

**UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO**

Facultad de Ciencias



**Análisis y Proyecciones de Becarios del
CONACYT**

TESIS

**Para obtener el título de Actuario
presentada por:**

Abigail Avilés Ruiz

Tutor

Dr. Leonardo Ignacio Martínez Sandoval

**Ciudad Universitaria, Ciudad de México,
Mayo 2024**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Índice general

1. Introducción	5
1.1. Idea general	5
1.2. Los métodos que utilizaremos	6
2. El CONACYT y las bases de datos de becarios	9
2.1. El CONACYT	9
2.2. Las bases de datos que utilizaremos	9
3. La descripción estadística	11
3.1. Número de becarios	11
3.2. Becarios por área de conocimiento, entidad y tipo de estudios . .	12
3.3. ¿Cuántos becarios hay por entidad?	13
3.4. Becarios por grado y entidad	14
3.5. Becarios por sexo	17
3.6. Becarios y habitantes	18
3.7. Becarios y habitantes por entidad	19
3.8. Transición de becarios entre grados	23
4. El modelo de Markov	27
4.1. Procesos Estocásticos	27
4.1.1. Cadenas de Markov	28
4.2. Máxima verosimilitud aplicada a cadenas de Markov	30
4.3. Aplicación a los datos	37
5. Análisis de los resultados y conclusiones	41

Capítulo 1

Introducción

1.1. Idea general

El principal tema de estudio de la presente tesis es el análisis de bases de datos y el manejo de la información obtenida. Esto nos da oportunidad de generar estadísticas, proyecciones, modelos o simplemente extraer datos específicos.

Para el desarrollo de este proyecto, usaremos información del CONACYT, la cual es publicada en su página web en bases de datos en Excel. Dichas bases de datos cuentan con información sin ningún tipo de análisis, y esta puede resultar complicada de entender para el público en general. También daremos una breve explicación de qué es el CONACYT y cómo funciona, para poder entender el porqué de nuestro análisis.

Tomaremos específicamente las bases de datos de Beneficiarios Nacionales del 2017 al 2019 [1]. Con dicha información haremos un análisis de estadística descriptiva para el estudio del número de becarios por año, por entidad, por grado de estudio y sexo, que nos ayudará a entender mejor esta información. Esto lo haremos mediante tablas dinámicas de Excel.

En esta parte del trabajo también haremos un breve análisis poblacional con bases de datos del INEGI [2], para conocer el número de becarios por entidad per cápita, así conoceremos la entidad con mayor aportación relativa de becarios.

Finalmente, otra de las incógnitas que nos gustaría analizar de manera descriptiva es la transición de becarios entre grados de estudio, es decir, ¿cuántos becarios pasan de un grado de estudio a otro?

Es importante conocer esta información, ya que la idea principal de esta tesis es hacer una proyección mediante modelos estocásticos, específicamente cadenas de Markov.

Antes de empezar con la parte matemática y gráfica de cadenas de Markov, daremos una breve explicación de qué son, para qué sirven, cómo las podemos utilizar, y cómo las vamos a utilizar específicamente para nuestro tema de estudio.

Para el estudio del modelo predictivo, usaremos Python. De esta manera po-

dremos realizar una proyección y hacer más visual nuestra tesis con el contenido gráfico de nuestro código.

Para finalizar, daremos conclusiones de lo analizado en las bases de datos y nuestro modelo, y a la par podremos dar sugerencias para mejorar la transición de becarios.

Haremos un conteo de las ventajas y desventajas de utilizar nuestro modelo, las limitaciones con las que nos encontramos a lo largo del proceso y finalmente, daremos nuestra conclusión general del proceso llevado a cabo.

1.2. Los métodos que utilizaremos

Primero haremos un análisis de **estadística descriptiva** con nuestras bases de datos. Empezaremos por lo básico, ¿cuántos becarios tenemos por año?, ¿cuántos becarios tenemos por entidad?, conocer la distribución de los becarios por grado de estudio, sexo, etcétera. Después analizaremos los cambios de estos becarios a través de los años, es decir, si un becario cambia de grado de estudio. Esto será la base para nuestro modelo predictivo.

Para este proyecto, consideramos que el mejor modelo a utilizar para predecir los cambios de un grado de estudios a otro, es el modelo de cadenas de Markov. Este es un proceso estocástico discreto en donde podemos calcular la probabilidad de que un objeto tome un estado dependiendo sólo del estado anterior. El estado que estudiamos es el grado de estudios del becario.

Es importante conocer más a detalle ¿qué son las cadenas de Markov?, ¿cómo surgieron?, y ¿para qué se utilizan? **Las cadenas de Markov** fueron introducidas por el matemático ruso Andrey Markov en 1905. Su intención principal era crear un modelo probabilístico para analizar la frecuencia con la que aparecen las vocales en poemas y textos literarios.

El éxito del modelo radica en que es lo suficientemente complejo como para describir ciertas características no tan simples 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.

Para estimar nuestras matrices de transición dentro de la metodología de Markov, utilizaremos el método de **máxima verosimilitud**, para justificar nuestras elecciones de parámetros. Pero antes de entrar en materia, ¿de qué se trata el método de máxima verosimilitud?

A grandes rasgos, el método de máxima verosimilitud nos permite ajustar un modelo a través de la estimación de sus parámetros bajo ciertas condiciones y restricciones, encontrando el valor que haga máxima la probabilidad de dicho sistema.

Esto nos llevará a utilizar el método de **multiplicadores de Lagrange** para resolver un sistema de ecuaciones generado de nuestras matrices de transición y de esta manera justificar la metodología que seguiremos para elegir las probabilidades de dichas matrices, las cuales más adelante explicaremos a detalle.

En los problemas de optimización, el método de los multiplicadores de Lagrange es un procedimiento para encontrar los máximos y mínimos de funciones de múltiples variables sujetas a restricciones. Este método reduce un problema de n variables y k restricciones a uno sin restricciones y con $n + k$ variables. Estas nuevas variables escalares desconocidas, una para cada restricción, son llamadas multiplicadores de Lagrange.

El método dice que los puntos donde la función tiene un extremo condicionado con k restricciones, están entre los puntos estacionarios de una nueva función sin restricciones construida como una combinación lineal de la función y las funciones implicadas en las restricciones, cuyos coeficientes son los multiplicadores.

La demostración usa derivadas parciales y la regla de la cadena para funciones de varias variables. Se trata de extraer una función implícita de las restricciones, y encontrar las condiciones para que las derivadas parciales con respecto a las variables independientes de la función sean iguales a cero.

Conforme vayamos avanzando en la tesis, todos estos conceptos serán explicados con mayor detalle, agregando algunas gráficas y ejemplos sencillos para entenderlos más fácilmente.

Capítulo 2

El CONACYT y las bases de datos de becarios

2.1. El CONACYT

El Consejo Nacional de Ciencia y Tecnología, mejor conocido por sus siglas en español CONACYT, es un organismo público con autonomía técnica, operativa y administrativa, cuyo objetivo principal es promover el desarrollo de la investigación científica, el desarrollo tecnológico y la innovación a fin de impulsar la modernización tecnológica del país.

El Consejo Nacional de Ciencia y Tecnología formula y financia programas de becas y, en general, brinda apoyo para la formación de investigadores y científicos en sus diversas modalidades, así como también integra la información de los programas de becas que ofrezcan otras instituciones nacionales, organismos internacionales y gobiernos extranjeros, a fin de optimizar los recursos en esta materia y establece esquemas de coordinación, en los términos de las convocatorias que para el efecto se emitan.

El CONACYT es financiado con fondos públicos, es por esto que toda su información es pública y de libre acceso. También por esta razón es importante que sea analizada y entendida.

2.2. Las bases de datos que utilizaremos

CONACYT publica sus bases de datos en formato Excel. Estas tienen información que para el público en general puede ser un poco complicada de entender, ya que dichas bases muestran la información sin ningún análisis. Es por esto que al hacer el análisis de dichas bases, esperamos que la información sea más comprensible. Inclusive al crear proyecciones, esperamos poder realizar algunas sugerencias de estrategia.

