UNIDAD IZTAPALAPA

División de Ciencias Básicas e Ingeniería
Posgrado en Ciencias (Matemáticas Aplicadas e Industriales)

"Un modelo de particiones aleatorias para observaciones espaciales"

Tesis que presenta

Luis Enrique Lara Pérez
Matrícula: 2212801666
Correo: luiz.lp.09@gmail.com

Para obtener el grado de Maestro en Ciencias (Matemáticas Aplicadas e Industriales)

Directores de Tesis:

Dra. Blanca Rosa Pérez Salvador
Dr. Asael Fabian Martínez Martínez

Jurado:

Prsidente: Dr. Andrey Novikov
Secretario: Dr. Asael Fabian Martínez Martínez
Vocal: Dra. Blanca Rosa Pérez Salvador
Vocal: Jose Antonio Perusquia Cortes

Iztapalapa, Ciudad de Mexico,
26 de Marzo del 2024

Agradezco a mis padres Víctor y Virginia, por todo el apoyo recibido, a mis maestros de la Maestría en Ciencias Matemáticas e industriales en especial al Dr. Fabian y la Dra. Blanca por su paciencia y enseñanzas.

II

Índice general

Introducción

El problema fundamental del progreso científico, y uno fundamental en la vida diaria, es el de aprender de la experiencia. El conocimiento obtenido de esta manera es precisamente una descripción de lo que ya hayamos observado, pero una parte consiste en la realización de inferencias de la experiencia pasada para predecir la experiencia futura. Una de las principales tareas de la estadística es poder realizar algún tipo de inferencia sobre la distribución de probabilidad que mejor ayude a modelar el fenómeno que se desea estudiar todo esto con ayuda de un conjunto de observaciones. Debemos observar que los modelos que se usan deberán tener algún sustento probabilístico esto para que se puedan tomar como válidos.

Uno de los enfoques estadísticos que estaremos tomando es el enfoque bayesiano el cual trabaja con la actualización de la evidencia considerando los conocimientos adquiridos previos a una investigación (llamada inicial) y utiliza el teorema de Bayes para actualizar la información. El enfoque bayesiano ha tomado una gran fuerza en los últimos años, debido a su gran efectividad para resolver problemas que no se pueden atacar con otros métodos.

Por otro lado la estadística espacial estudia procesos que ocurren en el espacio y de los que, de una manera u otra, se conoce su ubicación. Trata de los métodos estadísticos que utilizan relaciones espaciales (tales como distancia, longitud, orientación, centralidad, entre otros) directamente en sus cómputos matemáticos. La estadística espacial se emplea para toda una variedad de tipos de análisis, modelado y predicciones. Entre los muchos tipos de estadísticas espaciales están las estadísticas descriptivas, inferenciales, exploratorias, geoestadísticas y econométricas.

Adicionalmente el análisis de conglomerados lo contribuye un conjunto de técnicas mediante las cuales se clasifican datos en grupos relativamente homogéneos llamados conglomerados.

Con base a todo lo que se ha mencionados y sabiendo que este trabajo se

enfocará principalmente en analizar datos espaciales para encontrar agrupaciones sobre redes lineales, entonces algunos problemas de interés están dados por las siguientes preguntas ¿Será muy difícil poder encontrar agrupaciones teniendo datos espaciales?¿Qué tanto afectara trabajar con redes lineales en estadística espacial? ¿De qué manera podremos hacer uso de la estadística bayesiana para encontrar agrupaciones?

En el primer capítulo se da una breve introducción a la estadística bayesiana, posteriormente una vez entendido cómo es que podemos usar la estadística desde el enfoque bayesianos y algunas de sus aplicaciones se estudia un par de modelos probabilísticos.

En la sección 1.3 podremos introducir un poco a las bases para los algoritmos Monte Carlo vía cadenas de Márkov (MCMC) más utilizados en la estadística bayesiana que son: el Metropolis-Hastings y un caso partícular, Gibbs-Sampler.

Terminando el capítulo mostrando el tipo de datos sobre los que vamos a estar trabajando y dedicarnos a comprender un concepto que es fundamental para el análisis de este tipo de datos el cual es el de conglomerados partiendo de la definición de partición (la cual será usada para el modelo).Una vez se entiende este concepto clave pasamos a analizar y explicar el proceso por el cual se pueden generar estos conglomerados y algunas características que posee, como poder calcular la probabilidad de obtener cierta partición u obtener el conjunto de todas las particiones.

El segundo capítulo está dedicado al planteamiento, desarrollo y análisis del modelo, pasando así a la sección 2.5 donde se realiza un análisis de sensibilidad de algunos parámetros para ver los efectos que tienen sobre el modelo, esto con datos simulados.

El tercer capítulo se tiene la aplicación del modelo que se propuso en el capitulo dos esto para un tema de inseguridad en la Ciudad de México tomando datos reales recabados de la página de la PGR.

Finalmente, el último capítulo consta de las conclusiones finales sobre el trabajo realizado así como futuras aplicaciones y líneas de investigación.

Se incluye un apéndice que sirve como guía rápida para la elaboración de una red, como cargar los datos reales y una pequeño ejemplo de cómo se realizó la simulación de datos.

Capítulo 1

Marco teórico

Uno de los propósitos de la estadística es el estudio de fenómenos desarrollados en ambiente de incertidumbre (aleatorios). El estudio se lleva a cabo a partir de una cierta cantidad de observaciones que se recolectan y analizan con el fin de poder describir dicho fenómeno, gracias a la estadística logramos realizar inferencia para saber cuál es su comportamiento. Para esto podemos hacer estimaciones sobre parámetros y pruebas de hipótesis lo que nos ayudará a obtener conclusiones.

Uno de los enfoques de la estadística es la llamada estadística bayesiana la cual surgió gracias al matemático Thomas Bayes, que nació en Londres, Inglaterra en 1702 y murió en Tunbridge Wells, Kent, Inglaterra el 17 de Abril de 1761.

Bayes fue uno de los primeros en utilizar la probabilidad inductivamente y establecer una base matemática para la inferencia probabilística. Actualmente se ha desarrollado una poderosa teoría que ha conseguido notables aplicaciones en las más diversas áreas del conocimiento. Estudió el problema de la determinación de la probabilidad de las causas a través de los efectos observados. El teorema que lleva su nombre se refiere a la probabilidad de un suceso condicionado por la ocurrencia de otro suceso. Más específicamente, con su teorema se resuelve el problema conocido como "la probabilidad inversa"; esto es, valorar probabilísticamente las posibles condiciones que rigen bajo el supuesto que se ha observado cierto suceso.

El principal objetivo de este teorema es determinar la probabilidad de que ocurra el evento A, dado que ha ocurrido B, en función de la probabilidad de que ocurra B dado que ha ocurrido A, de la probabilidad de A y de la probabilidad de B. Este se describe mediante el siguiente resultado:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}, \tag{1.1}$$

este teorema tiene muchas aplicaciones como en la tecnología, en los negocios y la medicina por mencionar algunos. A continuación se presenta un ejemplo del uso de este teorema en un contexto deportivo.

El Ejemplo 1 ilustra el funcionamiento del teorema de Bayes y la relación que tienen las probabilidades para poder resolver el problema de la probabilidad inversa.

Ejemplo 1. *En una escuela, la probabilidad de que a un alumno seleccionado al azar le guste el fútbol es del* $50\,\%$*, por otro lado la probabilidad de que a un alumno le guste el basquetbol es del* $26\,\%$*. Además, sabemos que la probabilidad de que a un alumno le guste el basquetbol dado que le guste el fútbol es del* $30\,\%$*. Deseamos calcular la probabilidad de que a un alumno le guste el fútbol, dado que le guste el basquetbol.*

Solución. Sabemos que hay 2 eventos :

- A : Un alumno le gusta el fútbol
- B : Un alumno le gusta el basquetbol

Además, tenemos la siguiente información

- $P(A) = 0.5$
- $P(B) = 0.26$
- $P(B \mid A) = 0.3$

Por lo que basta calcular $P(A \mid B)$,usando el teorema de Bayes obtenemos:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{0.3 * 0.5}{0.26} = 0.5769.$$

de esta manera la probabilidad de que a un alumno le guste el fútbol, dado que le guste el basquetbol es del $57.69\,\%$.

□

Desde la perspectiva bayesiana de la estadística todas las cantidades que se desconocen van a hacer tratadas como variables aleatorias. Se asumirá que las observaciones son generadas en base a variables aleatorias condicionalmente independientes con una distribución de probabilidad en común, la cual depende de parámetros desconocidos que va a pertenecer a alguna familia paramétrica de funciones de distribución, indexada por un parámetro con dimensión finita donde este último parámetro es desconocido. Para una revisión más profunda sobre el tema el lector puede dirigirse a [1, pag.17-28] y [2, pag.1-28]. A continuación abordaremos los conceptos principales sobre la estadística bayesiana así como los elementos que la conforman.

1.1. Estadística bayesiana

El proceso de inferencia estadístico, tanto en el enfoque frecuentista como en el enfoque bayesiano, se basa en la presencia de observaciones y cuyos valores son inicialmente inciertos y que son descritos por una distribución de probabilidad con función de densidad o función de masa $f_{y|\theta}(y \mid \theta)$. La cantidad θ es un parámetro que indexa la familia de posibles distribuciones para las observaciones, éste representa características de interés que el investigador desearía conocer con la finalidad de obtener una descripción completa del proceso.

La diferencia entre el enfoque frecuentista y bayesiano, ya que en el enfoque frecuentista no admite la incorporación de alguna información inicial (el cual podría ser cualquier tipo de conocimientos previos sobre el tema por parte del investigador). Por otro lado el enfoque bayesiano permite la incorporación, de manera formal, este tipo de información inicial a través de una distribución con función de densidad $f_\theta(\theta)$, sin importar que la información sea imprecisa, mucha o poca.

Recordemos que trabajaremos con observaciones los cuales son descritas por una distribución de probabilidad con función de densidad, por lo que para nuestros intereses tendremos que escribir la ecuación (1.1) en términos de estas distribuciones, de esta forma obtenemos la ecuación (1.2).

$$f_{\theta|y}(\theta \mid y) = \frac{f_{y|\theta}(y \mid \theta) f_\theta(\theta)}{f_y(y)}. \tag{1.2}$$

Para fines prácticos estaremos denotando las funciones de densidad como

$p(y \mid \theta)$ entonces podremos reescribir la ecuación (1.2) como se muestra en ecuación (1.3).

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)}. \tag{1.3}$$

Los elementos usados en este enfoque bayesiano a partir del teorema de Bayes son:

- θ el cual es un parámetro que representa características de interés que desearíamos conocer con la finalidad de obtener una descripción completa de algún fenómeno .
- $p(\theta)$ denota la distribución inicial, la cual modela conocimientos previos sobre el parámetro.
- $p(y \mid \theta)$ es la función de verosimilitud de θ dados los datos
- $p(\theta \mid y)$ la distribución final de θ dados los datos, esta es la resultante de combinar la información contenida en distribución inicial y la verosimilitud de los datos.

El objetivo es obtener una distribución final $p(\theta \mid y)$ la cual actualizará a $p(\theta)$ dado los datos y. Hasta el momento tomamos a y como una sola observación pero en general no tendremos sólo una observación sino n observaciones, es decir, $y \in R^n$ que será un vector de observaciones, entonces ahora tomaremos $y = (y_1, \ldots, y_n)$.

Por lo que ahora tendremos un problema con la verosimilitud ya que ahora cada una de las observaciones dependerán del parámetro θ, por lo que introduciremos la Definición 1.

Definición 1 (independencia condicional)**.** *Las variables aleatorias $y_1, \ldots, y_n$ se denominan condicionalmente independientes dado θ, si se tiene:*

$$p(y_1, \ldots, y_n \mid \theta) = p(y_1 \mid \theta) \times \cdots \times p(y_n \mid \theta) = \prod_{i=1}^{n} p(y_i \mid \theta).$$

Entonces decimos que las variables aleatorias $y_1, \ldots, y_n$ se generan a partir de un proceso común dado θ.

