



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

PREDICCIÓN DE CRIMEN USANDO MODELOS DE MARKOV OCULTOS

TESIS PARA OPTAR AL GRADO DE
MAGÍSTER EN GESTIÓN DE OPERACIONES

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL

PAZ OBRECHT IHL

PROFESOR GUÍA:
RICHARD WEBER HAAS

MIEMBROS DE LA COMISIÓN:
FERNANDO ORDOÑEZ PIZARRO
RAÚL MANASEVICH TOLOSA
ENRIQUE BASSALETTI RIESS

SANTIAGO DE CHILE
2014

PREDICCIÓN DE CRIMEN USANDO MODELOS DE MARKOV OCULTOS

La prevención del crimen ha ganado cada vez más espacio e importancia entre las políticas públicas en seguridad ciudadana, tanto en Chile como en el mundo. Durante la investigación realizada en este trabajo, se desarrolla un modelo para predecir los crímenes sobre una ciudad, que incluye el efecto de intervenciones preventivas y que permite además estudiar el fenómeno de desplazamiento que se le atribuye a este tipo de medidas. Ambos aspectos incluidos rara vez en los modelos de predicción revisados en la literatura. La estructura utilizada corresponde a un modelo de Markov oculto, donde el atractivo de un lugar para cometer un tipo específico de crimen se considera oculto y se estudia a través de el registro de crímenes observados en dicho lugar, considerando el efecto que intervenciones policiales podrían tener.

De manera de demostrar el tipo de información y uso que se puede hacer del modelo desarrollado, se aplicó éste en un caso de estudio. Los datos de los crímenes y vigilancia policial utilizados se obtuvieron mediante un simulador del crimen sobre una ciudad ficticia. El modelo estimado, permitió comparar el efecto de la vigilancia en el lugar donde es ubicada, así como en las áreas aledañas, según el atractivo de cada lugar. Encontrándose que las celdas más atractivas son más susceptibles a esta vigilancia, tanto en la reducción de crímenes esperados al posicionarse un vigilante en un lugar, como en el aumento de la tasa de crímenes cuando un policía es ubicado en lugares aledaños.

A partir de las matrices de transición se clasificaron las unidades de estudio, que componen la ciudad virtual, según su potencial para pasar a un estado de alta atractividad. Donde el grupo más numeroso corresponde al de celdas, de Bajo y Mediano Potencial, que permanecen en el mínimo estado de atractividad, reportando pocos crímenes en el lugar. Por el contrario, aquellas celdas, de Alto Potencial, que tienen probabilidades significativas de llegar y permanecer en estados de alta atractividad constituyen el grupo menos numeroso, y el que además suele concentrar los crímenes. Esto se alinea con lo que sugiere la literatura respecto a unos pocos lugares concentrado la mayoría de los crímenes.

Para validar el modelo se comparó su ajuste y predicciones con los obtenidos de otros cuatro modelos con diferentes especificaciones y estructuras (HMM Homogéneo, Clases Latentes, Regresión de Poisson y Persistencia), obteniendo mejores tasas de aciertos en la predicción de los crímenes futuros, de alrededor del 97%. Además el modelo destaca prediciendo los crímenes de las celdas de Alto Potencial, respecto a los modelos alternativos, alcanzando tasas de aciertos de 97% en comparación con las obtenidas por los otros cuatro modelos: 78%, 92%, 48% y 34% respectivamente. Se concluye además, en el experimento, que la inclusión del efecto de la policía permite capturar mejor el fenómeno delictivo, mejorando el desempeño al predecir el número de crímenes.

Finalmente, en relación a los objetivos planteados en este trabajo, se puede concluir que el modelo HMM desarrollado logra incorporar de forma efectiva los dos atributos que se deseaban estudiar en el fenómeno delictivo: considerar la atractividad de forma dinámica, actualizándose período a período, e incluir el efecto de la vigilancia en la predicción.

Dedicado a mi padres, por todo su esfuerzo y amor, porque sin ellos no sería lo que soy. Y a Felipe, que me ayudó siempre a poner las cosas en perspectiva, que me aguantó, me exigió y me contuvo cuando fue necesario. Por todos los gestos de cariño y confianza le dedico esta tesis a ustedes, familia.

Agradecimientos

Quiero agradecer a mi familia por su apoyo incondicional y cariño. Y a mis amigos por la linda e inolvidable etapa universitaria, la disfruté enormemente debido a ustedes: Loretto Gatica, Diego Vergara, Esteban Reyes, Florencia Sotomayor, Claudio Contreras, Pablo Reyes, Dante Pérez, María Elena Zegers, Edison Marrese y Mariana Villalobos. Mención especial para los compañeros de la salita 103, que me acompañaron durante todo del proyecto de tesis: Claudio Kuhlmann, Loretto Gatica, Rodrigo Arriagada, Nicolás Padilla, Ignacio Ríos y Víctor Bucarey. Muchas gracias a todos ustedes.

De igual forma quiero agradecer a Richard Weber por su apoyo y confianza, y las diferentes oportunidades de desarrollo que me brindó el haber trabajado con él como profesor guía.

Quiero agradecer también el apoyo otorgado por el Departamento de Postgrado y Postítulo de la Vicerrectoría de Asuntos Académicos de la Universidad de Chile (Beca de Estadías Cortas en el Extranjero) que me permitió trabajar durante algunos meses mi proyecto de tesis en Canadá. Por último mencionar, y agradecer, que esta tesis fue posible gracias al financiamiento de CONICYT (Beca de Magíster en Chile).

Tabla de Contenido

1. Introducción	7
1.1. Objetivos	9
1.1.1. Objetivo General	9
1.1.2. Objetivos Específicos	9
1.2. Metodología	9
1.3. Estructura de la Tesis	10
2. Criminología	11
2.1. Introducción	11
2.2. Teoría de Elecciones Racionales	12
2.3. Criminología Ambiental	13
2.4. Localización y Desplazamiento del Crimen	14
2.4.1. Generadores del Crimen	14
2.4.2. Atractores del Crimen	15
2.4.3. Desplazamiento del Crimen	15
2.5. Predicción del Crimen	16
3. Modelos de Markov Ocultos y Estadística Bayesiana	18
3.1. Introducción	18
3.2. Modelo de Markov Oculto	19
3.2.1. Sistemas de Markov	20
3.2.2. Objetivo de un HMM	21
3.2.3. Elementos de un HMM	22
3.2.4. Los Tres Problemas de un HMM	23
3.3. Estadística Bayesiana	25

3.3.1.	Teorema de Bayes	25
3.3.2.	Distribución a Posteriori y Conjugados	26
3.4.	Metodos de Simulación de Monte Carlo con Cadenas de Markov	27
4.	Modelo de Markov Oculto para la Predicción del Crimen	29
4.1.	Introducción	29
4.2.	Atractividad en el Crimen y Modelo de Markov Oculto	30
4.2.1.	El Efecto de Intervenciones en el Corto y Mediano Plazo	31
4.2.2.	Heterogeneidad en el Modelo	34
4.3.	Modelo de Markov Oculto para el Crimen	34
4.3.1.	Divisiones Geográficas y Discretización temporal	34
4.3.2.	Elementos del Modelo	35
4.4.	Los Tres Problemas del Modelo HMM para el Crimen	39
4.4.1.	Problema 1: Definir una Medida de Ajuste	39
4.4.2.	Problema 2: Maximizar la Medida de Ajuste	40
4.4.3.	Problema 3: Predecir a Partir del Modelo	42
5.	Implementación	44
5.1.	Introducción	44
5.2.	Descripción de la Ciudad Virtual	45
5.2.1.	Características Estructurales	45
5.2.2.	Descripción de los Agentes	46
5.3.	Datos Obtenidos del Simulador	47
5.4.	Transformación de Datos	49
5.4.1.	Discretización del Espacio Físico y Temporal	49
5.4.2.	Generación de las Secuencias de Datos	51
5.5.	Estimación Modelo	54
5.5.1.	Detalles de la Estimación	54
5.5.2.	Elección del Número de Estados	54
6.	Resultados	56
6.1.	Introducción	56
6.2.	Resultados Específicos	56

6.2.1.	Predicción a Partir del Modelo Estimado	61
6.2.2.	Validación del Modelo	62
6.3.	Resultados Generales del Modelo	65
7.	Conclusiones y Trabajo Futuro	66
7.1.	Conclusiones	66
7.2.	Trabajo Futuro	68
	Bibliografía	69
A.	Apéndice A	I
A.1.	Algoritmo de Escalamiento de la Función de Verosimilitud	I
A.2.	Priors Conjugados de un Modelo de Regresión Multivariada	II
A.3.	Sampleos MCMC	III
A.3.1.	Sampleo de Gibbs	III
A.3.2.	Algoritmo de Metropolis	IV
B.	Apéndice B	VI
B.1.	Algoritmo de Estimación Jerárquico Bayesiano	VI
B.2.	Estadísticos para la Selección del Número de Estados	VII
B.2.1.	Densidad Marginal y Factor de Bayes	VII
B.2.2.	Criterio de la Desviación de la Información (DIC)	VIII
B.2.3.	Criterio de Selección de Modelos Markovianos (MSC)	VIII
C.	Apéndice C	X
C.1.	Transformación de datos	X
C.1.1.	Condiciones Ambientales con Efecto en el Corto Plazo	X
C.1.2.	Condiciones Ambientales con Efecto en el Largo Plazo	XII
C.2.	Resultados de la Estimación de Parámetros	XIII
C.2.1.	Parámetros de Efecto Fijo	XIII
C.2.2.	Parámetros de Efecto Aleatorio	XIV
C.2.3.	Parámetros Capturando la Heterogeneidad	XVI
C.3.	Modelos de Comparación para la Validación	XVII
C.3.1.	HMM sin Heterogeneidad	XVII

C.3.2. Clases Latentes	XVII
C.3.3. Regresión de Poisson	XVIII
C.3.4. Modelo de Persistencia	XVIII
C.4. RPS: Ranked Probability Score	XIX

Índice de figuras

3.1. Diagrama de un modelo de Markov oculto	20
3.2. Diagrama de un sistema de Markov	20
3.3. Experimento de lanzamiento de monedas	21
4.1. Modelo HMM para el crimen	31
4.2. Relación entre atractividad y condiciones ambientales	32
4.3. Proceso de actualización de la atractividad	33
5.1. Uso de suelo de la ciudad.	45
5.2. Agentes circulando en días hábiles y fines de semana.	47
5.3. Distribución del total de robos sobre la ciudad.	48
5.4. División de la ciudad en una grilla de 19x15.	50
5.5. Selección de celdas relevantes para el análisis.	51
6.1. Tasa de crímenes bajo vigilancia directa	57
6.2. Tasa de crímenes bajo vigilancia indirecta	58
6.3. Clasificación de celdas sobre el mapa de la ciudad.	60
C.1. Vigilancia directa sobre una celda.	XI
C.2. Vigilancia indirecta sobre una celda.	XII
C.3. Histograma valor promedio de la probabilidad de no observar delitos α	XV
C.4. Histogramas valor promedio de umbrales en el estado 1.	XV
C.5. Histogramas valor promedio de umbrales en el estado 2.	XVI
C.6. Histogramas valor promedio de umbrales en el estado 3.	XVI

Índice de tablas

5.1. Composición de categorías de uso de suelo.	46
5.2. Registros obtenidos en la simulación.	48
5.3. Elección del número de estados.	55
6.1. Clasificación de celdas en tres categorías.	59
6.2. Matriz de transición promedio entre las celdas de cada categoría.	59
6.3. Matriz de transición promedio según clasificación de uso de suelo.	60
6.4. Capacidad de predicción del modelo HMM para el crimen	62
6.5. Ajuste y capacidad de predicción de los modelos en la etapa de validación. . . .	64
C.1. Estimaciones de los parámetros de efecto fijo relacionados a las probabilidades de emisión	XIII
C.2. Estimaciones de los parámetros de efecto fijo relacionados a la matriz de transición	XIV
C.3. Parámetros relacionados a las características físicas que capturan la heterogeneidad entre las celdas.	XVII

Capítulo 1

Introducción

La prevención del crimen ha ganado cada vez más espacio e importancia en el debate público, siendo una de las tareas primordiales dentro de las políticas públicas en seguridad ciudadana, tanto en Chile como en el mundo [12]. La elaboración y aplicación de estas políticas públicas, así como estrategias e intervenciones orientadas a la prevención del crimen, requiere de conocimiento y comprensión específica del fenómeno delictual. La criminología y otros estudios académicos afines, han tratado históricamente de llenar este vacío estudiando el fenómeno y mecanismos de prevención, en un principio abordando la problemática desde la persona y sus motivaciones al delinquir [22], la prevención social, y durante las últimas décadas desde el dónde y por qué sucede [6, 10, 39, 45], la prevención situacional.

Este trabajo abordará la prevención y estudio del crimen a partir de la prevención situacional, donde las características del lugar y las circunstancias que rodean la ocurrencia de un crimen son utilizadas para explicar los patrones observados. La vigilancia existente, las ventajas y obstáculos físicos del lugar, el número de potenciales víctimas y sus hábitos juegan un rol fundamental a la hora de ofrecer oportunidades a potenciales infractores de la ley. A través de este enfoque es posible obtener conocimiento sobre lo que hay que evitar y lo que hay que potenciar en cada lugar para hacer del crimen una actividad menos atractiva. Siendo los recursos policiales limitados e insuficientes para intervenir toda la superficie de la ciudad, tomar buenas decisiones es crucial para una buena gestión de estos recursos.

En particular, una estrategia eficiente requiere información relacionada al lugar donde ocurrirán los crímenes con mayor probabilidad, identificar los patrones y causales que definen la ocurrencia del crimen y conocer el impacto que las intervenciones tendrán sobre la ciudad. El identificar los lugares con mayor propensión a registrar crímenes y generar una predicción de estos eventos, ha sido abordado desde numerosos y variados enfoques. Los modelos más simples utilizan el historial de crímenes para identificar las zonas donde más se concentran los eventos delictuales en el futuro [16, 25], las áreas identificadas son escogidas luego, para implementar algún tipo de intervención [27, 29, 39]. Una desventaja, de basar la predicción solamente en el historial de crímenes pasados, es que se debe asumir que las zonas que han registrado las mayores concentraciones de crímenes serán las mismas que registrarán la mayoría de los crímenes en el futuro. La experiencia de los delincuentes, las rutinas dentro de la ciudad y su estructura son dinámicas, sin embargo, cambiando estos puntos de concentración según cambia la ciudad y la rutina de sus habitantes. Esto implica que las predicciones generadas se ajustan solamente a un horizonte de corto, en un período tal que la ciudad no haya registrado cambios en sus rutinas.

Algunos investigadores han logrado incrementar el ajuste de las predicciones del crimen incluyendo las características de los lugares y explicitando el proceso de elección de los delincuentes y sus preferencias [11, 21, 34, 46]. Varios de estos modelos intentan modelar las preferencias de los delincuentes a través de la estimación del atractivo de los lugares en una ciudad, o su potencial para generar oportunidades para delinquir [18, 21, 46]. El atractivo de un lugar, o la utilidad esperada que le reporta el mismo, se relaciona a los costos y beneficios esperados al cometer ahí un crimen. Esta valoración incluye todos los riesgos al cometer un delito, y por tanto considera: la presencia de vigilancia, vías de escape, características físicas y otras determinantes ambientales. Los delincuentes, siendo individuos racionales, un supuesto comúnmente utilizado en este tipo de modelos, intentarán maximizar su beneficio, buscando aquellos lugares que les reporten una mayor utilidad esperada.

El mayor obstáculo al enfrentar el crimen urbano con un modelo basado en las preferencias de los delincuentes, es la estimación de los parámetros que le dan sentido al mismo. La utilidad esperada se ve afectada por múltiples variables, factores y rutinas de la ciudad. Aplicar el modelo en una ciudad real implicaría la estimación de la utilidad esperada de cada lugar de la ciudad, lo que involucra un enorme número de variables e información. Los modelos que intentan estimar el atractivo, o riesgo, de los lugares tienden a ser exhaustivos en las variables explicativas, requiriendo una enorme cantidad de información, no eliminando la posibilidad en el proceso, de dejar alguna variable relevante fuera del estudio.

Por otro lado, las intervenciones en un lugar, destinadas a la prevención del delito o que afecte las rutinas y comportamiento de los habitantes de la ciudad, tienen impacto sobre el atractivo percibido para cometer un crimen, por lo que parece importante el incluirlas en su estimación. Al influir sobre los riesgos y utilidades esperadas que ofrece un lugar, algunos delincuentes podrían decidir moverse a otro de mejor prospecto, en un fenómeno conocido como desplazamiento [1, 5, 7, 36]. Pareciera importante considerar este efecto cuando se intenta predecir el crimen, y más aún, cuando se debe escoger entre diversas estrategias de vigilancia. El conocer cómo cambiará la dinámica de la ciudad, al intervenirla, permitiría anticiparse al posible desplazamiento, pudiendo otorgar mayor precisión en la predicción de la ubicación y cantidad de eventos criminales en los siguientes períodos.

Existe pequeña evidencia en la bibliografía revisada, de modelos de predicción del crimen que intenten estimar el efecto de la policía, u otras intervenciones, en el atractivo de un lugar. Lazzati *et. al* [27] desarrolla un modelo basado en agentes donde se incorpora la policía y su efecto en el crimen, pero el modelo requiere la estimación previa del atractivo de los lugares y la cuantificación del impacto de la vigilancia, información que utiliza como input.

La idea principal de este trabajo es generar una metodología para la predicción del crimen que incorpore el efecto de la policía en el atractivo percibido en un lugar y reducir la cantidad de información necesaria para poder aplicarla. En particular, se desea explotar la relación de dependencia entre el número de delitos registrados en un lugar y el atractivo del mismo. En el modelo desarrollado se intentará utilizar el registro de delitos, información que se encuentra al alcance de la policía, a través de las denuncias de estos crímenes, para inferir la utilidad esperada de los lugares.

Se espera que la inclusión de intervenciones en el lugar permita predicciones más ajustadas y que faciliten la toma de decisiones de la policía, u otras instituciones, al proveer información sobre su efecto, cuales intervenciones son mejores que otras y en qué casos.

1.1. Objetivos

1.1.1. Objetivo General

El objetivo principal de este trabajo de tesis es proporcionar predicciones para un tipo de crimen específico, así como estimaciones del efecto de diferentes intervenciones policiales, o de otro tipo, en la percepción de delincuentes y la tasa esperada de crímenes.

1.1.2. Objetivos Específicos

- Modelar el proceso de aprendizaje y actualización del atractivo percibido en un lugar, incluyendo el efecto de intervenciones orientadas al lugar.
- Desarrollar un modelo que relacione el atractivo percibido en un lugar con el número de crímenes registrados en el mismo, incluyendo el efecto de intervenciones orientadas al lugar.
- Generar una metodología para predecir crímenes en el futuro a partir del modelo, dada alguna estrategia de prevención o intervención orientada al lugar.

1.2. Metodología

De forma de lograr los objetivos detallados anteriormente, se desarrolló el trabajo siguiendo la metodología, y sus etapas, que se describe a continuación.

- **Revisión Bibliográfica Criminología:** se realiza una revisión de teorías y tendencias en criminología, para ser utilizadas como base de los supuestos que se utilizaran para estructurar los modelos. También se realiza una revisión de otros trabajos en la literatura que han tenido como objetivo generar predicciones del crimen en de una ciudad.
- **Revisión Bibliográfica Modelo de Markov Oculto :** se revisa la estructura y teoría que definen los Modelos de Markov Ocultos, así como los algoritmos utilizados para su estimación.
- **Definición del concepto atraktividad y su proceso de actualización :** se define el concepto atraktividad, relacionado a la utilidad esperada de cometer un delito en un determinado lugar, y se describe su proceso de aprendizaje y actualización, tomando en consideración el efecto de intervenciones para la prevención del delito.
- **Desarrollo de un Modelo de Markov Oculto :** se desarrolla un modelo de Markov oculto que relaciona la atraktividad de un lugar, que se considera no observable, con los delitos observados en el mismo, incluyendo el efecto de intervenciones para la prevención del delito. Este modelo permite además generar predicciones para el crimen estudiado, a partir de una estrategia de prevención.
- **Aplicación de la metodología en un caso de estudio:** se aplica la metodología y modelos desarrollados sobre los datos simulados en una ciudad virtual en el simulador del crimen CRIMSIM. Se valida el modelo comparándolo con otras alternativas y metodologías.

1.3. Estructura de la Tesis

La estructura utilizada en este documento para exponer el trabajo realizado es la siguiente:

- **Capítulo 1. Introducción:** Corresponde a la descripción del tema, la motivación de éste y los alcances y objetivos del trabajo realizado.
- **Capítulo 2. Criminología:** Corresponde a la revisión de las teorías de elecciones racionales, criminología ambiental y prevención situacional. Se incluye también una revisión de algunos trabajos que han intentado estimar la atractividad de un lugar y predecir el número de crímenes a partir de estos.
- **Capítulo 3. Modelos de Markov Ocultos y Estadística Bayesiana:** Se explica y enuncia las principales características y estructura de un Modelo de Markov Oculto , así como algunos algoritmos comúnmente utilizados para su estimación.
- **Capítulo 4. Modelo de Markov Oculto para la Predicción del Crimen:** Se presenta y justifica un Modelo de Markov Oculto para la estimación de la atractividad y los crímenes observados en una ciudad, incluyendo el efecto de intervenciones destinadas a la prevención del crimen, y se propone un proceso de estimación mediante simulaciones MCMC.
- **Capítulo 5. Implementación:** Se presenta la implementación del modelo de Markov oculto desarrollado en un caso de estudio, utilizando datos obtenidos del simulador del crimen CRIMSIM en una ciudad virtual, de manera de ejemplificar las posibilidades del modelo.
- **Capítulo 6. Resultados:** Se incluyen los resultados y análisis realizados sobre las estimaciones obtenidas en la implementación en el caso de estudio, así como resultados y conclusiones generales del modelo de Markov Oculto para el crimen.
- **Capítulo 7. Conclusiones y Trabajo Futuro:** Se discuten los resultados del modelo, sus ventajas y debilidades, y se compara con otros modelos de estimación, para finalmente, presentar las conclusiones.

Capítulo 2

Criminología

2.1. Introducción

El estudio de los fenómenos delictivos, área de investigación conocida hoy en día como criminología, ha captado la atención de diferentes especialistas y se ha abordado desde una gran variedad de disciplinas. Inicialmente se enfoca en comprender las causales del crimen, clasificar los diferentes tipos de delincuentes y modelar el proceso delictivo [22]. Sin embargo, la utilización de una perspectiva económica para analizar el delito, al asumir racionalidad en los individuos, le dio un nuevo impulso a esta área de investigación integrando nuevas disciplinas y permitiendo la generación de modelos de corte más cuantitativo [6, 10, 39, 45]. Junto con la criminología ambiental y la prevención situacional del delito, otras dos corrientes dentro de la criminología, más que estudiar al delincuente y el origen de su comportamiento criminal, se enfocan en el evento criminal y las condiciones ambientales que lo favorecen o entorpecen. De alguna manera, se traslada el interés desde la persona que comete el delito, al lugar y el momento donde este ocurre.

Adicionalmente, durante las últimas décadas se ha observado un incremento sustancial en la capacidad de almacenamiento y procesamiento de datos. Lo anterior ha permitido abordar la problemática delictual a través de la estadística y geo-referenciación [16, 25]. La identificación de patrones en la distribución del crimen en las ciudades ha permitido generar modelos de predicción del crimen, así como generar conocimiento respecto al proceso de elección que realizan los delincuentes para escoger el lugar adecuado para delinquir. Esto se ha traducido en el desarrollo de herramientas de soporte para la policía, u otras instituciones, para la toma de decisiones destinadas a la prevención del delito.

A lo largo de este capítulo se describirán algunas de las corrientes teóricas de la criminología que actúan como base de los modelos y supuestos que serán expuestos más adelante. Además se realizará un revisión de algunos patrones identificados en la distribución de delitos y la forma en que esta evoluciona en el tiempo, para terminar con una exposición respecto al trabajo y modelos de otros autores en la estimación y predicción del crimen.

2.2. Teoría de Elecciones Racionales

La teoría de elecciones racionales asume una evaluación económica por parte de los individuos al enfrentarse a una situación en la que podrían o no cometer un crimen [3]. La relación entre los costos y beneficios de dicha acción son los que determinan su ocurrencia. Según esta teoría, una persona decidirá cometer un delito cuando la utilidad esperada sea mayor que otra actividad que pudiera realizar en cambio [3]. Entre los costos que percibe un sujeto al cometer un delito se encuentra el castigo esperado, así como el perjuicio en su imagen social, las dificultades para encontrar un trabajo legítimo, acceder a una mejor remuneración, obtención de beneficios sociales o cualquier otra desventaja o daño que podría percibir si es atrapado [8]. Los beneficios tienen relación con la ganancia económica que le reporta el crimen, los beneficios sociales en la pandilla a la que pertenezca, respeto por parte de otros criminales, satisfacción personal o cualquier otro factor favorable relacionado a cometer el crimen (ya sea físico, social o psicológico).

Cuando se habla de beneficio y costos esperados, refiere a la existencia de probabilidades involucradas en la ocurrencia de algunos eventos. En este caso particular, la probabilidad asociada corresponde a la de ser sorprendido, capturado y condenado por el delito que se está cometiendo.

Entre los factores que la policía y las instituciones judiciales de un determinado país pueden afectar en la función de utilidad de un delincuente, se encuentran los costos y las probabilidades de aprehensión. Aumentar las penas asociadas a los delitos es una de las maneras de afectar los costos asociados a un delito. Por otro lado aumentar los recursos policiales, aplicar mejores estrategias de vigilancia, intervenir la ciudad para hacer más difícil el acto de delinquir sin ser sorprendido, modificar el proceso de condena por un determinado delito son algunas de las maneras de afectar las probabilidades involucradas en la ecuación económica. Aumentar los costos de oportunidad del delito también es una manera de afectar su ocurrencia, es decir las probabilidades de empleo legítimo, aumento en los salarios esperados, u otros.

Cabe destacar que estos costos, beneficios y probabilidades son los percibidos por las personas dispuestas a delinquir. No basta con aumentar los años de cárcel para un determinado crimen o proponer pena de máxima para todos los delitos, el delincuente debe percibir que dichas penas o las probabilidades de aprehensión han aumentado.

Becker [3] propone el modelo que se describe a continuación, para representar la función de utilidad de un delincuente:

$$EU_j = p_j U_j(Y_j - f_j) + (1 - p_j) U_j(Y_j)$$

Donde EU_j es la utilidad esperada del delito j , p_j la probabilidad de aprehensión, Y_j es el beneficio, tanto físico como psicológico, en términos monetarios de cometer el crimen y f_j es el equivalente monetario del castigo en caso de ser aprehendido.

Luego, la utilidad de un delito disminuye cada vez que la probabilidad de aprehensión aumenta (siempre que exista castigo de por medio):

$$\frac{\partial EU_j}{\partial p_j} = U_j(Y_j - f_j) - U_j(Y_j) < 0$$

Como también disminuye toda vez que el castigo aumenta, siempre que la utilidad marginal del beneficio menos el castigo sea positiva:

$$\frac{\partial EU_j}{\partial f_j} = -p_j U'_j(Y_j - f_j) < 0$$

En el modelo desarrollado en el Capítulo 4 se utiliza esta formulación económica para establecer que la vigilancia, y otras intervenciones orientadas a la prevención del crimen, tienen efecto en la percepción de los delincuentes respecto a los riesgos involucrados al cometer un delito en un lugar y por tanto en la utilidad esperada que le reporta dicha actividad.

2.3. Criminología Ambiental

La principal premisa de la criminología ambiental es que la ocurrencia de un evento criminal se entiende como la confluencia de víctimas, victimarios y la ley, de forma muy específica, y en un momento y lugar en particular. El objetivo es buscar patrones del crimen, y explicarlos a través de factores e influencias ambientales que puedan ser usados luego para generar estrategias que permitan prevenirlo.

Según Wortley [45], la criminología ambiental se basa en tres premisas:

1. El comportamiento criminal se ve significativamente influenciado por la naturaleza del lugar donde este ocurre. La perspectiva ambiental se basa en el hecho de que un crimen es el resultado de la interacción de una persona con el ambiente en una determinada situación. El ambiente puede tanto propiciar como desfavorecer la ocurrencia de un evento criminal.
2. La distribución del crimen en el tiempo y espacio no es aleatoria. Debido a que el comportamiento criminal depende de factores situacionales, los eventos criminales se localizan en aquellos lugares que presentan características que favorecen la convergencia de víctimas, victimarios en ausencia de vigilancia. El crimen se concentrará en aquellos lugares donde existan mayores oportunidades de producirse. Estas concentraciones varían, entonces, geográficamente en una ciudad según el tipo de calle, el barrio o los eventos geográficos contenidos, como también presentan diferencias según la hora del día, el día de la semana y el mes del año donde se observe.
3. Comprender la manera en la que los factores ambientales afectan los patrones delictivos es una información poderosa para la investigación y prevención del crimen. Este conocimiento permite a la policía concentrar los recursos destinados al crimen, siempre limitados, en lugares bien delimitados y específicos.

La prevención del crimen mediante la intervención de variables y condiciones ambientales es el objetivo primordial de la prevención situacional. La idea principal es modificar el ambiente, de manera de reducir las ganancias del crimen y aumentar los riesgos involucrados [9]. De esta forma, la prevención situacional del crimen toma elementos de la criminología ambiental y los aplica en el contexto de la economía del delito, revisada en la subsección anterior. En particular, se busca disminuir las oportunidades que pudieran conducir a un crimen ubicando algún recurso policial, modificando la distribución de los objetos y espacios en un determinado lugar, mejorando la iluminación, instalando alarmas y cámaras de vigilancia u otras intervenciones orientadas al lugar. Basándose en esto, se originan diferentes estrategias policiales cuyo objetivo principal es la prevención, más que la captura y detección del crimen.