Para este proyecto, utilizaremos las bases de datos de Beneficiarios de CO-

10CAPÍTULO 2. EL CONACYT Y LAS BASES DE DATOS DE BECARIOS

NACYT con Becas Nacionales del 2017 al 2019 las cuales cuentan con los siguientes datos:

- Nombre: Nombre completo del becario.
- Fecha de inicio de beca
- Fecha de fin de beca
- Grado: Nos dice si el becario se encuentra estudiando una maestría, doctorado, especialidad o estancia técnica, siendo estos los únicos grados de estudio con los que se cuenta información.
- Institución: Institución donde el becario se encuentra estudiando.
- Entidad: Entidad federativa de la República donde el becario se encuentra estudiando.
- Área de conocimiento: Se refiere al ramo que se encuentra estudiando, medicina, matemáticas, ingenierías, entre otros.
- Programa: El nombre específico del programa que estudia de acuerdo al grado de estudios.
- Importe pagado: Este campo se divide en periodos trimestrales, pero para nuestro estudio, haremos un periodo total de un año.
- Sexo: Solo 2019 cuenta con este dato, el cual tendremos que agregar a los años anteriores mediante una asignación de género por nombre.

Las bases de datos utilizadas se encuentran disponibles en la página:

www.conacyt.gob.mx/index.php/becas-y-posgrados/padron-de-beneficiarios,

descargando los archivos con los nombres **Becas Nacionales de Enero a Diciembre 2019, 2018 y 2017** respectivamente.

Capítulo 3

La descripción estadística

A continuación presentaremos una descripción breve de todos los datos contenidos en las bases de datos a utilizar. De esta manera podremos utilizar dicha información para conocer con más detalle las bases y de esta manera podremos saber cómo utilizar esta información en nuestros modelos de proyección.

Al ser estadística descriptiva básica, podemos apoyarnos totalmente de las herramientas que nos proporciona Excel. Para los datos numéricos realizamos tablas dinámicas con los datos específicos que quisimos analizar.

Por ejemplo, para la primera tabla mostrada más adelante, solo necesitamos el Año y el número de becarios, los cuales contabilizamos por el campo de Nombre.

3.1. Número de becarios

Es importante conocer datos básicos como el número de becarios, y su distribución en los años que estamos analizando. De esta forma podemos ver de manera muy simple si los becarios van aumentando o no.

El total de becarios por año se puede ver en la siguiente tabla:

Año	Becarios
2017	72,108
2018	72,733
2019	72,905
Total	217,746

Como podemos ver, el número de becarios ha ido en aumento; de 2017 a 2018 hubo un crecimiento de 625 becarios, los cuales representan el 0.87% del total de becarios en 2017, y de 2018 a 2019 sólo hubo un aumento de 172 becarios, es decir, sólo 0.24% del total de becarios en 2018. Estos números son casi exactos, pero en la siguiente sección veremos que hay unas pocas excepciones.

3.2. Becarios por área de conocimiento, entidad y tipo de estudios

Para conocer más a detalle la información del total de becarios por año, a continuación haremos un análisis de la distribución de becarios por grado. El Conacyt contempla becas para los siguientes grados: Doctorado, Maestría, Especialidad y Estancia Técnica. En las tablas que tenemos también hay filas sin grado, o bien con una categoría llamada Convenio, la cual presenta estudiantes que se encuentran dentro de algún programa como PROGRAMA DE BECAS DE POSGRADO PARA ÍNDIGENAS (PROBEPI) y APOYO DE PROGRAMAS DUALES DE MAESTRÍA Y DOCTORADO, y programas varios. Es difícil saber a qué se refieren estas filas y además son únicamente casos excepcionales. Por esta razón, los incluiremos en la siguiente tabla, pero después de eso los omitiremos de nuestro estudio. El renglón de Total con grado ahora sí nos habla únicamente de becarios.

Grado	2017	2018	2019
Maestría	45,190	44,953	44,329
Doctorado	23,628	24,038	24,377
Especialidad	3,188	3,640	4,100
Estancia Técnica	99	97	97
Total con grado	72,105	72,728	72,903
NA/Convenio	3	5	2

Cuadro 3.1: Becarios por grado

Como podemos ver, Maestría es el grado con mayor número de becarios, el cual de 2017 a 2018 decreció 237 becarios, es decir, un 0.52% en relación a 2017, y decreció un 1.39% de 2018 a 2019, es decir, un total de 624 becarios.

En el punto anterior de número de becarios notamos que al pasar los años los becarios van en aumento. Pero en el grado de Maestría pasa lo contrario. Esto se compensa pues dicho crecimiento se ha presentado en los grados de Doctorado y Especialidad. Las Estancias Técnicas permanecen estables.

De 2017 a 2018 hubo un crecimiento en Doctorado de 410 becarios y en Especialidad de 452. De 2018 a 2019 hubo un crecimiento en doctorado de 339 becarios y en Especialidad de 460.

3.3. ¿Cuántos becarios hay por entidad?

También es importante conocer la distribución de becarios por entidad. De esta manera podemos conocer la entidad con mayor y menor número de becarios. Para dar más significado a estos números, podemos calcular también el número de becarios por cada cierta cantidad de habitantes. En la versión absoluta podemos a simple vista podemos decir cuáles son las tres entidades con mayor número de becarios, pero no arrojará las mismas entidades si lo hacemos en contraste con el número de habitantes por entidad.

Entidad	2017	2018	2019
AGUASCALIENTES	520	629	646
BAJA CALIFORNIA	2,904	2,941	2,929
BAJA CALIFORNIA SUR	628	666	653
CAMPECHE	160	197	196
CHIAPAS	819	856	917
CHIHUAHUA	1,407	1,471	1,602
CIUDAD DE MEXICO	22,780	22,795	21,629
COAHUILA	1,504	1,492	1,594
COLIMA	336	354	303
DURANGO	534	511	558
ESTADO DE MEXICO	4,270	4,286	4,298
GUANAJUATO	2,450	2,329	2,465
GUERRERO	695	875	965
HIDALGO	770	822	918
JALISCO	4,265	4,521	4,867
MICHOACAN	2,191	2,178	2,274
MORELOS	2,473	2,436	2,546
NAYARIT	312	287	303
NUEVO LEON	3,737	3,662	3,616
OAXACA	716	737	754
PUEBLA	3,488	3,468	3,544
QUERETARO	2,339	2,177	2,229
QUINTANA ROO	298	311	321
SAN LUIS POTOSI	2,013	1,907	1,840
SINALOA	1,470	1,476	1,545
SONORA	1,743	1,899	1,815
TABASCO	604	731	780
TAMAULIPAS	975	826	952
TLAXCALA	510	560	581
VERACRUZ	3,032	3,066	3,147
YUCATAN	1,611	1,750	1,553
ZACATECAS	551	512	563
Total general	72,105	72,728	72,903

Como podemos ver, la entidad con mayor número de becarios es Ciudad de México durante los 3 años, aunque podemos notar que hay una disminución de becarios entre el inicio y fin del periodo.

En 2017 hay 22,780 becarios en Ciudad de México, el cual representa un 31.59% del total nacional. En 2018 hay 22,795 becarios, es decir, aumentaron 15 becarios y representan el 31.34% del total de 2018. En 2019 hay 21,629 becarios. Es aquí donde podemos ver la disminución anteriormente mencionada, es decir, hay 1,166 becarios menos en relación a 2018, los cuales representan un 5.12% menos. De cualquier forma, Ciudad de México sigue dominando el total nacional con un porcentaje de 29.67%.

Sería interesante conocer si existe alguna explicación para dicha disminución. Por ejemplo, podría atribuirse a becarios que ahora participan en un programa de becas internacionales. Esta hipótesis puede explorarse con más datos. Necesitaríamos utilizar las bases de datos de Becarios al Extranjero de 2018 y 2019, para poder buscar alguna relación entre los Becarios Nacionales e Internacionales, y de ahí podremos concluir si la baja de becarios se debió a un cambio de Programa de Becas.

3.4. Becarios por grado y entidad

En esta sección estudiamos con más detalle la distribución de las becas a lo largo de los grados y del territorio nacional. La siguiente tabla muestra el número de becarios por Entidad y Grado en 2017.

Entidad	Doctorado	Maestría	Especialidad	Est Tec
AGS	199	318	3	
BCA	788	1,948	168	
BCS	249	379		
CAM	16	144		
CHS	243	549	27	
CHI	297	1,061	49	
CMX	8,943	13,022	716	99
COA	433	1029	42	
COL	133	179	24	
DUR	172	362		
MEX	1,439	2,509	322	
GTO	763	1,648	39	
GRO	67	559	69	
HGO	216	535	19	
JAL	1,293	2,585	387	
MIC	830	1,344	17	
MOR	881	1,562	30	
NAY	102	194	16	
NLO	976	2,477	284	

OAX	162	554		
PUE	1,183	2,284	21	
QUE	640	1,554	145	
QNR	72	226		
SLP	663	1,181	169	
SIN	472	857	141	
SON	647	1,082	14	
TAB	151	414	39	
TAM	195	722	58	
TLA	121	379	10	
VER	667	2,110	255	
YUC	465	1,041	105	
ZAC	150	382	19	
Total general	23,628	45,190	3,188	99

Para 2017 podemos ver que las entidades con mayor número de becarios como Ciudad de México, Jalisco, Estado de México y Puebla, muestran una distribución mayor en Maestría, seguida de Doctorado.

