándwich, se deduce:

$$P \left(\lim_{n \rightarrow \infty} \frac{n}{N_n(y)} = m_y \right) = 1$$

Sea nuevamente y un estado recurrente y se supone ahora que X_0 tiene una distribución arbitraria. Puede ocurrir que la cadena nunca visite y . En caso de que lo visite, el argumento anterior es válido y se cumple el ítem i) del teorema. por lo tanto:

$$P \left(\lim_{n \rightarrow \infty} \frac{N_n(y)}{n} = \frac{I_{\{T_y < \infty\}}}{m_y} \right) = 1$$

Ahora, por definición, se sabe que $0 \leq N_n(y) \leq 1$, entonces por el Teorema de Convergencia Acotada:

$$\begin{aligned} \lim_{n \rightarrow \infty} E_x \left[\frac{N_n(y)}{n} \right] &= E_x \left[\frac{I_{T_y < \infty}}{m_y} \right] \\ &= \frac{P_x(T_y < \infty)}{m_y} \\ &= \frac{\rho_{xy}}{m_y} \end{aligned} \tag{2.67}$$

y por lo tanto, el ítem ii) de la tesis del teorema es válido. ■

En el teorema anterior se afirma que, si se considera una cadena de Markov irreducible y finita, tal que todos sus estados son recurrentes y se comunican, entonces con probabilidad 1, todos los estados serán visitados y $\rho_{xy} = 1$ cualesquiera sean $x, y \in \mathcal{S}$.

Por lo tanto, el tiempo medio que la cadena permanece en el estado y cuando n es lo suficientemente grande, es aproximadamente $1/m_y = 1/E_y [T_y]$, es decir, el inverso multiplicativo del tiempo medio de retorno.

A continuación, se analiza una clasificación de los estados recurrentes y algunos resultados útiles que relacionan esta clasificación con temas ya abordados.

Definición 2.17. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} . Un estado recurrente $y \in \mathcal{S}$ es *recurrente nulo* si $m_y = \infty$.

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

Proposición 2.12. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Si $y \in \mathcal{S}$ es un estado recurrente nulo, entonces:

$$\begin{aligned} \forall x \in \mathcal{S} : \lim_{n \rightarrow \infty} \frac{G_n(x, y)}{n} &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n P^{(m)}(x, y) \\ &= 0 \end{aligned} \quad (2.68)$$

Demostración. La prueba de este resultado es inmediata por el ítem ii) del Teorema 8, porque como y es recurrente nulo, entonces $m_y = \infty$, es decir que la cadena regresa a y infinitas veces. Luego: $\frac{\rho_{xy}}{m_y} \rightarrow 0$ cuando $m_y \rightarrow \infty$. ■

Definición 2.18. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} . Se dice que un estado $y \in \mathcal{S}$ es *recurrente positivo* si $m_y < \infty$.

Proposición 2.13. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} . Si $y \in \mathcal{S}$ es un estado recurrente positivo entonces:

$$\lim_{n \rightarrow \infty} \frac{G_n(y, y)}{n} = \frac{1}{m_y} > 0$$

Demostración. Por el Teorema 8, se tiene que:

$$\lim_{n \rightarrow \infty} \frac{G_n(y, y)}{n} = \frac{\rho_{yy}}{m_y}$$

Ahora, por ser y un estado recurrente, se cumple por definición que, $\rho_{yy} = 1$ y, como y es recurrente positivo, también se cumple por definición que, $m_y < \infty$. Por lo tanto,

$$\frac{\rho_{yy}}{m_y} = \frac{1}{m_y} > 0$$

Luego, la tesis del teorema es válida. ■

Si se tiene una cadena de Markov que comienza en un estado recurrente y , a partir del Teorema 8 puede verse que, si y es recurrente nulo entonces, con probabilidad 1, la proporción del tiempo que la cadena está en el estado y , durante las primeras n unidades de tiempo, tiende a cero cuando $n \rightarrow \infty$, mientras que si y es recurrente positivo, con probabilidad 1, esta proporción tiende al límite positivo $1/m_y$ cuando $n \rightarrow \infty$.

Teorema 9. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Si $x, y \in \mathcal{S}$, se cumple:

- i) Si desde x se puede llegar a y , pero desde y no se puede llegar a x , entonces y es transitorio.
- ii) Si x es recurrente y desde x se puede llegar a y , entonces y es recurrente y, además, $\rho_{xy} = \rho_{yx} = 1$.
- iii) Si $x \in \mathcal{S}$ es un estado recurrente positivo y desde x se puede acceder a y , entonces y es recurrente positivo.

Demostración. Sean $x, y \in \mathcal{S}$

- i) Si $x \rightarrow y$ y se define:

$$n = \text{mín}\{n/P^{(n)}(x, y) > 0\}$$

Es decir que, se puede llegar desde x hasta y en un número finito de pasos y n es el menor número de pasos para lograr esto.

Como $P^{(n)}(x, y) > 0$ por hipótesis, existe una sucesión de estados y_i , $i = 1, \dots, n-1$, tal que:

$$P(x, y_1)P(y_1, y_2) \dots P(y_{n-1}, y) > 0 \quad (*)$$

Como n es el mínimo, resulta que $y_r \neq x$ cualquiera sea r tal que $1 \leq r < n$, pues de lo contrario se tendría una sucesión más corta.

Por otro lado, analizando la probabilidad de que partiendo de x no se regrese a x , puede escribirse:

$$P_x(T_x = \infty) \geq P(x, y_1)P(y_1, y_2) \dots P(y_{n-1}, y)(1 - \rho_{xy}) \quad (**)$$

Como por hipótesis, desde y no se puede llegar a x , entonces $\rho_{yx} = 0$ y por lo tanto, $1 - \rho_{yx} = 1$.

Sustituyendo según (*) y (**) se tiene que $P_x(T_x = \infty) > 0$ y entonces, por probabilidad del complemento de un evento, se tiene que $P_x(T_x < \infty) < 1$, es decir $\rho_{xx} < 1$.

Por lo tanto, por definición, se deduce que x es un estado transitorio.

- ii) Si x es recurrente entonces, si la cadena comienza en x es imposible que el tiempo de regreso a x no esté definido. Es decir:

$$P_x(T_x = \infty) = 0$$

Suponiendo que $\rho_{yx} \neq 1$, es decir que $\rho_{yx} < 1$ (ya que $\rho_{yx} > 1$ no puede ser por tratarse de una probabilidad). Entonces en (**), como el producto de probabilidades en el segundo miembro no puede ser negativo, debiera ser 0, lo cual es un absurdo puesto que por (*) no lo es y, además $1 - \rho_{yx} > 0$, porque se había supuesto que $\rho_{yx} < 1$. El absurdo provino de suponer que $\rho_{yx} \neq 1$, por lo tanto, $\rho_{yx} = 1$ y, en consecuencia, $y \rightarrow x$.

Falta ver que y es recurrente y que $\rho_{xy} = 1$.

Como ya se demostró que $\rho_{yx} = 1$, entonces se cumple, por definición, que:

$$\exists N \in \mathbb{N}_0 / P^{(N)}(y, x) > 0$$

porque de y es seguro llegar a x .

Ahora, como se demostró en *i*) que $y \rightarrow x$, se supone que desde y se puede llegar a x en un número finito de pasos. Sea $n + k$ dicho número de pasos. Entonces, si la cadena comienza en y y llega a x en N pasos y luego regresa a x en $n + k$ pasos, se tiene:

$$\begin{aligned} P^{(N+n+k)}(y, y) &= P_y(X_{N+n+k} = y) \\ &\geq P_y(X_N = x, X_{N+n} = x, X_{N+n+k} = y) \\ &= P^{(N)}(y, x)P^{(n)}(x, x)P^{(k)}(x, y) \end{aligned} \quad (2.69)$$

Por lo tanto:

$$\begin{aligned} G(y, y) &= \sum_{m=1}^{\infty} P^{(m)}(y, y) \quad (1) \\ &\geq \sum_{m=N+n+k}^{\infty} P^{(m)}(y, y) \quad (2) \\ &= \sum_{n=1}^{\infty} P^{(N+n+k)}(y, y) \quad (3) \\ &\geq P^{(N)}(y, x)P^{(k)}(x, y) \sum_{n=1}^{\infty} P^{(n)}(x, x) \quad (4) \\ &= P^{(N)}(y, x)P^{(k)}(x, y) \underbrace{G(x, x)}_{=\infty} \quad (5) \\ &= \infty \end{aligned} \quad (2.70)$$

En (1) se aplicó definición de la función G (esperanza de que comenzando en y la cadena regrese a y). En (2) no se consideran en la suma los $N + n + k - 1$ términos anteriores y el hecho de que la suma de números no negativos es mayor que sus sumandos. En (3) se reescribe la sumatoria considerando N y k fijos. En (4) se aplica lo demostrado en 2.69 y en (5) se utiliza la hipótesis de que x es recurrente.

Por lo tanto, y es recurrente.

Ahora, como y es recurrente y $y \rightarrow x$, por lo ya demostrado, resulta que $\rho_{yx} = 1$.

- iii) Como x es un estado recurrente y desde x se puede acceder a y , por lo demostrado en *ii*), se cumple que desde y se puede acceder a x . Luego, existen enteros positivos n_1 y n_2 tales que:

$$P^{(n_1)}(y, x) > 0 \text{ y } P^{(n_2)}(x, y) > 0$$

Entonces, como de y se puede acceder a x y de x se puede acceder a y , existe un número entero positivo m tal que:

$$P^{(n_1+m+n_2)}(y, y) \geq P^{(n_1)}(y, x)P^{(m)}(x, x)P^{(n_2)}(x, y)$$

Considerando n desigualdades como la anterior donde $m = 1, 2, \dots, n$ y se suma miembro a miembro las n inecuaciones:

$$\sum_{m=1}^n P^{(n_1+m+n_2)}(y, y) \geq \sum_{m=1}^n [P^{(n_1)}(y, x)P^{(m)}(x, x)P^{(n_2)}(x, y)]$$

Multiplicando en ambos miembros de la desigualdad por el inverso multiplicativo de $n > 0$, se obtiene:

$$\frac{1}{n} \sum_{m=1}^n P^{(n_1+m+n_2)}(y, y) \geq \frac{1}{n} \sum_{m=1}^n [P^{(n_1)}(y, x)P^{(m)}(x, x)P^{(n_2)}(x, y)]$$

Ahora, por definición de valor esperado de que ocurra la visita en n pasos, las sumatorias de ambos miembros pueden expresarse:

$$\frac{1}{n} (G_{n_1+n+n_2}(y, y) - G_{n_1+n_2}(y, y)) \geq P^{(n_1)}(y, x)P^{(n_2)}(x, y)\frac{1}{n}G_n(x, x)$$

Tomando límite para $n \rightarrow \infty$ en ambos miembros:

$$\lim_{n \rightarrow \infty} \frac{1}{n} (G_{n_1+n+n_2}(y, y) - G_{n_1+n_2}(y, y)) \geq \lim_{n \rightarrow \infty} P^{(n_1)}(y, x) P^{(n_2)}(x, y) \frac{1}{n} G_n(x, x)$$

y aplicando el Teorema 8:

$$\frac{1}{m_y} \geq \frac{P^{(n_1)}(y, x) P^{(n_2)}(x, y)}{m_x} > 0$$

En consecuencia, $m_y < \infty$ y, por lo tanto, y es recurrente positivo. ■

Proposición 2.14. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y con espacio de estados \mathcal{S} . Sea C una clase de comunicación, entonces todos los elementos de C son transitorios o recurrentes nulos o recurrentes positivos.

Demostración. En primer lugar, se rescata el hecho de que todo estado $x \in \mathcal{S}$ es transitorio o recurrente (en sentido excluyente) puesto que, por axioma de la teoría de probabilidades, resulta $\rho_{xx} < 1$ o $\rho_{xx} = 1$.

Sea $x \in C$, entonces por definición de clase de comunicación, se cumple:

$$\forall y \in C : x \leftrightarrow y$$

Ahora bien, si x es recurrente y cualquiera sea el estado $y \in C$ se cumple que $x \leftrightarrow y$, se tiene en particular que $x \rightarrow y$ y por la parte *ii*) del Teorema 9 se deduce inmediatamente que y es recurrente.

En el caso en que x sea transitorio, se supone que existe algún elemento de C que no sea transitorio. Entonces, por la contrarrecíproca de la parte *i*) del Teorema 9, desde x no se puede llegar a y o desde y se puede acceder a x , lo cual es un absurdo, puesto que x se comunica con y . Este absurdo provino de suponer que existe en C un estado que no sea transitorio; por lo tanto, en este caso, todos los estados de C son transitorios.

Sea $x \in C$ un estado recurrente positivo. Al ser C un conjunto cerrado e irreducible, x se comunica con cualquier estado $y \in C$, es decir que entonces, desde x se puede acceder a y y por el Teorema 9 se tiene que y es recurrente positivo. Luego, la propiedad es válida cualquiera sea el estado $y \in C$.

Ahora, sea $x \in C$ un estado recurrente nulo y si se supone que existe $y \in C$ tal que y es recurrente positivo, al ser C un conjunto cerrado e irreducible, x se comunica con y y, en particular, desde y se puede acceder a x y por el Teorema

9 debería cumplirse que x es recurrente positivo, lo cual es una contradicción. El absurdo provino de suponer que habiendo un estado recurrente nulo en C , existía al menos un estado recurrente positivo en C . Luego, todos los estados de C son recurrentes nulos. ■

Proposición 2.15. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} . Si la cadena de Markov es irreducible entonces es una cadena transitoria o una cadena recurrente.

Demostración. Si una cadena de Markov es irreducible entonces su espacio de estados está particionado por una única clase de comunicación. Luego, por la Proposición 2.14, resulta que todos los estados de \mathcal{S} son transitorios o todos son recurrentes. Por lo tanto, la cadena de Markov es transitoria o recurrente. ■

Proposición 2.16. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Si $C \subset \mathcal{S}$ es finito y cerrado, entonces tiene al menos un estado recurrente positivo.

Demostración. Como se cumple que cualquiera sea el estado $x \in C$:

$$\sum_{y \in \mathcal{S}} P^{(m)}(x, y) = 1$$

Sumando sobre $m = 1, 2, \dots, n$ y dividiendo por n , cualquiera sea el estado $x \in \mathcal{S}$ se obtiene:

$$\sum_{y \in \mathcal{S}} \frac{G_n(x, y)}{n} = 1$$

Si C es finito y todos los estados de C fuesen transitorios o recurrentes nulos, por el Teorema 4 y la Proposición 2.12 se cumpliría:

$$\begin{aligned} 1 &= \lim_{n \rightarrow \infty} \sum_{y \in \mathcal{S}} \frac{G_n(x, y)}{n} \\ &= \sum_{y \in \mathcal{S}} \lim_{n \rightarrow \infty} \frac{G_n(x, y)}{n} \\ &= 0 \end{aligned} \tag{2.71}$$

lo cual es una contradicción. Este absurdo provino de suponer que todos los estados de C eran recurrentes nulos o transitorios. Luego, C tiene al menos un estado recurrente positivo. ■

Teorema 10. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Sea $C \subset \mathcal{S}$ finito, cerrado e irreducible. Entonces todos los estados de C son recurrentes positivos.

Demostración. Como C es finito y cerrado, existe en C al menos un estado recurrente positivo, por la Proposición 2.16. Por ser C cerrado e irreducible, todos los estados se comunican y, por el Teorema 9, todos los estados de C deben ser recurrentes positivos. ■

Corolario 2. Una cadena de Markov irreducible y con espacios de estado finito es recurrente positiva.

Demostración. Si la cadena de Markov es irreducible, entonces está formada por una única clase de comunicación, es decir que todos sus estados se comunican. Además, dicha clase de comunicación es un conjunto cerrado, pues sus estados no se comunican con otros fuera de esta clase, al ser dicha clase única. Luego, se cumple la hipótesis del Teorema 10 y entonces, todos los estados de la cadena de Markov son recurrentes positivos. En consecuencia, la cadena es recurrente positiva. ■

Corolario 3. Una cadena de Markov con un número finito de estados no tiene estados recurrentes nulos.

Demostración. Si y es un estado recurrente, entonces por el Teorema 5, y pertenece a un conjunto cerrado e irreducible C_y de estados recurrentes. Como C_y es finito, por el teorema anterior todos los estados en C_y son recurrentes positivos. Además, por el teorema citado, \mathcal{S}_R es una unión de estos conjuntos cerrados, irreducibles y disjuntos. Por lo tanto, en \mathcal{S} no hay estados recurrentes nulos. ■

Ejemplo. Jugador ambicioso. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ la cadena de Markov, con función de transición P y espacio de estados \mathcal{S} , analizada en el ejemplo del *Jugador Ambicioso*. Como hipótesis se sabía que, comenzando en cualquier estado $x \in \mathcal{S}$, la cadena visita al estado $x + 1$ con probabilidad p , con $0 < p < 1$, y visita el estado 0 con probabilidad $1 - p$.

- Sean dos estados cualesquiera $y, z \in \mathcal{S}$ tal que $y, y + 1, y + 2, \dots, y + k$ (con $y + k = z$) son los estados intermedios entre y y z . Como la cadena, comenzando en y , visita al estado $y+1$, existe $n \in \mathbb{N}$ tal que $P^{(n)}(y, y+1) > 0$. Análogamente, desde $y + 1$ se puede acceder a $y + 2$ y, luego de un número finito de pasos, se puede acceder desde $y+k-1$ a $y+k = z$. Por lo tanto, desde y se puede acceder a z . Ahora, desde z se puede acceder a 0 con probabilidad $1 - p$ y, análogamente a como se procedió anteriormente, desde 0 se puede acceder a 1 de donde, luego de un número finito de pasos, desde $y - 1$ se puede acceder a y . En consecuencia, desde z se puede acceder a y . Como desde y se puede acceder a z y desde z se puede acceder a y , entonces y se comunica con z y, al ser dos estados cualesquiera de \mathcal{S} , resulta que la cadena tiene una sola clase comunicación. En conclusión, la cadena es irreducible.
- ¿Cuál es la probabilidad de que el jugador empezando sin dinero, vuelva a quedarse sin dinero en n jugadas?

Si el jugador vuelve a quedarse sin dinero, entonces significa que en alguna de las n jugadas perdió. Se había establecido en el ejemplo, que en este caso:

$$P^{(n)}(x, z) = (1 - p)p^z$$

Entonces: $P^{(n)}(0, 0) = (1 - p)p^0 = 1 - p$.

Para el caso de las apuestas al 7, $P^{(n)}(0, 0) = \frac{5}{6}$.

- También, puede decirse que la cadena es recurrente. Esto puede justificarse porque por el hecho de que, según ítems anteriores, la cadena es irreducible, entonces es transitoria o recurrente. Basta entonces con analizar si alguno de los estados es transitorio o recurrente y para ello se considera el estado 0:

$$\begin{aligned} \rho_{00} &= P_0(T_0 < \infty) \\ &= 1 - P_0(T_0 = \infty) \end{aligned} \tag{2.72}$$

Si $T_0 = \infty$ significa que partiendo desde \$0, el jugador gana indefinidamente, entonces:

$$\begin{aligned} \rho_{00} &= 1 - \lim_{n \rightarrow \infty} p^n \\ &= 1 \end{aligned} \tag{2.73}$$

Por lo tanto, 0 es recurrente y entonces la cadena es recurrente.

Ejemplo. Cadena de Ehrenfest con tres cajas y 10 bolas. Si se considera la matriz de transición de la cadena correspondiente a este ejemplo presentado anteriormente:

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

- $\{0, 1, 2, 3, 4, 5\}$ es una clase de comunicación, puesto que todos los estados de dicho conjunto se comunican entre sí.
- Como la clase de comunicación no es \mathcal{S} , resulta que esta cadena no es irreducible.
- Los estados 6, 7, 8 y 9 son estados transitorios ya que, por el ítem *i)* del Teorema 9, se cumple que desde cualquiera de ellos no se puede acceder al siguiente, pero sí puede accederse al anterior.
- Desde ningún estado de $\{0, 1, 2, 3, 4, 5\}$ se tiene acceso a elementos fuera de él, por lo tanto, este conjunto es cerrado. Desde el conjunto unitario $\{10\}$ tampoco se tiene acceso a otros estados, por lo que también es cerrado. Los conjuntos $\{0, 1, 2, 3, 4, 5, x\}$, con $x = 6, 7, 8, 9$, también son cerrados por el mismo argumento. Pero sólo en los dos primeros mencionados al comienzo del ítem, todos sus estados se comunican, luego estos dos conjuntos cerrados son irreducibles.
- Por el teorema que dice que si $C \subset \mathcal{S}$ es un conjunto finito, cerrado e irreducible, entonces todos los estados de C son recurrentes positivos, resulta que 0, 1, 2, 3, 4, 5 y 10 son estados recurrentes positivos.
- El estado 10 es además absorbente, puesto que $P(10, 10) = 1$.

Esta información puede obtenerse usando el paquete “markovchain” de R. Si se inserta previamente la matriz de transición P y un vector con los estados de la cadena, se crea la cadena de Markov y se piden sus características:

```
> MC=new("markovchain",states=estados,transitionMatrix=P,
+       name="Cadena Ehrenfest 3 Cajas")
> summary(MC)
Cadena Markov chain that is composed by:
Closed classes:
0 1 2 3 4 5
10
Recurrent classes:
{0,1,2,3,4,5},{10}
Transient classes:
{6},{7},{8},{9}
The Markov chain is not irreducible
The absorbing states are: 10
```

Con esta salida no pueden determinarse directamente si los estados recurrentes son positivos o nulos.

2.1.9. Representación canónica de la matriz de transición

Anteriormente, se había abordado que si el espacio de estados de una cadena de Markov es finito, puede construirse una matriz cuyas entradas sean las probabilidades de transición en un paso de un estado a otro. Es posible reordenar las filas y columnas de esta matriz de transición separando los estados recurrentes de los transitorios. A su vez, los estados recurrentes pueden reordenarse agrupando aquellos que se comunican entre sí.

Según el Teorema 5, los estados recurrentes se pueden agrupar de manera única en conjuntos disjuntos, cerrados e irreducibles. Dentro de cada uno de estos conjuntos, todos los estados se comunican (por ser los conjuntos irreducibles) y son del mismo tipo, recurrentes nulos o recurrentes positivos.

Si dos estados están en distintos conjuntos cerrados, no pueden comunicarse y, por lo tanto, $\rho_{xy} = \rho_{yx} = 0$. Y si dos estados x y y están en el mismo conjunto cerrado e irreducible entonces se cumple $\rho_{xy} = \rho_{yx} = 1$.

De acuerdo con estos resultados puede reordenarse al espacio de estados \mathcal{S} y, de esta manera, las filas y columnas de la matriz de transición P , colocando los estados transitorios en último lugar, quedan dispuestas de la siguiente manera:

$$P = \begin{pmatrix} R_1 & 0 & \dots & 0 & \dots & 0 \\ 0 & R_2 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & R_r & \dots & 0 \\ S_1 & S_2 & \dots & S_r & \dots & Q \end{pmatrix}$$

donde:

- las submatrices R_i , con $i = 1, \dots, r$, están asociadas a subcadenas recurrentes cerradas;
- las submatrices S_i , con $i = 1 \dots, r$ indican las probabilidades de transición de estados transitorios hacia recurrentes; y,
- Q es la submatriz que indica las probabilidades de transición desde un estado transitorio hacia otro transitorio.

Puede abreviarse aún más, escribiendo:

$$P = \begin{pmatrix} R & 0 \\ S & Q \end{pmatrix}$$

donde:

- R es la submatriz cuadrada correspondiente a los estados recurrentes (donde R puede descomponerse nuevamente).

2.1. CADENAS DE MARKOV A TIEMPO DISCRETO

- 0 es la matriz de ceros que representa las probabilidades de ir de estados recurrentes a estados transitorios.
- Q es la matriz cuadrada que representa las probabilidades de ir de estados transitorios a estados transitorios.
- S representa la matriz de ir de estados transitorios a estados recurrentes.

Ejemplo. Cadena de Ehrenfest con 3 cajas y 10 bolas. El espacio de estados de la cadena de Ehrenfest que se viene trabajando, es:

$$\mathcal{S} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

que puede partitionarse en: $\mathcal{S}_T = \{6, 7, 8, 9\}$ y $\mathcal{S}_R = \{0, 1, 2, 3, 4, 5, 10\}$. Por su parte \mathcal{S}_R puede partitionarse en dos conjuntos cerrados e irreducibles: $C_1 = \{0, 1, 2, 3, 4, 5\}$ y $C_2 = \{10\}$.

De esta manera, la representación canónica de la matriz de transición es:

$$P = \left(\begin{array}{cccccc|cccccc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{100} & \frac{10}{100} & \frac{88}{100} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{8}{90} & \frac{16}{90} & \frac{66}{90} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{18}{80} & \frac{18}{80} & \frac{44}{80} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{24}{70} & \frac{24}{70} & \frac{22}{70} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{30}{60} & \frac{30}{60} & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & \frac{24}{50} & 0 & \frac{26}{50} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{28}{40} & \frac{12}{40} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{16}{30} & \frac{14}{30} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{18}{20} & \frac{2}{20} \end{array} \right)$$

donde:

- R_1 es la matriz cuadrada 6×6 que muestra las probabilidades de transición entre los estados de C_1 ,
- R_2 la matriz 1×1 que muestra la probabilidad de transición del estado de C_2 consigo mismo,
- S_1 es la matriz 4×6 que muestra las probabilidades de transición desde los estados de \mathcal{S}_T hacia los estados de C_1 ,
- S_2 es la matriz 4×1 que muestra las probabilidades de transición desde los estados de \mathcal{S}_T hacia el estado de C_2 y, por último,

- Q es la matriz cuadrada 4×4 que muestra las probabilidades de transición entre estados de \mathcal{S}_T .

En R puede pedirse esta representación canónica de la matriz de transición, utilizando el siguiente comando del paquete “markovchain”:

```
> canonicForm(MC)
```

2.2. Distribuciones Estacionarias

Dada una cadena de Markov a tiempo discreto, interesa conocer si existe una distribución a la que converjan los estados de la cadena y que sea independiente de la distribución inicial. En distintas aplicaciones de las cadenas de Markov, es útil averiguar la probabilidad de que la variable aleatoria X_n tome un valor x (entre d estados posibles) al cabo de un número de etapas o transiciones que se hace cada vez más grande.

Definición 2.19. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Sea $\pi(x)$, con $x \in \mathcal{S}$ una distribución de probabilidad. Si

$$\sum_{x \in \mathcal{S}} \pi(x)P(x, y) = \pi(y), \quad y \in \mathcal{S}$$

se dice que π es una *distribución estacionaria* para la cadena de Markov.

Para una cadena con espacio de estados finito, se observa que si P es la matriz de transición y π es el vector que representa la distribución estacionaria, la relación establecida en la definición anterior, puede escribirse en forma matricial como $\pi' \cdot P = \pi$, donde π es un vector columna y π' su transpuesto.

Dada una cadena de Markov, no siempre es posible encontrar una distribución estacionaria. Más adelante se establecerán condiciones para su existencia.

Proposición 2.17. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . La distribución de X_n es independiente de n si y sólo si la distribución inicial es una distribución estacionaria.

Demostración. En primer lugar se demostrará que, si la distribución inicial es una distribución estacionaria entonces la distribución de X_n es independiente de n . Realizando inducción sobre n se tiene:

- Para $n = 1$ se cumple:

$$\sum_{x \in \mathcal{S}} \pi(x)P(x, y) = \pi(y), \quad y \in \mathcal{S}$$

lo cuál es válido ya que, por hipótesis, la distribución inicial es una distribución estacionaria.

- Se supone válida la tesis para $n = k$, es decir:

$$\sum_{x \in \mathcal{S}} \pi(x)P^{(k)}(x, y) = \pi(y), \quad y \in \mathcal{S}$$

Para $n = k + 1$ se tiene que, si $y, z \in \mathcal{S}$:

$$\begin{aligned} \sum_{x \in \mathcal{S}} \pi(x)P^{(k+1)}(x, y) &= \sum_{x \in \mathcal{S}} \pi(x) \sum_{z \in \mathcal{S}} P^{(k)}(x, z)P(z, y) \quad (1) \\ &= \sum_{z \in \mathcal{S}} \left(\sum_{x \in \mathcal{S}} \pi(x)P^{(k)}(x, z) \right) P(z, y) \quad (2) \\ &= \sum_{z \in \mathcal{S}} \pi(z)P(z, y) \quad (3) \\ &= \pi(y) \quad (4) \end{aligned} \tag{2.74}$$

En (1) se aplicó la ecuación de Chapman-Kolmogorov, en (2) propiedades de sumatorias, en (3) la hipótesis inductiva y en (4) la hipótesis de que π es una distribución estacionaria.

Luego, la propiedad es válida cualquiera sea $n \in \mathbb{N}_0$. Por lo tanto, si la distribución del estado inicial X_0 es π , se tiene que $\forall n \in \mathbb{N}_0$:

$$P(X_n = y) = \pi(y), \quad \text{con } y \in \mathcal{S}$$

y, en consecuencia, la distribución de X_n es independiente de n .

Para demostrar la implicación recíproca a la demostrada, se supone que X_n es independiente de n , entonces la distribución inicial $\pi(0)$ debe satisfacer:

$$\begin{aligned} \pi_0(y) &= P(X_0 = y) \quad (1) \\ &= P(X_n = y) \quad (2) \\ &= \sum_{x \in \mathcal{S}} \pi_0(x)P(x, y) \quad (3) \end{aligned} \tag{2.75}$$

En (1) se aplicó la definición de distribución inicial, en (2) el hecho de que X_n es independiente de n y en (3) el Corolario 1.

En consecuencia, π_0 satisface la condición de la definición de distribución estacionaria. ■

2.2. DISTRIBUCIONES ESTACIONARIAS

Definición 2.20. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Si existe una distribución ν tal que

$$\forall x \in \mathcal{S} : \lim_{n \rightarrow \infty} P^{(n)}(x, y) = \nu(y), \quad \text{con } y \in \mathcal{S}$$

entonces la distribución de X_n se aproxima a ν cuando n es lo suficientemente grande, sin importar cuál es la distribución inicial de la cadena. En este caso, se dice que la cadena de Markov tiene *distribución asintótica*.

Proposición 2.18. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Suponiendo que la cadena de Markov tiene distribución asintótica y que π_0 es la distribución inicial de dicha cadena, entonces:

$$\lim_{n \rightarrow \infty} P(X_n = y) = \nu(y)$$

Además, si la cadena de Markov tiene una distribución estacionaria y una distribución asintótica, ambas coinciden.

Demostración. Suponiendo que la cadena de Markov tiene distribución asintótica y que π_0 es la distribución inicial de dicha cadena, entonces:

$$P(X_n = y) = \sum_{x \in \mathcal{S}} \pi_0(x) P^{(n)}(x, y)$$

Tomando límite en ambos miembros, para $n \rightarrow \infty$, suponiendo que estos límites son finitos para cada término de la sumatoria y que en este caso el límite de una suma de funciones es igual a la suma de los límites de sus términos, se obtiene:

$$\begin{aligned} \lim_{n \rightarrow \infty} P(X_n = y) &= \lim_{n \rightarrow \infty} \sum_{x \in \mathcal{S}} \pi_0(x) P^{(n)}(x, y) \\ &= \sum_{x \in \mathcal{S}} \pi_0(x) \lim_{n \rightarrow \infty} P^{(n)}(x, y) \\ &= \sum_{x \in \mathcal{S}} \pi_0(x) \nu(y) \end{aligned} \tag{2.76}$$

y como $\sum_{x \in \mathcal{S}} \pi_0(x) = 1$, se obtiene:

$$\lim_{n \rightarrow \infty} P(X_n = y) = \nu(y)$$

2.2. DISTRIBUCIONES ESTACIONARIAS

Ahora, si la cadena tiene distribución estacionaria π y se comienza la cadena con dicha distribución, se tiene:

$$\begin{aligned} P(X_n = y) &= \sum_{x \in \mathcal{S}} \pi(x) P^{(n)}(x, y) \\ &= \pi(y) \quad \text{con } y \in \mathcal{S} \end{aligned} \tag{2.77}$$

y de aquí se deduce que $\pi = \nu$. ■

La proposición anterior indica que, sin importar cuál sea la distribución inicial de la cadena, para valores lo suficientemente grandes de n , la distribución de X_n es aproximadamente igual a la distribución asintótica ν .

Teorema 11. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados finito \mathcal{S} y función de probabilidad P . Si para algún $x \in \mathcal{S}$ se cumple que:

$$\forall y \in \mathcal{S} : \lim_{n \rightarrow \infty} P^{(n)}(x, y) = \pi(y)$$

entonces el vector $\pi(y)$ es una distribución de probabilidad estacionaria.

Demostración. Se cumple que $\forall n \in \mathbb{N}, x, y \in \mathcal{S} : 0 \leq P^{(n)}(x, y) \leq 1$ por ser probabilidades de transición. En consecuencia, $\forall y \in \mathcal{S} : 0 \leq \pi(y) \leq 1$.

Ahora, como \mathcal{S} es finito, por hipótesis y aplicando propiedades de límites finitos, se obtiene:

$$\begin{aligned} \sum_{y \in \mathcal{S}} \pi(y) &= \sum_{y \in \mathcal{S}} \lim_{n \rightarrow \infty} P^{(n)}(x, y) \\ &= \lim_{n \rightarrow \infty} \sum_{y \in \mathcal{S}} P^{(n)}(x, y) \\ &= 1 \end{aligned} \tag{2.78}$$

Luego, π es un vector de probabilidad.

Finalmente, por la ecuación de Chapman-Kolmogorov, se tiene que $\forall y \in \mathcal{S}$:

$$\begin{aligned} \pi(y) &= \lim_{n \rightarrow \infty} P^{(n+1)}(x, y) \\ &= \lim_{n \rightarrow \infty} \sum_{z \in \mathcal{S}} P^{(n)}(x, z) P(z, y) \\ &= \sum_{z \in \mathcal{S}} \lim_{n \rightarrow \infty} P^{(n)}(x, z) P(z, y) \\ &= \sum_{z \in \mathcal{S}} \pi(z) P(z, y) \end{aligned} \tag{2.79}$$

Por lo tanto, por definición, π es una distribución estacionaria. ■

2.2. DISTRIBUCIONES ESTACIONARIAS

En el teorema anterior se observa que $\pi(z) > 0$ para algún $z \in \mathcal{S}$ pues, al ser π un vector de probabilidad, no podría ser un vector nulo y no se cumpliría el axioma de que la suma de sus componentes es igual a 1. En el caso en que \mathcal{S} es infinito, esto no se puede garantizar. A modo de ejemplo, si se considera una cadena de Markov con $\mathcal{S} = \mathcal{S}_T$ infinito, se cumple $\forall y \in \mathcal{S}$:

$$\begin{aligned} \lim_{n \rightarrow \infty} P^{(n)}(x, y) &= 0 \\ &= \pi(y) \end{aligned} \tag{2.80}$$

ya que todos los estados son transitorios. Se cumple que el vector π , en este caso, es estacionario, puesto que $0P = 0$. Sin embargo, no es un vector de probabilidad ya que la suma de sus componentes es 0.

También se observa que, la relación establecida en la hipótesis del teorema no es necesario que se cumpla para cualquier estado de \mathcal{S} sino que, se especifica que se cumpla para algún $x \in \mathcal{S}$. Si el límite existe, cualquiera sea $x \in \mathcal{S}$, y no depende del valor inicial, entonces π es también una distribución asintótica además de estacionaria y, por lo tanto, es única.

2.2.1. Existencia y Unicidad de Distribuciones Estacionarias

En este apartado se estudiarán las condiciones que debe cumplir una cadena de Markov para que exista la distribución estacionaria y sea única.

Para ello, se utilizará otra versión del Teorema de Convergencia Acotada, que se enuncia a continuación.

Teorema 12. (Teorema de Convergencia Acotada) Sea $\{a_i\}$ con $i \in \mathcal{S}$ una sucesión de números no negativos con $\sum_{i \in \mathcal{S}} a_i < \infty$ y sean $b_{i,n}$ con $i \in \mathcal{S}$ y $n \in \mathbb{N}$ tales que $|b_{i,n}| \leq 1$ para $i \in \mathcal{S}$ y $n \in \mathbb{N}$ y para todo $i \in \mathcal{S}$:

$$\lim_{n \rightarrow \infty} b_{i,n} = b_i$$

Entonces:

$$\lim_{n \rightarrow \infty} \sum_{i \in \mathcal{S}} a_i b_{i,n} = \sum_{i \in \mathcal{S}} a_i b_i$$

Proposición 2.19. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Sea π una distribución estacionaria y $m \in \mathbb{N}$, se cumple:

$$\sum_{z \in \mathcal{S}} \pi(z) \frac{G_n(z, x)}{n} = \pi(x)$$

2.2. DISTRIBUCIONES ESTACIONARIAS

Demostración. Por definición de distribución estacionaria se tiene:

$$\sum_{z \in \mathcal{S}} \pi(z) P^{(m)}(z, x) = \pi(x), \quad x \in \mathcal{S}$$

Entonces:

$$\sum_{z \in \mathcal{S}} \sum_{m=1}^n \pi(z) P^{(m)}(z, x) = \pi(x)$$

con lo cual:

$$\sum_{z \in \mathcal{S}} \pi(z) G_n(z, x) = n\pi(x)$$

Y, multiplicando en ambos miembros por el inverso multiplicativo de n :

$$\sum_{z \in \mathcal{S}} \pi(z) \frac{G_n(z, x)}{n} = \pi(x)$$

■

Teorema 13. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Sea π una distribución estacionaria. Si $x \in \mathcal{S}$ es un estado transitorio o recurrente nulo, entonces $\pi(x) = 0$.

Demostración. Si x es un estado transitorio o recurrente nulo, por la ecuación 4 y la Proposición 2.12, se cumple:

$$\lim_{n \rightarrow \infty} \frac{G_n(z, x)}{n} = 0, \quad z \in \mathcal{S}$$

Entonces, por la Proposición 2.19 y el Teorema de Convergencia Acotada:

$$\begin{aligned} \pi(x) &= \lim_{n \rightarrow \infty} \sum_{z \in \mathcal{S}} \pi(z) \frac{G_n(z, x)}{n} \\ &= 0 \end{aligned} \tag{2.81}$$

■

Una observación inmediata que puede realizarse del teorema anterior es que, si una cadena de Markov no tiene estados recurrentes positivos, no tiene distribución estacionaria.

2.2. DISTRIBUCIONES ESTACIONARIAS

Teorema 14. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Si $\{X_n\}$ es una cadena irreducible y recurrente positiva, entonces tiene una única distribución estacionaria dada por:

$$\pi(x) = \frac{1}{m_x}, \quad x \in \mathcal{S}$$

Demostración. De la hipótesis y el Teorema 8 se sigue:

$$\lim_{n \rightarrow \infty} \frac{G_n(z, x)}{n} = \frac{1}{m_x}, \quad x, z \in \mathcal{S}$$

Suponiendo que π es una distribución estacionaria, por la Proposición 2.19 y el Teorema de Convergencia Acotada:

$$\begin{aligned} \pi(x) &= \lim_{n \rightarrow \infty} \sum_{z \in \mathcal{S}} \pi(z) \frac{G_n(z, x)}{n} \\ &= \frac{1}{m_x} \sum_{z \in \mathcal{S}} \pi(z) \\ &= \frac{1}{m_x} \end{aligned} \tag{2.82}$$

Luego, si el proceso tiene distribución estacionaria, entonces está dada por la relación establecida en la tesis del teorema.

Ahora, ¿ $\pi(x)$ es una distribución estacionaria?

$\pi(x)$ es no negativa, puesto que es el inverso multiplicativo del tiempo medio de retorno, siendo este último positivo.