Por lo que el problema de la verosimilitud para un vector de observaciones se soluciona, además en este caso diremos que $y = (y_1, \ldots, y_n)$ son condicionalmente independientes e idénticamente distribuidas (iid) condicionadas a θ y lo denotaremos como se muestra en la ecuación (1.4)

$$y_i \mid \theta \sim p(y_i \mid \theta) \qquad [iid] \qquad i = 1, .., n. \tag{1.4}$$

Gracias a estos conceptos básicos de la estadística bayesiana en conjunto de la Definición 1 podemos relacionar la distribución inicial y la distribución final para distribuciones de probabilidad con función de densidad.

Podemos observar en la ecuación (1.3) que la marginal $p(y)$ es una constante pues no depende de θ. Por lo que podemos tener:

$$p(\theta \mid y) = \frac{p(\theta)p(y \mid \theta)}{p(y)} = \frac{p(\theta)p(y \mid \theta)}{\int p(y \mid \theta)p(\theta)d\theta},$$

$$p(\theta \mid y) \propto p(\theta)p(y \mid \theta).$$

Lo cual denota una equivalencia salvo una constante (la cual es la marginal $p(y)$), es decir, ambos valores son proporcionales Este ultimo es usado en ocasiones o por simplicidad pues en la mayoría de los casos no se necesita calcular $p(y)$ (esto cuando conocemos esta constante).

Cuando conocemos la distribución final se podrá realizar cualquier tipo de inferencia sobre el parámetro θ, por ejemplo podremos calcular la media o varianza.

Una vez entendido cuales son los elementos de la estadística bayesiana, analizaremos un modelo en particular.

1.1.1. Modelo Poisson

La distribución de Possion fue creada por el matemático y filósofo francés Simeón-Denis Poisson en el siglo XVII, esta es una distribución de probabilidad discreta que esta relacionada a la ocurrencia de algún evento, es decir, esta distribución solo necesita saber los eventos y su frecuencia media de que ocurra para que de esta manera podamos conocer la probabilidad de que ocurra dicho evento. En general esta distribución se utiliza cuando deseamos

calcular alguna probabilidad donde requerimos contar el número de veces que produce algún suceso.

Por lo antes mencionado el modelo de conteo más utilizado es el "modelo Poisson",ya que algunas medidas, como lo son el numero de carros accidentados en alguna ciudad o el número de asaltos, tienen valores que son números enteros y obtenemos ocurrencia la cual es una de las principales características de este modelo.

Se dice que una variable Y tiene distribución de probabilidad de Poisson con parámetro θ si y solo si

$$P(Y = y) = \frac{\theta^y}{y!}e^{-\theta} \qquad y = 0, 1, ... \qquad \theta > 0,$$

si Y es una variable aleatoria que posee una distribución Poisson con parámetro θ, entonces se tiene que $E[Y \mid \theta] = Var[Y \mid \theta] = \theta$.

Para comprender mejor los conceptos que se han visto hasta ahora y cuál es el funcionamiento del teorema de Bayes procederemos a realizar un pequeño ejemplo donde involucre el modelo de Poisson.

Supongamos que tenemos datos y (recordemos que $y = (y_1, \dots y_n)$) y los modelamos con una distribución $Poisson(\theta)$, ahora para la distribución inicial tomaremos una gamma $p(\theta) = gamma(\theta \mid a, b)$ esto para hacer uso del teorema de Bayes, entonces tenemos:

- Modelo de muestreo Poisson

$$y_i \mid \theta \sim Poisson(\theta) \qquad iid \qquad i = 1, ..., n$$

- Distribución inicial

$$\theta \sim gamma(a, b)$$

Se dice que una variable θ tiene distribución de probabilidad $gamma(a, b)$ si:

$$p(\theta \mid a, b) = \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta}, \qquad a, b > 0.$$

Si θ es una variable aleatoria que posee una distribución gamma con parámetro (a, b), entonces obtenemos que $E[\theta] = a/b$ y $Var[\theta] = a/b^2$.

Deseamos obtener mediante el teorema de Bayes una distribución final de la forma

$$p(\theta \mid y) \propto p(\theta)p(y \mid \theta) \tag{1.5}$$

Primero veamos que la verosimilitud $p(y_1, .., y_n \mid \theta)$ de $(y_1, .., y_n)$ asociada a θ está dada por:

$$p(y_1, .., y_n \mid \theta) = \prod_{i=1}^{n} \frac{\theta^{y_i}}{y_i!} e^{-\theta} = \frac{\theta^{\sum y_i}}{\prod y_i!} e^{-n\theta}$$

Por otro lado sabemos que la inicial $p(\theta)$ tiene distribución gamma por lo que (1.5) queda de la forma

$$p(\theta \mid y) \propto \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \frac{\theta^{\sum y_i}}{\prod y_i!} e^{-n\theta}$$

Se puede observar que $\frac{b^a}{\prod y_i! \Gamma(a)}$ las podemos tomar como constantes pues no dependen de θ, obteniendo así:

$$p(\theta \mid y) \propto \theta^{a-1} e^{-b\theta} \theta^{\sum y_i} e^{-n\theta}$$

$$= \theta^{\sum y_i + a - 1} e^{-\theta(n+b)}$$

Si tomamos $a_n = \sum y_i + a$ y $b_n = b + n$ se deduce:

$$p(\theta \mid y) \propto \theta^{a_n - 1} e^{-b_n \theta}$$

Es decir que podemos concluir que nuestra final tiene una distribución gamma con parámetros a_n y b_n

$$\theta \mid y \sim gamma(a_n, b_n)$$

Ya entendimos cómo se usa la función de verosimilitud y una distribución inicial para poder obtener nuestra distribución final, por lo que ahora realizaremos un ejemplo práctico para poder ver cómo es que funciona. Este ejemplo fue tomado de [2, pag. 39].

Ejemplo 2 (Caso de accidentalidad)**.** *En cierta localidad el número de accidentes anuales de automóvil se puede modelar como una v.a.* $Y \sim Poisson(\theta)$*, donde* θ *depende del conductor y asumimos una distribución inicial gamma digamos* $gamma(a = 10, b = 2)$*. Si se observan dos registros de accidentes, supongamos que los datos son* $y_1 = 12$ *y* $y_2 = 10$*, ¿Cuál es la distribución final de* θ*, dado* $y_1 = 12$ *y* $y_2 = 10$*?.*

Solución. Al saber que la distribución inicial es $gamma(10, 2)$ podemos deducir cual es el número promedio de accidentes anuales. Recordemos que la esperanza de una distribución gamma con parámetros (a, b) esta dada por a/b, entonces $E[\theta] = 10/2 = 5$.

Observamos que tenemos una distribución inicial gamma y el problema se puede modelar como una distribución Poisson, es decir, la verosimilitud será Poisson, entonces podremos hacer uso de lo realizado anteriormente, es decir, podremos usar la ecuación (2.1) para obtener nuestra distribución final, de esta manera se tiene:

$$\theta \mid y \sim gamma(a_n, b_n),$$

donde $y = \{y_1, y_2\}$, $a_n = y_1 + y_2 + a$ y $b_n = b + n$.

Entonces obtenemos:

$$\theta \mid y \sim gamma(32, 4),$$

Ahora observemos que los parámetros de nuestra distribución inicial se actualizaron por lo que podríamos volver a obtener la esperanza de los accidentes anuales, entonces $E[\theta \mid y] = 32/4 = 8$.

□

Del ejemplo anterior podemos resaltar que una vez obtenida la distribución final se puede hacer el calculo de la esperanza y varianza.

Ahora en este ejemplo particular al actualizar la información del parámetro θ (para este ejemplo son los parámetros de la distribución gamma), podemos realizar el cálculo de la esperanza que esta dado por $E[\theta \mid y] = 8$

y compararla con la esperanza obtenida inicialmente podemos observar que hubo un cambio.

1.2. Modelos jerárquicos

Siguiendo con el Ejemplo 2 del caso de accidentalidad dados los datos que se tienen que dependen del parámetro θ y que corresponde a una localidad se pudo obtener información, pero podríamos extender esta idea aún más, podríamos realizar inferencia sobre mas localidades supongamos que son k localidades y cada una cuenta con m_k datos, entonces con base en los resultados obtenidos obtendríamos $\theta_1, ..., \theta_k$ parámetros, ahora estos parámetros los utilizaremos como datos los cuales se podrán modelar a través de una distribución π con hiperparámetro ϕ, este hiperparámetro es el que nos podrá dar información sobre las k localidades.

A este tipo de modelos son llamados modelos jerárquicos e involucran precisamente varios parámetros de tal manera que uno de los parámetros depende de los valores de otros. En otras palabras, los modelos jerárquicos ayudan a modelar problemas en los cuales tenemos uno o varios parámetros los cuales dependen de otro parámetro. Veremos que esto se logra si usamos una distribución inicial en la que los θ_k's se ven como una muestra de una distribución poblacional común.

Esta estructura jerárquica resulta muy útil al analizar modelos con al menos un par de parámetros que dependan uno del otro. Por ejemplo en el estudio de accidentes automovilísticos para pronosticar el número medio de accidentes anuales podemos considerar factores a nivel local, nacional, regional, etc.

Precisamente los modelos jerárquicos involucran varios parámetros de tal manera que unos de los parámetros dependen de los valores de otros parámetros como se muestra en la Figura 1.1. Para más información sobre el tema el lector puede dirigirse a [3, Cap. 5].

El modelo jerárquico que corresponde a la Figura 1.1 es el que se muestra a continuación, en la primer fila podremos observar la distribución de nuestra muestra (los cuales podríamos pensarlos como los accidentes registrados), posteriormente en se muestra la distribución de los parámetros (las cuales podríamos pensar las k localidades) y finalmente en la ultima fila podremos observar la distribución del hiperparametro.

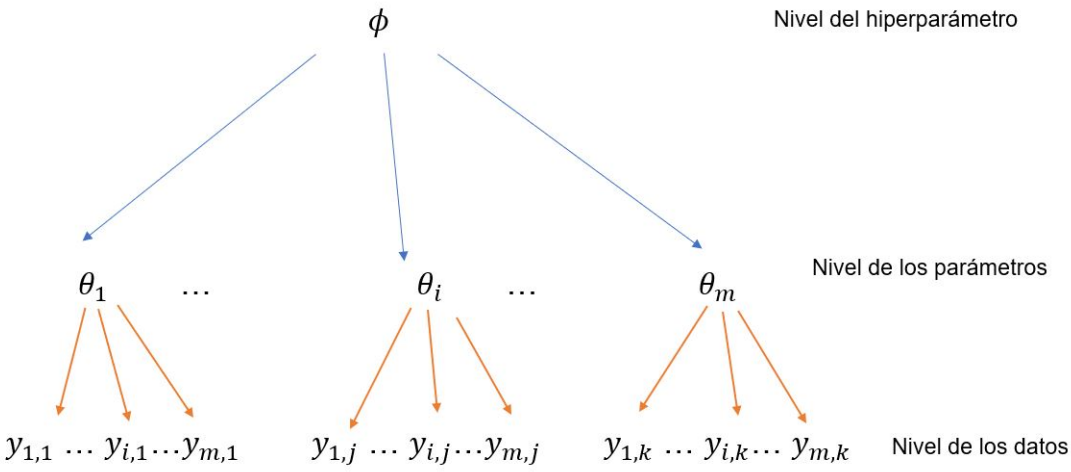


Figura 1.1: Modelo Jerárquico.

$$\begin{array}{lll} y_{i,j} \mid \theta_i \sim Poisson(\theta_i) & [iid] & i = 1,..,m; j = 1,..,k \\ \theta_i \mid \phi \sim gamma(\phi) & [iid] & \\ \phi \sim \pi & & \end{array} \tag{1.6}$$

1.3. Monte Carlo vía cadenas de Márkov

En esta sección presentaremos un método de simulación para generar muestras de cualquier distribución final, conocido como Monte Carlo vía cadenas de Márkov (MCMC). El método MCMC simula valores de manera iterativa, observaciones de distribuciones multivariadas que difícilmente se podrían simular usando otros métodos numéricos. La idea general se basa en construir una cadena de Márkov tal que esta pueda ser fácil de simular y la cual su distribución estacionaria va a corresponder a la distribución que tendremos como objetivo.

1.3.1. Monte Carlo

Uno de los problemas que aparecen en estadística es la integración, el cual está ligado a la obtención de valores esperados. Dado que no siempre podremos obtener el resultado de forma analítica, entonces una alternativa será obtener el resultado de forma numérica.