2.4. Localización y Desplazamiento del Crimen

Tal como se ha planteado hasta ahora, el contexto ambiental y situacional tienen un gran impacto en la ocurrencia de un evento delictivo. La forma en que está construida la ciudad y en la que la rutina diaria se desarrolla en este espacio físico son determinantes fundamentales en las oportunidades y riesgos que presentan diferentes lugares para un potencial infractor de la ley. Siguiendo esta línea, identificar dónde ocurren los eventos delictivos y cuáles son las características que hacen de ellos áreas atractivas o deseables para algunos delincuentes, es fundamental para generar una estrategia de prevención del crimen. P. Brantingham y P. Brantingham [6] caracterizan e identifican lugares en la ciudad como *generadores* y *atractores* como una forma de comprender cómo el crimen se localiza y distribuye, tanto física como temporalmente. En las Secciones 2.4.2 y 2.4.1 se desarrollan de forma explícita estas dos caracterizaciones.

Cuando se interviene un lugar, ya sea aumentando los riesgos, los esfuerzos necesarios para cometer un crimen o disminuyendo la recompensa que se podría recibir al cometerlo, es natural que algunos delincuentes traten de encontrar nuevos lugares que les permitan seguir con sus actividades. Esta idea constituye la mayor crítica a la prevención situacional, pues se argumenta que, al intervenir un lugar, el efecto no es una reducción del crimen, sino un desplazamiento de éste. Algunos estudios demuestran que si bien el desplazamiento del crimen ocurre, este es solo parcial, es decir, existe efectivamente una reducción del número de delitos, desplazándose otro resto [5]. La idea principal de estudiar el desplazamiento del crimen es comprender el efecto global de una intervención en un determinado lugar e intentar anticiparse a dicho desplazamiento. En la Sección 2.4.3 se ahondará sobre este tema en particular.

2.4.1. Generadores del Crimen

Brantingham and Brantingham [6] clasifican como generadores del crimen a aquellas áreas donde una gran cantidad de gente es atraída por razones que pudieran no estar relacionadas con ninguna motivación criminal en particular, o con el delito que pudieran terminar cometiendo. Entre estos lugares se pueden encontrar, por ejemplo, centros comerciales, barrios con concentración de oficinas o destinados a la entretención (restaurantes, pubs, etc.), estadios o cualquier otro lugar que produzca altas concentraciones de personas.

Los generadores del crimen, al proveer una gran concentración de personas en un determinado momento y lugar, provocan frecuentemente situaciones que podrían propiciar diferentes actos delictivos, y por tanto suelen generar altos números de eventos criminales. Mezclados entre la gente, y probablemente sin intenciones previas de cometer delito alguno, los potenciales delincuentes podrían notar las oportunidades que se presentan en dichos lugares y aprovecharlas si poseen la motivación criminal suficiente, en ese mismo momento, o bien, posteriormente.

Luego, estos lugares generan crimen al proveer las oportunidades a los potenciales delincuentes que por ahí pasan y que podrían no tener ninguna intención previa de cometer delitos en ese lugar.

2.4.2. Atractores del Crimen

Los atractores del crimen son lugares particulares, áreas, barrios, comunas en las que es sabido que existen buenas oportunidades para cometer un delito. Lugares donde delincuentes, fuertemente motivados y con serias intenciones de delinquir, son atraídos por las bien sabidas oportunidades para algún tipo particular de delito [6]. En este caso, los delincuentes acuden al lugar explícitamente con intenciones criminales, a diferencia de los generadores del crimen donde las oportunidades se notan casualmente al estar haciendo otra cosa. Como ejemplos se pueden poner barrios con alta concentración de bares, áreas de prostitución, áreas con tráfico de drogas, centros comerciales de gran tamaño, estacionamientos, u otros.

En estos casos, el delincuente se dirige específicamente a un determinado barrio para solicitar un acto de prostitución, a un área de tráfico de drogas para comprar o venderla, o bien a un centro comercial a asaltar una determinada joyería. Estos delitos en general son cometidos por personas que no viven en el mismo barrio, sino que viajan a través de la ciudad para llegar a estos puntos de interés. La atracción es generada por el pasado personal y experiencia del delincuente, que ha calificado el lugar como uno atractivo para cometer dicho delito.

2.4.3. Desplazamiento del Crimen

Una de las mayores críticas que se le hace a la criminología ambiental y la prevención situacional es que dichas técnicas no reducen en realidad el crimen, sino que meramente los desplazan a otro lugar, otro horario o bien otro tipo de crimen [1,36]. Otros estudios argumentan que la prevención situacional efectivamente reduce el crimen y que el desplazamiento observado es solo parcial, es decir, solo una parte del crimen erradicado en la zona intervenida sufre un desplazamiento de algún tipo [5,29].

El desplazamiento del crimen ante una intervención puede ocurrir entonces en menor o mayor grado, lo interesante sería poder anticiparse a dicho evento y generar medidas para controlarlo. De esta forma se podría obtener una visión más global del efecto producido por una intervención y por tanto generar mejores medidas para combatir y prevenir el crimen.

Brantingham y Brantingham [7] describen algunos patrones para anticipar el desplazamiento del crimen luego de una intervención en determinado lugar. El efecto será diferente dependiendo si el lugar es un atractor o generador del crimen o bien uno neutral. Un generador del crimen, dado que la ocurrencia de los delitos se debe a un factor de casualidad al presentarse las condiciones adecuadas cuando el potencial delincuente se encontraba haciendo otra cosa, no debería presentar desplazamiento del crimen, puesto que la intervención elimina la oportunidad de delinquir. Un atractor del crimen por otro lado debería presentar desplazamiento, puesto que los delincuentes se trasladan específicamente a esos lugares para cometer un crimen, en caso de intervenirlos, es probable que se movilen a otro atractor de características similares.

En general, se ofrecen los siguientes patrones [7] de desplazamiento en caso de intervenir un atractor del crimen:

1. Es probable observar un desplazamiento de la actividad criminal desde un atractor del crimen a los barrios aledaños a este, si existen los candidatos adecuados. Dado que el delincuente se traslada para cometer el crimen, una intervención, más que detener el comportamiento delictivo, lo llevará a buscar un nuevo lugar que le ofrezca mejores

oportunidades, siendo los barrios aledaños los lugares donde se minimiza el esfuerzo de la búsqueda.

2. Otra opción es que la actividad criminal se desplace al barrio donde el delincuente vive, antes que otros barrios cercanos. Esto minimiza la incomodidad psicológica que le produce estar en barrios poco familiares o con características muy disimiles con los que observa en el propio.
3. Como última alternativa, la actividad criminal puede desplazarse desde el atractor del crimen intervenido a otro nodo atractor del crimen. Esta es otra opción que reduce el esfuerzo de búsqueda y la incomodidad psicológica de estar en un lugar poco familiar. Esta alternativa es menos satisfactoria que las dos anteriores, pero podría ocurrir en caso de que no le sea posible al delincuente actuar de las otras dos maneras.

2.5. Predicción del Crimen

Según las premisas de la criminología ambiental, las características de un lugar influyen en el comportamiento criminal, los crímenes, por tanto, se concentran en aquellos lugares que presentan mejores condiciones y oportunidades. Si se denomina atractividad a la valoración de los delincuentes respecto a un lugar, se esperaría que los lugares más atractivos posean una mayor concentración de crímenes. La atractividad de un lugar debería sintetizar todas aquellas características, percibidas por los delincuentes, que podrían facilitar o entorpecer un delito: las vías de escape, la cantidad y el tipo de gente circulando, la presencia de vigilancia, características estructurales y los riesgos involucrados. Entendiendo y analizando la atractividad, se estudia y comprenden las preferencias y percepciones de los delincuentes respecto a cada lugar. A partir de estos análisis, es posible realizar predicciones respecto a los crímenes, y sus distribución, en el futuro.

La idea de una capa de atractividad describiendo la ciudad no es nueva en la literatura, técnicas comúnmente utilizadas son el mapeo de densidades y cálculos de kernels sobre el número de crímenes registrados en cada sector de la ciudad [16]. El número de crímenes registrados se relaciona directamente con la atractividad de un sector: mientras mayor el atractivo, mayor número esperado de crímenes. Estas técnicas utilizan como hipótesis que el historial de crímenes ocurridos es el mejor predictor de crímenes en el futuro, en otras palabras, asumen que la atractividad se mantiene relativamente constante en el tiempo, o que la valoración de los delincuentes respecto a un lugar es estable.

Algunos investigadores han aplicado algunas técnicas de minería de datos más avanzadas para analizar la distribución de crímenes y generar una predicción. Xue *et al.* incluye el proceso de elección de los delincuentes y sus preferencias, considerando las características y ventajas de un lugar, para explicar la concentración de crímenes en la ciudad y predecir su ocurrencia en el futuro [46]. Corcoran *et al.* realiza una clusterización de la ciudad, luego, a través de una red neuronal, utiliza los crímenes ocurridos en un determinado período para predecir los crímenes esperados en el período siguiente [11]. Brown presenta algunos modelos jerárquicos para estimar la probabilidad de observar un delito en un lugar, en base a una regresión lineal sobre sus características, donde los coeficientes de la regresión se distribuyen como variables multinomiales de media y varianza desconocidas, que también se deben estimar [21].

Lazzati *et al.* desarrolla un modelo basado en teoría de juegos para predecir el lugar

donde los delincuentes se deberían concentrar, tomando en consideración el atractivo que posee cada lugar, la vigilancia existente y la presencia de otros delincuentes [27]. Este modelo, que da por conocido el atractivo de los lugares y el efecto de la vigilancia en la utilidad esperada, incluye el efecto desplazamiento que una intervención pudiera provocar, algo no muy común en la literatura. Una metodología denominada modelamiento del riesgo del terreno (Terrain Risk Modeling) incorpora diferentes niveles de riesgo generados por distintas variables o características presentes en el lugar, para generar un único indicador que se utiliza para determinar los lugares más riesgosos de la ciudad [25].

El común objetivo de la mayoría de los modelos y metodologías revisadas en la literatura es proveer información relevante para la toma de decisiones de la policía u otras organizaciones, de manera de poder localizar su recursos de manera eficiente en la ciudad. En general se utiliza el historial de crímenes pasados, junto a las características físicas de los lugares, como base de predicción de los crímenes futuros. Algunos pocos modelos consideran el efecto de una intervención policial u de otra índole, en la atraktividad de la ciudad, como lo es el efecto desplazamiento.

Capítulo 3

Modelos de Markov Ocultos y Estadística Bayesiana

3.1. Introducción

Es común enfrentarse a un problema donde se desea explicar alguna variable, o fenómeno, que se distribuye según alguna distribución de probabilidad. Los modelos empleados generalmente intentan estimar el valor de los parámetros para ajustar lo mejor posible la distribución de probabilidad a las observaciones de la variable estudiada. En un modelo tradicional se asume que estos parámetros estimados, y por tanto la distribución de probabilidad que genera la variable observable, son constantes en el tiempo. Sin embargo, existen casos donde la distribución de una variable objetivo puede cambiar en el tiempo, dependiendo de alguna condición o característica particular del fenómeno. En una situación del estilo, un modelo tradicional podría ser incapaz de explicar el fenómeno a cabalidad a causa de esto.

A modo de ejemplo en el contexto del crimen, una unidad policial desea estudiar el número de crímenes registrados en una calle en particular. La calle en estudio es tranquila y de baja concurrencia la mayor parte del tiempo, pero es sede de una masiva feria de la comunidad dos días de cada semana. Los días de feria, la calle se convierte en un atractor y generador del crimen debido a la multitud que ahí se reúne. Un modelo tradicional probablemente subestime la tasa de crímenes en los días de feria y la sobrestime el resto de la semana, tomando un valor intermedio. En esta situación, un modelo con parámetros estáticos podría no ser el más adecuado para capturar el fenómeno.

Es fácil notar en este ejemplo, que la condición que genera las diferencias en las tasas esperadas de crímenes corresponde a la feria comunitaria que genera una mayor cantidad de oportunidades para delinquir. Lo adecuado sería estimar los parámetros del modelo por separado los días de feria y el resto de la semana. En muchos casos, sin embargo, la condición que genera un cambio en la tasa de crímenes no es tan simple de identificar y de medir. Suponiendo en el ejemplo anterior, que la policía envía los registros de los robos a un equipo de expertos para su análisis. El equipo de expertos intentará modelar la distribución de crímenes utilizando solo el registro de crímenes, desconociendo la existencia de la feria. En esta situación capturar las diferencias de tasas de crímenes durante el tiempo se convierte en una tarea complicada. En resumen, se tiene una señal observable: los registros de robos, y una condición en el sistema que genera cambios en la tasa esperada de crímenes, que es inobservable para los expertos:

la feria comunitaria. Esta caracterización es la que da origen a un modelo de Markov oculto, HMM por sus siglas en inglés.

En un modelo HMM se asume que el sistema que se desea estudiar se encuentra entre algún conjunto de estados que no se puede observar. Para los expertos en el ejemplo, la calle estudiada se encuentra entre dos posibles estados: de alta o de baja atraktividad, observándose una gran cantidad de crímenes en el primer caso y una baja cantidad en el segundo. La idea principal de un HMM es relacionar el estado oculto del sistema a través de una variable observable que depende de dicho estado. Si la atraktividad de un lugar se considera oculta, el número de crímenes ahí registrados, que depende del estado del sistema, sí se puede observar: mientras más atractivo, más crímenes se esperan. En un modelo HMM los parámetros que intentan explicar el número de crímenes diarios tomarán diferentes valores en cada estado del sistema, pudiendo ajustarse mejor al fenómeno que se intenta explicar: un valor para los días de feria y otro para un día normal.

En el modelo HMM que se desarrollará En el Capítulo 4, se desarrolla un modelo HMM para modelar el número de crímenes esperados en un lugar y donde se incluyen variables como la vigilancia policial, u otras intervenciones orientadas a la prevención situacional. Esta inclusión genera un cambio en la estructura clásica utilizada en los modelos HMM, inhabilitando los métodos de estimación más estándar que se suelen utilizar, como el método de Baum & Welch. En estos casos, donde el modelo posee una estructura más compleja y aumenta el número de parámetros a estimar, la utilización de métodos basados en estadística Bayesiana y simulaciones de Monte Carlo se han vuelto cada vez más populares.

En lo que sigue de este capítulo se realizará una revisión de los contenidos y base teórica necesaria para la construcción y estimación de los modelos HMM. En particular, se comenzará con la descripción y estructura de estos modelos, para finalizar explicando los fundamentos básicos de la estadística Bayesiana y algunas metodologías de estimación mediante simulaciones de Monte Carlo.

3.2. Modelo de Markov Oculto

Un modelo de Markov oculto representa un fenómeno que considera dos tipo de señales relacionadas entre sí: los estados del sistema y la variable observada. Se asume que el sistema se encuentra en uno dentro de un conjunto de posibles estados. Cada período el sistema cambia de estado, tras lo cual emite una señal de la variable observada según una distribución de probabilidad propia de dicho estado. Como resultado se obtienen dos secuencias a lo largo del tiempo: una secuencia de estados y otra de la variable observada. El estado en que se encuentra el sistema se considera latente, no se puede conocer con certeza en qué estado se encuentra, y dinámico, al cambiar período a período. La Figura 3.1 da cuenta de la relación entre las señales consideradas en un HMM.

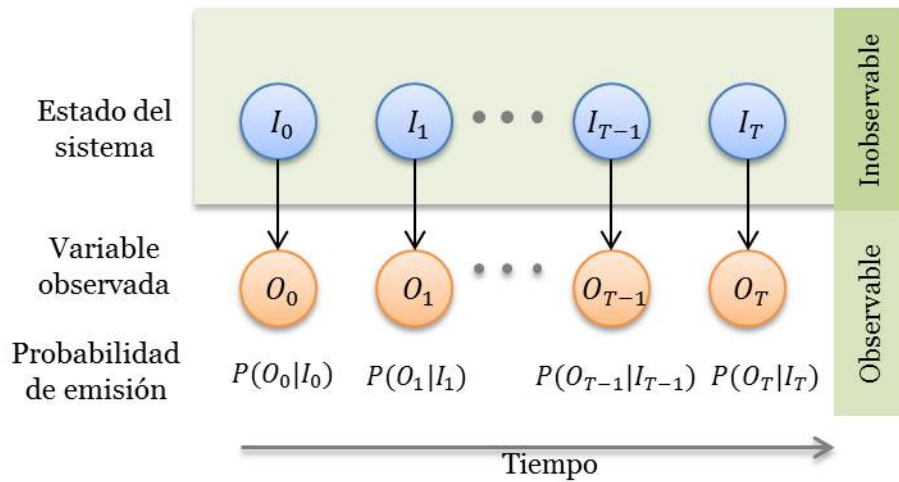


Figura 3.1: Diagrama de la secuencia de señales obtenidas a partir de un HMM.

3.2.1. Sistemas de Markov

En un modelo HMM se asume que los estados del sistema siguen un proceso de Markov, con la particularidad que los estados de dicho proceso permanecen ocultos. Un proceso se dice de Markov, cuando la probabilidad del sistema de encontrarse en un determinado estado, solo depende del estado del sistema en el período anterior.

$$P(S_t = s_t | s_1, s_2, \dots, s_{t-1}) = P(S_t = s_t | s_{t-1})$$

Un sistema de Markov se define mediante un conjunto de estados y las probabilidades de transición entre estos estados [30], tal como se puede observar en el ejemplo de la Figura 3.2 que representa un sistema de Markov con tres estados.

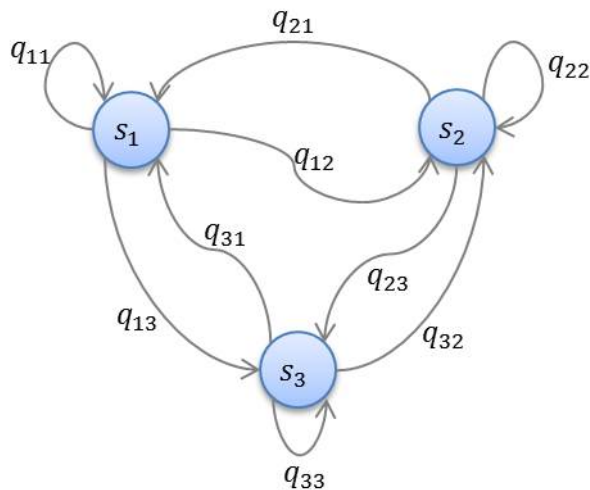


Figura 3.2: Representación de un sistema de Markov con tres estados y las probabilidades de transición entre estos.

3.2.2. Objetivo de un HMM

La idea principal de un HMM es intentar descubrir las probabilidades de transición entre estados y la función de distribución de la variable observada en cada uno de ellos, a partir de la información disponible: la secuencia de la variable observada a lo largo del tiempo.

Un ejemplo común para ilustrar el concepto HMM es el del lanzamiento de la moneda, utilizado por Rabiner y Juang [35]. En este experimento se ubica un observador en una habitación que contiene una barrera visual, una cortina por ejemplo, por lo que es incapaz de saber qué está ocurriendo tras ella. Al otro lado de la cortina se encuentra otra persona, que realiza un experimento de múltiples lanzamientos de moneda. Esta persona no le dará ninguna información al observador, salvo la secuencia de resultados cara o sello obtenida de los lanzamientos. Es decir, se realiza una secuencia de lanzamientos ocultos, desde el punto de vista del observador, que solo conoce el resultado de dichos lanzamientos.

Dado el experimento anterior, se busca construir un HMM que explique la secuencia de caras y sellos observada. Se podría considerar por ejemplo, que el lanzador de monedas está utilizando dos monedas equilibradas y que escoge cada una de ellas con igual probabilidad. En este caso el estado del sistema es la moneda escogida para lanzar, teniendo dos estados posibles: una para cada moneda. Otra alternativa es que las monedas estén cargadas en vez de equilibradas, o que sean más de dos monedas. Los siguientes diagramas ilustran la alternativa de dos monedas, equilibradas en el primer caso y cargadas en el segundo.

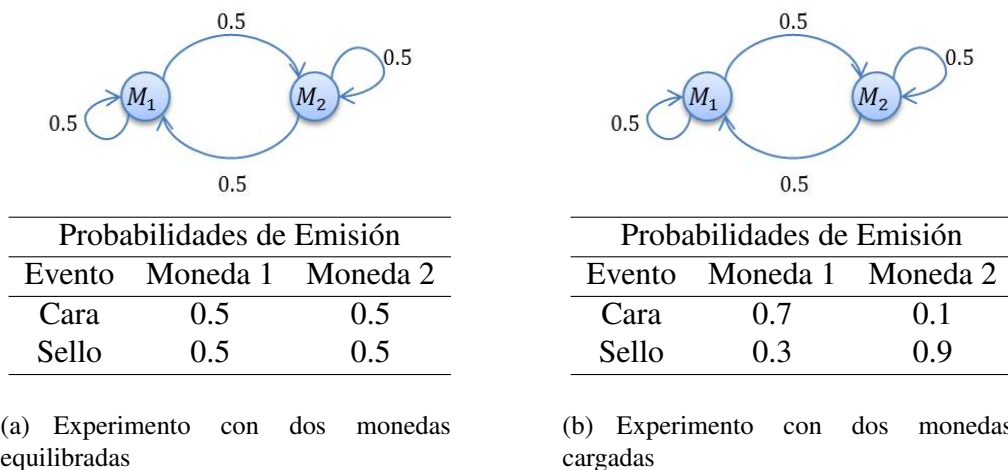


Figura 3.3: Posibles estructuras de HMM para el experimento del lanzamiento de la moneda.

El ejemplo anterior podría haber sido modelado con tres, cuatro o veinte estados, asumiendo que la persona arrojando monedas estuviera escogiendo con cierta probabilidad entre un conjunto, de tal número, de monedas para lanzar. La elección del número de estados no es trivial para el problema y constituye otra de las incógnitas que se desea resolver. Podría haberse considerado también que la probabilidad de escoger una moneda fuera mayor que escoger la segunda, es decir considerar otras probabilidades de transición.

Para estructurar un modelo HMM se deben determinar los valores, de manera de explicar de la mejor forma posible la secuencia de la variable observada, los siguientes elementos:

- El número de estados a considerar

- Probabilidades de transición entre estados.
- Distribuciones de probabilidad de la variable observada en cada estado.

3.2.3. Elementos de un HMM

De manera de formalizar las ideas anteriores y establecer la notación que será utilizada en lo que sigue de este trabajo, corresponde establecer las propiedades y elementos que componen un modelo HMM.

Las propiedades de los HMM que se utilizarán a lo largo de este trabajo corresponden a las enunciadas por Rabiner y Juang [35]:

1. Se considera un número finito de estados. Los estados de un sistema pueden ser interpretados de forma diferente según el problema que se intente modelar, pero se puede decir que, mientras el sistema está en un estado específico, el proceso de la variable observable posee algunas características distintivas.
2. En cada período de tiempo, de duración d , el sistema cambia a un nuevo estado según las probabilidades de transición, siguiendo la propiedad Markoviana. No restringiendo que el sistema pueda permanecer en el mismo estado durante dos o más períodos consecutivos.
3. Después de ocurrida la transición, ocurre una realización de la variable observable de acuerdo a una función de distribución que depende del estado en el que se encuentre el sistema. Existen, entonces, tantas distribuciones de probabilidad de la variable observable, como estados en el sistema.

Por otro lado, la notación que se utilizará para denotar los diferentes elementos del modelo, en lo que sigue de este trabajo, será:

d	es la duración de los períodos.
T	es el número total de períodos.
N	es el número de estados en el modelo.
$S = \{s_1, s_2, \dots, s_N\}$	es el conjunto de estados
$Q = \{q_{ij}\}_{i,j=1}^N$	es la matriz de probabilidades de transición entre estados, donde $q_{ij} = P(S_{t+1} = s_j S_t = s_i)$
$B = \{f_1, f_2, \dots, f_N\}$	es el vector con las funciones de densidad de probabilidad de la variable observable, en cada uno de los estados del sistema.
$\pi = \{\pi_1, \pi_2, \dots, \pi_N\}$	es el vector de probabilidades para el estado inicial.

Según esta descripción, el sistema comienza en el período inicial en algún estado $i_0 \in S$ con probabilidad π_{i_0} , generando una observación que se distribuye según f_{i_0} . Al final del período realizará una transición a algún estado $i_1 \in S$, con probabilidad $q_{i_0 i_1}$, donde generará una nueva observación para la variable observada distribuida como f_{i_1} . Siguiendo este proceso, se obtendrá como resultado una secuencia de la variable observable $Y = \{Y_0, Y_1, \dots, Y_T\}$ y una secuencia de estados, inobservable, $I = \{i_0, i_1, \dots, i_T\}$. En particular, se dirá que un HMM está definida de forma compacta por $\lambda = (Q, B, \pi)$, que son los elementos que determinan el proceso descrito.

En un modelo tradicional de HMM, λ es constante en el tiempo, pero en los modelos desarrollados en este trabajo se incluirán algunas variables que afectarán sus componentes. Lo

anterior implica que las probabilidades de transición y las de emisión dependerán del valor que tomen estas variables durante cada período, $\lambda_t = (Q_t, B_t, \pi)$.

3.2.4. Los Tres Problemas de un HMM

Tal como se explicó en el experimento de las monedas, muchos modelos pueden ser utilizados para intentar explicar una misma secuencia de la variable observada. Lo ideal en este caso, sería contar con una medida que establezca la idoneidad del modelo para explicar la secuencia de la variable observada, que corresponde al primer problema de un HMM. Teniendo una medida como la anterior, es necesario encontrar el modelo que la maximice, es decir, un algoritmo o metodología que permita encontrar el mejor modelo que explique la variable observada, lo que constituye el segundo problema de un HMM. Finalmente, una vez que se obtiene el mejor modelo que explique la secuencia observada, es necesario utilizar el modelo para predecir a partir del mismo, lo que constituye el tercer y último problema de un HMM [35].

La medida a utilizar para comparar los modelos, la verosimilitud del mismo, los algoritmos y metodologías para la estimación de los parámetros y para la predicción a partir de estos últimos, serán revisados con mayor detalle en lo que sigue de esta sección.

Problema 1: Definir una Medida de Ajuste

Para evaluar un modelo HMM usualmente se calcula la probabilidad de observar la secuencia observada $Y = \{Y_0, Y_1, \dots, Y_T\}$, dado los parámetros y estructura del modelo, a lo que se denomina verosimilitud. La verosimilitud, $L(\lambda)$, indica cuán probable es que la secuencia Y haya sido producida por el sistema determinado por $\lambda = (Q, B, \pi)$, y es la medida de ajuste que se considerará en lo que sigue de este trabajo para comparar diferentes modelos y especificaciones.

La probabilidad de observar la secuencia Y dado el modelo λ y una secuencia específica I de estados, puede expresarse a través de la distribuciones de probabilidad de la variable observada, $B = \{f_1, f_2, \dots, f_N\}$.

$$P(Y|\lambda, I) = f_{i_0}(Y_0)f_{i_1}(Y_1) \dots f_{i_T}(Y_T)$$

Por otro lado, la probabilidad de observar una secuencia de estados I dado el modelo λ se puede expresar a partir de la matriz de transición y las probabilidades iniciales.

$$P(I|\lambda) = \pi_{i_0}q_{i_0i_1}q_{i_1i_2} \dots q_{i_{T-1}i_T}$$

Luego, la probabilidad de observar la secuencia Y dado el modelo λ , $P(Y|\lambda)$, se puede escribir a partir de la suma de las probabilidades condicionales sobre todas las posibles secuencias de estados I^* .

$$L(\lambda) = P(Y|\lambda) = \sum_{I \in I^*} P(Y|\lambda, I)P(I|\lambda)$$

El problema de esta representación es que el número de elementos en la suma es exponencial, del orden de N^T : la cardinalidad del conjunto I^* . Esto significa que, incluso para

pequeños valores de N y T , se requiere de grandes recursos computacionales y eventualmente ser imposible de manejar.

Para evitar lo anterior se utiliza la representación matricial de la función de verosimilitud propuesta por MacDonald y Zucchini [28], que permite su cómputo de manera simplificada y con bajos costos computacionales.

$$L(\lambda) = \pi \tilde{m}_1 Q \tilde{m}_2 \dots Q \tilde{m}_T 1'$$

Donde \tilde{m}_i es una matriz de $N \times N$, con los elementos del vector $m_i = \{f_1(Y_i), f_2(Y_i), \dots, f_N(Y_i)\}$ en su diagonal y $1'$ es un vector de unos en \mathbb{R}^N traspuesto.

Problema 2: Maximizar la Medida de Ajuste

Cuando se estima un modelo HMM se desea ajustar los parámetros $\lambda = (Q, B, \pi)$ de manera de obtener el modelo que explique de la mejor forma posible la data observada. Esto se traduce en un problema de optimización: maximizar la medida de ajuste, la verosimilitud del modelo, dado los parámetros.

$$\max_{\lambda} L(\lambda) = \max_{\lambda} P(Y|\lambda)$$

Para encontrar los valores óptimos de λ , el método más tradicional es el algoritmo EM (Expectation-Maximization) denominado Baum-Welch, que utiliza el mecanismo de inducción hacia adelante y hacia atrás para encontrar un punto fijo que maximiza la verosimilitud [2]. Este método de estimación requiere de una estructura específica y distribuciones de probabilidad conocidas para ser implementado, limitando el tipo de modelos que se pueden estimar a partir de ella.

Durante las últimas décadas, sin embargo, y con el aumento de la capacidad computacional, se han visto potenciados otros métodos para la estimación de los modelos HMM, principalmente basados en estadística Bayesiana. El uso de muestreos de Monte Carlo, como lo son el muestreo de Gibbs y el algoritmo de Metropolis, han ido ganando cada vez más popularidad para estimar modelos HMM, permitiendo mayor flexibilidad en su estructura y la inclusión de un mayor número de parámetros [37]. En las secciones 3.3 y 3.4 se revisarán las bases de la estadística Bayesiana y los muestreos de Monte Carlo, para luego ser usadas, en el Capítulo 4, en la estimación de los modelos desarrollados en este trabajo.

Problema 3: Predecir a partir del Modelo

Dada una secuencia observada $Y = \{Y_0, Y_1, \dots, Y_T\}$, se desea encontrar la secuencia óptima de estados $S = \{S_0, S_1, \dots, S_T\}$ que pudiera haberla originado. Dado que el criterio de optimalidad puede ser diferente en cada caso, existen diversas maneras de resolver un problema de este estilo.