La siguiente tabla nos muestra el número de becarios por Entidad y Grado en 2018:

Entidad	Doctorado	Maestría	Especialidad	Est Tec
AGS	220	399	10	
BCA	773	1,981	187	
BCS	253	411	2	
CAM	24	173		
CHS	283	546	27	
CHI	369	1,045	57	
CMX	8,779	13,072	847	97
COA	452	1,003	37	
COL	170	175	9	
DGO	177	334		
MEX	1,558	2,421	307	
GTO	760	1,507	62	
GRO	78	697	100	
HGO	262	544	16	
JAL	1,352	2,738	431	
MIC	831	1,328	19	
MOR	845	1,544	47	
NLO	1,010	2,299	353	
OAX	182	551	4	
PUE	1,221	2,195	52	
QRO	608	1,415	154	
QNR	69	242		

SLP	627	1,091	189	
SIN	501	826	149	
SON	620	1,266	13	
TAB	192	499	40	
TAM	217	573	36	
TLA	125	426	9	
VER	694	2,113	259	
YUC	497	1,070	183	
ZAC	174	313	25	
Total general	24,038	44,953	3,640	97

En Ciudad de México aumentan 50 becarios en Maestría, pero disminuyen 164 becarios en doctorado, y para Especialidad aumentan 131 becarios, lo cual nos lleva a una de las preguntas de nuestra tesis, ¿este aumento se debe a una transición de becarios entre un grado de estudios a otro, o simplemente son becarios nuevos? Esto lo estudiaremos más adelante.

Finalmente, la siguiente tabla nos muestra el número de becarios por Entidad y Grado en 2019:

Entidad	Doctorado	Maestría	Especialidad	Est Tec
AGS	213	414	19	
BCA	759	1,972	198	
BCS	268	380	5	
CAM	26	170		
CHS	330	553	34	
CHI	418	1,136	48	
CMX	8,464	12,018	1,050	97
COA	487	1,073	34	
COL	144	159		
DGO	184	374		
MEX	1,618	2,383	297	
GTO	800	1,592	73	
GRO	86	766	113	
HGO	317	578	23	
JAL	1,362	2,968	537	
MIC	892	1,363	19	
MOR	855	1,646	45	
NAY	112	175	16	
NLO	1,033	2,146	437	
OAX	205	542	7	
PUE	1,341	2,134	69	
QRO	640	1,407	182	
QNR	77	244		
SLP	653	983	204	

SIN	509	881	155	
SON	626	1,177	12	
TAB	199	537	44	
TAM	227	662	63	
TLA	130	442	9	
VER	777	2,130	240	
YUC	469	943	141	
ZAC	156	381	26	
Total general	24,377	44,329	4,100	97

Para 2019 hay una disminución significativa de becarios en Ciudad de México, para Doctorado hay una disminución de 315 becarios, para Maestría disminuye en 1,054 becarios, pero para Especialidad aumentan 203 becarios, lo que en total nos da una disminución de 1,166.

Pero como vimos anteriormente, en total, el número de becarios aumentó entre un año y otro, entonces este resultado nos lleva otra incógnita. Si hay mayor número de becarios en relación al año anterior, y casi todos los grados de estudio presentan una disminución en Ciudad de México, ¿significa que hay transición de becarios entre entidades? ¿esta disminución se debe a un cambio de entidad o simplemente los nuevos becarios entran a otra entidad y no necesariamente es una transición? Esta pregunta no la estudiaremos con tanta profundidad, pero más adelante veremos cómo nuestros métodos nos permitirían decir algo al respecto.

3.5. Becarios por sexo

Como mencionamos anteriormente, sólo la base de datos de 2019 tiene el campo de Sexo, así que nos dimos a la tarea de agregar este campo en las bases de 2018 y 2017, por medio de una asignación de sexo por nombre.

Este análisis se hizo en varias etapas, primero buscamos los becarios en común entre 2019 y 2018, para que a esos becarios se les asignara el mismo sexo que tienen en la base de 2019.

Después buscamos una lista de nombres asignados con sexo, y buscamos dichos nombres entre los becarios que aún no tenían sexo asignado.

En algunos casos, los nombres principales no eran encontrados en la base de datos de nombres con sexo, así que buscamos el segundo nombre que también podía ayudarnos a asignar un sexo a ese becario en específico.

Revisamos que el resultado final de dicha asignación mostrara datos coherentes en cuanto a nombres y cifras comparadas con los datos mostrados en 2019, es decir, que el porcentaje de Femenino y Masculino luciera similar en 2018 y 2017, y las cifras no se dispararan de manera incongruente.

Dichos resultados se presentan en la siguiente tabla:

	Masculino	Femenino	Total
2019	38,076	34,827	72,903
2018	37,367	35,361	72,728
2017	36,214	35,891	72,105

3.6. Becarios y habitantes

En esta sección haremos un análisis de la relación proporcional entre becarios y habitantes en la República Mexicana, por lo que para esta sección utilizamos algunos datos obtenidos de la página oficial del INEGI.

Primero necesitamos tomar en cuenta nuevamente el número de becarios en 2017, 2018 y 2019. Recordemos que descartamos las filas con campo de grado inusual.

Año	Becarios
2017	72,105
2018	72,728
2019	72,903

Al iniciar con este análisis, se pudo observar que algunos becarios se encontraban duplicados, ya que estos contaban con una beca de algún grado de estudios a principio de año y al finalizar ese mismo año ya tenían una nueva beca en otro grado de estudios, por lo que al quitar esos duplicados, la tabla anterior resulta de la siguiente manera:

Año	Becarios
2017	71,795
2018	72,549
2019	71,210

Según el INEGI, la población de México en 2015 era de 119,938,473 habitantes. Debido a que la información del INEGI está reportada en bloques de 5 años y la información está actualizada hasta el año 2015, primero utilizamos el histórico de la información a partir de 1995 para poder realizar el cálculo de crecimiento hasta 2015.

Después calculamos el promedio de los resultados anteriores, y este se sumó a la cifra de 2015, el cuál daría como resultado un estimado poblacional para 2020.

Para calcular la población de los años entre 2015 y 2020, simplemente dividimos el promedio anterior entre 5 años y sumamos ese resultado a cada año. Usando esta extrapolación básica, nuestras predicciones se ven de la siguiente manera:

Para corroborar que la información obtenida del método anterior sea congruente con la información del INEGI, utilizamos una página que calcula el número aproximado de la población en los años 2017, 2018 y 2019.

Año	Población estimada
2017	123,033,579
2018	124,581,132
2019	126,128,685
2020	127,676,239

La página se llama Statista, y se puede encontrar en el siguiente enlace:

<https://es.statista.com/estadisticas/635250/poblacion-total-de-mexico-en-2020/>

De acuerdo con la página Statista, la proyección es la siguiente:

Año	Población estimada
2017	123,520,000
2018	124,740,000
2019	125,930,000
2020	127,090,000

Como podemos observar, el resultado de nuestra proyección luce similar a los resultados de la página Statista. Tomaremos los datos de nuestra proyección para continuar con nuestro estudio, y encontrar una relación entre el número de becarios y habitantes.

Originalmente, habíamos calculado el número de becarios per cápita, que se obtiene al dividir el total de becarios entre el total de habitantes. Sin embargo, esto dio números menores a uno con poco significado. Para obtener un indicador más entendible, dividimos el total de la población entre el número de becarios. Esto es más entendible.

El resultado nos queda de la siguiente forma:

- En 2017 hay aproximadamente 1 becario por cada 1,706 habitantes.
- En 2018 hay aproximadamente 1 becario por cada 1,713 habitantes.
- En 2019 hay aproximadamente 1 becario por cada 1,730 habitantes.

Notemos cómo es son más entendibles estas expresiones que decir que hay, por ejemplo, 0.00058 becarios per cápita.

3.7. Becarios y habitantes por entidad

Ahora bien, tenemos información previa del número de becarios por entidad, y pensamos que también nos sería de mucha ayuda saber el número de becarios relativo por entidad.