La idea básica de cualquier método de Monte Carlo es escribir la integral que se desea como el valor esperado de alguna función con respecto a alguna distribución de probabilidad. Para esto necesitamos encontrar una variable aleatoria X con soporte en $[a, b]$ y distribución π, con densidad g y una función $h = \frac{f(x)}{g(x)}$ tales que cumpla

$$I = \int_a^b f(x)dx = \int_a^b h(x)g(x)dx = E[h(X)] \tag{1.7}$$

Ahora si tenemos $x_1, \ldots x_n$ una muestra de tamaño n con distribución π y usamos la ley de los grandes números podremos hacer una estimación de (1.7), ya que

$$\widehat{I} = \frac{1}{n}\sum_{i=1}^{n} h(x_i) \to E[h(X)] \qquad \text{cuando} \qquad n \to \infty$$

Este método es mucho más factible pues una de sus ventajas es que se puede extender fácilmente al caso multidimensional.

Como se mencionó anteriormente en el análisis bayesiano se requiere realizar integraciones sobre distribuciones de probabilidad de dimensiones grandes para poder hacer inferencias sobre los parámetros de algún modelo y una solución serán los modelos MCMC la cual es una herramienta muy efectiva para la modelación estadística. Ya vimos un poco de lo que trata el método de Monte Carlo, ahora introduciremos algunos conceptos sobre cadenas de Márkov los cuales serán utilizados para comprender de mejor manera el método MCMC.

1.3.2. Cadenas de Márkov

Ya sabemos que el método de Monte Carlo es utilizado cuando no se pueden obtener integrales directamente de una distribución π pero podemos extender este método un poco más a los métodos conocidos como Monte Carlo vía cadenas de Markov (MCMC).

Estos son métodos que simulan valores sucesivamente para generar muestras de la distribución final y así estimar cantidades de la misma. Estas cantidades que se simulan de manera sucesiva dependerán del valor simulado anteriormente, de ahí viene la noción de cadenas de Markov.

La idea de estos métodos es construir una cadena de Markov con una distribución estacionaria. Esto con la finalidad de poder asegurar la conver-

gencia de la cadena sin importar en que estado empecemos lo cual será muy útil para métodos que serán usados posteriormente.

A continuación, se presentan algunos conceptos básicos. Para más información sobre resultados el lector se puede dirigir a [4] & [5].

Sea $\{X_n\}$ una sucesión de variables aleatorias con espacio de estados S $n = 0, 1, 2, ...,$. Para todo $x_0, x_1, .., x_{n-1}, x, y \in S$ para todo $n \geqslant 0$..Diremos que $\{X_n\}$ es una cadena de Márkov si se cumple:

$$p[X_{n+1} = y \mid X_0 = x_0, ..., X_{n-1} = x_{n-1}, X_n = x] = p[X_{n+1} = y \mid X_n = x] = p(x, y).$$

A esta distribución se le llama "kernel de transición" o "kernel de Márkov".

Sean $x, y \in S$. Diremos que x alcanza a y si existe $n \geqslant 1$ tal que $p^n(x, y) = 0$, donde p denota la probabilidad de transición, es decir, la probabilidad de pasar del estado x al estado y. En este caso escribimos

$$x \to y.$$

Diremos que x y y se comunican si $x \to y$ y $y \to x$ y lo denotaremos $x \longleftrightarrow y$

La distribución de probabilidad estacionaria f es tal que si $x^{(t+1)} \sim f$, entonces, el kernel de transición y la distribución estacionaria satisfacen la siguiente igualdad:

$$\int_x k(x, y) f(x) dx = f(y).$$

La existencia de la distribución estacionaria f impone algunas restricciones en k :

- Irreducibilidad, el kernel nos permite "movernos libremente" por todo el espacio de estados.
- Recurrencia, la cadena puede regresar a cualquier estado.

Para cadenas recurrentes, la distribución estacionaria también es la distribución límite, esto es, la distribución límite de $x^{(t)}$ es f para casi todos los valores iniciales $x^{(0)}$. Esto se conoce también como ergocicidad.

Teorema 1 (érgodico)**.** *Sea $Y_1, Y_2, \ldots$ una secuencia de variables aleatorias no negativas e idénticamente distribuidas con $E[Y_1] = \mu$. Entonces*

$$P\left(\frac{Y_1 + Y_2 + ... + Y_n}{n} \to \mu\right) = 1 \qquad cuando \quad n \to \infty.$$

1.3.3. Metropolis-Hasting

Este algoritmo genera una muestra de una distribución. Este algoritmo construye una cadena de Markov que converge a la distribución final i.e., al final del algoritmo también se generará una muestra de dicha distribución.

Esta cadena se genera de la siguiente manera. Dado $\theta^{(t)}$

1. Generar $Y_t \sim q(y \mid \theta^{(t)})$ donde q es una densidad auxiliar
2. Hacer

$$\theta^{[t+1]} = \begin{cases} y_t; con\ prob. & \rho(\theta^{[t]}, y_t) \\ \theta^{[t]}; con\ prob. & 1 - \rho(\theta^{[t]}, y_t) \end{cases}$$

donde $\rho(\theta^{[t]}, y_t) = \text{mín}\left\{\frac{f(y)}{f(\theta)}\frac{q(\theta|y)}{q(y|\theta)}, 1\right\}$ y f la distribución final.

1.3.4. Muestreo de Gibbs

El algoritmo de muestreo de Gibbs nos permite simular una cadena de Márkov con distribución $\pi(x)$ de la siguiente manera. Para valores iniciales, $\theta_j^{[0]}$ $j = 1, ..., k$, $\phi^{[0]}$, para $t = 1, ..., T$, simular

$$\begin{gathered} \theta_1^{[t]} \sim p(\theta_1 \mid y, \theta_2^{[t-1]}, \theta_3^{[t-1]}, ..., \theta_k^{[t-1]}, \phi^{[t-1]}) \\ \theta_2^{[t]} \sim p(\theta_2 \mid y, \theta_1^{[t]}, \theta_3^{[t-1]}, ..., \theta_k^{[t-1]}, \phi^{[t-1]}) \\ \vdots \\ \theta_k^{[t]} \sim p(\theta_k \mid y, \theta_1^{[t]}, \theta_2^{[t]}, ..., \theta_{k-1}^{[t]}, \phi^{[t-1]}) \\ \phi_k^{[t]} \sim p(\phi^{[t]} \mid y, \theta_1^{[t]}, \theta_2^{[t]}, ..., \theta_k^{[t]}) \end{gathered}$$

El proceso anterior completa una transición de $\theta^{[t-1]} = (\theta_1^{[t-1]}, ..., \theta_k^{[t-1]})$ a $\theta^{[t]} = (\theta_1^{[t]}, ..., \theta_k^{[t]})$. Por lo que al realizar iteraciones sucesivas nos produciría la

serie $\theta^{[0]}, \theta^{[1]}, .., \theta^{[T]}$ la cual es una cadena de Markov y dado que este proceso se repite para cada uno de los j parámetros así como para ϕ entonces al final de las T iteraciones tendremos que $(\theta_1^{[t]}, .., \theta_k^{[t]}, \phi^{[t]})$ es una muestra de los parámetros.

El objetivo principal de estos dos algoritmos presentados es el poder generar una muestra de una distribución final para que con base a esta se pueda realizar cálculos como lo es la esperanza, varianza, etc. Para una revisión más profunda sobre le tema el lector puede dirigirse a [2]

1.4. Datos espaciales

El objetivo principal del presente trabajo es poder realizar un modelo de particiones aleatorias para observaciones espaciales, es decir, que deseamos trabajar particularmente con datos espaciales. Entonces nos podríamos preguntar ¿Qué es un dato espacial?

Un dato espacial es un dato que tiene asociada una referencia geográfica a través de coordenadas, de tal manera que se pueda localizar en un mapa, por lo que este tipo de datos ayudan a dar información sobre hechos sobre el espacio. Un dato espacial puede ser cualquier dato que hace referencia a una localización o zona geográfica especifica.

La representación visual de la información espacial se puede hacer en una esfera o en una superficie plana como un mapa. A este proceso de asignar un punto de la superficie de la Tierra sobre un mapa se conoce como proyección cartográfica la cual es una conversión de coordenadas desde un sistema de coordenada geodésicas a uno plano, por tanto, permiten transformar una superficie esférica irregular (la Tierra), en una lisa como en un mapa, el principal objetivo es poder representar una parte de la superficie terrestre (o toda), en un mapa convertido en coordenadas geográficas. Por lo que se representan mediante este tipo de coordenadas geográficas (longitud, latitud, altura) las cuales no suelen ser universales y dependen de un sistema Geodésico de Referencia o Datum, pero para nuestro caso estaremos usando $WGS8$ que es un Datum global.

Entonces si nosotros introdujéramos una ubicación en un GPS (o Google Maps), esta ubicación quedaría transformada directamente en un vector con dos posiciones: latitud y longitud. Por ejemplo “Av. Juárez S/N, Centro Histórico de la Cdad. de México, Centro, Cuauhtémoc, 06050 Ciudad de México, CDMX” (Palacio de Bellas Artes) se convirtió en $(19.435292, -99.141187)$

y “Av. San Rafael Atlixco 186, Leyes de Reforma 1ra Secc, Iztapalapa, 09340 Ciudad de México, CDMX” en $(19.362581, -99.0728495)$.

Las latitudes y longitudes se expresan en grados, así que ya podemos establecer una diferencia cuantitativa entre nuestras dos posiciones incluso podemos expresarlo en una unidad de distancia como metros o kilómetros, es decir, ya tenemos una unidad de medida para poder trabajar con los datos espaciales. Para más información sobre el tema el lector puede dirigirse a [6], [7] & [8]

1.5. Análisis de conglomerados

El análisis de conglomerados está orientado a la colección de métodos especializados en descubrir grupos, los cuales serán llamados conglomerados, alrededor de una colección de datos.

El poder descubrir conglomerados conduce a menudo a una mejor compresión del conjunto de datos, mostrando posibles relaciones que se tienen entre las observaciones.

El principal objetivo del análisis de conglomerados se basa en encontrar grupos cuyos elementos tengan algo en común y que no comparten con elementos de otros grupos. El grado de (di)similitud puede ser definido por alguna métrica o de acuerdo con algún modelo probabilístico asignado a cada grupo, Entre los métodos más usados para hacer conglomerados están el de agrupación jerárquica, agrupación de $k-$medias y agrupación basada en algún modelo (mediante alguna estructura probabilística).

Por otro lado, podemos definir una clasificación como un ordenamiento o una organización de cosas en una serie de categorías o clases. Es decir dadas la categorías o clases ya establecidas se irán anexando los datos de acuerdo a las características del mismo, por ejemplo podríamos clasificar los datos por edad, tamaño, color, etcétera.

En conclusión, para generar los conglomerados tendremos que recurrir a métodos para descubrir las agrupaciones mientras que para hacer clasificación bastara con ordenar los datos dependiendo de las categorías ya establecidas (una característica que se desee destacar sobre los datos). Para una revisión más profunda sobre el tema el lector puede dirigirse a [9, Cap.1].

Un concepto que viene acompañado del análisis de conglomerados es la división de estos por grupos es el de partición, como lo muestra [10] será de gran ayuda para entender más este tema.

Definición 2 (Partición)**.** *Una partición del conjunto K, no vacío, que está formada por los subconjuntos $\pi_1, \pi_2, \pi_3, ..., \pi_k$, los cuales deben cumplir:*

- $\pi_i \subseteq K$ *y* $\pi_i \neq \emptyset$*, para* $i = 1, .., k$
- *Para cada par* $i \neq j$*,* $\pi_i \cap \pi_j = \emptyset$
- $\cup_{i=1}^{k} \pi_i = K$

Para ilustrar la Definición 3 tomaremos el siguiente ejemplo.