Una alternativa es escoger, en cada período, el estado más probable que haya generado la observación, dada la información hasta ese momento.

$$i_t = \arg \max_{i \in S} P(i_t = i | Y^t = \{Y_1, Y_2, \dots, Y_t\}, \lambda)$$

Donde $P(i_t = i|Y^t, \lambda)$ para un estado i cualquiera corresponde a la i -ésima componente del vector:

$$p^t = \frac{\pi \tilde{m}_1 Q \tilde{m}_2 \dots Q \tilde{m}_t}{P(Y|\lambda)} = \frac{\pi \tilde{m}_1 Q \tilde{m}_2 \dots Q \tilde{m}_t}{L(\lambda)_t}$$

Donde

$$L(\lambda)_t = \pi \tilde{m}_1 Q \tilde{m}_2 \dots Q \tilde{m}_t 1'$$

Luego

$$i_t = \arg \max_{i \in S} [p_i^t]$$

Otra alternativa es escoger, en cada período, el estado más probable que haya generado la observación, dada la secuencia observable completa Y .

$$i_t = \arg \max_{i \in S} P(i_t = i|Y, \lambda)$$

Esta probabilidad puede expresarse como la probabilidad de llegar al estado i dada las observaciones hasta ese momento $Y^t = \{Y_0, Y_1, \dots, Y_t\}$ y la probabilidad de observar la secuencia $Y^{-t} = \{Y_{t+1}, Y_{t+2}, \dots, Y_T\}$ a partir del estado i . Como este es un proceso de Markov, solo el período $(t+1)$ tiene dependencia con el estado en t , por lo que se puede desestimar el resto de las observaciones.

$$P(i_t = i|Y, \lambda) = \frac{P(i|Y^t, \lambda)P(Y^{-t}|i_t = i, \lambda)}{P(Y|\lambda)} = \frac{p_i^t Q_{(i \cdot)} \tilde{m}_{t+1}(Y_{t+1})}{L(\lambda)}$$

3.3. Estadística Bayesiana

El principal objetivo de la estadística Bayesiana es usar la información disponible para hacer inferencia sobre cantidades desconocidas. Se le denomina información disponible a cualquier conocimiento que le de estructura a lo desconocido: data, conocimiento experto respecto a un tema, información a priori sobre segmentos, tipos o categorías, etc. En general, se clasifica la información en dos tipos: (1) información que se obtiene de la data y (2) información independiente o *a priori* de la data. Por otro lado cuando se habla de cantidades desconocidas, refiere a cualquier parte de un modelo del que no se posee información: parámetros, efectos de alguna variable, predicciones de alguna señal en el futuro, etc. La inferencia bayesiana implica inferir probabilidades sobre las variables desconocidas, condicional a la data y la información a priori, más que sobre los valores exactos de las variables (como en un modelo tradicional). Los teoremas, definiciones y descripciones contenidas en esta sección están basadas en [37], fuente que puede ser revisada en caso de necesitar una revisión más profunda de estos temas.

3.3.1. Teorema de Bayes

Sea θ un conjunto de valores desconocidos que, según la información *a priori* que se posee, sigue una distribución de probabilidad $p(\theta)$. La información es introducida en el modelo a través de la distribución de probabilidad de la data $p(D|\theta)$, donde D corresponde a la data observable. Cualquier función proporcional a la probabilidad de observar la data condicional a los valores desconocidos, $p(D|\theta)$, se le denomina verosimilitud $\ell(\theta)$.

El objetivo principal es hacer inferencia sobre la distribución de probabilidad de θ condicional a la data e información disponible: $p(\theta|D)$. El teorema de Bayes establece, sobre las probabilidades condicionales que:

$$p(\theta|D) = \frac{p(D, \theta)}{p(D)} = \frac{p(D|\theta)p(\theta)}{p(D)}$$

La combinación de la data y la información a priori, dan como resultado la distribución a *posteriori* de θ : $p(\theta|D)$. La fórmula anterior puede ser expresada en función de la verosimilitud del modelo:

$$p(\theta|D) \propto \ell(\theta)p(\theta)$$

3.3.2. Distribución a Posteriori y Conjugados

En general, los problemas abordados desde la estadística Bayesiana suelen tener distribuciones a posteriori de alta dimensionalidad, por lo que es necesario sintetizar estas distribuciones para poder visualizar los resultados de mejor manera. Donde, en la estadística clásica se reportan las medias y desviación estándar, en la estadística Bayesiana se reportan los momentos de la distribución a *posteriori*: medias y desviaciones estándar a posteriori. Sin embargo, dado que en muchas ocasiones las distribuciones a posteriori no siguen una distribución normal, estos estadísticos podrían no ser de mayor relevancia. El problema general al intentar generar estadísticos y valores describiendo la distribución de la variable, sin definir un el estadístico en particular, es encontrar el valor esperado de una función $h(\theta)$ dada la distribución a posteriori. Escogiendo la función $h(\cdot)$ adecuada, se puede obtener cualquier tipo de estadístico: percentiles, distribuciones marginales, momentos, u otros. El problema a resolver se traduce entonces a calcular:

$$E_{\theta|D}[h(\theta)] = \int h(\theta)p(\theta|D)d\theta = \int h(\theta)\frac{\ell(\theta)p(\theta)}{p(D)}d\theta$$

Usualmente, solo se dispone de la función de distribución sin normalizar, $\ell(\theta)p(\theta)$, para representar $p(\theta|D)$. Sin embargo se tiene que

$$\int \frac{\ell(\theta)p(\theta)}{p(D)}d\theta = 1$$

Luego dividiendo la ecuación por esta expresión, calcular $E_{\theta|D}[h(\theta)]$ se reduce a:

$$E_{\theta|D}[h(\theta)] = \int h(\theta)p(\theta|D)d\theta = \frac{\int h(\theta)\ell(\theta)p(\theta)d\theta}{\int \ell(\theta)p(\theta)d\theta}$$

Antes de la introducción de los métodos de Monte Carlo, que se revisará en la Sección 3.4, solo los problemas donde estas dos integrales tenían solución analítica podían ser abordadas desde la estadística Bayesiana. Esto restringía el conjunto de priors y verosimilitudes a aquellos que producían distribuciones a posteriori conocidas y cuyas integrales, y la de sus estadísticos descriptivos, podían ser evaluadas analíticamente. Esto impone restricciones en ambos: la función de verosimilitud y priors. En particular se requiere que la distribución a priori sea un conjugado de la función de verosimilitud.

Un prior se dice un conjugado de la verosimilitud si la función a posteriori, derivada de la función de verosimilitud y el prior, tienen la misma clase de distribución de probabilidad

que el prior. Es decir, dado un prior que sigue una distribución F , $p(\theta) \sim F$, y una función de verosimilitud $\ell(\theta)$, si la distribución a posteriori es de clase F , $\ell(\theta)p(\theta) \sim F$, entonces F y $\ell(\theta)$ son conjugados.

En la sección A.2 del Apéndice se puede estudiar el ejemplo de priors conjugados en un modelo de regresión multivariado.

3.4. Metodos de Simulación de Monte Carlo con Cadenas de Markov

El problema de la utilización de conjugados para la estimación e inferencia en modelos Bayesianos, es lo restrictivo en su estructura y las limitaciones que tiene su uso a medida que la dimensión de los parámetros aumenta. Un modelo que considere heterogeneidad en los parámetros, por ejemplo, quedaría fuera del alcance con una estimación de esta forma [37]. Los métodos de simulación, que explotan la estructura de los modelos, han ganado cada vez más espacio, en la medida que se desarrollan algoritmos que reducen los costos computacionales asociados a una estimación de este estilo.

La idea detrás de los métodos de simulación MCMC, por su siglas en inglés, es formular una cadena de Markov para el espacio de los parámetros Θ . Cada valor θ en el espacio vectorial Θ corresponde a un estado de la cadena. Para cumplir con la propiedad Markoviana, la simulación de los parámetros θ en una iteración solo dependen de su valor, estado, en la iteración anterior. Si la cadena tiene una distribución de equilibrio de largo plazo $\pi(\cdot)$, entonces se pueden estimar las integrales deseadas, sobre las distribuciones a posteriori de las variables, a partir de la cadena simulada.

Partiendo desde un punto inicial θ_0 , una cadena de Markov especifica las probabilidades de moverse desde un estado a otro, generando una secuencia de variables aleatorias $\{\theta_0, \theta_1, \dots, \theta_r, \dots\}$. Las probabilidades, o kernels si el modelo considera espacios Θ continuos, de transición en la cadena de Markov pueden ser continuas, discretas o una mixtura entre ambas. En particular, en una cadena de Markov de espacio continuo, el kernel de la cadena, $K(\theta, A)$, indica la probabilidad de que los parámetros tomen valores en el conjunto A en una iteración, dado que se encuentran en el estado θ . Este kernel se define mediante una densidad de probabilidad $p(\cdot, \cdot)$ sobre el espacio vectorial.

$$K(\theta, A) = \int_A p(\theta, \vartheta) d\vartheta$$

Una distribución $\pi(\cdot)$ se dice estacionaria si la probabilidad de que θ se encuentre dentro del conjunto $A \in \Theta$ es equivalente a la probabilidad de estar en cualquier punto del espacio vectorial Θ y pasar al conjunto A según la probabilidad de transición de la cadena, dado que θ se distribuye según π :

$$\int_A \pi(\theta) d\theta = \int_{\Theta} K(\theta, A) \pi(\theta) d\theta = \int_{\Theta} \left[\int_A p(\theta, \vartheta) d\vartheta \right] \pi(\theta) d\theta$$

Otra forma de encontrar la distribución estacionaria de una cadena de Markov es a través de su reversibilidad. Una cadena se dice reversible respecto a una distribución de densidad $w(\theta)$

si la función de transición, $p(\theta, \vartheta)$, satisface:

$$w(\theta)p(\theta, \vartheta) = w(\vartheta)p(\vartheta, \theta)$$

Si se satisface reversibilidad en una cadena, la distribución estacionaria tiene densidad $w(\theta)$.

Si se logra construir una cadena que converja a una distribución estacionaria, entonces se puede calcular el valor esperado a posteriori de cualquier función, para así obtener el valor del estadístico deseado.

$$E[h(\theta)] = \frac{1}{R} \sum h(\theta_r)$$

Cuando la cadena proporciona una distribución estacionaria el límite de $\frac{1}{R} \sum h(\theta_r)$, cuando R tiende a infinito, es equivalente al valor esperado de la función $h(\cdot)$ a posteriori. Existen familias de algoritmos que proporcionan cadenas, para distribuciones a posteriori arbitrarias, que gozan de propiedades de convergencia a la distribución de equilibrio bajo condiciones leves y verificables [37]. En la sección A.3 del Apéndice se revisan dos de ellas, los sampleos de Gibbs y el algoritmo de Metropolis.

Para simular desde una cadena de Markov es necesario:

1. Inicializar θ_0 .
2. Simular θ_1 a partir del estado anterior: $\theta_1 | \theta_0 \sim F(\theta_0)$.
3. Reemplazar θ_0 con θ_1 y repetir el procedimiento R veces.

Normalmente, la cadena pasa, en primera instancia, por un período de transición antes de converger a la distribución estacionaria, por lo que es común descartar las iteraciones iniciales, de manera de difuminar el efecto de las condiciones iniciales θ_0 . Se debe tener en cuenta, que los sampleos en un método de MCMC no son independientes, pues cada iteración se basa en los valores de la iteración anterior, esto puede provocar que la convergencia de la cadena sea muy lenta, pero no tiene mayor trascendencia en la estimación de funciones a posteriori mientras esta dependencia no sea patológica [37].

Capítulo 4

Modelo de Markov Oculto para la Predicción del Crimen

4.1. Introducción

En la Sección 2.5 del Capítulo 2 se desarrolla la idea del atractivo de un lugar como una manera de describir las preferencias de los delincuentes. Para continuar es necesario definir de forma explícita este concepto.

Atractividad: es un atributo del lugar, que resume la utilidad esperada por los delincuentes al cometer un crimen ahí. Considera todos los costos, beneficios y riesgos involucrados que los delincuentes perciben y evalúan según su actividad.

Las percepciones de los delincuentes respecto a las condiciones de un lugar y riesgos involucrados van cambiando en el tiempo, en la medida que ese lugar sufra cambios: estructurales, de uso de suelo, en las medidas de seguridad y en función de su experiencia, y la de sus pares. La atractividad, que resume estas percepciones, puede considerarse dinámica: va cambiando en la medida que la percepción de los delincuentes cambia [23]. Esta dinámica es un factor relevante a considerar cuando se desea conocer el impacto de alguna intervención en el lugar, estudiando la forma en que la atractividad percibida evoluciona con o sin dicha intervención.

Sin embargo, tal como lo enuncia Clarke [9], la estimación de la atractividad, o cualquier parámetro que intente emular la utilidad percibida por un delincuente al cometer un crimen, requiere de data imposible de obtener o la asunción de supuestos poco realistas. Se podría considerar entonces, que el atractivo de un lugar es un estado latente, no se puede observar debido a que las variables y factores que la determinan son demasiados o imposibles de medir.

En este capítulo se propone la utilización de un Modelo de Markov Oculto (HMM), revisado en el Capítulo 3, para la estimación de la atractividad percibida en los diferentes sectores de una ciudad para cometer un determinado crimen. Los delitos registrados son utilizados como la variable observable que depende del estado oculto del sistema, la atractividad. Uno de los objetivos de esta investigación es proporcionar un mecanismo de evaluación y estudio del impacto que una intervención o medida de prevención tiene sobre la atractividad percibida y el número de delitos registrados. Por esta razón se incluyen también variables exógenas en el modelo representando intervenciones o cambios en las condiciones

ambientales, tanto en las probabilidades de transición, como en las de emisión.

La metodología y el modelo presentados a continuación se basan en el trabajo realizado por Netzer [32] en el estudio de la relación del cliente con la empresa. Se busca lograr dos objetivos, primero determinar la forma en que la atractividad de un lugar evoluciona en el tiempo, relacionándolo con el número de crímenes esperados, y segundo, la evaluación del impacto de diferentes alternativas (intervenciones) para la prevención del crimen.

Es importante recalcar la importancia de enfocar el estudio solo en un tipo de crimen, pues la valoración del atractivo de un lugar puede diferir en gran medida entre un tipo crimen u otro. Las características deseadas y buscadas por los delincuentes son específicas a la actividad y delito que se quiera realizar, por lo que se deben estudiar de forma independiente. Solo por citar un ejemplo, se puede observar que un carterista preferirá un lugar con bastante flujo de personas, donde sea fácil pasar desapercibido y desaparecer en la multitud, y otro que intenta un asalto con violencia buscará un lugar solitario y poco concurrido donde no haya testigos del acto. A pesar de que pudiera existir una correlación entre distintos tipos de crímenes que requieran características similares, aun así el impacto de una medida de prevención podría ser diferente en cada uno ellos. En el modelo desarrollado en este capítulo se considerará que la variable observada corresponde al número de delitos de un único tipo de crimen, aunque en el Capítulo 7 se argumentará la posibilidad de considerar más de un tipo de crimen adaptando el modelo.

4.2. Atractividad en el Crimen y Modelo de Markov Oculto

La unidad de estudio considerada en el modelo que se desarrollará en este capítulo corresponde a un lugar dentro de la ciudad, o un sector acotado de la misma. Para definir el tamaño de la unidad de estudio se debe determinar primero cuál el nivel al que se desea enfocar el fenómeno delictual [16]: nivel de calle, manzana, comuna o región, y el radio del impacto que se espera de las medidas incluidas en el modelo.

Cada lugar se encontrará en algún estado de atractividad para la materialización de algún crimen específico. En este modelo se considerará que la atractividad de un lugar viene determinada por un proceso de aprendizaje por parte de los delincuentes, que considera el conocimiento previo respecto a un lugar y las condiciones ambientales o eventos que ocurran durante un período. Las condiciones ambientales incluyen por ejemplo la existencia de vigilancia, el número de personas circulando u otras variables que se relacionan más con el ambiente y rutina de las personas, que con la infraestructura física permanente en el lugar. Cada período las condiciones ambientales confirmarán o negarán las creencias previas que tienen los delincuentes sobre la idoneidad del lugar para cometer un crimen.

La atractividad se puede considerar dinámica en el tiempo, al actualizarse período a período. Asumiendo un conjunto discreto de estados para representar los diferentes niveles de atractividad, luego de cada período el lugar cambia de estado o se mantiene en el mismo. Este cambio se produce según las diferencias entre las creencias previas sobre el lugar y la experiencia real observada en el mismo. A modo de ejemplo, un delincuente cree que determinada esquina es ideal para realizar un robo, considera que es un lugar de gran atractivo, y acude al mismo para aprovechar las oportunidades para delinquir. Estando allí detecta algunas cámaras de seguridad que se instalaron después de su última visita, cambiando su valoración de los riesgos y por tanto bajando el estado de atractividad a uno menor. Según esto y asumiendo

que la atractividad resume todo el proceso de aprendizaje de los delincuentes, la probabilidad de encontrarse en un estado cualquiera depende solo del estado en el que se encontraba en el período anterior y de las condiciones ambientales reinantes en ese período. Definido así, el proceso de actualización de la atractividad cumple la propiedad Markoviana por lo que se puede modelar como una cadena de Markov.

A pesar de considerar de forma no observable el estado de atractividad de un lugar, sí se posee registros y conocimiento sobre el comportamiento delictual a través de las denuncias de los delitos. Asumiendo la teoría de elecciones racionales y considerando los supuestos de la criminología ambiental, revisados en el Capítulo 2, los delitos tienden a concentrarse en los lugares más atractivos y con mayores probabilidades de éxito. Considerando esto, el número de delitos observados en un período depende del estado de atractividad en que se encuentre cada lugar: mientras más atractivo un lugar, más crímenes observados en el mismo. Esta relación entre un sistema Markoviano latente y una variable observable estocástica que depende del estado del sistema, puede modelarse como un HMM.

A través del modelo HMM que se desarrolla a lo largo de este capítulo se busca estudiar la atractividad de un lugar, que se considera latente, a través de la secuencia del número de delitos ahí registrados período a período, que será la variable observada. Variables exógenas que describan las condiciones ambientales permitirán estudiar su impacto en la dinámica del crimen, tanto en las transiciones entre estados como en las probabilidades de emisión.

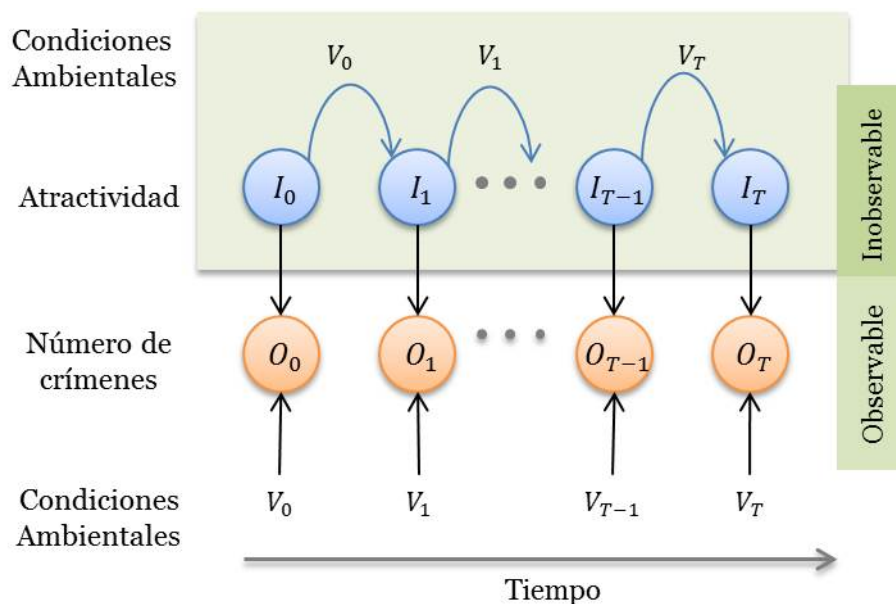


Figura 4.1: Representación del modelo de Markov propuesto para modelar la relación entre la atractividad de un lugar y los crímenes ahí registrados. El modelo incluye el efecto de las condiciones ambientales sobre las transiciones período a período y el número de crímenes observados.

4.2.1. El Efecto de Intervenciones en el Corto y Mediano Plazo

La atractividad de un lugar está determinada entonces por las diferencias entre las expectativas y las características y ventajas reales para cometer un crimen. Las condiciones ambientales, mencionadas en la sección 4.2, resumen todos los factores que un delincuente

considera relevantes, y tal como la atraktividad, se considera inobservable dada la dificultad de determinar y medir todos ellos. Sin embargo, pudieran existir algunas intervenciones, o cambios en las características del lugar que se sospecha tienen incidencia sobre los riesgos, oportunidades y ventajas percibidas por los delincuentes. Entre estas intervenciones se pueden incluir estrategias de vigilancia policial, la instalación de cámaras de seguridad o luminarias, cambios en el uso de suelo de la zona o rutina de las personas que lo habitan, instalación de infraestructura, centros comerciales o cualquier otra medida que afecte las probabilidades de cometer un crimen exitosamente.

De esta forma, aun considerando el estado de atraktividad como algo inobservable en su completa dimensión, existen factores y condiciones que sí se pueden observar que pudieran explicar sus cambios parcialmente. Por esta razón, se incluyen en el modelo variables que describan alguna condición específica en el ambiente y que se sospecha tienen incidencia en la atraktividad percibida o en el número de delitos registrados.

Con esta inclusión se pretende determinar el efecto de las medidas e intervenciones consideradas en los cambios de estado de la atraktividad en un lugar, como también en el número de crímenes ahí registrados. Lo anterior requiere clasificar estas intervenciones entre aquellas que tienen un impacto sólo en el corto plazo y las que, además, afectan el sistema en el mediano plazo.

Una intervención se dirá de corto plazo cuando no tenga repercusiones para el sistema en el siguiente período y, por tanto, tiene incidencia solo en las probabilidades de emisión del delito. Al ser las condiciones mejores, o peores, para delinquir durante ese período en particular, se verá un efecto en el número de delitos observados, pero que no repercutirá mayormente en el proceso de actualización de la atraktividad. A modo de ejemplo, se puede suponer que en un sector medianamente atractivo se organiza un evento musical masivo, durante ese día se observa un aumento sustancial de hurtos, en relación al promedio observado en los períodos previos, debido al gran número de personas circulando en el lugar, que representa mayores oportunidades para robar desapercibidamente. Sin embargo, es de conocimiento general que ese evento dura solo un día y que el flujo de personas volverá a las magnitudes promedios conocidas al día siguiente. Este evento no tiene un efecto en el proceso de actualización de la atraktividad percibida, pero sí tiene un efecto sobre la probabilidad de observar un número determinado de crímenes durante el período.

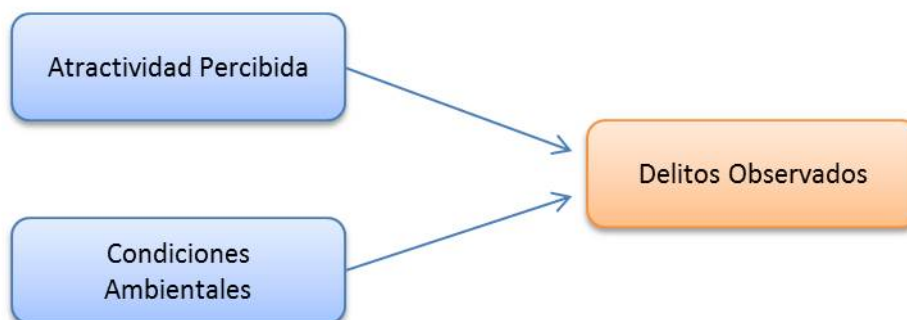


Figura 4.2: Relación entre la atraktividad percibida, y las condiciones ambientales reinantes, en un período determinado y los delitos observados durante el mismo.

Cuando algún suceso o intervención afecta las expectativas sobre la atraktividad de un

lugar para el siguiente período, se observa un cambio en las probabilidades de transición entre estados. Un lugar determinado, con gran afluencia de personas, considerado muy atractivo entre los delincuentes para robar, es poco probable que cambie su estado de atraktividad a uno bajo, sin intervenciones significantes de por medio. Suponiendo que la municipalidad instala una garita de vigilancia en la esquina, el primer día los delincuentes irán al lugar tal como siempre, pero actualizarán sus expectativas sobre la atraktividad y el siguiente período tal vez escojan otro lugar para delinquir. La probabilidad de que el lugar cambie su estado de atraktividad, a uno menor, aumenta.

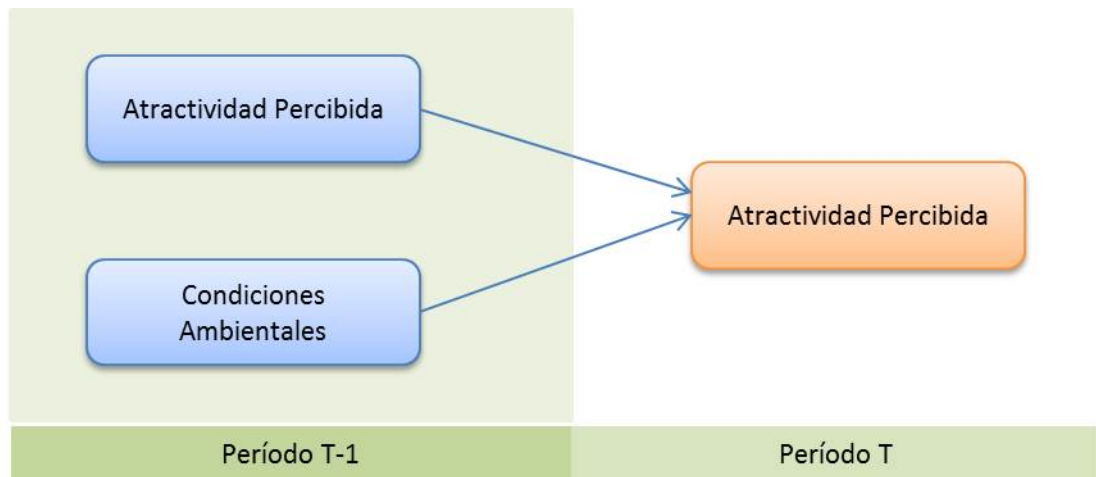


Figura 4.3: Proceso de actualización de la atraktividad percibida a lo largo del tiempo, ante cambios en las condiciones ambientales.

Con estas consideraciones, en un lugar se modelan las probabilidades de transición entre estados considerando una parte observable, a través de las variables incluidas, y otra parte inobservable, que explica los cambios de estados debido al resto de los factores que no se están controlando.

Estas intervenciones en el ambiente, pueden ser de diversa índole, orientadas específicamente a la prevención del crimen o bien relacionadas a cambios estructurales y de hábitos en la ciudad. Se puede estudiar, entonces, el impacto en el crimen que significa la apertura de un centro comercial en determinado barrio, la estructuración de un plan de vigilancia, la reforma y recuperación de áreas verdes o muchas otras. De esta manera, se provee al tomador de decisiones de información respecto a la magnitud y sentido que sus medidas tienen sobre la dinámica del crimen, permitiéndole escoger las más adecuadas según el caso o aquellas que justifiquen mejor el esfuerzo necesario para implementarlas.

El efecto desplazamiento, detallado en la sección 2.4 del Capítulo 2, que a veces ocurre al intervenir un lugar puede también incluirse en el modelo. Esto, considerando variables que describan las intervenciones y condiciones ambientales de lugares cercanos, con características similares u atractores del crimen, en las probabilidades de emisión y transición de un lugar. Considerando el efecto que una medida implementada en un barrio vecino pudiera tener sobre otro barrio en particular, permitiría generar estrategias globales e integradas sobre la ciudad.

4.2.2. Heterogeneidad en el Modelo

Dado que la atraktividad percibida tiene relación con las características físicas y rutinas presentes en un lugar, se esperaría que diferentes lugares tuvieran diferentes comportamientos, tanto en el proceso de actualización de la atraktividad, como en el impacto de diferentes medidas en el mismo.

Se esperaría que una zona residencial tenga una baja probabilidad de pasar a un estado altamente atractivo para robos en la vía pública, debido al bajo número de transeúntes, y posea, por otro lado, una alta probabilidad de permanecer en un estado de baja o mediana atraktividad. Por otro lado, un sector densamente comercial se espera que tenga una alta probabilidad de pasar y permanecer en estados de alta atraktividad, debido a la afluencia de público y la facilidad para pasar desapercibido entre el mismo, y una probabilidad pequeña de pasar a un estado de baja atraktividad.

Por estas razones, y de manera de aislar de forma efectiva la dinámica de la atraktividad en los diferentes lugares de una ciudad, separándola de las posibles diferencias entre lugares, es que se considera heterogeneidad en las probabilidades de transición entre estados. Cada lugar poseerá su propia matriz de transición.

Los estados de atraktividad serán los mismos para cada lugar, y al igual que el efecto de las intervenciones, tanto en las matrices de transición, como en las probabilidades de emisión. De esta manera, dado un mismo conjunto de estados, diferentes lugares tendrán distintas propensiones a permanecer o moverse entre ellos. Distinguiendo el estado en que se encuentran, el efecto de las intervenciones es el mismo para todos los lugares, así como el número esperado de crímenes. Esa formulación permite estudiar una intervención y sus efectos, según el estado en que se encuentre un lugar. Evitándose, de esta forma, el esfuerzo de analizar cada uno de los lugares por sí mismo y permitiendo incorporar estas diferencias.

4.3. Modelo de Markov Oculto para el Crimen

El objetivo del modelo desarrollado en esta sección es proveer información relevante para los tomadores de decisiones, relacionada a la dinámica del crimen y cómo se ve afectada cuando se realiza una intervención en la ciudad. Estos tomadores de decisiones pueden ser tanto organismos públicos, como privados, que pudieran generar impacto con sus medidas en las rutinas y hábitos de una ciudad: ministerios de urbanismo y transporte, departamentos de policías, empresas relacionadas al retail, organizaciones de la comunidad u otros. En específico, el objetivo es estimar y comprender el proceso de actualización de la atraktividad en una ciudad, considerando el efecto de diferentes intervenciones en la misma y además proveer información respecto al efecto que algunas condiciones ambientales específicas tienen en este proceso.