La información por entidad en el INEGI también se encuentra actualizada hasta 2015 y por bloques de 5 años. Así, hicimos un ejercicio análogo al de la sección anterior.

Primero sacamos el porcentaje de población de cada entidad, dividiendo la población de la entidad entre el la población total, después sacamos un promedio de estos resultados desde los años 1995 hasta 2015, y este promedio se aplicó por entidad en los años que estamos estudiando.

Las siguientes tablas son el resultado de un cruce entre la tabla de becarios por entidad y el número de habitantes por entidad.

En la columna de Población pusimos tal cual la predicción del número de habitantes de cada entidad. En la columna de Becarios relativos pusimos uno de cada cuántos habitantes es becario.

La siguiente tabla nos muestra el resultado de 2017:

Entidad	Población	Becarios per Capita
Aguascalientes	1,253,003	2,410
Baja California	3,327,953	1,146
Baja California Sur	614,365	978
Campeche	890,566	5,566
Chiapas	5,105,778	6,234
Chihuahua	3,815,312	2,712
Ciudad de México	10,315,043	453
Coahuila	2,961,135	1,969
Colima	691,364	2,058
Durango	1,804,941	3,380
Estado de México	16,613,179	3,891
Guajuato	5,908,252	2,412
Guerrero	3,769,965	5,424
Hidalgo	2,844,965	3,695
Jalisco	8,024,818	1,882
Michoacán	4,840,350	2,209
Morelos	1,943,697	786
Nayarit	1,160,387	3,719
Nuevo León	4,979,633	1,333
Oaxaca	4,227,423	5,904
Puebla	6,383,744	1,830
Querétaro	1,892,826	809
Quintana Roo	1,302,920	4,372
San Luis Potosí	2,868,544	1,425
Sinaloa	3,113,633	2,118
Sonora	2,855,803	1,638
Tabasco	2,403,466	3,979
Tamaulipas	3,552,624	3,644
Tlaxcala	1,256,338	2,463
Veracruz	8,520,766	2,810
Yucatán	2,133,936	1,325
Zacatecas	1,656,848	3,007
Total	123,033,579	1,706

Baja California Sur, Ciudad de México, Morelos y Puebla son las entidades con un número de becarios relativos.

La siguiente tabla muestra los resultados para 2018:

Entidad	Población	Becarios per Capita
Aguascalientes	1,268,764	2,017
Baja California	3,369,813	1,146
Baja California Sur	622,092	934
Campeche	901,768	4,578
Chiapas	5,170,000	6,040
Chihuahua	3,863,303	2,626
Ciudad de México	10,444,789	458
Coahuila	2,998,381	2,010
Colima	700,060	1,978
Durango	1,827,644	3,577
Estado de México	16,822,145	3,925
Guanajuato	5,982,567	2,568
Guerrero	3,817,385	4,363
Hidalgo	2,880,750	3,505
Jalisco	8,125,756	1,797
Michoacán	4,901,234	2,250
Morelos	1,968,145	808
Nayarit	1,174,983	4,094
Nuevo León	5,042,268	1,377
Oaxaca	4,280,597	5,808
Puebla	6,464,040	1,864
Querétaro	1,916,635	880
Quintana Roo	1,319,309	4,242
San Luis Potosí	2,904,626	1,523
Sinaloa	3,152,797	2,136
Sonora	2,891,724	1,523
Tabasco	2,433,698	3,329
Tamaulipas	3,597,310	4,355
Tlaxcala	1,272,141	2,272
Veracruz	8,627,942	2,814
Yucatán	2,160,778	1,235
Zacatecas	1,677,688	3,277
Total	124,581,132	1,713

Para 2018, Querétaro es una nueva entidad con mayor número de becarios relativos.

Finalmente, la siguiente tabla muestra el resultado para el año 2019:

Entidad	Población	Becarios per Capita
Aguascalientes	1,284,524	1,988

Baja California	3,411,673	1,165
Baja California Sur	629,820	965
Campeche	912,970	4,658
Chiapas	5,234,222	5,708
Chihuahua	3,911,293	2,442
Ciudad de México	10,574,535	489
Coahuila	3,035,627	1,904
Colima	708,756	2,339
Durango	1,850,347	3,316
Estado de México	17,031,110	3,963
Guanajuato	6,056,883	2,456
Guerrero	3,864,804	4,005
Hidalgo	2,916,535	3,177
Jalisco	8,226,695	1,690
Michoacán	4,962,117	2,182
Morelos	1,992,594	783
Nayarit	1,189,579	3,926
Nuevo León	5,104,903	1,412
Oaxaca	4,333,771	5,748
Puebla	6,544,337	1,847
Querétaro	1,940,444	871
Quintana Roo	1,335,697	4,161
San Luis Potosí	2,940,707	1,598
Sinaloa	3,191,962	2,066
Sonora	2,927,645	1,613
Tabasco	2,463,929	3,159
Tamaulipas	3,641,996	3,826
Tlaxcala	1,287,943	2,217
Veracruz	8,735,119	2,776
Yucatán	2,187,619	1,409
Zacatecas	1,698,529	3,017
Total	126,128,685	1,730

Podemos decir que para 2019, las entidades con mayor número de becarios relativos son las mismas que en 2018.

Concluimos que las entidades con mayor número de becarios relativos son:

- Ciudad de México con 1 becario por cada 466 habitantes en promedio.
- Morelos con 1 becario por cada 792 habitantes en promedio.
- Querétaro con 1 becario por cada 853 habitantes en promedio.
- Baja California Sur con 1 becario por cada 959 habitantes en promedio.

3.8. Transición de becarios entre grados

Para analizar la transición de becarios entre un año y otro, primero debemos tener en cuenta que existen programas multi-anales, por lo que para este análisis tomamos en cuenta que la duración de la beca es por periodos anuales. Es decir, para calcular la transición de becarios de 2017 a 2018, tomé en cuenta solo los programas con fin de beca de 2017, y lo mismo para la transición de 2018 a 2019.

Las siguientes tablas de transición nos muestran el número de becarios que cambiaron de un grado de estudios a otro o inclusive si se quedaron en el mismo grado de estudios. La manera en la que calculamos los números es por persona: cruzamos la tabla de un año con el siguiente mediante el nombre de becario, nos preguntamos en qué grado estaba y a qué grado pasó. Algunos becarios aparecen más de una vez en cada tabla, correspondiente a que tuvieron beca de más de un tipo. Consideraremos únicamente la primer aparición que tengan.

La tabla tiene un par de columnas especiales que se refieren a las personas fuera del sistema CONACYT: “Posibles ingresos” (PI) y la columna “Salieron de CONACYT” (S). En la columna “Posibles Ingresos”, estamos pensando en todas la personas que nunca han estado dentro del sistema CONACYT y que pudieran entrar. Supondremos que inicialmente son la población mexicana a 2017 menos los 71,208 becarios (sin repetición) que ya están dentro del CONACYT en 2017, es decir, 122,962,371 de acuerdo al resultado obtenido en la sección de becarios por población.

En la columna “Salieron” pondremos a todas las personas que tuvieron una beca y después salieron del sistema CONACYT, donde vamos a suponer que todas esas personas ya no regresarán en un futuro al sistema.

		2017					
		PI	D	M	E	ET	S
2018	PI	122,937,780	0	0	0	0	0
	D	5,329	17,169	1,439	3	0	0
	M	17,423	19	27,358	71	0	0
	E	1,777	3	6	1,851	0	0
	ET	62	0	1	0	34	0
	S	0	5,660	16,274	1,255	65	0
	Total	122,962,371	22,851	45,078	3,180	99	0

Dividiendo cada celda entre el total de su columna, obtenemos la siguiente tabla relativa, a cinco cifras significativas. Como en el caso de la columna “Salieron” esta operación no está bien definida, estamos usando la suposición de que los becarios que están ahí ya no regresan al sistema.

		2017					
		PI	D	M	E	ET	S
2018	PI	0.99980	0.00000	0.00000	0.00000	0.00000	0.00000
	D	0.00004	0.75135	0.03192	0.00094	0.00000	0.00000

M	0.00014	0.00083	0.60690	0.02233	0.00000	0.00000
E	0.00001	0.00013	0.00013	0.58208	0.00000	0.00000
ET	0.00000	0.00000	0.00002	0.00000	0.34343	0.00000
S	0.00000	0.24769	0.36102	0.39465	0.65657	1.00000
Total	1	1	1	1	1	1

Analizando estos valores, vemos una tendencia a que los becarios de Maestría, Doctorado y Especialidades permanezcan en el mismo grado. En el caso de Estancias Técnicas, lo más común es que al año siguiente los estudiantes salgan.