Ejemplo 3. *Sea* $K = \{1, 2, 3\}$ *las posibles particiones de* K *son:*
$\{\{1\}, \{2\}, \{3\}\}$ *,*$\{\{1, 2\}, \{3\}\}$*,* $\{\{1, 3\}, \{1\}\}$*,* $\{\{2, 3\}, \{1\}\}$*,* $\{\{1, 2, 3\}\}$
Así podemos observar que cada una de las posibles particiones cumple con nuestra definición.Observemos que en los casos particulares $\{\{1, 2, 3\}\}$ *significa que todos los elementos pertenecen a un mismo grupo y por otra parte para la partición conformada por* $\{\{1\}, \{2\}, \{3\}\}$ *cada elemento pertenece a su propio grupo.*

Todas estas particiones forman una clase combinatoria conocida como particiones de conjuntos (de K); su cardinalidad viene dada por el número de Bell (B_n) con n el tamaño de K. El número de Bell es el número de particiones de un conjunto K con n elementos, para el Ejemplo 3 tenemos que $B_3 = 5$, ya que se obtuvieron 5 particiones a partir de los tres elementos de K. Denotaremos esta clase por $\rho_{[n]}$.

Una vez que se ha entendido que es el análisis de conglomerados y sabiendo que existen varios métodos para poder encontrarlos una pregunta natural que se puede hacer es ¿Qué método podremos usar para encontrar conglomerados? En la siguiente sección daremos una forma de encontrar conglomerados para un conjunto especial de datos.

1.5.1. Proceso del restaurante chino

Como se mencionó en la sección anterior tendremos que usar algún método para encontrar conglomerados el cual permita trabajar con dichos datos, por lo que usaremos un método basado en modelo particularmente usaremos el proceso del restaurante chino (PRC), se llama específicamente proceso de restaurante chino ya que los creadores vieron la gran cantidad de gente que se atendía y el proceso de sentar clientes en las mesas de los restaurantes en

un barrio chino.Para más información sobre el tema el lector puede dirigirse a [11] & [12]

Entre las aplicaciones de este proceso se encuentran en las finanzas, medicina, detección de objetos en imágenes, etc.

Supongamos que se tiene un restaurante chino el cual puede tener número infinito de mesas, donde en cada mesa se pueden sentar un infinito número de clientes. El primer cliente entra al restaurante y se sienta en la primera mesa, el segundo cliente entra y decide sentarse ya sea con el primer cliente o en una nueva mesa. En general el n-ésimo cliente se sienta en una mesa k ya ocupada previamente o en una nueva mesa.

La probabilidad para el primer caso es $\frac{n_k}{n-1+\alpha}$ donde n_k es el número de clientes ya sentados y si decide sentarse en una nueva mesa la probabilidad es $\frac{\alpha}{n-1+\alpha}$.

Identificando los clientes con enteros $i = 1, 2, ...n$, los grupos que se irán formando los representaremos como π_j con $j = 1, \ldots, k$. Para poder denotar que algún cliente se sienta en una mesa estaremos usando las variables de pertenencia $d = (d_1, \ldots, d_n)$ que definiremos como $d_i = j$ si y solo si $i \in \pi_j$ para algún $1 \leq j \leq k$ y para $i = 1, .., n$, lo cual servirá para referirnos a que el cliente i se sentó en la mesa j.

En la ecuación (1.8) se muestra cual será la probabilidad de que el n_k-ésimo cliente se siente en la mesa j o decida sentarse en una nueva mesa.

$$P(d_i = \delta \mid d_1, ..., d_{n-1}) = \begin{cases} \frac{\alpha}{n-1+\alpha}; & \delta \notin \{d_1, ..., d_{n-1}\} \\ \frac{n_k}{n-1+\alpha}; & \delta = d_l \quad p.a. \ d_l \in \{d_1, ..., d_{n-1}\} \end{cases} \tag{1.8}$$

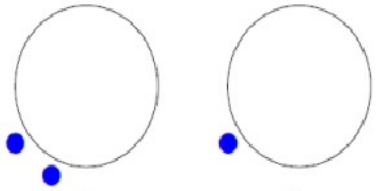
Para una revisión más profunda sobre le tema el lector puede dirigirse a [13, pag. 65-68]

Ejemplo 4. *Realicemos el ejemplo cuando tenemos 3 clientes. Sea* $\alpha > 0$.

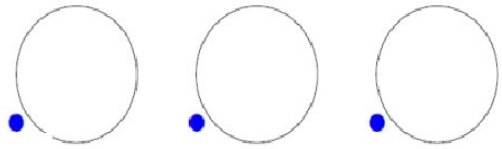
Solución. Primero suponemos que el primer cliente se sienta en la primera mesa. Ahora para el segundo cliente se tiene la probabilidad mostrada en la ecuación (1.9)

$$P(d_2 = \delta \mid d_1) = \begin{cases} \frac{\alpha}{1+\alpha}; & \delta \neq d_1 \\ \frac{1}{1+\alpha}; & \delta = d_1 \end{cases} \tag{1.9}$$

Aquí tenemos dos opciones, la opción de sentarse con con el primer cliente o la opción de sentarse en una nueva mesa.



(a) Primer caso



(b) Segundo caso

Figura 1.2: Ejemplos de partición usando PRC

Finalmente para el tercer cliente. Si suponemos que el cliente 2 se sentó con el cliente 1 se tendrá la probabilidad que se muestra en la ecuación (1.10)

$$P(d_3 = \delta \mid d_2, d_1) = \begin{cases} \frac{\alpha}{2+\alpha}; & \delta \notin \{d_1, d_2\} \\ \frac{2}{2+\alpha}; & \delta = d_l \quad p.a. \ d_l \in \{d_1, d_2\} \end{cases} \tag{1.10}$$

Para este primer caso podemos observar que las mesas quedarían como se muestra en la Figura 1.2 a).

Así una posible partición de este grupo para este caso es $\{\{1, 2\}, \{3\}\}$

Ahora, si suponemos que el segundo cliente decidió sentarse en una nueva mesa. se tendrá la probabilidad que se muestra en la ecuación (1.11)

$$P(d_3 = \delta \mid d_2, d_1) = \begin{cases} \frac{\alpha}{2+\alpha}; & \delta \notin \{d_1, d_2\} \\ \frac{1}{2+\alpha}; & \delta = d_1 \\ \frac{1}{2+\alpha}; & \delta = d_2 \end{cases} \tag{1.11}$$

La Figura 1.2 b) ilustra el segundo caso.

Entonces otra posible partición de este grupo para este caso es $\{\{1\}, \{2\}, \{3\}\}$.

□

Del ejemplo anterior podemos concluir que para la asignación del último cliente en el PRC dependerá del anterior. Por otro lado, al final de este proceso encontramos una partición de nuestros clientes.

En la siguiente tabla mostramos las probabilidades:

Partición	*	Resultado para cada posible valor de *
$\{\{1,2,3\}\}$	(π_1,π_1,π_1)	$1(\frac{1}{1+\alpha})(\frac{2}{2+\alpha}) = \frac{2}{(1+\alpha)(2+\alpha)}$
$\{\{1\},\{2,3\}\}$	(π_1,π_2,π_2)	$1(\frac{\alpha}{1+\alpha})(\frac{1}{2+\alpha}) = \frac{\alpha}{(1+\alpha)(2+\alpha)}$
$\{\{1,2\},\{3\}\}$	(π_1,π_1,π_2)	$1(\frac{1}{1+\alpha})(\frac{\alpha}{2+\alpha}) = \frac{\alpha}{(1+\alpha)(2+\alpha)}$
$\{\{1,3\},\{2\}\}$	(π_1,π_2,π_1)	$1(\frac{\alpha}{1+\alpha})(\frac{1}{2+\alpha}) = \frac{\alpha}{(1+\alpha)(2+\alpha)}$
$\{\{1\},\{2\},\{3\}\}$	(π_1,π_2,π_3)	$1(\frac{\alpha}{1+\alpha})(\frac{\alpha}{2+\alpha}) = \frac{\alpha^2}{(1+\alpha)(2+\alpha)}$

Recordemos que en la sección 1.5 definimos el conjunto de todas las particiones de un conjunto como $\rho_{[n]}$, el Ejemplo 4 ilustra el caso particular cuando tenemos 3 elementos. Ahora denotaremos como $\rho^k_{[n]}$ para obtener un número exacto de grupos de una partición, por ejemplo retomando el Ejemplo 4 si buscamos las particiones que constan de 1 grupo lo podremos denotar como $\rho^1_{[3]}$, entonces tendríamos que:

$$\rho^1_{[3]} = \{\{1,2,3\}\}.$$

Para $\rho^k_{[n]}$ podemos encontrar la probabilidad (que es la probabilidad del numero de grupos) el cual está dado por:

$$P(K_n = k) = \sum_{\pi \in \rho^k_{[n]}} P(\Pi = \pi).$$

Donde Π es una partición aleatoria es decir una variable aleatoria que toma valores en $\rho_{[n]}$ y $P(\Pi = \pi)$ es la probabilidad de que π sea un grupo dado ($\pi = \{\pi_1, \pi_2, .., \pi_k\} \in \rho_{[n]}$) que se calcula como:

$$P(\Pi = \pi) = \frac{\alpha^{k-1}\prod_{j=1}^{k}(n_j - 1)!}{(1+\alpha)(2+\alpha)\cdots(n-1+\alpha)}.$$

Donde $n_j = \#\pi_j$ para $j = 1, ..., k$.

Mientras que su esperanza está dada por:

$$E[K_n] = \sum_{i=1}^{n} \frac{\alpha}{\alpha + i - 1}.$$

Retomando el Ejemplo 4 dado que tenemos $n = 3$ (podemos tener a lo más tres grupos) calculamos que la probabilidad de obtener el número de grupos es:

k	$P(K_n = k)$	Particiones
1	$\frac{2}{(1+\alpha)(2+\alpha)}$	(π_1, π_1, π_1)
2	$\frac{3\alpha}{(1+\alpha)(2+\alpha)}$	(π_1, π_2, π_2),(π_1, π_1, π_2),(π_1, π_2, π_1)
3	$\frac{\alpha^2}{(1+\alpha)(2+\alpha)}$	(π_1, π_2, π_3)

Capítulo 2

Modelo propuesto

Este capítulo está dedicado a construir un modelo que esta enfocado en particiones aleatorias para observaciones espaciales. Como se mencionó en la sección 1.4 estaremos usando el PRC para poder encontrar conglomerados para datos espaciales en particular estaremos trabajando con delitos sobre las calles de la ciudad de México los cuales serán nuestros datos espaciales.

A lo largo de este capítulo podremos observar los conceptos previamente analizado en el capitulo anterior como lo son conglomerados y cómo se hace uso del PRC para este tipo de problemas. Así como las complicaciones que se tienen al trabajar con los datos espaciales y cuáles fueron las características principales que se necesitarán para poder construir el modelo.

2.1. Desarrollo del modelo

Sabemos que los datos con los que vamos a trabajar son datos espaciales específicamente para el ejemplo práctico se estará trabajando con eventos (delitos) sobre alguna región de la ciudad de México para probar el modelo resultante, en general los delitos se hacen sobre las calles y avenidas de algún lugar de la ciudad la cual tiene una forma particular sobre la superficie.

Observemos que una red lineal esta incrustado en alguna región $U \subset \mathbb{R}^2$, es decir, puede ser una manzana, una colonia, etc. en la ciudad de México.

Por lo que se mencionó anteriormente será necesario tomar en cuenta la parte geométrica ya que los datos estarán sobre las calles y avenidas.

Para esto tendremos que empezar por definir una red lineal, L, la cual consta de segmentos de línea que terminan en vértices. Los segmentos de

línea sólo pueden encontrarse en los vértices. Si dos caminos se cruzan, su punto de cruce debe ser un vértice de la red, y cada uno de los caminos debe dividirse en segmentos que terminan en el punto de cruce y con un conjunto finito de aristas:

$$E = \{\xi'_1, .., \xi'_m\}$$

Entonces los datos pertenecen a la red L, por lo que al realizar las agrupaciones se estarán haciendo sobre L, para los cuales tenemos dos conjuntos de medidas que son: variables de ubicación espacial digamos $v_i \in L$ para $i = 1, .., n$ y sus respectivas respuestas x_i. Estas últimas pueden ser constantes por ejemplo $x_i = 1$, lo que significa que ocurrió un evento en la ubicación v_i, gracias a estas respuestas podremos identificar el tipo de evento que se presentará para análisis posteriores.