Por ser $P^{(m)}$ una función de probabilidad de transición, se cumple:

$$\sum_{x \in \mathcal{S}} P^{(m)}(z, x) = 1$$

Sumando sobre $m = 1, \dots, n$ y multiplicando en ambos miembros por el inverso multiplicativo de n , se obtiene:

$$\sum_{x \in \mathcal{S}} \frac{G_n(z, x)}{n} = 1, \quad z \in \mathcal{S} \tag{2.83}$$

Como \mathcal{S} es finito, tomando límite para $n \rightarrow \infty$ en ambos miembros y teniendo en cuenta la ecuación establecida al comienzo de la demostración:

$$\begin{aligned} 1 &= \lim_{n \rightarrow \infty} \sum_{x \in \mathcal{S}} \frac{G_n(z, x)}{n} \\ &= \sum_{x \in \mathcal{S}} \frac{1}{m_x} \end{aligned} \tag{2.84}$$

2.2. DISTRIBUCIONES ESTACIONARIAS

Luego, se cumple que

$$\sum_{x \in \mathcal{S}} \frac{1}{m_x} = 1$$

Por último, por la ecuación de Chapman-Kolmogorov:

$$\sum_{x \in \mathcal{S}} P^{(m)}(z, x)P(x, y) = P^{(m+1)}(z, y)$$

Sumando nuevamente sobre $m = 1, \dots, n$ y multiplicando en ambos miembros por el inverso multiplicativo de n :

$$\sum_{x \in \mathcal{S}} \frac{G_n(z, x)}{n} P(x, y) = \frac{G_{n+1}(z, y)}{n} - \frac{P(z, y)}{n}$$

Tomando límite para $n \rightarrow \infty$ se obtiene:

$$\sum_{x \in \mathcal{S}} \frac{1}{m_x} P(x, y) = \frac{1}{m_y}, \quad y \in \mathcal{S}$$

Por lo tanto, la distribución estacionaria de la cadena, con espacio de estados finito, es:

$$\pi(x) = \frac{1}{m_x}, \quad x \in \mathcal{S}$$

Cuando \mathcal{S} es infinito, la justificación anterior no es válida, ya que no se puede aplicar el Teorema de Convergencia Acotada.

Sea \mathcal{S}_1 un subconjunto de \mathcal{S} , entonces de 2.83 se tiene:

$$\sum_{x \in \mathcal{S}} \frac{G_n(z, x)}{n} \leq 1, \quad z \in \mathcal{S}$$

Tomando límite cuando n tiende a infinito:

$$\lim_{n \rightarrow \infty} \sum_{x \in \mathcal{S}} \frac{G_n(z, x)}{n} \leq \lim_{n \rightarrow \infty} 1$$

De donde se obtiene:

$$\sum_{x \in \mathcal{S}_1} \frac{1}{m_x} \leq 1, \quad \forall \mathcal{S}_1 \subset \mathcal{S}$$

lo cual a su vez implica que:

$$\sum_{x \in \mathcal{S}} \frac{1}{m_x} \leq 1$$

2.2. DISTRIBUCIONES ESTACIONARIAS

ya que de lo contrario la suma sobre algún conjunto finito de \mathcal{S} , también sería mayor que 1.

Análogamente, se concluye que si \mathcal{S}_1 es un subconjunto finito de \mathcal{S} , entonces:

$$\sum_{x \in \mathcal{S}_1} \frac{G_n(z, x)}{n} P(x, y) \leq \frac{G_{n+1}(z, y)}{n} - \frac{P(z, y)}{n}$$

lo cual implica:

$$\sum_{x \in \mathcal{S}_1} \frac{1}{m_x} P(x, y) \leq \frac{1}{m_y}$$

y entonces:

$$\sum_{x \in \mathcal{S}} \frac{1}{m_x} P(x, y) \leq \frac{1}{m_y} \quad y \in \mathcal{S}$$

Ahora se demuestra que la igualdad se cumple en la relación anterior. Suponiendo que no se cumple, se tendría que:

$$\begin{aligned} \sum_{y \in \mathcal{S}} \frac{1}{m_y} &> \sum_{y \in \mathcal{S}} \left(\sum_{x \in \mathcal{S}} \frac{1}{m_x} P(x, y) \right) \\ &= \sum_{x \in \mathcal{S}} \frac{1}{m_x} \left(\sum_{y \in \mathcal{S}} P(x, y) \right) \\ &= \sum_{x \in \mathcal{S}} \frac{1}{m_x} \end{aligned} \tag{2.85}$$

lo cual es una contradicción, que provino de suponer que la igualdad planteada no era cierta. Por lo tanto:

$$\sum_{x \in \mathcal{S}} \frac{1}{m_x} P(x, y) = \frac{1}{m_y}, \quad y \in \mathcal{S}$$

Ahora se define:

$$c = \sum_{x \in \mathcal{S}} \frac{1}{m_x} \quad y \quad \pi_0(x) = \frac{1}{cm_x}, \quad x \in \mathcal{S}$$

Se observa que π_0 es una distribución de probabilidad y es estacionaria, es decir:

$$\begin{aligned} \sum_{x \in \mathcal{S}} \pi_0(x) P(x, y) &= \frac{1}{c} \sum_{x \in \mathcal{S}} \frac{1}{m_x} P(x, y) \\ &= \frac{1}{c} \frac{1}{m_y} \\ &= \pi_0(y), \quad \forall y \in \mathcal{S} \end{aligned} \tag{2.86}$$

Entonces, $c = 1$ y con esto queda demostrada la tesis del teorema. ■

Corolario 4. Una cadena de Markov irreducible con espacio de estados finito, es recurrente positiva si y sólo si tiene una distribución estacionaria.

Demostración. Si la cadena de Markov es irreducible y recurrente positiva entonces, por el Teorema 14, tiene una distribución estacionaria. Ahora, si la cadena tiene una distribución estacionaria y es irreducible, por el Corolario 2, dicha cadena es recurrente positiva. ■

Corolario 5. Si una cadena de Markov con espacio de estados finito es irreducible, entonces tiene una única distribución estacionaria.

Demostración. Si una cadena con espacio de estados finito es irreducible, por el Corolario 2, dicha cadena es recurrente positiva y entonces, por el Teorema 13, la cadena tiene una única distribución estacionaria. ■

Corolario 6. Si una cadena de Markov con espacio de estados finito, es irreducible, recurrente positiva y con distribución estacionaria π , entonces $\forall x \in \mathcal{S}$:

$$P \left(\lim_{n \rightarrow \infty} \frac{N_n(x)}{n} = \pi(x) \right) = 1$$

Demostración. Por la hipótesis del corolario y el Teorema 14, la cadena tiene una única distribución estacionaria dada por:

$$\forall x \in \mathcal{S} : \pi(x) = \frac{1}{m_x}$$

Por el Teorema 8 se cumple que, si $x \in \mathcal{S}$ es un estado recurrente entonces:

$$P \left(\lim_{n \rightarrow \infty} \frac{N_n(x)}{n} = \frac{I_{\{T_x < \infty\}}}{m_x} \right) = 1$$

Como x es recurrente positivo entonces $T_x < \infty$, luego $I_{\{T_x < \infty\}} = 1$, por lo tanto:

$$P \left(\lim_{n \rightarrow \infty} \frac{N_n(x)}{n} = \frac{1}{m_x} \right) = 1$$

Y, por la distribución estacionaria de la cadena, establecida al comienzo de la demostración, se tiene que $\forall x \in \mathcal{S}$:

$$P \left(\lim_{n \rightarrow \infty} \frac{N_n(x)}{n} = \pi(x) \right) = 1$$

■

Proposición 2.20. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov, con función de transición P y espacio de estados \mathcal{S} . Si $P(x, y) = \alpha_y$, con $x, y \in \mathcal{S}$ y α_y constantes. Entonces la cadena tiene una única distribución estacionaria π , dada por $\pi(y) = \alpha_y$ con $y \in \mathcal{S}$.

Demostración. Sea $\pi(x)$, con $x \in \mathcal{S}$, una función de probabilidad, es decir que $\pi(x) \geq 0$, $\forall x \in \mathcal{S}$ y $\sum_{x \in \mathcal{S}} \pi(x) = 1$.

Como se cumple que:

$$\begin{aligned} \sum_{x \in \mathcal{S}} \pi(x)P(x, y) &= \sum_{x \in \mathcal{S}} \pi(x)\alpha_y \\ &= \alpha_y \sum_{x \in \mathcal{S}} \pi(x) \\ &= \alpha_y \end{aligned} \tag{2.87}$$

entonces π es una distribución estacionaria para la cadena y está dada por:

$$\pi(y) = \alpha_y, \quad y \in \mathcal{S}$$

Además, considerando que $P(x, y) = \alpha_y$, $\forall x, y \in \mathcal{S}$, entonces $P^{(n)}(x, y) = \alpha_y$, por lo tanto:

$$\begin{aligned} \lim_{n \rightarrow \infty} P^{(n)}(x, y) &= \alpha_y \\ &= \pi(y) \end{aligned} \tag{2.88}$$

con lo que $\pi(y)$, con $y \in \mathcal{S}$, es una distribución de probabilidad invariante.

A continuación, se demostrará por inducción sobre n que

$$P^{(n)}(x, y) = \alpha_y, \quad \forall n \in \mathbb{N}, y \in \mathcal{S}$$

- Para $n = 1$ se cumple ya que, por la definición de la función P , se tiene que $P(x, y) = \alpha_y$.
- Se supone válida la propiedad para $n = k$, es decir:

$$P^{(k)}(x, y) = \alpha_y$$

Para $n = k + 1$ se tiene:

$$\begin{aligned} P^{(k+1)}(x, y) &= \sum_{z \in \mathcal{S}} P(x, z)P^{(k)}(z, y) \\ &= \sum_{z \in \mathcal{S}} P(x, z)\alpha_y \\ &= \alpha_y \sum_{z \in \mathcal{S}} P(x, z) \\ &= \alpha_y \end{aligned} \tag{2.89}$$

2.2. DISTRIBUCIONES ESTACIONARIAS

En la anterior serie de igualdades se utilizó la ecuación de Chapman-Kolmogorov, la hipótesis inductiva y el hecho de que P es una función de probabilidad.

Luego, la afirmación es cierta cualquiera sea $n \in \mathbb{N}$.

Para finalizar, se demuestra que $\pi(y)$ es una distribución de probabilidad. Por un lado, se cumple que para cualquier $y \in \mathcal{S}$: $\pi(y) = \alpha_y = P(x, y) > 0$, y además:

$$\begin{aligned} \sum_{y \in \mathcal{S}} \pi(y) &= \sum_{y \in \mathcal{S}} \alpha_y \\ &= \sum_{y \in \mathcal{S}} P(x, y) \\ &= 1 \end{aligned} \tag{2.90}$$

■

Definición 2.21. Sea π una distribución de probabilidad sobre \mathcal{S} y sea $C \subset \mathcal{S}$. Se dice que π está *concentrada* en C , si $\pi(x) = 0$, siempre que $x \notin C$.

Teorema 15. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Sea C un conjunto cerrado e irreducible de estados recurrentes positivos. Entonces $\{X_n\}$ tiene una única distribución estacionaria π concentrada en C que está dada por:

$$\pi(x) = \begin{cases} \frac{1}{m_x} & \text{si } x \in C, \\ 0 & \text{si } x \notin C. \end{cases}$$

Demostración. Como C es cerrado e irreducible y todos sus estados son recurrentes positivos, entonces se cumple, por el Teorema 14, que:

$$\forall x \in C : \pi(x) = \frac{1}{m_x}$$

Como los elementos de C son todos recurrentes positivos, entonces cualquier estado que no pertenezca a C es transitorio o recurrente nulo, en cuyo caso $\forall x \notin C : \pi(x) = 0$. Luego, por definición, la distribución estacionaria está concentrada en C y se cumple la tesis del teorema. ■

Corolario 7. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Sea \mathcal{S}_{RP} el conjunto formado por todos los estados recurrentes positivos de $\{X_n\}$:

- i) Si $\mathcal{S}_{RP} = \emptyset$ entonces $\{X_n\}$ no tiene distribución estacionaria.
- ii) Si $\mathcal{S}_{RP} \neq \emptyset$ es un conjunto irreducible, entonces $\{X_n\}$ tiene una única distribución estacionaria.
- iii) Si $\mathcal{S}_{RP} \neq \emptyset$ es no irreducible, entonces $\{X_n\}$ tiene infinitas distribuciones estacionarias.

Demostración. Teniendo en cuenta los resultados establecidos anteriormente:

- i) Si $\mathcal{S}_{RP} = \emptyset$, entonces \mathcal{S} está formado por estados transitorios o recurrentes nulos, en cuyo caso ya se demostró que $\forall x \in \mathcal{S} : \pi(x) = 0$.
- ii) Es una consecuencia del Teorema 15.
- iii) Sea $\mathcal{S}_{RP} \neq \emptyset$ no irreducible y sean C_1, C_2, \dots conjuntos distintos, cerrados e irreducibles de estados recurrentes positivos. Por el teorema anterior se sabe que la cadena tiene una distribución estacionaria π_1 concentrada en C_1 , otra distribución estacionaria π_2 concentrada en C_2 y, en general, una distribución estacionaria π_n concentrada en C_n . Entonces las distribuciones π_α , definidas para $0 \leq \alpha \leq 1$, por:

$$\pi_\alpha(x) = \sum_{i=1}^n \alpha_i \pi_i(x), \quad x \in \mathcal{S} \quad \text{y} \quad \sum_i \alpha_i = 1$$

son distribuciones estacionarias distintas. $\pi_\alpha(x)$ es una combinación lineal convexa de distribuciones estacionarias.

Para demostrar que efectivamente son distribuciones estacionarias, basta ver que:

- a) $\sum_{x \in \mathcal{S}} \pi_\alpha(x) = 1$
- b) $\pi_\alpha(x) = \sum_{y \in \mathcal{S}} \pi_\alpha(y) P(y, x)$

Entonces, para demostrar a):

$$\begin{aligned}
 \sum_{x \in \mathcal{S}} \pi_\alpha(x) &= \sum_{x \in \mathcal{S}} \sum_{i=1} \alpha_i \pi_i(x) \\
 &= \sum_{i=1} \alpha_i \sum_{x \in \mathcal{S}} \pi_i(x) \\
 &= \sum_{i=1} \alpha_i \cdot 1 \\
 &= 1
 \end{aligned} \tag{2.91}$$

donde la tercera igualdad se cumple porque π_i es una distribución estacionaria para todo i .

Para demostrar b), se parte del segundo miembro, obteniendo:

$$\begin{aligned}
 \sum_{x \in \mathcal{S}} \pi_\alpha(y) P(y, x) &= \sum_{x \in \mathcal{S}} \sum_{i=1} \alpha_i \pi_i(y) P(y, x) \\
 &= \sum_{i=1} \alpha_i \sum_{x \in \mathcal{S}} \pi_i(y) P(y, x) \\
 &= \sum_{i=1} \alpha_i \pi_i(x) \\
 &= \pi_\alpha(x)
 \end{aligned} \tag{2.92}$$

Luego, b) se cumple y de esta manera $\pi_\alpha(x)$ es una distribución estacionaria. ■

Observación: es importante notar que si la cadena de Markov tiene espacio de estados finito, entonces tiene al menos un estado recurrente positivo, es decir $\mathcal{S}_{RP} \neq \emptyset$ y, en consecuencia, por el corolario anterior, tiene al menos una distribución estacionaria.

Ejemplo. Jugador Ambicioso. Se quiere analizar ahora la existencia de una distribución estacionaria o una distribución asintótica para esta cadena.

Ya se había concluido que, para $n \geq 2$ y $0 \leq z \leq n - 1$:

$$P^{(n)}(x, z) = (1 - p)p^z$$

Cuando n es lo suficientemente grande, entonces para $x \in \mathbb{N}_0$ y $z \in \mathbb{N}_0$ se obtiene:

$$\lim_{n \rightarrow \infty} P^{(n)}(x, z) = (1 - p)p^z$$

2.2. DISTRIBUCIONES ESTACIONARIAS

Este límite no depende del estado x del cual parte la cadena. También se puede asegurar, por contrarrecíproca del Teorema 13, que la cadena es recurrente positiva y que, por el Teorema 11, $\pi(z)$ es una distribución de probabilidad estacionaria dada por $\pi(z) = (1 - p)p^z$, $z \geq 0$.

Por el Teorema 14 también puede decirse que la distribución estacionaria obtenida anteriormente es única, y está dada por $\pi(z) = \frac{1}{m_z}$, por lo que:

$$\frac{1}{m_z} = (1 - p)p^z \quad \Rightarrow \quad m_z = \frac{1}{(1 - p)p^z}$$

Es decir, que el tiempo medio de retorno al estado z es el inverso multiplicativo de $(1 - p)p^z$.

Para el caso del jugador que apuesta al 7, si empieza sin dinero, el número medio de jugadas que debe realizar para quedarse sin dinero nuevamente es:

$$m_0 = \frac{1}{(1 - \frac{1}{6}) \cdot (\frac{1}{6})^0} = \frac{6}{5}$$

Si obtiene una racha de 3 jugadas, el número medio de jugadas que debe transcurrir para volver a obtener esa racha es:

$$m_3 = \frac{1}{(1 - \frac{1}{6}) \cdot (\frac{1}{6})^3} = 259,2$$

Si análogamente se calculan los otros tiempos medios de recurrencia para el modelo teórico y se los compara con los de la simulación de las 10000 apuestas al 7, puede verse que los valores obtenidos son bastante parecidos.

En R, utilizando el paquete “markovchain”, puede pedirse la distribución estacionaria de la cadena con el comando “steadyStates”, y para calcular el tiempo medio de retorno se realiza el inverso multiplicativo de las entradas del vector distribución estacionaria:

```
> DistEst=steadyStates(mc);DistEst
      0      1      2      3      4
[1,] 0.8322 0.1393 0.0245 0.0038 2e-04
> M=1/DistEst;M
      0      1      2      3      4
[1,] 1.201634 7.178751 40.81633 263.1579 5000
```

Número medio de jugadas: m_z					
Longitud de la Racha: z	0	1	2	3	4
Modelo Teórico	1,2	7,2	43,2	259,2	1555,2
Simulación 10000 jugadas	1,2016	7,1788	40,816	263,16	5000

Ejemplo. Cadena de Ehrenfest con 3 cajas y 10 bolas. Retomando este ejemplo, se había concluido que el espacio de estados \mathcal{S} está particionado de la siguiente manera: $\mathcal{S} = \mathcal{S}_T \cup C_1 \cup C_2$, donde \mathcal{S}_T es el conjunto de estados transitorios $\{6, 7, 8, 9\}$ y $C_1 = \{0, 1, 2, 3, 4, 5\}$ y $C_2 = \{10\}$ son dos conjuntos cerrados e irreducibles de estados recurrentes positivos.

Por el Teorema 15 y el Corolario 7, como la cadena tiene espacio de estados finito, puede decirse que tiene dos distribuciones estacionarias π_1 y π_2 concentradas en C_1 y C_2 , respectivamente.

Es sencillo obtener la distribución estacionaria concentrada en C_2 , puesto que este conjunto está formado por un único estado y se sabe que $m_{10} = 1$. Entonces, por el Teorema 15, se tiene que:

$$\pi_2(x) = \begin{cases} 1, & \text{si } x = 10 \\ 0, & \text{si } x \neq 10 \end{cases}$$

Luego,

$$\pi_2 = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1)$$

Si en R se utiliza el comando “steadyStates”, se obtienen las dos distribuciones estacionarias para esta cadena y los tiempos medios de retorno en cada caso:

```
> DistEst = steadyStates(MC);DistEst #Distribuciones Estacionarias
#pi_2 y pi_1, respectivamente
      0      1      2      3      4      5 6 7 8 9 10
[1,] 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0 0 0 0 1
[2,] 0.00015 0.00784 0.07765 0.25310 0.40602 0.25522 0 0 0 0 0

> M = 1/DistEst;M #Tiempos medios de retorno
      0      1      2      3      4      5      6      7      8      9      10
[1,]   Inf   1
[2,] 6374.17 127.48 12.88 3.95 2.46 3.92 Inf Inf Inf Inf Inf
```

De donde se puede ver fácilmente que:

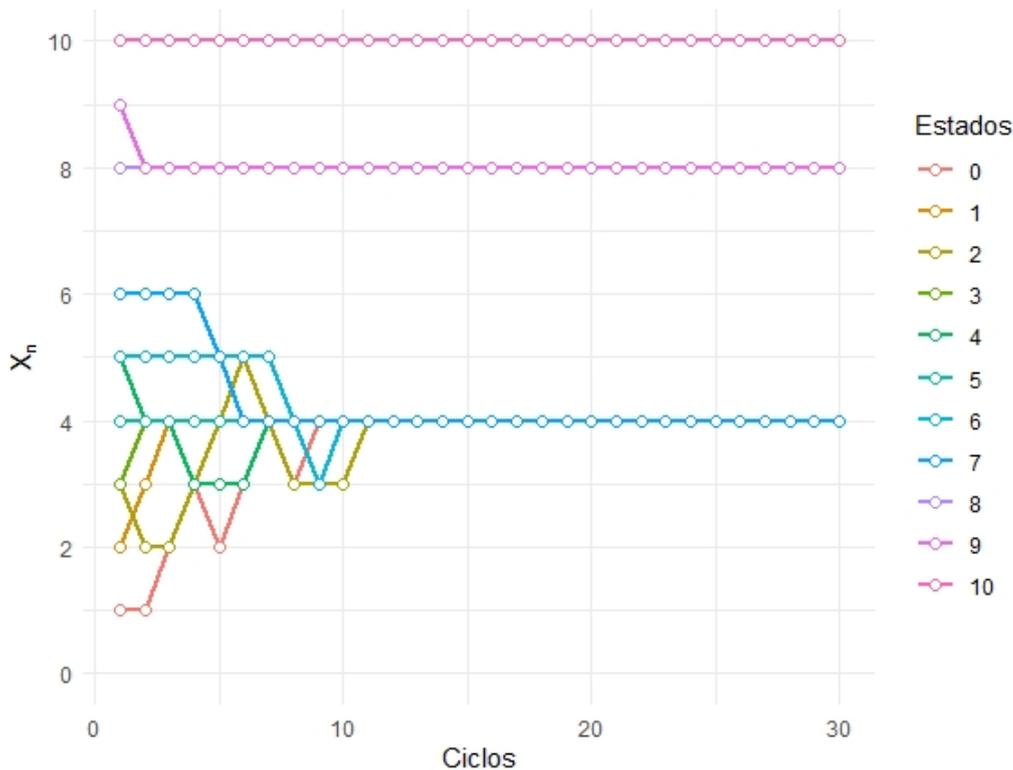
$$\pi_1 = (0, 00015; 0, 00784; 0, 07765; 0, 25310; 0, 40602; 0, 25522; 0; 0; 0; 0; 0)$$

Estos resultados son coherentes, porque en el modelo planteado, si en la caja B , inicialmente hay 10 bolas, ya se dedujo que indefinidamente permanecerá con 10 bolas; y en el caso en que inicialmente no tenga 10 bolas, para un número de transiciones lo suficientemente grande, la caja B tenderá a tener 0, 1, 2, 3, 4 o 5 bolas, con mayor probabilidad de tener 4 bolas.

2.3. CADENAS ABSORBENTES

En el siguiente gráfico se muestran los resultados de la simulación en R para 30 instancias y para la cadena de Ehrenfest empezando desde los diferentes estados iniciales posibles.

Puede verse que si la cadena comienza en cualquiera de los estados de C_1 , la caja B tiende a tener 4 bolas, conforme $n \rightarrow \infty$. Si la cadena comienza en el estado de C_2 , es decir con 10 bolas, permanecerá en dicho estado.



2.3. Cadenas absorbentes

En muchas aplicaciones ocurre que una cadena de Markov tiene, entre sus estados recurrentes, algunos o todos absorbentes. Ya se abordó anteriormente, que si un proceso llega hasta un estado absorbente, entonces no puede salir de él.

En particular, en el ámbito educativo, estas cadenas son de gran importancia para analizar la deserción estudiantil. En este caso, se pueden modelar las distintas etapas por las que pasa un estudiante, como el ingreso a un trayecto académico, la aprobación de espacios curriculares, entre otros posibles estados. Además, se pueden definir estados absorbentes, como el abandono de la carrera, para analizar la probabilidad de que un estudiante abandone sus estudios en cualquier momento.

De esta manera, se pueden obtener estimaciones precisas de las tasas de deserción, lo que puede ayudar a las instituciones educativas a implementar medidas para reducir este problema y mejorar la retención de estudiantes.

Además de su utilidad en el análisis de la deserción estudiantil, las cadenas de Markov absorbentes son ampliamente utilizadas en otras aplicaciones, como la modelización de procesos de decisiones en negocios y finanzas, la evaluación de la eficacia de tratamientos médicos en estudios clínicos y la predicción de la evolución de sistemas complejos en la ingeniería y la física. En el campo de la economía, por ejemplo, se pueden utilizar cadenas de Markov absorbentes para modelar el comportamiento de los mercados financieros y estimar la probabilidad de que se produzcan cambios bruscos en las tendencias económicas. En la biología, estas cadenas se pueden utilizar para analizar la dinámica de poblaciones de especies y predecir la extinción de ciertas especies en un ecosistema determinado.

En esta sección se estudiarán algunos resultados útiles para calcular las probabilidades en que la cadena pasa de un estado transitorio a uno absorbente y el tiempo medio en que esto se produce.

2.3.1. Probabilidades de absorción

En este apartado, interesa analizar cuál es la probabilidad de que comenzando en determinado estado, una cadena de Markov llegue a un subconjunto cerrado e irreducible del espacio de estados.

Definición 2.22. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Sea $C \subset \mathcal{S}$ un conjunto cerrado e irreducible de estados recurrentes y sea

$$\rho_C(x) = P_x(T_C < \infty)$$

la probabilidad de que la cadena de Markov que comienza en x llegue en algún momento a C . Se dice que $\rho_C(x)$ es la probabilidad de que la cadena que comienza en x sea *absorbida* por C .

Como consecuencia de esta definición surgen dos interrogantes:

- i) ¿Cuál es la probabilidad de que la cadena que comienza en x sea absorbida por C , sabiendo que $x \in C$?; y
- ii) ¿Cuál es la probabilidad de que la cadena que comienza en x sea absorbida por C , sabiendo que $x \notin C$?

2.3. CADENAS ABSORBENTES

En el caso en que $x \in C$, la cadena de Markov que comienza en x ya está en C , por lo tanto $\rho_C(x) = 1$.

También, puede verse fácilmente que si $x \notin C$ y siendo x un estado recurrente, entonces $\rho_C(x) = 0$. Esto se debe a que, si se supone que comenzando en x se puede acceder a algún estado de C , como C es cerrado e irreducible, la cadena no podría regresar a x y, por lo tanto, $\rho_{xx} \neq 1$, lo cual contradice el hecho de que x es recurrente.

¿Qué sucede con esta probabilidad si $x \notin C$ y x es un estado transitorio?

En el caso en que \mathcal{S}_T sea un conjunto finito, y en particular \mathcal{S} sea finito, es posible calcular $\rho_C(x)$, con $x \in \mathcal{S}_T$, resolviendo un sistema de ecuaciones lineales con tantas ecuaciones como incógnitas.

Si $x \in \mathcal{S}_T$ y $x \notin C$, la cadena que comienza en x puede acceder a C luego de un paso o permanecer en \mathcal{S}_T luego de un paso y acceder a C luego de un número finito de pasos. Entonces

$$\forall x \in \mathcal{S}_T : \rho_C(x) = \sum_{y \in C} P(x, y) + \sum_{z \in \mathcal{S}_T} P(x, z) \rho_C(z)$$

Teorema 16. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con espacio de estados \mathcal{S} y función de probabilidad P . Si el conjunto \mathcal{S}_T es finito y C es un conjunto cerrado e irreducible de estados recurrentes, entonces el sistema de ecuaciones:

$$f(x) = \sum_{y \in C} P(x, y) + \sum_{y \in \mathcal{S}_T} P(x, y) f(y), \quad x \in \mathcal{S}_T$$

tiene una única solución $f(x) = \rho_C(x)$ donde $x \in \mathcal{S}_T$.

Demostración. Si la igualdad del teorema es válida, entonces debería cumplirse para $y \in \mathcal{S}_T$, es decir:

$$f(y) = \sum_{z \in C} P(y, z) + \sum_{z \in \mathcal{S}_T} P(y, z) f(z), \quad y \in \mathcal{S}_T$$

Sustituyendo $f(y)$ en la igualdad del teorema, para $y \in \mathcal{S}_T$ se obtiene:

$$f(x) = \sum_{y \in C} P(x, y) + \sum_{y \in \mathcal{S}_T} \sum_{z \in C} P(x, y) P(y, z) + \sum_{y \in \mathcal{S}_T} \sum_{z \in \mathcal{S}_T} P(x, y) P(y, z) f(z)$$

La suma de los dos primeros términos del segundo miembro es $P_x(T_C \leq 2)$ y el tercer término es

$$\sum_{z \in \mathcal{S}_T} P^{(2)}(x, z) f(z) = \sum_{y \in \mathcal{S}_T} P^{(2)}(x, y) f(y)$$

De esta manera:

$$f(x) = P_x(T_C \leq 2) + \sum_{y \in \mathcal{S}_T} P^{(2)}(x, y)f(y)$$

Repitiendo este proceso, se tiene que cualquiera sea $n \in \mathbb{N}_0$:

$$f(x) = P_x(T_C \leq n) + \sum_{y \in \mathcal{S}_T} P^{(n)}(x, y)f(y)$$

Como los estados y son transitorios:

$$\lim_{n \rightarrow \infty} P^{(n)}(x, y) = 0, \quad x \in \mathcal{S}, y \in \mathcal{S}_T$$

Por hipótesis, \mathcal{S}_T es finito, entonces:

$$\lim_{n \rightarrow \infty} \sum_{y \in \mathcal{S}_T} P^{(n)}(x, y)f(y) = 0$$

Luego:

$$\begin{aligned} \forall x \in \mathcal{S}_T : f(x) &= \lim_{n \rightarrow \infty} P_x(T_C \leq n) \\ &= P_x(T_C < \infty) \\ &= \rho_C(x) \end{aligned} \tag{2.93}$$

■

2.3.2. Distribución estacionaria de cadenas absorbentes

En este apartado se define una nueva clase de cadenas de Markov en la que todos sus estados recurrentes son absorbentes, y se analiza cuál es la distribución estacionaria concentrada en los conjuntos cerrados e irreducibles incluidos en su espacio de estados.

Definición 2.23. Una cadena de Markov se llama absorbente si cada uno de sus estados recurrentes es también absorbente.

Retomando la representación canónica de la matriz de transición, en el caso que una cadena de Markov sea absorbente, la submatriz cuadrada R que representa la probabilidad de transición en un paso, de ir de un estado recurrente a otro estado

recurrente, es la matriz identidad, por ser dichos estados absorbentes. Por lo tanto, la matriz P puede escribirse como:

$$P = \begin{pmatrix} I & 0 \\ S & Q \end{pmatrix}$$

Teorema 17. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov absorbente con espacio de estados \mathcal{S} y matriz de probabilidad de transición P . Entonces:

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} I & 0 \\ NS & 0 \end{pmatrix}$$

donde $N = (I - Q)^{-1}$

NS representa a la matriz de probabilidad de pasar eventualmente desde un estado transitorio a un estado absorbente. A la matriz N se la llama **matriz fundamental**.

Demostración. Se denota con Q^m a la m -ésima potencia de la matriz Q y por:

$$S_m = (I + Q + Q^2 + \dots + Q^m)S = \left(\sum_{k=0}^m Q^k \right) S$$

Esto puede justificarse por el hecho de que Q es la matriz cuyas entradas son las probabilidades de transición en un paso entre estados transitorios. La probabilidad de transición en m pasos entre estados transitorios, está dada por la ecuación de Chapman-Kolmogorov:

$$P^{(m)}(x, y) = \sum_{z \in \mathcal{S}_T} P^{(m-k)}(x, z)P^{(k)}(z, y)$$

donde \mathcal{S}_T es el espacio de estados transitorios de la cadena.

De la ecuación de Chapman-Kolmogorov, y la definición de producto de matrices, se tiene que:

$$P^m = \begin{pmatrix} I & 0 \\ S_m & Q^m \end{pmatrix}$$

Además, $Q^m \rightarrow 0$ cuando $m \rightarrow \infty$, ya que Q representa la submatriz cuadrada de probabilidades de transición en un paso desde un estado transitorio a otro transitorio.

Ahora se considera la identidad:

$$(I - Q)(I + Q + Q^2 + \dots + Q^{n-1}) = I - Q^n \tag{2.94}$$

2.3. CADENAS ABSORBENTES

Como $(I - Q^n)$ tiende a I para n que tiende a infinito, entonces para n suficientemente grande $(I - Q^n)$ tiene determinante distinto de 0, porque la matriz identidad es inversible. Además, como el determinante de un producto de matrices es igual al producto de los determinantes de dichas matrices, se tiene que, el determinante de $I - Q$ es distinto de 0. Se concluye que la matriz $(I - Q)$ admite inversa.

Si se multiplica a izquierda en ambos miembros de 2.94 por $(I - Q)^{-1}$, se obtiene:

$$I + Q + Q^2 + \dots + Q^{n-1} = (I - Q)^{-1}(I - Q^n)$$

Pero como el segundo miembro tiende a $(I - Q)^{-1}$ cuando $n \rightarrow \infty$, entonces:

$$(I - Q)^{-1} = I + Q + Q^2 + \dots$$

para n lo suficientemente grande.

Sea

$$N = (I - Q)^{-1} = \sum_{k=0}^{\infty} Q^k$$

entonces:

$$\begin{aligned} \lim_{m \rightarrow \infty} S_m &= \lim_{m \rightarrow \infty} \left(\sum_{k=0}^m Q^k \right) S \\ &= \left(\sum_{k=0}^{\infty} Q^k \right) S \\ &= NS \end{aligned} \tag{2.95}$$

Por lo tanto, se obtiene:

$$\lim_{n \rightarrow \infty} P^n = \begin{pmatrix} I & 0 \\ NS & 0 \end{pmatrix}$$

■

Ahora, si se considera la cadena de Markov restringida al conjunto de estados recurrentes y se supone que el conjunto de estados recurrentes está formado por sólo un conjunto cerrado e irreducible, entonces, por las relaciones establecidas, R es una submatriz correspondiente a probabilidades de transición en un paso de un estado recurrente a otro recurrente, en una cadena con espacio de estados recurrentes e irreducible.

2.3. CADENAS ABSORBENTES

Luego, por el Teorema 15, se sigue que la única distribución estacionaria π^* concentrada en \mathcal{S}_R satisface:

$$\lim_{n \rightarrow \infty} P^{(n)}(x, y) = \pi^*(y), \quad x, y \in \mathcal{S}_R$$

Sea

$$\pi = \begin{cases} \pi^*(x) & \text{si } x \in \mathcal{S}_R \\ 0 & \text{si } x \notin \mathcal{S}_R \end{cases}$$

la única distribución estacionaria con respecto a P .

Si se denota por $\mathbf{1}$ al vector columna con unos en cada una de sus entradas, entonces $\mathbf{1}\pi$ representa la matriz cuadrada con filas idénticas a la distribución π .

Teorema 18. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov y matriz de probabilidad de transición P . Si la restricción a la clase recurrente es irreducible, sea π^* la distribución estacionaria concentrada en la clase de estados recurrentes y π dada como se definió en el párrafo anterior. Entonces:

$$\lim_{n \rightarrow \infty} P^n = \mathbf{1}\pi$$

Demostración. Si se representa a P como en la hipótesis del teorema y Q^n y R^n representan la n -ésima potencia de las matrices Q y R , respectivamente, entonces de la ecuación de Chapman-Kolmogorov, se sigue:

$$P = \begin{pmatrix} R^n & 0 \\ S_n & Q^n \end{pmatrix}$$

donde

$$S_{n+1} = \sum_{k=0}^n Q^k S R^{n-k} = \sum_{k=0}^n Q^{n-k} S R^k$$

Sean

$$M = R - \mathbf{1}\pi^* \quad \text{y} \quad M^n = R^n - \mathbf{1}\pi^*$$

Por lo tanto, se tiene que S_{k+1} se puede escribir de la siguiente manera:

$$S_{k+1} = \sum_{n=0}^k Q^{k-n} S \mathbf{1}\pi^* + \sum_{n=0}^k Q^{k-n} S M^n$$

2.3. CADENAS ABSORBENTES

Entonces, existe $0 < \alpha_1 < 1$, tal que cada entrada de la matriz M^n está acotada por α_1^n . Luego cada entrada de la submatriz Q^n , está acotada por α_2^n con $0 < \alpha_2 < 1$. Por lo tanto cada entrada de la matriz:

$$\sum_{n=0}^k Q^{k-n} S M^n$$

está acotada por:

$$\sum_{n=0}^k \alpha_1^{k-n} \alpha_2^n$$

En consecuencia:

$$\lim_{n \rightarrow \infty} \sum_{n=0}^k Q^{k-n} S M^n = 0$$

Luego, tomando límite para $k \rightarrow \infty$ se obtiene:

$$\begin{aligned} \lim_{k \rightarrow \infty} S_{k+1} &= \sum_{n=0}^{\infty} Q^n S \mathbf{1} \pi^* \\ &= (I - Q)^{-1} S \mathbf{1} \pi^* \\ &= (I - Q)^{-1} (I - Q) \mathbf{1}^* \pi^* \\ &= \mathbf{1}^* \pi^* \end{aligned} \tag{2.96}$$

La segunda igualdad se obtiene de relaciones obtenidas en el teorema anterior. Además, el vector columna $S \mathbf{1}$ se puede escribir como el vector columna $(I - Q) \mathbf{1}^*$, ya que la suma de las entradas de la i -ésima fila de la matriz S es igual a 1 menos la suma de las entradas de la i -ésima fila de la matriz Q (S y Q tienen el mismo número de filas). También, los vectores columna $\mathbf{1}$ y $\mathbf{1}^*$, no tienen necesariamente la misma dimensión.

$$\begin{aligned}
 S \times \mathbf{1}_{s \times 1} &= \begin{pmatrix} S_{11} & S_{12} & \dots & S_{1r} \\ S_{21} & S_{22} & \dots & S_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ S_{s1} & S_{s2} & \dots & S_{sr} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\
 &= \begin{pmatrix} S_{11} + S_{12} + \dots + S_{1r} \\ S_{21} + S_{22} + \dots + S_{2r} \\ \vdots \\ S_{s1} + S_{s2} + \dots + S_{sr} \end{pmatrix} \\
 &= \begin{pmatrix} 1 - q_{11} - q_{12} - \dots - q_{1r} \\ 1 - q_{21} - q_{22} - \dots - q_{2r} \\ \vdots \\ 1 - q_{s1} - q_{s2} - \dots - q_{sr} \end{pmatrix}
 \end{aligned} \tag{2.97}$$

Como

$$I - Q = \begin{pmatrix} 1 - q_{11} & -q_{12} & -q_{13} & \dots & -q_{1s} \\ -q_{21} & 1 - q_{22} & -q_{23} & \dots & -q_{2s} \\ -q_{31} & -q_{32} & 1 - q_{33} & \dots & -q_{3s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -q_{s1} & -q_{s2} & -q_{s3} & \dots & 1 - q_{ss} \end{pmatrix}$$

se tiene que:

$$(I - Q)\mathbf{1}^* = \begin{pmatrix} 1 - q_{11} - q_{12} - q_{13} - \dots - q_{1s} \\ 1 - q_{21} - q_{22} - q_{23} - \dots - q_{2s} \\ 1 - q_{31} - q_{32} - q_{33} - \dots - q_{3s} \\ \vdots \\ 1 - q_{s1} - q_{s2} - q_{s3} - \dots - q_{ss} \end{pmatrix}$$

Por lo tanto, se concluye que:

$$(I - Q)^{-1}S\mathbf{1}^* = (I - Q)^{-1}(I - Q)\mathbf{1}^*\pi^*$$

■

2.3.3. Probabilidades de transición en cadenas absorbentes

En este apartado, se estudia cómo calcular las probabilidades de transición en una cadena de Markov con espacio de estados finito, teniendo en cuenta la clasificación de los estados y los resultados de la subsección anterior. En particular, se

2.3. CADENAS ABSORBENTES

abordan estos cálculos desde la representación canónica de la matriz de transición y se estudia el caso en que el espacio de estados recurrentes se particiona en más de un conjunto cerrado e irreducible.

Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y espacio de estados finito \mathcal{S} . Si se quiere calcular la probabilidad de que la cadena visite un estado y desde el estado x (ρ_{xy}), teniendo en cuenta la clasificación de los estados:

- Si x y y son estados recurrentes, entonces ya se concluyó que $\rho_{xy} = 1$ cuando ambos pertenecen al mismo conjunto cerrado y $\rho_{xy} = 0$ cuando pertenecen a conjuntos cerrados diferentes (por el Teorema 9).
- Si x es recurrente y y es transitorio entonces $\rho_{xy} = 0$, ya que no se puede acceder a ningún estado transitorio desde uno recurrente.
- Los casos en que x y y son ambos transitorios, o x es transitorio y y recurrente, se trabajarán en los siguientes teoremas:

Teorema 19. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y espacio de estados finito \mathcal{S} . Sean $x, y \in \mathcal{S}$ estados transitorios y sea $N = (I - Q)^{-1}$ la matriz fundamental con entradas n_{ij} , con $i, j = 1, \dots, \#\mathcal{S}_T$, entonces:

i) $\rho_{xy} = \frac{n_{xy}}{n_{yy}}$

ii) $\rho_{yy} = 1 - \frac{1}{n_{yy}}$

Demostración. por el Teorema 3, se tenía que:

$$G(x, y) = \frac{\rho_{xy}}{1 - \rho_{yy}} \quad \text{y} \quad G(y, y) = \frac{1}{1 - \rho_{yy}}$$

puede obtenerse:

$$\rho_{xy} = \frac{G(x, y)}{G(y, y)} \quad \text{y} \quad \rho_{yy} = 1 - \frac{1}{G(y, y)}$$

Considerando la representación canónica de la matriz de transición de la cadena, y sabiendo que:

$$G(x, y) = \sum_{n=1}^{\infty} P^{(n)}(x, y)$$

2.3. CADENAS ABSORBENTES

entonces puede escribirse a la matriz G , cuyas entradas son los valores esperados de los números de visitas a y , de la cadena que comienza en x :

$$G = I + P + P^2 + P^3 + \dots = \sum_{n=1}^{\infty} P^n(x, y)$$

y como:

$$P^n = \begin{pmatrix} R^n & 0 \\ S_n & Q^n \end{pmatrix}$$

Entonces:

$$G = \begin{pmatrix} \sum_{n=1}^{\infty} R^n & \sum_{n=1}^{\infty} 0 \\ \sum_{n=1}^{\infty} S_n & \sum_{n=1}^{\infty} Q_n \end{pmatrix}$$

Como solo interesa considerar estados x y y transitorios, basta tomar la matriz fundamental $N = (I - Q)^{-1}$.

Ya se dedujo que:

$$\rho_{xy} = \frac{G(x, y)}{G(y, y)} \quad \text{y} \quad \rho_{yy} = 1 - \frac{1}{G(y, y)}$$

entonces:

$$\begin{aligned} \rho_{xy} &= \frac{G(x, y)}{G(y, y)} \\ &= \frac{n_{xy}}{n_{yy}} \end{aligned} \tag{2.98}$$

porque como x e y son estados transitorios, las probabilidades se obtienen de la submatriz $\sum_{n=1}^{\infty} Q_n = N$ de la matriz G .

Análogamente:

$$\begin{aligned} \rho_{yy} &= 1 - \frac{1}{G(y, y)} \\ &= 1 - \frac{1}{n_{yy}} \end{aligned} \tag{2.99}$$

Y con esto quedan probados *i*) y *ii*). ■

Teorema 20. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y espacio de estados finito \mathcal{S} . Sea Q la matriz obtenida a partir de la representación canónica P , eliminando las filas y columnas pertenecientes a los estados recurrentes y sea S la matriz construida a partir de los vectores columna s_y , con $y = 1, \dots, r$, donde r es el número de conjuntos cerrados e irreducibles de estados recurrentes; entonces la matriz $NS = (I - Q)^{-1}S$ devuelve las probabilidades de absorción de los estados transitorios por los conjuntos cerrados e irreducibles.

Demostración. Ya se demostró que el conjunto de estados recurrentes se particiona en conjuntos cerrados e irreducibles (Teorema 5). Por ello, todo estado recurrente se encuentra en un conjunto cerrado e irreducible. De esta manera, si x es un estado transitorio y y pertenece a un conjunto cerrado e irreducible de estados recurrentes C :

$$\forall y, z \in C : \rho_{xy} = \rho_{xz}$$

puesto que:

$$\forall y, z \in C : \rho_{yz} = \rho_{zy} = 1$$

Debido a esto, para calcular ρ_{xy} sólo basta con calcular la probabilidad de absorción $\rho_C(x)$, y de aquí resulta que, cada conjunto cerrado e irreducible de una cadena es tomado como conjunto unitario formado por un estado absorbente.

Si el espacio de estados \mathcal{S} tiene r conjuntos cerrados e irreducibles y m estados transitorios, entonces la submatriz S de la representación canónica P , puede escribirse como:

$$S = \begin{pmatrix} s_{11} & \dots & s_{1r} \\ \vdots & \ddots & \vdots \\ s_{m1} & \dots & s_{mr} \end{pmatrix}$$

siendo en cada caso:

$$s_{xy} = \sum_{z \in C_y} P(x, z), \quad \forall x \in \mathcal{S}_T$$

donde C_y es el conjunto cerrado e irreducible que contiene al estado y .

Así, se suman las probabilidades de transición en un paso del estado transitorio x a cada uno de los estados recurrentes pertenecientes al conjunto cerrado e irreducible correspondiente C_y , y por definición de producto de matrices y los pasos empleados en el cálculo de NS en la demostración del Teorema 17, estas sumas son las entradas de la matriz NS . ■

2.3.4. Tiempo medio de absorción

En esta subsección, se estudia el número de pasos esperados que se necesitan para que una cadena de Markov visite un estado absorbente a partir de un estado determinado. Posteriormente, se analiza el tiempo medio en que esta cadena visita alguno de los estados absorbentes. Sin embargo, en algunas aplicaciones, como la que se desarrolla en este trabajo, puede ser útil encontrar el tiempo medio de absorción hacia un estado absorbente en particular, en lugar de cualquiera de ellos. Por lo tanto, se propone una solución para este problema, el cual no ha sido abordado en la bibliografía consultada.

Definición 2.24. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y espacio de estados finito \mathcal{S} . Se define el tiempo medio de absorción desde un estado transitorio x , como el número medio de pasos que la cadena, que comienza en x , realizará para pasar a cualquier estado absorbente.

Teorema 21. Sea $\{X_n\}_{n \in \mathbb{N}_0}$ una cadena de Markov con función de transición P y espacio de estados finito \mathcal{S} . Sea Q la matriz obtenida a partir de la representación canónica P , eliminando las filas y columnas pertenecientes a los estados recurrentes y sea $N = (I - Q)^{-1}$ la matriz fundamental con entradas n_{ij} , con $i, j = 1, \dots, m$, m el cardinal de \mathcal{S}_T . Si $x \in \mathcal{S}_T$ y A_x es el número medio de pasos que la cadena realiza para pasar a un estado absorbente desde el estado x , entonces

$$A_x = \sum_{j=1}^m n_{xj}$$

Demostración. Como una vez que la cadena visite a un conjunto cerrado e irreducible, no podrá salir de él, entonces el número de pasos necesarios hasta llegar a un estado absorbente, comenzando en uno transitorio, será la suma de los valores esperados de las visitas que se realizan a todos los estados transitorios.

Si se denota con A_x al número de pasos necesarios para que la cadena pase del estado transitorio x a un estado absorbente, entonces:

$$A_x = \sum_{y \in \mathcal{S}_T} G(x, y)$$

Como la matriz fundamental $N = (I - Q)^{-1}$ de la representación canónica de P tiene como entradas a los valores esperados de las visitas que se realizan a

2.3. CADENAS ABSORBENTES

estados transitorios desde estados transitorios, basta fijar la fila x y sumar en y los elementos n_{xy} , cuando $y \in \mathcal{S}_T$. Así, A_x se calcula sumando los elementos de la fila x de la matriz fundamental N , es decir:

$$A_x = \sum_{j=1}^m n_{xj}$$

■

Se puede generalizar esta idea, y calcular el número medio de pasos que hay que recorrer antes de visitar a un estado transitorio y por primera vez. En este caso, bastará solo con sumar aquellos $G(x, z)$ donde $z \neq y$, es decir que, si se impone la visita a y por primera vez, el número medio de etapas será:

$$\sum_{z \neq y} G(x, z), \quad z \in \mathcal{S}_T$$

La desventaja del Teorema 21 es que la suma de los elementos de una fila de la matriz fundamental devuelve el tiempo medio de absorción en que la cadena pasa del estado transitorio asociado a dicha fila hacia cualquier estado absorbente, y lo que puede llegar a interesar es el tiempo de absorción hacia un estado absorbente determinado.

Una solución para este problema, que se propone en este trabajo, es particionar al conjunto de estados absorbentes en conjuntos unitarios (o en su defecto considerar la partición del conjunto de estados recurrentes en los r conjuntos cerrados e irreducibles) y construir r matrices fundamentales asociadas a cadenas con espacio de estados $\mathcal{S}_T \cup C_i$, con $i = 1, \dots, r$.

De esta manera, el tiempo medio de absorción para cada una de estas r cadenas, corresponderá al tiempo medio de absorción hacia cada uno de los conjuntos C_i .

Ejemplo. Cadena de Ehrenfest con tres cajas y 10 bolas. Retomando el ejemplo que se viene trabajando para esta cadena, se había deducido que la matriz

2.3. CADENAS ABSORBENTES

de transición P en su representación canónica es:

$$P = \left(\begin{array}{cccccc|c|cccc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{100} & \frac{10}{100} & \frac{88}{100} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{8}{90} & \frac{16}{90} & \frac{66}{90} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{18}{80} & \frac{18}{80} & \frac{44}{80} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{24}{70} & \frac{24}{70} & \frac{22}{70} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{30}{60} & \frac{30}{60} & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & \frac{24}{50} & 0 & \frac{26}{50} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{28}{40} & \frac{12}{40} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{16}{30} & \frac{14}{30} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{18}{20} & \frac{2}{20} \end{array} \right)$$

Esta cadena no es absorbente puesto que no todos los estados recurrentes son absorbentes, pero si se restringe el espacio de estados a cada uno de los conjuntos cerrados e irreducibles de estados recurrentes C_1 y C_2 :

$$\mathcal{S}_{C_1} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

y

$$\mathcal{S}_{C_2} = \{10, 6, 7, 8, 9\}$$

puede usarse el Teorema 18 para conseguir las dos distribuciones estacionarias de la cadena.

Si se calculan las potencias n -ésimas de las matrices de transición en sus representaciones canónicas de acuerdo a las restricciones a C_1 y C_2 , es decir P_1^n y P_2^n , respectivamente; y se hace n lo suficientemente grande, se obtiene mediante el software R:

$$P_1^n = \left(\begin{array}{cccccccccc} 0,00015 & 0,00784 & 0,07765 & 0,25351 & 0,40602 & 0,25521 & 0 & 0 & 0 & 0 \\ 0,00015 & 0,00784 & 0,07765 & 0,25351 & 0,40602 & 0,25521 & 0 & 0 & 0 & 0 \\ 0,00015 & 0,00784 & 0,07765 & 0,25351 & 0,40602 & 0,25521 & 0 & 0 & 0 & 0 \\ 0,00015 & 0,00784 & 0,07765 & 0,25351 & 0,40602 & 0,25521 & 0 & 0 & 0 & 0 \\ 0,00015 & 0,00784 & 0,07765 & 0,25351 & 0,40602 & 0,25521 & 0 & 0 & 0 & 0 \\ 0,00015 & 0,00784 & 0,07765 & 0,25351 & 0,40602 & 0,25521 & 0 & 0 & 0 & 0 \\ 0,00015 & 0,00784 & 0,07765 & 0,25351 & 0,40602 & 0,25521 & 0 & 0 & 0 & 0 \\ 0,00015 & 0,00784 & 0,07765 & 0,25351 & 0,40602 & 0,25521 & 0 & 0 & 0 & 0 \\ 0,00015 & 0,00784 & 0,07765 & 0,25351 & 0,40602 & 0,25521 & 0 & 0 & 0 & 0 \\ 0,00015 & 0,00784 & 0,07765 & 0,25351 & 0,40602 & 0,25521 & 0 & 0 & 0 & 0 \end{array} \right)$$

y

$$P_2^n = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

En el primer caso se observa que la matriz obtenida corresponde a $\mathbf{1}\pi_1^*$ donde $\mathbf{1}$ es un vector columna de dimensión 10×1 con entradas todas iguales a 1. Entonces:

$$\pi_1^* = (0,00015 \quad 0,00695 \quad 0,06886 \quad 0,22445 \quad 0,36006 \quad 0,33949 \quad 0 \quad 0 \quad 0 \quad 0)$$

Como $\pi_1(x)$ vale $\pi_1^*(x)$ si $x \in C_1$ y 0 en caso contrario, entonces para la cadena sin restricciones se obtiene:

$$\pi_1 = (0,00015 \quad 0,00695 \quad 0,06886 \quad 0,22445 \quad 0,36006 \quad 0,33949 \quad 0 \quad 0 \quad 0 \quad 0)$$

Análogamente, en el segundo caso se observa que la matriz obtenida corresponde a $\mathbf{1}\pi_2^*$ donde $\mathbf{1}$ es un vector columna de dimensión 5×1 con entradas todas iguales a 1. Entonces:

$$\pi_2^* = (1 \quad 0 \quad 0 \quad 0 \quad 0)$$

Como $\pi_2(x)$ vale $\pi_2^*(x)$ si $x \in C_2$ y 0 en caso contrario, entonces para la cadena sin restricciones se obtiene:

$$\pi_2 = (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0)$$

Ambas distribuciones estacionarias coinciden con las encontradas en la sección anterior.

Se consideran ahora las submatrices de la matriz de transición P en su representación canónica.

Por un lado, la matriz R devuelve las probabilidades de pasar de un estado recurrente a otro recurrente y puede subdividirse para considerar a las dos matrices R_1 y R_2 cuadradas, correspondientes a las probabilidades de transición entre estados recurrentes que pertenecen a C_1 y C_2 , respectivamente, siendo estos dos últimos conjuntos cerrados e irreducibles.

$$R = \left(\begin{array}{cccccc|c} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{100} & \frac{10}{100} & \frac{88}{100} & 0 & 0 & 0 & 0 \\ 0 & \frac{8}{90} & \frac{16}{90} & \frac{66}{90} & 0 & 0 & 0 \\ 0 & 0 & \frac{18}{80} & \frac{18}{80} & \frac{44}{80} & 0 & 0 \\ 0 & 0 & 0 & \frac{24}{70} & \frac{24}{70} & \frac{22}{70} & 0 \\ 0 & 0 & 0 & 0 & \frac{30}{60} & \frac{30}{60} & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right)$$

$$R_1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{2}{100} & \frac{10}{100} & \frac{88}{100} & 0 & 0 & 0 \\ 0 & \frac{8}{90} & \frac{16}{90} & \frac{66}{90} & 0 & 0 \\ 0 & 0 & \frac{18}{80} & \frac{18}{80} & \frac{44}{80} & 0 \\ 0 & 0 & 0 & \frac{24}{70} & \frac{24}{70} & \frac{22}{70} \\ 0 & 0 & 0 & 0 & \frac{30}{60} & \frac{30}{60} \end{pmatrix} \quad \text{y} \quad R_2 = (10)$$

Por su parte, la matriz S cuyas entradas son las probabilidades de transición entre estados transitorios y recurrentes es:

$$S = \left(\begin{array}{cccccc|c} 0 & 0 & 0 & 0 & 0 & \frac{24}{50} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

Finalmente, la matriz Q cuyas entradas son las probabilidades de transición entre estados transitorios es:

$$Q = \begin{pmatrix} \frac{26}{50} & 0 & 0 & 0 \\ \frac{28}{40} & \frac{12}{40} & 0 & 0 \\ 0 & \frac{16}{30} & \frac{14}{30} & 0 \\ 0 & 0 & \frac{18}{20} & \frac{2}{20} \end{pmatrix}$$

Si se calcula la matriz fundamental $N = (I - Q)^{-1}$ se obtiene:

$$N = \begin{pmatrix} 2,0833 & 0 & 0 & 0 \\ 2,0833 & 1,4286 & 0 & 0 \\ 2,0833 & 1,4286 & 1,875 & 0 \\ 2,0833 & 1,4286 & 1,875 & 1,111 \end{pmatrix}$$

Luego, pueden obtenerse de esta matriz, las probabilidades de transición de que eventualmente una cadena visite un estado transitorio, comenzando en otro estado transitorio, puesto que por el Teorema 19, permite crear la siguiente tabla que contiene las probabilidades ρ_{xy} , con $x, y \in \mathcal{S}_T$:

ρ_{xy}	6	7	8	9
6	0,52	0	0	0
7	1	0,3	0	0
8	1	1	0,47	0
9	1	1	1	0,01

En primer lugar, se observa que si la caja B tiene 6, 7, 8 o 9 bolas, es seguro que no va a incrementarse este número de bolas en ningún momento y, con probabilidad

1, este número de bolas disminuirá en algún momento. Si la caja B tiene entre 6 y 9 bolas inclusive, la probabilidad de que vuelva a tener el mismo número de bolas en algún momento, coincide con la probabilidad de transición en un paso de que esto suceda. El resultado es coherente, ya que, con probabilidad 1, este número de bolas disminuirá eventualmente.

Si se realiza el producto entre N y S se obtiene:

$$NS = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

donde puede verse que si la caja B tiene 6, 7, 8 o 9 bolas, con probabilidad 1 será absorbida por el conjunto cerrado e irreducible C_1 y habría que esperar un número medio de 1 paso para que esto ocurra.

Podrían calcularse las matrices fundamentales correspondientes a P_1 y P_2 descritas anteriormente, para calcular las probabilidades de absorción desde un estado transitorio, pero no tiene sentido puesto que, si la caja B tiene 6, 7, 8 o 9 bolas (estados transitorios), con probabilidad 1, la caja B pasará a tener un número de bolas perteneciente a C_1 y es imposible que llegue a tener 10 bolas, es decir que sea absorbida por $C_2 = \{10\}$.

Otra forma de obtener las probabilidades de que esta cadena, que comienza en un estado transitorio, sea absorbida por los conjuntos cerrados e irreducibles de estados recurrentes C_1 y C_2 , es resolver el sistema de ecuaciones planteado en el Teorema 16.

Para ello, se plantean las siguiente cuatro ecuaciones correspondientes a los cuatro estados transitorios:

$$f(6) = \sum_{y \in C_1} P(x, y) + \sum_{y \in \mathcal{S}_T} P(x, y)f(y)$$

De la primera sumatoria del segundo miembro de la igualdad, todos los términos son iguales a 0, excepto el correspondiente a $P(6, 5)$, ya que en una transición, el número de bolas de la caja no puede disminuir más de dos unidades. Análogamente, el único término de la segunda sumatoria, distinto de 0, es $P(6, 6)f(6)$. Luego:

$$f(6) = P(6, 5) + P(6, 6)f(6)$$

Entonces:

$$f(6) = \frac{24}{50} + \frac{26}{50}f(6)$$

Realizando un análisis similar:

$$\begin{aligned}
 f(7) &= P(7, 6)f(6) + (7, 7)f(7) \\
 &= \frac{28}{40}f(6) + \frac{12}{40}f(7) \\
 f(8) &= P(8, 7)f(7) + (8, 8)f(8) \\
 &= \frac{16}{30}f(7) + \frac{14}{30}f(8) \\
 f(9) &= P(9, 8)f(8) + (9, 9)f(9) \\
 &= \frac{18}{20}f(8) + \frac{2}{20}f(9)
 \end{aligned} \tag{2.100}$$

Luego, el sistema de ecuaciones queda:

$$\left\{ \begin{array}{l} \frac{24}{50}f(6) \\ -\frac{28}{40}f(6) + \frac{28}{40}f(7) \\ -\frac{16}{30}f(7) + \frac{16}{30}f(8) \\ -\frac{18}{20}f(8) + \frac{18}{20}f(9) \end{array} \right. \begin{array}{l} = \frac{24}{50} \\ = 0 \\ = 0 \\ = 0 \end{array}$$

Cuyo conjunto solución es $S = \{(1, 1, 1, 1)\}$, y como $f(x) = \rho_{C_1}(x)$, con $x \in \mathcal{S}_T$, entonces la probabilidad de que la cadena comenzando en x sea absorbida por C_1 es igual a 1, como ya se había obtenido anteriormente.

Para calcular la probabilidad de que la cadena comenzando en un estado transitorio sea absorbida por C_2 , se plantea el mismo sistema de ecuaciones y, en este caso, la única diferencia que habrá es en la primera sumatoria de $f(6)$ que es nula: Luego, el sistema de ecuaciones queda:

$$\left\{ \begin{array}{l} \frac{24}{50}f(6) \\ -\frac{28}{40}f(6) + \frac{28}{40}f(7) \\ -\frac{16}{30}f(7) + \frac{16}{30}f(8) \\ -\frac{18}{20}f(8) + \frac{18}{20}f(9) \end{array} \right. \begin{array}{l} = 0 \\ = 0 \\ = 0 \\ = 0 \end{array}$$

Cuyo conjunto solución es $S = \{(0, 0, 0, 0)\}$, y como $f(x) = \rho_{C_2}(x)$, con $x \in \mathcal{S}_T$, entonces la probabilidad de que la cadena comenzando en x sea absorbida por C_2 es igual a 0, como ya se había obtenido anteriormente.

Capítulo 3

Aplicación de las cadenas de Markov al problema planteado

En este capítulo, se mostrará una aplicación de las cadenas de Markov a tiempo discreto, en la que se modelará la dinámica de la trayectoria de estudiantes a lo largo de una carrera, tomando como caso de estudio al Profesorado para la Educación Secundaria en Matemática que se dicta en el Instituto de Enseñanza Superior Simón Bolívar de la Provincia de Córdoba.

Asimismo, se explica la metodología empleada para construir cada uno de los modelos propuestos y se evalúa su adecuación a los supuestos necesarios para su aplicación. Además, se presentan los resultados obtenidos y se procede con su interpretación y discusión. En particular, se describe paso a paso cómo obtener y analizar los resultados para permitir al lector comprender y replicar el proceso.

De esta manera, se podrán analizar y comparar las probabilidades de transición por cada campo de la formación superior, hasta que se produce la absorción, para identificar en qué momento de la trayectoria académica se llega a estos estados. Por otra parte, se calculan los tiempos medios de absorción de la trayectoria modelada, para poder aportar claridad sobre la problemática abordada y favorecer el planteo de líneas de acción, con la finalidad de mejorar estas trayectorias.

3.1. Introducción

Las instituciones educativas establecen, como uno de sus objetivos claves, el poder formar a sus estudiantes con un perfil de egresado específico. Para lograr esto, es fundamental que la gran mayoría de los estudiantes puedan completar los trayectos formativos que la institución ofrece.

Para lograr esta continuidad en la trayectoria estudiantil, las instituciones educativas implementan diversas acciones para evitar el abandono o interrupción de

3.1. INTRODUCCIÓN

los estudios. Un primer paso para desarrollar estas acciones es conocer la dinámica de transición de los estudiantes dentro del trayecto formativo, identificando los momentos en que se produce mayor deserción y ralentización en el avance del estudiante.

Al conocer esta información, las instituciones pueden indagar sobre los factores más significativos en la deserción estudiantil y desarrollar estrategias efectivas para abordarlos. Así, se busca lograr una continuidad en la trayectoria estudiantil que permita a los estudiantes alcanzar su perfil de egresado y completar los trayectos formativos que la institución ofrece.

Tal como se mencionó en la sección de *Antecedentes*, existen diferentes enfoques para abordar la problemática de la deserción estudiantil. Algunos estudios aplican técnicas de estadística descriptiva, mientras que otros utilizan modelos que permiten identificar y analizar los factores más significativos que influyen en la deserción. No obstante, muchas veces, esta información no está disponible para las instituciones educativas, ya que requiere de una instancia de recolección de datos, que no siempre es de fácil acceso una vez que se produce la deserción. A pesar de ello, es posible desarrollar acciones para prevenir la deserción estudiantil si se conoce la dinámica en la trayectoria estudiantil. Para ello, se puede utilizar la información disponible en los registros académicos de cada estudiante, para modelar su trayectoria y detectar oportunamente los momentos críticos en los que se puede producir la deserción. De esta manera, se busca implementar medidas preventivas antes de que el abandono ocurra y asegurar que los estudiantes puedan completar satisfactoriamente su trayecto formativo.

Es por ello, que cobran relevancia las cadenas de Markov a tiempo discreto, las cuales permiten modelar estas transiciones.

Este trabajo constituye un estudio de caso cuantitativo y el diseño de la investigación es no experimental, longitudinal de evolución de grupo o cohorte, puesto que se toman los datos de cohortes completas y se analiza su comportamiento a lo largo de los ciclos lectivos.

En este capítulo, se busca estimar las probabilidades de transición entre campos de formación por año de la cursada sobre los cuales se estructura el plan de estudios del Profesorado para la Educación Secundaria en Matemática de la Provincia de Córdoba, así como también las probabilidades de desertar, pedir un pase (para cursar la misma carrera en otra institución de educación superior) o egresar de dicha carrera, desde diferentes estados de la formación; y, el tiempo medio que le lleva a los estudiantes llegar a dichos estados. También, para profundizar aún más en estas trayectorias, se analizan y comparan las probabilidades de transición sobre los distintos campos de la formación, propuestos en el Diseño Curricular y los tiempos medios en que los cursantes desertan, piden pase o egresan.

Para ello, se analizaron las trayectorias de los estudiantes del Profesorado para

la Educación Secundaria en Matemática, que se dicta en el Instituto de Enseñanza Superior Simón Bolívar. Se modelarán estas trayectorias mediante cadenas de Markov a tiempo discreto, tomando diferentes espacios de estados que permitirán estimar las probabilidades mencionadas en el párrafo anterior.

De esta manera, se podrán identificar en qué años del plan de estudios y en qué campos de la formación se presenta mayor probabilidad de que el estudiante abandone sus estudios en esta carrera y en esta institución, así como también, aquellos estadíos en los que la transición al estadío siguiente es más lenta. Con esta información, la institución educativa podrá tender las líneas de acción necesarias para acompañar a estos estudiantes durante su trayectoria académica, así como también evitar que se produzca la deserción.

3.2. Descripción del problema

En el transcurso del Profesorado para la Educación Secundaria en Matemática, los estudiantes deben cursar y acreditar los diferentes espacios curriculares que se ofrecen en el plan de estudios de dicha carrera, para obtener el correspondiente título.

En todos los Institutos de Enseñanza Superior de la Provincia de Córdoba, el plan de estudios vigente para esta carrera, comenzó a implementarse desde el año 2011, sufriendo una ligera modificación de contenidos en el espacio curricular *Sujetos de la Educación* que pasó a llamarse *Sujetos de la Educación y ESI*, para incorporar formación relativa a la Educación Sexual Integral; y el espacio curricular *Unidad de Definición Institucional II* que pasó a llamarse *Problemáticas y Desafíos de la Educación Secundaria*, en donde en esta institución no sufrió cambios en cuanto al contenido que se trabaja.

Los diferentes espacios curriculares de este plan de estudios se dividen en cuatro años de cursado y se clasifican en tres grupos: *campo de la Formación General*, *campo de la Formación Específica*, y *campo de la Práctica Docente*.

Para ingresar a la carrera, los aspirantes realizan los respectivos trámites administrativos y deben aprobar dos seminarios: uno de Formación General y el otro de Formación Específica. Dichos seminarios no son eliminatorios sino que pueden recuperarse hasta su aprobación.

Una vez el aspirante haya ingresado, debe cursar y aprobar los diferentes espacios curriculares que se presentan en diferentes formatos: *asignaturas*, *talleres* y *seminarios*. Los talleres y seminarios se acreditan únicamente promocionándolos sin instancias de examen final, mientras que las asignaturas pueden regularizarse o promocionarse.

En caso de que se regularice una asignatura o el estudiante se quede libre en la

3.2. DESCRIPCIÓN DEL PROBLEMA

misma, puede acreditarla mediante un examen final en los turnos de exámenes disponibles en diciembre, febrero/marzo y julio/agosto. El estudiante libre o regular en una asignatura puede optar por recurrir el espacio curricular, lo que produce a su vez un retardo en el tiempo de egreso.

En relación a las correlatividades entre espacios curriculares, cabe mencionar que según este plan de estudios, los alumnos podrán cursar y acreditar un espacio curricular, habiendo regularizado el espacio correlativo inmediato anterior (si éste existiera). Esto genera la situación de que el estudiante pueda seguir avanzando sin tener espacios aprobados de los primeros años de cursado.

Un aspecto importante a destacar, es la ausencia de la recopilación de información acerca de los ingresantes al momento en que se realiza la inscripción. Conocer estos datos permitiría explorar algunos factores significativos en la dinámica de las trayectorias académicas.

Por su parte, conforme los estudiantes van avanzando en la carrera, se comienza a retrasar el avance y por esta razón, entre otras más variadas, toman la decisión de pedir un pase para continuar sus estudios en otra institución. Esto genera una nueva problemática para la institución, que se debe analizar y atender.

Debido a esta gran diversidad, es que el mayor desafío de retención y acompañamiento de estudiantes en la carrera, se encuentra en los dos primeros años de cursado. Es por esto, que se modeló la dinámica de la trayectoria estudiantil mediante cadenas de Markov a tiempo discreto, con las cuales es posible estimar las probabilidades de transición entre los diferentes campos de la formación a lo largo del tiempo, hasta que se produce la deserción, el estudiante pide un pase hacia otra institución o logra egresar.

En este estudio, se construirán diversas cadenas de Markov cuyos espacios de estados estarán conformados por los que se detallan a continuación, junto con su referencia:

- **I**: el alumno ingresa a la carrera y aún no ha aprobado ningún espacio curricular.
- **D**: el alumno desertó de la carrera en esta institución.
- **P**: el estudiante pide un pase a otra Institución de Educación Superior en donde se dicta la misma carrera.
- **E**: el alumno egresó de la carrera en esta institución.
- **FG i** : el estudiante aprueba todos los espacios curriculares del campo de la Formación General correspondiente al i -ésimo año de cursado ($i = 1, \dots, 4$).
- **PD i** : el alumno aprueba el espacio curricular del campo de la Práctica Docente correspondiente al i -ésimo año de cursado ($i = 1, \dots, 4$).

- **FE i** : el estudiante aprueba todos los espacios curriculares del campo de la Formación Específica correspondiente al i -ésimo año de cursado ($i = 1, \dots, 4$).

Los diferentes espacios curriculares que conforman a cada año de cursado de los campos de la formación, se detallan en el Cuadro 3.1, señalando el formato curricular y el campo al que pertenecen.

Cabe mencionar que los espacios curriculares *Sujetos de la Educación y ESI y Filosofía de las Ciencias* fueron reubicados, en este trabajo, en el campo de la Formación General, debido al enfoque que se les da en esta institución.

3.3. Metodología de trabajo

Para la recolección de datos, se presentó una nota a la Dirección del Instituto de Enseñanza Superior Bolívar, solicitando acceso a los datos disponibles en el Sistema de Gestión de Alumnos.

La Dirección del instituto se mostró interesada en la propuesta, brindando acceso de lector al sistema de donde se pudo descargar una base de datos con la actuación de cada uno de los estudiantes que transitó por este plan de estudios. Las bases de datos reúnen la siguiente información de 875 estudiantes matriculados desde 2011 hasta 2019 inclusive:

- Año de ingreso a la carrera (Cohorte).
- Si egresó o no.
- Apellido y nombre del cursante.
- Tipo y número de documento.
- Cantidad de materias (por año) aprobadas hasta la actualidad.
- Espacio curricular en los que se matriculó (si recurrió el espacio curricular aparece tantas veces como se matriculó, así como también se repite de acuerdo con la cantidad de veces en que se inscribió para rendir durante los turnos de exámenes).
- Fecha de finalización de la cursada para cada espacio curricular.
- Calificación obtenida al finalizar la cursada (si corresponde).
- Condición en la que finalizó la cursada.

3.3. METODOLOGÍA DE TRABAJO

Espacio Curricular	Formato	Campo
Elementos de la Aritmética y el Álgebra	Asignatura	FE1
Problemáticas de la Geometría I	Asignatura	
Modelización Matemática	Taller	
Lenguaje Digital y Audiovisual	Taller	FG1
Pedagogía	Asignatura	
Problemáticas Socioantropológicas en Educación	Seminario	
Práctica Docente I	Seminario	PD1
Problemáticas del Álgebra I	Asignatura	FE2
Problemáticas del Análisis Matemático I	Asignatura	
Problemáticas de la Geometría II	Asignatura	
Didáctica de la Matemática I	Asignatura	
Didáctica General	Asignatura	FG2
Sujetos de la Educación y ESI	Seminario	
Psicología y Educación	Asignatura	
Práctica Docente II	Seminario	PD2
Problemáticas del Álgebra II	Asignatura	FE3
Problemáticas del Análisis Matemático II	Asignatura	
Problemáticas de la Probabilidad y Estadística	Asignatura	
Didáctica de la Matemática II	Asignatura	
Unidad de Definición Institucional I	Asignatura	
Filosofía de las Ciencias	Taller	FG3
Historia y Política de la Educación Argentina	Asignatura	
Práctica Docente III	Seminario	PD3
Problemáticas del Álgebra III	Asignatura	FE4
Modelización Matemática en las Ciencias	Taller	
Problemáticas de la Geometría III	Asignatura	
Didáctica de la Matemática III	Asignatura	
Historia y Epistemología de la Matemática	Asignatura	
Ética y Construcción Ciudadana	Seminario	FG4
Problemáticas y Desafíos de la Educación Secundaria	Seminario	
Práctica Docente IV	Seminario	PD4

Cuadro 3.1: Espacios curriculares por campo de la formación y formato.

3.3. METODOLOGÍA DE TRABAJO

- Si solicitó equivalencia en el espacio curricular, la respectiva resolución y la calificación (si corresponde).
- Si el espacio está aprobado, por haberlo hecho en otra institución educativa de nivel superior de la provincia.
- Condición y nota actual en cada espacio curricular.
- Fechas de exámenes en que se matriculó para rendir cada espacio curricular (en los casos que corresponda).
- Nota de exámenes en que se matriculó para rendir cada espacio curricular (en los casos que corresponda).

Una vez obtenida esta base de datos se procedió a realizar una “limpieza de datos”, acorde a la información necesaria para construir las matrices de transición correspondientes a las cadenas de Markov a considerar. Para ello, se registró para cada estudiante, la fecha de ingreso a la carrera y las fechas de aprobación de cada espacio curricular. Se registraron aquellos espacios curriculares aprobados por equivalencia los cuales no se tuvieron en cuenta a la hora de plantear las transiciones, puesto que no fueron aprobados en esta institución. En este caso, no se consideraron los campos afectados por estas equivalencias, para estos estudiantes.

Además, se registró la fecha de egreso (en aquellos estudiantes que completaron el plan de estudios) y la última fecha en la que se registró actividad en el sistema, como por ejemplo, matriculación en exámenes finales o cursadas, considerando esta última fecha como deserción.

Cabe destacar, que se considera como estudiante que deserta aquel que no registró actuación provisoria en el Sistema de Gestión de Alumnos del Instituto, estableciendo como período de gracia el de un año, por si el estudiante decidió retomar sus estudios.

Para incluir a los alumnos que decidieron continuar sus estudios en otra institución, se solicitó información adicional a la Secretaría del Instituto, ya que en el Sistema de Gestión de Alumnos, no se menciona este estado. Se consiguió una nueva planilla en donde se registraba el año en que se produjo el pase y la institución y carrera al cual se cambia. De esta manera, era posible observar y registrar aquellos que decidían cambiarse de institución para seguir cursando la misma carrera.

Seguidamente, se consideró como periodos a los ciclos lectivos de cursada que van aproximadamente desde la tercera semana de marzo de un año hasta que finaliza el turno de exámenes de febrero/marzo del año siguiente.

Luego, se construyeron nuevas columnas con las transiciones entre estados a lo largo del tiempo de permanencia en la carrera; y, con estas nuevas columnas, se procedió a realizar los conteos de cada una de las transiciones entre los diferentes estados a considerar.

Cabe mencionar, que todos los estudiantes inician cada cadena desde el estado I , el cual para evitar que se superponga con los estados alcanzados en el primer año de cursada, se coloca en el año anterior al inicio. Por su parte, los estados absorbentes D , P y E , se cuentan en el periodo inmediato superior a la última actuación registrada por cada estudiante.

En caso de que algún estudiante alcanza dos o más estados en un mismo periodo se define el mismo número de transiciones desde el periodo anterior a cada uno de estos estados.

Antes de realizar los ajustes definitivos que se presentan en este trabajo, se intentó plantear un modelo más sencillo en el que se consideraban solamente los estados I , R , D , P y E , siendo R el estado de “Alumno Regular” que se obtiene al haber aprobado al menos un espacio curricular. Con este modelo se quería estudiar los tiempos medios en que los estudiantes se mantienen regulares en la carrera, así como el tiempo medio empleado en desertar, pedir un pase o egresar. Esta cadena no satisfacía los supuestos de una cadena de Markov y por ello se descartó.

Seguidamente, se probó considerar periodos más cortos. En lugar de un ciclo lectivo completo, se trabajó con cuatrimestres, los cuales cada uno contenían un turno de exámenes finales. Debido a la presencia de espacios curriculares con formato de seminario o taller, los cuales no pueden rendirse en julio, el modelo planteado no satisfacía el supuesto de estacionariedad.

En la siguiente sección se describe el procedimiento utilizado para construir las cadenas de Markov. Se presenta el mecanismo para probar los supuestos de una cadena de Markov y se estiman las probabilidades de transición.

3.4. Ajustes, resultados y análisis

3.4.1. Transición por el campo de la Formación General

El modelo que se genera a continuación, se basa en cadenas de Markov a tiempo discreto. Se considera como espacio de estados al conjunto:

$$\mathcal{S} = \{I, FG1, FG2, FG3, FG4, D, P, E\}$$

y cada paso o etapa de la cadena se corresponde con un ciclo lectivo, desde que inicia en la tercera semana de marzo hasta la semana previa a ésta en el siguiente año.

Con estas consideraciones, se trata de predecir de manera general las probabilidades de que el estudiante que ingresa o permanece en la carrera, abandone sus estudios en la institución, pida un pase a otra institución para cursar la misma carrera o egrese. De acuerdo a lo expuesto en el *Capítulo 2*, los estados I , $FG1$,

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

$FG2$, $FG3$ y $FG4$ son transitorios, mientras que los estados D , P y E son considerados como recurrentes absorbentes, puesto que una vez se accede a ellos ya no se puede salir.

Las transiciones posibles son las siguientes:

- $I \longrightarrow I$: el estudiante ingresante no aprobó todos los espacios curriculares de ningún FGi ($i = 1, 2, 3, 4$), durante el ciclo lectivo, permaneciendo en el mismo estado en el siguiente ciclo.
- $I \longrightarrow FGi$: el estudiante ingresante aprueba todos los espacios curriculares que conforman FGi ($i = 1, 2, 3, 4$), pasando al estado FGi en el siguiente ciclo lectivo. (Debe recordarse que tal como están planteadas las correlatividades puede darse esta situación para cualquier $i = 1, 2, 3, 4$).
- $I \longrightarrow D$: el estudiante ingresante abandona sus estudios en esta institución sin aprobar ningún FGi , pasando al estado de *deserción* en el siguiente ciclo lectivo.
- $I \longrightarrow P$: el estudiante ingresante pide un pase a otra institución para continuar la misma carrera, sin haber aprobado ningún FGi .
- $FGi \longrightarrow FGi$: el alumno que aprobó todos los espacios curriculares de FGi , y continúa en la misma condición durante el ciclo lectivo, permanece en el mismo estado en el siguiente ciclo.
- $FGi \longrightarrow FGj$ ($i \neq j$): el alumno que aprobó todos los espacios curriculares de FGi , aprueba todos los espacios curriculares de FGj durante el ciclo lectivo, pasando al estado FGj en el siguiente ciclo.
- $FGi \longrightarrow D$: el estudiante, habiendo aprobado FGi , abandona sus estudios en esta institución, pasando al estado de *deserción* en el siguiente ciclo lectivo.
- $FGi \longrightarrow P$: el estudiante pide un pase a otra institución para continuar la misma carrera, habiendo aprobado FGi , en el ciclo anterior.
- $FGi \longrightarrow E$: el estudiante termina de aprobar todos los espacios curriculares del plan de estudios en el ciclo lectivo y pasa al estado de *egreso* en el siguiente ciclo lectivo.

Las transiciones no permitidas entre dos ciclos lectivos consecutivos son:

- $I \longrightarrow E$: el estudiante ingresante no puede egresar en el ciclo lectivo siguiente, puesto que por la estructura del plan de estudios, debe permanecer en la carrera al menos 4 ciclos lectivos, desde que ingresa.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

- $FGi \rightarrow I$: el estudiante ya aprobó algún FGi del plan de estudios y es imposible que vuelva al estado de ingresante.
- $D \rightarrow I$: no se registraron casos en que el estudiante haya abandonado la carrera y haya regresado en el siguiente ciclo lectivo para comenzar nuevamente desde el inicio. Si se diera ese caso, se reconocería el trayecto previo del estudiante por medio de equivalencias.
- $P \rightarrow I$: no se registraron casos en que el estudiante se haya cambiado de institución y haya regresado en el siguiente ciclo lectivo para comenzar nuevamente desde el inicio. Si se diera ese caso, se reconocería el trayecto previo del estudiante por medio de equivalencias.
- $E \rightarrow I$: el estudiante que egresó no puede volver al estado de ingresante, salvo que desee cursar nuevamente la carrera desde el inicio.
- $D \rightarrow FGi$: esta transición, si bien es posible, no se registró que ningún estudiante que haya abandonado sus estudios en un ciclo lectivo, los haya retomado en el siguiente ciclo.
- $P \rightarrow FGi$: esta transición, si bien es posible, no se registró que ningún estudiante que se haya cambiado de institución en un ciclo lectivo, haya regresado a esta en el siguiente ciclo.
- $E \rightarrow FGi$: el estudiante que egresó en un ciclo lectivo, no regresa al estado de estudiante regular en el siguiente ciclo lectivo.
- Las transiciones entre estados absorbentes tampoco son posibles.