4.3.1. Divisiones Geográficas y Discretización temporal

En lo que sigue de este capítulo, se dividirá la ciudad en una grilla de M celdas para simplificar la notación, pero el modelo puede utilizar cualquier división de la ciudad que considere sectores disjuntos. Un modelo HMM requiere además de una discretización en la dimensión temporal, de manera de definir los períodos en que ocurren los cambios de estado en

el proceso de actualización de la atraktividad de un lugar. El tamaño de estos intervalos pueden ser de días, semanas o meses, según el tipo de crimen que se desea estudiar. En particular el largo de un período debería considerar el tiempo promedio que los delincuentes necesitan para actualizar sus expectativas. Una manera de determinar estos intervalos es analizar gráficamente la distribución de crímenes sobre un mapa de la ciudad y determinar el tiempo que demora para que ocurran cambios en dicha distribución. En lo que sigue de este capítulo se denominará como períodos, a los intervalos discretos de tiempo que conforman el horizonte de evaluación.

4.3.2. Elementos del Modelo

Matriz de Transición

La matriz de transición del modelo representa las probabilidades de cambios en la atraktividad percibida por los delincuentes respecto a algunos lugares. Si se sospecha que la atraktividad de un lugar no suele cambiar durante el tiempo, observándose una distribución similar de crímenes sobre la ciudad período tras período, asumir la matriz de transición como una matriz identidad en el número de estados, I_N , puede ser una buena idea para la especificación del modelo. Esta especificación asume que los lugares se mantienen en un mismo estado de atraktividad indefinidamente. Se define la probabilidad de un lugar i de pertenecer a un estado s , π_{is} , como:

$$\pi_{is} = \frac{\exp(\mu_{is})}{1 + \sum_{k=1}^{N-1} \exp(\mu_{ik})} \quad s = 1, \dots, N-1$$

$$\pi_{iN} = 1 - \sum_{k=1}^{N-1} \pi_{ik}$$

Donde el parámetro μ_{is} es propio de cada celda, reflejando la heterogeneidad entre diferentes lugares.

Por el contrario, si se observan cambios significativos en la distribución de crímenes, considerar probabilidades de transición no nulas entre estados permite estudiar el impacto de las intervenciones y cambios estructurales del lugar que podrían haberlos provocado. Como ya se discutió antes, una transición de un estado de atraktividad a otro se debe a las diferencias entre la percepción previa que poseen los delincuentes sobre el lugar y lo que realmente experimentan en el mismo. Lo anterior podría ser interpretado como que los cambios de estado ocurren cuando estas diferencias superan cierto umbral. Si la experiencia confirma las expectativas de un delincuente, es poco probable que un lugar cambie su estado de atraktividad. Por otro lado, si la experiencia es diametralmente opuesta a las expectativas, un cambio de estado podría ocurrir con mayor probabilidad, resultado del proceso de actualización de estas expectativas: pasando a un estado de mayor atraktividad, si la experiencia fue mejor que la expectativa, o a uno de menor atraktividad, si ocurre lo contrario.

Lo anterior, y asumiendo que los errores son independientes e idénticamente distribuidos como valor extremo, permitiría modelar las probabilidades de transición como un modelo de logit ordenado. Esta forma de modelar preferencias es común en la literatura [17], usado especialmente para casos donde se desea modelar elecciones del consumidor entre un conjunto de categorías ordenadas, en este caso los estados de atraktividad. En este modelo se asume que la variable, en este caso la atraktividad de un lugar, puede ser descrita con el siguiente sistema lineal:

$$s_t^* = \beta a_t + \varepsilon$$

Donde s_t^* es el valor exacto de la atractividad del lugar si pudiera observarse, considerando su valor continuo, a_t corresponde al vector de variables describiendo las intervenciones en el lugar y ε representa el error de la estimación, o bien la parte no observable de la atractividad. El valor exacto de la variable s_t^* , la atractividad, es inobservable en este caso, sin embargo, pueden detectarse categorías de respuesta:

$$s_t = \begin{cases} 1 & \text{si } s_t^* \leq \mu_1 \\ 2 & \text{si } \mu_1 < s_t^* \leq \mu_2 \\ \vdots & \\ N & \text{si } \mu_{N-1} < s_t^* \end{cases}$$

En este caso, donde las categorías estarían definidas por los estados de atractividad, si el efecto de las intervenciones y los factores inobservables durante un período es suficiente para superar alguno de los umbrales, se puede observar una transición a algún otro estado. Estos umbrales μ_s determinan los rangos de atractividad, en su valor continuo, que definen los estados. Si la atractividad percibida en un momento t , dadas las intervenciones, $s_t^* = \beta a_t + \varepsilon$, esta entre el rango determinado por los umbrales de un estado, el lugar cambia a aquel estado de atractividad. Dado que se debe considerar la información y aprendizaje de períodos anteriores, resumidos en el estado de atractividad en que se encuentra el lugar, estos umbrales μ_s dependen del estado en que se encuentre el sistema.

Por ejemplo, si los delincuentes consideran que un lugar se encuentra en el estado de baja atractividad, debido a su experiencia previa, las condiciones ambientales y la experiencia actual deben ser significativamente superiores para que cambien sus expectativas a un estado de atractividad mayor. Al contrario, una experiencia ligeramente desfavorable podría tener un gran impacto, bajando las expectativas, en un lugar que se consideraba en un estado de atractividad alto. Los umbrales por tanto, son diferentes según el estado de atractividad en que se encuentre un lugar. De esta forma, las probabilidades de transición, en caso de encontrarse en un estado s , quedan determinadas por el siguiente sistema de ecuaciones.

$$\begin{aligned} q_{tis1} &= \frac{\exp(\mu_{is1} - a'_{it}\rho_{is})}{1 + \exp(\mu_{is1} - a'_{it}\rho_{is})} \\ q_{tisr} &= \frac{\exp(\mu_{isr} - a'_{it}\rho_{is})}{1 + \exp(\mu_{isr} - a'_{it}\rho_{is})} - \frac{\exp(\mu_{is(r-1)} - a'_{it}\rho_{is})}{1 + \exp(\mu_{is(r-1)} - a'_{it}\rho_{is})} \quad \forall r = 2, \dots, N-1 \\ q_{tisN} &= 1 - \frac{\exp(\mu_{is(N-1)} - a'_{it}\rho_{is})}{1 + \exp(\mu_{is(N-1)} - a'_{it}\rho_{is})} \end{aligned}$$

Donde q_{tisr} es la probabilidad de observar una transición, en la celda i durante el período t , desde el estado s al r , dado el efecto de las variables describiendo las intervenciones a_{it} . Los parámetros μ_{isr} representan los valores umbrales del modelo de logit ordenado para una celda en estado s , y el vector de parámetros ρ_{is} captura el efecto de las variables a_{it} en dicho estado. En este modelo se impone que:

$$\mu_{is1} \leq \mu_{is2} \leq \dots \leq \mu_{is(N-1)}$$

Como se menciona en la Sección 4.2.2, es conveniente considerar heterogeneidad en las matrices de transición, de manera de aislar la dinámica del proceso de actualización de la atractividad, sin posibles distorsiones debido a las diferencias entre lugares. Para incluir

heterogeneidad en las transiciones, se considerará que los umbrales que determinan las probabilidades de transición son propios de cada lugar.

Así, si un lugar destinado a fines comerciales, con alta afluencia de público y turistas, que usualmente presenta múltiples oportunidades para cometer un hurto, puede poseer mayor probabilidad de permanecer, o pasar, en estados de mayor atraktividad, que una zona residencial que suele permanecer en estados de atraktividad bajos.

Estas diferencias se intentarán representar mediante variables que describan las características físicas y las rutinas que se desarrollan en cada lugar, utilizando la siguiente representación:

$$\tilde{\mu}_{isr} = \delta_{sr}' Z_i + \varepsilon_i$$

Para imponer valores crecientes en los umbrales, éstos se calculan mediante la siguiente formulación:

$$\begin{aligned} \mu_{i1r} &= \tilde{\mu}_{i1r} \\ \mu_{isr} &= \tilde{\mu}_{isr} + \sum_{r=2}^s \exp(\tilde{\mu}_{isr}) \quad \forall s = 2, 3, \dots, N-1 \end{aligned}$$

Donde el vector de variables $Z_i \in \mathbb{R}^{nz}$ representa las características físicas de la celda i , como podría ser el uso de sueldo, si contiene un centro comercial o una estación de metro, la distancia a vías principales o puntos relevantes de transporte, o cualquier otra particularidad que podría convertir a un lugar en un generador o atraktor del crimen. El vector de parámetros $\delta_{sr} \in \mathbb{R}^{nz}$ captura el efecto de estas características en los umbrales de transición a cada estado, considerando además un efecto aleatorio que da cuenta de información no observable que también podría afectarlos, $\varepsilon_i \sim N(0, \Sigma_\theta)$.

Probabilidades de Emisión

Como ya se argumentó en la Sección 4.2 de este capítulo, la variable observada corresponde al número de crímenes registrados en una celda a lo largo de un período, Y_{it} . Se asumirá que el número de crímenes registrados en un período se distribuye como una Poisson, un supuesto común al intentar predecir cantidades discretas y, en particular, utilizado para estudiar las tasas de crímenes en la literatura [17, 24, 34]. Esta variable, se asumirá condicionalmente independiente dado el estado de atraktividad s en el que se encuentre el lugar, determinando el valor de la media de la distribución λ_s .

$$Y_{it|s_t} \sim \text{Poisson}(\lambda_s)$$

Es probable, según el tipo de crimen y la data con la que se trabaje este modelo, que existan lugares donde rara vez se producen delitos. Esta característica puede producir una sobre dispersión de los datos y el mal ajuste de un modelo de Poisson, que no tiene dicho comportamiento. En estos casos, la utilización de un modelo de Poisson cero-inflado es una alternativa común en la literatura [14, 19, 26], de manera de capturar el hecho de que existen lugares donde no es frecuente que se registren crímenes. La distribución de probabilidad de una Poisson cero-inflada es una distribución mixta, que considera una masa de probabilidad en el valor 0. Con probabilidad α_i un lugar no registrará crímenes durante el período, el resto de las veces seguirá una Poisson dependiente del estado de atraktividad en el que se encuentre.

$$Y_{it|s_t} \sim \begin{cases} 0 & \text{con probabilidad } \alpha_i \\ \text{Poisson}(\lambda_{is_t}) & \text{con probabilidad } (1 - \alpha_i) \end{cases}$$

Dado que no todos los lugares poseen la misma propensión a registrar delitos, es conveniente asumir heterogeneidad en el parámetro α_i , que será propio de cada celda. De forma análoga a la utilizada en los valores umbrales de la matriz de transición en la Sección 4.3.2, se utilizarán las mismas características físicas y rutinas presentes en el lugar ahí mencionadas, para tratar de explicar el valor de este parámetro.

En una formulación similar a la propuesta por Hall [19], se modela el parámetro α_i de la siguiente manera:

$$\hat{\alpha}_i = \delta'_\alpha Z_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$$

$$\alpha_i = \frac{\exp(\hat{\alpha}_i)}{1 + \exp(\hat{\alpha}_i)}$$

En la Sección 4.2.2 se menciona el efecto de intervenciones en un lugar, sobre cambios en los riesgos y oportunidades para delinquir durante el período en que son implementados. Estas intervenciones tienen impacto en el número de crímenes esperado para el período, en otras palabras, un efecto en los parámetros λ_{is_t} . Se incluyen entonces estas variables, describiendo las intervenciones y condiciones ambientales del lugar, en la formulación de la siguiente forma:

$$\log(\lambda_{is_t}) = \bar{\beta}_{0s} + x'_{it} \beta_s$$

En esta representación $\bar{\beta}_{0s}$ es el coeficiente específico del estado s , x_{it} corresponde a las variables que afectan las probabilidades de cometer un delito en el corto plazo y β_s es el vector de coeficientes que captura su efecto en una celda en estado s .

Para asegurar la identificación de los estados, evitando el problema conocido como *switching labels* (intercambio de etiquetas) [42] al estimar el modelo, se impone que el número esperado de delitos sea no decreciente en los estados. Es decir, se busca que una celda en un estado de atractividad dado posea un número esperado de delitos mayor que cuando se encuentra en un estado de atractividad menor. Para lograrlo, se centran las variables independientes x_{it} en el promedio y se impone que $0 \leq \bar{\beta}_{01} \leq \bar{\beta}_{02} \leq \dots \leq \bar{\beta}_{0N}$. Con esta construcción, λ_{is_t} toma el valor del intercepto $\bar{\beta}_{0s}$, cada vez que las variables independientes se encuentran en su valor promedio.

Una variable \tilde{x}_{it} se centra en el promedio mediante la siguiente construcción.

$$\tilde{x}_{it} = x_{it} - \bar{x} \quad \bar{x} = \frac{1}{m \times T} \sum_{i,t} x_{it}$$

Para lograr interceptos crecientes en los estados se impone que:

$$\bar{\beta}_{01} = \beta_{01}$$

$$\bar{\beta}_{0s} = \beta_{01} + \sum_{r=2}^s \exp(\beta_{0r}) \quad \forall s = 2, 3, \dots, N$$

Distribución de Probabilidad Inicial

Para estimar la distribución de probabilidades inicial se utiliza el método sugerido por Netzer [32], resolviendo el siguiente sistema de ecuaciones para cada una de las celdas.

$$\begin{aligned}\pi_i \bar{Q}_i &= \pi_i \\ \sum_{s=1}^N \pi_{is} &= 1\end{aligned}$$

Donde \bar{Q}_i es la matriz de transición considerando todas las variables independientes fijas en su valor promedio, entre todos los individuos y períodos. Lo anterior, debido a que esta matriz va cambiando al depender de variables dinámicas en el tiempo. En un modelo HMM tradicional, la matriz de transición es estática y las probabilidades iniciales corresponden a las probabilidades estacionarias del sistema. Al ser la matriz de transición propia de cada lugar, las probabilidades iniciales son también intrínsecas de la celda.

4.4. Los Tres Problemas del Modelo HMM para el Crimen

Siguiendo la estructura utilizada en la Sección 3.2.4 del Capítulo 3, se desarrolla a continuación la solución para los tres problemas clásicos de los modelos HMM aplicado al modelo para el crimen desarrollado en este capítulo. Se explicita la medida de ajuste, específicamente la verosimilitud, para esta aplicación en particular y la metodología propuesta para la estimación de los parámetros utilizando samplers MCMC, que fueron también descritos en el Capítulo 3. Finalmente se describe el mecanismo utilizado para predecir el número de crímenes del siguiente período, dada las intervenciones.

4.4.1. Problema 1: Definir una Medida de Ajuste

Utilizando el desarrollo de la Sección 3.2.4 del Capítulo 3, la medida de ajuste corresponde a la verosimilitud del modelo, que en un HMM viene dado por la probabilidad de observar una secuencia de la variable observada condicional a una secuencia de estados oculta, sobre todas las posibles secuencias de estados. La verosimilitud de una secuencia de T observaciones consecutivas, en una determinada celda i , dados los parámetros, corresponde a:

$$\begin{aligned}L_{iT} &= P(Y_{i1} = y_{i1}, \dots, Y_{iT} = y_{iT}) \\ &= \sum_{s_1=1}^N \sum_{s_2=1}^N \dots \sum_{s_T=1}^N \left[\pi_{is_1} \prod_{\tau=2}^T q_{is_\tau s_{\tau-1}} \cdot \prod_{t=1}^T \left((1 - \alpha_i) \frac{\exp(-\lambda_{is_t}) \lambda_{is_t}^{y_{it}}}{y_{it}!} + \alpha_i \mathbb{1}_{y_{it}=0} \right) \right]\end{aligned}$$

Y utilizando la representación matricial propuesta por [28], desarrollada también en la Sección 3.2.4, la verosimilitud en una celda queda expresada como:

$$L_{iT} = \pi_i \tilde{m}_{i1} Q_{i,1 \rightarrow 2} \tilde{m}_{i2} \dots Q_{i,T-1 \rightarrow T} \tilde{m}_{iT} \mathbf{1}'$$

En donde $\tilde{m}_{it|s} = \left((1 - \alpha_i) \frac{\exp(-\lambda_{is}) \lambda_{is}^{y_{it}}}{y_{it}!} + \alpha_i \mathbb{1}_{y_{it}=0} \right)$ y \tilde{m}_{it} es una matriz diagonal en $\mathbb{R}^{N \times N}$ con los elementos de $\tilde{m}_{it|s}$ en la diagonal, finalmente $\mathbf{1}'$ es un vector de unos en \mathbb{R}^N .

La verosimilitud del modelo, considerando todas las celdas de la ciudad, es entonces:

$$\ell(\theta) = \prod_{i=1}^M L_{iT}$$

Que también puede ser expresado en función de las log-verosimilitudes de las celdas:

$$\log[\ell(\theta)] = \sum_{i=1}^M \log[L_{iT}]$$

Los valores de la verosimilitud pueden ser muy pequeños, tanto así, que un ordenador ser incapaz de manejar números de tales magnitudes, aproximándolos al valor 0. Para evitar este fenómeno, conocido como *under flow*, se utiliza un mecanismo de escalamiento, que permite manejar computacionalmente los exiguos valores de la función de verosimilitud, propuesto por MacDonald y Zucchini [28] (revisar la sección A.1 del Apéndice A para más detalle).

4.4.2. Problema 2: Maximizar la Medida de Ajuste

En esta sección se presenta la metodología de estimación basada en un método MCMC, cuya teoría fue revisada en la Sección 3.4, para la estimación jerárquica Bayesiana de los parámetros [32] que permita maximizar la verosimilitud del modelo.

En el modelo desarrollado en la Sección 4.3 se tienen dos tipo de parámetros, aquellos que poseen un mismo valor para todas las celdas, y otros utilizados para incluir heterogeneidad, cuyos valores son propios de cada celda, lugar. Para su estimación, se utiliza el procedimiento propuesto por Rossi [38], donde se separan estos parámetros en dos grupos: aquellos que tienen el mismo valor para todos las celdas, Ψ de efecto fijo, y aquellos parámetros propios de cada lugar, θ_i de efecto aleatorio.

$$\theta_i = (\alpha_i, \mu_{i1s}, \dots, \mu_{i(N-1)s} \quad s = 1, \dots, N)$$

$$\Psi = (\rho_1, \rho_2, \dots, \rho_N, \beta_{01}, \beta_{02}, \dots, \beta_{0N}, \beta_1, \beta_1, \dots, \beta_N)$$

El proceso de estimación jerárquica, para estimar los parámetros y verosimilitud, considera un muestreo de Gibbs, revisado en la Sección A.3 del Apéndice, en el que se estima: la verosimilitud dado los parámetros, los parámetros de efecto aleatorio dado los de efecto fijo y posteriormente los de efecto fijo dado los parámetros de efecto aleatorio.

Verosimilitud:	$p(y_i \theta_i, \Psi)$
Priors de Primera Etapa:	$p(\theta_i \Psi)$
Priors de Segunda Etapa:	$p(\Psi \theta_i)$

Si siguiendo esta secuencia se van realizando muestreos de los parámetros a partir de las distribuciones condicionales a posteriori, que se van actualizando en cada iteración. Las distribuciones condicionales a posteriori de θ y Ψ no tienen forma cerrada, por lo que no se pueden obtener analíticamente, y se utiliza, para poder muestrear a partir de estas, el algoritmo Metropolis, también revisado en la Sección 3.4 del Capítulo 3.

El primer paso de estimación corresponde a muestrear el vector de parámetros θ_i de efecto aleatorio, usando el algoritmo de Metropolis. Una vez obtenidos los valores de θ_i para cada

celda, se asume un modelo aparentemente no relacionado, revisado en el Apéndice A.3.1, donde θ es la variable a explicar y las características de la celda Z_i son las variables independientes. Utilizando un sampleo de Gibbs se estima el vector de regresores y la matriz de covarianzas, δ y Σ_θ .

Finalmente se samplean los parámetros de efecto fijo Φ a través del algoritmo de Metropolis. La secuencia del algoritmo MCMC utilizado corresponde entonces a:

1. **Generar** $\{\theta_i\} | Y_i, X_i, a_i, Z_i, \Psi, \delta, \Sigma_\theta$
2. **Generar** $\delta | \{\theta_i\}, Z, \Sigma_\theta$
3. **Generar** $\Sigma_\theta | \{\theta_i\}, Z, \delta$
4. **Generar** $\Psi | Y, X, a, \{\theta_i\}$

En el Apéndice B.1 se puede estudiar el algoritmo utilizado con mayor detalle.

Elección del Número de Estados

La elección del número de estados es una decisión no trivial cuando se estima un Modelo de Markov Oculto y, muchas veces un proceso crítico en el ajuste y éxito del modelo. Esta decisión implica estimar el modelo con diferente número de estados y escoger aquel que mejor explique los datos observados.

El principio de Occam es, en general, el principio que guía la selección de modelos, donde un modelo debe ser lo suficientemente simple para la eficiencia computacional y lo suficientemente complejo como para poder capturar las particularidades de los datos. Normalmente un modelo demasiado complejo tiende a caer en el sobreajuste sobre los datos y uno demasiado simple no es capaz de explicar correctamente las características y particularidades de los mismos. El criterio de selección por lo general posee dos partes: un término que represente el ajuste del modelo y otro que caracterice la complejidad del mismo [4].

En este caso, desde un enfoque Bayesiano, es posible definir un conjunto de M_1, M_2, \dots, M_k modelos y calcular la probabilidad a posteriori de cada uno de ellos. La probabilidad a posteriori de un modelo se puede obtener a través de la verosimilitud de la data y la probabilidad a priori del modelo de la siguiente manera [37]:

$$p(M_i|y) = \frac{p(y|M_i)p(M_i)}{p(y)}$$

La probabilidad de la data, $p(y)$, es usualmente desconocida por lo que el ratio de probabilidades a posteriori es una alternativa para comparar y elegir entre dos modelos.

$$\frac{p(M_1|y)}{p(M_2|y)} = \frac{p(y|M_1)p(M_1)}{p(y|M_2)p(M_2)} = \frac{p(y|M_1)}{p(y|M_2)} \times \frac{p(M_1)}{p(M_2)}$$

Donde el primer término del lado derecho de la ecuación corresponde al Factor de Bayes y el segundo término corresponde a la probabilidad del modelo dada la estructura y distribuciones a priori.

En el Apéndice A en la Sección B.2 se detallan diferentes estadísticos y ratios utilizados para escoger el modelo más adecuado para explicar los datos, en este caso particular, decidir sobre el número de estados del sistema.

4.4.3. Problema 3: Predecir a Partir del Modelo

Obtenidas las distribuciones a posteriori de todos los parámetros del modelo a través de los valores obtenidos en los sampleos del algoritmo MCMC utilizado. Se calculan los valores esperados a posteriori de las matrices de transición, las tasas de crímenes en cada estado, la probabilidad α_i de no observar crímenes y las probabilidades iniciales, para cada celda y cada período, incluyendo el período a predecir $t = T + 1$:

$$Q_{it} = \frac{1}{R} \sum_{r=1}^R Q_{it}^r \quad \forall t = 1, \dots, T, \quad \forall i = 1, \dots, M$$

$$\lambda_{ist} = \frac{1}{R} \sum_{r=1}^R \lambda_{ist}^r \quad \forall t = 1, \dots, T + 1, \forall s \in S, \quad \forall i = 1, \dots, M$$

$$\alpha_i = \frac{1}{R} \sum_{r=1}^R \alpha_i^r \quad \forall i = 1, \dots, M$$

$$\pi_i = \frac{1}{R} \sum_{r=1}^R \pi_i^r \quad \forall i = 1, \dots, M$$

Donde Q_{it}^r es la matriz de transición de la celda i en el período t , utilizando el valor de los parámetros de la iteración r del sampleo MCMC de estimación, considerando que dicho sampleo contiene R iteraciones. Análogo para el resto de los componentes del modelo.

Teniendo estos valores, se calcula para cada celda la probabilidad de encontrarse en cada estado durante el último período sobre el que se tienen datos, dado el historial de crímenes hasta dicho período. Para esto, se utiliza la fórmula desarrollada en la Sección 3.2.4.

$$p^t = \frac{\pi \tilde{m}_1 Q_1 \tilde{m}_2 \dots Q_{t-1} \tilde{m}_t}{P(Y|\lambda)} = \frac{\pi \tilde{m}_1 Q_1 \tilde{m}_2 \dots Q_{t-1} \tilde{m}_t}{L(\lambda)_t}$$

Donde

$$L(\lambda)_t = \pi \tilde{m}_1 Q_1 \tilde{m}_2 \dots Q_{t-1} \tilde{m}_t 1'$$

Teniendo este vector de probabilidades, se puede calcular para cada celda la probabilidad de encontrarse en cada estado durante el período a predecir.

$$p^{(T+1)} = p^T Q_{iT}$$

Con estos valores se pueden calcular los límites del intervalo de predicción. Estos límites se definen de manera de que el número de crímenes a predecir se encuentre con un 90% de probabilidad dentro del intervalo, $P(Lim_{inf} \leq Y_{i(T+1)} \leq Lim_{sup}) = 0.9$, asumiendo que el número de crímenes se distribuye como la Poisson cero-inflada estimada en el modelo.

$$p_{inf} = \text{máx}\{0.05 - \alpha_i, 0\}$$

$$p_{sup} = \text{máx}\{0.95 - \alpha_i, 0\}$$

$$Lim_{inf} = \text{mín}\{x \in \mathbb{N} : \sum_{s=1}^N \sum_{k=0}^x P(Y_{it} = k | S_t = s) P(S_{it} | Y_{i1}, Y_{i2}, \dots, Y_{i(t-1)}) \geq \frac{p_{inf}}{1 - \alpha_i}\}$$

$$Lim_{sup} = \min\{x \in \mathbb{N} : \sum_{s=1}^N \sum_{k=0}^x P(Y_{it} = k | S_t = s) P(S_{it} | Y_{i1}, Y_{i2}, \dots, Y_{i(t-1)}) \geq \frac{P_{sup}}{1 - \alpha_i}\}$$

Donde las probabilidades de observar el número de crímenes, dado un estado y que este número se distribuye como una Poisson, corresponden a:

$$P(Y_{it} = k | S_t = s) = \frac{\exp(-\lambda_{is(T+1)}) \lambda_{is(T+1)}^k}{k!}$$

Capítulo 5

Implementación

5.1. Introducción

El modelo presentado en el Capítulo 4 requiere de data sensible y de difícil acceso, como los registros de delitos denunciados, las coordenadas de dichos eventos, fecha y hora de ocurrencia, así como registros respecto a intervenciones policiales o de otro tipo, en el intervalo de tiempo a evaluar. Lo anterior puede significar una barrera sustancial para su implementación, o mirándolo desde otro punto de vista, una oportunidad para colaborar con las instituciones que posean o recolecten dichos datos, y que finalmente son los que pondrían en práctica los aprendizajes y conclusiones obtenidos a partir de un estudio de este estilo.

El objetivo de este capítulo es proveer un caso de estudio que permita demostrar el potencial y desempeño del modelo HMM para el crimen, desarrollado en este trabajo. Dado que no se cuenta con datos reales para realizar este experimento, se utilizarán datos obtenidos desde el simulador del crimen CRIMSIM [13], basado en agentes, en base a la estructura del centro de la ciudad de Santiago de Chile. Los parámetros, sin embargo, que definen el comportamiento de los agentes del simulador, no se calibran para emular los de la ciudad real. La ciudad simulada es en esencia una ciudad ficticia, y se denominará en lo que sigue como ciudad virtual. Los datos obtenidos a partir del simulador son utilizados para implementar y estimar el modelo HMM para el crimen desarrollado en el Capítulo 4.

Es importante destacar que el uso de datos simulados se justifica por la ausencia de datos reales para este propósito. En caso de tener acceso a estos datos, se podría replicar esta metodología sin mayores cambios en su procedimiento, ya que los datos y registros obtenidos del simulador se encuentran al alcance de la policía u otras instituciones gubernamentales.

De manera de darle un tratamiento similar al que se le daría a una ciudad real, en la Sección 5.2 se realiza un análisis y descripción de la ciudad de manera de tener una idea general de su dinámica y estructura. Con el objetivo de explicitar el tipo de registros necesarios para implementar el modelo HMM para el crimen, en la Sección 5.3 se describen los datos e información que se obtienen del simulador, en relacionados a los crímenes y la vigilancia registrados durante la simulación. Estos registros son similares a lo que podría obtener la policía mediante las denuncias de los delitos y los registros sobre la vigilancia realizada en la ciudad. La transformación y manejo de estos datos para generar las variables requeridas por el modelo, como lo son la secuencia de crímenes y las variables describiendo las condiciones ambientales, es detallada en la Sección 5.4.2. Finalmente la Sección 5.5 da cuenta de los detalles del proceso

de estimación del modelo, como la elección del número de estados.

5.2. Descripción de la Ciudad Virtual

La simulación realizada para la obtención de datos considera 329 días de la ciudad virtual, equivalente a 47 semanas. El tipo de crimen estudiado corresponde al de hurtos en la vía pública, un crimen favorecido por las multitudes y lugares donde es fácil pasar desapercibido y escapar una vez cometido el delito. Los agentes (ciudadanos, delincuentes y policías) circulan aleatoriamente, siguiendo sus rutinas y actividades. La forma en que se mueven, así como desde dónde y hacia dónde van los agentes tiene fuerte relación con las características estructurales de la ciudad.

Para definir el tamaño de las unidades de estudio y los intervalos temporales es necesario estudiar estas características, así como el comportamiento de los agentes. En esta sección se procede a describir a rasgos generales la dinámica observada.

5.2.1. Características Estructurales

La ciudad virtual donde se realiza la simulación tiene la estructura del centro de la ciudad de Santiago de Chile, que puede ser observada, junto al uso de suelo de cada zona, en la Figura 5.1. El área estudiada está destinada principalmente a infraestructura de uso comercial, que se intensifica en centro mismo del mapa en base a dos ejes perpendiculares, en un área calificada como súper comercial. Existen también algunos barrios estudiantiles desperdigados por la periferia de la ciudad. Un parque y un río delimitan superiormente la ciudad, y una gran avenida la parte inferior.



Figura 5.1: Uso de suelo de la ciudad virtual.

La clasificación de uso de suelo, consistente en tres categorías, se define según las actividades destinadas a los diferentes edificios y áreas que componen un determinado sector.

En la Tabla 5.1 se pueden apreciar las probabilidades de observar algún tipo de edificación en cada una de las categorías de uso de suelo.