Estas variables de ubicación van a influir en la agrupación, en el sentido de que es más probable que las respuestas cercanas entre si estén en un mismo grupo por lo que se nos da un patrón de puntos en L, i.e., un conjunto de ubicaciones $v_j \in L$, $j = 1, \ldots, m$ que indican la ocurrencia de algún evento, entonces vamos a definir y_i la variable aleatoria definida como el numero de eventos ocurridos en la arista ξ'_i, es decir,

$$y_i = \#\{v_j \in \xi'_i : j = 1, ..., m\}$$

Aquí contamos el número de eventos ocurridos en cada una de las aristas sobre la red (esta variable nos ayudará a contar el número de eventos sobre las calles) pero sólo requerimos trabajar con las variables y_i distintas de cero (ya que es donde existen los eventos) de esta manera por simplicidad solo tomaremos $y_i > 0$ para $i = 1, .., n$ para algún n.

Observamos que el principal objetivo es analizar el número de eventos ocurridos por lo que al tener que realizar procesos de conteos escogeremos una distribución Poisson desplazada de parámetro λ. Será una distribución desplazada ya que nos interesan los eventos tales que $y_i > 0$ por lo que el caso cuando $y_i = 0$ no nos interesará. Esta distribución está dada por :

$$P(Y = y) = \frac{\lambda^{y-1}}{(y-1)!} \qquad y = 1, 2... \tag{2.1}$$

Hasta el momento sólo tendremos el número de eventos en cierta región

pero no sabemos donde están ubicados por lo que es necesario definir una variable de localización espacial para y_i, digamos e_i.

2.2. Un problema topológico

Si $v_{i_1}, .., v_{i_m} \in L$ son tal que $v_{i_l} \in e_i' \; l = 1, .., m$ para algún m definimos el centroide como (ubicación promedio de los eventos sobre una misma arista)

$$e_i = \frac{1}{m} \sum_{j=1}^{m} v_{ij}.$$

Gracias a la variable e_i ya no será necesario preocuparnos por la topología de la red lineal (la cual es muy complicada de analizar), sino solo por la región U , además, agrupar las aristas de la red tiene sentido, ya que nos permitirá detectar los grupos de eventos.

Una pregunta natural que nos podríamos hacer es ¿Por qué es difícil analizar la topología de la red? la cual se responde en seguida. Tengamos en cuenta que el espacio donde estamos trabajando es una red lineal $U \in \mathbb{R}^2$ por lo que una métrica que se podría usar es la métrica Manhattan (o métrica del taxista) que es muy útil para las redes lineales y funcionaría muy bien para poder generar los grupos dependiendo la distancia. Para más información sobre el tema el lector puede dirigirse a [14, pag. 12-16].

Uno de los problemas que principales de seguir trabajando con la topología en la red lineal es poder definir una distribución de probabilidad sobre todos los elementos que conforman a L, si pensamos cuando solo tenemos una arista no existe problema alguno ya que se puede definir una distribución, así como se muestra en la Figura 2.1.

Cuando tenemos dos o más segmentos la construcción de dicha distribución se vuelve más complicada, podemos pensar en dar una distribución normal bivariada sobre L y hacer un "recorte" o proyección sobre cada uno de los elementos de L, de esta manera obtendremos algo como se muestra en la Figura 2.2.

Posteriormente se tendrá que estandarizar para que de esta manera la intersección de cada una de las curvas obtenidas sean los vértices, así como se muestra en la Figura 2.3, además tendríamos que cerciorarnos que la suma

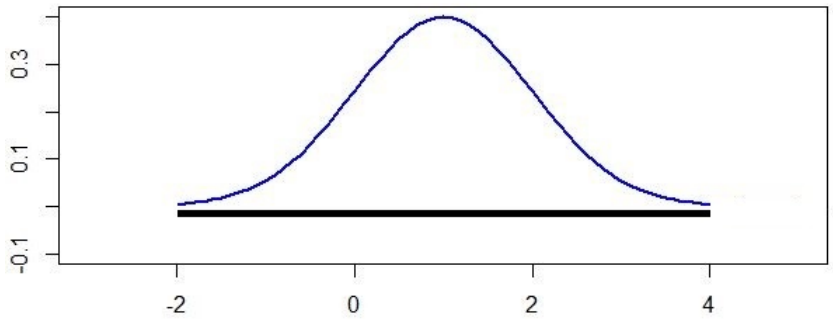


Figura 2.1: Distribución normal sobre elemento de la red lineal.

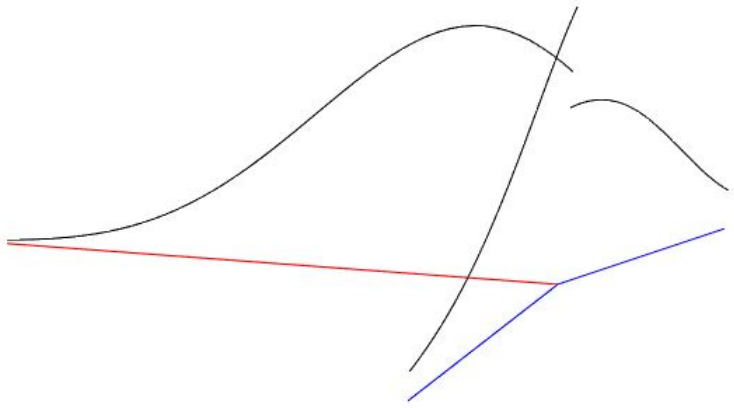


Figura 2.2: Distribución sobre la red lineal.

del área bajo la curva obtenida nos de como resultado 1. Todo esto para verificar que se definió correctamente la distribución sobre L.

Si se tuviera de manera exitosa la distribución sobre L viene otro problema y es el de poder hacer conglomerados ya que para esto se necesitarían varias distribuciones de este tipo y el problema está en que si existen dos conglomerados donde tomen uno o más datos entonces existiría un traslape entre las distribuciones y no sabríamos a donde irían los datos.

Por lo que se menciona anteriormente es que se trabaja dejando de lado este método e introducir la variable e_i para trabajar en el método basado en modelos (que es precisamente como se ha estado trabajando).

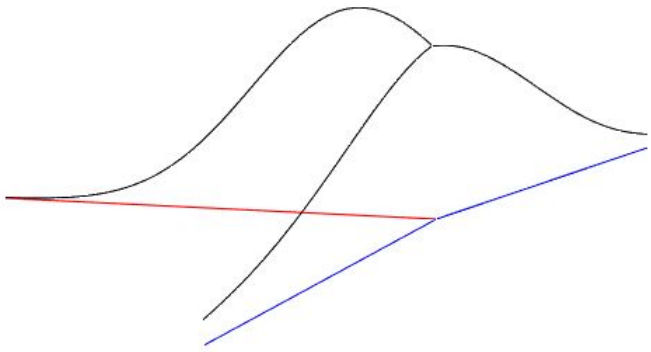


Figura 2.3: Distribución sobre la red lineal.

2.3. Función de penalización

Ahora nuestro interés es agrupar las aristas E de L a través de su correspondiente y_i usando e_i como ubicación espacial (aquí tomamos aquellas aristas donde existe algún evento).

Para un agrupamiento $\pi = \pi_1/.../\pi_k$ se introduce una variable de ubicación u_j, $j = 1, .., k$ asociado a cada grupo π_j, esta nos ayudara a ver la ubicación de cada grupo.

Por lo tanto, deseamos que los conteos y_l, cuyos puntos asociados e_l están mas cercanos de la ubicación u_j se junten en el grupo correspondiente π_j (pues tienen una misma ubicación).

Una forma de medir la cercanía de un punto e, y una localización u es a través de una función de penalización, por ejemplo

$$w(e, u \mid \tau) = \exp\{-\tau(e-u)'(e-u)\} \tag{2.2}$$

para algún $\tau > 0$.

Esta función de penalización nos ayudará a castigar los puntos que están más alejados de la ubicación u, es decir, si dos puntos tienen una distancia muy grande entre ellos la probabilidad de estar en el mismo grupo será chica.

Como se vio en la sección 1.4.1 al querer encontrar agrupamientos el proceso con la cual se trabajará es el del restaurante Chino.

Por lo que para este caso el proceso en conjunto con la función de penalización queda de la forma:

$$p(d_i = \delta \mid d_{-i}, e, u, u^*, \tau) \propto \begin{cases} \frac{\alpha}{n-1+\alpha} w(e_l, u^* \mid \tau) & \text{si } \delta \notin d_{-i} \\ \frac{n_k}{n-1+\alpha} w(e_l, u_l \mid \tau) & \text{si } \delta = d_l \quad p.a. \quad d_l \in d_{-i} \end{cases} \tag{2.3}$$

Donde $d_{-i} = \{d_1, ..., d_{i-1}, d_{i+1}, ..., d_n\}$ para $1 \leq i \leq n$, $e = \{e_1, .., e_n\}$ es el conjunto de variables espaciales, u es el conjunto de ubicaciones y u^* es un valor simulado sobre U. Este proceso lo llamaremos el proceso del restaurante chino con penalización (PRCP).

2.4. Planteamiento del modelo

Del análisis realizado anteriormente podemos observar que nuestro modelo a usar se trata de un modelo jerárquico así como se muestra en el Capítulo 1, por lo que nuestro modelo es de la forma:

$$\begin{array}{lll} y_i \mid \pi, \lambda, e, u \sim \kappa(\lambda_j) 1(i \in \pi_j) & [ind] & i = 1, .., n \\ \pi \mid e, u \sim \rho_0(e, u) & & \\ \lambda_j \mid \pi \sim \nu_0 & [iid] & j = 1, .., k \\ u_j \sim \mu_0 & [iid] & \end{array} \tag{2.4}$$

Observemos que para nuestro problema κ al realizar conteos tomaremos la distribución de $Poisson(\lambda)$, ρ_0 ayudara a generar los grupos y tomaremos el proceso del restaurante chino con penalización, ν_0 será una distribución $gamma(a, b)$ y finalmente μ_0 es la distribución $uniforme(U)$ por lo que (2.4) queda de la forma:

$$\begin{array}{lll} y_i \mid \pi, \lambda, e, u \sim Poisson(\lambda_j) 1(i \in \pi_j) & [ind] & i = 1, .., n \\ \pi \mid e, u \sim PRCP(e, u) & & \\ \lambda_j \mid \pi \sim gamma(a, b) & [iid] & j = 1, .., k \\ u_j \sim Unif(U) & [iid] & \end{array} \tag{2.5}$$

Deseamos encontrar la distribución final que está dada por

$$p(u, \lambda, \pi \mid y, e) \propto L(y \mid \pi, \lambda, e, u) p(\lambda \mid \pi) p(\pi \mid e, u) p(u)$$

Bastará con calcular las densidades condicionales completas, así como la verosimilitud entonces calculemos

1. $L(y \mid \pi, \lambda, e, u)$

   Sabemos que esta función de verosimilitud tiene la forma

$$\begin{aligned} L(y \mid \pi, \lambda, e, u) &= \prod_{i=1}^{n} \frac{\lambda^{y_i - 1} e^{-\lambda_i}}{(y_i - 1)!} 1(i \in \pi_i) \\ &= \frac{1}{\prod_{i=1}^{n} (y_i - 1)!} \left[\lambda^{\sum_{i=1}^{n} y_i - 1} e^{-\sum_{i=1}^{n} \lambda_i} \right] 1(i \in \pi_i). \end{aligned}$$

   Dada la función indicadora para cada una de la partición $\pi = \pi_1 / \ldots / \pi_k$ se cumple que $d_i = j$ si y solo si $i \in \pi$ para algún $1 \leq j \leq k$, es decir, a cada observación le vamos a asignar una partición j entonces tendremos que:

$$L(y \mid \pi, \lambda, e, u) = \prod_{j=1}^{k} \frac{1}{\prod_{i \in \pi_j} (y_i - 1)!} \left[\lambda^{\sum_{i \in \pi_j} y_i - n_j} e^{n_j \lambda_j} \right].$$

   Donde $n_j = |\pi_j|$ es el tamaño del $j-$ésimo grupo

2. $p(\lambda \mid \pi)$

$$\begin{aligned} p(\lambda \mid \pi) &= p(\lambda_1, ..., \lambda_k \mid \pi) \\ &= \prod_{j=1}^{k} p(\lambda_j \mid \pi). \end{aligned}$$

   Esto es para cada λ_j, $j = 1, ..., k$, entonces

$$p(\lambda_j \mid \lambda_{-j}, u, d, y, e) \propto p(\lambda_j \mid \pi) L(y \mid \pi, \lambda, e, u)$$