Podemos ver también que, sin contar la población inicial que se queda fuera, la entrada más grande de 2017 a 2018 fue el ingreso de nuevos becarios al grado de estudio de Maestría. De quienes estaban en maestría, el 60 % no cambió de grado de estudios de un año a otro.

La siguiente tabla nos muestra ahora la transición de becarios de un grado de estudios a otro del 2018 al 2019. Posteriormente mostramos la tabla relativa con los valores calculados de igual manera que el ejercicio anterior.

		2018					
		PI	D	M	E	ET	S
2019	PI	122,881,926	0	0	0	0	0
	D	20,611	3,730	13	3	0	0
	M	31,926	2	11,325	79	0	0
	E	3,220	0	1	785	0	0
	ET	97	0	0	0	0	0
	S	23,021	20,208	33,532	2,770	97	0
	Total	122,937,780	23,940	44,871	3,637	97	0

		2018					
		PI	D	M	E	ET	S
2019	PI	0.99935	0.00000	0.00000	0.00000	0.00000	0.00000
	D	0.00017	0.16323	0.00029	0.00094	0.00000	0.00000
	M	0.00026	0.00009	0.25123	0.02484	0.00000	0.00000
	E	0.00003	0.00000	0.00002	0.24686	0.00000	0.00000
	ET	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	S	0.00019	0.88434	0.74387	0.87107	0.97980	1.00000
	Total	1	1	1	1	1	1

Esta tabla muestra una tendencia muy distinta a la del año pasado. Tras una inspección, nos dimos cuenta que hubo cambios en cómo se nombran a los becarios de 2018 a 2019, de modo que en muchas situaciones parecería que hay becarios nuevos entrando, y muchos saliendo, cuando en realidad son los mismos becarios, pero siendo reordenado sus nombres y apellidos. Esto nos lleva a descartar esta última tabla y para nuestros análisis posteriores usaremos la tabla de 2017 a 2018, en donde este problema es menor. Usaremos esta tabla

para predecir lo que sucederá a futuro con nuestro modelo.

Capítulo 4

El modelo de Markov

4.1. Procesos Estocásticos

Para poder conocer a detalle el modelo de Markov y todo lo que implica, es importante conocer primero ¿qué es un modelo o proceso estocástico?

Un **proceso estocástico** es una colección de variables aleatorias y su evolución de un estado a otro con el paso del tiempo aleatoriamente, dichas variables pueden tener correlación entre ellas o ser independientes.

Existen diferentes tipos de procesos estocásticos, estacionarios y no estacionarios, discretos y continuos.

Un **proceso estacionario** es un proceso estocástico cuya distribución de probabilidad en un instante de tiempo fijo o una posición fija es la misma para todos los instantes de tiempo o posiciones. En consecuencia, parámetros tales como la media y la varianza, si existen, no varían a lo largo del tiempo o la posición. Un **proceso estocástico no estacionario** es aquel cuya distribución de probabilidad varía.

Un **proceso estocástico discreto** es aquel en donde la transición de un estado a otro tiene un tiempo determinado, es decir, nos dice en qué estado se encuentra en determinado tiempo y en qué estado se encuentra en el siguiente bloque de tiempo. Un **proceso estocástico continuo** es aquel en donde la transición de un estado a otro se continuamente dentro de un intervalo de tiempo continuo.

Podemos tener las siguientes combinaciones:

Estacionario	Discreto
Estacionario	Continuo
No estacionario	Discreto
No estacionario	Continuo

Un ejemplo de un proceso estocástico estacionario y discreto, es el proceso Bernoulli.

El proceso Bernoulli se trata de la observación de resultados al repetir varias veces un proceso, el cual se denomina como “ensayo Bernoulli”, el cual nos ayudará para calcular posteriormente la probabilidad de cada uno de dichos resultados.

El ejemplo más común para ilustrar el proceso Bernoulli, es el de lanzar varias veces una moneda, y posteriormente observar cuántas veces salió “águila” o “sol” en dicho experimento, para después calcular la probabilidad de cada resultado.

Es importante que el proceso Bernoulli cumpla que:

- La probabilidad de éxito permanece constante ensayo tras ensayo.
- Los ensayos deben de ser independientes entre sí.

Una definición más formal de un proceso estocástico es la siguiente:

Definición. Un proceso estocástico es una colección de variables aleatorias (X_t) $t \in T$ indexadas por un conjunto T y definidas en algún espacio de probabilidad (Ω, F, \mathbb{P}) .

Interpretamos al conjunto de índices T como un parámetro temporal; para nosotros T será $\{1, 2, 3, \dots, n\}$, algún intervalo $[0, t]$ ó $[0, \infty)$. Interpretamos a un proceso estocástico como la evolución en el tiempo de algún fenómeno cuya dinámica se rige por el azar.

4.1.1. Cadenas de Markov

Las cadenas de Markov, también conocidas como modelo de Markov o proceso de Markov, es un concepto desarrollado dentro de la teoría de la probabilidad y la estadística que establece una fuerte dependencia entre un evento y el suceso inmediato anterior. Su principal utilidad es el análisis del comportamiento de procesos estocásticos.

Las **cadenas de Markov** son propiamente un proceso estocástico discreto, y cuentan con diversas aplicaciones. Una cadena de Márkov es una secuencia $X_0, X_1, X_2, \dots, X_n$ de variables aleatorias.

El dominio de estas variables es llamado espacio estado; el valor de X_n es el estado del proceso en el tiempo n . Si la distribución de probabilidad condicional de X_{n+1} en estados pasados es una función de X_n por sí sola, entonces:

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_1 = x_1, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n).$$

La identidad mostrada es la **propiedad de Márkov**, la cual significa que el cambio de un estado a otro sólo depende del estado inmediatamente anterior, y no depende de los demás estados anteriores.

Existen dos tipos de cadenas de Markov, homogéneas y no homogéneas.

Una cadena de Markov se dice **homogénea** si la probabilidad de ir del estado i al estado j en un paso no depende del tiempo en el que se encuentra la cadena, esto es

$$\mathbb{P}(X_n = j | X_{n-1} = i) = \mathbb{P}(X_1 = j | X_0 = i).$$

Si para alguna pareja de estados y para algún tiempo n la propiedad antes mencionada no se cumple diremos que la cadena de Markov es **no homogénea**.

Matriz de Transición

Una **matriz de transición** se utiliza para describir la forma en la que un sistema cambia de un estado a otro en un periodo de tiempo.

Esta matriz se puede calcular de formas distintas, entre ellas están el método directo, calculando la matriz diagonal, o calculando analítica o gráficamente a partir de la ecuación de estado.

Una **matriz de transición de Markov** se define como una matriz cuadrada $k \times k$ de la siguiente forma:

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1k} \\ p_{21} & p_{22} & \dots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \dots & p_{kk} \end{pmatrix}$$

El elemento en la fila i , columna j , $p_{ij} = \mathbb{P}(X_n = s_i | X_{n-1} = s_j)$, representa la probabilidad de transición de un paso de s_j a s_i .

La matriz de transición cumple las siguientes propiedades:

- $p_{ij} \geq 0$
- $\sum_{i=1}^k p_{ij} = 1$

Ejemplo 1

Supongamos que tenemos una cadena de Markov, en donde el espacio de estados es el conjunto $S = \{0, 1, 2\}$, es decir, S solo tiene tres estados. Supongamos además, que la cadena comienza en el estado 1, es decir, $X_0 = 1$.

Pensemos que nuestra matriz de probabilidades de transición en un paso, está dada de la siguiente manera:

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & 2 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \end{matrix} & \begin{pmatrix} 0.2 & 0.5 & 0.4 \\ 0.3 & 0 & 0.3 \\ 0.5 & 0.5 & 0.3 \end{pmatrix} \end{matrix}$$

Supongamos que nuestra cadena se mueve de acuerdo a la Figura ??