En las siguientes tablas se presentan las transiciones entre estados para dos periodos consecutivos. En las filas se omiten los estados absorbentes D , P y E puesto que sólo representarán los estudiantes que en esa transición abandonaron la carrera en la institución, pidieron pase o egresaron. Se consideraron para el análisis los periodos desde 2014 hasta 2019, puesto que para los ciclos 2011, 2012 y 2013 no debieran registrarse egresados.

De esta manera, puede observarse que en el año 2014, de los estudiantes que estaban en el estado I (es decir ingresantes o alumnos que no hayan aprobado los espacios curriculares de algún FGi): 33 permanecen en el mismo estado, 16 aprobaron todos los espacios curriculares de la Formación General de primer año, alcanzando el estado $FG1$, 44 abandonaron la carrera y un estudiante pidió un pase para cursar la carrera en otra institución.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

De los estudiantes que habían aprobado *FG1* en 2013, se observa que en 2014, 18 continúan en el mismo estado, 15 aprobaron los espacios de la Formación General de segundo año. es decir pasaron al estado *FG2*, 4 abandonaron la carrera y 2 pidieron un pase para seguir cursándola en otra institución.

2014	<i>I</i>	<i>FG1</i>	<i>FG2</i>	<i>FG3</i>	<i>FG4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	33	16	0	0	0	44	1	0
<i>FG1</i>	0	18	15	0	0	4	2	0
<i>FG2</i>	0	1	18	10	2	1	2	0
<i>FG3</i>	0	0	0	0	3	0	0	0
<i>FG4</i>	0	0	0	1	0	0	0	0

2015	<i>I</i>	<i>FG1</i>	<i>FG2</i>	<i>FG3</i>	<i>FG4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	47	23	1	0	0	55	4	0
<i>FG1</i>	0	18	9	1	0	7	2	0
<i>FG2</i>	0	2	20	5	2	1	5	0
<i>FG3</i>	0	0	0	5	5	1	1	0
<i>FG4</i>	0	0	0	1	4	0	1	0

2016	<i>I</i>	<i>FG1</i>	<i>FG2</i>	<i>FG3</i>	<i>FG4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	43	22	0	1	0	45	0	0
<i>FG1</i>	0	14	20	0	0	5	1	0
<i>FG2</i>	0	1	10	10	1	5	1	0
<i>FG3</i>	0	0	1	5	7	0	0	0
<i>FG4</i>	0	0	0	2	4	0	1	4

2017	<i>I</i>	<i>FG1</i>	<i>FG2</i>	<i>FG3</i>	<i>FG4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	38	19	2	1	1	55	2	0
<i>FG1</i>	0	15	16	0	1	6	1	0
<i>FG2</i>	0	1	15	9	1	4	1	0
<i>FG3</i>	0	0	1	7	7	2	1	1
<i>FG4</i>	0	0	0	2	4	0	0	5

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

2018	<i>I</i>	<i>FG1</i>	<i>FG2</i>	<i>FG3</i>	<i>FG4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	48	22	5	0	0	55	0	0
<i>FG1</i>	0	9	16	0	0	5	4	0
<i>FG2</i>	0	1	18	6	0	6	2	1
<i>FG3</i>	0	1	0	6	8	1	2	1
<i>FG4</i>	0	1	0	1	5	0	2	5

2019	<i>I</i>	<i>FG1</i>	<i>FG2</i>	<i>FG3</i>	<i>FG4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	31	31	5	0	0	63	0	0
<i>FG1</i>	0	12	12	0	1	7	1	0
<i>FG2</i>	0	3	17	8	4	7	3	0
<i>FG3</i>	0	0	0	3	6	2	0	1
<i>FG4</i>	0	0	0	0	11	0	0	1

Cuadro 3.2: Transiciones entre estados por año - Campo de la Formación General.

A continuación, se describe el procedimiento que se realizó para probar que se cumplen los supuestos para ajustar una cadena de Markov: estacionariedad y la propiedad de Markov.

Para que se cumpla la propiedad de estacionariedad, se requiere homogeneidad de las probabilidades de transición durante todos los periodos. Para verificar esta propiedad se debe asegurar que la frecuencia de estudiantes en cada uno de los estados no sea dependiente del periodo en el cual se toma la medición.

La tabla de contingencia debe probarse para cada una de las posibles transiciones $x \rightarrow y$, sin embargo, debido a que los estados *D*, *P* y *E* son absorbentes, sólo se determinará la independencia entre los periodos académicos y las frecuencias de estudiantes que cambian de estado.

A continuación, se detallan las tablas construidas a partir de las frecuencias de transiciones entre cada estado transitorio y todos los demás estados, para cada período observado:

Periodo	<i>I</i>	<i>FG1</i>	<i>FG2</i>	<i>FG3</i>	<i>FG4</i>	<i>D</i>	<i>P</i>	<i>E</i>
2014	36	16	0	0	0	44	1	0
2015	47	23	1	0	0	55	4	0
2016	43	22	0	1	0	45	0	0
2017	38	19	2	1	1	55	2	0
2018	48	22	5	0	0	55	0	0
2019	31	31	5	0	0	63	0	0

Cuadro 3.3: Frecuencia de transiciones entre el estado *I* y todos los estados para cada periodo.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Periodo	<i>I</i>	<i>FG1</i>	<i>FG2</i>	<i>FG3</i>	<i>FG4</i>	<i>D</i>	<i>P</i>	<i>E</i>
2014	0	18	15	0	0	4	2	0
2015	0	18	9	1	0	7	2	0
2016	0	14	20	0	0	5	1	0
2017	0	15	16	0	1	6	1	0
2018	0	9	16	0	0	5	4	0
2019	0	12	12	0	1	7	1	0

Cuadro 3.4: Frecuencia de transiciones entre el estado *FG1* y todos los estados para cada periodo.

Periodo	<i>I</i>	<i>FG1</i>	<i>FG2</i>	<i>FG3</i>	<i>FG4</i>	<i>D</i>	<i>P</i>	<i>E</i>
2014	0	1	18	10	2	1	2	0
2015	0	2	20	5	2	1	5	0
2016	0	1	10	10	1	5	1	0
2017	0	1	15	9	1	4	1	0
2018	0	1	18	6	0	6	2	1
2019	0	3	17	8	4	7	3	0

Cuadro 3.5: Frecuencia de transiciones entre el estado *FG2* y todos los estados para cada periodo.

Periodo	<i>I</i>	<i>FG1</i>	<i>FG2</i>	<i>FG3</i>	<i>FG4</i>	<i>D</i>	<i>P</i>	<i>E</i>
2014	0	0	0	0	3	0	0	0
2015	0	0	0	5	5	1	1	0
2016	0	0	1	5	7	0	0	0
2017	0	0	1	7	7	2	1	1
2018	0	1	0	6	8	1	2	1
2019	0	0	0	3	6	2	0	1

Cuadro 3.6: Frecuencia de transiciones entre el estado *FG3* y todos los estados para cada periodo.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Periodo	I	$FG1$	$FG2$	$FG3$	$FG4$	D	P	E
2014	0	0	0	1	0	0	0	0
2015	0	0	0	1	4	0	1	0
2016	0	0	0	2	4	0	1	4
2017	0	0	0	2	4	0	0	5
2018	0	1	0	1	5	0	2	5
2019	0	0	0	0	11	0	0	1

Cuadro 3.7: Frecuencia de transiciones entre el estado $FG4$ y todos los estados para cada periodo.

Como debe asegurarse que estas frecuencias sean homogéneas, a través de cada uno de los periodos, se utilizará el test exacto de Fisher. Este test se utiliza cuando se quiere analizar si existe asociación entre dos variables cualitativas, es decir si las proporciones de una variable son diferentes dependiendo del valor que adquiera la otra variable. En la gran mayoría de casos, el test de Fisher, se aplica para comparar dos variables categóricas con dos niveles cada una (tabla 2×2). Es posible utilizarlo con tablas $m \times n$ niveles pero los requerimientos de cálculo son altos (Requena Guerrero, 1999).

El test de Fisher es más preciso que sus equivalentes aproximados (test chi-cuadrado de independencia o G-test de independencia) cuando el número de eventos esperado por nivel es pequeño. Se recomienda utilizarlo siempre que sea posible (tiempo de computación), aunque para observaciones totales mayores a 1000 los resultados de los test aproximados son muy parecidos.

En este caso, las hipótesis que se plantean son:

- H_0 : Las variables son independientes, vs.
- H_1 : Las variables no son independientes.

El test exacto de Fisher se basa en la distribución hipergeométrica múltiple, que permite calcular la probabilidad exacta de obtener una determinada distribución de eventos en una tabla.

Sea una tabla de contingencia X para dos variables categóricas, con m y n niveles cada una y conteos por celdas x_{ij} , totales fila F_i , totales columna C_j , $i = 1, \dots, m$, $j = 1, \dots, n$, y total de elementos de la muestra N . Bajo la hipótesis nula de independencia entre las dos variables, la distribución exacta de X contiene parámetros desconocidos que se eliminan condicionando sobre sus estadísticos suficientes, los totales marginales.

De esta manera, la distribución exacta de X , se conoce como distribución hi-

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

pergeométrica multivariante múltiple, y está dada por:

$$P(X) = \frac{\prod_{i=1}^m F_i! \prod_{j=1}^n C_j!}{N! \prod_{i,j} x_{ij}!}$$

Esta función de probabilidad ya no contiene parámetros desconocidos y la inferencia exacta es posible.

El dominio de $P(X)$ es el conjunto de dimensión $(m-1)(n-1)$ formado por todas las tablas $m \times n$ con totales marginales de filas F_i , $i = 1, \dots, m$ y totales marginales de columnas C_j , $j = 1, \dots, n$. A este conjunto se lo conoce como familia isomarginal para dichos totales marginales.

Si se denota por \mathbf{x} al vector aleatorio de dimensión $(m-1)(n-1)$ con componentes x_{ij} , $i = 1, \dots, m-1$ y $j = 1, \dots, n-1$, entonces la función de probabilidad anterior es la función de probabilidad para dicho vector.

El test de Fisher calcula las probabilidades de todas las posibles tablas \mathbf{x} y suma las de aquellas tablas que tengan probabilidades menores o iguales que la tabla observada, generando así el p -valor para una prueba bilateral.

En R puede obtenerse el cálculo del p -valor correspondiente a esta prueba, mediante el siguiente comando:

```
fisher.test(x = tabla, alternative = "two.sided")
```

Aplicando el test exacto de Fisher, en las tablas de contingencia 3.4.1, para distintos grupos de años consecutivos, se obtiene que para el periodo 2014-2019, las frecuencias de transiciones son estadísticamente independientes del periodo.

Estado	p -valor
<i>I</i>	0,1074
<i>FG1</i>	0,6567
<i>FG2</i>	0,5037
<i>FG3</i>	0,9820
<i>FG4</i>	0,0587

Cuadro 3.8: Resultados de la prueba de independencia entre periodos y transiciones (2014-2019).

En este caso, si se desean obtener estimaciones a largo plazo, del comportamiento de las transiciones, debe tenerse en cuenta el periodo 2014-2019.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Otro supuesto a asegurar es que las frecuencias de transición obtenidas cumplan con la principal característica de las cadenas de Markov, deben ser dependientes únicamente de su estado anterior.

Para verificar este supuesto, se realizó el test exacto de Fisher en la tablas de contingencia para los años que pertenecen al periodo 2014-2019, obteniendo en todos los casos un p -valor de 0,0005.

Los resultados de la prueba indican que se rechaza la hipótesis nula entre la independencia de estados, por lo que se puede garantizar que los resultados obtenidos en un estado, dependen del estado en el periodo anterior.

Se estiman las matrices de transición P , utilizando el estimador de máxima verosimilitud del resultado del Teorema 2.

$$\hat{P}_{MV} = \frac{n_{xy}}{\sum_{z \in \mathcal{S}} n_{xz}}$$

análogamente a como se trabajó con los ejemplos desarrollados en el marco teórico.

Los resultados de dichas estimaciones, se muestran en la siguiente tabla:

Estados	I	$FG1$	$FG2$	$FG3$	$FG4$	D	P	E
I	0,3366	0,1865	0,0182	0,0028	0,0014	0,4446	0,0098	0,0000
$FG1$	0,0000	0,3874	0,3964	0,0045	0,0090	0,1532	0,0495	0,0000
$FG2$	0,0000	0,0441	0,4804	0,2353	0,0490	0,1176	0,0686	0,0049
$FG3$	0,0000	0,0128	0,0256	0,3333	0,4615	0,0769	0,0513	0,0385
$FG4$	0,0000	0,0182	0,0000	0,1273	0,5091	0,0000	0,0727	0,2727
D	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000
P	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000
E	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000

Cuadro 3.9: Matriz de transición - Campo de la Formación General.

En el siguiente gráfico, pueden observarse las probabilidades de transición en un año:

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

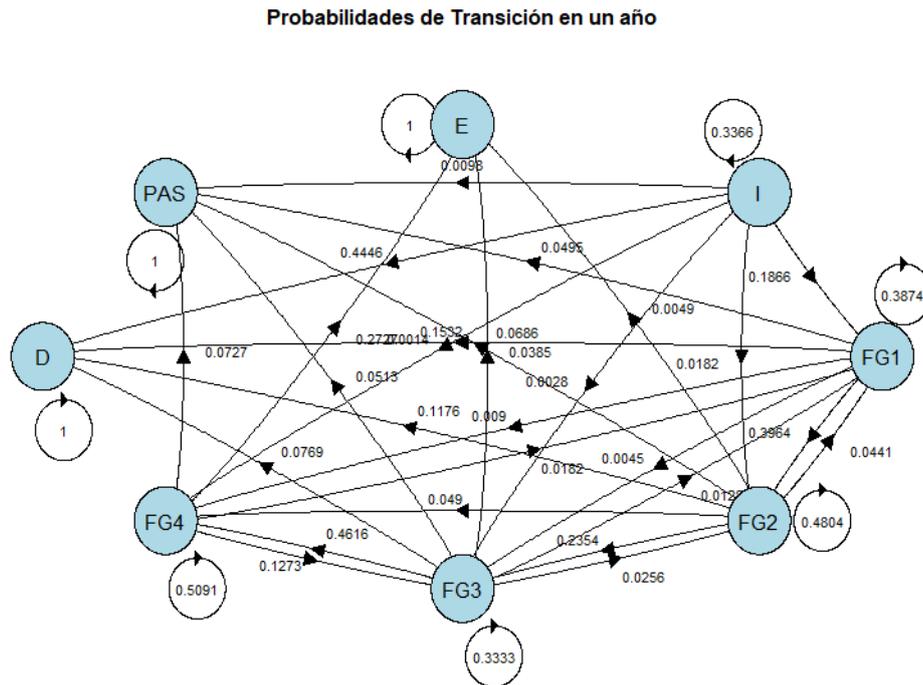


Figura 3.1: Probabilidades de transición en un año.

Se pueden obtener las probabilidades de transición en n pasos, o más específicamente, en n ciclos lectivos, calculando las potencias n -ésimas de la matriz de transición P .

Interesa conocer entonces, ¿cuál es la probabilidad de que un estudiante que ingresa, egrese en los 4 años propuestos en el plan de estudios de la carrera? y, ¿cuál es la probabilidad de llegar a algún estado absorbente, estando en otros estados de la carrera, luego de 4 años.

Para ello, se calcula P^4 :

Estados	I	$FG1$	$FG2$	$FG3$	$FG4$	D	P	E
I	0,0128	0,0422	0,0797	0,0335	0,0230	0,7498	0,0516	0,0073
$FG1$	0	0,0450	0,1481	0,1049	0,1067	0,3824	0,1677	0,0452
$FG2$	0	0,0243	0,0860	0,1062	0,1687	0,2838	0,1929	0,1381
$FG3$	0	0,0145	0,0217	0,0798	0,1974	0,1413	0,1733	0,3719
$FG4$	0	0,0112	0,0150	0,0568	0,1453	0,0349	0,1686	0,5683

Cuadro 3.10: Matriz de transición en 4 pasos - Transición por el campo de la Formación General.

Se puede observar que la probabilidad de que un estudiante recién ingresado

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

se egrese en cuatro años es muy baja, solo del 0,73 %. Además, la probabilidad de cambiar de institución es algo mayor, del 5,16 %, y la probabilidad de abandonar los estudios es significativamente más alta, del 74,98 %.

En cuanto a los estudiantes que aprobaron los espacios curriculares del primer año de la Formación General, se observa que la probabilidad de egreso aumenta a un 4,52 %. Esto sugiere que para lograr su egreso, los estudiantes necesitan más de cuatro años de estudio. Por otro lado, la probabilidad de abandonar los estudios disminuye considerablemente, hasta un 38,24 %, mientras que la probabilidad de cambiar de institución aumenta hasta un 16,77 %.

Para los estudiantes que aprobaron los espacios curriculares del campo de la Formación General de segundo año, se observa que la probabilidad de egreso aumenta significativamente, alcanzando un 13,81 %. Asimismo, la probabilidad de abandonar los estudios disminuye a un 28,38 %, mientras que la probabilidad de cambiar de institución aumenta a un 19,29 %.

Para los estudiantes que aprobaron los espacios curriculares del campo de la Formación General de tercer año, se observa que la probabilidad de egreso aumenta significativamente, alcanzando un 37,19 %. Además, la probabilidad de abandonar los estudios disminuye a un 14,13 %, aunque la probabilidad de cambiar de institución se reduce ligeramente a un 17,33 %.

Para los estudiantes que aprobaron los espacios curriculares del campo de la Formación General de cuarto año, se observa que la probabilidad de egreso sigue aumentando significativamente, llegando a un 56,83 %. Por otro lado, la probabilidad de abandonar los estudios disminuye notablemente a un 3,49 %, mientras que la probabilidad de cambiar de institución se reduce ligeramente a un 16,86 %.

Se puede decir que, conforme los estudiantes van aprobando los espacios curriculares del campo de la Formación General de la carrera, se consolidan en la misma y aumenta la probabilidad de graduarse, disminuyendo la de abandonar los estudios. Asimismo, se observa que a medida que avanzan por los espacios de la Formación General, aumenta la probabilidad de cambiar de institución, lo que sugiere que estas transferencias ocurren principalmente en los cursos superiores.

En el siguiente cuadro se muestran las probabilidades de transición en más de 4 ciclos lectivos para visualizar cuáles son las transiciones más probables por el campo de la Formación General y las probabilidades de llegar a algún estado absorbente:

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

5 Ciclos Lectivos								
	I	FG1	FG2	FG3	FG4	D	P	E
I	0,0043	0,0231	0,0561	0,0331	0,0315	0,7740	0,0627	0,0152
FG1	0	0,0273	0,0917	0,0836	0,1104	0,4148	0,1933	0,0791
FG2	0	0,0176	0,0537	0,0772	0,1393	0,3058	0,2177	0,1886
FG3	0	0,0112	0,0182	0,0569	0,1385	0,1523	0,1940	0,4289
FG4	0	0,0084	0,0131	0,0410	0,1010	0,0427	0,1836	0,6102

6 Ciclos Lectivos								
	I	FG1	FG2	FG3	FG4	D	P	E
I	0,0015	0,0132	0,0370	0,0284	0,0343	0,7886	0,0717	0,0254
FG1	0	0,0177	0,0570	0,0636	0,0995	0,4362	0,2132	0,1128
FG2	0	0,0127	0,0348	0,0562	0,1094	0,3207	0,2364	0,2299
FG3	0	0,0084	0,0147	0,0409	0,0978	0,1605	0,2088	0,4690
FG4	0	0,0062	0,0106	0,0296	0,0711	0,0487	0,1944	0,6394

7 Ciclos Lectivos								
	I	FG1	FG2	FG3	FG4	D	P	E
I	0,0004	0,0080	0,0238	0,0226	0,0325	0,7978	0,0789	0,0360
FG1	0	0,0120	0,0360	0,0474	0,0830	0,4505	0,2285	0,1427
FG2	0	0,0092	0,0232	0,0409	0,0834	0,3311	0,2502	0,2620
FG3	0	0,0062	0,0114	0,0296	0,0695	0,1667	0,2194	0,4973
FG4	0	0,0045	0,0083	0,0215	0,0504	0,0531	0,2021	0,6600

8 Ciclos Lectivos								
	I	FG1	FG2	FG3	FG4	D	P	E
I	0,0002	0,0051	0,0152	0,0173	0,0282	0,8038	0,0844	0,0458
FG1	0	0,0083	0,0233	0,0349	0,0660	0,4602	0,2400	0,1673
FG2	0	0,0066	0,0158	0,0297	0,0626	0,3383	0,2604	0,2865
FG3	0	0,0045	0,0087	0,0214	0,0496	0,1712	0,2271	0,5174
FG4	0	0,0033	0,0063	0,0156	0,0360	0,0565	0,2077	0,6746

Cuadro 3.11: Matrices de transición en más de 4 pasos - Campo de la Formación General.

Los siguientes gráficos muestran la convergencia de las probabilidades de transición para 12 años lectivos sobre cada uno de los estados futuros basado en un estado actual.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

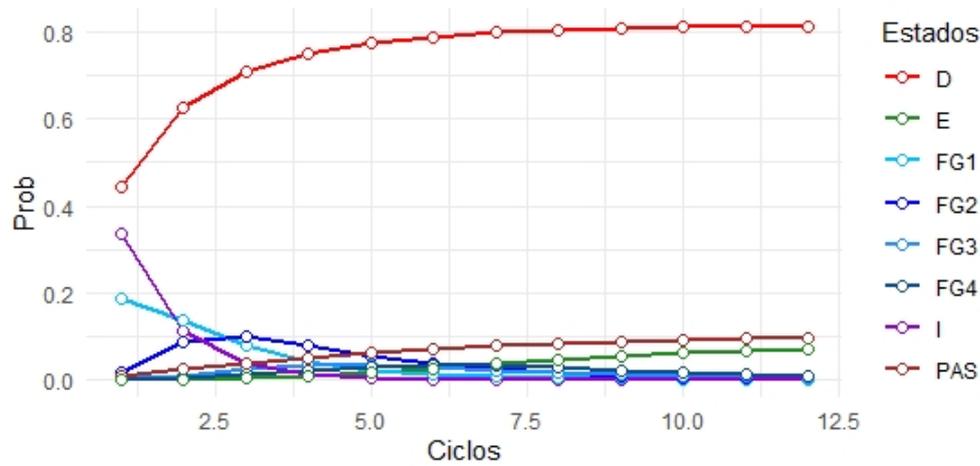


Figura 3.2: Probabilidades de transición por año lectivo, dado que actualmente es Ingresante.

Si se dejan solamente los estados $I, FG1, D, P, E$, es decir el estado consecutivo a I y los estados absorbentes:

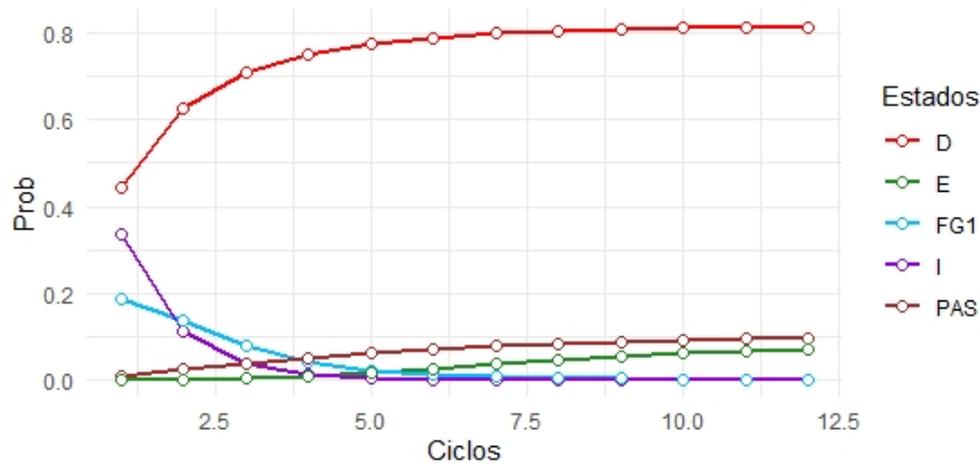


Figura 3.3: Probabilidades de transición por año lectivo, dado que actualmente es Ingresante.

En la Figura 3.2, se puede apreciar que la probabilidad de abandono de los estudiantes ingresantes es mayor que la de aprobación de otros espacios de la Formación General, y esta probabilidad aumenta a medida que transcurren los años, llegando a superar ligeramente el 80 %.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Se observa una mayor probabilidad de que los estudiantes no completen el primer año del campo de la Formación General después de su primer año de estudio, pero esta situación se revierte después de dos años.

A medida que transcurren los ciclos lectivos, es bastante similar la probabilidad de egresar y la de pedir un pase de institución, siendo ligeramente mayor esta última.

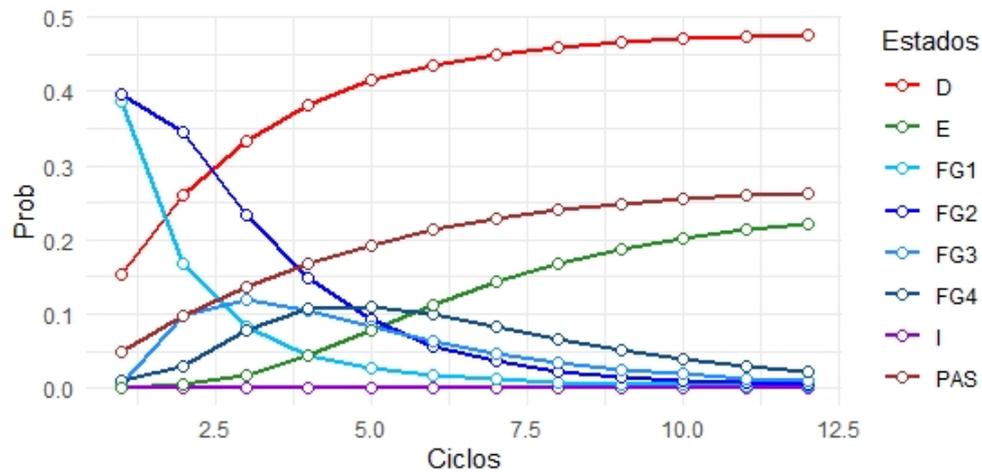


Figura 3.4: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el primer año del campo de la Formación General.

Si se dejan solamente los estados $FG1$, $FG2$, D , P , E , es decir el estado consecutivo a $FG1$ y los estados absorbentes:

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

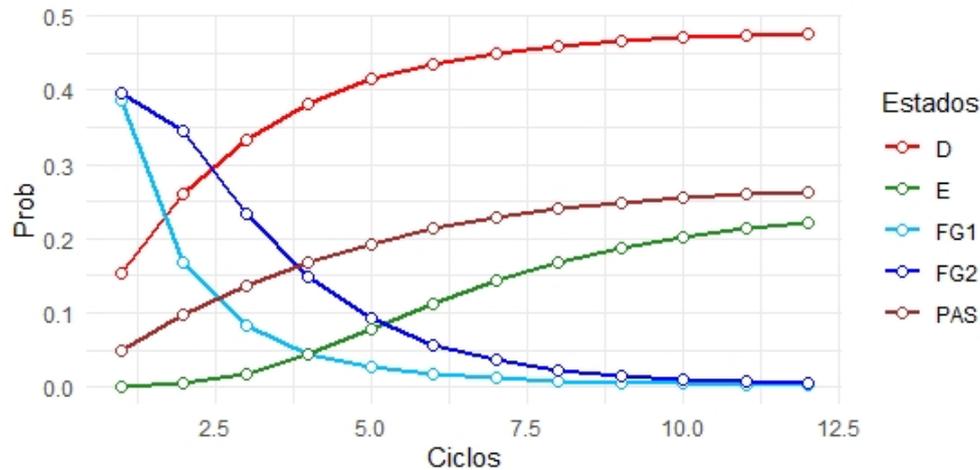


Figura 3.5: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el primer año del campo de la Formación General.

En la Figura 3.4, se puede observar que la probabilidad de abandono de los estudiantes que aprobaron el primer año del Campo de la Formación General es mayor que la de aquellos que aprueban otros espacios curriculares. Esta probabilidad aumenta a medida que pasan los años, llegando a un 48% al final del ciclo lectivo. Sin embargo, cabe destacar que la probabilidad de abandono ha disminuido significativamente en comparación con la tasa de abandono de los estudiantes ingresantes.

También se puede observar que la probabilidad de que los estudiantes completen el segundo año del campo de la Formación General después de transcurrido un año de estudio es mayor. Esta tendencia se mantiene a medida que pasa el tiempo.

En el segundo año, se observa una gran diferencia entre los estudiantes que habiendo aprobado *FG1*, aprueban *FG2* y los que permanecen en *FG1*.

A medida que avanzan los ciclos lectivos, la probabilidad de que los estudiantes soliciten un pase de institución es similar a la probabilidad de que egresen, siendo ligeramente mayor la primera. Cabe destacar que estas dos probabilidades han aumentado en comparación con las probabilidades analizadas anteriormente para los estudiantes ingresantes, y se acercan cada vez más a la probabilidad de abandono de los estudios.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

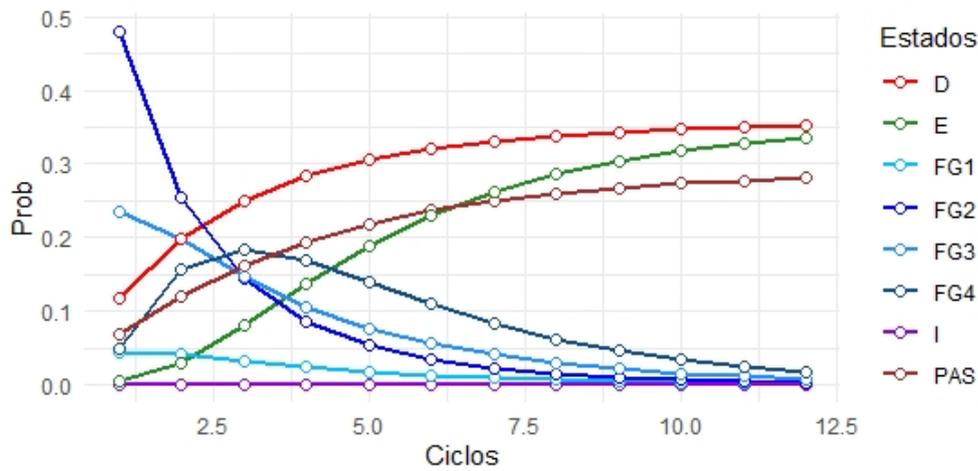


Figura 3.6: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el segundo año del campo de la Formación General.

Si se dejan solamente los estados $FG2$, $FG3$, D , P , E , es decir el estado consecutivo a $FG2$ y los estados absorbentes:

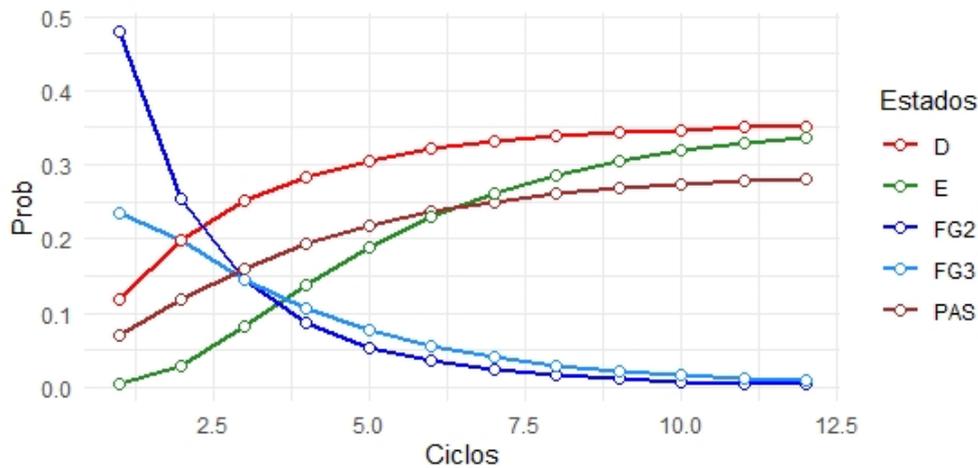


Figura 3.7: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el segundo año del campo de la Formación General.

En la Figura 3.6, se puede observar que la probabilidad de abandono de los estudiantes que aprobaron el segundo año del campo de la Formación General es mayor que la de aquellos que aprueban otros espacios curriculares, y esta probabilidad aumenta a medida que pasan los años, llegando a un 35%.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

A diferencia de las probabilidades de los casos anteriores, esta situación se da a partir de los dos años de este estado. Se aprecia una disminución de esta probabilidad de deserción comparada con la del estudiante que aprobó *FG1*.

Además, es mayor la probabilidad de que transcurrido un año de estudio, permanezca en el mismo estado a corto plazo, situación que se mantiene transcurridos tres años, en donde pasa a ser más probable aprobar el tercer año de la Formación General.

A medida que transcurren los ciclos lectivos, es bastante similar la probabilidad de egresar, la de pedir un pase de institución y la de desertar, siendo ligeramente mayor esta última.

Y cabe mencionar que, las probabilidades de egresar y de pedir un pase, aumentaron respecto de las probabilidades dado que el estudiante aprobó *FG1*, analizada anteriormente, acercándose aún más a la probabilidad de abandonar los estudios.

Transcurridos seis años, la probabilidad de egresar supera ligeramente a la de pedir un pase.

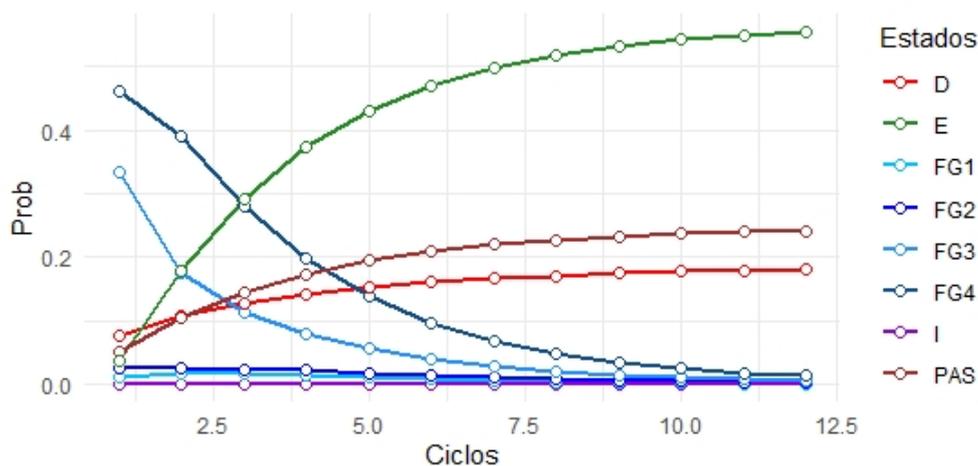


Figura 3.8: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el tercer año del campo de la Formación General.

Si se dejan solamente los estados *FG3*, *FG4*, *D*, *P*, *E*, es decir el estado consecutivo a *FG3* y los estados absorbentes:

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

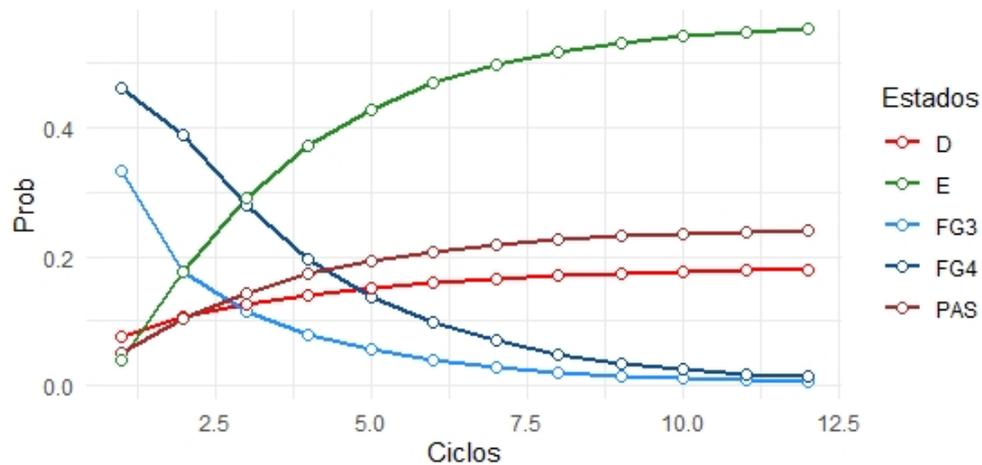


Figura 3.9: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el tercer año del campo de la Formación General.

En la Figura 3.8, se observa que la probabilidad de que el estudiante, que aprobó el tercer año del campo de la Formación General, apruebe el cuarto año del mismo campo es mayor que la de que siga en el mismo estado y, por su parte, la probabilidad de abandonar los estudios decrece notablemente. Se aprecia una disminución importante de esta última probabilidad comparada con la del estudiante que aprobó *FG2*.

Se observa que aumenta considerablemente la probabilidad de egresar, la de cambiarse de institución crece levemente y la de desertar se mantiene por debajo de éstas últimas dos.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

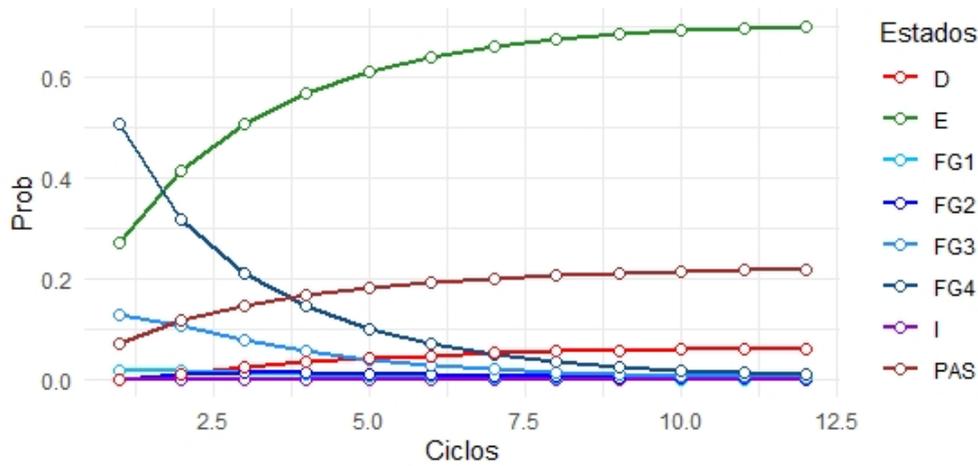


Figura 3.10: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el cuarto año del campo de la Formación General.

Si se dejan solamente los estados $FG4$, D , P , E :

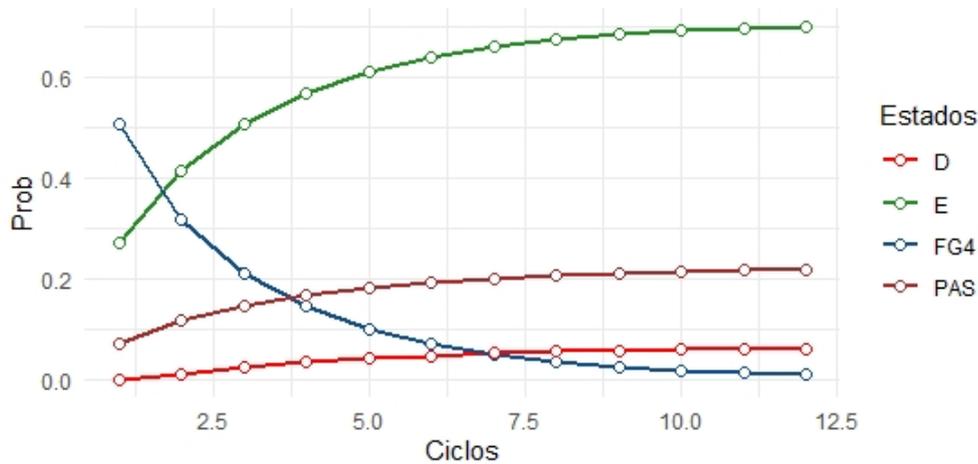


Figura 3.11: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el cuarto año del campo de la Formación General.