   Por otro lado, sabemos que se cumple

$$L(y \mid \pi, \lambda, e, u) = \prod_{j=1}^{k} \prod_{d_i=j} \kappa(y_i \mid \lambda_j).$$

Por lo que para este caso tendremos que

$$\begin{aligned}
p(\lambda_j \mid \lambda_{-j}, u, d, y, e) &\propto p(\lambda_j \mid \pi) \prod_{d_i=j} \kappa(y_i \mid \lambda_j) \qquad j = 1, .., k \\
&\propto \lambda_j^{a-1} e^{-b\lambda_j} \prod_{d_i=j} \lambda_j^{y_i-1} e^{-\lambda_i} \qquad j = 1, .., k \\
&\propto \lambda_j^{a-1} e^{-b\lambda_j} \lambda_j^{\sum_{d_i=j}(y_i-1)} e^{-\lambda_i n_j} \qquad j = 1, .., k \\
&\propto \lambda_j^{a-1} e^{-b\lambda_j} \lambda_j^{(\sum_{d_i=j} y_i)-n_j} e^{-\lambda_i n_j} \qquad j = 1, .., k \\
&\propto \lambda_j^{(\sum_{d_i=j} y_i)-n_j+a-1} e^{-\lambda_i n_j - b\lambda_j} \qquad j = 1, .., k \\
&\propto \lambda_j^{(\sum_{d_i=j} y_i)-n_j+a-1} e^{-\lambda_i (n_j+b)} \qquad j = 1, .., k
\end{aligned}$$

$$\therefore p(\lambda_j \mid \lambda_{-j}, u, d, y, e) \propto \lambda_j^{(\sum_{d_i=j} y_i)-n_j+a-1} e^{-\lambda_i (n_j+b)} \qquad j = 1, .., k$$

La cual es una distribución *gamma* con parámetros $(\sum_{d_i=j} y_i - n_j + a - 1, n_j + b)$

3. $p(\pi \mid e, u)$

   Este se realizará usando las variables de pertenencia d_i, $i = 1, ...n$. Para esta probabilidad tendremos dos casos ya que para estas variables de pertenencia se utilizó el proceso del restaurante chino con penalización (el cual tiene dos probabilidades) por tanto

   - Caso 1
     Cuando la variable de pertenencia está en un nuevo grupo, entonces tomaremos $\delta \notin d_{-i}$

     $$p(d_i = \delta \mid d_{-i}, u, u^*, \lambda, \lambda^*, y, e, \tau)\kappa(y_i, \lambda^*)$$

   - Caso 2
     Cuando la variable de pertenencia se queda en algún grupo ya predeterminado, entonces tomaremos $\delta = d_l$ para algún $d_l \in d_{-i}$

$$p(d_i = \delta \mid d_{-i}, u, u^*, \lambda, \lambda^*, y, e, \tau) \propto w(e_i, u_l \mid \tau)\kappa(y_i, \lambda_l)$$

4. $p(u)$

   Recordemos que este parámetro es una de las ubicaciones u_j , $j = 1, .., k$ por lo que la probabilidad anterior se tiene para cada ubicación, es decir, para cada $j = 1, .., k$, entonces

$$\begin{aligned} p(u_j \mid u_{-j}, \lambda, d, y, e) &\propto p(u_j)L(y \mid \pi, \lambda, e, u) \\ &\propto p(u_j) \prod_{d_i=j} w(e_i, u_j \mid \tau) \end{aligned}$$

   Dado que $p(u_j)$ para cada $j = 1, .., k$ es una uniforme se tiene que

$$p(u_j \mid u_{-j}, \lambda, d, y, e) \propto \prod_{d_i=j} w(e_i, u_j \mid \tau)1(u_j \in U)$$

   Pero también sabemos que $w(e_i, u_j \mid \tau) = e^{-\tau(e-u)'(e-u)}$

   Por lo que al realizar la multiplicación obtendremos que

$$\begin{aligned} \prod_{d_i=j} e^{-\tau(e-u)'(e-u)} &= e^{-\tau \sum_{d_i=j}(e_i-u)'(e_i-u)} \\ &= e^{-\tau \sum_{d_i=j}(e_i^2-2e_iu'-u'u)} \\ &= e^{-\tau \sum_{d_i=j} e_i^2} e^{(-2\tau u' \sum_{d_i=j} e_i - \tau n_j u'u)} \end{aligned}$$

$$\therefore p(u_j \mid u_{-j}, \lambda, d, y, e) \propto e^{(-2\tau u' \sum_{d_i=j} e_i - \tau n_j u'u)} 1(u_j \in U)$$

Una vez obtenido cada una de las densidades condicionales podemos observar que el modelo con el que trabajaremos es de la forma:

$$p(\pi, \lambda, u \mid y) \propto l(y \mid \pi, \lambda, e, u)p(\lambda \mid \pi)p(\pi \mid e, u)p(u)$$

Ya obtenidas las densidades condicionales completas procederemos a usar el algoritmo de Gibbs-Sampler entonces:

1. Iniciamos λ^0, u^0, d^0 y hacemos $t = 1$
2. Obtenemos $\lambda^{(t)}, u^{(t)}, d^{(t)}$ con ayuda de $\lambda^{(t-1)}, u^{(t-1)}, d^{(t-1)}$ y generamos

$$\lambda_1^{(t)} \sim p(\lambda_1 \mid \lambda_2^{(t-1)}, ..., \lambda_k^{(t-1)}, d_1^{(t-1)}, ..., d_n^{(t-1)}, u_1^{(t-1)}, ..., u_k^{(t-1)}, y, e)$$

$$\vdots$$

$$d_i^{(t)} \sim p(d_i = r \mid \lambda_1^{(t)}, ..., \lambda_k^{(t)}, d_1^{(t)}, d_2^{(t)}, ..., d_{i-1}^{(t)}, d_{i+1}^{(t-1)}, ..., d_n^{(t-1)}, u_1^{(t-1)}, ..., u_k^{(t-1)}, y, e)$$

$$\vdots$$

$$u_j^{(t)} \sim p(u_2 \mid \lambda_1^{(t)}, ..., \lambda_k^{(t)}, d_1^{(t)}, d_2^{(t)}, ..., d_{n-1}^{(t)}, u_1^{(t)}, ..., u_{j-1}^{(t)}, u_{j+1}^{(t-1)}, ..., u_k^{(t-1)}, y, e)$$

3. Hacer $t = t + 1$ y repetir 2)

Los parámetros τ y α los podremos tomar como aleatorios y así incluirlos en el algoritmo de Gibbs-Sampler.

Estos se podrán actualizar en el paso 2).

Para el parámetro τ vamos a obtener la distribución condicional tomando como distribución inicial una $gamma(c, d)$.

$$\begin{aligned}
p(\tau \mid \pi, u, y, e) &\propto p(\tau)p(d \mid \tau) \\
&\propto p(\tau) \prod_{i=1}^{n} w(e_i, u, d_i \mid \tau) \\
&\propto \tau^{c-1} e^{-b\tau} \exp\{-\tau \sum_{i=1}^{d} (e_i - u)'(e_i - u)\} \\
&\propto \tau^{c-1} e^{-b\tau} \exp\{-\tau \sum_{j=1}^{k} \sum_{d_i=j} (e_i - u)'(e_i - u)\} \\
&\propto \tau^{c-1} \exp\{-\tau (b \sum_{j=1}^{k} \sum_{d_i=j} (e_i - u)'(e_i - u))\}
\end{aligned}$$

Ahora para el parámetro α se asumirá como en [15] que los datos $y_1, ..., y_n$ serán inicialmente condicionalmente independientes de α dado los demás parámetros obteniendo $p(\alpha \mid k) \propto p(\alpha)p(k \mid \alpha)$, para la distribución inicial se tomará como $gamma(a, b)$ y la función de verosimilitud será $\frac{c_{n,k}\alpha^k}{(\alpha)_{n\uparrow}}$ para n el tamaño de la muestra, donde $c_{n,k}$ es el valor absoluto del número de Stirling de primera clase y $(\alpha)_{n\uparrow}$ es el símbolo de Pochhammer.

Sea $\alpha > 0$, el reciproco del símbolo de Pochhammer se puede escribir como:

$$\frac{1}{(\alpha)_{n\uparrow}} = \frac{\Gamma(\alpha)}{\Gamma(\alpha+n)} = \frac{(\alpha+n)\beta(\alpha+1,n)}{\alpha\Gamma(n)}.$$

Donde $\beta(.,.)$ es la función beta.Entonces tendremos

$$\begin{aligned}
p(\alpha \mid k) &= p(\alpha)p(k \mid \alpha) \qquad k = 1, ..., n \\
&= p(\alpha)\frac{c_{n,k}\alpha^k(\alpha+n)\beta(\alpha+1,n)}{\alpha\Gamma(n)} \qquad k = 1, ..., n \\
&\propto p(\alpha)\alpha^{k-1}(\alpha+n)\beta(\alpha+1,n) \qquad k = 1, ..., n \\
&\propto p(\alpha)\alpha^{k-1}(\alpha+n)\int_0^1 \eta^\alpha(1-\eta)^{n-1}d\eta \qquad k = 1, ..., n.
\end{aligned}$$

Por tanto $p(\alpha \mid k)$ es la distribución marginal de la distribución conjunta para α y una variable continua η, con $0 < \eta < 1$, por lo que ahora tendremos que calcular $p(\alpha, \eta \mid k) \propto p(\alpha)\alpha^{k-1}(\alpha+n)\eta^\alpha(1-\eta)^{n-1}$.Bastará con calcular las distribuciones condicionales.

1. $p(\alpha \mid \eta, k)$

$$\begin{aligned}
p(\alpha \mid \eta, k) &= p(\alpha)\alpha^{k-1}(\alpha+n)\eta^\alpha(1-\eta)^{n-1} \qquad k = 1, ..., n \\
&= \frac{\alpha^{a-1}b^a}{\Gamma(a)}e^{-b\alpha}\alpha^{k-1}(\alpha+n)\eta^\alpha(1-\eta)^{n-1} \qquad k = 1, ..., n \\
&\propto \alpha^{a+k-2}(\alpha+n)\eta^\alpha e^{-b\alpha} \qquad k = 1, ..., n \\
&\propto \alpha^{a+k-2}(\alpha+n)\exp(-\alpha(b-log\eta)) \qquad k = 1, ..., n \\
&\propto \alpha^{a+k-1}\exp(-\alpha(b-log\eta)) + n\alpha^{a+k-2}\exp(-\alpha(b-log\eta)) \qquad k = 1, ..., n
\end{aligned}$$

Entonces $p(\alpha \mid \eta, k) \propto gamma(\alpha \mid a+k, b-log\eta) + gamma(\alpha \mid a+k-, b-log\eta)$

2. $p(\eta \mid \alpha, k)$

$$p(\eta \mid \alpha, k) \propto \eta^{\alpha}(1-\eta)^{n-1}.$$

Entonces $\eta \sim beta(\eta \mid \alpha + 1, n)$ con $0 < \eta < 1$.

2.5. Análisis de sensibilidad del modelo

A lo largo de este capítulo hemos realizado el diseño del modelo que estaremos usando así como los valores que se involucran pero existen dos valores que no son fijos como lo son τ y λ el primero usado en la función de penalización (2.2) y el segundo valor está incluido en la distribución Poisson.

Al tener valores que nosotros podemos manipular será necesario realizar un análisis de sensibilidad sobre estos valores pues al estar dentro de nuestro modelo puede influir con el número de grupos que se genera o que tan grande sea la distancia que tome para formar grupos.

En la sección 2.4 se realizó el planteamiento del modelo, así como su implementación mediante el muestreador de Gibbs, el cual será usado para poder realizar dicho análisis de sensibilidad de los valores antes mencionados.

2.5.1. Análisis de τ

Para el primer análisis procederemos a tomar 200 puntos sobre nuestra red lineal previamente elaborada (en el apéndice A.1 se muestra cómo se realiza la red lineal), posteriormente al querer observar el impacto que tiene τ sobre nuestro modelo tendremos que fijar los siguientes parámetros, $\alpha = 4.8986$ y los parámetros de la distribución inicial Gamma los tomaremos como $(1.1, 0.1)$.