Para calcular la probabilidad

$$\mathbb{P}(X_0 = 1, X_1 = 2, X_2 = 0)$$

haremos los siguientes pasos:

$$\begin{aligned} & \mathbb{P}(X_0 = 1, X_1 = 2, X_2 = 0) \\ &= \mathbb{P}(X_2 = 0 | X_1 = 2, X_0 = 1) \mathbb{P}(X_1 = 2, X_0 = 1) \\ &= \mathbb{P}(X_2 = 0 | X_1 = 2) \mathbb{P}(X_1 = 2, X_0 = 1) \\ &= \mathbb{P}(X_2 = 0 | X_1 = 2) \mathbb{P}(X_1 = 2 | X_0 = 1) \mathbb{P}(X_0 = 1) \\ &= P_{0,2} P_{2,1} \mathbb{P}(X_0 = 1) \end{aligned}$$

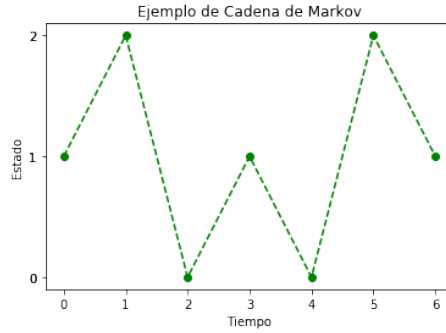


Figura 4.1: Ejemplo 1

Para calcular el siguiente paso, usaremos nuestra matriz de probabilidades de transición P , donde:

$$\begin{aligned} P_{0,2} &= 0.4 \\ P_{2,1} &= 0.5 \\ \mathbb{P}(X_0 = 1) &= 1 \end{aligned}$$

Por lo tanto, nuestro resultado es el siguiente:

$$P_{0,2}P_{2,1}\mathbb{P}(X_0 = 1) = (0.4)(0.5)(1) = 0.2.$$

El usar esta representación en forma matricial nos facilita el cómputo de las probabilidades de transición en más de un paso. En general $P \times P = P^2$ corresponde a las probabilidades de transición en dos pasos, y P^3, P^4, \dots, P^m corresponden a las probabilidades de transición en 3, 4, ..., m pasos respectivamente. De hecho, la matriz P^m se conoce como la **matriz de transición en m pasos** de la cadena de Markov.

4.2. Máxima verosimilitud aplicada a cadenas de Markov

Máxima Verosimilitud

En nuestro caso de estudio, modelaremos el cambio de estudios de becarios CONACYT mediante una cadena de Markov. Entonces decimos que dicho fenómeno se ajusta a una distribución de probabilidad, y queremos encontrar los parámetros que mejor se adaptan al modelo.

Para justificar nuestras elecciones de probabilidades de transición, usaremos el proceso de máxima verosimilitud, el cual nos ayuda a encontrar los parámetros más adecuados para una distribución de probabilidad. La idea general de la

4.2. MÁXIMA VEROSIMILITUD APLICADA A CADENAS DE MARKOV 31

técnica es encontrar los parámetros para maximizar la probabilidad de obtener los datos que ya tenemos.

Veamos un ejemplo sencillo de usar máxima verosimilitud en un proceso de Bernoulli.

Ejemplo 2

Tenemos una moneda que se distribuye Bernoulli con parámetro p , donde p es la probabilidad de que salga Sol y $(1-p)$ la probabilidad de que salga Águila. Queremos estimar p a partir de un experimento.

Digamos que lanzamos la moneda 5 veces y que nuestro experimento muestra los siguientes resultados:

Sol, Águila, Sol, Sol, Sol.

Este resultado tiene la siguiente probabilidad de ocurrir:

$$\mathbb{P}(\text{Experimento}) = (1-p) * p * p * p * p = p^4 * (1-p).$$

De acuerdo al proceso de máxima verosimilitud, eliremos a p como el parámetro que maximiza esta expresión.

Si graficamos dicha ecuación, vemos que el punto máximo lo alcanza cuando $p = 0.8$. Esto también se puede verificar analíticamente. Entonces podemos decir que ese es el valor del parámetro que maximiza nuestra ecuación y es lo que estamos buscando.

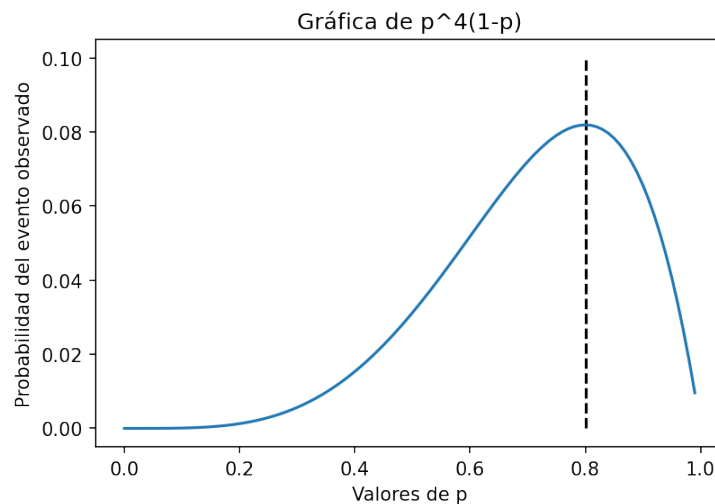


Figura 4.2: Grafica Ejemplo 2

Ejemplo 3

Ahora veamos un ejemplo de máxima verosimilitud para cadenas de Markov, aplicado a un proceso con dos estados A y B . Los parámetros a determinar son

las probabilidades de transición, donde nuestra matriz M se ve de la siguiente forma:

$$M = \begin{matrix} & \begin{matrix} A & B \end{matrix} \\ \begin{matrix} A \\ B \end{matrix} & \begin{pmatrix} p & q \\ 1-p & 1-q \end{pmatrix} \end{matrix}$$

Supongamos que cierto experimento muestra que se dieron los siguientes cambios de estado:

1. de A a A
2. de A a A
3. de A a B
4. de A a A
5. de B a B
6. de B a A
7. de B a B
8. de B a A
9. de A a A

Tomando en cuenta dichos resultados, la probabilidad de que esto ocurra es la siguiente:

$$\begin{aligned} \mathbb{P}(\text{Experimento}) &= p * p * (1-p) * p * (1-q) * q * (1-q) * q * p \\ &= p^4 * (1-p) * q^2 * (1-q)^2 \end{aligned}$$

Si realizamos la gráfica de la ecuación anterior, tenemos que el valor de los parámetros que la maximiza son cuando $p = 0.8$ y $q = 0.5$. Esto también puede verificarse analíticamente.

Entonces podemos decir que estos valores son adecuados para nuestros parámetros, ya que maximizan nuestra expresión, por lo tanto, la matriz queda de la siguiente manera:

$$M = \begin{matrix} & \begin{matrix} A & B \end{matrix} \\ \begin{matrix} A \\ B \end{matrix} & \begin{pmatrix} 0.8 & 0.5 \\ 0.2 & 0.5 \end{pmatrix} \end{matrix}$$

4.2. MÁXIMA VEROSIMILITUD APLICADA A CADENAS DE MARKOV33

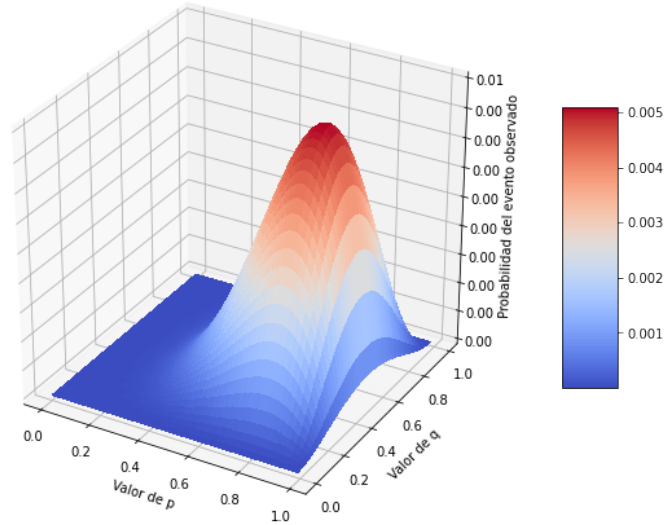


Figura 4.3: Grafica Ejemplo 3

Máxima verosimilitud para cadenas de Markov generales

Para nuestro modelo en particular, necesitaremos conocer la aplicación de máxima verosimilitud en cadenas de Markov en general.

Supongamos que tenemos n estados, por lo tanto tendremos que estimar una matriz de probabilidades M de tamaño $n \times n$, la cual se ve de la siguiente forma:

$$M = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2n} \\ p_{31} & p_{32} & p_{33} & \cdots & p_{3n} \\ \vdots & \vdots & \ddots & \vdots & \\ p_{n1} & p_{n2} & p_{n3} & \cdots & p_{nn} \end{pmatrix}$$

Aquí nuestra restricción principal es $\sum_{i=1}^n p_{ij} = 1$ para todo j en $\{1, 2, \dots, n\}$.