Edificios	Categorías Uso de Suelo		
	Comercial	Súper Comercial	Estudiantil
Residencias	3 %	0 %	2 %
Escuelas	2 %	0 %	55 %
Bancos	10 %	10 %	3 %
Transporte	25 %	20 %	15 %
Oficinas	20 %	20 %	10 %
Comercio	40 %	50 %	15 %

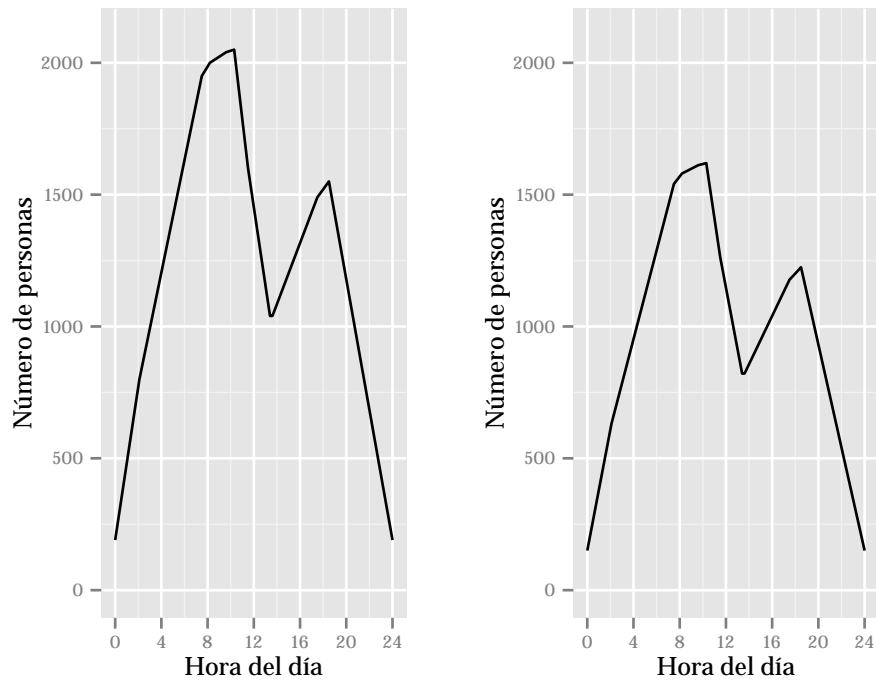
Tabla 5.1: Composición de cada categoría de uso de suelo, en cuanto a la infraestructura presente en la zona.

5.2.2. Descripción de los Agentes

La simulación realizada considera una población permanente de 40 delincuentes y 20 efectivos policiales. El número de ciudadanos va cambiando en el tiempo, en la medida que estos aparecen y desaparecen de las calles de la ciudad, cuando entran y salen de los edificios, siguiendo sus rutinas y visitando sus lugares de interés.

Se realizan dos tipos de intervenciones durante el horizonte de simulación, uno de corto y otro de largo plazo. Se destinan diez policías para la intervención de corto plazo, donde los agentes son ubicados en un puesto fijo escogido aleatoriamente dentro de la ciudad. En dichas ubicaciones realizarán una vigilancia permanente durante una semana de duración, tras lo cual, son reubicados al azar en una nueva locación. El mecanismo es similar para los diez policías restantes destinados a las intervenciones de largo plazo, con la diferencia que estos permanecen en su lugar de vigilancia por cuatro semanas, en vez de una, antes de ser relocalizados.

La dinámica de la ciudad virtual, considerando el número de personas circulando por sus calles, contiene dos horas pico que se distinguen durante la mañana alrededor de las 7:00 hrs, cuando la gente se traslada desde su hogar al trabajo, y las 18:00 hrs, cuando vuelven desde el trabajo a su hogar. Durante la noche apenas se observan peatones circulando. Los fines de semana se observan los mismos dos horarios de máxima circulación peatonal, sin embargo el número de peatones es significativamente menor. La figura 5.2 da cuenta de dicho fenómeno durante los días hábiles y los fines de semana.



(a) Agentes circulando en día hábil. (b) Agentes circulando en fin de semana.

Figura 5.2: Número de agentes promedio, circulando en la ciudad virtual según la hora del día, tanto en un día hábil como en un día de fin de semana.

Los delincuentes por otro lado circulan permanentemente por la ciudad, moviéndose aleatoriamente según las oportunidades que les reporta cada lugar, y robando en la medida que puedan localizar y alcanzar a sus víctimas de forma exitosa.

No es el caso de esta aplicación, pero en otro estudio se podría estudiar el crimen en intervalos de un par horas e incluir los horarios punta y valles como variables describiendo las condiciones ambientales, de manera de comprender diferencias en la dinámica delictual respecto a esta característica temporal.

5.3. Datos Obtenidos del Simulador

El simulador del crimen entrega como resultado una serie de archivos que contienen información relacionada a los crímenes cometidos en la ciudad y a la posición de los policías durante la simulación.

La data relacionada a los robos contiene las coordenadas, la hora y el día en que ocurrió cada uno de estos eventos. En cuanto a la data describiendo la vigilancia a la que es sometida la ciudad, consta de las coordenadas del lugar en el que son relocalizados los policías, el tipo de vigilancia que realizan (de corto o largo plazo), así como el día y la hora en que este cambio de posición ocurre.

Evento	Coordenada X	Coordenada Y	Día	Hora	Tipo de Vigilancia
Eventos Delictivos	✓	✓	✓	✓	
Vigilancia Policial	✓	✓	✓	✓	✓

Tabla 5.2: Registros obtenidos en la simulación.

Además, se obtiene una lista con el tipo de estructura física que contiene cada celda interna del simulador, que puede ser: oficina, escuela, comercio, infraestructura de transporte, residencia, parque, río o calle.

Las coordenadas de los eventos delictuales, posiciones de la policía y el tipo de estructura física son reportadas en base a la grilla interna de 91x75 celdas cuadradas que el simulador CRIMSIM utiliza para generar las simulaciones.

Como un análisis preliminar de la data relacionada a los robos, y usando la discretización geográfica que se describe en la Sección 5.4.1, la Figura 5.3 describe la distribución sobre la ciudad virtual de los crímenes registrados durante las 47 semanas de simulación. Del total de 82,367 robos, cerca del 80 % de ellos se concentran en solo 15 celdas, de las 207 que representan la ciudad. Entre estos 15 puntos de alto atractivo, existen cinco de ellos que acumulan más de la mitad del total de robos. El fenómeno de concentración de los crímenes en pocos lugares, descrito en la literatura, es claramente observable en este experimento.

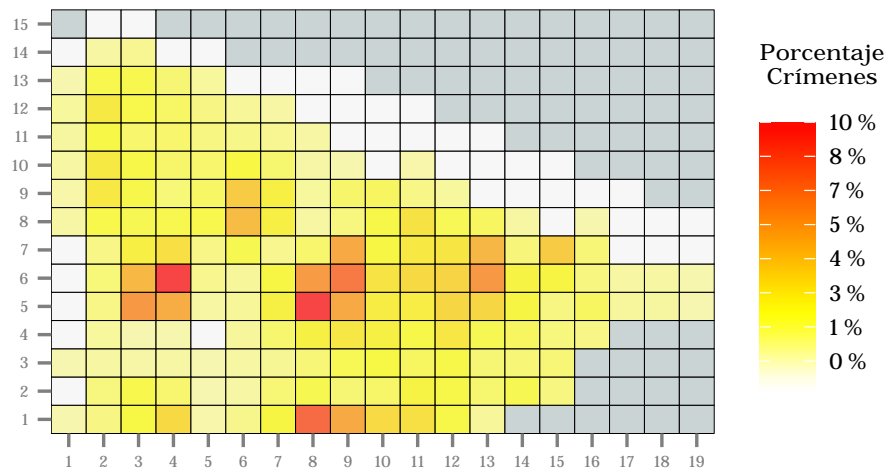


Figura 5.3: Distribución sobre la ciudad virtual del total de los 82,367 robos registrados durante la simulación.

Aplicando el modelo HMM para el crimen, desarrollado en el Capítulo 4, en este caso de estudio se intentará explicar y predecir el número de crímenes sobre la ciudad semana a semana, a partir de la vigilancia policial observada y la estructura física de la ciudad.

5.4. Transformación de Datos

Los datos obtenidos en la simulación son similares a los que podría poseer la policía mediante la denuncia de delitos y el registro de la vigilancia realizada. Tanto en un caso real, como en este caso de estudio, es necesario realizar un tratamiento y transformación de estos datos para obtener los registros en el formato preciso que se requiere para aplicar el modelo HMM para el crimen.

La primera consideración es escoger el tamaño y forma de las unidades de estudio, así como el largo de los intervalos temporales que definirán los cambios de estado de atractividad de las unidades de estudio. En función de lo anterior, se deben generar las secuencias de datos necesarias para el modelo. En particular, para cada unidad de estudio se requiere la siguiente información:

1. **Secuencia Observable (Y_i):** Secuencia del número de robos, agregados según el intervalo temporal escogido, registrados en la celda durante el período de simulación.
2. **Condiciones Ambientales Corto Plazo (x_i):** Secuencias de los atributos describiendo las condiciones ambientales durante cada intervalo temporal, con efecto en el corto plazo.
3. **Condiciones Ambientales Largo Plazo (a_i):** Secuencias de los atributos describiendo las condiciones ambientales de cada semana, con efecto en el largo plazo.
4. **Características Estructurales (Z_i):** Set de atributos de la unidad de estudio describiendo sus características estructurales.

En las siguientes secciones se describe el manejo y transformación de la data disponible obtenida del simulador, para obtener las variables y secuencias requeridas para la implementación del modelo HMM.

5.4.1. Discretización del Espacio Físico y Temporal

La generación de datos y variables para el modelo requiere primeramente la sectorización de la ciudad y la elección de los intervalos de tiempo que definirán los períodos del modelo. Es conveniente dividir la ciudad en sectores de áreas de similar tamaño, de manera que el número de crímenes observados puedan ser comparables entre un sector y otro, y que las diferencias existentes puedan ser explicadas por las variables incluidas en el modelo y no simplemente por la diferencia del tamaño de las áreas.

Para determinar cualquier división espacial sobre la ciudad se requiere conocer primero una medida de distancia que permita medirla. Se define una medida de distancia de acuerdo a esta grilla interna del simulador, para describir las distancias y áreas en la ciudad, donde:

Definición: Se define la unidad de distancia *celda simulador* [c.s] como el ancho de una celda cuadrada del simulador.

Tal como se menciona en la Sección 4.2, el tamaño de la unidad de estudio se debe definir según el enfoque que se le quiere dar al estudio y al alcance estimado de las intervenciones

incluidas en la investigación. En este caso, las intervenciones corresponden a la vigilancia realizada por los policías ubicados aleatoriamente dentro de la ciudad.

El radio de visión de la policía, en el simulador, es de 3[c.s], lo que le permite vigilar de forma directa lo que ocurre a lo largo de una cuadra si está ubicado en el centro de la misma (la longitud promedio de las cuadras de la ciudad, que corresponde a la distancia entre dos intersecciones de calles consecutivas, es de aproximadamente 5[c.s]). Parece natural escoger un tamaño de celda tal que permita capturar el efecto directo de la vigilancia policial, es decir el área efectivamente vigilada por un policía. En un caso real, el radio de visión de la policía pudiera ser difícil de obtener de forma tan directa, pero se puede estimar aproximando la distancia efectiva que un policía puede observar desde una posición determinada, y no menos importante, la máxima distancia desde la cual es visible para los delincuentes.

Por esta razón se divide la ciudad en una grilla de 19x15 celdas, donde cada una de estas celdas tiene un ancho promedio de 5[c.s], la Figura 5.4 muestra la división realizada sobre la ciudad. Cada celda corresponderá a una unidad de estudio en la aplicación.

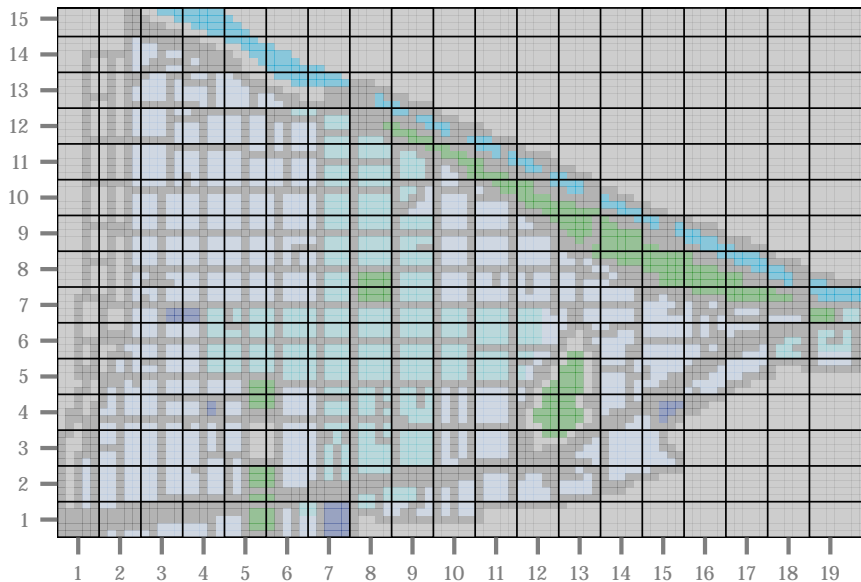


Figura 5.4: División de la ciudad en una grilla de 19x15.

Se puede notar de la Figura 4.2 que la ciudad ocupa parcialmente la grilla propuesta, por lo que se requiere identificar las celdas que son relevantes para el problema y dejar fuera del análisis al resto. Se consideran en el estudio solo las celdas que contienen alguna estructura de la ciudad, o bien, donde los agentes puedan circular: edificaciones, pavimento o parque. La Figura 5.5 muestra las 207 celdas, en blanco, que serán estudiadas y analizadas efectivamente en este capítulo.

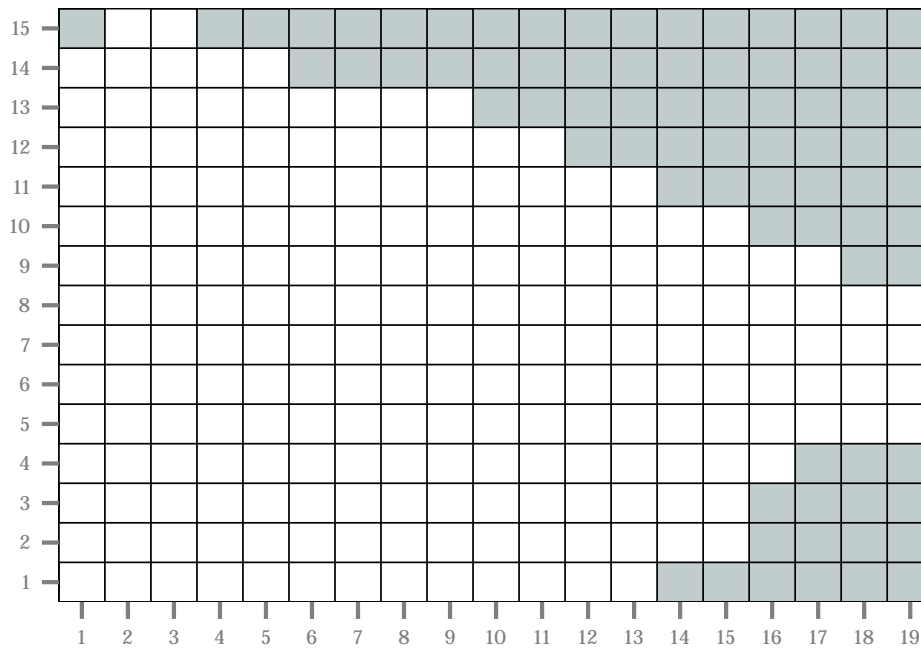


Figura 5.5: Selección de celdas relevantes para el análisis. Las celdas blancas son seleccionadas para el estudio.

Por otro lado, y refiriéndose a la discretización temporal, los intervalos deben ser, dentro de lo posible, homogéneos, de manera que la dinámica observada dentro de los mismos sea similar entre un período y el siguiente. Por esta razón se escoge la semana como intervalo temporal, de manera de poder incluir tanto la dinámica urbana de días hábiles, como la de los fines de semana, en un mismo período. Una alternativa pudiera haber sido considerar solo días hábiles en el análisis, u homológamente, solo fines de semanas, sin embargo dado que las estrategias policiales, descritas en la Sección 5.2.2, están estructuradas semanalmente y la experiencia de los delincuentes es continua en el tiempo, incluyendo lo que ocurre durante los fines de semanas, utilizar la semana completa como intervalo temporal parece adecuado para este caso.

5.4.2. Generación de las Secuencias de Datos

Es importante recordar que las coordenadas entregadas por el simulador están basada en su grilla interna, que es distinta y más fina que la discretización espacial escogida para llevar a cabo el análisis. Es preciso entonces, agregar la información al nivel de las unidades de estudio e intervalos temporales escogidos. De manera de formalizar la agregación de la data se definen algunos conjuntos y subconjuntos:

Definición: Denominando A al conjunto de celdas j escogidas como unidades de estudio, caracterizadas en la Figura 5.5, y B el conjunto de celdas de la grilla interna del simulador, se define $B(j) \subset B$ como el conjunto de celdas de B contenidas en la unidad de estudio j .

Utilizando la notación anterior, se describe a continuación el procedimiento realizado para construir las series de datos necesarias para estimar el modelo HMM para el crimen.

Serie Observable: Secuencia de Robos

La secuencia de robos requiere agregar los delitos registrados en cada unidad de estudio semanalmente. Para cada unidad de estudio j , se suman los robos registrados en las celdas $i \in B(j)$ durante cada semana, generando una secuencia de 47 observaciones.

Condiciones Ambientales Corto Plazo: Vigilancia Policial

Uno de los principales objetivos de este caso de estudio es analizar el efecto de la vigilancia policial en el número de crímenes registrados en un lugar. Este efecto podría analizarse desde dos perspectivas, la primera sería enfocar el estudio en el área que se encuentra directamente vigilada, es decir dentro del radio de visión de un policía. Un segundo enfoque, sería analizar el efecto desplazamiento que la vigilancia policial podría provocar, estudiando las unidades de estudio que se encuentran alrededor del área directamente vigilada, que corresponde a una de las posibles opciones de desplazamientos enunciadas en la Sección 2.4.3.

De forma de considerar ambos efectos, se incluyen dos atributos describiendo las condiciones ambientales con efecto en el corto plazo, que serán incluidas en las probabilidades de emisión:

- **Vigilancia Directa (VD_{jt}):** Este atributo considera la vigilancia policial directa observada durante cada semana t en la unidad de estudio j .
- **Vigilancia Indirecta (VI_{jt}):** Este atributo mide la vigilancia observada en lugares cercanos a la unidad de estudio j durante cada semana t .

En el simulador, tal como se explicó anteriormente, los policías tienen un radio de visión de 3[c.s], y por tanto todas las celdas de la grilla interna del simulador que se encuentren dentro del radio están siendo vigiladas de forma directa. En algunos casos una unidad de estudio pudiera estar siendo vigilada solo parcialmente, y parece correcto asumir que mientras menos área de la celda se encuentre vigilada, menor será el efecto de la vigilancia en el número de crímenes observados en ese lugar. De manera de caracterizar las condiciones ambientales de forma diferenciada, cuando se encuentre completa o parcialmente vigilada, se contabilizará el área de la unidad de estudio que se encuentra efectivamente vigilada por la policía.

Por otro lado, la estrategia de vigilancia implementada durante la simulación no restringe la ubicación de los policías de ninguna manera, por lo que una unidad de estudio j podría estar siendo vigilada por más de un policía simultáneamente. Asumiendo que un lugar con más policías vigilando un sector presenta, para un delincuente, más riesgos de ser sorprendido, es relevante considerar el número de policías que vigila cada punto de la ciudad en la descripción de las condiciones ambientales.

En el Apéndice C.1 se puede analizar con mayor detalle el tratamiento y cálculos realizados sobre los datos disponibles para obtener la secuencia $x_i = (VD_i, VI_i)$ de las condiciones ambientales con efecto en el corto plazo para cada unidad de estudio.

Condiciones Ambientales Largo Plazo: Vigilancia Policial

Una de las hipótesis que se desea testear sobre la vigilancia policial en la ciudad virtual, es que la estrategia de vigilancia de largo plazo implementada en el simulador tiene incidencia en la percepción de la atractividad en los lugares vigilados. Al confirmar, período a período la experiencia de los delincuentes sobre una unidad de estudio como un lugar constantemente vigilado, se esperaría que el nivel de atractividad cambie de estado en determinado momento. Para tomar esto en consideración se contabiliza para cada celda $i \in B$, el número de períodos seguidos en que la celda ha estado bajo vigilancia. Si la celda ha estado bajo vigilancia policial por al menos cuatro semanas corridas, se considera que dicho lugar se encuentra bajo una estrategia de vigilancia de largo plazo durante ese período.

De forma análoga a la utilizada para las condiciones ambientales con efecto en el corto plazo, se consideran dos atributos representando la vigilancia policial prolongada en cada unidad de estudio, que serán incluidas en las probabilidades de transición entre estados:

- Vigilancia Directa (LD_{jt}): Este atributo considera la vigilancia policial directa, de largo plazo, observada durante cada semana t en la unidad de estudio j .
- Vigilancia Indirecta (LI_{jt}): Este atributo mide la vigilancia de largo plazo observada en lugares cercanos a la unidad de estudio j durante cada semana t .

En el Apéndice C.1 se puede analizar con mayor detalle el tratamiento y cálculos realizados sobre los datos disponibles para obtener la secuencia $a_i = (LD_i, LI_i)$ de las condiciones ambientales con efecto en el largo plazo para cada unidad de estudio.

Características estructurales

Tal como se menciona en la Sección 4.2.2, el modelo HMM para el crimen desarrollado permite incluir heterogeneidad entre los diferentes lugares de una ciudad. En particular se intenta estudiar estas diferencias en las probabilidades transición y de emisión a través de un conjunto de características estructurales del lugar. En este experimento, se utilizará el conjunto de edificios que se encuentran ubicados en una unidad de estudio, por lo que se calcula el número de: escuelas, residencias, puntos de transporte, bancos, oficinas y comercio existentes.

Como estos edificios, y las actividades que ahí se realizan, son los que determinan las rutinas de los habitantes de una ciudad, se esperaría que pudieran explicar las diferencias entre un lugar u otro respecto a la atractividad que posean.

Estas seis variables serán incluidas en las regresiones que se realizan sobre los parámetros relacionados a los umbrales de transición, μ_{isr} , y la probabilidad de no observar crímenes durante un período, $\tilde{\alpha}_i$, que se detallan en las Secciones 4.3.2 y 4.3.2.

5.5. Estimación Modelo

5.5.1. Detalles de la Estimación

De manera de afrontar la posibilidad que el simulador CRIMSIM atravesase por un período de transición, antes de llegar a un estado estacionario, se descartan las primeras 3 semanas simuladas, tomando las 44 semanas restantes como data real a ser estudiada. La data, se divide luego en dos: la primeras 35 semanas de datos para estimar y las últimas 9 semanas para testear y validar el modelo.

La estimación y luego el análisis se realizan sobre las 207 celdas seleccionadas en la Figura 5.5, que corresponden a las unidades de estudio del modelo. Se pretende estudiar la atraktividad percibida en cada una de estas celdas, a partir de los crímenes registrados en ellas a lo largo del tiempo. El algoritmo de estimación, desarrollado en detalle en la Sección B.1 del Apéndice A, es implementado en R, un lenguaje de programación ampliamente utilizado para este tipo de problemas [37].

Una cadena, con puntos de partida aleatorios, es estimada a partir de la secuencia de delitos registrados durante las primeras 35 semanas. De las 300,000 iteraciones de la cadena, las primeras 290,000 fueron descartadas de manera de eliminar el efecto de las condiciones iniciales, y las últimas 10,000 son utilizadas para calcular las densidades a posteriori de los parámetros y estadísticos a posteriori.

De manera de atenuar la correlación entre los sampleos de las iteraciones, siempre presente en los algoritmos de Metropolis, es que se utilizan los valores de una de cada diez iteraciones. De esta manera, se cuenta con 1,000 sampleos de la distribución a posteriori para cada parámetro.

A continuación se describe el proceso de selección del número de estados, los resultados de la estimación y su análisis, así como la validación del modelo, sin embargo, se presentan con detalle en el Capítulo 6.

5.5.2. Elección del Número de Estados

El proceso de selección del número de estados es uno basado en pruebas y comparaciones entre una cantidad finita de modelos. En la Sección B.2 del Apéndice se detallan los diferentes estadísticos que serán utilizados a continuación como herramientas para decidir el número de estados a utilizar.

Asumiendo desde 1 a 7 estados, se estima el modelo desarrollado en el Capítulo 4, considerando que la variable observada se distribuye como una Poisson cero-inflada, sobre los datos generados por el simulador CRIMSIM y las variables descritas en la Sección 5.4.2. Los valores de los estadísticos se pueden observar en la Tabla 5.3.

Número de Estados	-2 Log-densidad marginal	Log-Factor de Bayes	DIC	MSC	-2 Log Verosimilitud de validación
1	10,780,845.3	-	10,780,840.9	10,781,075.4	2,774,986.2
2	75,521.3	5,352,662.0	75,659.4	82,705.2	21,459.0
3	33,812.0	20,854.7	33,828.8	41,085.7	9,754.4
4	28,751.1	2530.5	28,680.6	36,077.1	7,863.1
5	26,846.2	952.5	26,860.1	34,284.2	8,772.2
6	28,106.8	-630.3	28,116.2	36,284.3	7,623.6
7	28,666.3	-279.8	28,652.2	22,628.8	10,065.9

Tabla 5.3: Elección del número de estados.

El modelo de 5 estados presenta mejor desempeño en todos los estadísticos, a excepción de la verosimilitud de validación donde el modelo de 4 y el de 6 estados son capaces de explicar mejor la data reservada para la validación del modelo. Esto podría implicar un sobreajuste a la data de calibración en el modelo de 5 estados, que por tanto, no es considerado para la aplicación.

Si bien el modelo de 6 estados posee mejor desempeño en la mayoría de los estadísticos, el de 4 estados destaca en el estadístico MSC, específico de este tipo de modelos. Las mejoras observadas en el resto de los estadísticos no parecen justificar el uso de 2 estados adicionales, que aumentan la complejidad y los tiempos necesarios para la estimación del modelo.

Este experimento además, tiene como principal objetivo demostrar la aplicación y aprendizajes que se pueden obtener a partir la aplicación del modelo HMM para el crimen. El desarrollo y presentación de resultados para el modelo de 6 estados resulta bastante más engorroso y tedioso debido a la gran cantidad de parámetros estimados que considerar tan solo 4 estados, perdiéndose el objetivo principal del ejercicio.

Por esta razón, se decide utilizar el modelo de 4 estados, que presenta un número de parámetros más reducido y por tanto más adecuado para el propósito de esta sección, y que además es uno de los modelos estimados con mejor desempeño en los estadísticos observados.

Capítulo 6

Resultados

6.1. Introducción

Este capítulo da cuenta de los resultados obtenidos a partir del proceso de estimación a partir del caso de estudio descrito en el Capítulo 5, así como aprendizajes y conclusiones que se pueden obtener sobre el modelo a partir del experimento.

Cabe destacar, dado que se asume que la atractividad de un lugar es inobservable, que las predicciones y validación del modelo solo se pueden medir respecto a la variable observable: la secuencia de crímenes. Es por eso que no se intenta predecir el estado de atractividad en que se encuentra una celda, sino la probabilidad de que se encuentre en cada uno de ellos y a partir de aquello generar una predicción para el número de crímenes esperado para el siguiente período.

La Sección 6.2 contiene el análisis de los resultados obtenidos de la estimación del caso de estudio, a partir de los parámetros estimados, además se reportan los resultados al intentar predecir los crímenes registrados en la data de validación, así como una comparación del desempeño del modelo HMM para el crimen respecto a otras formulaciones alternativas. En la Sección 6.3 se enuncian las principales conclusiones del modelo HMM para el crimen, a partir de la experiencia en la implementación.

6.2. Resultados Específicos

Luego de estimar el modelo HMM para el crimen con 4 estados seleccionado en el Capítulo 5.5, asumiendo que los delitos semanales en cada celda se distribuyen como una Poisson cero-inflada cuya tasa depende del estado de atractividad del sistema. Se utilizan los valores generados cada 10 iteraciones para obtener las distribuciones a posteriori de cada parámetro.

La Tabla C.1 en el Apéndice C muestra las medias y desviaciones estándar a posteriori de los parámetros relacionados con las tasas semanales de crímenes en cada estado. Pudiéndose identificar al estado 4 como aquel que presenta una mayor cantidad de crímenes esperados por período y el estado 1, como aquel con menor cantidad de crímenes esperados. El promedio de las tasas de crímenes para cada estado, del estado 1 al 4, sin vigilancia policial alguna, corresponden a $\lambda = (1.41, 6.87, 21.68, 51.53)$ respectivamente.

Un resultado esperado, es que la vigilancia policial tenga un mayor impacto en una zona

donde existen múltiples oportunidades para delinquir, que en un lugar donde es raro que ocurra un crimen. Recordando además el fenómeno conocido como desplazamiento y los atractores del crimen, revisados en la Sección 2.4 del Capítulo 2, cuando los delincuentes se ven desplazados por la presencia de un policía lo más probable es que se trasladen al atractor del crimen más cercano, prestando poco atención a celdas de bajo atractivo.

Los resultados obtenidos del modelo están de acuerdo con estas ideas sugeridas por la literatura, donde las celdas que se encuentran en el estado de mayor atractividad (estados 3 y 4) son más susceptibles a la vigilancia: disminuyendo en mayor proporción el número de crímenes promedio cuando se apostea un vigilante en el lugar y aumentando el número de crímenes esperados cuando se ubica un policía en lugares aledaños. Celdas en estados 1 y 2, al contrario, tienen una susceptibilidad menor a la vigilancia directa y prácticamente la vigilancia indirecta no tiene impacto en el número de crímenes esperados.