Al observar la Figura 3.10, se puede apreciar que la probabilidad de que un estudiante, que ha aprobado el cuarto año del campo de la Formación General, egrese es ligeramente menor que la probabilidad de que permanezca en el mismo campo durante el siguiente año. Sin embargo, a partir del segundo año, esta situación se revierte, incrementándose progresivamente la probabilidad de que el estudiante egrese.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Se aprecia una disminución importante de la probabilidad de desertar comparada con la del estudiante que aprobó *FG3*.

La probabilidad de cambiarse de institución crece levemente, sin embargo es mucho menor que la de que el estudiante egrese.

En síntesis, la mayor probabilidad de desertar se da cuando el alumno está en el estado de ingresante, disminuyendo considerablemente conforme se va asentando en la carrera, aprobando los diferentes campos de la Formación General.

Los campos *FG1* y *FG2* tienen mayor probabilidad de aprobarse durante los primeros años, y se observa un pequeño retardo en el caso de *FG3* y *FG4*.

Se observa también que en el caso de aprobar *FG4* hay un pequeño retraso en llegar al estado de egreso, lo que indicaría que los estudiantes, en general, se reciben con materias de otros campos.

Para obtener las probabilidades de absorción por cada uno de los estados absorbentes *D*, *P*, *E*, así como los tiempos medios de absorción por cada uno de dichos estados, se consideran los espacios $\mathcal{S}_D = \{I, FG1, FG2, FG3, FG4, D\}$, $\mathcal{S}_P = \{I, FG1, FG2, FG3, FG4, P\}$ y $\mathcal{S}_E = \{I, FG1, FG2, FG3, FG4, E\}$.

Debido a que no se registraron estudiantes que hayan desertado después de aprobar *FG4*, se omite este estado en el espacio \mathcal{S}_D . Es decir:

$$\mathcal{S}_D = \{I, FG1, FG2, FG3, D\}$$

Para estimar las matrices de probabilidades de transición se consideraron nuevas bases de datos considerando, por un lado, a los estudiantes que desertaron; por otro a los que pidieron un pase y; finalmente, con los egresados.

Los conteos de transiciones según cada caso, son los siguientes:

Periodo	<i>I</i>	<i>FG1</i>	<i>FG2</i>	<i>FG3</i>	<i>D</i>
<i>I</i>	145	40	1	0	254
<i>FG1</i>	0	32	17	0	29
<i>FG2</i>	0	0	29	4	17
<i>FG3</i>	0	0	0	9	4

Cuadro 3.12: Conteo de transiciones entre estados del campo de la Formación General para el caso de estudiantes que desertaron.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Periodo	<i>I</i>	<i>FG1</i>	<i>FG2</i>	<i>FG3</i>	<i>FG4</i>	<i>P</i>
<i>I</i>	12	7	0	0	0	7
<i>FG1</i>	0	31	18	0	0	6
<i>FG2</i>	0	7	19	7	1	9
<i>FG3</i>	0	0	0	0	0	6
<i>FG4</i>	0	0	0	0	0	4

Cuadro 3.13: Conteo de transiciones entre estados del campo de la Formación General para el caso de estudiantes que solicitaron un pase.

Periodo	<i>I</i>	<i>FG1</i>	<i>FG2</i>	<i>FG3</i>	<i>FG4</i>	<i>E</i>
<i>I</i>	3	8	1	1	0	0
<i>FG1</i>	0	1	11	0	1	0
<i>FG2</i>	0	1	8	20	3	1
<i>FG3</i>	0	0	1	6	24	2
<i>FG4</i>	0	0	0	5	14	14

Cuadro 3.14: Conteo de transiciones entre estados del campo de la Formación General para el caso de estudiantes que egresaron.

Asegurando que se cumplen los supuestos para realizar el ajuste de las cadenas de Markov, se tiene que las respectivas matrices de probabilidades de transición en un paso son:

$$P_D = \begin{pmatrix} 0,3295 & 0,0909 & 0,0023 & 0 & 0,5773 \\ 0 & 0,4103 & 0,2179 & 0 & 0,3718 \\ 0 & 0 & 0,5800 & 0,0800 & 0,3400 \\ 0 & 0 & 0 & 0,6923 & 0,3077 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$P_P = \begin{pmatrix} 0,4616 & 0,2692 & 0 & 0 & 0 & 0,2692 \\ 0 & 0,5636 & 0,3273 & 0 & 0 & 0,1091 \\ 0 & 0,1628 & 0,4419 & 0,1628 & 0,0233 & 0,2093 \\ 0 & 0 & 0 & 0,1667 & 0,5 & 0,3333 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$P_E = \begin{pmatrix} 0,2308 & 0,6154 & 0,0769 & 0,0769 & 0 & 0 \\ 0 & 0,0769 & 0,8462 & 0 & 0,0769 & 0 \\ 0 & 0,0303 & 0,2424 & 0,6061 & 0,0909 & 0,0303 \\ 0 & 0 & 0,0303 & 0,1818 & 0,7273 & 0,0606 \\ 0 & 0 & 0 & 0,1515 & 0,4242 & 0,4242 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Según el Teorema 17, las respectivas matrices fundamentales se detallan a continuación. Debe señalarse que se agrega una columna con la suma de los elementos de cada fila, según el Teorema 21; donde dicha suma aparece disminuida en 1, ya que se había supuesto, para cada ajuste, que el estudiante ingresó en un periodo anterior al verdadero.

	I	FG1	FG2	FG3	(Suma por fila)-1
I	1,49	0,23	0,13	0,03	0,88
FG1	0	1,70	0,88	0,23	1,80
FG2	0	0	2,38	0,62	2,00
FG3	0	0	0	3,25	2,25

Cuadro 3.15: Matriz Fundamental de la cadena del campo de la Formación General para el caso de estudiantes que desertaron.

	I	FG1	FG2	FG3	FG4	(Suma por fila)-1
I	1,85	1,47	0,86	0,17	0,10	3,46
FG1	0	2,93	1,72	0,34	0,21	4,20
FG2	0	0,86	2,29	0,45	0,28	2,87
FG3	0	0	0	1,20	0,60	0,80
FG4	0	0	0	0	1	0

Cuadro 3.16: Matriz Fundamental de la cadena del campo de la Formación General para el caso de estudiantes que solicitaron un pase.

	I	FG1	FG2	FG3	FG4	(Suma por fila)-1
I	1,30	0,91	1,20	1,39	2,07	5,87
FG1	0	1,13	1,31	1,36	2,07	4,86
FG2	0	0,05	1,43	1,44	2,05	3,97
FG3	0	0	0,07	1,66	2,11	2,85
FG4	0	0	0,02	0,44	2,29	1,75

Cuadro 3.17: Matriz Fundamental de la cadena del campo de la Formación General para el caso de estudiantes que egresaron.

Se puede observar que, en promedio, los estudiantes abandonan sus estudios antes de finalizar el primer ciclo lectivo, lo que se refleja en un tiempo medio de deserción de menos de un año desde que el alumno ingresa. Aprobando los espacios curriculares de la Formación General de primero, segundo y tercer año, se esperan que transcurran 2 años para que decida abandonar sus estudios.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

En lo que respecta a los cambios de institución por parte de los ingresantes, se estima que en promedio transcurren 3,46 años antes de que se llegue a esa situación. Si el estudiante logra aprobar el primer año del campo de la Formación General, se estima que pasarán alrededor de 4,2 años antes de que solicite el pase a otra institución. En cambio, si aprueba el segundo año de la Formación General, se espera que la solicitud de pase se presente en casi tres años, y si aprueba el tercer año, se espera que esto sucederá en menos de un año.

Para que un estudiante egrese, se estima que deben transcurrir casi 6 años desde el momento en que ingresa a la institución. A pesar de que no existen retrasos significativos en la aprobación de los años de la Formación General, se observa que, si el estudiante aprueba el cuarto año de esta formación, aún transcurre un tiempo promedio de 1,75 años hasta que finalmente egresa. De esto último se puede concluir que, en general, los estudiantes aún deben aprobar espacios curriculares de otros campos, después de completar la Formación General, para poder graduarse.

A continuación, se calcula el producto entre la matriz fundamental y la submatriz de probabilidades de transición en un paso desde estados transitorios a recurrentes, con el objetivo de obtener una matriz de acceder eventualmente a los estados absorbentes desde uno transitorio, de acuerdo con el Teorema 17.

	D	P	E
I	0.8168	0.1007	0,0824
FG1	0,4832	0,2722	0,2444
FG2	0,3567	0,2881	0,3552
FG3	0,1837	0,2471	0,5692
FG4	0,0656	0,2222	0,7121

Cuadro 3.18: Matriz de acceso eventual a los estados absorbentes desde uno transitorio, para el campo de la Formación General.

Según los resultados obtenidos, es evidente que el alumno ingresante enfrenta altas probabilidades de abandonar la carrera. Concretamente, existe un 81,68 % de probabilidad de deserción, un 10,07 % de cambiar de institución y solo un 8,24 % de egresar. No obstante, aquellos estudiantes que logran aprobar el primer año del campo de la Formación General, presentan una disminución significativa en la probabilidad eventual de abandono, pasando a un 48,32 %, un 27,22 % de cambiarse de institución y un 24,44 % de egresar. Si bien aún hay margen de mejora, aprobar el segundo año del campo de la Formación General disminuye la probabilidad de deserción eventual a un 35,67 %, pero aumenta la probabilidad de cambiar de institución a un 28,81 % y la de egresar también se eleva a un 35,52 %. Sin embargo, si el estudiante logra aprobar el tercer año de este campo, la probabilidad de

abandonar la carrera disminuye a un 18,37 %, mientras que la probabilidad de pedir un pase aumenta a un 24,71 %, y la de egresar asciende al 56,92 %. Finalmente, aquellos que aprueban el cuarto año tienen una probabilidad de deserción eventual de sólo el 6,56 %, de pedir un pase del 22,22 % y de egresar del 71,21 %.

Aunque las probabilidades de egresar mejoran a medida que el estudiante avanza en los campos de la Formación General, es importante tener en cuenta que estos resultados no reflejan el tiempo promedio que se necesita para alcanzar cada estado. Además, es interesante notar que, la probabilidad de egresar eventualmente aumenta después de que el estudiante aprueba el primer año del campo de la Formación General.

3.4.2. Transición por el campo de la Formación Específica

En esta subsección se ajusta otro modelo basado en cadenas de Markov a tiempo discreto, en el cual cada ciclo lectivo se considera como un paso, y el espacio de estados:

$$\mathcal{S} = \{I, FE1, FE2, FE3, FE4, D, P, E\}$$

Como se puede observar, en este espacio de estados se incluyen los espacios curriculares correspondientes al campo de la Formación Específica, organizados por año. De esta forma, es posible identificar a partir de cuál campo de estudio se incrementa la probabilidad de alcanzar los estados absorbentes.

Las transiciones posibles son las siguientes:

- $I \rightarrow I$: el estudiante ingresante no aprobó todos los espacios curriculares de ningún FEi ($i = 1, 2, 3, 4$), durante el ciclo lectivo, permaneciendo en el mismo estado en el siguiente ciclo.
- $I \rightarrow FEi$: el estudiante ingresante aprueba todos los espacios curriculares que conforman FEi ($i = 1, 2, 3, 4$), pasando al estado FEi en el siguiente ciclo lectivo. (Debe recordarse que tal como están planteadas las correlatividades puede darse esta situación para cualquier $i = 1, 2, 3, 4$).
- $I \rightarrow D$: el estudiante ingresante abandona sus estudios en esta institución sin aprobar ningún FEi , pasando al estado de *deserción* en el siguiente ciclo lectivo.
- $I \rightarrow P$: el estudiante ingresante pide un pase a otra institución para continuar la misma carrera, sin haber aprobado ningún FEi .
- $FEi \rightarrow FEi$: el alumno que aprobó todos los espacios curriculares de FEi , y continúa en la misma condición durante el ciclo lectivo, permanece en el mismo estado en el siguiente ciclo.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

- $FEi \rightarrow FEj$ ($i \neq j$): el alumno que aprobó todos los espacios curriculares de FEi , aprueba todos los espacios curriculares de FEj durante el ciclo lectivo, pasando al estado FEj en el siguiente ciclo.
- $FEi \rightarrow D$: el estudiante, habiendo aprobado FEi , abandona sus estudios en esta institución, pasando al estado de *deserción* en el siguiente ciclo lectivo.
- $FEi \rightarrow P$: el estudiante pide un pase a otra institución para continuar la misma carrera, habiendo aprobado FEi , en el ciclo anterior.
- $FEi \rightarrow E$: el estudiante termina de aprobar todos los espacios curriculares del plan de estudios en el ciclo lectivo y pasa al estado de *egreso* en el siguiente ciclo.

Las transiciones no permitidas entre dos ciclos lectivos consecutivos son:

- $I \rightarrow E$: el estudiante ingresante no puede egresar en el ciclo lectivo siguiente, puesto que por la estructura del plan de estudios, debe permanecer en la carrera al menos 4 ciclos, desde que ingresa.
- $FEi \rightarrow I$: el estudiante ya aprobó algún FEi del plan de estudios y es imposible que vuelva al estado de ingresante.
- $D \rightarrow I$: no se registraron casos en que el estudiante haya abandonado la carrera y haya regresado en el siguiente ciclo lectivo para comenzar nuevamente desde el inicio. Si se diera ese caso, se reconocería el trayecto previo del estudiante por medio de equivalencias.
- $P \rightarrow I$: no se registraron casos en que el estudiante se haya cambiado de institución y haya regresado en el siguiente ciclo lectivo, ni en ningún otro de los considerados, para comenzar nuevamente desde el inicio. Si se diera ese caso, se reconocería el trayecto previo del estudiante por medio de equivalencias.
- $E \rightarrow I$: el estudiante que egresó no puede volver al estado de ingresante, salvo que desee cursar nuevamente la carrera desde el inicio.
- $D \rightarrow FEi$: no se registró que ningún estudiante que haya abandonado sus estudios en un ciclo lectivo, los haya retomado en el siguiente ciclo.
- $P \rightarrow FEi$: no se registró que ningún estudiante que se haya cambiado de institución en un ciclo lectivo, haya regresado a esta en el siguiente ciclo.
- $E \rightarrow FEi$: el estudiante que egresó en un ciclo lectivo, no regresa al estado de estudiante regular en el siguiente ciclo lectivo.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

- Las transiciones entre estados absorbentes tampoco son posibles. Las que sí se podrían definir (transformando a los estados en no absorbentes) son $D \rightarrow E$ y $P \rightarrow E$, de las cuales no hay registro de casos.

Para plantear este nuevo modelo, se realizó un procedimiento similar al anterior, para probar los supuestos de estacionariedad y la propiedad de Markov. Se trabajó sobre los periodos comprendidos entre 2014 y 2019, como se mencionó con anterioridad.

Para que se cumpla la propiedad de estacionariedad, se requiere homogeneidad de las probabilidades de transición durante todos los periodos. Para verificar esta propiedad se debe asegurar que la frecuencia de estudiantes en cada uno de los estados no sea dependiente del periodo en el cual se toma la medición. Para esto se utilizó una prueba de independencia mediante el test exacto de Fisher.

En las siguientes tablas se presentan las transiciones entre estados para dos periodos consecutivos.

2014	<i>I</i>	<i>FE1</i>	<i>FE2</i>	<i>FE3</i>	<i>FE4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	94	11	0	0	0	54	5	0
<i>FE1</i>	0	18	2	0	0	0	0	0
<i>FE2</i>	0	0	2	0	0	0	0	0
<i>FE3</i>	0	0	0	0	0	0	0	0
<i>FE4</i>	0	0	0	0	0	0	0	0

2015	<i>I</i>	<i>FE1</i>	<i>FE2</i>	<i>FE3</i>	<i>FE4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	101	18	0	0	0	54	5	0
<i>FE1</i>	0	21	3	0	0	1	4	0
<i>FE2</i>	0	0	3	1	1	0	0	0
<i>FE3</i>	0	0	0	0	0	0	0	0
<i>FE4</i>	0	0	0	0	0	0	0	0

2016	<i>I</i>	<i>FE1</i>	<i>FE2</i>	<i>FE3</i>	<i>FE4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	93	16	0	0	0	55	2	0
<i>FE1</i>	0	27	7	3	1	3	2	0
<i>FE2</i>	0	0	0	5	3	0	0	0
<i>FE3</i>	0	0	0	0	0	0	0	1
<i>FE4</i>	0	0	0	0	0	0	0	1

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

2017	<i>I</i>	<i>FE1</i>	<i>FE2</i>	<i>FE3</i>	<i>FE4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	94	10	0	0	0	66	2	0
<i>FE1</i>	0	31	5	0	0	6	1	0
<i>FE2</i>	0	0	2	1	3	0	1	1
<i>FE3</i>	0	0	0	0	4	0	0	5
<i>FE4</i>	0	0	0	0	0	0	0	5

2018	<i>I</i>	<i>FE1</i>	<i>FE2</i>	<i>FE3</i>	<i>FE4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	102	14	5	0	0	73	5	0
<i>FE1</i>	0	23	8	1	1	3	6	0
<i>FE2</i>	0	0	4	3	1	0	0	0
<i>FE3</i>	0	0	0	0	0	0	0	1
<i>FE4</i>	0	0	0	0	0	0	0	5

2019	<i>I</i>	<i>FE1</i>	<i>FE2</i>	<i>FE3</i>	<i>FE4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	92	18	1	0	0	76	2	0
<i>FE1</i>	0	19	8	2	1	6	4	0
<i>FE2</i>	0	0	4	6	6	0	2	0
<i>FE3</i>	0	0	1	0	2	0	0	1
<i>FE4</i>	0	0	0	0	1	0	0	1

Cuadro 3.19: Transiciones entre estados por año - Campo de la Formación Específica.

En consecuencia, las tablas que muestran las frecuencias de transiciones entre un estado transitorio determinado y todos los demás estados, para cada periodo, son las siguientes:

Periodo	<i>I</i>	<i>FE1</i>	<i>FE2</i>	<i>FE3</i>	<i>FE4</i>	<i>D</i>	<i>P</i>	<i>E</i>
2014	94	11	0	0	0	54	5	0
2015	101	18	0	0	0	66	7	0
2016	93	16	0	0	0	55	2	0
2017	94	10	0	0	0	66	2	0
2018	102	14	0	0	0	73	5	0
2019	92	18	1	0	0	76	2	0

Cuadro 3.20: Frecuencia de transiciones entre el estado *I* y todos los estados para cada periodo.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Periodo	<i>I</i>	<i>FE1</i>	<i>FE2</i>	<i>FE3</i>	<i>FE4</i>	<i>D</i>	<i>P</i>	<i>E</i>
2014	0	18	2	0	0	0	0	0
2015	0	21	3	1	0	1	4	0
2016	0	27	7	3	1	3	2	0
2017	0	31	5	0	0	6	1	0
2018	0	23	8	1	1	3	6	0
2019	0	19	8	2	1	6	4	0

Cuadro 3.21: Frecuencia de transiciones entre el estado *FE1* y todos los estados para cada periodo.

Periodo	<i>I</i>	<i>FE1</i>	<i>FE2</i>	<i>FE3</i>	<i>FE4</i>	<i>D</i>	<i>P</i>	<i>E</i>
2014	0	0	2	0	0	0	0	0
2015	0	0	3	1	1	0	0	0
2016	0	0	0	5	3	0	0	0
2017	0	0	2	1	3	0	1	1
2018	0	0	4	3	1	0	0	0
2019	0	0	4	6	6	0	2	0

Cuadro 3.22: Frecuencia de transiciones entre el estado *FE2* y todos los estados para cada periodo.

Periodo	<i>I</i>	<i>FE1</i>	<i>FE2</i>	<i>FE3</i>	<i>FE4</i>	<i>D</i>	<i>P</i>	<i>E</i>
2014	0	0	0	0	0	0	0	0
2015	0	0	0	0	0	0	0	0
2016	0	0	0	0	0	0	0	1
2017	0	0	0	0	4	0	0	5
2018	0	0	0	0	0	0	0	1
2019	0	0	1	0	2	0	0	1

Cuadro 3.23: Frecuencia de transiciones entre el estado *FE3* y todos los estados para cada periodo.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Periodo	I	$FE1$	$FE2$	$FE3$	$FE4$	D	P	E
2014	0	0	0	0	0	0	0	0
2015	0	0	0	0	0	0	0	0
2016	0	0	0	0	0	0	0	1
2017	0	0	0	0	0	0	0	5
2018	0	0	0	0	0	0	0	5
2019	0	0	0	0	1	0	0	1

Cuadro 3.24: Frecuencia de transiciones entre el estado $FE4$ y todos los estados para cada periodo.

Aplicando el test exacto de Fisher, en las tablas de contingencia 3.20 a 3.24, para distintos grupos de años consecutivos, se obtiene que para el periodo 2014-2019, las frecuencias de transiciones son estadísticamente independientes del periodo.

Estado	p -valor
I	0,6972
$FE1$	0,2229
$FE2$	0,2859
$FE3$	0,5287
$FE4$	0,2344

Cuadro 3.25: Resultados de la prueba de independencia entre periodos y transiciones (2014-2019).

En este caso, si se desean obtener estimaciones a largo plazo, del comportamiento de las transiciones, debe tenerse en cuenta el periodo 2014-2019.

Otro supuesto a asegurar es que, las frecuencias de transición obtenidas cumplan con la principal característica de las cadenas de Markov, deben ser dependientes únicamente de su estado anterior. Para verificar este supuesto, se realizó el test exacto de Fisher en las tablas de contingencia para los años que pertenecen al periodo 2014-2019, obteniendo en todos los casos un p -valor de 0,0005. Los resultados de la prueba indican que se rechaza la hipótesis nula entre la independencia de estados, por lo que se puede garantizar que los resultados obtenidos en un estado, dependen del estado en el periodo anterior.

Se estiman las matrices de transición P , utilizando el estimador de máxima verosimilitud del resultado del Teorema 2.

$$\hat{P}_{MV} = \frac{n_{xy}}{\sum_{z \in \mathcal{S}} n_{xz}}$$

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

análogamente a como se trabajó con el primer modelo.

Los resultados de dichas estimaciones, se muestran en la siguiente tabla:

Estados	<i>I</i>	<i>FE1</i>	<i>FE2</i>	<i>FE3</i>	<i>FE4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	0,5348	0,0808	0,0009	0,0000	0,0000	0,3621	0,0214	0,0000
<i>FE1</i>	0,0000	0,6406	0,1521	0,0276	0,0138	0,0876	0,0783	0,0000
<i>FE2</i>	0,0000	0,0000	0,3061	0,3265	0,2857	0,0000	0,0612	0,0204
<i>FE3</i>	0,0000	0,0000	0,0667	0,0000	0,4000	0,0000	0,0000	0,5333
<i>FE4</i>	0,0000	0,0000	0,0000	0,0000	0,0769	0,0000	0,0000	0,9231
<i>D</i>	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000
<i>P</i>	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000
<i>E</i>	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000

Cuadro 3.26: Matriz de transición - Campo de la Formación Específica.

En el siguiente gráfico pueden observarse las probabilidades de transición en un año:

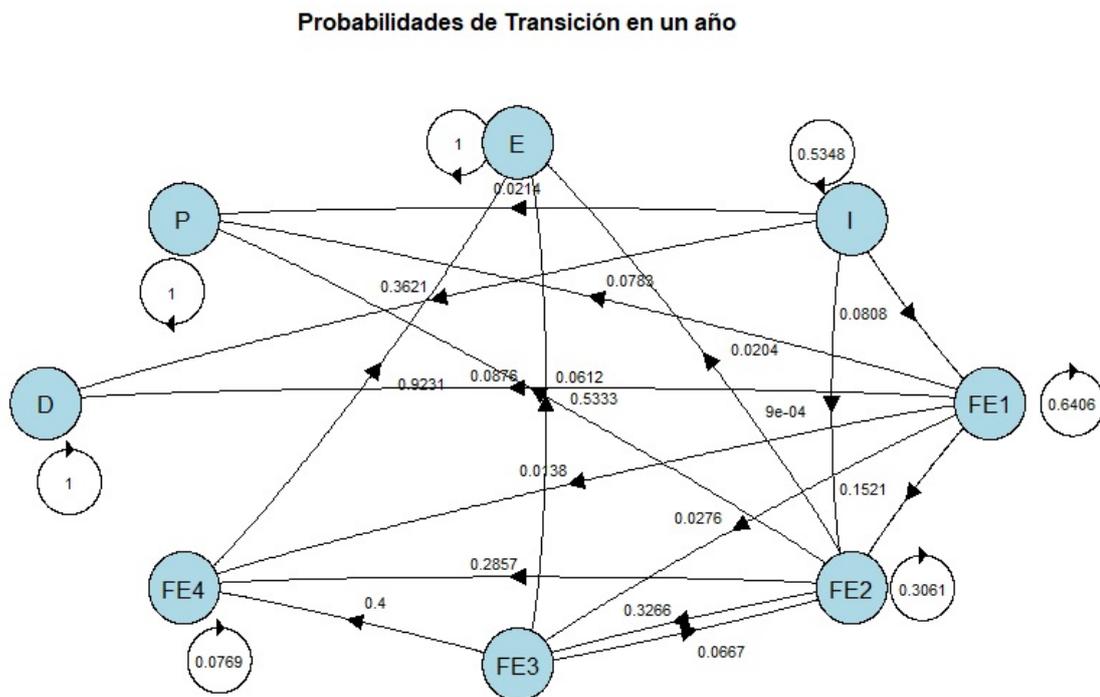


Figura 3.12: Probabilidades de transición en un año - Campo de la Formación Específica.

Se pueden obtener las probabilidades de transición en n pasos, o más específi-

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

camente, en n ciclos lectivos, calculando las potencias n -ésimas de la matriz de transición P .

Interesa conocer entonces, ¿cuál es la probabilidad de que un estudiante que ingresa, egrese en los 4 años propuestos en el plan de estudios de la carrera? y, ¿cuál es la probabilidad de llegar a algún estado absorbente, estando en otros estados de la carrera, luego de 4 años?.

Para ello, se calcula P^4 :

Estados	I	$FE1$	$FE2$	$FE3$	$FE4$	D	P	E
I	0,0818	0,0661	0,0191	0,0084	0,0009	0,7375	0,0646	0,0126
$FE1$	0	0,1684	0,0781	0,0437	0,0651	0,2027	0,2062	0,2357
$FE2$	0	0	0,0154	0,0137	0,0341	0	0,0895	0,8472
$FE3$	0	0	0,0028	0,0025	0,0062	0	0,0058	0,9825
$FE4$	0	0	0	0	0	0	0	1

Cuadro 3.27: Matriz de transición en 4 pasos - Transición por el campo de la Formación Específica.

En este caso, se puede apreciar que la probabilidad de que un estudiante finalice la carrera en los cuatro años es muy reducida, con un valor del 1,26 %. En cambio, la probabilidad de cambiar de institución es algo mayor, con un 6,46 %, mientras que la probabilidad de abandonar la carrera es significativamente alta, con un valor del 73,75 %.

En cuanto a los estudiantes que han aprobado los espacios curriculares del primer año del campo de la Formación Específica, se puede observar que la probabilidad de egresar aumenta hasta un 23,57 %, lo cual sugiere que es poco probable que el estudiante logre finalizar la carrera en los cuatro años establecidos. Por otro lado, la probabilidad de abandonar los estudios disminuye significativamente a un 20,27 %, mientras que la probabilidad de cambiar de institución se eleva a un 20,62 %.

En lo que respecta a los estudiantes que han aprobado los espacios curriculares del segundo año del campo de la Formación Específica, se puede apreciar un incremento significativo en la probabilidad de egresar, llegando a un valor de 84,72 %. Cabe destacar que la probabilidad de abandonar los estudios se reduce a cero, mientras que la probabilidad de cambiar de institución disminuye nuevamente a un valor de 8,95 %.

En relación a los estudiantes que han aprobado los espacios curriculares del tercer año del campo de la Formación Específica, se puede observar un aumento significativo en la probabilidad de egresar, alcanzando un valor de 98,25 %. Es importante señalar que la probabilidad de abandonar los estudios sigue siendo nula y la probabilidad de cambiar de institución continúa disminuyendo, llegando

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

a un valor de 0,58 %.

Los resultados muestran que para los estudiantes que han aprobado los espacios curriculares del cuarto año del campo de la Formación Específica, la probabilidad de egresar es del 100 %. Es importante destacar que tanto la probabilidad de abandonar los estudios como la probabilidad de cambiarse de institución son nulas en este caso, lo que indica que los estudiantes que han alcanzado este nivel tienen una alta probabilidad de completar exitosamente su carrera.

Se puede decir que conforme los estudiantes van aprobando los espacios curriculares del campo de la Formación Específica por año, van consolidando su avance en la carrera, aumenta la probabilidad de egreso y disminuye la de deserción. También se observa que aprobando los campos de la Formación Específica de segundo año, se puede asegurar la no deserción y se aprecia una disminución en la probabilidad de cambiarse de institución.

Al comparar estos resultados con los del modelo anterior, se podría conjeturar que los estudiantes suelen aprobar primero las materias pertenecientes al campo de la Formación General. En cambio, se tardan más en aprobar los espacios pertenecientes al campo de la Formación Específica durante los dos primeros años.

El siguiente cuadro presenta las probabilidades de transición en más de 4 ciclos lectivos, lo cual permite visualizar las transiciones más probables en el campo de la Formación Específica. Asimismo, se presentan las probabilidades de llegar a algún estado absorbente, lo que brinda información relevante acerca de la trayectoria de los estudiantes en este campo de estudio.

5 Ciclos Lectivos								
	I	FE1	FE2	FE3	FE4	D	P	E
I	0,0437	0,0489	0,0165	0,0080	0,0105	0,7728	0,0726	0,0266
FE1	0	0,1078	0,0524	0,0301	0,0471	0,2174	0,2241	0,3207
FE2	0	0	0,0056	0,0050	0,0125	0	0,0905	0,8863
FE3	0	0	0,0010	0,0009	0,0022	0	0,0059	0,9897
FE4	0	0	0	0	0	0	0	1

6 Ciclos Lectivos								
	I	FE1	FE2	FE3	FE4	D	P	E
I	0,0233	0,0349	0,0131	0,0067	0,0094	0,7929	0,0785	0,0409
FE1	0	0,0691	0,0344	0,0200	0,0321	0,2268	0,2358	0,3814
FE2	0	0	0,0020	0,0018	0,0045	0	0,0908	0,9006
FE3	0	0	0,0003	0,0003	0,0008	0	0,0060	0,9999
FE4	0	0	0	0	0	0	0	1

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

7 Ciclos Lectivos								
	I	FE1	FE2	FE3	FE4	D	P	E
I	0,0125	0,0242	0,0097	0,0052	0,0076	0,8045	0,0825	0,0535
FE1	0	0,0442	0,0224	0,0131	0,0213	0,2329	0,2433	0,4225
FE2	0	0	0,0007	0,0006	0,0016	0	0,0909	0,9059
FE3	0	0	0,0001	0,0001	0,0003	0	0,0060	0,9934
FE4	0	0	0	0	0	0	0	1

8 Ciclos Lectivos								
	I	FE1	FE2	FE3	FE4	D	P	E
I	0,0067	0,0165	0,0070	0,0038	0,0058	0,8111	0,0852	0,0635
FE1	0	0,0283	0,0144	0,0085	0,0139	0,2368	0,2482	0,4496
FE2	0	0	0,0002	0,0002	0,0006	0	0,0910	0,9078
FE3	0	0	0,0001	0,0000	0,0001	0	0,0061	0,9937
FE4	0	0	0	0	0	0	0	1

Cuadro 3.28: Matrices de transición en más de 4 pasos - Campo de la Formación Específica.

En los siguientes gráficos se puede apreciar la convergencia de las probabilidades de transición a lo largo de 12 años lectivos para cada uno de los estados futuros, partiendo de un estado actual determinado. Estos gráficos ofrecen información valiosa acerca de la evolución temporal de los distintos estados y su probabilidad de ocurrencia en el futuro.

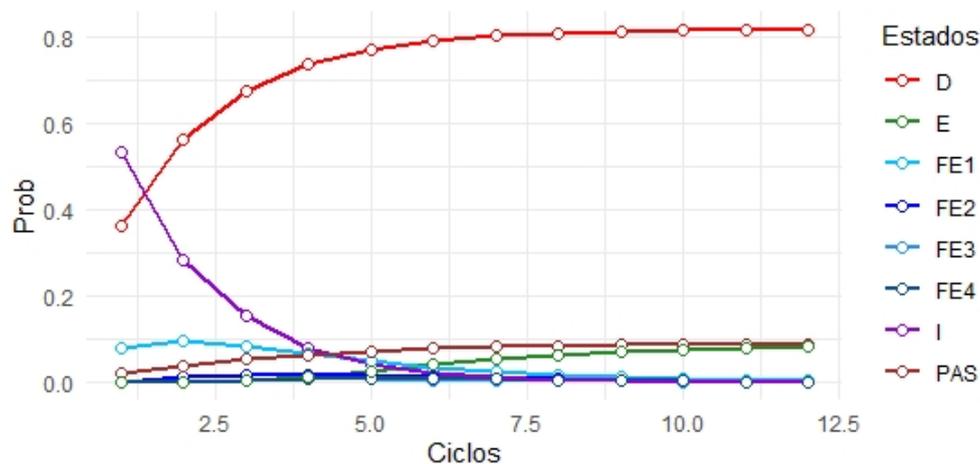


Figura 3.13: Probabilidades de transición por año lectivo, dado que actualmente es Ingresante.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Si se dejan solamente los estados I , $FE1$, D , P , E , es decir el estado consecutivo a I y los estados absorbentes:

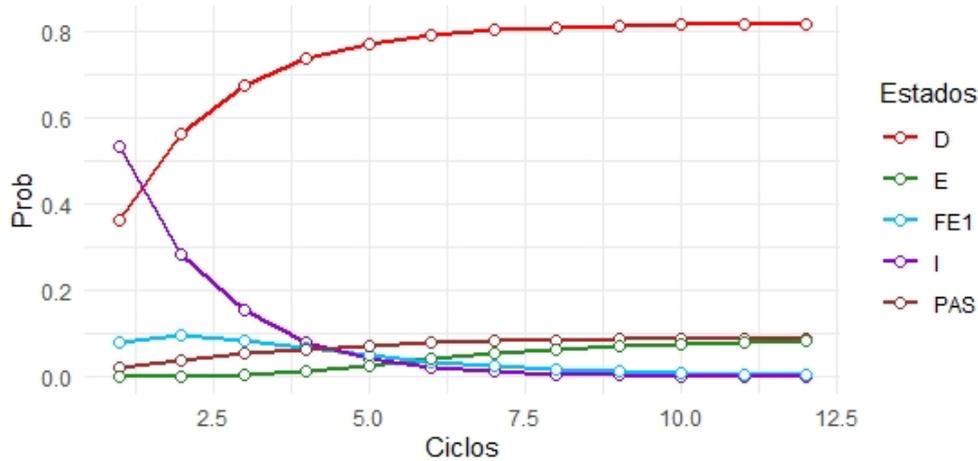


Figura 3.14: Probabilidades de transición por año lectivo, dado que actualmente es Ingresante.

En la Figura 3.13, se puede observar que la probabilidad de abandono de los estudios por parte de los estudiantes ingresantes es significativamente mayor que la probabilidad de aprobación de espacios curriculares pertenecientes al campo de la Formación Específica. Además, esta probabilidad de abandono tiende a aumentar a medida que pasan los años, llegando a alcanzar el 82 %.

La probabilidad de que un estudiante no complete el primer año del campo de la Formación Específica, después de haber transcurrido el primer año de estudios, es mayor que la de avanzar en dicho campo, lo que lo mantendría en el estado de *Ingresante*. No obstante, esta situación se revierte después de transcurridos cinco años, ya que la probabilidad de aprobar el primer año de la Formación Específica es ligeramente superior a la probabilidad de permanecer en el estado de *Ingresante*.

A medida que transcurren los ciclos lectivos, se observa que la probabilidad de solicitar un pase de institución es ligeramente mayor que la probabilidad de egresar. Sin embargo, la diferencia entre ambas probabilidades es relativamente pequeña.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

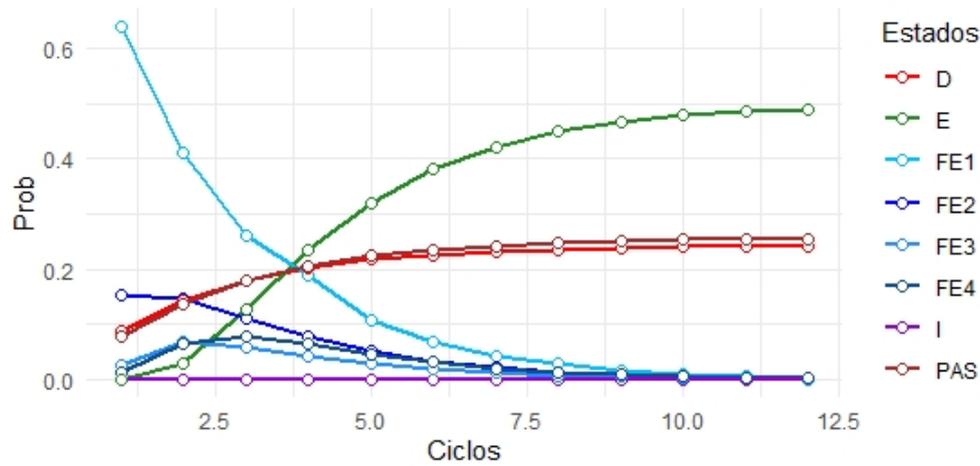


Figura 3.15: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el primer año del campo de la Formación Específica.

Si se dejan solamente los estados $FE1$, $FE2$, D , P , E , es decir el estado consecutivo a $FE1$ y los estados absorbentes:

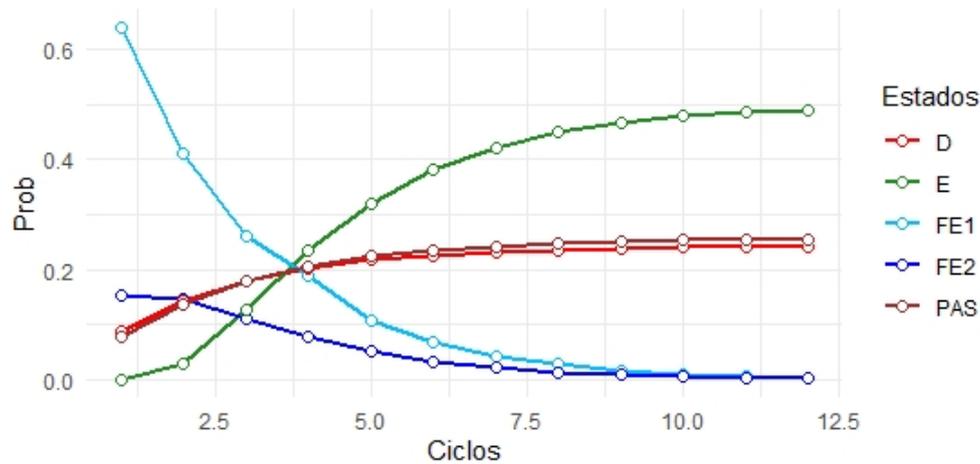


Figura 3.16: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el primer año del campo de la Formación Específica.

En la Figura 3.15, se puede observar que la probabilidad de abandono de los estudios, por parte de un estudiante que ha aprobado el primer año del campo de la Formación Específica, disminuye significativamente en comparación con la de los estudiantes que no han completado dicho campo. Sin embargo, a medida que transcurren los años, esta probabilidad tiende a aumentar ligeramente, llegando a

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

ser similar a la probabilidad de cambiar de institución, la cual también tiende al 25 %.

La probabilidad de mantener aprobado el primer año del Campo de la Formación Específica es significativamente mayor que la de aprobar el segundo año de este campo. A medida que pasa el tiempo, ambas probabilidades disminuyen hasta volverse similares. Cabe destacar que la probabilidad de mantener aprobado solo el primer año del campo de la Formación Específica decrece rápidamente a medida que transcurren los años. Al comparar esta información con el gráfico 3.13, se puede afirmar que tanto los espacios curriculares del primer año como los del segundo año tienen una baja probabilidad de ser aprobados, lo cual puede ser la causa de retrasos, abandono o la solicitud de un pase a otra institución.