Esta es la matriz que queremos estimar a partir de un experimento realizado. Supondremos que en dicho experimento hubo n_{ij} transiciones independientes del estado i al estado j , así que para calcular la probabilidad de que esto ocurra, tendremos la siguiente formula.

$$\mathbb{P}(\text{Experimento}) = \prod_{i,j} p_{ij}^{n_{ij}}.$$

De acuerdo al método de Máxima Verosimilitud, para estimar los valores de nuestra matriz de transición debemos encontrar los p_{ij} que maximizan esta expresión. Para hacerlo más sencillo, necesitamos maximizar una simple suma, pero para esto necesitaremos sacar el logaritmo del producto anterior:

$$\begin{aligned} L(\mathbb{P}(\text{Experimento})) &= L\left(\prod_{i,j} p_{ij}^{n_{ij}}\right) \\ &= \sum_{i,j} n_{ij} L(p_{ij}) \end{aligned}$$

Multiplicadores de Lagrange

Los multiplicadores de Lagrange son una manera de optimizar una expresión con ciertas restricciones sobre los parámetros.

Para una definición más detallada de multiplicadores de Lagrange, el lector se puede guiar con la definición descrita en el libro de Cálculo de Spivak, ya que para fines prácticos dentro de este estudio, sólo necesitaremos conocer a grandes rasgos las restricciones descritas en la siguiente situación.

En la sección anterior llegamos a un problema de este estilo, pues queremos maximizar

$$L(P) := \sum_{i,j} n_{ij} L(p_{ij}),$$

sujeto a las n ecuaciones

$$\sum_{i=1}^n p_{ij} = 1,$$

para todo j en $\{1, 2, \dots, n\}$.

Esto significa que necesitamos n multiplicadores de Lagrange, $\lambda_1, \lambda_2, \dots, \lambda_n$.

De acuerdo al método de multiplicadores de Lagrange la nueva función objetivo es:

$$L(P) - \sum_{j=1}^n \lambda_j \left(\sum p_{ij} - 1 \right).$$

Derivando con respecto a λ_j nos da la restricción de la ecuación j . Si derivamos con respecto a p_{ij} e igualamos a cero, obtenemos lo siguiente:

$$0 = n_{ij}/p_{ij} - \lambda_j$$

$$\lambda_j = n_{ij}/p_{ij}$$

$$p_{ij} = n_{ij}/\lambda_j$$

4.2. MÁXIMA VEROSIMILITUD APLICADA A CADENAS DE MARKOV 35

Esto nos da condiciones sobre las p_{ij} que al sustituir en las restricciones iniciales se traducen a:

$$\sum_{i=1}^n n_{ij}/\lambda_j = 1$$
$$\sum_{i=1}^n n_{ij} = \lambda_j$$

De esta forma

$$p_{ij} = \frac{n_{ij}}{\sum_{i=1}^n n_{ij}}$$

Estos son los valores de probabilidad que estábamos buscando. Hagamos un ejemplo concreto.

Ejemplo 4

Tenemos los estados A , B y C , y nuestro experimento expulsa los siguientes resultados:

1. de A a A aparece 7 veces
2. de A a B aparece 2 veces
3. de A a C aparece 3 veces
4. de B a A aparece 2 veces
5. de B a B aparece 1 vez
6. de B a C aparece 3 veces
7. de C a A aparece 7 veces
8. de C a B aparece 0 veces
9. de C a C aparece 7 veces

Nuestra matriz propuesta es:

$$M = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

Entonces nuestra expresión para calcular la probabilidad del experimento es:

$$\mathbb{P}(\text{Experimento}) = p_{11}^7 * p_{21}^2 * p_{31}^3 * p_{12}^2 * p_{22}^1 * p_{32}^3 * p_{13}^7 * p_{23}^0 * p_{33}^7$$

¿Qué valores maximizan la expresión anterior?

Primero aplicamos logaritmo, por lo que la expresión nos queda de la siguiente forma:

$$\begin{aligned} L(\text{Experimento}) = & 7\text{Log}(p_{11}) + 2\text{Log}(p_{21}) + 3\text{Log}(p_{31}) + \\ & 2\text{Log}(p_{12}) + \text{Log}(p_{22}) + 3\text{Log}(p_{32}) + \\ & 7\text{Log}(p_{13}) + 7\text{Log}(p_{33}) \end{aligned}$$

Como utilizaremos el método de Multiplicadores de Lagrange, tenemos nuestras restricciones de la siguiente manera:

$$\begin{aligned} R1 &= p_{11} + p_{21} + p_{31} - 1 = 0 \\ R2 &= p_{12} + p_{22} + p_{32} - 1 = 0 \\ R3 &= p_{13} + p_{23} + p_{33} - 1 = 0. \end{aligned}$$

Por lo que nuestra expresión nos queda de la siguiente forma:

$$L - (\lambda_1 R1 + \lambda_2 R2 + \lambda_3 R3)$$

Al derivar la expresión anterior respecto a $p_{11}, p_{12}, \dots, p_{33}$ e igualando a cero, encontraremos los máximos que estamos buscando. Por ejemplo, derivando sobre p_{11} , tenemos el siguiente resultado:

$$0 = \frac{d(L(p)) - (\lambda_1 R1 + \lambda_2 R2 + \lambda_3 R3)}{dp_{11}} = \frac{n_{11}}{p_{11}} - \lambda_1$$

Como $n_{11} = 7$, entonces el máximo se alcanzará cuando $p_{11} = \frac{7}{\lambda_1}$

Posteriormente podemos realizar este paso análogamente para los demás términos de la expresión, lo que nos deja los siguientes resultados:

$$\begin{aligned} p_{21} &= \frac{2}{\lambda_1} \\ p_{31} &= \frac{3}{\lambda_1} \\ p_{12} &= \frac{2}{\lambda_2} \\ p_{22} &= \frac{1}{\lambda_2} \\ p_{32} &= \frac{3}{\lambda_2} \\ p_{13} &= \frac{7}{\lambda_3} \\ p_{23} &= \frac{0}{\lambda_3} \\ p_{33} &= \frac{7}{\lambda_3} \end{aligned}$$

Sabemos que $p_{11} + p_{21} + p_{31} = 1$, por lo tanto $\lambda_1 = 12$. De manera similar $\lambda_2 = 6$ y $\lambda_3 = 14$.

Por lo tanto, la matriz de transición que maximiza la probabilidad de los eventos observados es:

$$M = \begin{pmatrix} \frac{7}{12} & \frac{2}{6} & \frac{7}{14} \\ \frac{2}{12} & \frac{1}{6} & 0 \\ \frac{3}{12} & \frac{3}{6} & \frac{7}{14} \end{pmatrix}$$

4.3. Aplicación a los datos

Para esta sección, es importante tomar en cuenta todo lo que estudiamos anteriormente, ya que nuestra información recabada del CONACYT será clasificada de acuerdo a las transiciones de Markov.

Por ejemplo, nuestros estados serán cada uno de los grados de estudios, y de esta forma analizaremos la transición de un estado a otro según las probabilidades calculadas con el método de Lagrange.

De acuerdo con la información presentada en las tablas de transición, podemos decir que nuestro estado absorbente es cuando un becario ya no pasa a ningún otro grado de estudios y sale del programa CONACYT, es decir, nuestro estado “Salieron” es nuestro estado absorbente.

Algo que también tenemos que tomar en cuenta, son todos los becarios nuevos para los años 2018 y 2019, es decir, que en 2017 no están en ningún grado de estudios y en 2018 sí, y también los que en 2018 no están y en 2019 sí. A esos becarios los identificaremos como “Nuevos”, y este estado no es absorbente.

Tenemos que considerar que nuestro modelo no es del todo realista, ya que un becario en la vida real puede hacer cambios de estado en cualquier momento, e inclusive puede regresar de nuestro estado absorbente “Salieron” más adelante.

Tomaremos la tabla relativa de transición de grados de 2017 a 2018 del capítulo 3 y la transformaremos en nuestra matriz de probabilidades de transición de un grado a otro. Esto se justifica con el argumento que explicamos para estimar las probabilidades con máxima verosimilitud.

Usaremos esta matriz para dos cosas. Una de ellas será predecir el número de becarios que podría haber en cada grado de estudios en 2020. La otra será hacer una gráfica con la proyección de becarios a 5 y 20 años.