Una vez fijados estos parámetros realizaremos diferentes corridas cambiando los valores de τ. Estas corridas serán de $10,000$ iteraciones de las cuales se tomaron las ultimas $5,000$.

Observamos que la Figura 2.4 corresponde a un conjunto de datos simulados (en el apéndice A.3 se muestra cómo se realiza la simulación de los datos sobre la red lineal) estos puntos (los cuales representan nuestros delitos) se muestran en color gris y el centroide e se presenta de color azul. El tamaño de cada círculo corresponde al número de eventos contabilizados en cada arista.

Figura 2.4: τ.

Ya entendido cómo funcionan los elementos (gráficos) dentro de la red, procederemos a realizar el análisis del parámetro τ el cual está incluido en nuestra función de penalización w. Recordemos que esta nos ayuda a determinar cuáles son puntos cercanos entre si y de esta manera podemos observar cómo es que afecta en la distribución final, es decir como quedarán las agrupaciones sobre la red lineal.

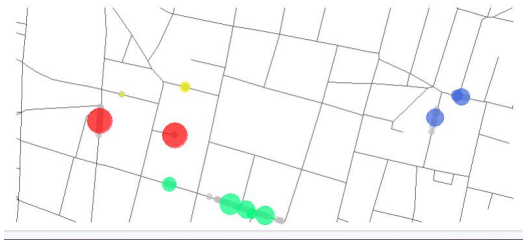
Una vez que se fijaron los parámetros y se entendió cómo funcionan los elementos (gráficos) dentro de la red, procederemos a tomar diferentes valores de τ obteniendo así la Figura 2.5.

Podemos observar que conforme el parámetro τ va aumentando las agrupaciones van mejorando, es decir, si nos fijamos en la Figura 2.5 a) podemos observar en la parte inferior izquierda existen tres tipos de agrupaciones para $\tau = 10^2$ (que es un valor pequeño) el cual no es muy bueno ya que en ese segmento de la red debido a su cercanía debe corresponder a un mismo color (agrupación).

En comparación a esta última Figura 2.6 con $\tau = 10^9$ podemos enfocarnos nuevamente en la parte inferior izquierda para observar que efectivamente obtenemos una misma agrupación es decir los cuatro círculos que corresponden a esa zona son del mismo color.

En conclusión, el parámetro τ incorporado en la función de penalización w es de mucha utilidad ya que como se mencionó antes efectivamente es muy útil para poder determinar las agrupaciones dependiendo su distancia. Conforme se va aumentando τ la función de penalización w es mejor y en consecuencia nuestro modelo es más preciso.

(a) $\tau = 10^2$ (b) $\tau = 10^5$



(c) $\tau = 10^7$

Figura 2.5: Resultados de las agrupaciones para diferentes valores de τ.

2.5.2. Análisis de λ

Para realizar el análisis de λ a diferencia del análisis anterior se fijarán los parámetros para $\alpha = 1$ y $\tau = 1$. Estaremos modificando los hiperparametros a y b de tal forma que el valor esperado inicial λ vaya aumentando, de esta manera podremos observar cual es el efecto que tiene este parámetro en el modelo.

A diferencia de las corridas realizadas para el análisis de τ aquí tomaremos corridas de 10,000 iteraciones de las cuales se tomaron las ultimas 5,000.

Una vez fijados los parámetros procederemos a dar diferentes valores para a y b y como afecta en el modelo.Los resultados se muestran en la Cuadro 2.1.

En este cuadro podremos observar tres columnas, en la primera se encuentran los valores que se estuvo asignando a a y b, en el segundo se puede ver el $E(\lambda)$ para cada valor y finalmente en la última columna se muestra el número de agrupaciones que se estuvieron encontrando.

En las Figuras 2.7 (a) y 2.7 (b) no se logra observar ningún cambio debido a que solo difieren por 2 grupos. Cuando movemos más los valores a y b para que la esperanza aumente podremos observar un cambio como el que se

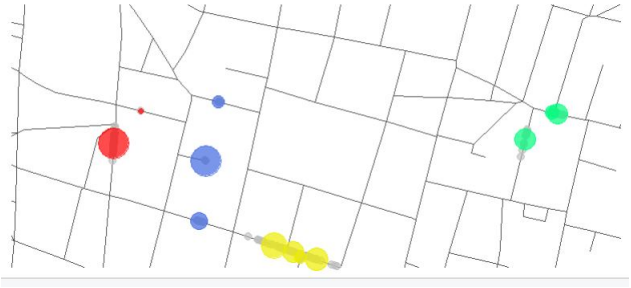


Figura 2.6: $\tau = 10^9$.

(a, b)	$E(\lambda)$	No. Agrupaciones
$(4, 0.1)$	40	46
$(5, 0.05)$	100	38
$(5, 0.010)$	200	11
$(15, 0.01)$	1500	7

Cuadro 2.1: Resultados obtenidos para diferentes valores de λ

muestra en la Figura 2.7 (c).

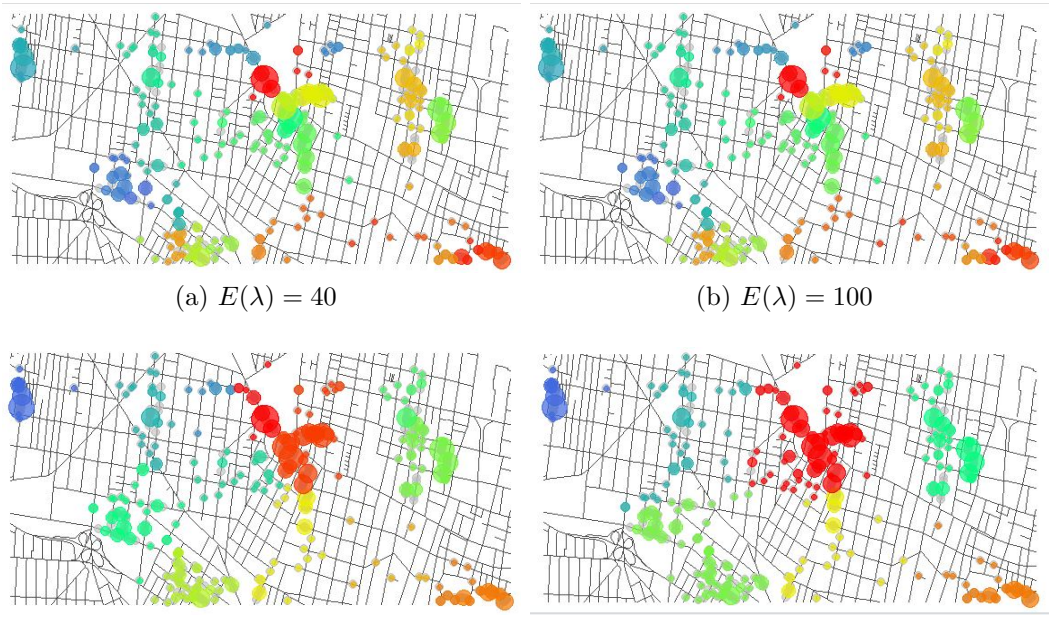
Finalmente para los últimos valores de a y b mostrados en la Figura 2.7 (d) el número de grupos disminuyó a 7 y se ve en comparación como los colores disminuyeron considerablemente pues para este último caso $E(\lambda)$ es mucho mayor. Por lo que podemos observar mediante el análisis de λ que entre más aumenta $E(\lambda)$ el número de grupos son menores es decir el parámetro λ nos ayudara a determinar el número de agrupaciones dentro de L.

En conclusión, los parámetros τ y λ influyen en el modelo para que sea más preciso mediante la función de penalización y con el número de agrupaciones que se encuentran, es decir, nos ayuda a encontrar la menor cantidad de agrupaciones, pero en distancias grandes.

2.6. Estimadores puntuales

Debido a que las poblaciones están caracterizadas por medidas descriptivas numéricas llamadas parámetros, el objetivo de muchas investigaciones estadísticas es calcular el valor de uno o más parámetros relevantes.

Por ejemplo, podríamos estimar la media del tiempo de espera μ en una



(a) $E(\lambda) = 40$ (b) $E(\lambda) = 100$

(c) $E(\lambda) = 200$ (d) $E(\lambda) = 1500$

Figura 2.7: Resultados de las agrupaciones para diferentes valores de λ.

caja registradora del supermercado o la desviación estándar del error de medición σ de un instrumento electrónico. Para simplificar nuestra terminología, al parámetro de interés le llamaremos parámetro objetivo en el experimento. La información de la muestra se puede emplear para calcular el valor de una estimación puntual.

Un estimador es una regla, a menudo expresada como una fórmula, que indica cómo calcular el valor de una estimación con base en las mediciones contenidas en una muestra.

Recordemos que el principal objetivo de nuestro modelo es poder realizar agrupaciones aleatorias con datos espaciales. Por lo que para el modelo establecido anteriormente nuestro parámetro "objetivo" es la partición π sobre la red (mientras que los demás son parámetros auxiliares). Al tener este parámetro en nuestro modelo y mediante la muestra obtenida por el mismo podemos obtener el estimador puntual de la siguiente manera.

Para las agrupaciones π hay un problema particular que a continuación ejemplificaremos para el caso cuando tenemos 3 datos, es decir, $1, 2, 3$ y supongamos que tenemos $m = 5$ el número de agrupaciones obtenida, por ejemplo, podríamos pensar que el modelo encontró las siguientes agrupaciones (el cual será nuestra muestra).

- $\pi^{(1)} = 1/2, 3$
- $\pi^{(2)} = 1/2/3$
- $\pi^{(3)} = 1/2, 3$
- $\pi^{(4)} = 1/2, 3$
- $\pi^{(5)} = 1, 2/3$

Como se muestra en el ejemplo anterior las muestras que se obtengan serán la forma de agrupar los centroides,por lo que el calcular la media de una seria de agrupaciones no es posible (el problema es que los elementos en este caso son grupos y no un número que se pueda promediar) por lo que recurriremos a usar la moda. Podemos hacer uso de la moda ya que tenemos un conjunto discreto el cual se puede contar, de esta manera al usar la moda será más fácil identificar que grupo se repite más veces y así es más probable que se genere.

Sabiendo que la moda es el elemento que se repite la mayor cantidad de veces en un conjunto de datos para nuestro ejemplo es claro que la moda

(que tomaremos como estimador puntual) es el dado por $\widehat{\pi} = 1/2, 3$, pues aparece en tres ocasiones (para $\pi^{(1)}, \pi^{(3)}$ y $\pi^{(4)}$).

Capítulo 3

Aplicación: Detección de focos rojos

Una vez entendido cómo se comporta el modelo, así como el funcionamiento de los parámetros pondremos en práctica lo estudiado en los capítulos anteriores en un problema común en nuestro país que es la inseguridad como se muestra en [16] & [17] la violencia e inseguridad que se sufre hoy actualmente en la Ciudad de México se volvieron temas importantes para los ciudadanos. Según la Encuesta Nacional de Victimización y Percepción sobre Seguridad Pública (ENVIPE) 2022 del INEGI, en septiembre del 2022, el 64.4 % de la población de 18 años en adelante se considera inseguro en la ciudad donde vive. Así como se ilustra en la Figura 3.1 fueron encuestas de la percepción de los ciudadanos sobre la inseguridad por lo que podemos observar que una de las principales necesidades y más urgentes es erradicar el problema.

Una de las demandas más urgentes a los políticos y al alcalde de la Ciudad de México es implementar mejores métodos de vigilancia y control del crimen el cual para poder realizar esto uno de los pasos iniciales es poder identificar los principales lugares con mayor incidencia delictiva en la ciudad para que posteriormente los encargados puedan controlar estas incidencias.