En cuanto a la probabilidad de abandonar los estudios o cambiar de institución, se observa que estas son ligeramente menores que la probabilidad de quedarse solo con el primer año del campo de la Formación Específica aprobado y de aprobar el segundo año. Sin embargo, a partir del segundo año, la probabilidad de deserción y cambio de institución supera la probabilidad de aprobar el segundo año de este campo. Además, a partir de los cuatro años, también supera la probabilidad de mantener aprobado solo el primer año del campo.

No obstante, si el estudiante cuenta solamente con el primer año del campo de la Formación Específica aprobado, la probabilidad de egresar supera a la probabilidad de transitar a cualquier otro estado a partir del cuarto año. Esta probabilidad aumenta significativamente y se aproxima cada vez más a un 50 %.

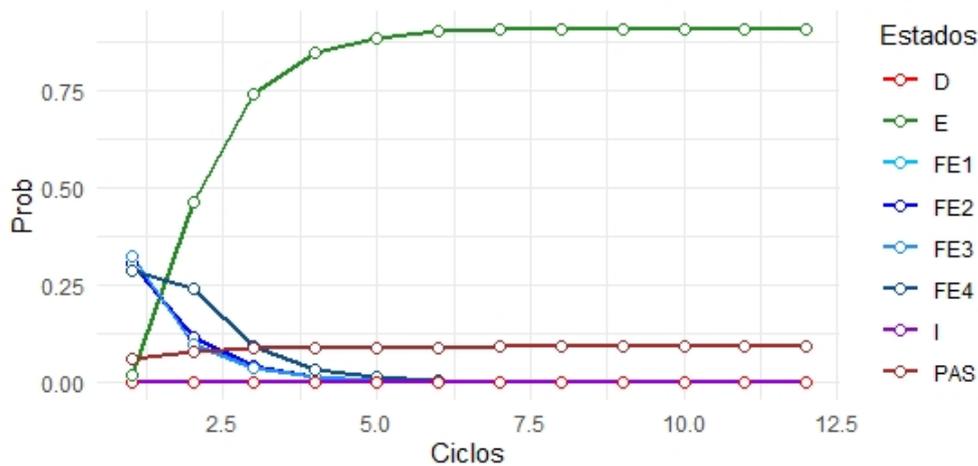


Figura 3.17: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el segundo año del campo de la Formación Específica.

Si se dejan solamente los estados $FE2$, $FE3$, D , P , E , es decir el estado

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

consecutivo a $FE2$ y los estados absorbentes:

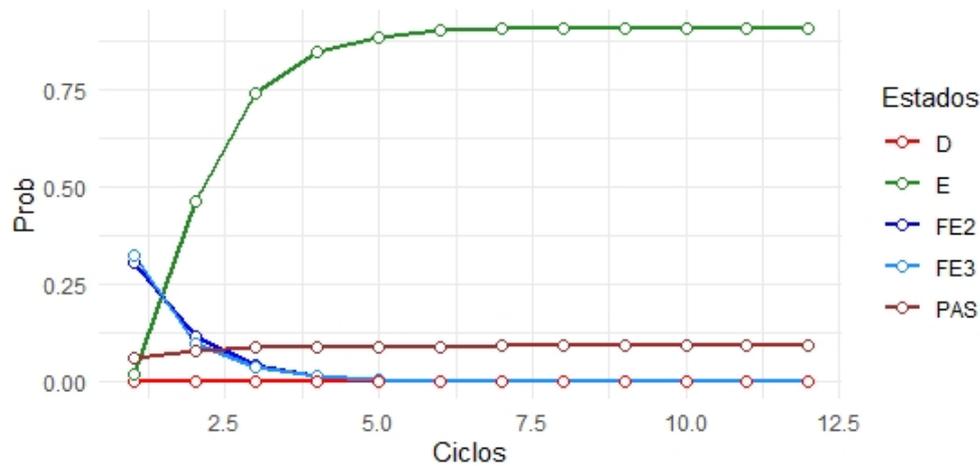


Figura 3.18: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el segundo año del campo de la Formación Específica.

En la figura 3.17, se puede apreciar que la tasa de abandono de los estudiantes que aprobaron el segundo año del campo de la Formación Específica se reduce a 0, mientras que la probabilidad de cambio de institución se mantiene prácticamente constante, aumentando ligeramente a medida que transcurren los años y estabilizándose en torno al 9%.

Se observa en los resultados que la probabilidad de aprobar el tercer año de la Formación Específica es prácticamente idéntica a la de mantener el segundo año aprobado. Esta situación se mantiene constante a medida que transcurre el tiempo, y ambas probabilidades disminuyen gradualmente hasta acercarse a cero.

En cuanto a la probabilidad de cambio de institución, se observa que, después de transcurridos dos años, se sitúa ligeramente por debajo de la de mantener aprobado el segundo año o de aprobar el tercer año del mismo campo. No obstante, a partir del tercer año, la probabilidad de cambio de institución supera a la de mantener aprobado el segundo año, sugiriendo que algunos estudiantes que no lograron avanzar deciden cambiar de institución.

No obstante, si el estudiante ha aprobado el segundo año de la Formación Específica, a partir de los 2 años, la probabilidad de graduarse supera la probabilidad de cualquier otro estado y aumenta significativamente, acercándose progresivamente al 91%. A pesar de que los estudiantes adeudan materias específicas del segundo año, esto sugiere que muchos de ellos optan por rendir esas materias hacia el final de su trayectoria, en un período más corto.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS



Figura 3.19: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el tercer año del campo de la Formación Específica.

Si se dejan solamente los estados $FE3$, $FE4$, D , P , E , es decir el estado consecutivo a $FE3$ y los estados absorbentes:



Figura 3.20: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el tercer año del campo de la Formación Específica.

En la Figura 3.19, se puede apreciar que la probabilidad de abandono de los estudiantes que aprobaron el tercer año del campo de la Formación Específica es nula. Además, se observa que la probabilidad de cambio de institución permanece prácticamente constante, manteniéndose lo suficientemente cercana a cero y aumentando ligeramente a medida que transcurren los años.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Se observa una diferencia significativa entre la probabilidad de aprobar el cuarto año de la Formación Específica y la de mantener aprobado el tercer año del mismo campo. Esta tendencia se mantiene a lo largo del tiempo y ambas probabilidades disminuyen gradualmente hasta acercarse lo suficiente a cero.

En cuanto a la probabilidad de cambio de institución, se puede observar que esta es ligeramente inferior a la probabilidad de mantener aprobado el tercer año del campo de la Formación Específica o de aprobar el cuarto año. Sin embargo, a partir de los tres años, la probabilidad de cambio de institución supera la de mantener aprobado el tercer año, lo que sugiere que algunos estudiantes que no logran aprobar optan por cambiar de institución. A partir de los cuatro años, la probabilidad de cambio de institución supera ligeramente la probabilidad de aprobar el cuarto año, pero sigue siendo prácticamente nula.

Sin embargo, si un estudiante tiene aprobado el tercer año, a partir de los 2 años la probabilidad de egresar de la institución supera la probabilidad de pasar a cualquier otro estado y aumenta considerablemente, aproximándose cada vez más al 100%. Esto sugiere que la mayoría de los estudiantes que adeudan materias específicas del campo las rinden hacia el final de su trayectoria, en un tiempo más acotado. Cabe destacar que, aunque solo se hayan aprobado materias del tercer año, el estudiante tiene altas probabilidades de egresar.



Figura 3.21: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobado el cuarto año del campo de la Formación Específica.

La Figura 3.21 muestra que la probabilidad de abandono o cambio de institución para los estudiantes que aprobaron el cuarto año del campo de la Formación Específica es de 0, lo que sugiere que la gran mayoría de los estudiantes que llegan a esta etapa logran completar sus estudios en la misma institución donde los

comenzaron.

Es bastante mayor la probabilidad de egresar (casi 100 %) que la de permanecer en el estado de aprobación del cuarto año de la Formación Específica. Esto sugiere que la mayoría de los estudiantes que completan el cuarto año específico, egresan del programa. Sin embargo, hay un pequeño porcentaje (alrededor del 8 %) que permanece en este estado de aprobación. Este porcentaje disminuye a prácticamente cero después de dos años, lo que podría indicar que estos estudiantes están esperando aprobar materias de otros campos o del mismo campo de años anteriores.

En resumen, la probabilidad de abandono disminuye significativamente a medida que el estudiante avanza en su carrera y aprueba los diferentes campos de la Formación Específica. Los campos FE1 y FE2 tienen una baja probabilidad de aprobación durante los primeros años, lo que provoca un gran retraso en estos estados. Una vez aprobado el primer año de la Formación Específica, la probabilidad de egreso es mayor que la de cambiarse de institución o abandonar los estudios. Sin embargo, es importante prestar atención a la transición de la aprobación del primer año hacia la aprobación del segundo año.

Una vez que el estudiante aprueba las materias del campo de Formación Específica del segundo año, la probabilidad de desertar disminuye significativamente. Sin embargo, existe una probabilidad baja de que el estudiante se cambie de institución. Al aprobar las materias del tercer y cuarto año de este campo, el egreso del estudiante se vuelve prácticamente seguro. Por lo tanto, se puede concluir que la transición del segundo al tercer año es fundamental para evitar la deserción y enfocarse en el éxito académico.

Tanto el tercer como el cuarto año de esta carrera tienen altas probabilidades de ser aprobados en un corto plazo, lo que sugiere que la mayoría de los estudiantes que completan sus estudios, lo hacen con estas materias que se rinden masivamente hacia el final de su trayectoria académica.

Para obtener las probabilidades de absorción por cada uno de los estados absorbentes D , P , E , así como los tiempos medios de absorción por cada uno de dichos estados, se consideran los espacios $\mathcal{S}_D = \{I, FE1, D\}$ (debido a que no hay estudiantes que hayan desertado habiendo aprobado los espacios curriculares de $FE2$, $FE3$ y $FE4$), $\mathcal{S}_P = \{I, FE1, FE2, P\}$ (pues no hay estudiantes que hayan pedido un pase de institución habiendo aprobado los espacios curriculares de $FE3$ y $FE4$) y $\mathcal{S}_E = \{I, FE1, FE2, FE3, FE4, E\}$.

Para estimar las matrices de probabilidades de transición se consideraron nuevas bases de datos considerando, por un lado, a los estudiantes que desertaron; por otro a los que pidieron un pase y; finalmente, con los egresados.

Los conteos de transiciones según cada caso, son los siguientes:

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Periodo	<i>I</i>	<i>FE1</i>	<i>D</i>
<i>I</i>	269	12	314
<i>FE1</i>	0	36	13

Cuadro 3.29: Conteo de transiciones entre estados del campo de la Formación Específica para el caso de estudiantes que desertaron.

Periodo	<i>I</i>	<i>FE1</i>	<i>FE2</i>	<i>P</i>
<i>I</i>	68	10	0	20
<i>FE1</i>	0	26	3	12
<i>FE2</i>	0	0	2	2

Cuadro 3.30: Conteo de transiciones entre estados del campo de la Formación Específica para el caso de estudiantes que solicitaron un pase.

Periodo	<i>I</i>	<i>FE1</i>	<i>FE2</i>	<i>FE3</i>	<i>FE4</i>	<i>E</i>
<i>I</i>	3	12	0	0	0	0
<i>FE1</i>	0	25	19	4	2	0
<i>FE2</i>	0	0	7	17	15	1
<i>FE3</i>	0	0	0	0	7	7
<i>FE4</i>	0	0	0	0	0	12

Cuadro 3.31: Conteo de transiciones entre estados del campo de la Formación Específica para el caso de estudiantes que egresaron.

Asegurando que se cumplen los supuestos para realizar el ajuste de las cadenas de Markov, se tiene que las respectivas matrices de probabilidades de transición en un paso son:

$$P_D = \begin{pmatrix} 0,4521 & 0,0202 & 0,5277 \\ 0 & 0,7347 & 0,2653 \\ 0 & 0 & 1 \end{pmatrix}$$

$$P_P = \begin{pmatrix} 0,6939 & 0,1020 & 0 & 0,2041 \\ 0 & 0,6341 & 0,0732 & 0,2927 \\ 0 & 0 & 0,5000 & 0,5000 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

$$P_E = \begin{pmatrix} 0,2000 & 0,8000 & 0 & 0 & 0 & 0 \\ 0 & 0,5000 & 0,3800 & 0,0800 & 0,0400 & 0 \\ 0 & 0 & 0,1750 & 0,4250 & 0,3750 & 0,0250 \\ 0 & 0 & 0 & 0 & 0,5000 & 0,5000 \\ 0 & 0 & 0 & 0 & 0,0000 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Análogamente a como se procedió en el análisis para el campo de la Formación General, las respectivas matrices fundamentales son las siguientes:

	I	FE1	(Suma por fila)-1
I	1,83	0,14	0,97
FE1	0	3,77	2,77

Cuadro 3.32: Matriz Fundamental de la cadena del campo de la Formación Específica para el caso de estudiantes que desertaron.

	I	FE1	FE2	(Suma por fila)-1
I	3,27	0,91	0,13	3,31
FE1	0	2,73	0,40	2,13
FE2	0	0	2,00	1

Cuadro 3.33: Matriz Fundamental de la cadena del campo de la Formación Específica para el caso de estudiantes que solicitaron un pase.

En la siguiente, no se resta 1 a la suma de los elementos de la fila, puesto que al ser *FE4* un estado absorbente, se conseguirán los tiempos promedios para llegar a ese estado.

	I	FE1	FE2	FE3	FE4	Suma por fila
I	1,25	2	0,92	0,55	0,70	5,42
FE1	0	2	0,92	0,55	0,70	4,17
FE2	0	0	1,21	0,51	0,71	2,44
FE3	0	0	0	1	0,5	1,5
FE4	0	0	0	0	1	1

Cuadro 3.34: Matriz Fundamental de la cadena del campo de la Formación Específica para el caso de estudiantes que egresaron.

En promedio, se espera que un estudiante abandone la institución en menos de un año desde su ingreso. Si el alumno solicita un traslado a otra institución, el tiempo promedio estimado es de 3,31 años, mientras que para obtener el egreso se

requiere un tiempo promedio de 5,42 años. Al comparar estos resultados con los obtenidos para las materias del campo de Formación General, se puede observar que el tiempo promedio de deserción desde el momento del ingreso es bastante similar.

Se ha observado que el tiempo promedio estimado para que un estudiante egrese es de 5,42 años desde el momento en que ingresa a la institución. Este tiempo es ligeramente menor que el obtenido para las transiciones en el campo de Formación General, lo cual podría explicarse por la omisión de algunos estudiantes en la muestra que presentaron equivalencias en espacios curriculares específicos. Cabe destacar que estos casos son más frecuentes que aquellos con equivalencias en espacios curriculares de Formación General. A pesar de esto, los tiempos medios estimados son idénticos cuando se habla en términos del ciclo lectivo.

Cuando un estudiante aprueba el primer año del campo de Formación Específica, disminuye la estimación de la probabilidad de deserción, por lo que se espera que pase un tiempo promedio de 2,77 años desde esa aprobación hasta la posible deserción. Este tiempo no coincide con el de los estudiantes que aprueban el primer año del campo de Formación General. De esta diferencia se deduce que una gran parte de los estudiantes aprueba primero el primer año del campo de Formación General.

Por otro lado, se ha observado que el tiempo promedio estimado para que un estudiante decida cambiarse de institución disminuye a 2,13 años, lo que sugiere que estas asignaturas tienen un mayor impacto en la decisión de cambio de institución por parte del estudiante.

Al ser mayor la probabilidad de egreso en este caso, se espera que pasen 4,17 años desde que el estudiante completó el primer año de la Formación Específica, para que egrese.

En cuanto a los estudiantes que han aprobado el segundo año de la Formación Específica, se ha observado que se espera un tiempo promedio de 1 año para que decidan cambiarse de institución, lo que es ligeramente menor que el valor obtenido en el estado anterior, debido al pequeño aumento en la probabilidad de pase. Por otro lado, como la probabilidad de egreso es mucho mayor, se espera para estos estudiantes, que después de aprobar el segundo año de la Formación Específica, el tiempo promedio para egresar sea de 2,44 años.

Aprobando las materias de tercer año, el estudiante egresa en un tiempo promedio de 1,45 años comparado con el del año que correspondería si se sigue el plan de estudios. Se observa un tiempo promedio de 0 años desde que aprueban el cuarto año de este campo hasta el egreso, lo que indicaría que los estudiantes se reciben con estas materias. Debe tenerse en cuenta que $FE4$ era un estado absorbente en esta cadena.

A continuación, se calcula el producto entre la matriz fundamental y la sub-

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

matriz de probabilidades de transición en un paso desde estados transitorios a recurrentes, con el objetivo de obtener una matriz de acceder eventualmente a los estados absorbentes desde uno transitorio.

	D	P	E
I	0.8207	0.0907	0,0884
FE1	0,2437	0,2568	0,4993
FE2	0	0,0910	0,9089
FE3	0	0,0061	0,9939
FE4	0	0	1

Cuadro 3.35: Matriz de acceso eventual a los estados absorbentes desde uno transitorio, para el campo de la Formación Específica.

De esta manera, el alumno ingresante tiene una probabilidad de 82,07 % de abandonar eventualmente la carrera sin aprobar materias del campo de la Formación Específica; 9,07 % de cambiarse de institución y solo un 8,84 % de egresar.

Por su parte, aquellos estudiantes que logran aprobar el primer año del campo de la Formación Específica, tienen una probabilidad de 24,37 % de eventualmente dejar la carrera, 25,68 % de cambiarse de institución y 49,93 % de egresar.

Aprobando el segundo año del campo de la Formación Específica, la probabilidad de desertar disminuye a 0, la de cambiarse de institución disminuye a 9,1 % y la de egresar se eleva considerablemente a 90,89 %.

Si el estudiante aprueba el tercer año de este campo, tiene una probabilidad de 0,61 % de pedir un pase y 99,39 % de egresar.

Finalmente, si aprueba el cuarto año, tiene una probabilidad de egresar del 100 %.

Si bien las probabilidades de egreso aumentan notoriamente a medida que el estudiante va aprobando los campos de la Formación Específica, y disminuyen considerablemente las de deserción y las de pedir un pase, estas probabilidades no consideran el tiempo promedio para alcanzar esos estados, los cuales se alejan bastante de los cuatro años propuestos en el plan de estudios.

Se puede apreciar, que la probabilidad de egresar eventualmente aumenta si el estudiante aprueba el primer año de las materias específicas.

3.4.3. Transición por el campo de la Práctica Docente

A continuación se considera otro modelo empleando cadenas de Markov a tiempo discreto, en el que se considera nuevamente a cada ciclo lectivo como un paso, y el espacio de estados

$$\mathcal{S} = \{I, PD1, PD2, PD3, PD4, D, P, E\}$$

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Como puede verse, en este espacio de estados, se consideran los espacios curriculares del campo de la Práctica Docente por año. De esta manera, se podrá visualizar a partir de qué campo aumenta la probabilidad de acceder a los estados absorbentes.

Las transiciones posibles son las siguientes:

- $I \rightarrow I$: el estudiante ingresante no aprobó la Práctica Docente I, durante el ciclo lectivo, permaneciendo en el mismo estado en el siguiente ciclo lectivo.
- $I \rightarrow PD1$: el estudiante ingresante aprueba la Práctica Docente I en el siguiente ciclo lectivo. Debe recordarse que al tener formato de seminario, estos espacios curriculares se aprueban o el estudiante queda libre, en donde en el último caso no puede cursar el espacio correlativo.
- $I \rightarrow D$: el estudiante ingresante abandona sus estudios en esta institución sin aprobar $PD1$, pasando al estado de *deserción* en el siguiente ciclo lectivo.
- $I \rightarrow P$: el estudiante ingresante pide un pase a otra institución para continuar la misma carrera, sin haber aprobado $PD1$.
- $PDi \rightarrow PDi$ ($i = 1, 2, 3, 4$): el alumno que aprobó PDi , y continúa en la misma condición durante el ciclo lectivo, permanece en el mismo estado en el siguiente ciclo.
- $PDi \rightarrow PD(i+1)$ ($i = 1, 2, 3$): el alumno que aprobó PDi , aprueba $PD(i+1)$ durante el ciclo lectivo, pasa al estado $PD(i+1)$ en el siguiente ciclo.
- $PDi \rightarrow D$ ($i = 1, 2, 3, 4$): el estudiante habiendo aprobado PDi abandona sus estudios en esta institución, pasando al estado de *deserción*, en el siguiente ciclo lectivo.
- $PDi \rightarrow P$ ($i = 1, 2, 3, 4$): el estudiante pide un pase a otra institución para continuar la misma carrera, habiendo aprobado PDi , en el ciclo anterior.
- $PD4 \rightarrow E$: el estudiante termina de aprobar todos los espacios curriculares del campo de la Práctica Docente, en el ciclo lectivo, y pasa al estado de *egreso* en el siguiente ciclo lectivo.

Las transiciones no permitidas entre dos ciclos lectivos consecutivos son:

- $I \rightarrow E$: el estudiante ingresante no puede egresar en el ciclo lectivo siguiente, puesto que por la estructura del plan de estudios, debe permanecer en la carrera al menos 4 ciclos lectivos, desde que ingresa.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

- $PDi \rightarrow I$ ($i = 1, 2, 3, 4$): el estudiante regular ya aprobó algún PDi del plan de estudios y es imposible que vuelva al estado de ingresante.
- $PDi \rightarrow PDj$ (con $j \neq i + 1$): el estudiante ya aprobó alguna PDi del plan de estudios y es imposible que apruebe una Práctica Docente de años anteriores, puesto que para acceder a la actual, debe haber aprobado las anteriores correlativas. Al mismo tiempo, no puede acceder a la acreditación de una Práctica Docente de años siguientes, excepto la inmediata superior.
- $D \rightarrow I$: no se registraron casos en que el estudiante haya abandonado la carrera y haya regresado en el siguiente ciclo lectivo para comenzar nuevamente desde el inicio. Si se diera ese caso, se reconocería el trayecto previo del estudiante por medio de equivalencias.
- $P \rightarrow I$: no se registraron casos en que el estudiante se haya cambiado de institución y haya regresado en el siguiente ciclo lectivo para comenzar nuevamente desde el inicio. Si se diera ese caso, se reconocería el trayecto previo del estudiante por medio de equivalencias.
- $E \rightarrow I$: el estudiante que ingresó no puede volver al estado de ingresante, salvo que desee cursar nuevamente la carrera desde el inicio.
- $D \rightarrow PDi$ ($i = 1, 2, 3, 4$): esta transición, si bien es posible, no se registró que ningún estudiante que haya abandonado sus estudios en un ciclo lectivo, los haya retomado en el siguiente ciclo.
- $P \rightarrow PDi$: esta transición, si bien es posible, no se registró que ningún estudiante que se haya cambiado de institución en un ciclo lectivo, haya regresado a esta en el siguiente ciclo.
- $E \rightarrow PDi$: el estudiante que egresó en un ciclo lectivo, no regresa al estado de estudiante regular en el siguiente ciclo lectivo.
- Las transiciones entre estados absorbentes tampoco son posibles.

Para plantear este nuevo modelo se realizó un procedimiento similar al modelo anterior para probar los supuestos de estacionariedad y la propiedad de Markov. Se trabajó sobre los periodos comprendidos entre 2014 y 2019, como se mencionó con anterioridad.

Para que se cumpla la propiedad de estacionariedad, se requiere homogeneidad de las probabilidades de transición durante todos los periodos. Para verificar esta propiedad se debe asegurar que la frecuencia de estudiantes en cada uno de los estados no sea dependiente en el cual se toma la medición. Para esto se utilizó una prueba de independencia mediante el test exacto de Fisher.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

En las siguientes tablas se presentan las transiciones entre estados para dos periodos consecutivos.

2014	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	16	36	0	0	0	39	0	0
<i>PD1</i>	0	14	23	0	0	9	2	0
<i>PD2</i>	0	0	20	9	0	4	3	0
<i>PD3</i>	0	0	0	4	5	0	0	0
<i>PD4</i>	0	0	0	0	0	0	0	0

2015	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	11	42	0	0	0	54	6	0
<i>PD1</i>	0	23	17	0	0	11	1	0
<i>PD2</i>	0	0	26	6	0	3	7	0
<i>PD3</i>	0	0	0	3	8	1	1	0
<i>PD4</i>	0	0	0	0	0	0	0	5

2016	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	13	28	0	0	0	34	0	0
<i>PD1</i>	0	17	32	0	0	14	1	0
<i>PD2</i>	0	0	18	13	0	11	1	0
<i>PD3</i>	0	0	0	5	4	0	0	0
<i>PD4</i>	0	0	0	0	0	0	0	8

2017	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	10	34	0	0	0	44	0	0
<i>PD1</i>	0	7	22	0	0	15	1	0
<i>PD2</i>	0	0	27	13	0	9	1	0
<i>PD3</i>	0	0	0	6	9	2	1	0
<i>PD4</i>	0	0	0	0	0	0	0	4

2018	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	9	49	0	0	0	48	0	0
<i>PD1</i>	0	9	25	0	0	6	2	0
<i>PD2</i>	0	0	26	7	0	12	4	0
<i>PD3</i>	0	0	0	5	9	2	3	0
<i>PD4</i>	0	0	0	0	0	0	1	8

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

2019	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	7	43	0	0	0	43	0	0
<i>PD1</i>	0	15	22	0	0	20	0	0
<i>PD2</i>	0	0	27	7	0	14	3	0
<i>PD3</i>	0	0	0	5	5	1	1	0
<i>PD4</i>	0	0	0	0	0	0	0	9

Cuadro 3.36: Transiciones entre estados por año - Campo de la Práctica Docente.

En consecuencia, las tablas que muestran las frecuencias de transiciones entre un estado transitorio determinado y todos los demás estados, para cada periodo, son las siguientes:

Periodo	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>D</i>	<i>P</i>	<i>E</i>
2014	16	36	0	0	0	39	0	0
2015	11	42	0	0	0	54	6	0
2016	13	28	0	0	0	34	0	0
2017	10	34	0	0	0	44	0	0
2018	9	49	0	0	0	48	0	0
2019	7	43	0	0	0	43	0	0

Cuadro 3.37: Frecuencia de transiciones entre el estado *I* y todos los estados para cada periodo.

Periodo	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>D</i>	<i>P</i>	<i>E</i>
2014	0	14	23	0	0	9	2	0
2015	0	23	17	0	0	11	1	0
2016	0	17	32	0	0	14	1	0
2017	0	7	22	0	0	15	1	0
2018	0	9	25	0	0	6	2	0
2019	0	15	22	0	0	20	0	0

Cuadro 3.38: Frecuencia de transiciones entre el estado *PD1* y todos los estados para cada periodo.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Periodo	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>D</i>	<i>P</i>	<i>E</i>
2014	0	0	20	9	0	4	3	0
2015	0	0	26	6	0	3	7	0
2016	0	0	18	13	0	11	1	0
2017	0	0	27	13	0	9	1	0
2018	0	0	26	7	0	12	4	0
2019	0	0	27	7	0	14	3	0

Cuadro 3.39: Frecuencia de transiciones entre el estado *PD2* y todos los estados para cada periodo.

Periodo	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>D</i>	<i>P</i>	<i>E</i>
2014	0	0	0	4	5	0	0	0
2015	0	0	0	3	8	1	1	0
2016	0	0	0	5	4	0	0	0
2017	0	0	0	6	9	2	1	0
2018	0	0	0	5	9	2	3	0
2019	0	0	0	5	5	1	1	0

Cuadro 3.40: Frecuencia de transiciones entre el estado *PD3* y todos los estados para cada periodo.

Periodo	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>D</i>	<i>P</i>	<i>E</i>
2014	0	0	0	0	0	0	0	0
2015	0	0	0	0	5	0	0	0
2016	0	0	0	0	7	0	1	5
2017	0	0	0	0	6	1	0	4
2018	0	0	0	0	8	0	1	6
2019	0	0	0	0	15	0	0	1

Cuadro 3.41: Frecuencia de transiciones entre el estado *PD4* y todos los estados para cada periodo.

Aplicando el test exacto de Fisher, en las tablas de contingencia 3.37 a 3.41, para distintos grupos de años consecutivos, se obtiene que para el periodo 2014-2019, las frecuencias de transiciones son estadísticamente independientes del periodo.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Estado	p -valor
I	0,0659
$PD1$	0,0594
$PD2$	0,0659
$PD3$	0,9625
$PD4$	0,0659

Cuadro 3.42: Resultados de la prueba de independencia entre periodos y transiciones (2014-2019).

En este caso, si se desean obtener estimaciones a largo plazo, del comportamiento de las transiciones, debe tenerse en cuenta el periodo 2014-2019.

Otro supuesto a asegurar es que las frecuencias de transición obtenidas, cumplan con la principal característica de las cadenas de Markov, deben ser dependientes únicamente de su estado anterior.

Para verificar este supuesto, se realizó el test exacto de Fisher en las tablas de contingencia para los años que pertenecen al periodo 2014-2019, obteniendo en todos los casos un p -valor de 0,0005.

Los resultados de la prueba indican que se rechaza la hipótesis nula entre la independencia de estados, por lo que se puede garantizar que los resultados obtenidos en un estado, dependen del estado del periodo anterior.

Se estiman las matrices de transición P , utilizando el estimador de máxima verosimilitud del resultado del Teorema 2

$$\hat{P}_{MV} = \frac{n_{xy}}{\sum_{z \in \mathcal{S}} n_{xz}}$$

análogamente a como se trabajó con los modelos anteriores.

Los resultados de dichas estimaciones, se muestran en la siguiente tabla:

Estados	I	$PD1$	$PD2$	$PD3$	$PD4$	D	P	E
I	0,1166	0,4099	0,0000	0,0000	0,0000	0,4629	0,0106	0,0000
$PD1$	0,0000	0,2760	0,4578	0,0000	0,0000	0,2435	0,0227	0,0000
$PD2$	0,0000	0,0000	0,5314	0,2030	0,0000	0,1956	0,0701	0,0000
$PD3$	0,0000	0,0000	0,0000	0,3500	0,5000	0,0750	0,0750	0,0000
$PD4$	0,0000	0,0000	0,0000	0,0000	0,6833	0,0167	0,0333	0,2667
D	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000
P	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000
E	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000

Cuadro 3.43: Matriz de transición - Campo de la Práctica Docente.

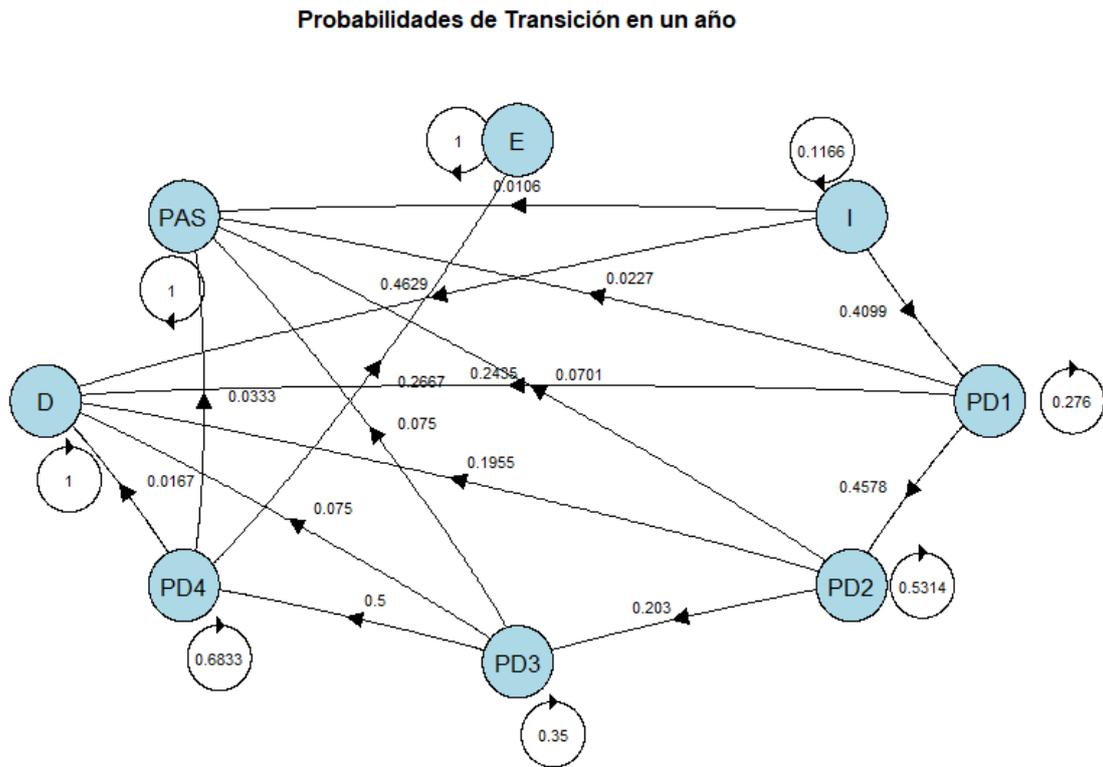


Figura 3.22: Probabilidades de transición en un año.

En el gráfico 3.22 pueden observarse las probabilidades de transición en un año.

Se pueden obtener las probabilidades de transición en n pasos, o más específicamente, en n ciclos lectivos, calculando las potencias n -ésimas de la matriz de transición P .

Interesa conocer entonces, ¿cuál es la probabilidad de que un estudiante que ingresa, egrese en los 4 años propuestos en el plan de estudios de la carrera? y, ¿cuál es la probabilidad de llegar a algún estado absorbente, estando en otros estados de la carrera, luego de 4 años.

Para ello, se calcula P^4 :

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Estados	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>D</i>	<i>P</i>	<i>E</i>
<i>I</i>	0,0001	0,0144	0,1150	0,0485	0,0190	0,7485	0,0542	0,0000
<i>PD1</i>	0	0,0058	0,1325	0,0845	0,0855	0,5571	0,1219	0,0123
<i>PD2</i>	0	0	0,0797	0,0724	0,1684	0,4259	0,1839	0,0694
<i>PD3</i>	0	0	0	0,0150	0,3045	0,1375	0,1613	0,3816
<i>PD4</i>	0	0	0	0	0,2179	0,0412	0,0822	0,6585

Cuadro 3.44: Matriz de transición en 4 pasos - Transición por el campo de la Práctica Docente.

Aquí se puede observar que la probabilidad de que un estudiante se gradúe en cuatro años es nula. La probabilidad de que se cambie de institución es ligeramente mayor, con un 5,42 %, mientras que la probabilidad de abandono es significativamente más alta, con un 74,85 %.

Si consideramos a los estudiantes que aprobaron la Práctica Docente I, la probabilidad de que se gradúen en cuatro años aumenta ligeramente a un 1,23 %, mientras que la probabilidad de abandonar los estudios disminuye a un 55,71 %. Además, se observa un aumento en la probabilidad de cambiar de institución, que alcanza un 12,19 %.

En el caso de los estudiantes que aprobaron la Práctica Docente II, se observa un leve aumento en la probabilidad de graduarse en cuatro años, alcanzando un 6,94 %. Asimismo, la probabilidad de abandonar los estudios disminuye a un 42,59 %, mientras que la probabilidad de cambiar de institución aumenta a un 18,39 %.

Para los estudiantes que aprobaron la Práctica Docente III, se observa un significativo aumento en la probabilidad de graduarse en cuatro años, llegando a un 38,16 %. Además, la probabilidad de abandonar los estudios disminuye a un 13,75 %, mientras que la probabilidad de cambiar de institución también disminuye ligeramente a un 16,13 %.

Respecto de los estudiantes que aprobaron la Práctica Docente IV, la probabilidad de graduarse en cuatro años experimenta un significativo aumento, llegando a un 65,85 %. Además, tanto la probabilidad de abandonar los estudios como la probabilidad de cambiar de institución disminuyen a 4,12 % y 8,22 %, respectivamente.

Puede afirmarse que a medida que los estudiantes van aprobando los espacios curriculares del campo de la Práctica Docente, se afianzan en su carrera y aumenta la probabilidad de graduarse, mientras que disminuye la probabilidad de abandonar los estudios. Además, se observa que la aprobación de la Práctica Docente III y, en mayor medida, de la Práctica Docente IV, incrementa significativamente la probabilidad de egreso. Esto podría deberse a que para aprobar estas prácticas, los estudiantes deben haber regularizado varios espacios curriculares de la Formación

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Específica de los primeros dos años, tal y como se observó en el modelo anterior. No obstante, es importante tener en cuenta que la regularidad de los espacios específicos no garantiza el aumento de la probabilidad de graduarse, sino que es necesario aprobarlos.

Al comparar estos resultados con los modelos anteriores, se puede plantear la hipótesis de que los estudiantes, en general, aprueban primero las materias del campo de la Formación General y la Práctica Docente (con una mayor probabilidad en las últimas) y posteriormente las del campo de la Formación Específica.

En el siguiente cuadro se muestran las probabilidades de transición en más de 4 ciclos lectivos para visualizar cuáles son las transiciones más probables por el campo de la Práctica Docente y las probabilidades de llegar a algún estado absorbente:

5 Ciclos Lectivos								
	I	PD1	PD2	PD3	PD4	D	P	E
I	0	0,0040	0,0677	0,0403	0,0372	0,7785	0,0669	0,0050
PD1	0	0,0016	0,0730	0,0565	0,1007	0,5922	0,1405	0,0352
PD2	0	0	0,0423	0,0415	0,1513	0,4497	0,2006	0,1143
PD3	0	0	0	0,0052	0,2155	0,1437	0,1725	0,4628
PD4	0	0	0	0	0,1489	0,0448	0,0894	0,7167

6 Ciclos Lectivos								
	I	PD1	PD2	PD3	PD4	D	P	E
I	0	0,0011	0,0378	0,0279	0,0456	0,7964	0,0760	0,0150
PD1	0	0,0004	0,0395	0,0346	0,0971	0,6128	0,1533	0,0621
PD2	0	0	0,0225	0,0231	0,1241	0,4636	0,2117	0,1547
PD3	0	0	0	0,0018	0,1499	0,1477	0,1801	0,5203
PD4	0	0	0	0	0,1017	0,0473	0,0944	0,7564

7 Ciclos Lectivos								
	I	PD1	PD2	PD3	PD4	D	P	E
I	0	0,0003	0,0206	0,0174	0,0451	0,8069	0,0823	0,0271
PD1	0	0,0001	0,0212	0,0201	0,0836	0,6249	0,1619	0,0879
PD2	0	0	0,0119	0,0126	0,0964	0,4718	0,2191	0,1878
PD3	0	0	0	0,0006	0,1033	0,1503	0,1852	0,5603
PD4	0	0	0	0	0,0695	0,0491	0,0978	0,7835

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

		8 Ciclos Lectivos							
		I	PD1	PD2	PD3	PD4	D	P	E
I		0	0	0,0111	0,0102	0,0395	0,8132	0,0865	0,0392
PD1		0	0	0,0113	0,0113	0,0672	0,6320	0,1677	0,1102
PD2		0	0	0,0063	0,0068	0,0722	0,4767	0,2241	0,2135
PD3		0	0	0	0,0002	0,0709	0,1521	0,1887	0,5878
PD4		0	0	0	0	0,0475	0,0502	0,1001	0,8021

Cuadro 3.45: Matrices de transición en más de 4 pasos - Campo de la Práctica Docente.

Los siguientes gráficos muestran la convergencia de las probabilidades de transición para 12 años lectivos sobre cada uno de los estados futuros basado en un estado actual.

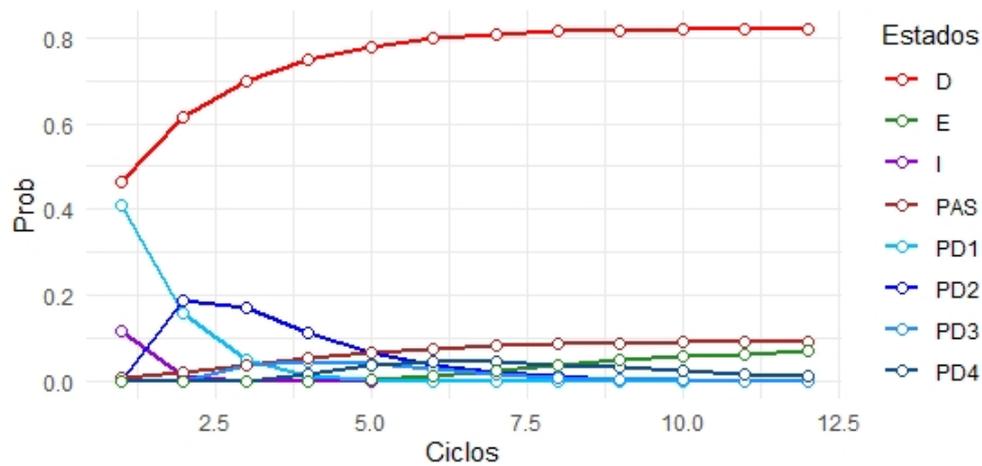


Figura 3.23: Probabilidades de transición por año lectivo, dado que actualmente es Ingresante.