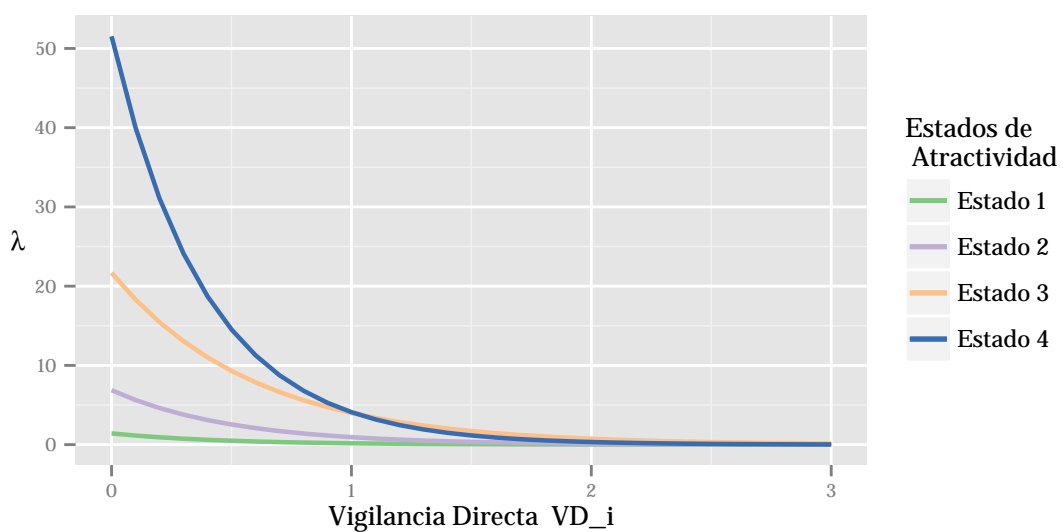


Figura 6.1: Variación de la tasa semanal esperada de crímenes en cada estado, variando el grado de vigilancia directa en la celda VD_i

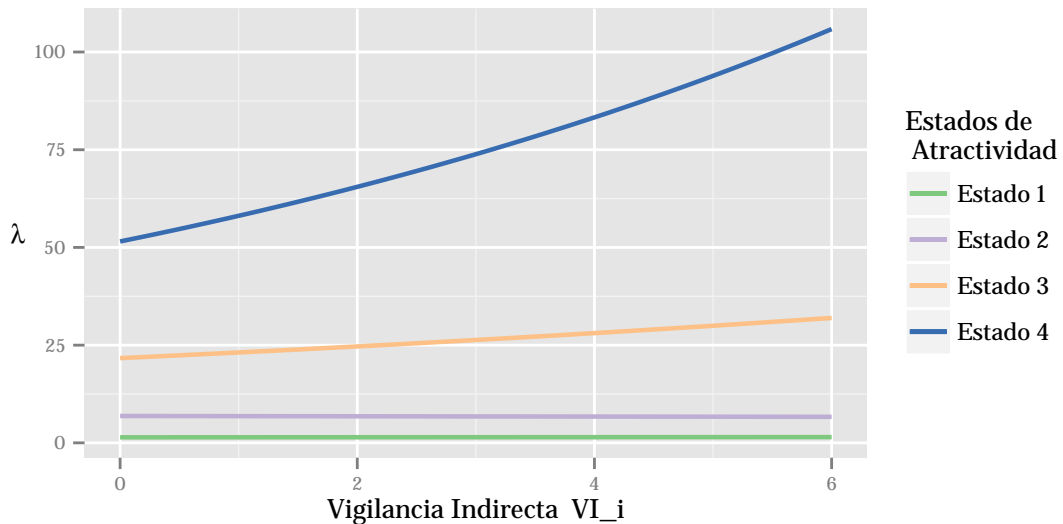


Figura 6.2: Variación de la tasa semanal esperada de crímenes en cada estado, variando el grado de vigilancia en las celdas aledañas VI_i

Los parámetros que capturan el efecto de una vigilancia prolongada en las matrices de transición entre estados se pueden observar en la Tabla C.2 en el Apéndice C. Esta condición ambiental tiene efectos significativos, con 95% de confianza, en los umbrales de transición, excepto para el estado 2 de atractividad.

La vigilancia directa tiene efecto positivo en los umbrales de transición, aumentando sus valores y por tanto, las probabilidades de pasar a estados de baja atractividad. La vigilancia indirecta prolongada tiene el efecto contrario en los umbrales de transición, para los estados de atractividad 1 y 4, no así para el estado 3, aumentando las probabilidades de pasar a estados de mayor atractividad.

Estos resultados se condicen con lo que se esperaría de estas situaciones, donde una vigilancia prolongada podría confirmar las expectativas de los delincuentes período a período, aumentando las probabilidades de permanecer en estados de baja atractividad, en el caso de la vigilancia directa por ejemplo. Importante también destacar, que el efecto de la vigilancia prolongada, tanto directa como indirecta, es mucho mayor en el estado de mayor atractividad, estado 4, que en cualquiera de los otros estados. Por lo que una estrategia orientada a celdas en estado 4 tendría un mayor impacto en la reducción de la atractividad percibida. Por otro lado, la ubicación de recursos policiales en celdas cercanas a estos lugares en el máximo estado de atractividad, genera un aumento en la atractividad percibida de estas mismas.

Los parámetros de efecto aleatorio, umbrales de transición entre estados, μ_{isr} y la probabilidad de observar cero delitos en la Poisson cero-inflada α_i , son propios de cada celda, permitiendo expresar la heterogeneidad entre ellas. Los histogramas de los promedios a posteriori de estos parámetros, en el Apéndice C, sugieren la existencia de dos o tres clases de celdas. De manera de identificar estas categorías, se realiza una clasificación utilizando el método de k-means [20] sobre los umbrales de transición, resultando tres grupos cuyas características se describen en la Tabla 6.1.

Clasificación mediante K-Means			
Clase	Número de Celdas	Parámetro α	
		Media	Desviación Estándar
Bajo Potencial	71	0.50	0.38
Potencial Medio	78	0.73	0.34
Alto Potencial	58	0.00	0.00

Tabla 6.1: Clasificación de las celdas entre aquellas con alto, medio y bajo potencial de registrar crímenes

La primera categoría de celdas, denominada de Bajo Potencial, presenta una probabilidad promedio del 0.50 de no registrar delitos durante una semana, sus matrices de transición, además, llevan a la celda a permanecer en el menor estado de atractividad la mayor parte del tiempo. La segunda categoría, denominada de Potencial Medio, a pesar de que en promedio presenta una probabilidad del 0.73 de no registrar crímenes durante un período, sus matrices de transición le permiten llegar a estados de alta atractividad aunque posee una tendencia a pasar a los estados 1 y 2 con mayor probabilidad. El último segmento, denominado de Alto Potencial, tiene una probabilidad del 0.00 no registrar crímenes y es el grupo de celdas más dinámico, cuyas probabilidades de transición les permite visitar los diferentes estados de atractividad y permanecer en ellos durante algunos períodos. La Figura 6.3 describe esta clasificación de celdas sobre el mapa de la ciudad.

	Matriz de Transición Celdas Según su Potencial											
	Bajo Potencial				Potencial Medio				Alto Potencial			
	1	2	3	4	1	2	3	4	1	2	3	4
Estado 1	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.64	0.31	0.00	0.06
Estado 2	0.95	0.00	0.00	0.05	0.85	0.02	0.00	0.13	0.22	0.71	0.01	0.06
Estado 3	0.81	0.00	0.00	0.19	0.78	0.00	0.00	0.22	0.38	0.02	0.13	0.47
Estado 4	0.31	0.00	0.00	0.69	0.65	0.00	0.00	0.35	0.29	0.00	0.00	0.70

Tabla 6.2: Matriz de transición promedio entre las celdas calificadas como de Bajo Potencial, Potencial Medio y Alto Potencial. Considerando que no existe vigilancia policial a largo plazo de ningún tipo ($LD = 0$ y $LI = 0$).

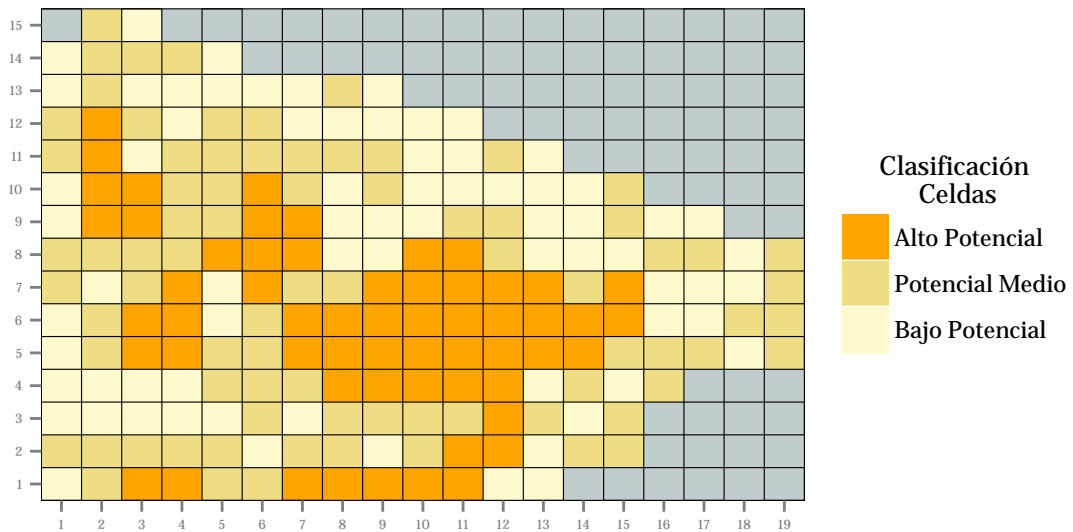


Figura 6.3: Clasificación de celdas de Alto, Medio y Bajo Potencial sobre el mapa de la ciudad.

Respecto a las características físicas de un lugar que capturan la heterogeneidad de cada celda, cuyos parámetros se pueden observar en la Tabla C.3 del Apéndice C, se puede concluir que el número de bancos y escuelas en un lugar reduce la probabilidad de no observar delitos durante un período, α_i . Lo que indica que estos edificios proveen buenas oportunidades para delinquir, ya sea propiciando su ocurrencia a través de su configuración física o bien por el número de personas que atraen estos lugares.

Por otro lado, la presencia de oficinas en un lugar tiende a aumentar la probabilidad de pasar a estados de menor atractividad (al aumentar los umbrales de transición). Al contrario, lugares con un mayor número de residencias, escuelas y bancos tienden a mostrar mayores probabilidades de pasar a estados de mayor atractividad. El número de edificios destinados al comercio solo es significativo en los umbrales de transición del estado cuatro, disminuyendo estos umbrales y aumentando la probabilidad de permanecer en el estado.

La Tabla 6.3 describe las matrices de transición promedio según uso de suelo, usando la composición de la celda en cuanto a estructuras físicas detalladas en la Tabla 5.1 del Capítulo 5 y los parámetros estimados para estos atributos. La probabilidad de no observar delitos en la Poisson cero-inflada, α_i , corresponde a 0.68, 0.42 y 0.00 para las categorías Comercial, Súper Comercial y Estudiantil, respectivamente.

Matriz de Transición Celdas Según Uso de Suelo												
	Comercial				Súper Comercial				Estudiantil			
	1	2	3	4	1	2	3	4	1	2	3	4
Estado 1	0.99	0.00	0.00	0.01	0.99	0.00	0.00	0.01	0.98	0.01	0.00	0.01
Estado 2	0.85	0.10	0.00	0.05	0.88	0.09	0.00	0.03	0.36	0.47	0.01	0.16
Estado 3	0.70	0.00	0.01	0.29	0.69	0.00	0.01	0.30	0.24	0.00	0.04	0.72
Estado 4	0.52	0.00	0.00	0.48	0.39	0.00	0.00	0.61	0.05	0.00	0.00	0.95

Tabla 6.3: Matriz de transición promedio entre las celdas calificadas según uso de suelo: Comercial, Súper Comercial y Estudiantil. Considerando que no existe vigilancia policial a largo plazo de ningún tipo ($LD = 0$ y $LI = 0$).

Es posible concluir que sectores destinados a barrios estudiantiles son foco para la concentración de crímenes en la ciudad, presentando altas posibilidades de pasar y mantenerse en estados de alta atraktividad. No ocurre lo mismo en los sectores destinados a usos comerciales, que corresponden a la mayor parte de la superficie en la ciudad virtual. Es importante destacar, en todo caso, que la composición de uso de suelo corresponde a valores promedios para cada tipo de edificios, por lo que pudieran existir sectores calificados como de uso comercial, pero que aun así contengan algunas escuelas, convirtiéndose en puntos de alto potencial para cometer delitos.

6.2.1. Predicción a Partir del Modelo Estimado

Usando los datos de las 9 últimas semanas, reservadas para este propósito, se valida el modelo estimado en la Sección 5.5, midiendo su capacidad de predicción para estos períodos. La predicción del modelo, realizada a partir de los parámetros estimados, consta de dos partes: la primera se centra en encontrar para cada unidad de estudio las probabilidades de encontrarse en cada estado durante cada período de validación, dada la información disponible hasta el período anterior. La segunda parte consiste en calcular un intervalo, para cada celda, entre el que se pronostica se encontrará el número de crímenes del siguiente período, dadas las condiciones ambientales que se esperan para ese período.

La atraktividad percibida por los delincuentes es inobservable, por lo que no se puede comprobar en qué estado se encuentra efectivamente una celda en cualquiera de los períodos. Sin embargo el número de crímenes a predecir sí es observable y puede ser usado para comprobar la capacidad de predicción del modelo. Como medida para la capacidad de predicción, se contabilizan los períodos en que el número real de crímenes registrados durante el período estuvo efectivamente dentro del intervalo predicho.

$$\text{acierto}_{it} = \begin{cases} 1 & \text{si } Lim_{inf} \leq Y_{it} \leq Lim_{sup} \\ 0 & \text{si no.} \end{cases}$$

Se define la tasa de aciertos como el porcentaje de aciertos en el total de períodos y celdas.

$$\text{Tasa aciertos} = \frac{1}{T_{val} \times M} \sum_{t=1}^{T_{val}} \sum_{i=1}^M \text{acierto}_{it}$$

La predicción de los intervalos se realiza a partir de las distribuciones a posteriori de las tasas semanales de crímenes, λ_{ist} , pronosticadas para cada período y las probabilidades de encontrarse en cada uno de los estados en el período a pronosticar. Los límites del intervalo se definen de manera de que el número de crímenes a predecir se encuentre con un 90% de probabilidad dentro de este, asumiendo que este número de crímenes se distribuye como la Poisson cero-inflada estimada en el modelo.

$$p_{inf} = \text{máx}\{0.05 - \alpha_i, 0\}$$

$$p_{sup} = \text{máx}\{0.95 - \alpha_i, 0\}$$

$$Lim_{inf} = \text{mín}\left\{x \in \mathbb{N} : \sum_{s=1}^N \sum_{k=0}^x P(Y_{it} = k | S_t = s) P(S_{it} | Y_{i1}, Y_{i2}, \dots, Y_{i(t-1)}) \geq \frac{p_{inf}}{1 - \alpha_i}\right\}$$

$$Lim_{sup} = \min\{x \in \mathbb{N} : \sum_{s=1}^N \sum_{k=0}^x P(Y_{it} = k | S_t = s) P(S_{it} | Y_{i1}, Y_{i2}, \dots, Y_{i(t-1)}) \geq \frac{P_{sup}}{1 - \alpha_i}\}$$

Donde las probabilidades de observar el número de crímenes, dado un estado y que este número se distribuye como una Poisson, corresponden a:

$$P(Y_{it} = k | S_t = s) = \frac{\exp(-\lambda_{is}) \lambda_{is}^k}{k!}$$

Las probabilidades de encontrarse en cualquiera de los estados en el período $t - 1$ dada la secuencia de crímenes hasta ese período, p_i^t , fue calculada en la Sección 3.2.4 del Capítulo 3. La probabilidad de encontrarse en un estado s en el período t puede ser calculada, entonces, utilizando lo anterior y la matriz de transición de la celda en el período $(t - 1)$, $Q_{i(t-1)}$.

$$P(S_t = s | Y_{i1}, Y_{i2}, \dots, Y_{i(t-1)}) = (p_i^t)' Q_{i(t-1)}$$

En la siguiente tabla se puede observar las tasas de aciertos obtenidas a partir de la predicción del modelo, tanto de forma general como según la clasificación realizada según su potencial.

Capacidad de Predicción del Modelo	
	Tasas de Acierto
Total Celdas	97 %
Celdas Bajo Potencial	99 %
Celdas Potencial Medio	97 %
Celdas Alto Potencial	97 %

Tabla 6.4: Tasas de acierto del modelo HMM para el crimen sobre la data de validación.

El modelo HMM para el crimen estimado tiene tasas de aciertos mayores al 90% exigido, de forma general y en cada uno de los tipos de celdas. Esto sugiere que la distribución estimada para el número de crímenes le asigna mayor probabilidad a valores de mayor magnitud que la distribución real de los crímenes. Implica además, que los intervalos pronosticados son más anchos de los requeridos para acertar con un 90% de confianza. Se hace necesario comparar los resultados de este modelo con el de otros de estructura diferente antes de concluir sobre su desempeño, tarea que se aborda en la Sección 6.2.2.

6.2.2. Validación del Modelo

Usando los datos de las 9 últimas semanas, se valida el modelo estimado en la Sección 5.5, midiendo su capacidad de predicción para estos períodos y comparándolo con otros cuatro modelos alternativos.

La probabilidad que el número de crímenes se encuentre dentro del intervalo de predicción es mayor mientras más ancho sea este intervalo, por lo que es importante considerar dicho punto

cuando se desea comparar distintos modelos. Un atributo deseable es que el modelo sea capaz de predecir a cierto nivel de confianza con el intervalo de predicción más pequeño posible.

Una manera de comparar la agudeza de la predicción de diferentes modelos es la utilización de reglas de puntaje para la medición de este atributo. Estos puntajes se estructuran de manera que mientras más cercana es a la distribución real, de la variable observada, con la calibrada en el modelo, menor es el puntaje que recibe el modelo. De cierta manera, es una medida de distancia entre las distribuciones real y pronosticada, que se desea lo más pequeña posible. Mientras más ajustada la distribución real con la estimada, los intervalos de predicción tienden a ser más pequeños y se conoce con mayor certeza el rango en que se mueve la variable observada. Se utilizará la puntuación CRPS(*continous ranking probablity score*), usada para medir el ajuste de distribuciones de Poisson, para medir el ajuste de las predicciones de los diferentes modelos [15]. El cálculo de este puntaje se puede estudiar con más detalle en la Sección C.4 del Apéndice.

La verosimilitud de los modelos en la data de validación, es decir, la probabilidad que la data observada fuera producida por la estructura propuesta por los modelos, será también utilizada como medida de comparación de la capacidad de predicción.

Los modelos estimados alternativos corresponden a:

- HMM homogéneo: modelo de estructura equivalente al modelo HMM heterogéneo, pero que considera las mismas matrices de transición para todas las celdas e igual probabilidad de no registrar delitos.
- Clases Latentes: modelo que considera cuatro categorías, con tasas esperadas de crímenes específicas para cada una, y que clasifica cada celda en alguna de estas categorías (las celdas no pueden cambiar de categoría).
- Regresión de Poisson: modelo que intenta estimar la tasa esperada de crímenes a partir de las condiciones ambientales, donde el efecto de estas es el mismo para todas las celdas.
- Persistencia: modelo tomado desde la literatura [31] que considera que existen dos tipos de crímenes, aquellos que son espontáneos y que surgen aleatoriamente, y las secuelas, que son crímenes que se generan en torno a los crímenes espontáneos con cierta tasa.

En el Apéndice C se encuentran las descripciones más detalladas de cada uno de ellos. A pesar que todos estos modelos están basados en distribuciones de Poisson y Poisson cero-infladas, pronosticando tasas de crímenes para cada celda en cada período, la estructura es diferente para cada uno de ellos.

La predicción de los intervalos se realiza a partir de las distribuciones a posteriori de las tasas semanales de crímenes, λ_{ist} , pronosticadas para cada período y las probabilidades de encontrarse en cada uno de los estados en el período a pronosticar. Los límites del intervalo se definen de manera de que el número de crímenes a predecir se encuentre con un 90% de probabilidad dentro del este, asumiendo que este número de crímenes se distribuye como la Poisson, o Poisson cero-inflada, estimada en el modelo.

Los resultados de las predicciones sobre la data de validación se pueden observar en la Tabla 6.5.

Ajuste y Capacidad de Predicción de los Modelos

Medida	Modelos				
	Modelo 1 HMM heterogéneo	Modelo 2 HMM homogéneo	Modelo 3 Clase Latente	Modelo 4 Regresión Poisson	Modelo 5 De Persistencia
Verosimilitud					
-2 Log-Verosimilitud	7863.1	9433.7	9376.1	34084.5	15711.5
Validación					
Tasas de Aciertos					
% Aciertos Total	97 %	92 %	97 %	82 %	71 %
% Aciertos Bajo Potencial	99 %	99 %	99 %	97 %	90 %
% Aciertos Potencial Medio	97 %	97 %	99 %	94 %	81 %
% Aciertos Alto Potencial	97 %	78 %	92 %	48 %	34 %
CRPS					
Puntaje Total	3.57	3.62	6.53	13.48	4.57
Puntaje Bajo Potencial	2.07	1.90	4.65	12.38	0.37
Puntaje Potencial Medio	2.08	1.94	3.4	11.30	1.08
Puntaje Alto Potencial	7.50	8.02	12.59	17.24	13.01
Intervalos de Predicción					
Ancho Total	24.27	17.2	36.45	12.28	3.56
Ancho Bajo Potencial	2.66	10.54	9.45	9.35	0.85
Ancho Potencial Medio	5.62	8.58	18.78	14.52	1.87
Ancho Alto Potencial	75.27	34.21	89.8	13.52	9.76

Tabla 6.5: Medidas de ajuste y predicción de los modelos a comprar, sobre la data de validación.

El modelo HMM con heterogeneidad presenta mejores tasas de aciertos generales que cualquiera de los otros cuatro modelos, en particular se destaca prediciendo con mejor desempeño el número de crímenes en las celdas de Alto Potencial. Todos los modelos pueden predecir casi con total certeza el rango de crímenes entre las celdas de Bajo Potencial, que tienden a permanecer en el menor estado de atractividad la mayor parte del tiempo y el que para efectos de la policía, existe menor interés en pronosticar. Este segmento siendo el más numeroso, tiene un peso mayor en la función de verosimilitud, es de esperarse entonces que los modelo HMM homogéneo, que no considera heterogeneidad entre las celdas, y la Regresión de Poisson, que considera un único estado, ajusten sus parámetros en función de este segmento, pudiendo predecir bastante bien su comportamiento. Siendo además el segmento de Bajo Potencial un conjunto de celdas prácticamente estáticas, el modelo de Clases Latentes puede explicar su comportamiento y predecir con gran certeza el número de crímenes que ocurrirán en esas celdas, sin embargo, este desempeño decae en el segmento de Alto Potencial, que contiene celdas más dinámicas, algo que el modelo no puede capturar.

El modelo de Persistencia, que considera heterogeneidad en las celdas pero no el efecto de la vigilancia en las tasas esperadas de crímenes, presenta tasas de aciertos inferiores a las del resto de los modelos, en cada uno de los segmentos.

En cuanto al puntaje CRPS, el modelo HMM con heterogeneidad presenta predicciones más ajustadas en general, y en cada segmento en particular. Esto se puede apreciar en el ancho promedio de los intervalos de predicción, significativamente menores que el modelo de Clases Latentes, que es el que posee tasas de aciertos más cercanas entre los modelos de

comparación. Los modelos HMM homogéneo y Regresión de Poisson, presentan intervalos aún más ajustados, pero la tasa de aciertos, para los segmentos de Alto y Mediano Potencial, que presentan estos modelos están por debajo del 90% pronosticado y del desempeño mostrado por el modelo HMM heterogéneo.

6.3. Resultados Generales del Modelo

En el ámbito del crimen los delitos tienden a concentrarse en algunos pocos lugares donde las oportunidades son mayores [45], esta característica de los datos con numerosas celdas de Bajo Potencial y escasas celdas de Alto Potencial podría darse con bastante regularidad en datos reales. En dicho caso, la capacidad del modelo HMM con heterogeneidad para capturar las diferencias entre estos grupos de celdas permite mejores predicciones en las celdas de interés: las que tienden a concentrar estos crímenes.

En cuanto a las características físicas, es claro que su inclusión solo vale la pena si el área que se está estudiando tiene una composición no homogénea en su superficie de estas características físicas. Puesto que su función es capturar las diferencias entre las celdas, debe existir una varianza significativa de estos atributos para que cumplan dicha función.

Por otro lado, parece importante incluir el efecto que la vigilancia, u otras intervenciones, pudieran tener en las tasas esperadas de crímenes y en la atractividad percibida por los delincuentes. Los modelos estudiados que incluyen estos efectos en su estructura, tienden a predecir mejor, que el modelo que no los incluye (al menos en el experimento realizado). En este sentido, la utilización de un modelo HMM como el presentado en este trabajo, no solo entrega información respecto al efecto de intervenciones policiales u de otro estilo en la dinámica del crimen, como desplazamientos por ejemplo, sino que las predicciones que ofrece son más ajustadas y con mejor desempeño que las de los modelos alternativos evaluados que también incluyen estos efectos.

Esta aplicación demuestra la utilidad que el modelo HMM desarrollado en este trabajo tiene para los tomadores de decisiones en cuanto a la prevención y combate contra el crimen. La habilidad para incorporar variables describiendo las condiciones ambientales en la matriz de transición brinda la oportunidad de investigar cuáles son las actividades, intervenciones y estrategias de prevención más efectivas para reducir la atractividad de los lugares para cometer un determinado tipo de crimen y para determinar los lugares óptimos donde enfocar estas intervenciones. En el experimento realizado se pudo constatar, por ejemplo, que una vigilancia prolongada tiene un efecto mayor en la atractividad percibida por los delincuentes en lugares que ya se encuentran en un estado de alta atractividad, que celdas percibidas como bajo atractivo.

El modelo además, no solo entrega el sentido en que las intervenciones pueden afectar la tasa esperada de crímenes, sino que también le asigna una magnitud cuantitativa a estos efectos. En este experimento, se consideraron las intervenciones como un dato en los períodos a predecir, para un tomador de decisiones, sin embargo, el contar con una predicción sobre el estado de atractividad en que se encuentra cada celda le permitiría decidir qué tipo de intervención utilizar en qué lugar en dicho momento. Pudiéndose generar una estrategia óptima dada esta información, según algún objetivo específico que se desee alcanzar.

Capítulo 7

Conclusiones y Trabajo Futuro

7.1. Conclusiones

La prevención del crimen ha ganado cada vez más espacio e importancia entre las políticas públicas en seguridad ciudadana, tanto en Chile como en el mundo. En línea con esto, comprender el fenómeno delictivo y cómo este se ve afectado cuando se realizan intervenciones destinadas a su prevención, es crucial para generar estrategias más efectivas, dado los recursos disponibles para la lucha contra el crimen.

En el modelo desarrollado a lo largo de este trabajo se utiliza los registros de crímenes en una ciudad para estimar un modelo de Markov oculto, considerando la atractividad de los lugares percibida por los delincuentes como la señal oculta del modelo. El modelo HMM, estructurado como un modelo jerárquico Bayesiano de manera de considerar heterogeneidad en los parámetros, permite entender y estudiar las variables y la estructura que rige el proceso de actualización de la atractividad de un lugar.

La principal contribución de este trabajo, es la sugerencia de un modelo que permite incluir en su formulación el efecto de intervenciones en un lugar, ligadas a la prevención del crimen, en el proceso de actualización de la atractividad y en el número de crímenes esperados para cada período. Este aspecto de la problemática de la prevención del crimen y su predicción ha sido abordada solo en unos pocos trabajos, según la bibliografía revisada. Conocer el impacto que una intervención puede provocar en las tasas esperadas de crímenes y en la atractividad percibida por los delincuentes le permite a la policía, u otras instituciones, fundamentar las decisiones en cuanto a qué tipo de intervención se debe realizar en qué lugar. Además, el modelo permite la clasificación de los lugares cada período según el estado de atractividad de los mismos y enfocar de mejor manera los esfuerzos, así como segmentarlos según el potencial que tengan para convertirse en puntos de concentración de crímenes.

La aplicación en un caso de estudio, sobre los datos simulados mediante el simulador del crimen CRIMSIM, permite visualizar de mejor manera el potencial uso del modelo en el estudio de la dinámica del crimen y la subyacente atractividad percibida. A través del modelo se pudo estudiar el efecto de la vigilancia policial en el número de crímenes registrados, así como en el impacto sobre la atractividad percibida por los delincuentes. Se pudo constatar que la vigilancia tiene un mayor impacto en lugares que se encuentran en estados de gran atractividad, disminuyendo el número de crímenes esperados para el período en mayor magnitud que en el resto de los estados. Por otro lado, se pudo analizar el efecto desplazamiento que tiene

esta medida de vigilancia, tal como se constata en la literatura, los lugares en estados de baja atraktividad apenas se ven afectados por la vigilancia en sectores aledaños, sin embargo el número de crímenes esperado aumenta significativamente para lugares en estados de alta atraktividad.

En términos de capacidad de predicción, se demuestra que, en el experimento realizado, el modelo HMM con heterogeneidad es superior que el resto de los modelos evaluados (HMM homogéneo, Clases Latentes, Regresión de Poisson y Persistencia), tanto en las tasas de aciertos como en el tamaño de los intervalos de predicción. Si bien todos los modelos evaluados son capaces de predecir con gran certeza los crímenes para las celdas clasificadas como de Bajo Potencial, el modelo HMM heterogéneo es capaz de predecir con gran precisión los delitos para las celdas de Alto Potencial, que suelen ser los lugares con mayor interés para la policía, cuyo comportamiento el resto de los modelos tiene dificultades para capturar.

Esta aplicación demuestra la utilidad que el modelo HMM desarrollado en este trabajo tiene para los tomadores de decisiones en cuanto a la prevención y combate contra el crimen. Permitiéndoles período a período contar con un rango de predicción para los crímenes en cada lugar de la ciudad y la estimación de los estados de atraktividad en los que se encuentran, de manera de escoger la intervención que genere el mayor impacto, dado algún objetivo específico. La habilidad para incorporar variables describiendo las condiciones ambientales en la matriz de transición brinda la oportunidad de investigar cuáles son las actividades, intervenciones y estrategias de prevención más efectivas para reducir la atraktividad de los lugares para cometer un determinado tipo de crimen y para determinar los lugares óptimos donde enfocar estas intervenciones.