Comenzamos con la predicción de 2020. Tomaremos como base la tabla absoluta de transición de 2017 a 2018, la cual al convertirla a matriz la llamaremos M .

$$M = \begin{pmatrix} 122,937,780 & 0 & 0 & 0 & 0 & 0 \\ 5,329 & 17,169 & 1,439 & 3 & 0 & 0 \\ 17,423 & 19 & 27,358 & 71 & 0 & 0 \\ 1,777 & 3 & 6 & 1,851 & 0 & 0 \\ 62 & 0 & 1 & 0 & 34 & 0 \\ 0 & 5,660 & 16,274 & 1,255 & 65 & 0 \end{pmatrix}$$

A partir de ella, usamos el método de máxima verosimilitud para estimar la matriz de transición para una cadena de Markov. Para la última columna usamos que “Salieron” es estado absorbente. A esta matriz la llamaremos L .

Usaremos como población inicial de nuestro proceso de Markov a la distribución de becarios por grado en 2017 (recordemos que estamos quitando duplicados), es decir:

Población inicial: 122,962,371 Doctorado: 22,851 Maestría: 45,078 Especialidad: 3,180 Estancia técnica: 99

Lo que nos deja el vector denominado como u_0 de la siguiente forma:

$$u_0 = \begin{pmatrix} 122,962,371 \\ 22,851 \\ 45,078 \\ 3,180 \\ 99 \\ 0 \end{pmatrix}$$

Para realizar la predicción de un año al siguiente, multiplicamos por la matriz de transición, lo que nos da el siguiente resultado:

$$L * u_0 = \begin{pmatrix} 122,937,780 \\ 23,940 \\ 44,871 \\ 3,637 \\ 97 \\ 23,255 \end{pmatrix}$$

Para realizar la predicción para 2020, la multiplicaremos 3 veces, lo que nos da el siguiente resultado, redondeado al entero más cercano

$$L^3 * u_0 \approx \begin{pmatrix} 122,888,613 \\ 25,356 \\ 44,685 \\ 4,057 \\ 96 \\ 70,774 \end{pmatrix}$$

Ahora sí, veamos la tendencia a mediano y largo plazo.

A continuación realizamos una predicción a cinco años a partir de 2017 usando las potencias de nuestra tabla en porcentajes y la visualizamos de la siguiente manera:

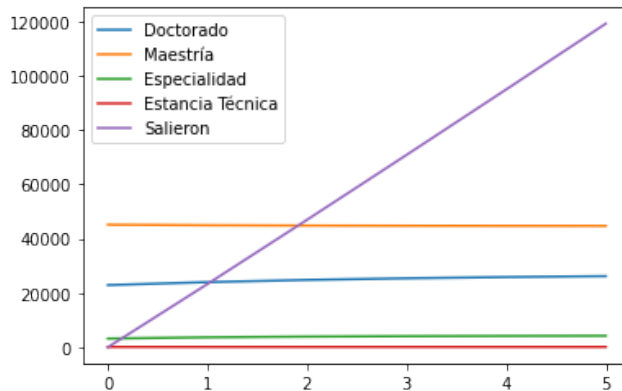


Figura 4.4: Proyección a 5 años

Como podemos ver en la Figura 4.4, la tendencia es que el número de becarios que salen del CONACYT vaya en aumento, lo cual se relaciona con la disminución de becarios en Maestría, dado que este grado de estudios es el que muestra mayor número de becarios en general. En este periodo de tiempo parece conservarse en general la cantidad de becarios para cada grado.

Para la siguiente proyección realizamos el ejercicio con 20 años. Se puede ver en la Figura 4.5. Aquí podemos deducir por el comportamiento de la gráfica que en algún momento el número de becarios de Maestría y Doctorado son muy parecidas, pues parecen acercarse demasiado.

Para poder visualizar mejor qué tanto se parecen las proyecciones de Maestría y Doctorado, quitamos el grupo "Salieron". El resultado nos queda en la Figura 4.6. Aquí vemos que si bien estos grados se acercan, más bien parece que la distribución de grados se estabiliza en el mediano plazo. Esto podría estudiarse más adelante.

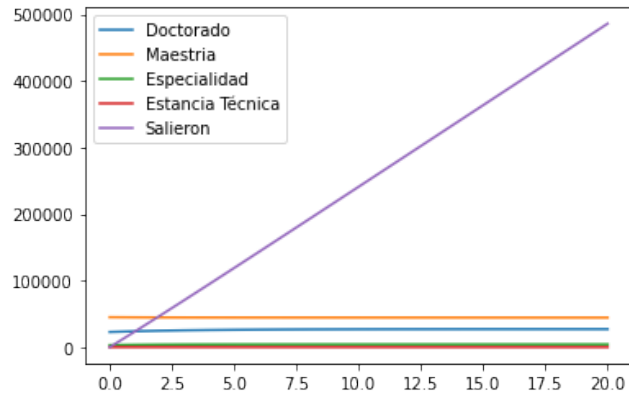


Figura 4.5: Proyección a 20 años

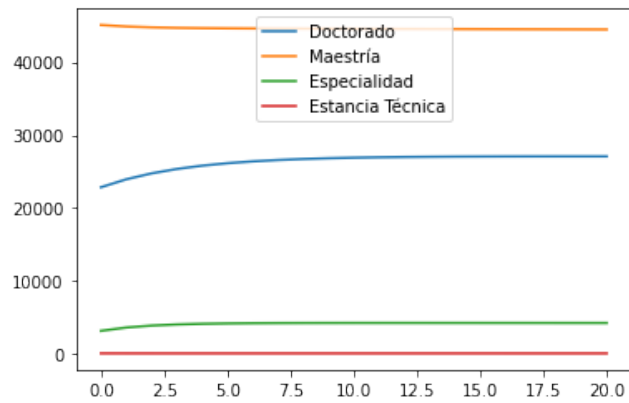


Figura 4.6: Proyección a 20 años sin grupo “Salieron”

Capítulo 5

Análisis de los resultados y conclusiones

En este capítulo analizaremos todos los resultados obtenidos a lo largo de esta tesis, donde podremos enlazar cada uno de ellos para poder utilizarlo dentro de nuestro modelo de Markov.

Para analizar nuestros resultados, dividiremos en tres secciones nuestra tesis.

En la primera sección nuestro estudio fue totalmente con **Estadística Descriptiva**, donde analizamos principalmente las bases de datos de becarios de todos los años, dando como resultado un total de 217,746 becarios y un promedio por año de 72,582 becarios.

Posteriormente se analizó el número de becarios por grado de estudios, donde pudimos observar que Maestría es la especialidad con mayor número de becarios, lo cual representa el 61 % del total de becarios de los tres años, seguido por Doctorado, representando el 33 % del total.

En un análisis de becarios por entidad, encontramos que Ciudad de México es la entidad con mayor concentración de becarios, representando un 30 % del total de becarios en los tres años.

También pudimos observar que la distribución de becarios por género se distribuye en mayor proporción al sexo Masculino con un 52 %, aunque no es una diferencia muy grande en la distribución del sexo Femenino.

Para nuestro estudio de Becarios per cápita analizamos primeramente las bases de datos del INEGI y realizamos una proyección para poder obtener la población total del año 2020, ya que las bases publicadas por el INEGI, están actualizadas hasta el 2015.

De acuerdo a nuestros resultados, el total de becarios de los tres años, sólo representa un 0.17 % de la población total estimada para 2020.

Para la segunda sección, analizamos todo lo referente a **Inferencia Estadística**, donde comenzamos a realizar nuestro Modelo mediante matrices de transición, analizando el movimiento de becarios de una grado de estudios a otro.

Por último, se debe considerar un punto muy importante **¿Qué podría salir mal en nuestro modelo?**

Si bien sabemos que este modelo solo nos da una estimación del comportamiento en los siguientes años de la transición de alumnos de un grado de estudios a otro, sabemos que dicha estimación no será exactamente igual a un escenario real, ya que existen muchas variables externas que no se están considerando dentro de este modelo, tales como la reciente Pandemia por COVID-19, factores socio económicos e inclusive el comportamiento de cada uno de los estudiantes, el cual es impredecible y no se puede incluir en nuestro modelo.

De esta forma se puede concluir que de acuerdo a nuestro modelo, el crecimiento de los becarios es constante, sin embargo no muestra un crecimiento exponencial en lo absoluto, ya que la proporción de becarios nuevos no representan ni un 5%.

Bibliografía

- [1] CONACYT. Padrón de Beneficiarios. <https://www.conacyt.gob.mx/Padrón-de-Beneficiarios.html>, 2021. Consultado el 22 de mayo de 2021.
- [2] INEGI. Población. <https://www.inegi.org.mx/temas/estructura>, 2020. Consultado el 22 de mayo de 2021.