Si se dejan solamente los estados I , $PD1$, D , P , E , es decir el estado consecutivo a I y los estados absorbentes:

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

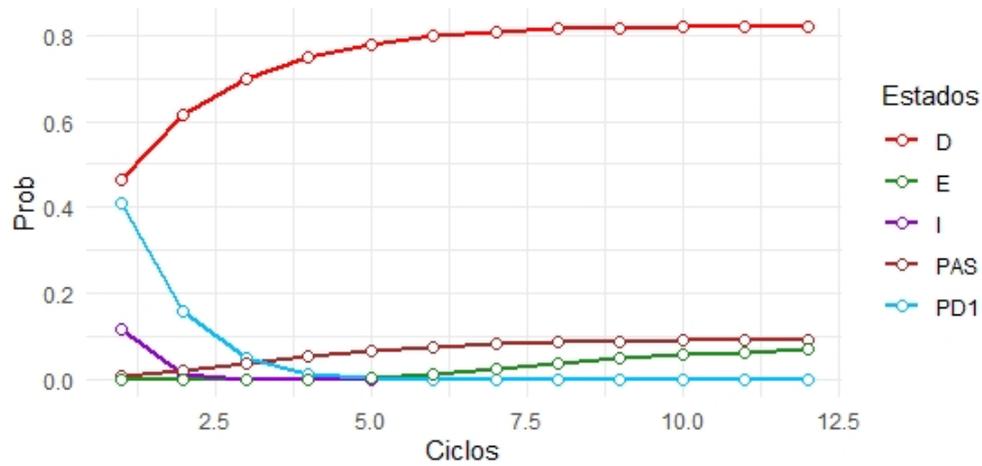


Figura 3.24: Probabilidades de transición por año lectivo, dado que actualmente es Ingresante.

En la Figura 3.23 se puede apreciar que la probabilidad de abandono de los estudiantes ingresantes es significativamente superior a la de aprobación de la Práctica Docente I. Esta tendencia se intensifica a medida que transcurren los años, llegando a alcanzar un valor del 82%.

La probabilidad de que un estudiante acredite la Práctica Docente I, en su primer año de estudios, es mayor que la de permanecer en el estado de ingresante. Sin embargo, esta probabilidad disminuye rápidamente a medida que pasan los años, lo que sugiere que los estudiantes salen rápidamente de este estado.

A medida que avanzan los ciclos lectivos, la probabilidad de egresar es bastante similar a la de solicitar un pase de institución, siendo esta última ligeramente mayor.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

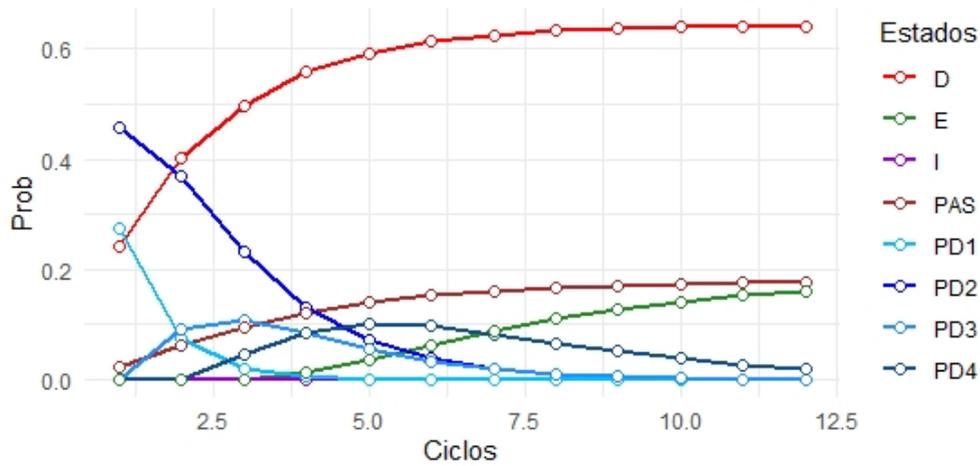


Figura 3.25: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente I.

Si se dejan solamente los estados $PD1$, $PD2$, D , P , E , es decir el estado consecutivo a $PD1$ y los estados absorbentes:

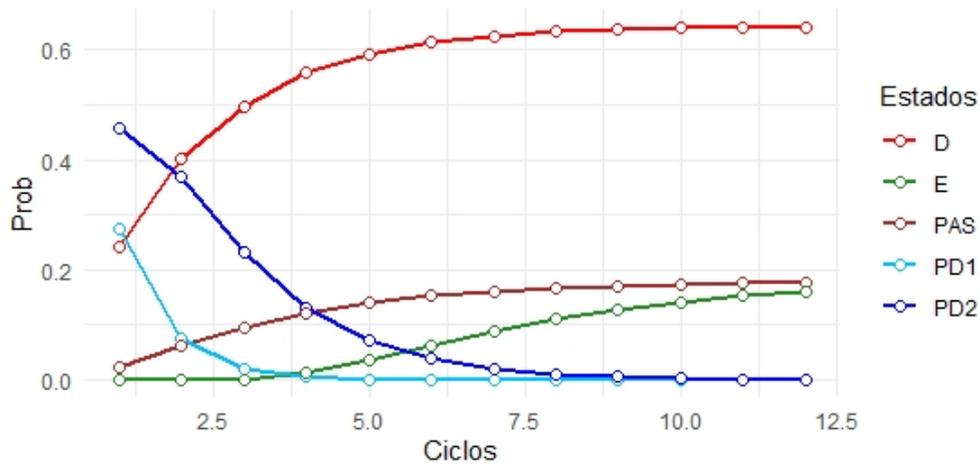


Figura 3.26: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente I.

En la Figura 3.25, se puede observar que la probabilidad de abandono de un estudiante que ha aprobado la Práctica Docente I disminuye en comparación con la de un estudiante ingresante. Sin embargo, esta probabilidad crece rápidamente durante los primeros años transcurridos y converge a un valor cercano al 65%.

La probabilidad de aprobar la Práctica Docente II es mucho mayor que la de permanecer con la Práctica Docente I aprobada, y ambas probabilidades disminuyen a medida que pasa el tiempo. La probabilidad de aprobar la Práctica Docente II, teniendo ya aprobada la Práctica Docente I, es mayor que la de abandonar los estudios después de 1 año; sin embargo, después de 2 años esta situación se revierte y la probabilidad de abandonar los estudios es mayor.

En comparación con el gráfico 3.23, se puede afirmar que tanto la Práctica Docente I como la II tienen una buena probabilidad de ser aprobados. Sin embargo, a diferencia de lo que ocurre con la aprobación de los espacios curriculares del primer año del campo de Formación Específica, no se logra reducir significativamente la probabilidad de deserción.

La probabilidad de solicitar un pase hacia otra institución es significativamente menor que la de aprobar la Práctica Docente II, aunque después de cinco años esta situación se revierte. La probabilidad de pedir un pase crece rápidamente durante los primeros años y converge a un valor cercano al 18 %, lo cual es ligeramente mayor que la probabilidad de pase del estudiante ingresante.

En cambio, si el estudiante ya ha aprobado la Práctica Docente I, la probabilidad de egresar es ligeramente mayor que la del estudiante ingresante y aumenta de manera casi lineal durante los primeros años, aproximándose progresivamente al 15 %.

La probabilidad de egreso a medida que transcurre el tiempo, es significativamente menor para los estudiantes que aprobaron la Práctica Docente I en comparación con aquellos que aprobaron las materias del campo de la Formación Específica de primer año y ligeramente menor que para los que aprobaron las materias de primer año del campo de la Formación General.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

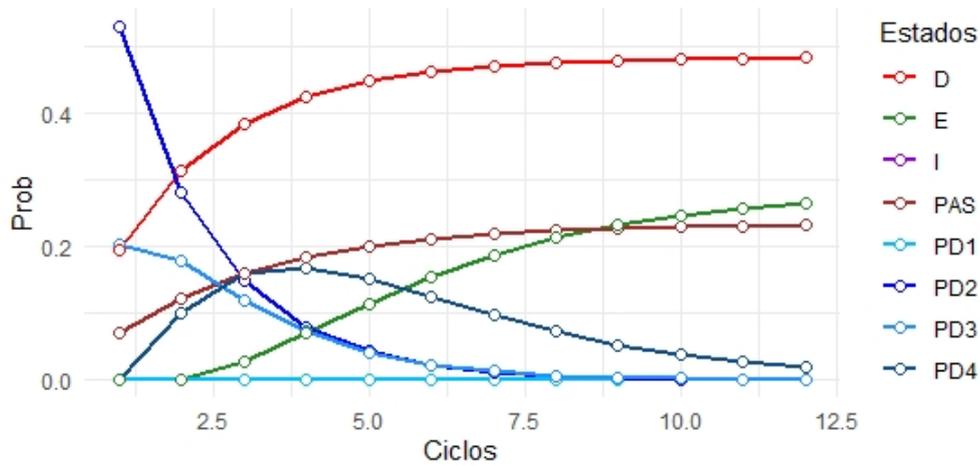


Figura 3.27: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente II.

Si se dejan solamente los estados $PD2$, $PD3$, D , P , E , es decir el estado consecutivo a $PD2$ y los estados absorbentes:

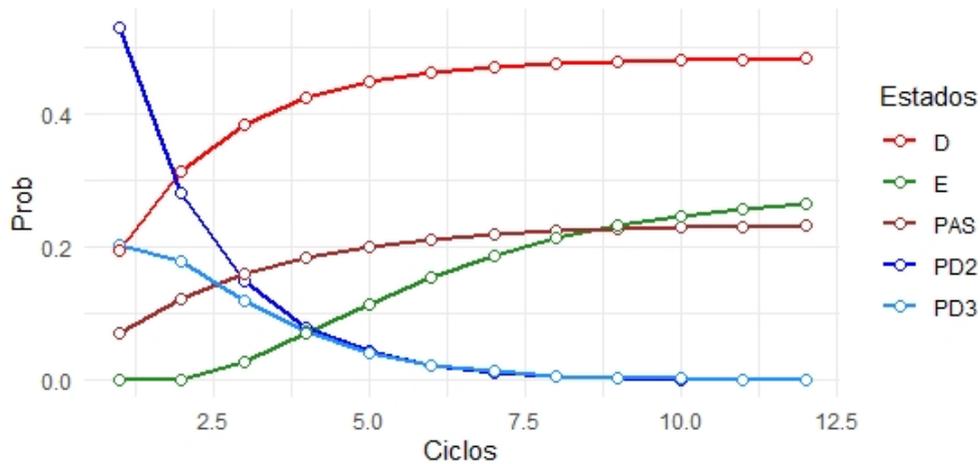


Figura 3.28: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente II.

En la Figura 3.27, se puede observar que después de transcurrido el primer año, la probabilidad de que un estudiante que aprobó la Práctica Docente II permanezca en ese estado es mayor que la de pasar a cualquier otro estado. Sin embargo, esta situación se revierte después de dos años, donde la probabilidad de abandonar

la carrera aumenta considerablemente y supera a la probabilidad de aprobar la Práctica Docente III.

La probabilidad de aprobar la Práctica Docente III es notablemente inferior a la de abandonar los estudios y quedarse en el estado de aprobación de la Práctica Docente II. Además, se puede observar que después de tres años, la probabilidad de aprobar la Práctica III solo supera ligeramente a la de permanecer en la Práctica II. En ambos casos, estas probabilidades disminuyen rápidamente a medida que pasa el tiempo, lo que sugiere que los estudiantes abandonan estos estados rápidamente.

Por otro lado, se puede apreciar que hay un retraso en la aprobación de la Práctica Docente III, ya que esta etapa es la primera dentro de este ámbito que exige la regularidad en las materias del campo de Formación Específica correspondientes al primer año.

Asimismo, se puede observar que la probabilidad de cambiar de institución es significativamente menor que la de mantener la Práctica Docente II aprobada o acreditar la III. No obstante, después de 4 años esta tendencia cambia, ya que la probabilidad de cambiar de institución supera la de mantener aprobada la Práctica Docente II, lo que sugiere que algunos estudiantes que no lograron avanzar en su trayectoria académica deciden cambiar de institución.

La probabilidad de cambiar de institución aumenta a medida que pasan los años en comparación con la probabilidad del estudiante que solo ha aprobado la Práctica Docente I, llegando a converger al 22 %.

No obstante, si el estudiante ha aprobado la Práctica Docente II, la probabilidad de egresar es menor que la de cambiar de institución, aunque esta diferencia disminuye a medida que pasan los años y la probabilidad de egreso aumenta más rápidamente durante los primeros años. Finalmente, después de 9 años, la probabilidad de egreso supera la de cambio de institución y aumenta aún más su diferencia, convergiendo a alrededor del 25 %.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

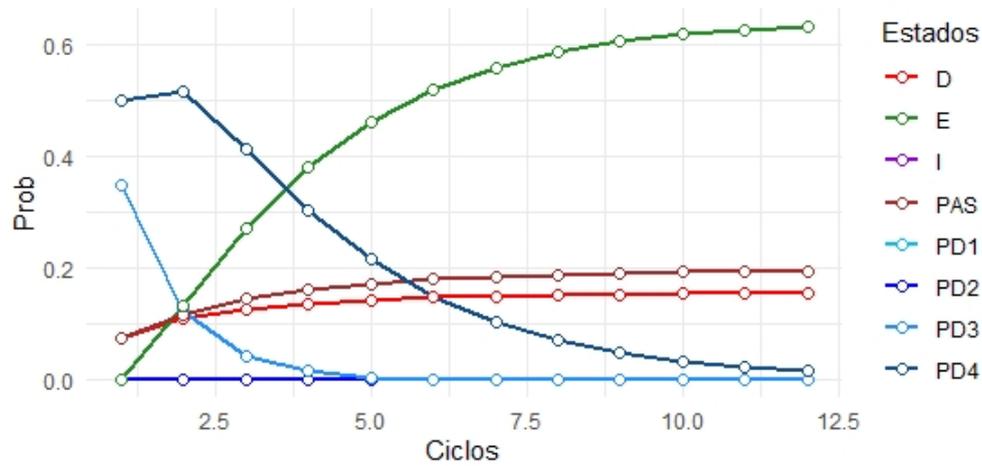


Figura 3.29: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente III.

Si se dejan solamente los estados $PD3$, $PD4$, D , P , E , es decir el estado consecutivo a $PD3$ y los estados absorbentes:

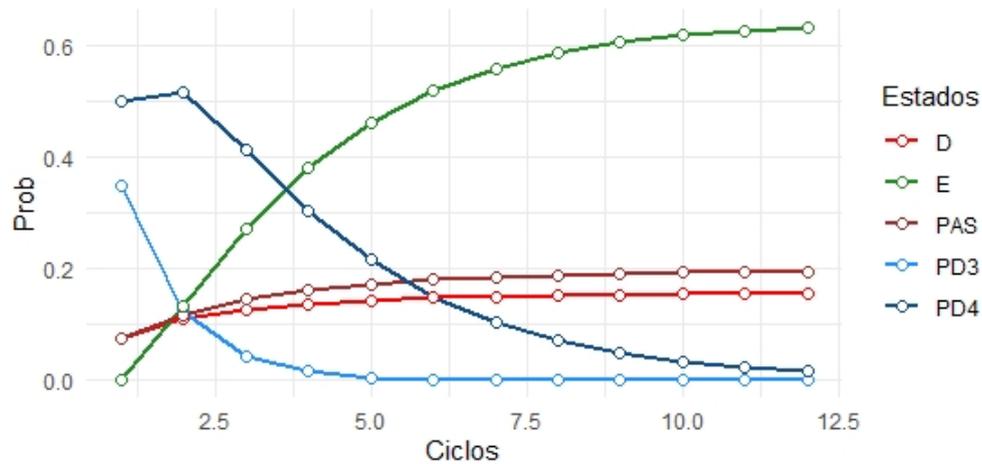


Figura 3.30: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente III.

En la Figura 3.29 se puede apreciar que después de un año, la probabilidad de que un estudiante que ha aprobado la Práctica Docente III avance a la Práctica Docente IV es mayor que la probabilidad de transitar a cualquier otro estado. Además, esta probabilidad es ligeramente superior a la de mantenerse en la Práctica Docente III.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

A medida que pasan los ciclos lectivos, se puede observar que las probabilidades de mantenerse en el estado de Práctica Docente III aprobada disminuyen rápidamente, lo que sugiere que los estudiantes abandonan ese estado con rapidez.

Después de 7 años, las probabilidades de abandonar los estudios y cambiar de institución superan a la de aprobar la Práctica Docente IV. Estas probabilidades aumentan lentamente y son muy similares, convergiendo a alrededor del 15% y 20%, respectivamente. Cabe destacar que la probabilidad de cambiar de institución es ligeramente mayor que la de abandonar los estudios.

Es importante destacar que la probabilidad de abandono de los estudios disminuye significativamente en comparación con la de abandonar después de haber aprobado las Prácticas Docentes I y II. Sin embargo, la probabilidad de cambiar de institución aumenta ligeramente.

La probabilidad de egresar, habiendo aprobado la Práctica Docente III, aumenta considerablemente en comparación con la de haber aprobado las Prácticas Docentes anteriores. Se observa que esta probabilidad crece rápidamente durante los primeros años y a partir del cuarto año ya supera la probabilidad de pasar a cualquier otro estado, convergiendo a casi un 63%.

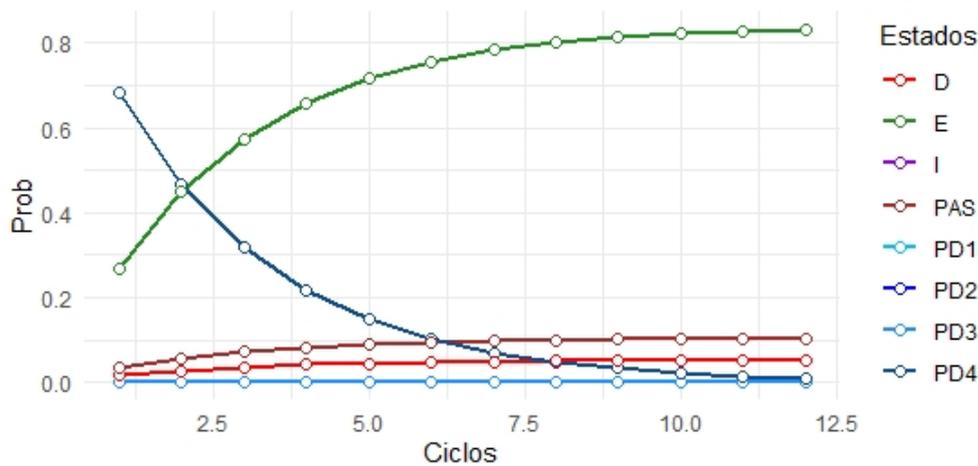


Figura 3.31: Probabilidades de transición por año lectivo, dado que actualmente tiene aprobada la Práctica Docente IV.

En la Figura 3.31, se observa que, si bien el estudiante completó este campo de la Práctica Docente, la probabilidad de permanecer en este estado es mayor que la de egresar transcurrido un ciclo lectivo.

A partir de los dos años, esta situación se revierte y la probabilidad de egresar se vuelve mayor. De hecho, esta probabilidad supera la de pasar a cualquier otro estado y converge a casi un 82%.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Por lo tanto, se podría afirmar que es más probable que se apruebe la Práctica Docente IV antes de terminar los otros campos de la formación, lo cual puede generar un pequeño retraso en el paso al estado de egreso.

No obstante, la probabilidad de pedir un pase o de desertar sigue presente. En ambos casos, se observa una disminución en comparación con los estudiantes que aprobaron la Práctica Docente III. En particular, ambas probabilidades crecen lentamente y convergen a aproximadamente un 10 % y un 6 %, respectivamente.

A partir de los 6 años, se observa que la probabilidad de pedir un pase supera a la de mantener la Práctica Docente IV aprobada, y a partir de los 8 años, la probabilidad de desertar supera a la de mantenerse en ese estado. Este resultado es coherente ya que, si el estudiante no logra egresar, es probable que se deba a otros campos de la formación ya que ha completado la Práctica Docente IV y, en consecuencia, el estudiante puede optar por cambiarse de institución o abandonar sus estudios.

En síntesis, la probabilidad de desertar es mayor cuando el alumno se encuentra en el estado de ingresante, pero disminuye significativamente al aprobar las Prácticas Docentes III y IV. Las Prácticas Docentes I y II tienen una alta probabilidad de ser aprobadas durante los primeros años de la carrera.

Se puede observar un retraso en el paso de la Práctica Docente II a la III, lo cual podría deberse a que la Práctica Docente III es correlativa de algunos espacios curriculares específicos del primer año.

Aprobando la Práctica Docente III, ya es mayor la probabilidad de egresar que la de cambiarse de institución o abandonar los estudios.

El riesgo de abandono de la carrera se concentra en los primeros años, pero dado que existe una alta probabilidad de aprobación de las primeras dos Prácticas Docentes, se puede inferir que su deserción no se debe principalmente a estos campos.

A medida que se avanza hacia las siguientes Prácticas Docentes, se observa un ligero aumento en la probabilidad de cambio de institución, pero esta disminuye una vez aprobada la Práctica Docente IV.

Para obtener las probabilidades de absorción por cada uno de los estados absorbentes D , P , E , así como los tiempos medios de absorción por cada uno de dichos estados, se consideran los espacios $\mathcal{S}_D = \{I, PD1, PD2, PD3, PD4, D\}$, $\mathcal{S}_P = \{I, PD1, PD2, PD3, P\}$ y $\mathcal{S}_E = \{I, PD1, PD2, PD3, PD4, E\}$.

Para estimar las matrices de probabilidades de transición se consideraron nuevas bases de datos considerando, por un lado, a los estudiantes que desertaron; por otro a los que pidieron un pase y; finalmente, con los egresados.

Los conteos de transiciones según cada caso, son los siguientes:

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

Periodo	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>D</i>
<i>I</i>	49	116	0	0	0	278
<i>PD1</i>	0	42	46	0	0	77
<i>PD2</i>	0	0	45	6	0	53
<i>PD3</i>	0	0	0	10	2	6
<i>PD4</i>	0	0	0	0	5	2

Cuadro 3.46: Conteo de transiciones entre estados del campo de la Práctica Docente para el caso de estudiantes que desertaron.

Periodo	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>P</i>
<i>I</i>	5	7	0	0	0	6
<i>PD1</i>	0	17	18	0	0	7
<i>PD2</i>	0	0	35	7	0	20
<i>PD3</i>	0	0	0	4	2	6
<i>PD4</i>	0	0	0	0	0	2

Cuadro 3.47: Conteo de transiciones entre estados del campo de la Práctica Docente para el caso de estudiantes que solicitaron un pase.

Periodo	<i>I</i>	<i>PD1</i>	<i>PD2</i>	<i>PD3</i>	<i>PD4</i>	<i>E</i>
<i>I</i>	0	7	0	0	0	0
<i>PD1</i>	0	1	12	0	0	0
<i>PD2</i>	0	0	9	19	0	0
<i>PD3</i>	0	0	0	3	25	0
<i>PD4</i>	0	0	0	0	22	16

Cuadro 3.48: Conteo de transiciones entre estados del campo de la Práctica Docente para el caso de estudiantes que egresaron.

Asegurando que se cumplen los supuestos para realizar el ajuste de las cadenas de Markov, se tiene que las respectivas matrices de probabilidades de transición en un paso son:

$$P_D = \begin{pmatrix} 0,1106 & 0,2619 & 0 & 0 & 0 & 0,6275 \\ 0 & 0,2545 & 0,2788 & 0 & 0 & 0,4667 \\ 0 & 0 & 0,4327 & 0,0577 & 0 & 0,5096 \\ 0 & 0 & 0 & 0,5556 & 0,1111 & 0,3333 \\ 0 & 0 & 0 & 0 & 0,7143 & 0,2857 \\ 0 & 0 & 0 & 0 & 0 & 1,0000 \end{pmatrix}$$

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

$$P_P = \begin{pmatrix} 0,2780 & 0,3890 & 0 & 0 & 0 & 0,3333 \\ 0 & 0,4050 & 0,4290 & 0 & 0 & 0,1670 \\ 0 & 0 & 0,5650 & 0,1130 & 0 & 0,3230 \\ 0 & 0 & 0 & 0,3333 & 0,1670 & 0,5000 \\ 0 & 0 & 0 & 0 & 0 & 1,0000 \end{pmatrix}$$

$$P_E = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0,0770 & 0,9230 & 0 & 0 & 0 \\ 0 & 0 & 0,3210 & 0,6790 & 0 & 0 \\ 0 & 0 & 0 & 0,1070 & 0,8930 & 0 \\ 0 & 0 & 0 & 0 & 0,5790 & 0,4210 \\ 0 & 0 & 0 & 0 & 0 & 1,0000 \end{pmatrix}$$

Análogamente a como se procedió en los análisis para los otros dos campos de la formación, las respectivas matrices fundamentales son las siguientes:

	I	PD1	PD2	PD3	PD4	(Suma por fila)-1
I	1,12	0,39	0,19	0,02	0,01	0,75
PD1	0	1,34	0,65	0,08	0,03	1,11
PD2	0	0	1,76	0,22	0,08	1,08
PD3	0	0	0	2,25	0,87	2,12
PD4	0	0	0	0	3,5	2,50

Cuadro 3.49: Matriz Fundamental de la cadena del campo de la Práctica Docente para el caso de estudiantes que desertaron.

	I	PD1	PD2	PD3	Suma por fila
I	1,38	0,90	0,88	0,15	2,32
PD1	0	1,67	1,65	0,27	2,60
PD2	0	0	2,29	0,38	1,68
PD3	0	0	0	1,49	0,49

Cuadro 3.50: Matriz Fundamental de la cadena del campo de la Práctica Docente para el caso de estudiantes que solicitaron un pase.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

	I	PD1	PD2	PD3	PD4	(Suma por fila)-1
I	1	1,08	1,47	1,11	2,37	6,05
PD1	0	1,08	1,47	1,11	2,37	5,05
PD2	0	0	1,47	1,11	2,37	3,96
PD3	0	0	0	1,11	2,37	2,49
PD4	0	0	0	0	2,37	1,37

Cuadro 3.51: Matriz Fundamental de la cadena del campo de la Práctica Docente para el caso de estudiantes que egresaron.

Al igual que en los modelos anteriores, se puede observar que se espera que el estudiante ingresante abandone los estudios en un tiempo promedio inferior a un año. Se espera que solicite un pase después de un tiempo promedio de 2,32 años desde su ingreso, y que egrese después de un tiempo promedio de 6,05 años.

Cuando el estudiante aprueba la Práctica Docente I, se espera que pase un tiempo promedio de 1,11 años antes de abandonar los estudios. El tiempo promedio para solicitar un pase aumenta ligeramente a 2,6 años. Además, se espera que pasen 5,05 años para que el estudiante egrese.

Si un estudiante ha aprobado la Práctica Docente II, se espera que tarde en promedio 1,08 años en abandonar la carrera y 1,68 años en decidir cambiarse de institución. El tiempo promedio para que un estudiante egrese es de 3,96 años, lo cual es más largo que los tiempos esperados para abandonar la carrera o cambiarse de institución.

Aprobando la Práctica Docente III, se espera que tarde en promedio 2,12 años en decidir abandonar sus estudios. Por otro lado, el tiempo promedio para que decida cambiarse de institución es de 0,49 años. En este punto de su carrera, es más probable que el estudiante decida cambiarse de institución que desertar. El tiempo promedio para que un estudiante egrese es de 2,49 años.

Al aprobar el seminario de la Práctica Docente IV, el estudiante completa este campo y no debería demorar en egresar. Sin embargo, se espera que transcurra un promedio de 1,37 años para que esto suceda, lo que indica que aún falta que apruebe otros espacios curriculares para lograr el egreso. Por otro lado, para la deserción se espera un promedio de 2,5 años. Esto sugiere que una vez que el estudiante finaliza la Práctica Docente IV, es más probable que egrese a que abandone los estudios.

A continuación, se calcula el producto entre la matriz fundamental y la submatriz de probabilidades de transición en un paso, desde estados transitorios a recurrentes, con el objetivo de obtener una matriz de acceder eventualmente a los estados absorbentes desde uno transitorio.

3.4. AJUSTES, RESULTADOS Y ANÁLISIS

	D	P	E
I	0.8222	0.0953	0,0823
PD1	0,6428	0,1797	0,1774
PD2	0,4847	0,2346	0,2826
PD3	0,1559	0,1962	0,6477
PD4	0,0527	0,1051	0,8421

Cuadro 3.52: Matriz de acceso eventual a los estados absorbentes desde uno transitorio, para el campo de la Práctica Docente.

De acuerdo con los resultados, el estudiante que ingresa a la carrera tiene una probabilidad del 82,22% de abandonar eventualmente sin haber aprobado Prácticas Docentes, un 9,53% de cambiarse de institución y solo un 8,23% de lograr el egreso.

Por su parte, aquellos estudiantes que logran aprobar la Práctica Docente I, tienen una probabilidad de 64,28% de eventualmente dejar la carrera, 17,97% de cambiarse de institución y 17,74% de egresar.

Después de aprobar la Práctica Docente II, las probabilidades de transición cambian. En concreto, la probabilidad de abandonar la carrera disminuye a 48,47%, mientras que la probabilidad de cambiarse de institución aumenta a 23,46% y la probabilidad de egresar también se eleva a 28,26%.

Después de aprobar la Práctica Docente III, la probabilidad de abandonar la carrera disminuye considerablemente a 15,59%. Por otro lado, la probabilidad de solicitar un pase aumenta ligeramente a 19,62%, mientras que la probabilidad de egresar aumenta significativamente a 64,77%.

Finalmente, si aprueba el cuarto año, aún existe la probabilidad de deserción de un 5,27%, disminuye la probabilidad de cambiarse a otra institución a un 10,51% y aumenta notoriamente la de egresar a un 84,21%.

Es importante tener en cuenta que, aunque las probabilidades de egreso aumentan a medida que el estudiante va aprobando las Prácticas Docentes, y disminuyen considerablemente las probabilidades de deserción, estas probabilidades no reflejan el tiempo promedio necesario para alcanzar estos estados. Además, es importante destacar que estos tiempos promedio se alejan bastante de los cuatro años propuestos en el plan de estudios. Las probabilidades de solicitar un pase van aumentando, excepto cuando el estudiante aprueba la Práctica Docente II. Es necesario tener en cuenta tanto las probabilidades de transición como los tiempos promedio para tener una visión más completa de la situación.

Se puede apreciar, que la probabilidad de egresar eventualmente aumenta si el estudiante aprueba la Práctica Docente III. Como se mencionó con anterioridad, la aprobación de este espacio curricular, requiere de la regularidad de espacios curriculares de la Formación Específica de primer año.

Capítulo 4

Un modelo alternativo: el Análisis de Supervivencia

En este capítulo, se realiza un estudio sobre otro método estadístico utilizado para analizar las probabilidades de que un estudiante abandone los estudios: el análisis de supervivencia. Este método aparece con frecuencia en la investigación bibliográfica llevada a cabo sobre deserción estudiantil. Se detalla brevemente el marco teórico necesario para construir un modelo que describa la problemática abordada en este trabajo. El objetivo es, al finalizar este capítulo, comparar el ajuste por cadenas de Markov y el análisis de supervivencia para enumerar las ventajas y desventajas de su aplicación en función del análisis que se pretende realizar en este trabajo.

4.1. Generalidades del análisis de supervivencia

El análisis de supervivencia es un conjunto de procedimientos estadísticos utilizados para analizar datos en los que la variable de interés es el tiempo hasta que ocurre un evento específico en una unidad de estudio. Por lo general, se analiza el tiempo de supervivencia hasta que los individuos experimentan un evento de fracaso, que puede incluir la muerte, la recaída en una enfermedad, el fallo de un dispositivo médico, entre otros.

La principal diferencia entre un modelo de regresión logística y un análisis de supervivencia es que, en el primero, se estudia una variable respuesta binaria que registra la ocurrencia o no de un evento particular en un periodo de tiempo fijo, es decir, sin tener en cuenta el momento exacto en que ha sucedido el evento. Por su parte, en el análisis de supervivencia se tiene en cuenta, no sólo si el evento ha sucedido o no, sino también cuándo ha ocurrido.

La mayoría de los análisis de supervivencia deben considerar un problema llamado censura. La censura ocurre cuando se conoce información sobre el tiempo de supervivencia individual, pero no de manera exacta.

En el presente trabajo, un ejemplo sería la trayectoria de un estudiante que se sigue hasta el periodo observado, y no ha accedido al evento de interés (deserción), entonces el tiempo de supervivencia de estos estudiantes se considera censurado. Se sabe que el tiempo de supervivencia es al menos tan largo como el periodo observado para el estudiante, pero si el mismo abandona la carrera después del periodo de observación, no se conoce el tiempo de supervivencia completo. También puede ocurrir, que el estudiante puede acceder a otro estado diferente al de interés, durante el periodo observado, como ser cambiarse de institución o egresar.

4.2. Notación y terminología

- T : variable aleatoria que indica el tiempo de supervivencia de un individuo.
En un análisis de supervivencia, la variable en estudio es el tiempo transcurrido hasta que ocurre un evento de interés. Por ello, debe estar claramente definido e indicar de forma precisa cuál es el momento inicial.
- t : valor específico que toma la variable aleatoria T . Es una realización de T .
- t_i+ : observación censurada o realización de la variable aleatoria T censurada, sobre el i -ésimo individuo.

En estudios longitudinales puede ocurrir que algún individuo lo abandone antes de que se produzca el fracaso, por lo que sólo se tendrá información parcial sobre el tiempo en que se produce el fracaso. Existen diversos tipos de censura: censura por la derecha, por la izquierda o sobre un intervalo.

- Censura por la derecha: se produce cuando el individuo no alcanzó el evento de fracaso, finalizado el tiempo de observación, pudiendo haber alcanzado un tiempo de supervivencia real que es mayor o igual al tiempo considerado durante el estudio.
- Censura por la izquierda: se produce cuando el individuo ingresa al estudio, sin conocerse el tiempo inicial y, en consecuencia, su tiempo de supervivencia no puede estimarse de manera precisa.
- Censura sobre un intervalo: se produce cuando el individuo se retira del estudio en un intervalo de tiempo incluido en el periodo de observación, perdiendo así información acerca de si accedió al evento de fracaso o no durante dicho periodo.

4.2. NOTACIÓN Y TERMINOLOGÍA

De esta manera, se incluye en el análisis, la siguiente variable aleatoria indicadora de la censura:

$$D = \begin{cases} 1, & \text{si ocurre el evento de interés sobre el individuo.} \\ 0, & \text{si el tiempo de supervivencia es censurado.} \end{cases} \quad (4.1)$$

- S : Función de Supervivencia. $S(t)$ indica la probabilidad de que un individuo sobreviva un tiempo mayor que un tiempo específico t . Es decir, $S(t)$ es la probabilidad de que la variable aleatoria T tome valores mayores que el tiempo especificado t .

$$S(t) = P(T > t)$$

En la práctica, la gráfica de la función de supervivencia (conocida como *curva de supervivencia*) resulta ser el de una función escalonada en los casos en que se considera un conjunto numerable de tiempos.

- h : Función de Riesgo. $h(t)$ indica la tasa de variación instantánea del tiempo de supervivencia, dado que el individuo ha sobrevivido hasta el tiempo t . Se observa que en el numerador se tiene la probabilidad de que sobre el individuo ocurra el evento de interés en $[t, t + \delta t]$ dado que ha sobrevivido hasta t . Por esto la función de riesgo también es conocida como tasa de fracaso condicional.

Si T es una variable aleatoria continua, la función de riesgo se define como:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

En caso de que T sea una variable aleatoria discreta, y tome los valores $t_1 < t_2 < \dots$, con función de masa de probabilidad $p(t_j)$, entonces la función de riesgo se define como:

$$h(t_j) = P(T = t_j \mid T > t_{j-1})$$

Mientras que la función de supervivencia modela el no-fracaso, la de riesgo se centra en el fracaso, es decir, en la ocurrencia del evento de interés.

De esta manera, la relación entre la función de supervivencia y la de riesgo es:

- Para el caso de T continua:

$$S(t) = e^{-\int_0^t h(u)du}$$

y

$$h(t) = -\frac{\frac{dS(t)}{dt}}{S(t)}$$

- Para el caso de T discreta:

$$\begin{aligned} h(t_j) &= \frac{p(t_j)}{S(t_{j-1})} \\ &= \frac{S(t_{j-1}) - S(t_j)}{S(t_{j-1})} \\ &= 1 - \frac{S(t_j)}{S(t_{j-1})} \end{aligned} \tag{4.2}$$

y

$$S(t) = \prod_{j/t_j \leq t} (1 - h(t_j)) \tag{4.3}$$

4.3. Curvas de supervivencia de Kaplan-Meier

El estimador de Kaplan-Meier se basa en los tiempos de supervivencia individuales y asume que la censura es independiente del tiempo de supervivencia. Es decir, la razón por la que se censura una observación no está relacionada con la causa por la que se produce el evento de interés.

Este estimador se relaciona con la descomposición de la función de supervivencia en un producto de probabilidades condicionadas (ecuación 4.3). Como método no paramétrico, no asume una distribución en particular para la variable tiempo:

$$\hat{S}(t) = \begin{cases} 1, & \text{si } t < Y_{(1)} \\ \prod_{j/t_j \leq t} (1 - \frac{d_j}{n_j}), & \text{si } t \geq Y_{(1)} \end{cases} \tag{4.4}$$

donde d_j representa el número de individuos que registraron el evento de interés en el momento t_j ; n_j representa el número total de individuos en riesgo en el momento t_j , es decir, que su tiempo es mayor o igual que t_j . Al número de individuos en riesgo se lo indica con $R(t)$.

4.4. UNA SOLUCIÓN COMPETITIVA EMPLEANDO ANÁLISIS DE SUPERVIVENCIA

Un diseño de datos alternativo, que constituye la base para obtener las curvas de supervivencia por el método de Kaplan-Meier, es obtener los tiempos de supervivencia ordenados de menor a mayor, los conteos de frecuencias de fracasos en cada tiempo de fracaso distinto. Determinar en cada caso, la frecuencia de aquellas personas censuradas en el intervalo de tiempo que comienza con el tiempo de fracaso hasta, pero sin incluir, el siguiente tiempo de fracaso. Y por último, considerar el número de individuos que han sobrevivido al menos hasta el momento de fracaso.

Para estimar la probabilidad de supervivencia en un momento, se hace uso del riesgo fijado para ese momento, para incluir la información que se tiene del individuo censurado hasta el momento de la censura, en lugar de descartar al sujeto censurado.

Si se desea evaluar si las curvas de Kaplan-Meier para dos o más grupos, son o no estadísticamente iguales, puede utilizarse el test Log-Rank. Este test, para una muestra lo suficientemente grande, utiliza el test chi-cuadrado como criterio de una prueba estadística que proporciona una comparación general de las curvas de Kaplan-Meier que se consideran. Como en otras pruebas, se analizan las discrepancias entre los valores observados frente a los esperados bajo supuesto de independencia.

Dados dos o más grupos, las hipótesis que se analizan con este test son:

$$H_0 : S_1(t) = S_2(t) = \dots = S_k(t), \quad \forall t \leq \tau$$

$$H_1 : S_i(t_0) \neq S_j(t_0), \quad \text{para algún par de valores } i \neq j \text{ y } t_0$$

4.4. Una solución competitiva empleando Análisis de Supervivencia

A continuación, se presentan los resultados obtenidos al aplicar el método de Kaplan-Meier al conjunto de datos que describe las trayectorias estudiantiles en el Profesorado de Educación Secundaria en Matemática del Instituto de Enseñanza Superior Simón Bolívar.

El propósito de esta sección es comparar las metodologías de trabajo basadas en cadenas de Markov y análisis de supervivencia, con el objetivo de analizar las ventajas y desventajas de ambas alternativas para abordar la problemática en estudio.

4.4. UNA SOLUCIÓN COMPETITIVA EMPLEANDO ANÁLISIS DE SUPERVIVENCIA

4.4.1. Metodología de Trabajo

La muestra utilizada en este estudio se compone de los estudiantes que se matricularon en la carrera entre los años 2014 y 2019, ambos inclusive, y se ha analizado la trayectoria del campo de la Formación Específica. Esta trayectoria se seleccionó debido a que presenta un mayor retraso en el avance por parte de los estudiantes y, además, cuenta con el mayor número de estudiantes que no presentaron equivalencias.