Utilizar una predicción para generar una estrategia de prevención o vigilancia, no es un problema trivial, que en la mayoría de los modelos revisados en la literatura se traduce en mantener vigiladas las zonas identificadas como de mayor riesgo o implementar políticas de prevención orientadas a esos lugares. Sin embargo, poseer información referente al efecto de la policía en la dinámica del crimen permite generar estrategias con objetivos específicos, como reducir el número total de crímenes, o el número de crímenes en sectores particulares, teniendo en consideración, además, efectos secundarios que se podrían observar, como por ejemplo un efecto desplazamiento.

Una limitación del modelo, sin embargo, recae en su complejidad para programar e implementar el algoritmo de estimación, así como el tiempo necesario para que los parámetros converjan a su distribución a posteriori. Estos tiempos de estimación crecen exponencialmente con el número de estados, por lo que se debe tener cuidado cuando se toma la decisión sobre cuántos estados considerar. En el experimento implementado, estimar el modelo para cuatro estados toma alrededor de 36 horas, y el de siete aproximadamente una semana. En caso de que no existieran cambios significativos en la atraktividad de un lugar según la data, la utilización un modelo simplificado de Clases Latentes, como el desarrollado en la sección de validación del Capítulo 5, puede ser una buena alternativa para ser utilizado, reduciendo los tiempos de estimación.

Finalmente, en relación a los objetivos planteados al inicio de este trabajo, se puede concluir que el modelo HMM, aquí desarrollado, logra incorporar de forma efectiva los dos atributos que se deseaban incluir al estudiar el fenómeno delictual: el modelo considera la atraktividad de forma dinámica, al moverse los lugares entre diferentes estados de atraktividad, e incluye el efecto de intervenciones en el lugar en las tasas de crímenes y en la percepción de los delincuentes sobre su atractivo.

7.2. Trabajo Futuro

La implementación de la metodología desarrollada en este trabajo en un caso con datos reales queda pendiente como trabajo futuro. Las intervenciones afectando probabilidades de emisión y transición en el modelo no son restrictivas y pudieran ser de cualquier tipo tal que tengan incidencia en la dinámica del crimen. De esta forma, el modelo HMM no solo puede ser utilizado para predecir crímenes y elaborar estrategias de prevención, sino que también puede ser utilizado para estudiar el impacto en el crimen de instalaciones que cambien la estructura física de la ciudad. Comprender la forma en que un centro comercial, estaciones de metro, cambios de usos de suelo, parques y áreas verdes, afectan las tasas de crímenes y la atraktividad de los lugares donde son implementadas, es información que puede ser utilizada para análisis de impacto ambiental y urbano.

Sería interesante comparar los resultados y el ajuste del modelo HMM, que se basa sólo en el registro de crímenes y variables de vigilancia, con otros modelos exhaustivos en variables, que describen de forma elaborada la estructura de cada lugar y las condiciones reinantes en cada período. Para así determinar si el supuesto de considerar la atraktividad como un atributo oculto, que los otros modelos intentan describir a partir de estos grandes conjuntos de variables, permite estudiar el crimen con una efectividad comparable, sin necesidad de tal volumen de información.

En el modelo presentado se considera un único tipo de crimen, la inclusión simultánea de dos o más tipos de crímenes relacionados sería interesante para estudiar fenómenos de desplazamiento distintos al geográfico, como el desplazamiento de tipo de crímenes. Suele suceder que diferentes tipos de crímenes tienen alta correlación en su ocurrencia, como delitos en tráfico de drogas que derivan en prostitución, pandillas y asesinatos. Intervenir un tipo de crimen puede llevar a un desplazamiento a otro tipo de crimen [36]. Considerar los estados de atraktividad como un doble atributo, que considere la atraktividad de cada tipo de crimen como un elemento de un vector que describe el estado, podría ser una forma interesante de incluir estos desplazamientos.

De igual manera, se considera en el modelo que los lugares son independientes unos de otros, algo que en la práctica es difícil de justificar. La inclusión de vigilancia en celdas aledañas permite, de alguna manera, incluir el efecto de los que ocurren en una celda en las que la rodean. Sin embargo, es probable que el atractivo que tenga un lugar pueda afectar la percepción que los delincuentes tengan sobre las celdas que lo rodean. La inclusión de dependencia, o de relación entre las celdas, sería una forma adecuada de estudiar fenómenos como el de la ventana rota, donde lugares descuidados y con altos niveles de delincuencia tienden a incentivar más delincuencia en el lugar y sus alrededores.

Bibliografía

- [1] Barr, Robert y Pease, Ken: *Crime Placement, Displacement, and Deflection*. Crime and Justice: A Review of Research, 12:277–318, 1990.
- [2] Baum, Leonard E., Petrie, Ted, Soules, George y Weiss, Norman: *A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains*. The Annals of Mathematical Statistics, 41(1):164–171, 1970.
- [3] Becker, Gary S: *Crime and Punishment: An Economic Approach*. The Journal of Political Economy, 76(2):169–217, Mar-Abr 1968.
- [4] Biem, Alain: *A Model Selection Criterion for Classification: Application to HMM Topology Optimization*. En *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on*, volumen 1, páginas 104–108, Ago 2003.
- [5] Braga, Anthony A.: *The Effects of Hot Spots Policing on Crime*. The ANNALS of the American Academy of Political and Social Science, 578(1):104–125, 2001.
- [6] Brantingham, Paul J. y Brantingham, Patricia L.: *Criminality of place*. European Journal on Criminal Policy and Research, 3:5–26, Sep 1995.
- [7] Brantingham, Paul J. y Brantingham, Patricia L.: *Anticipating the Displacement of Crime Using the Principles of Environmental Criminology*. Crime Prevention Studies, 16:119–148, 2003.
- [8] Bruce L. Benson y Paul R. Zimmerman: *Handbook in the Economics of Crime*, capítulo 1 - 4. Edward Elgar Publishing Limited, 2010.
- [9] Clarke, Ronald V.: *Building a Safer Society: Strategic Approaches to Crime Prevention*. Crime and Justice, 19:91–150, 1995.
- [10] Clarke, Ronald V.: *Situational Crime Prevention: Successful Case Studies*, capítulo 1 - 4. Harrow and Heston Publisher, Guilderland, New York, 2ª edición, 1997.
- [11] Corcoran, Jonathan J., Wilson, Ian D. y Ware, J.Andrew: *Predicting the geo-temporal variations of crime and disorder*. International Journal of Forecasting, 19(4):623–634, 2003.
- [12] Dammert, Lucía y Lunecke, Alejandra: *La Prevención del Delito en Chile. Una visión desde la Comunidad.*, capítulo 1 - 2. Serie Estudios : Colección Seguridad Ciudadana y Democracia. Centro de Estudios en Seguridad Ciudadana, 2004.
- [13] Devia, Nelson y Weber, Richard: *Generating crime data using agent-based simulation*. Computers, Environment and Urban Systems, 42:26–41, 2013.

- [14] Ghosh, Sujit K., Mukhopadhyay, Pabak y Jye-Chyi(JC) Lu: *Bayesian analysis of zero-inflated regression models*. Journal of Statistical Planning and Inference, 136(4):1360–1375, 2006.
- [15] Gneiting, Tilmann, Balabdaoui, Fadoua y Raftery, Adrian E.: *Probabilistic Forecasts, Calibration and Sharpness*. Journal of the Royal Statistical Society, Series B, Statistica Methodology, 69:243–268, 2007.
- [16] Gonzales, Alberto R., Schofield, Regina B. y Hart, Sarah V.: *Mapping Crime: Understanding Hot Spots*. U.S. Department of Justice: National Institute of Justice, 2005.
- [17] Greene, William H.: *Econometric Analysis*, capítulo 19. Prentice Hall, Upper Saddle River, New Jersey, 3ª edición, 1997.
- [18] Groff, Elizabeth R. y La Vigne, Nancy G.: *Mapping an Opportunity Surface of Residential Burglary*. Journal of Research in Crime and Delinquency, 38(3):257–277, 2001.
- [19] Hall, Daniel B.: *Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study*. Biometrics, 56(4):1030–1039, 2000.
- [20] Hartigan, J. A. y Wong, M. A.: *Algorithm AS 136: A K-Means Clustering Algorithm*. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1):100–108, 1979.
- [21] Huddleston, Samuel H. y Brown, Donald E.: *A Statistical Threat Assessment*. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 39(6):1307–1315, 2009.
- [22] Jeffery, Clarence R.: *The Historical Development of Criminology*. Journal of Criminal Law and Criminology, 50(1):4–19, 1951.
- [23] Jofré, M. P.: *Análisis del fenómeno delictual utilizando un modelo de regresión logística en base a atributos*. Tesis de Licenciatura, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial., 2011.
- [24] Kelly, Morgan: *Inequality and Crime*. The Review of Economics and Statistics, 82(4):530–539, 2000.
- [25] Keneddy, Leslie W., Caplan, Joel M. y Piza, Eric: *Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies*. Journal of Quantitative Criminology, 27(3):339–362, 2011.
- [26] Lambert, D.: *Zero-inflated Poisson regression, with an application to defects in manufacturing*. Technometrics, 34:1–14, 1992.
- [27] Lazzati, Natalia y Menichini, Amilcar A.: *Hot Spot Policing: A Theoretical Study of Place-Based Strategies to Crime Prevention*. Disponible en <http://dx.doi.org/10.2139/ssrn.2277876>, 2013.
- [28] MacDonald, Iain L. y Zucchini, Walter: *Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman and Hall, Londres, 1997.
- [29] Mazerolle, Lorraine, Price, James F. y Roehl, Jan: *Civil Remedies and Drug Control: A Randomized Field Trial in Oakland, California*. Evaluation Review, 24(2):212–241, 2000.

- [30] Meyn, S. y Tweedie, R. L.: *Markov Chains and Stochastic Stability*, capítulo 1. Wiley Series in Probability and Statistics. Cambridge University Press, Cambridge CB2 8RU, UK, 2ª edición, 2009.
- [31] Mohler, G. O., Short, M. B., Brantingham, P.J., Schoenberg, F. P. y Tita, G.E: *Self-exciting point processes explain spatial-temporal patterns in crime*.
- [32] Netzer, Oded, Lattin, James M. y Srinivasan, V.: *A Hidden Markov Model of Customer Relationship Dynamics*. *Marketing Science*, 27(2):185–204, mar 2008.
- [33] Newton, Michael A. y Raftery, Adrian E.: *Approximate Bayesian Inference with the Weighted Likelihood Bootstrap*. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(1):3–48, 1994.
- [34] Osgood, D. Wayne: *Poisson-Based Regression Analysis of Aggregate Crime Rates*. *Journal of Quantitative Criminology*, 16(1):21–43, 2000.
- [35] Rabiner, L. R. y Juang, B. H.: *An Introduction to Hidden Markov Models*. IEE ASSP Magazine, páginas 4–15, jan 1986.
- [36] Reppetto, Thomas A.: *Crime Prevention and the Displacement Phenomenon*. *Crime and Delinquency*, 22(2):166–177, 1976.
- [37] Rossi, P. E., Allenby, G. M. y McCulloch, R.: *Bayesian Statistics and Marketing*, capítulo 1-6. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd, Chichester, UK, 2006.
- [38] Rossi, Peter E. y Allenby, Greg M.: *Bayesian Statistics and Marketing*. *Marketing Science*, 22(3):304–328, Mar 2003.
- [39] Sherman, Lawrence W.: *Hot Spots of Crime and Criminal Careers of Places*. *Crime and Place. Crime Prevention Studies*, 24:35–52, 1995.
- [40] Smith, Aaron, Naik, Prasad A. y Tsai, Chih Ling: *Markov-switching model selection using Kullback-Leibler divergence*. *Journal of Econometrics*, 134(2):553–577, 2006.
- [41] Spiegelhalter, David J., Best, Nicola G., Carlin, Bardley P. y Van der Linde, Angelika: *Bayesian Measures of Model Complexity and Fit*. *Journal of the Royal Statistical Society. Series B (Methodological)*, 64(4):583–639, 2002.
- [42] Stephens, Matthew: *Dealing with Label Switching in Mixture Models*. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 62(4):795–809, 2000.
- [43] Tierney, Luke: *Markov Chains for Exploring Posterior Distributions*. *The Annals of Statistics*, 22(4):1701–1728, 1994.
- [44] Winter, Harold: *The Economics of Crime: An Introduction to Rational Crime Analysis*, capítulo 1. Routledge, 1ª edición, Jun 2008.
- [45] Wortley, Richard y Mazerolle, Lorrain: *Environmental Criminology and Crime Analysis*, capítulo 1 - 4. Crime Science Series. Willan Publishing, 2008.
- [46] Xue, Yifei y Brown, Donald E.: *A decision model for spatial site selection by criminals: a foundation for law enforcement decision support*. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 33(1):78–85, 2003.

Apéndice A

Apéndice A

A.1. Algoritmo de Escalamiento de la Función de Verosimilitud

Dado que la función de verosimilitud consiste en una multiplicación de probabilidades, todas ellas menores que 1, mientras mayor el número de períodos a considerar, mayor el riesgo que el valor obtenido sea demasiado pequeño para la capacidad computacional de almacenamiento de decimales, con lo que el valor retornado es 0. Para evitar este problema, se utiliza un escalamiento del logaritmo de la verosimilitud, de manera de poder manejar estos números de magnitudes tan pequeñas.

La función de verosimilitud del modelo de HMM puede ser escrita como:

$$L_i(y_i) = P(Y_{i1} = y_{i1}, \dots, Y_{iT} = y_{iT}) = \pi_i \tilde{m}_{i1} Q_{i,1 \rightarrow 2} \tilde{m}_{i2} \dots Q_{i,T-1 \rightarrow T} \tilde{m}_{iT} 1'$$

El mecanismo de escalamiento, propuesto por [28] y utilizado por [32] se describe a continuación mediante el siguiente algoritmo.

Algoritmo 1 Escalamiento del Logaritmo de la Verosimilitud

```
Ai = πi·m̃i1
s = 1/N · Ai1'
Ai = Ai/s
logLi = log(s)
for t = 2 → T do
  Ai = Ai · Qi,(t-1)→t · m̃it
  s = 1/N · Ai1'
  Ai = Ai/s
  logLi = logLi + log(s)
end for
Li = Ai · 1'
return ScaledLogLi = log(Li) + logLi
```

A.2. Priors Conjugados de un Modelo de Regresión Multivariada

En los modelos que se desarrollarán en el Capítulo 4 se utilizarán como priors distribuciones gaussianas, que tienen momentos simples, y cuyos conjugados se encuentran bien documentados en la literatura. En particular, se utilizará una regresión multivariada: un conjunto de m ecuaciones de regresión para un conjunto común de covariables X , cuyos errores se encuentran correlacionados.

$$Y = XB + E$$

Donde $Y, E \in \mathbb{R}^{n \times m}$, $X \in \mathbb{R}^{n \times k}$ y $B \in \mathbb{R}^{k \times m}$. Para cada observación $r = 1, \dots, n$, se tiene que:

$$y_r = B'x_r + \varepsilon_r, \quad \varepsilon_r \sim N(0, \Sigma)$$

Si se asume que solo existe información respecto a la varianza de Y , y no a su media, se tiene que $Y = E$. La verosimilitud del modelo, bajo esta restricción, es equivalente a la probabilidad de observar los errores dada la matriz de covarianzas.

$$p(E|\Sigma) \propto |\Sigma|^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \text{tr} E' E \Sigma^{-1}\right)$$

Al agregar el resto de la información disponible, las variables, X , y sus ponderadores, B , que explican el vector de medias de Y , la verosimilitud se expresa como:

$$p(E|\Sigma, X, B) \propto |\Sigma|^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \text{tr}(Y - XB)'(Y - XB)\Sigma^{-1}\right)$$

Descomponiendo la suma de cuadrados, usando la proyección del estimador por mínimos cuadrados $\hat{B} = (X'X)^{-1}X'Y$, la expresión anterior puede ser dividida en dos factores.

$$p(E|\Sigma, X, B) \propto \left\{ |\Sigma|^{-\frac{n-k}{2}} \exp\left(-\frac{1}{2} \text{tr} S \Sigma^{-1}\right) \right\} \left\{ |\Sigma|^{-\frac{k}{2}} \exp\left[-\frac{1}{2} \text{tr}(B - \hat{B})' X' X (B - \hat{B}) \Sigma^{-1}\right] \right\}$$

$$\text{Donde } S = (Y - X\hat{B})'(Y - \hat{B}).$$

Para encontrar los priors conjugados se debe cumplir que:

$$\begin{aligned} p(B|X, \Sigma) \propto p(B)p(E|\Sigma, X, B) &\sim F_1 & \text{con } p(B) &\sim F_1 \\ p(\Sigma|X, B) \propto p(\Sigma)p(E|\Sigma, X, B) &\sim F_2 & \text{con } p(\Sigma) &\sim F_2 \end{aligned}$$

La expresión de $p(E|\Sigma, X, B)$ sugiere que el conjugado natural de Σ es una distribución Wishart inversa [37] y que el conjugado natural de B es una distribución normal condicionada en Σ . En particular, los priors conjugados de una regresión multivariada son de la forma (para más detalle [37]):

$$\begin{aligned} \Sigma &\sim IW(v_0, V_0) \\ \text{vec}(B)|\Sigma &\sim N(\tilde{\beta}, \Sigma \otimes A^{-1}) \end{aligned}$$

Las distribuciones a posteriori, bien documentadas en la literatura y comúnmente usadas en modelos Bayesianos, corresponden a:

$$\begin{aligned} \Sigma|Y, X &\sim IW(v_0 + n, V_0 + S) \\ \text{vec}(B)|Y, X, \Sigma &\sim N(\tilde{\beta}, \Sigma \otimes (X'X + A)^{-1}) \\ \tilde{\beta} = \text{vec}(\tilde{B}) &\quad \tilde{B} = (X'X + A)^{-1}(X'Y + A\bar{B}) \\ S = (Y - X\tilde{B})'(Y - X\tilde{B}) &+ (\tilde{B} - \bar{B})'A(\tilde{B} - \bar{B}) \end{aligned}$$

Ahora bien, este modelo asume a priori que la varianza de B depende de la varianza de Y , si se quiere modelar asumiendo independencia entre Σ y B , no existen conjugados naturales para el problema. Lo que sí se tiene, son conjugados condicionales, que se pueden utilizar en un método de simulación de Monte Carlo.

A.3. Muestras MCMC

A continuación se describen los dos tipos de muestras MCMC que se utilizarán para estimar el modelo desarrollado en el Capítulo 4.

A.3.1. Muestra de Gibbs

Método de muestra de Gibbs es una cadena de Markov que se obtiene iterando a través de un conjunto de distribuciones condicionales de π . Descomponiendo θ en p grupos disjuntos de parámetros,

$$\theta = \{\theta_1, \theta_2, \dots, \theta_p\}$$

el muestreo de Gibbs se define a través de un muestreo iterativo según p distribuciones, condicionales a los valores muestreados previamente en esa iteración:

1. Inicializar θ_0 y $r = 1$
2. Muestrear en cada iteración

$$\begin{aligned}\theta_1^r &\sim f_1(\theta_1 | \theta_2^{r-1}, \dots, \theta_p^{r-1}) \\ \theta_2^r &\sim f_2(\theta_2 | \theta_1^r, \theta_3^{r-1}, \dots, \theta_p^{r-1}) \\ &\vdots \\ \theta_p^r &\sim f_p(\theta_p | \theta_1^r, \theta_2^r, \dots, \theta_{p-1}^r)\end{aligned}$$

3. Actualizar $r = r + 1$
4. Repetir, hasta que $r = R$.

En este algoritmo f_1, \dots, f_p son las densidades condicionales apropiadas, derivadas de π :

$$f_i = \frac{\pi(\theta)}{\int \pi(\theta) d\theta_i}$$

Si el muestreo de Gibbs es irreducible, es decir todo subconjunto A , en el espacio vectorial de θ , con probabilidad positiva en la distribución estacionaria puede ser alcanzado en un número finito de pasos, el teorema 1 de Tierney [43] asegura convergencia a la distribución estacionaria desde casi todos los puntos de partida θ_0 .

En el caso del modelo de regresión multivariada sin aparente correlación, B independiente de Σ , estructurar un muestreo de Gibbs es bastante sencillo a partir de las distribuciones conjugadas condicionales. Para una mejor notación, se cambia la estructura de los datos, quedando el modelo de la siguiente forma:

$$y = X\beta + \varepsilon \quad \varepsilon \sim N(0, \Sigma \otimes I_n)$$

Donde:

$$y' = (y'_1, y'_2, \dots, y'_m), \quad X = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & X_m \end{bmatrix}, \quad \beta' = (\beta'_1, \beta'_2, \dots, \beta'_m), \quad \varepsilon' = (\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_m)$$

Las distribuciones a priori corresponden a:

$$\begin{aligned} p(\Sigma, B) &= p(\Sigma)p(B) \\ \Sigma &\sim IW(v_0, V_0) \\ \text{vec}(B) &\sim N(\bar{\beta}, A^{-1}) \end{aligned}$$

Y las distribuciones a posteriori corresponden a:

$$\begin{aligned} \Sigma|Y, X, \beta &\sim IW(v_0 + n, V_0 + S) \\ \beta|Y, X, \Sigma &\sim N(\tilde{\beta}, \tilde{\Omega}) \\ S &= (y - X\beta)'(y - X\beta) \\ \tilde{\beta} &= (X'(\Sigma^{-1} \otimes I_n)X + A)^{-1}(X'(\Sigma^{-1} \otimes I_n)y + A\bar{\beta}) \\ \tilde{\Omega} &= (X'(\Sigma^{-1} \otimes I_n)X + A)^{-1} \end{aligned}$$

Separando θ en dos bloques $\theta = \{B, \Sigma\}$, se samplea desde las distribuciones condicionales:

1. Inicializar $\theta_0 = \{B_0, \Sigma_0\}$ y $r = 1$
2. En cada iteración
 - Samplear

$$\begin{aligned} \beta_r|Y, X, \Sigma_{r-1} &\sim N(\tilde{\beta}_r, \tilde{\Omega}_r) \\ \Sigma|Y, X, \beta_r &\sim IW(v_0 + n, V_0 + S_r) \end{aligned}$$

3. Actualizar $r = r + 1$
4. Repetir, hasta que $r = R$.

A.3.2. Algoritmo de Metropolis

A pesar de que el sampleo de Gibbs es muy útil para algunos modelos, requiere conocer la forma de las distribuciones condicionales para simular a partir de ellas. En muchas ocasiones la estructura de estas distribuciones no son conocidas y es necesaria la utilización de un método más general, que permita simular desde cualquier distribución a posteriori. La idea del algoritmo de Metropolis es generar una cadena con sampleos de la distribución estacionaria, π , a través de modificaciones a una cadena de Markov relacionada, desde la cual sea más sencillo simular. Dado un sampleo de la cadena relacionada, a través de decisiones de aceptación o rechazo, se genera un sampleo de la distribución objetivo.

Dado un kernel de transición, definido por una función de transición $q(\theta, \vartheta)$, el algoritmo de Metropolis considera los siguientes pasos:

1. Inicializar θ_0 y $r = 0$
2. Samplear en cada iteración $\vartheta \sim q(\theta_r, \cdot)$.
3. Computar

$$\alpha(\theta_r, \vartheta) = \min \left\{ 1, \frac{\pi(\vartheta)q(\vartheta, \theta_r)}{\pi(\theta_r)q(\theta_r, \vartheta)} \right\}$$

4. Con probabilidad $\alpha(\theta_r, \vartheta)$ setear $\theta_{r+1} = \vartheta$, si no $\theta_{r+1} = \theta_r$.
5. Actualizar $r = r + 1$
6. Repetir, hasta que $r = R$.

Con probabilidad $(1 - \alpha)$ la cadena no cambiará de estado en una iteración, repitiendo el valor anterior. Estas repeticiones son necesarias para asegurar la reversibilidad de la cadena respecto a π . Definiendo el kernel, es decir la probabilidad de que la cadena actualice sus parámetros a valores dentro de un conjunto A dado que se encuentra en el punto θ , como:

$$K(\theta, A) = \int_A p(\theta, \vartheta) d\vartheta + r(\theta) \mathbb{1}_{\theta \in A}$$

Donde la función de transición de la cadena de Markov corresponde a:

$$p(\theta, \vartheta) = \alpha(\theta, \vartheta)q(\theta, \vartheta)$$

Dado que la cadena puede permanecer en el mismo punto por varias iteraciones, la probabilidad de moverse del punto θ a cualquier otro punto del espacio vectorial Θ , se expresa como:

$$\int_{\Theta} p(\theta, \vartheta) d\vartheta = \int_{\Theta} \alpha(\theta, \vartheta)q(\theta, \vartheta) d\vartheta$$

Y la probabilidad de permanecer en el mismo punto θ es $r(\theta)$. Este kernel es una mixtura entre una densidad continua, $\frac{p(\theta, \vartheta)}{1-r(\theta)}$, y un punto de masa en θ . Para comprobar la reversibilidad de la cadena en el tiempo se tiene que:

$$\begin{aligned} \pi(\theta)p(\theta, \vartheta) &= \pi(\theta)\alpha(\theta, \vartheta)q(\theta, \vartheta) \\ &= \pi(\theta)\min\left\{1, \frac{\pi(\vartheta)q(\vartheta, \theta)}{\pi(\theta)q(\theta, \vartheta)}\right\}q(\theta, \vartheta) \end{aligned}$$

Si ocurre que $\pi(\vartheta)q(\vartheta, \theta) \geq \pi(\theta)q(\theta, \vartheta)$, se tiene que

$$\alpha(\theta, \vartheta) = 1 \rightarrow \pi(\theta)p(\theta, \vartheta) = \pi(\theta)q(\theta, \vartheta)$$

y que

$$\alpha(\vartheta, \theta) = \frac{\pi(\theta)q(\theta, \vartheta)}{\pi(\vartheta)q(\vartheta, \theta)} \rightarrow \pi(\vartheta)p(\vartheta, \theta) = \pi(\vartheta)q(\vartheta, \theta) \frac{\pi(\theta)q(\theta, \vartheta)}{\pi(\vartheta)q(\vartheta, \theta)} = \pi(\theta)q(\theta, \vartheta)$$

Luego

$$\pi(\theta)p(\theta, \vartheta) = \pi(\theta)p(\vartheta, \theta)$$

Comprobándose la reversibilidad de la cadena en el tiempo, el caso contrario cuando $\pi(\vartheta)q(\vartheta, \theta) < \pi(\theta)q(\theta, \vartheta)$ es análogo, la cadena converge a la distribución estacionaria π objetivo.

Apéndice B

Apéndice B

B.1. Algoritmo de Estimación Jerárquico Bayesiano

1. Generar $\{\theta_i\} | Y_i, X_i, a_i, Z_i, \Psi, \delta, \Sigma_\theta$

$$f(\{\theta_i\} | Y_i, X_i, a_i, Z_i, \Psi, \delta, \Sigma_\theta) \sim N(\{\theta_i\} | \Psi, Z_i, \delta, \Sigma_\theta) L_i(Y_i)$$

Lo que significa que la función de densidad condicionada de θ_i corresponde a:

$$f(\{\theta_i\} | Y_i, X_i, a_i, Z_i, \Psi, \delta, \Sigma_\theta) \sim |\Sigma_\theta|^{-1} \exp[-1/2(\theta_i - \delta'Z_i)' \Sigma_\theta^{-1} (\theta_i - \delta'Z_i)] L_i(Y_i)$$

Donde $L_i(Y_i)$ es la función de verosimilitud desarrollada en la subsección anterior. Esta densidad no tiene forma cerrada, por lo que se utiliza el algoritmo de Metropolitan-Hasting para samplear con la distribución condicional de θ_i .

Algoritmo 2 Iteración del algoritmo de Metropolitan-Hasting

Sea $\theta_i^{(k)}$ el sampleo de θ_i aceptado en la iteración k

Samplear $\Delta\theta \sim N(0, \sigma^2\Delta)$

Sea $q = \theta_i^{(k)} + \Delta\theta$

$$A = \text{mín} \left\{ \frac{\exp[-1/2(q - \delta'Z_i)' \Sigma_\theta^{-1} (q - \delta'Z_i)] L_i(Y_i | \theta_i = q)}{\exp[-1/2(\theta_i^{(k)} - \delta'Z_i)' \Sigma_\theta^{-1} (\theta_i^{(k)} - \delta'Z_i)] L_i(Y_i | \theta_i^{(k)})}, 1 \right\}$$

Samplear $\varepsilon \sim U[0, 1]$

if $\varepsilon \leq A$ **then**

$$\theta_i^{(k+1)} = q$$

else

$$\theta_i^{(k+1)} = \theta_i^{(k)}$$

end if

return $\theta_i^{(k+1)}$

2. Generar $\delta | \{\theta_i\}, Z_i, \Sigma_\theta$

En el modelo desarrollado $\theta_i = \delta'Z_i + \varepsilon_{\theta_i}$, por lo que $\delta'Z_i$ representa la media de θ_i . Teniendo los valores de θ_i , se pueden estimar los valores de δ mediante un sampleo de Gibbs, condicional en la matriz de covarianzas Σ_θ tal como se revisó en el ejemplo de la Sección A.3.1.

$$vec(\delta)|\{\theta_i\}, Z_i, \Sigma_\theta \sim MVN(u_n, V_n)$$

Donde:

$$\begin{aligned} V_n &= [\Sigma_\theta^{-1} \otimes (Z'Z)^{-1}] \\ u_n &= V_n[\Sigma_\theta^{-1} \otimes Z'\Theta^* + V_0^{-1}u_0] \\ Z &= (Z'_1, Z'_2, \dots, Z'_m) \in \mathbb{R}^{m \times nz} \text{ es una matriz de covariables} \\ \Theta &= (\theta'_1, \theta'_2, \dots, \theta'_m) \in \mathbb{R}^{m \times n\theta} \text{ es una matriz que contiene a los vectores } \{\theta_i\} \\ \Theta^* &= vec(\Theta) \\ n\theta &= dim(\theta_i) \\ nz &= dim(Z_i) \end{aligned}$$

V_0 y u_0 son hiperparámetros del modelo, de manera de escoger priors difusos (no informativos) se escoge u_0 como un vector de ceros en $\mathbb{R}^{n\theta \cdot nz}$ y $V_0 = 100I_{n\theta \cdot nz}$.