Con el fin de llevar a cabo el análisis, se creó una nueva base de datos en la que se registró información específica sobre cada uno de los 504 estudiantes considerados:

- Ciclos: tiempo (en años) que le demanda al estudiante, desde que ingresa a la carrera, alcanzar alguno de los estados absorbentes (*deserción, pedir un pase o egresar*). En caso de que el estudiante continúe en el trayecto al finalizar el periodo de observación, se considera el tiempo observado. Cabe aclarar que se establece el ingreso de cada estudiante en el año anterior al verdadero, para aquellos casos en los que se observe el alcance de algún estado durante el año verdadero de ingreso.
- Status: estado del estudiante alcanzado en el periodo de observación (D: *Deserción*, P: *Pide un pase para cursar la misma carrera en otra institución de educación superior*, E: *Egresar de la carrera* o C: *Continúa en el trayecto*).
- *FE_i*: Variable indicadora de aprobación de los espacios curriculares del campo de la Formación Específica correspondiente al *i*-ésimo año del plan de estudios ($i = 1, 2, 3, 4$). (1: *aprobó todos los espacios curriculares de FE_i*, 0: *no aprobó todos los espacios curriculares de FE_i*).

El tiempo de supervivencia de un estudiante se define como el período de tiempo que permanece en la cursada. En este estudio, el evento de interés es la deserción, y se considera censura cuando el estudiante abandona el trayecto por motivos distintos a la deserción, como el pase a otra institución, el egreso o la continuación en el trayecto.

4.4.2. Ajuste del modelo de Kaplan-Meier

El ajuste de la función de supervivencia por el método de Kaplan-Meier, devuelve los siguientes resultados:

4.4. UNA SOLUCIÓN COMPETITIVA EMPLEANDO ANÁLISIS DE SUPERVIVENCIA

time	n.risk	n.event	survival	std.err	lower 95 % CI	upper 95 % CI
1	504	227	0.550	0.0222	0.508	0.595
2	275	65	0.420	0.0220	0.379	0.465
3	174	55	0.287	0.0211	0.249	0.332
4	83	19	0.221	0.0210	0.184	0.267
5	43	3	0.206	0.0213	0.168	0.252
6	23	2	0.188	0.0229	0.148	0.239

Cuadro 4.1: Ajuste de la función de supervivencia.

En la salida, puede observarse el tiempo de observación, el número de estudiantes en riesgo de deserción, el número de estudiantes que desertaron, la estimación de la función de supervivencia, la desviación estándar de la estimación y los intervalos de confianza para dicha estimación.

El siguiente gráfico muestra la función de supervivencia ajustada:

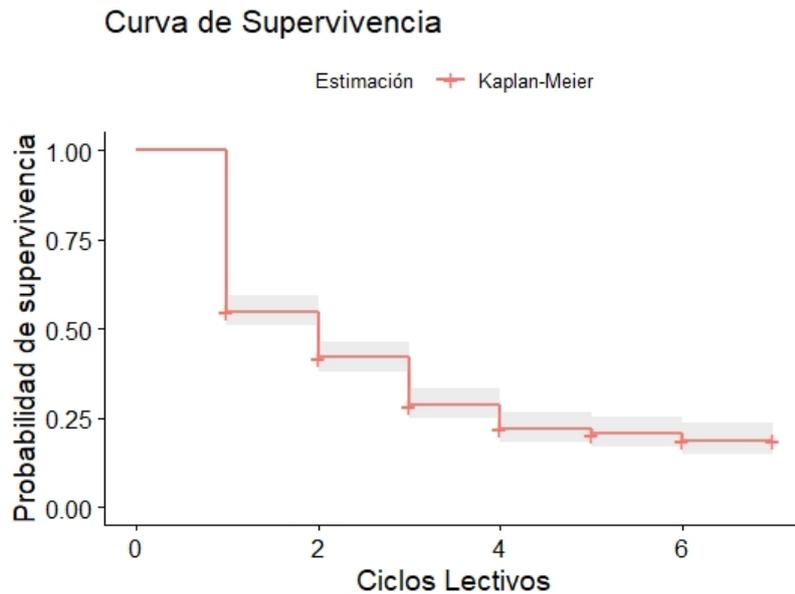


Figura 4.1: Función de supervivencia. Estimación: Kaplan-Meier.

A partir del ajuste del modelo, se pueden calcular tanto la media como la mediana de los tiempos de supervivencia. En este caso, el tiempo medio que transcurre hasta que los estudiantes desertan de la carrera es de 1,87 años. Asimismo, se encuentra que la mediana de los tiempos de supervivencia es de 1 año, lo que indica que el 50% de los estudiantes abandonan los estudios antes de finalizar su primer año de cursada.

4.4. UNA SOLUCIÓN COMPETITIVA EMPLEANDO ANÁLISIS DE SUPERVIVENCIA

n	events	*rmean	*se(rmean)	median	0.95LCL	0.95UCL
504	371	2.872	0.108	2	2	2

Cuadro 4.2: Media y mediana de los tiempos de supervivencia.

Si se incluye la variable indicadora $FE1$, se puede analizar si la aprobación del primer año del campo de la Formación Específica tiene un efecto significativo en los tiempos de supervivencia de los estudiantes. Para ello, se realiza un nuevo ajuste que considera a $FE1$ como un factor. La variable $FE1$ toma el valor 0 si el estudiante aún no ha aprobado todos los espacios curriculares del primer año de la Formación Específica, mientras que toma el valor 1 si ya los aprobó.

El ajuste devuelve los siguientes resultados:

	n	events	median	0.95LCL	0.95UCL
FE1=0	441	362	1	1	2
FE1=1	63	9	NA	NA	NA

Cuadro 4.3: Media y mediana de los tiempos de supervivencia habiendo aprobado o no $FE1$.

En el caso de $FE1 = 1$, es decir del grupo de estudiantes que aprobó $FE1$, no es posible definir a la mediana de los tiempos de supervivencia, puesto que la función de supervivencia no toma el valor 0,5 en algún valor de su dominio.

FE1=0						
time	n.risk	n.event	survival	std.error	lower 95 % CI	upper 95 % CI
1	441	227	0.4853	0.0238	0.4408	0.534
2	212	64	0.3388	0.0226	0.2973	0.386
3	117	51	0.1911	0.0201	0.1555	0.235
4	42	16	0.1183	0.0190	0.0864	0.162
5	19	3	0.0996	0.0188	0.0688	0.144
6	10	1	0.0897	0.0194	0.0587	0.137

Cuadro 4.4: Ajuste de la función de supervivencia sin aprobar $FE1$.

4.4. UNA SOLUCIÓN COMPETITIVA EMPLEANDO ANÁLISIS DE SUPERVIVENCIA

FE1=1						
time	n.risk	n.event	survival	std.error	lower 95 % CI	upper 95 % CI
2	63	1	0.984	0.0157	0.954	1.000
3	57	4	0.915	0.0364	0.846	0.989
4	41	3	0.848	0.0502	0.755	0.952
6	13	1	0.783	0.0780	0.644	0.952

Cuadro 4.5: Ajuste de la función de supervivencia aprobando $FE1$.

Para comparar si existen diferencias significativas entre los tiempos de supervivencia de los dos grupos, se realiza la prueba de comparación Log-Rank.

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
FE1=0	441	362	297.5	14.0	116
FE1=1	63	9	73.5	56.6	116
Chisq= 116 on 1 degrees of freedom, p= <2e-16					

Cuadro 4.6: Prueba de comparación Log-Rank entre el grupo que no aprobó $FE1$ y el que sí lo aprobó.

Como se obtuvo un p -valor menor a 0,05; se concluye que se rechaza la hipótesis nula de igualdad entre funciones de supervivencia. Luego, hay diferencias significativas entre estas dos funciones.

El gráfico de ambas curvas de supervivencia se muestra a continuación:

4.5. COMPARACIÓN ENTRE CADENAS DE MARKOV Y ANÁLISIS DE SUPERVIVENCIA

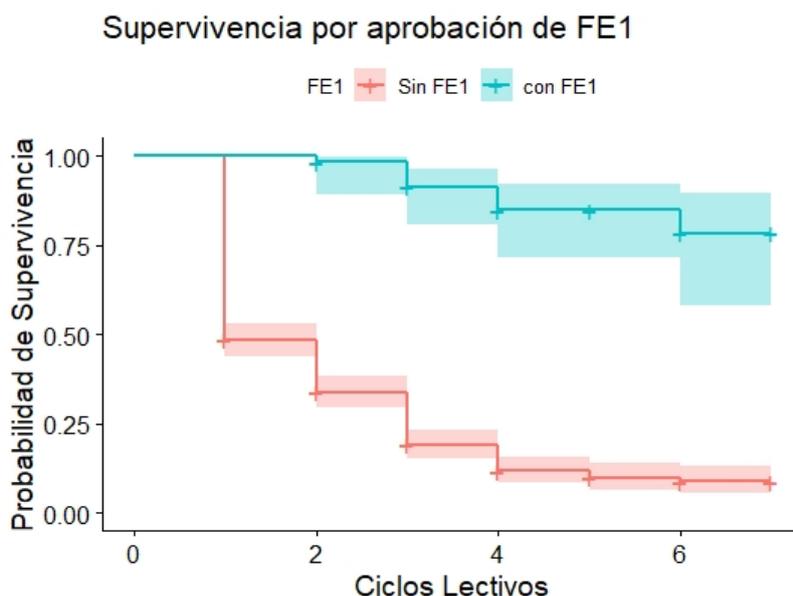


Figura 4.2: Curvas de supervivencia por aprobación de FE1.

4.5. Comparación entre Cadenas de Markov y Análisis de Supervivencia

En esta sección se realiza una comparación entre el uso de cadenas de Markov y el análisis de supervivencia para abordar la problemática analizada. Se evalúa el aporte de los resultados obtenidos a través de ambas metodologías, considerando la trayectoria estudiantil en el campo de la Formación Específica y la aprobación o no de sus espacios, así como el tiempo en que los estudiantes logran esas aprobaciones.

En primer lugar, es evidente la cantidad de información que se puede obtener mediante el uso de una cadena de Markov. En el caso de este conjunto de datos, se pueden calcular las probabilidades de alcanzar los diferentes estados a partir de uno dado, utilizando una única cadena. Además, al estimar la matriz de transición, es posible conocer las probabilidades de transición para más de un periodo. Por otro lado, en el análisis de supervivencia, solo se puede elegir un estado como evento de interés y estimar la probabilidad de no alcanzarlo. Si se quisiera estimar la probabilidad de que un estudiante alcance o no otros estados (como solicitar un pase o graduarse), se deberían utilizar otros modelos de supervivencia. Es importante destacar que el análisis de supervivencia no permite calcular la probabilidad

4.5. COMPARACIÓN ENTRE CADENAS DE MARKOV Y ANÁLISIS DE SUPERVIVENCIA

de alcanzar estados transitorios antes de llegar a cualquiera de los tres estados absorbentes.

Otro aspecto a considerar es que, aunque ambos modelos se refieren al mismo periodo de tiempo, existen diferencias en cuanto a la unidad de análisis utilizada. Las cadenas de Markov contabilizan las transiciones entre estados que ocurren en un intervalo de tiempo determinado, mientras que el análisis de supervivencia se construye considerando el trayecto completo de cada estudiante. Es decir, en el segundo caso se utiliza un menor número de información, ya que se pierden los datos correspondientes a los estudiantes que ya estaban en la carrera en el año inicial (2014). En otras palabras, para el ajuste de las cadenas de Markov, los individuos censurados no son relevantes, ya que solo importa la transición entre estados que los estudiantes hayan realizado.

En cuanto a los resultados obtenidos, se observa que la probabilidad de deserción de los estudiantes presenta una diferencia significativa entre ambos modelos durante el primer año de cursada (aproximadamente 9% desde que el estudiante ingresa). Sin embargo, a partir del segundo año de cursada, las estimaciones de la probabilidad de deserción no presentan diferencias significativas, ya que las obtenidas mediante cadenas de Markov se encuentran dentro del intervalo de confianza de las estimaciones obtenidas mediante el análisis de supervivencia.

Se observa un comportamiento similar si se comparan las estimaciones obtenidas mediante Cadenas de Markov, tomando en cuenta la aprobación de *FE1*, y las obtenidas mediante el Análisis de Supervivencia ajustado para los estudiantes que alcanzaron el mismo estado. Se registran diferencias significativas en el primer año después de la aprobación, mientras que a partir del segundo año de aprobación no se observan diferencias significativas en estas estimaciones.

Con ambos modelos, se puede concluir que la probabilidad de deserción disminuye considerablemente, si el estudiante aprueba *FE1*.

Se ajustó un modelo de supervivencia de Kaplan-Meier considerando al grupo de estudiantes que había aprobado los espacios de la Formación Específica de segundo año, *FE2*, obteniendo como resultados el mismo que con la cadena de Markov. En ambos ajustes, se obtiene que la probabilidad de abandonar los estudios es nula.

Se llevó a cabo un ajuste del Modelo de Regresión de Cox de Riesgos Proporcionales, el cual permitió obtener información específica sobre el riesgo de desertar incorporando covariables, como el número de años que tarda el estudiante en aprobar *FE1*. Este ajuste permitió evaluar cómo varía el riesgo por cada año de retraso en el avance por el trayecto. Cabe destacar que esta información es difícil de comparar con el ajuste de la Cadena de Markov. Sin embargo, se descartó el ajuste realizado debido a que no verificaba el supuesto de riesgos proporcionales. Además, como se mencionó anteriormente, sería necesario plantear más de un modelo de su-

4.5. *COMPARACIÓN ENTRE CADENAS DE MARKOV Y ANÁLISIS DE SUPERVIVENCIA*

pervivencia para observar detalladamente las probabilidades de acceso y no acceso a cada uno de los estados definidos.

Capítulo 5

Conclusiones y Recomendaciones

En este capítulo se aborda la metodología empleada en el estudio, evaluando su eficacia y limitaciones. A continuación, se interpretan los resultados obtenidos y se discuten en detalle las implicaciones de estos hallazgos. Es importante destacar las contribuciones que se han hecho al campo de la investigación educativa y cómo estos resultados pueden ser utilizados para mejorar el conocimiento y la práctica en este ámbito. Además, se presentan recomendaciones concretas para futuras investigaciones y se discuten las posibles limitaciones del estudio.

5.1. Importancia de las Cadenas de Markov

Una cadena de Markov a tiempo discreto es un proceso estocástico que se caracteriza por una propiedad clave, conocida como propiedad de Markov, según la cual la probabilidad de que ocurra un evento depende solamente del evento inmediatamente anterior. Esto significa que la evolución del proceso estocástico se realiza en etapas discretas, y en cada una de ellas se considera solamente la información más reciente. En contraposición, en los eventos independientes no hay relación entre los eventos, por lo que la ocurrencia de uno no tiene memoria de ningún otro evento anterior. Esta propiedad de Markov hace que las cadenas de Markov sean una herramienta poderosa para modelar sistemas complejos y procesos aleatorios en una amplia variedad de campos.

Las cadenas de Markov a tiempo discreto son una herramienta útil para modelar la dinámica de la trayectoria académica de un estudiante en una carrera determinada. Este tipo de modelo permite estimar las probabilidades de que un estudiante alcance diferentes estados de la carrera, como aprobar ciertos espacios curriculares o campos de formación, abandonar la institución o egresar, entre otros posibles estados. Al ingresar a la carrera, el estudiante se encuentra en un estado inicial, y el modelo calcula la probabilidad de que el estudiante pase a diferentes

estados a medida que avanza en la carrera. La principal ventaja de las cadenas de Markov, en este contexto, es que el estado actual sólo depende del estado inmediato anterior y no de los demás considerados, lo que es crucial para modelar la dinámica compleja de la trayectoria académica de un estudiante. De esta manera, esta metodología tiene un potencial modelizador en estudios de evaluación y seguimiento de la calidad educativa, así como en la toma de decisiones en materia de investigación educativa.

Para ilustrar el poder de las cadenas de Markov en el contexto educativo, se seleccionó la carrera de Profesorado para la Educación Secundaria en Matemática, ofrecida por el Instituto de Enseñanza Superior Simón Bolívar en la Provincia de Córdoba. Esta carrera enfrenta la problemática de registrar cada año altos índices de abandono escolar, lo que representa un desafío importante para la institución educativa y la política educativa en general. Muchos de los ingresantes a la carrera no logran completarla en el tiempo estipulado, y algunos optan por continuar sus estudios en otras instituciones. La aplicación de cadenas de Markov a la trayectoria académica de los estudiantes en esta carrera puede proporcionar información valiosa sobre los patrones de deserción, así como sobre los posibles caminos de progreso para aquellos estudiantes que logran superar los obstáculos iniciales. Este análisis podría servir como base para la implementación de políticas y programas de intervención efectivos para mejorar la calidad educativa y reducir la tasa de abandono en esta y otras carreras similares.

Estos porcentajes preocupan a la institución educativa mencionada y, por ello, propone diversas soluciones para abordar este problema. A pesar de que una vez que los estudiantes abandonan sus estudios no hay marcha atrás, se pueden identificar los momentos más críticos en su trayectoria académica para que la institución pueda prestarles mayor atención y así reducir la tasa de deserción.

Debido que los estudiantes de esta carrera acceden a tres estados: *deserción*, *pedir un pase para cursar la misma carrera en otra institución* y *egreso*, y de estos estados no se registran accesos a retomar la carrera; es que se los considera como *estados absorbentes*. Por su parte, los demás estados del alumnado durante la carrera, constituyen *estados transitorios*.

De acuerdo a como se planteó en el marco teórico, el espacio de estados de una cadena de Markov, se particiona en estados transitorios y recurrentes. En el caso en que todos los estados recurrentes sean absorbentes, la cadena de Markov recibe el nombre de *cadena absorbente*. Este modelo permitirá realizar el ajuste para predecir las probabilidades de acceso de los estudiantes a estos estados absorbentes, así como también el tiempo medio de absorción.

Para el estudio, se consideraron las trayectorias estudiantiles de las cohortes desde el año 2014 hasta el 2019. En este trabajo se plantearon modelos mediante cadenas de Markov a tiempo discreto, con espacio de estados finito, para describir

la dinámica de la trayectoria estudiantil por esta carrera.

En un primer momento, explorando alternativas que finalmente se descartaron, se ajustaron diferentes modelos los cuales no verificaban los supuestos de Markov o no permitían visualizar correctamente la trayectoria estudiantil por el plan de estudios del profesorado en cuestión.

A través de la visualización de los datos, se identificó que hay una mayor tasa de aprobación de materias en los campos de la Formación General y de la Práctica Docente, mientras que en el campo de la Formación Específica se presentan mayores retrasos en su aprobación o, en muchos casos, pocos espacios aprobados. Esto se ha observado al analizar las historias académicas de los estudiantes desertores o que cambiaron de institución para cursar la misma carrera.

Para estudiar en qué momentos de la carrera se produce el mayor porcentaje de deserción y la demora en que el estudiante avance, se consideraron cadenas de Markov, con las transiciones por cada campo de la formación propuesto en el plan de estudios de la carrera.

De esta manera, se ajustaron 3 modelos, para cada uno de los tres campos: de la Formación General, de la Formación Específica y de la Práctica Docente. Por su parte, en el espacio de estados se consideró el estado de ingresante I , y paquetes de materias de cada campo de acuerdo al año de cursado en el que aparece en el plan de estudios.

Como estados absorbentes se plantearon los estados D (el estudiante abandona sus estudios), P (el estudiante pide un pase para continuar sus estudios de la misma carrera en otra institución de educación superior) y E (el estudiante egresa al culminar la aprobación del plan de estudios propuesto para la carrera).

Las probabilidades de transición, y por consiguiente, las matrices de transición, se obtuvieron con el estimador de máxima verosimilitud de los conteos de estas transiciones. En este caso, cabe mencionar que estas probabilidades pueden ajustarse con algún otro modelo que incorpore diversos factores, los cuales podrían ser influyentes en que los estudiantes accedan a cada estado absorbente. Sin embargo, no fue posible realizar esta propuesta, ya que la Institución carece de datos de esta índole, de sus estudiantes ingresantes.

5.2. Modelo empleando cadenas de Markov

En cuanto a la trayectoria estudiantil en el campo de la Formación General, se ha observado un alto porcentaje de deserción al comienzo de la misma. Al analizar el tiempo promedio que los estudiantes tardan en abandonar los estudios, se obtuvo que es de 0,88 años, lo que significa que esto ocurre en algún momento del primer año de cursada. Por otro lado, el tiempo promedio para cambiar de institución es de 3,46 años, mientras que para egresar debería pasar una media de 5,87 años.

Cabe destacar que, incluso después de completar este campo, los estudiantes deben esperar para poder egresar, ya que todavía deben aprobar espacios curriculares de otros campos de la Formación.

En cuanto a la trayectoria estudiantil en el campo de la Formación Específica, se ha observado un alto porcentaje de deserción antes de completar el primer año de este campo, el cual alcanza un 82 %. Sin embargo, este porcentaje disminuye considerablemente al aprobar el primer año del campo y se vuelve nulo una vez que se han aprobado los espacios específicos correspondientes al segundo año. Desde que el estudiante ingresa hasta que alcanza alguno de los estados absorbentes, se obtuvo que el tiempo medio para abandonar los estudios es de 0,97 años, el de cambiarse de institución de 3,31 años mientras que para egresar debería pasar una media de 5,42 años.

Se observa que completado este campo, el estudiante logra recibirse, debido a que todos los estudiantes se reciben con materias del campo de la Formación Específica. Se observa también, que conforme el estudiante va avanzando por este campo, la probabilidad de cambiarse de institución va disminuyendo hasta que es nula a partir de la aprobación de las materias del tercer año.

Otro detalle que se observó es que las materias del tercer y cuarto año de este campo, se aprueban en tiempos muy cortos hacia el final de la trayectoria completa, con lo que indicaría que estos espacios son aprobados juntos hacia el final de la carrera.

Con respecto a la trayectoria por el campo de la Práctica Docente, se obtuvieron resultados similares a los obtenidos para el campo de la Formación General, observando un gran porcentaje de deserción hacia el comienzo de la carrera. Para abandonar los estudios se espera que transcurra un tiempo promedio de 0,75 años, para cambiarse de institución una media de 2,32 años y para egresar de 6,05 años. En este caso también se observa que completado este campo, el estudiante debe esperar algunos años para egresar, debido a que aún adeudan la aprobación de espacios curriculares de otros campos de la Formación. Cabe mencionar que se observa un pequeño retraso en la aprobación de las Prácticas Docentes III y IV, que son los espacios que requieren tener regularizadas algunas materias del campo de la Formación Específica.

Como se obtuvo un menor tiempo promedio para la deserción en el campo de la Formación Específica, se podría concluir que este campo es el que genera mayor porcentaje de estudiantes que abandonan la carrera.

En cuanto al tiempo promedio de cambios de institución, se ha observado que se mantienen en los campos de la Formación General y de la Práctica Docente, y se reducen a 0 una vez que se ha aprobado el tercer año de la Formación Específica. Esto indica que estos cambios de institución se producen debido a los largos tiempos de aprobación de las materias disciplinares.

Al analizar los tres modelos, se puede observar que los estudiantes, en general, aprueban en primer lugar las materias del campo de la Formación General y de la Práctica Docente, llegando hasta las de tercer año. Luego, continúan con los espacios curriculares específicos de primer año, seguidos por las materias generales de cuarto año y la Práctica Docente IV. Finalmente, se aprueban los espacios curriculares de la Formación Específica de segundo, tercer y cuarto año en un intervalo de tiempo más acotado, pero algo separado desde la última aprobación.

El campo de la Formación Específica es el que presenta un mayor retraso promedio para su culminación por lo que éste es el que indica cuál es el retraso promedio total para la obtención del título.

5.3. Cadenas de Markov vs. Análisis de Supervivencia

Se propuso un modelo que permitiera abordar la problemática de manera efectiva y competitiva en comparación con las cadenas de Markov. Para ello, se planteó una alternativa basada en un análisis de supervivencia mediante el ajuste de Kaplan-Meier. El objetivo principal de plantear esta otra metodología estadística fue establecer una comparación con el modelo de cadenas de Markov y determinar cuál de ellos sería más adecuado para la problemática en cuestión. En la búsqueda de material bibliográfico sobre investigaciones previas, este modelo competitivo se registró con mayor frecuencia.

Esta alternativa permite obtener la probabilidad de que un individuo acceda a un evento de interés (generalmente identificado como un evento de fracaso), en determinado tiempo. De esta manera, el ajuste realizado, consideró el evento de acceder al estado de *deserción*. Sin embargo, el planteo de este modelo, enfrenta el problema de las observaciones censuradas, es decir, el problema de conocer el tiempo de supervivencia individual, pero no de manera precisa. Por este motivo, se consideró un ajuste de Kaplan-Meier, el cual asume que la censura es independiente del tiempo de supervivencia.

La primera desventaja que se observa al comparar ambos modelos, es que el ajuste mediante análisis de supervivencia de Kaplan-Meier no permite obtener toda la información que brinda el ajuste de una cadena de Markov. Aquí se muestra la probabilidad de deserción en el tiempo transcurrido, mientras que con las cadenas es posible visualizar no sólo esta probabilidad, sino también la de acceder a otros estados absorbentes y la probabilidad de acceder a otros estados intermedios como lo son los estados transitorios considerados.

Otra desventaja a señalar de los análisis de supervivencia es que, a pesar de considerar el mismo intervalo de tiempo, debe considerarse la trayectoria del estu-

5.4. COMPARACIÓN DE LOS RESULTADOS OBTENIDOS CON LA PERCEPCIÓN INSTITUCIONAL

diente desde que ingresa y que puede presentar censura por la derecha. Por este motivo, para poder llevar a cabo este ajuste, se tuvo que considerar un número menor de estudiantes al considerado en el ajuste de Markov, ya que con este último, las unidades de observación son las transiciones entre estados de cada estudiante.

De todas maneras, al realizar las comparaciones de resultados mediante ambos modelos, se pudo ver que había diferencias estadísticamente significativas sólo en el primer año de cursado.

El ajuste mediante cadenas de Markov, es menos costoso en cuanto a tiempo de construcción de la base de datos y economía de resultados. Mediante un único modelo ajustado se puede visualizar mucha información relevante para la toma de decisiones frente a esta problemática. Por su parte, el análisis de supervivencia permite visualizar información de manera directa, en cuanto al tiempo transcurrido y el riesgo de acceder al evento de interés. Sin embargo, para obtener la misma información que se visualiza mediante la cadena de Markov, en este caso, sería necesario el planteo de varios modelos de análisis de supervivencia.

5.4. Comparación de los resultados obtenidos con la percepción Institucional

Antes de llevar a cabo este trabajo, se consultó al personal directivo, del Instituto de Enseñanza Superior Simón Bolívar, si cuenta con informes de las trayectorias académicas sobre el profesorado analizado. En cuanto a la respuesta obtenida, se comunicó que los únicos datos que manejan son los correspondientes al sistema de gestión de alumnos, los cuales no aparecen resumidos o procesados. Por este motivo, recomendaron que se entrevistase a la Coordinadora de Curso designada para la carrera, quien está más cerca de percibir la dinámica en la trayectoria de estudiantes, así como también las problemáticas que en la misma se detectan.

Comparando los resultados obtenidos con la percepción de la coordinadora de curso, puede verse que coinciden en un amplio porcentaje.

Tanto desde el punto de vista de la Coordinadora, como de los resultados obtenidos en este trabajo, se observa el gran porcentaje de deserción durante el primer año de cursada. La coordinadora, menciona dos momentos que no se pueden visualizar en el ajuste, el momento de inicio de la cursada y el regreso del receso invernal. Mientras que en el ajuste realizado, se observa en los tres campos de la formación, que el tiempo medio de deserción es durante el primer año de cursada, y se ubica aproximadamente a mediados del segundo cuatrimestre del primer año.

Al igual que con los resultados obtenidos en este trabajo, no se observan retrasos en el avance por el campo de la Formación General, observando en ambos casos, el retraso entre la aprobación de *FG4* y el egreso.

En cuanto al campo de la Formación Específica, hay pocas diferencias en cuanto a los resultados obtenidos y la percepción de la coordinadora. Se observan grandes retrasos en la aprobación de los espacios específicos de los primeros años y la aprobación de los espacios específicos de tercero y cuarto hacia el final de la trayectoria.

A diferencia de lo que observa la entrevistada, el riesgo de deserción es nulo, a partir de la acreditación de los espacios específicos de segundo año. Sin embargo, ella advierte que no hay riesgo de abandono, una vez que el alumno acreditó los espacios específicos de tercer y cuarto año. Esta pequeña discrepancia puede deberse a que gran parte del estudiantado se recibe al rendir en última instancias espacios específicos de segundo, tercero y cuarto año.

Al igual que en los resultados obtenidos, el riesgo de cambiarse de institución para continuar el cursado de la misma carrera, la coordinadora menciona que es nulo, una vez que el estudiante ha acreditado los espacios de la Formación Específica del tercer año.

En relación a la trayectoria estudiantil por el campo de la Práctica Docente, no se vislumbran diferencias significativas. Ambas partes señalan la similitud con el trayecto por el campo de la Formación General y el pequeño retraso en la acreditación de *PD3*, debido a que es el primer espacio que requiere de la regularidad de espacios de la Formación Específica.

En cuanto al tiempo medio en que el estudiante egresa, se percibe una cantidad de años que abarca los resultados estimados, entre 5 y 7 años. En relación al tiempo medio en que el estudiante pide un pase para cursar la misma carrera en otra institución, se presentan diferencias, puesto que desde la coordinación se percibe que transcurre un tiempo medio de entre 5 y 6 años; mientras que por medio del ajuste se estimaron entre 3 y 4 años.

Por último, la coordinadora de curso realiza un análisis del conjunto de estudiantes que logra egresar, similar al observado mediante las estimaciones por cadenas de Markov. Aquí se aprecia que los estudiantes avanzan, en general, según la trayectoria estipulada en el plan de estudios de la carrera, con leves retrasos en el campo de la Formación Específica.

5.5. Recomendaciones y futuras líneas de investigación

En virtud de los resultados obtenidos en este trabajo puede señalarse que el acompañamiento por parte de la Institución debiera realizarse en el primer año de cursado y sobre todo en los primeros dos años del campo de la Formación Específica.

5.5. RECOMENDACIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

Es en estos instantes de la trayectoria estudiantil donde se produce mayor deserción y donde se piden pases para continuar sus estudios en otra institución.

Estos altos porcentajes de deserción podrían deberse a muchos factores, y podría explorarse un poco más la transición del Nivel Medio hacia el Superior.

Para profundizar el análisis de este estudio se podría indagar sobre los siguientes ítems:

- Construir modelos que consideren cada uno de los espacios curriculares de los campos de la Formación, para poner la atención en aquellas materias que requieran mayor acompañamiento.
- ¿Qué factores producen el retraso en la aprobación de los espacios curriculares del campo de la Formación Específica?.
- Variables de mayor influencia en la decisión de abandonar los estudios, ya que esto puede explicar mejor esta problemática.
- Variables de mayor influencia en la decisión de cambiarse de institución para continuar esta carrera, ya que esto puede explicar mejor esta problemática.

La Institución Educativa podría documentar de manera completa estos aspectos, indagando en diversas cuestiones para conocer mayor información de los estudiantes que ingresan. Se recomienda realizar una encuesta exhaustiva al momento de la preinscripción, para recolectar información relevante y así enriquecer el modelo. De esta manera, luego se puede profundizar en qué características predominan entre los estudiantes que abandonan los estudios, así como de los que egresan. Con este análisis se puede realizar un mejor acompañamiento atendiendo a estos factores.

Para el ajuste de las cadenas de Markov, también es útil el conocimiento de estos factores influyentes, ya que con ellos se puede diseñar un mejor modelo en donde se estimen probabilidades de transición más adaptadas.

Análogamente, cuando se detecta que un estudiante no registra actuaciones en el Sistema de Gestión de Alumnos, se podría indagar sobre las causas de esta situación y analizar posibles soluciones para motivarlo a que continúe en la carrera, al igual que con los que deciden cambiarse de institución.

Sin embargo, el modelo así planteado, permite establecer un punto de partida para la toma de decisiones frente a esta problemática; además de ser plausible de replicar en otras trayectorias educativas similares y poder contribuir con resultados más precisos en el área de la investigación educativa.

Sería interesante también analizar características de los alumnos que lograron egresar, ya que este comportamiento podría ayudar para el acompañamiento de aquellos estudiantes que se retrasan o que están por decidir abandonar los estudios.

Bibliografía

- [1] Aguilera, N. (2009) Cadenas de Markov. OMA.
- [2] Adeleke, R. A., Oguntuase, K. A., & Ogunsakin, R. E. (2014). Application of Markov chain to the assessment of students' admission and academic performance in Ekiti State University. *International Journal of Scientific & Technology Research*, 3(7), 349-357.
- [3] Alawadhi, S., & Konsowa, M. (2010). Markov chain analysis and student academic progress: An empirical comparative study. *Journal of Modern Applied Statistical Methods*, 9(2), 26.
- [4] Álvarez, M., & Orrego Posada, R. (2000). Modelo markoviano para el estudio de evolución de cohortes de estudiantes de un programa académico. *Revista Universidad EAFIT*, 36(120), 44-56.
- [5] Amat, J. (2016). Test estadísticos para variables cualitativas. Recuperado de https://rstudio-pubs-static.s3.amazonaws.com/220579_704e04beed0a499c8b4ac26dec006cd1.html.
- [6] Barrio, O. S., & Carmona, A. G. (1981). Cadenas de markov aplicadas al diagnostico y tratamiento de conductas y procesos de aprendizaje. *Enseñanza & Teaching: Revista Interuniversitaria de Didáctica*, 8.
- [7] Barroeta Rojo, C. (2016). Modelos para el análisis de supervivencia en tiempos discretos: aplicación en el área de veterinaria.
- [8] Borges, P., & Eduardo, R. (2005). Análisis de supervivencia de pacientes con diálisis peritoneal. *Rev. Colomb. Estad.*, 243-259.
- [9] Bolivar-Cimé, A., Notario, C., & Pérez, A. (2016). Modelos de Markov para la trayectoria académica de estudiantes de la UJAT.
- [10] Casini, R., Gonzalez, M., Gertel, H., & Giuliadori, R. (2012, August). Métodos estadísticos aplicados para el análisis de deserción y rendimiento académico de

- estudiantes de tres facultades de la Universidad Nacional de Córdoba. In Décimo Congreso Latinoamericano de Sociedades de Estadística.
- [11] Brzezniak, Z., & Zastawniak, T. (2000). Basic stochastic processes: a course through exercises. Springer Science & Business Media.
- [12] Caicedo, R., Atuesta, S. B., & Caicedo, J. P. (2016). Análisis de la retención de estudiantes de ingeniería basado en la pérdida consecutiva de una misma asignatura. Un enfoque de Cadenas de Markov. *Ingeniería Industrial. Actualidad y Nuevas Tendencias*, (16), 7-18.
- [13] Delgado-Moya, E. M., & Marrero-Severo, A. (2018). Estudio estocástico con el uso de cadenas de Markov para la transmisión del dengue. *Uniciencia*, 32(1), 108-117.
- [14] Diseño Curricular. Profesorado de Educación Secundaria en Matemática. Diseño Curricular de la Provincia de Córdoba. (2010).
- [15] Escanés, G. A., Herrero, V., & Merlino, A. (2013). El Estudio de la Permanencia de Estudiantes en la Universidad: Elementos para la Aplicación de un Modelo de Duración.
- [16] Fox, J., & Carvalho, M. S. (2012). The RcmdrPlugin. survival package: Extending the R Commander interface to survival analysis. *Journal of Statistical Software*, 49, 1-32.
- [17] González-Campos, J. A., Carvajal-Muquillaza, C. M., & Aspeé-Chacón, J. E. (2020). Modelación de la deserción universitaria mediante cadenas de Markov. *Uniciencia*, 34(1), 129-146.
- [18] Gracia, E. D. (2000). TES 15 (Doctoral dissertation, UNIVERSIDAD DE SONORA).
- [19] Himmel, E. (2002). Modelo de análisis de la deserción estudiantil en la educación superior. *Calidad en la Educación*, (17), 91-108.
- [20] Hoel, P., Port, P., & Stone, C. J. (1972). *Introduction to Stochastic Processes*. Boston, MA: HoughtonMifflin.
- [21] Hosmer, D. W., & Lemeshow, S. (1999). *Applied survival analysis: time-to-event* (Vol. 317). Wiley-Interscience.
- [22] Ibarra, L. F. (2009). Predicciones de Markov Aplicadas en el Programa de Ingeniería Industrial de la Universidad Nacional Experimental del Táchira (UNET). *Ingeniería Industrial. Actualidad y Nuevas Tendencias*, 1(2), 39-51.

- [23] Isaacson, D. L., & Madsen, R. W. (1976). Markov chains theory and applications.
- [24] Kassambara, A., Kosinski, M., Biecek, P., & Fabian, S. (2017). Package ‘survminer’. Drawing Survival Curves using ‘ggplot2’(R package version 03 1).
- [25] Kleinbaum, D. G., & Klein, M. (2012). Survival analysis: a self-learning text (Vol. 3). New York: Springer.
- [26] Meyn, S., & Tweedie, R. L. (2009). Markov chains and stochastic stability (nd ed.).
- [27] Mo, J. (2010). Performance modeling of communication networks with Markov chains. *Synthesis Lectures on Data Management*, 3(1), 1-90.
- [28] Murtaugh, P. A., Burns, L. D., & Schuster, J. (1999). Predicting the retention of university students. *Research in higher education*, 40(3), 355-371.
- [29] Norris, J. R. (1997). *Markov Chains*, Cambridge Uni.
- [30] Ñaupa Ñaupa, D. M. (2017). Simulación de la dinámica de permanencia estudiantil en la Facultad de Ingeniería Estadística e Informática de la UNA-Puno, 2005–2020.
- [31] Ramírez, M. G., & Denazis, J. M. (2011). Dificultades de estudiantes de profesorado en matemática en la transición del nivel medio a la universidad. *Premisa*, 51, 23-36.
- [32] Requena Guerrero, F. (1999). Aportaciones a la mejora del algoritmo del test de Fisher en tablas $2 \times c$: Una extensión a tablas $r \times c$ (Doctoral dissertation, Universidad de Extremadura).
- [33] Resnick, S. I. (1992). *Adventures in Stochastic Processes* Birkhäuser. Boston et al.
- [34] Rincón, L. (2012). *Introducción a los procesos estocásticos*. UNAM, Facultad de Ciencias.
- [35] Rodríguez Rios, C. Y. (2012). Diseño de un modelo estocástico usando cadenas de Markov para pronosticar la deserción académica de estudiantes de Ingeniería caso: Escuela Colombiana de Ingeniería Julio Garavito.
- [36] Ross, S. M. (1996). *Stochastic Processes*. John Wiley & Sons. New York.

- [37] Spedicato, G. A., Kang, T. S., Yalamanchi, S. B., Yadav, D., & Cordón, I. (2016). The markovchain package: a package for easily handling Discrete Markov Chains in R. Accessed Dec.
- [38] Soukhov, I. M., & Kelbert, M. (2008). Probability and statistics by example: II Markov chains: A primer in random processes and their applications. Cambridge University Press.
- [39] Stevenson, M., & EpiCentre, I. V. A. B. S. (2009). An introduction to survival analysis. EpiCentre, IVABS, Massey University.
- [40] Stewart, W. J. (1994). Introduction of the numerical solution of Markov chains, Prince.
- [41] Stroock, D. W. (2013). An introduction to Markov processes (Vol. 230). Springer Science & Business Media.
- [42] Symeonaki, M., & Kalamatianou, A. (2011). Markov systems with fuzzy states for describing students' educational progress in Greek Universities. *Isi*, 1, 5956-5961.
- [43] Taylor, M. (1994). An Introduction to stochastic modeling. Michigan: Academic Press.
- [44] Tamir, A. (1998). Applications of Markov chains in chemical engineering. Elsevier.
- [45] Tijms, H. C. (2003). A first course in stochastic models. John Wiley and sons.
- [46] Tovar, R. J., & Salazar, J. C. (2009). Un modelo lineal mixto adaptado a una cadena de Markov con espacio de estados ordinal. Aplicación a datos sobre promedios académicos de estudiantes. *Revista Colombiana de Estadística*, 32(2), 213-230.
- [47] Verbeek, A., & Kroonenberg, P. M. (1985). A survey of algorithms for exact distributions of test statistics in $r \times c$ contingency tables with fixed margins. *Computational Statistics & Data Analysis*, 3, 159-185.
- [48] Wai-Ki, C., & Michael, K. N. (2006). Markov chains: Models, algorithms and applications.