3. Generar $\Sigma_\theta|\{\theta_i\}, Z, \delta$

De igual manera que en el paso anterior, se usando un sampleo de Gibbs se genera el valor de la matriz de covarianzas Σ_θ , condicional al valor de δ anteriormente obtenido.

$$\Sigma_\theta|\{\theta_i\}, Z, \delta \sim IW_{n\theta}(f_0 + N, G_0^{-1} + \sum_{i=1}^N (\theta_i - \delta'Z_i)'(\theta_i - \delta'Z_i))$$

Donde f_0 y G_0 son hiperparámetros del modelo, f_0 corresponde a los grados de libertad y G_0 es la matriz de escala de la distribución de Wishart inversa. Se definieron priors difusos escogiendo $f_0 = n\theta + 5$ y $G_0 = I_{n\theta}$.

Luego de samplear el vector de parámetros de efecto aleatorio, se procede a generar el vector de parámetros de efecto fijo Ψ .

4. Generar $\Psi|Y, X, a, \{\theta_i\}$

$$\Psi|Y, X, a, \{\theta_i\} \sim N(\Psi_0, V_{\Psi_0})L(Y)$$

Donde Ψ_0 y V_{Ψ_0} son priors difusos y $L(Y)$ es la función de verosimilitud para toda la muestra (todas las celdas y períodos). Esta distribución no tiene forma cerrada, por lo que se utiliza el algoritmo de Metropolitan-Hasting para samplear Ψ desde su distribución condicional. La probabilidad de aceptar un sampleo en la iteración $k+1$ es:

$$P(\text{aceptar } \Psi^{(k+1)}) = \min \left\{ \frac{\exp[-1/2(\Psi^{(k+1)} - \Psi_0)'V_{\Psi_0}^{-1}(\Psi^{(k+1)} - \Psi)]L(Y|\Psi^{(k+1)})}{\exp[-1/2(\Psi^{(k)} - \Psi_0)'V_{\Psi_0}^{-1}(\Psi^{(k)} - \Psi)]L(Y|\Psi^{(k)})}, 1 \right\}$$

B.2. Estadísticos para la Selección del Número de Estados

B.2.1. Densidad Marginal y Factor de Bayes

En un modelo paramétrico, el cálculo de la densidad marginal, $p(y|M_i)$ supone la integración de los parámetros, promediando la densidad marginal de la data dado los parámetros, sobre la densidad a priori:

$$p(y|M_i) = \int p(\theta|M_i)p(y|\theta, M_i)d\theta = \int \ell(\theta)p_i(\theta)d\theta$$

Cuando la estimación del modelo se realiza mediante un algoritmo MCMC, se pueden utilizar los valores de la verosimilitud obtenidos en cada iteración para obtener un valor estimado de esta densidad marginal, $p(y|M_i)$, usando el estimador de Newton y Raftery [33].

$$\hat{p}(y|M_i) = \left(\frac{1}{R} \sum_{r=1}^R \frac{1}{\ell(\theta^r|M_i)} \right)^{-1}$$

Donde R es el número de iteraciones del sampleo MCMC. El problema de este estimador, es que en muchas ocasiones, unos pocos valores del sampleo son los que determinan su valor. Si alguno de los valores de la verosimilitud en el sampleo es muy pequeño, puede cambiar el valor del estimador por completo, haciéndolo muy inestable. Este estimador es adecuado cuando la data es informativa respecto a los parámetros y la verosimilitud se mueve de forma estable dentro de un rango.

El Factor de Bayes es fácil de calcular una vez obtenidas las densidades marginales de la data dado los modelos, y consiste solo en su cociente.

$$FB = \frac{\hat{p}(y|M_1)}{\hat{p}(y|M_2)}$$

B.2.2. Criterio de la Desviación de la Información (DIC)

El criterio de la desviación de la información (DIC) [41] es un estadístico definido a partir de una estimación clásica de ajuste, penalizado por el número de parámetros incluidos en el modelo. Definiendo $D(\theta)$ como la desviación Bayesiana y p_D el número efectivo de parámetros en modelo en evaluación:

$$D(\theta) = -2\log\{p(y|\theta)\} + 2\log\{f(y)\} = -2\log\{\ell(\theta)\} + 2\log\{f(y)\}$$

$$p_D = \overline{D(\theta)} - D(\bar{\theta})$$

Donde $\bar{\theta}$ es el valor promedio de los parámetros obtenidos en el sampleo MCMC, $f(y)$ es una constante normalizadora y $\overline{D(\theta)}$ es el promedio de $D(\theta)$ sobre los valores de los parámetros en el sampleo. El criterio de ajuste DIC se define como:

$$DIC = D(\bar{\theta}) + 2p_D$$

Que es equivalente, usando las definiciones anteriores, a:

$$DIC = \overline{D(\theta)} + p_D$$

Escogiendo $f(y) = 1$ se obtiene el criterio DIC no estandarizado del modelo, con $D(\theta)$ equivalente a $-2\log\{\ell(\theta)\}$. En este caso el criterio DIC se expresa mediante la log verosimilitud promedio en el sampleo MCMC y la función log verosimilitud evaluada en el valor promedio de los parámetros.

B.2.3. Criterio de Selección de Modelos Markovianos (MSC)

La aplicación de algunos criterios clásicos de selección, como el criterio de selección Akaike, tiende a escoger modelos con más estados de los necesarios. Smith *et. al* [40] desarrollan

un criterio de selección especializado para el caso de regresiones con parámetros cambiantes según un proceso de Markov, de manera de escoger variables y estados para el modelo. Netzer [32] adapta el estadístico para considerar heterogeneidad en los parámetros y data no agregada. El criterio de selección de modelos markovianos (MSC, por sus siglas en inglés) corresponde a:

$$MSC = \overline{D(\theta)} + \sum_{s=1}^N \frac{\hat{T}_s(\hat{T}_s + \lambda_s K)}{\delta_s \hat{T}_s - \lambda_s K - 2}$$

Donde:

$\overline{D(\theta)}$ es equivalente al utilizado en el cálculo del criterio DIC.

$\hat{T}_s = \sum_{i=1}^M \sum_{t=1}^T P(S_{it} = s | y, M_i)$

$\lambda_s = N$ siguiendo la recomendación de Smith et. al, donde N es el número de estados del modelo.

$\delta_s = 1$ siguiendo la recomendación de Smith et. al, y finalmente

K es el número de variables incluidas en el modelo, tanto en la matriz de transición como en las probabilidades de emisión.

Apéndice C

Apéndice C

C.1. Transformación de datos

C.1.1. Condiciones Ambientales con Efecto en el Corto Plazo

Para obtener la secuencia de valores del atributo VD_{it} para cada unidad de estudio se realiza el procedimiento que se describe a continuación. Durante cada período, se contabiliza el número de policías que vigilan de forma directa cada celda de la grilla interna del simulador B :

$$hd_{ti} = \sum_{p \in P} vd_{tip} \quad i \in B, \quad t = 1, \dots, T$$

Donde P es el conjunto de policías y

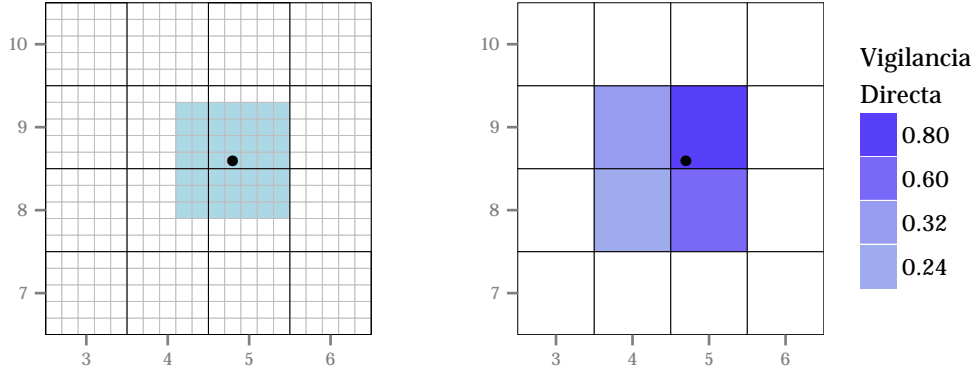
$$vd_{tip} = \begin{cases} 1 & \text{si la celda } i \text{ se encuentra bajo vigilancia directa de policía } p \text{ durante la semana } t. \\ 0 & \text{si no.} \end{cases}$$

Dado el radio de visión de la policía, todas las celdas en B en un rango de 3[c.s] desde el punto donde se encuentra el policía, se considera que se encuentran bajo vigilancia directa.

De manera de llevar esta información al nivel de la unidad de estudio $j \in A$, se considera el promedio de la variable vd_{tip} en la celda durante cada período.

$$VD_{tj} = \frac{1}{25} \sum_{i \in B(j)} hd_{ti} \quad k, \quad t = 1, \dots, T$$

Se puede estudiar un ejemplo en la Figura C.1, donde un policía es ubicado en la parte inferior de la celda (5,9). El primer cuadro, donde es visible la grilla interna del simulador, muestra el área con vigilancia directa, en color celeste. El segundo cuadro representa los valores de la variable VD_{tj} , que corresponde a la variable resumiendo la vigilancia directa, al agregar la información para todas las unidades de estudio afectadas por dicha vigilancia.



(a) Área con vigilancia directa del policía ubicado en la parte inferior de la celda (5,9). (b) Valores de la variable VD_{tk} al agregar la información.

Figura C.1: Construcción de la variable VD_{tk} , que determina la vigilancia directa a la que es sometida una celda.

En definitiva, la variable VD_{tk} recoge información tanto de la porción de la unidad de estudio que se encuentra vigilada, como del número de policías que se encuentran realizando vigilancia en el lugar. En esta aplicación ambas condiciones generan un impacto equivalente, es decir, la variable presenta un mismo valor si dos policías vigilan media celda cada uno, que un policía vigile él solo la celda completa.

De manera de considerar también el efecto desplazamiento, desarrollado en la Sección 2.4, que una medida de vigilancia podría provocar, es que se mide el grado de vigilancia en las celdas aledañas para comprender su efecto en el número de crímenes registrados. En particular, todas las celdas en B que se encuentren en un radio r tal que $3[c.s] < r \leq 8[c.s]$ de algún policía, se encuentran bajo vigilancia indirecta. De forma análoga a la vigilancia directa, se calcula primero el número de policías que vigilan las zonas aledañas de cada celda durante esa semana:

$$hi_{tj} = \sum_{p \in P} vi_{tjp} \quad j \in B, \quad t = 1, \dots, T$$

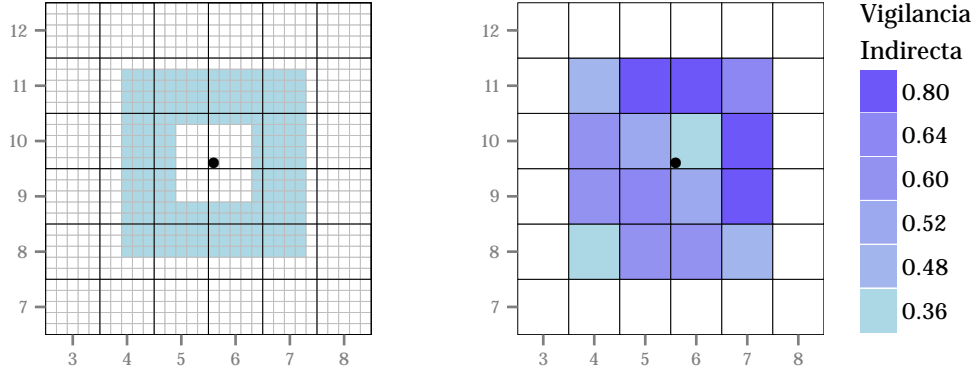
Donde P es el conjunto de policías y

$$vi_{tjp} = \begin{cases} 1 & \text{si celda } j \text{ se encuentra bajo vigilancia indirecta de policía } p \text{ durante la semana } t. \\ 0 & \text{en caso contrario.} \end{cases}$$

Para agregar la información al nivel de las unidades de estudio cada semana, se calcula el valor promedio de la variable hi_{tj} en cada celda:

$$VI_{tk} = \frac{1}{25} \sum_{j \in B(k)} hi_{tj} \quad k \in A, \quad t = 1, \dots, T$$

En la Figura C.2 se puede apreciar las celdas de la Grilla B aledañas a un punto de vigilancia, marcadas como lugares con vigilancia indirecta. En el cuadro siguiente se observa el valor de la variable VI para las unidades de estudios afectadas, al agregar la información.



(a) Área con vigilancia indirecta del policía ubicado en la parte inferior de la celda (6,10). (b) Valores de la variable VI_k al agregar la información.

Figura C.2: Construcción de la variable VI_k , que determina la vigilancia indirecta a la que es sometida una celda.

Al incluir estas dos variables de efecto en el corto plazo, VD y VI , en las probabilidades de emisión, se espera capturar el efecto de esta medida de vigilancia en el número de crímenes registrados y el efecto desplazamiento que pudieran tener.

C.1.2. Condiciones Ambientales con Efecto en el Largo Plazo

Para obtener la secuencia de valores de los atributos LD_{it} y Li_{it} para cada unidad de estudio se realiza el procedimiento que se describe a continuación. Durante cada período, se determina si cada unidad de estudio se encuentra bajo vigilancia directa e indirecta usando las variables hd_{sip} y hi_{sip} definidas en el caso de corto plazo:

$$ld_{tj} = \sum_{s=t-4, \dots, t-1} \mathbb{1}_{(hd_{sjp} \geq 1)} \quad j \in \text{Grilla B}, t = 5, \dots, T$$

$$li_{tj} = \sum_{s=t-4, \dots, t-1} \mathbb{1}_{(hi_{sjp} \geq 1)} \quad j \in \text{Grilla B}, t = 5, \dots, T$$

Se define entonces, una variable binaria, que indique si la celda se encuentra bajo una estrategia de vigilancia de largo plazo:

$$gd_{tj} = \begin{cases} 1 & \text{si } ld_{tj} \geq 4 \\ 0 & \text{en caso contrario} \end{cases}$$

$$gi_{tj} = \begin{cases} 1 & \text{si } li_{tj} \geq 4 \\ 0 & \text{en caso contrario} \end{cases}$$

Finalmente, es necesario agregar la información de manera de obtener una variable que indique el nivel de vigilancia de largo plazo para todas las unidades de estudio. En este caso, se

escogerá la proporción del área de la celda $k \in A$, que se encuentra bajo una vigilancia de largo plazo durante el período.

$$LD_{tk} = \frac{1}{25} \sum_{j \in B(k)} gd_{tj}$$

$$LI_{tk} = \frac{1}{25} \sum_{j \in B(k)} gi_{tj}$$

Se espera que estas dos variables, LD y LI , permitan entender la forma en que una vigilancia prolongada afecta las probabilidades que el lugar, y las celdas aledañas, cambien su estado de atractividad.

C.2. Resultados de la Estimación de Parámetros

C.2.1. Parámetros de Efecto Fijo

Tasa Esperada de Crímenes

La tasa esperada de crímenes, que determina las probabilidades de emisión, se define a partir de las condiciones ambientales observadas en cada unidad de estudio durante un período determinado. La Tabla C.1 da cuenta de los valores estimados para el caso de estudio del Capítulo 5, de los parámetros asociados a la vigilancia directa e indirecta de la policía.

Parámetro	Probabilidades de Emisión		Intervalo de Confianza	
	Promedio	Desv. Estándar	2.5 %	97.5 %
	Posteriori	Posteriori		
Intercepto				
β_{01} Intercepto Emisión - estado 1	0.060	0.011	0.04	0.084
β_{02} Intercepto Emisión - estado 2	1.652	0.008	1.635	1.667
β_{03} Intercepto Emisión - estado 3	2.882	0.011	2.860	2.901
β_{04} Intercepto Emisión - estado 4	3.665	0.008	3.649	3.681
Efecto Vigilancia en el Corto Plazo				
β_{11} Vigilancia Directa - estado 1	-2.124	0.035	-2.211	-2.071
β_{12} Vigilancia Directa - estado 2	-1.990	0.023	-2.030	-1.951
β_{13} Vigilancia Directa - estado 3	-1.693	0.052	-1.795	-1.625
β_{14} Vigilancia Directa - estado 4	-2.531	0.027	-2.590	-2.481
β_{21} Vigilancia Indirecta - estado 1	0.006	0.010	-0.015	0.026
β_{22} Vigilancia Indirecta - estado 2	-0.006	0.009	-0.024	0.012
β_{23} Vigilancia Indirecta - estado 3	0.065	0.006	0.053	0.075
β_{24} Vigilancia Indirecta - estado 4	0.120	0.004	0.112	0.128

Tabla C.1: Resultados de estimaciones de los parámetros de efecto fijo relacionados a las probabilidades de emisión. Se presenta el promedio y la desviación estándar, de cada uno de ellos, así como el intervalo en que se mueven los parámetros con un 95 % de confianza.

Matrices de Transición

Las matrices de transición entre estados se ven afectadas en el modelo propuesto, por la vigilancia continua de al menos cuatro semanas observada en cada unidad de estudio durante un período determinado. La Tabla C.2 da cuenta de los valores estimados para el caso de estudio del Capítulo 5, de los parámetros asociados a la vigilancia directa e indirecta de la policía.

Parámetro	Matriz de Transición		Intervalo de Confianza	
	Promedio Posteriori	Desv. Estándar Posteriori	2.5 %	97.5 %
Efecto Vigilancia en el Largo Plazo				
ρ_{11} Vigilancia Directa - estado 1	0.195	0.0320	0.144	0.255
ρ_{12} Vigilancia Directa - estado 2	-0.265	0.030	-0.326	-0.218
ρ_{13} Vigilancia Directa - estado 3	0.151	0.043	0.079	0.249
ρ_{14} Vigilancia Directa - estado 4	0.255	0.023	0.211	0.300
ρ_{21} Vigilancia Indirecta - estado 1	-0.137	0.023	-0.188	-0.097
ρ_{22} Vigilancia Indirecta - estado 2	0.079	0.04	-0.018	0.146
ρ_{23} Vigilancia Indirecta - estado 3	0.114	0.039	0.038	0.190
ρ_{24} Vigilancia Indirecta - estado 4	-0.638	0.045	-0.720	-0.552

Tabla C.2: Resultados de estimaciones de los parámetros de efecto fijo relacionados a la matriz de transición. Se presenta el promedio y la desviación estándar, de cada uno de ellos, así como el intervalo en que se mueven los parámetros con un 95 % de confianza.

Analizando los estadísticos, no se puede descartar la posibilidad de que posean un valor nulo con un 95 % de confianza. Esto sugiere que las variables incluidas no son capaces de explicar a cabalidad los cambios en la probabilidades de transición entre estados.

C.2.2. Parámetros de Efecto Aleatorio

Histograma Valores Promedio de Probabilidad de no Observar Delitos

La siguiente figura muestra la distribución, entre las celdas de las ciudad, del promedio de la probabilidad de observar cero delitos en un período. Se puede notar que existen al menos dos grupos de celdas, un segmento presentando probabilidad cercana a uno de no observar delitos y el otro, al contrario, con probabilidades casi nulas.

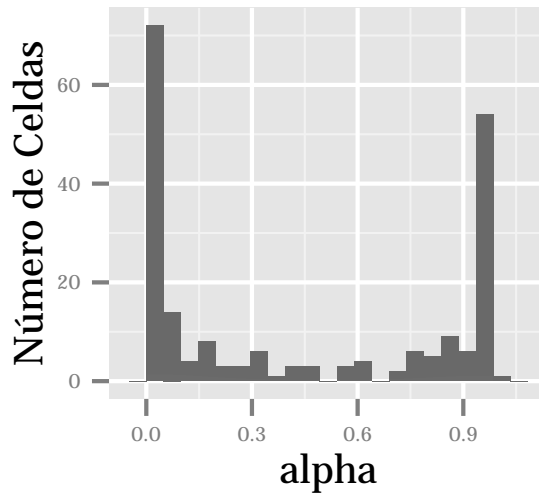
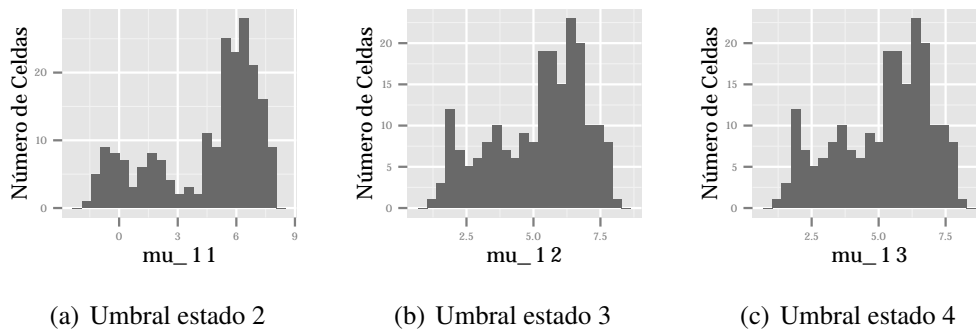


Figura C.3: Histograma del valor promedio de la probabilidad de no observar delitos α , entre todas las celdas

Histograma Valores Promedio de Umbrales de Transición

Las siguientes figuras muestran la distribución, entre las celdas de la ciudad, del promedio de los umbrales de transición entre estados μ_{sr} . Se puede notar que existen al menos dos grupos de celdas, con valores bastante diferenciados.



(a) Umbral estado 2

(b) Umbral estado 3

(c) Umbral estado 4

Figura C.4: Histograma del valor promedio de los umbrales de transición desde el estado 1, entre todas las celdas

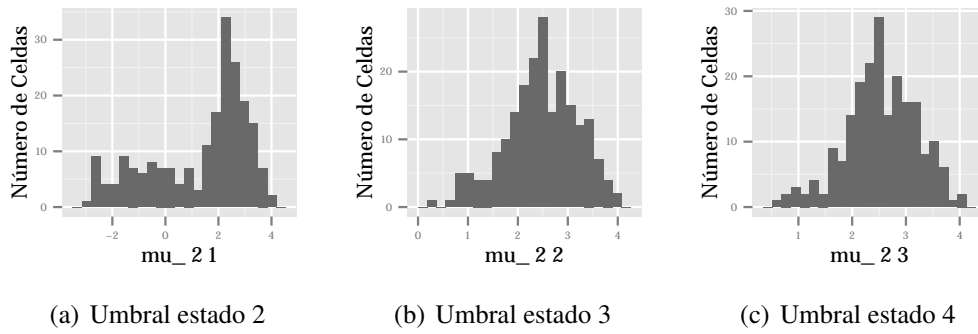


Figura C.5: Histograma del valor promedio de los umbrales de transición desde el estado 2, entre todas las celdas

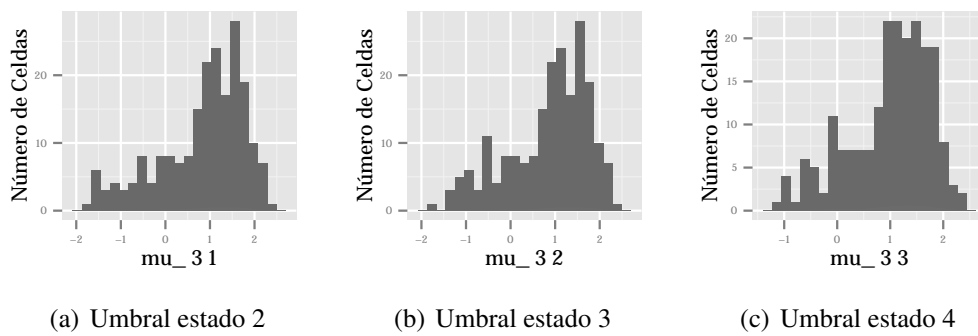


Figura C.6: Histograma del valor promedio de los umbrales de transición desde el estado 3, entre todas las celdas

C.2.3. Parámetros Capturando la Heterogeneidad

El número de edificios destinados a actividades específicas como escuelas, comercios, transporte, bancos y residencias fueron incluidos para intentar explicar la heterogeneidad de las celdas en los umbrales de transición y el parámetro α_i de la Poisson cero-inflada. Los parámetros asociados a estas características pueden observarse en la siguiente tabla.

Medias a Posteriori de Parámetros para Heterogeneidad Observable							
Parámetro	Intercepto	Banco	Comercio	Oficina	Residencia	Escuela	Transporte
$\tilde{\alpha}$	-1.99*	-1.19*	0.09	0.37*	-0.61	-1.30*	0.31
μ_{11}	3.98*	-0.15	0.00	0.30*	-0.83*	-0.08	0.08
μ_{12}	4.58*	0.00	0.03	0.23*	-0.70*	0.02	0.04
μ_{13}	4.58*	0.00	0.03	0.23*	-0.70*	0.02	0.04
μ_{21}	1.19*	-0.02	0.06	0.24*	-0.61*	-0.28	-0.18*
μ_{22}	2.17*	0.25*	0.12*	0.13*	-0.44*	-0.10*	-0.25*
μ_{23}	2.28*	0.25*	0.11*	0.11*	-0.36*	-0.09*	-0.25*
μ_{31}	0.64*	-0.15*	0.03	0.02	-0.46*	-0.24*	0.12*
μ_{32}	0.67*	-0.14*	0.03	0.02	-0.44*	-0.24*	0.13*
μ_{33}	0.89*	-0.15*	0.02	-0.02	-0.27*	-0.23*	0.14*
μ_{41}	-0.65*	-0.64*	-0.02*	-0.07*	0.65*	-0.44*	0.49*
μ_{42}	-0.65*	-0.64*	-0.02*	-0.07*	0.65*	-0.44*	0.49*
μ_{43}	-0.64*	-0.64*	-0.02*	-0.07*	0.65*	-0.44*	0.49*

* El intervalo de confianza del 90% no incluye el 0.

Tabla C.3: Parámetros relacionados a las características físicas que capturan la heterogeneidad entre las celdas.

C.3. Modelos de Comparación para la Validación

Los cuatro modelos que se describen a continuación corresponden a los utilizados para comparar, en la Sección 6.2.2, el ajuste y capacidad de predicción del modelo HMM para el crimen estimado en el Capítulo 5.

C.3.1. HMM sin Heterogeneidad

De manera de clarificar el beneficio de incorporar diferencias entre las celdas, se estima el mismo modelo especificado en la Sección 5.5 pero considerando que los umbrales de transición, μ_{is} , y la probabilidad de no observar delitos en la Poisson cero-inflada, α_i , son los mismos para todas las celdas. El modelo es estimado considerando los mismos cuatro estados de atractividad escogidos para el modelo HMM con heterogeneidad.

C.3.2. Clases Latentes

El modelo de Clases Latentes asume que las celdas no cambian de estado a lo largo del tiempo, lo que da origen a la especificación generada en la Sección 4.3.2 del Capítulo 4. Donde la matriz de transición se considera una matriz identidad y las probabilidades de pertenecer a un estado de atractividad corresponden a:

$$\pi_{is} = \frac{\exp(\mu_{is})}{1 + \sum_{k=1}^{N-1} \exp(\mu_{ik})} \quad s = 1, \dots, N-1$$

$$\pi_{iN} = 1 - \sum_{k=1}^{N-1} \pi_{ik}$$

En este modelo se considera heterogeneidad no observable en los parámetros μ_{is} y α_{is} . Si las secuencias de crímenes muestran variaciones poco significativas durante el tiempo, la verosimilitud de este modelo debería ser similar al del HMM con heterogeneidad.

C.3.3. Regresión de Poisson

El modelo Regresión de Poisson es el más simple de todos, considerando heterogeneidad en el parámetro de la Poisson cero-inflada α_i , no considera estados de atractividad en su estructura, o que es lo mismo, un único estado. El número de crímenes en una celda durante un período se distribuye como:

$$Y_{it} \sim \begin{cases} 0 & \text{con probabilidad } \alpha_i \\ \text{Poisson}(\lambda_{it}) & \text{con probabilidad } (1 - \alpha_i) \end{cases}$$

Definiendo α_i y λ_{it} tal como en el modelo desarrollado en la Sección 4.3.2:

$$\hat{\alpha}_i = \delta_{\alpha}' Z_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$$

$$\alpha_i = \frac{\exp(\hat{\alpha}_i)}{1 + \exp(\hat{\alpha}_i)}$$

$$\log(\lambda_{it}) = \bar{\beta}_0 + x_{it}' \beta$$

C.3.4. Modelo de Persistencia

De manera de comparar el modelo desarrollado en este trabajo con alguno con una formulación diferente, pero que utilice el mismo tipo de información, se escoge el modelo desarrollado por Mohler *et al.* [31]. Este modelo considera que existen dos tipo de eventos criminales, los eventos ambientales, que ocurren de forma independiente sobre la ciudad, y que generan eventos replicas a su alrededor, al detectar los delincuentes el potencial de los lugares, generando concentraciones de crímenes en la zona.

Los crímenes se asumen, siguen una distribución de Poisson con una tasa esperada que se distribuye en la celda i como:

$$\lambda(i, t) = \mu_i + \sum_{t_k < t} g(i, t - t_k)$$

Donde μ_i es la tasa esperada de eventos ambientales en la celda i y $g(i, t)$ es una kernel paramétrico que modela la distribución de los eventos réplica. En particular se define:

$$g(i, t) = K_0 w \exp(-wt)$$

El modelo es estimado maximizando la verosimilitud mediante una simulación de Monte Carlo.

C.4. RPS: Ranked Probability Score

Este puntaje para medir el ajuste del modelo se puede calcular mediante la siguiente formulación:

$$rps(P,x)_{it} = \sum_{k=0}^{\infty} (P_{itk} - 1_{y_{it} \leq k})^2$$

Donde P_{itk} es la probabilidad de que el número de crímenes en la celda i sea k durante el período t , y por otro lado, y_{it} es el número de crímenes observado en la celda i durante el período t . Finalmente el puntaje para el modelo, corresponde al promedio obtenido entre todos los períodos y celdas.

$$RPS = \frac{1}{M \times T} \sum_i \sum_t rps(P,x)_{it}$$