Observamos que al tratarse de delitos sobre algún lugar de la Ciudad de México podemos tomar estos como datos espaciales y el lugar a estudiar como nuestra red lineal (el cual podría ser una colonia, alcaldía o sección dentro de la ciudad) y al aplicar el modelo propuesto en el Capítulo 2 encontraremos las agrupaciones las cuales tienen mayores elementos y de esta manera podremos saber cuáles son los lugares con alta incidencia delictiva (que serán nuestros

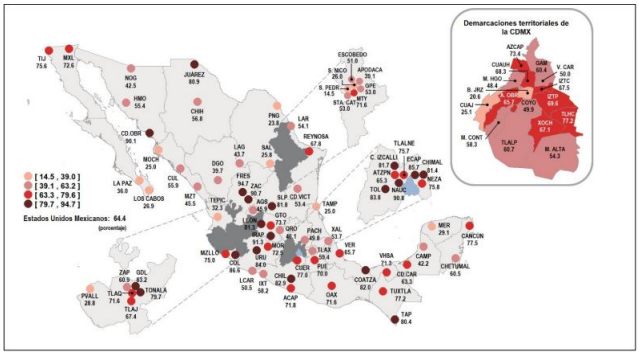


Figura 3.1: Percepción social sobre la delincuencia pública por ciudad de interés septiembre 2022 (porcentaje).

focos rojos).

3.1. Modelo propuesto en la aplicación

Primero obtendremos los datos (los cuales serán delitos de cualquier tipo) sobre algunas regiones de la CDMX, para obtener estos datos nos dirigiremos a la siguiente liga:

https://datos.cdmx.gob.mx/dataset/carpetas-de-investigacion-pgj-cdmx.

En esta página se encuentran las carpetas de investigación de la PGJ de la CDMX la cual nos proporciona los delitos ocurridos en diferentes Alcaldías de la CDMX. Cabe mencionar que esta información tiene una actualización del 23 de Diciembre del 2022.

Una vez recopilado los datos (este proceso se detalla en el apéndice A.4) los descargaremos sobre la red lineal (el cual se generó en apéndice A.1) con sus respectivos centroides obteniendo así la Figura 3.2.

Posteriormente tomamos en cuenta una corrida de 10,000 iteraciones tomando como observaciones las ultimas 5,000 así mismo se tomó una distribución gamma de parámetros $(1.1, 0.1)$ para el parámetro α, análogamente para el parámetro τ se fijó una distribución gamma con parámetros $(10^{11}, 10^4)$.

Ya que el objetivo de esta aplicación es poder encontrar los lugares con alto índice delictivo procederemos a tomar distintos valores para λ y de esta manera tener varios modelos y ver el comportamiento sobre nuestra red lineal. Los valores seleccionados, así como su esperanza teórica están dados en el cuadro 3.1.

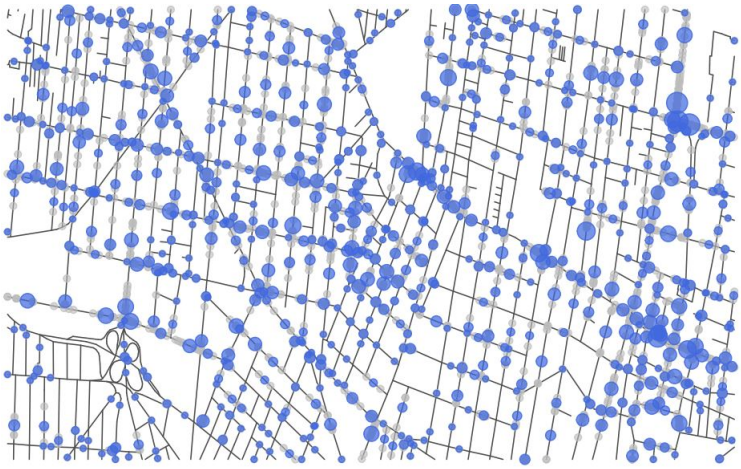


Figura 3.2: Datos reales sobre red lineal.

Modelo	(a, b)	$E(\lambda)$
A	$(4, 0.1)$	40
B	$(10, 0.10)$	100
C	$(15, 0.05)$	300
D	$(15, 0.03)$	500
E	$(10, 0.01)$	1000

Cuadro 3.1: Modelos propuestos

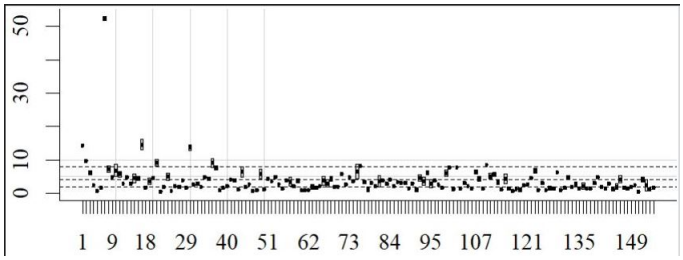


Figura 3.3: Diagrama de caja para el modelo A.

Los resultados obtenidos para cada uno de los modelos se mostrarán mediante un diagrama de caja donde tomaremos en cuenta el numero de grupos que se genera (estos sobre el eje x), así como la cantidad de incidentes en cada uno de los grupos.

Recordemos que para cada grupo π_j, $j = 1, .., k$ se tiene una muestra de tamaño m cada uno con parámetro de intensidad λ_j por lo que para cada uno de estas m muestras obtendremos λ_{jl} con $l = 1, .., m$ para cada j. De esta manera podremos calcular la intensidad media de cada grupo, $\widehat{\lambda_j}$, definida como:

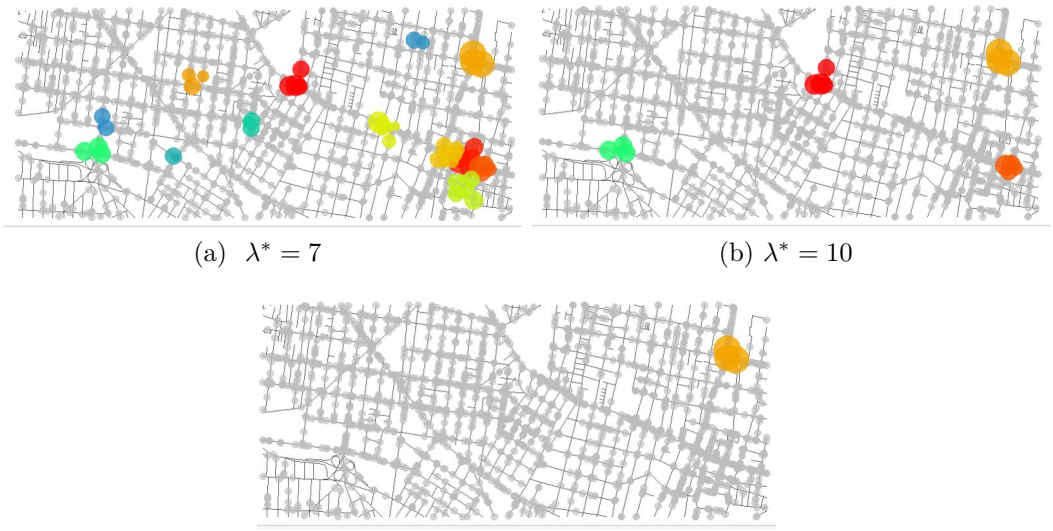
$$\widehat{\lambda_j} := \frac{1}{m} \sum_{l=1}^{m} \lambda_{jl} \qquad j = 1, .., k$$

Finalmente tomaremos diferentes valores (los cuales denotaremos por λ^*) y los valores que cumplan la desigualdad $\widehat{\lambda_j} > \lambda^*$ serán los lugares que vamos a considerar como las zonas peligrosas sobre la red lineal, estos resultados se ilustraran para observar cuales son zonas más peligrosas.

Para el modelo D, observamos que se obtuvieron 21 grupos. En este modelo podemos observar que disminuyeron bastante el número de grupos en comparación de los modelos anteriores, así como las intensidades de algunos grupos (aunque siempre existen un par que sobresale de los demás) lo que hace que el diagrama de caja se pueda visualizar de mejor manera.

Como se muestra en los modelos anteriores, conforme aumentamos el valor de λ^* las zonas que se considerarían como peligrosas serían cada vez menos pero con una mayor intensidad de delitos (es decir, que serán zonas con mayor incidencias).

Podemos observar que el grupo ubicado en la esquina superior de la derecha, así como el que está en la parte inferior derecha están en la mayoría las



(a) $\lambda^* = 7$ (b) $\lambda^* = 10$

(c) $\lambda^* = 15$

Figura 3.4: Detección de zonas "peligrosas" con $\lambda^* = 7, 10$ y 15.Modelo A.

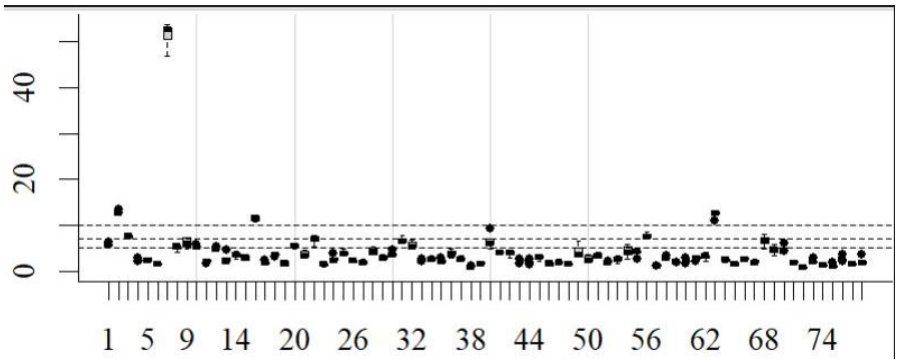
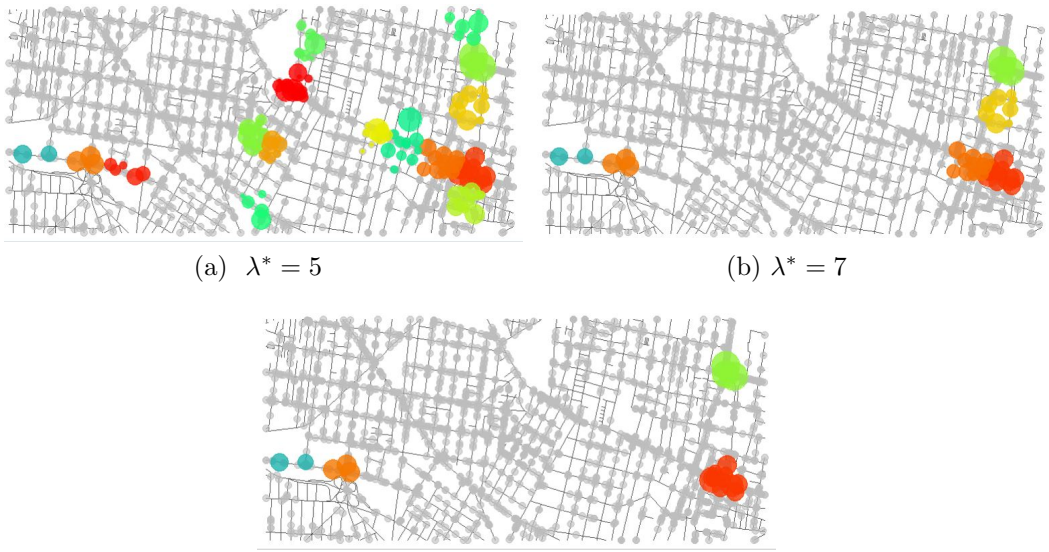


Figura 3.5: Diagrama de caja para el modelo B.



(a) $\lambda^* = 5$ (b) $\lambda^* = 7$

(c) $\lambda^* = 10$

Figura 3.6: Detección de zonas "peligrosas" con $\lambda^* = 5, 7$ y 10.Modelo B.

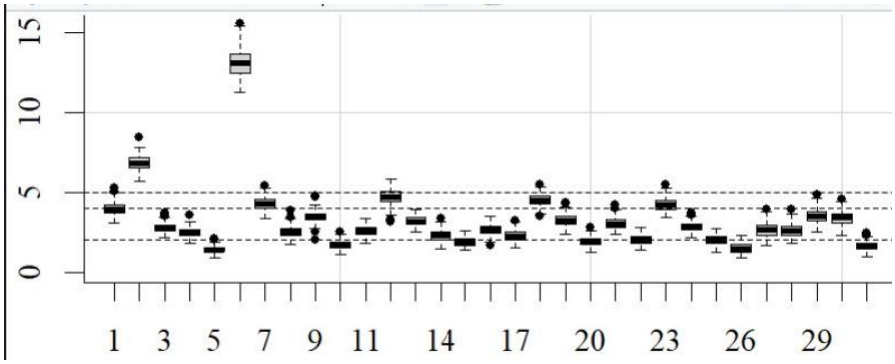
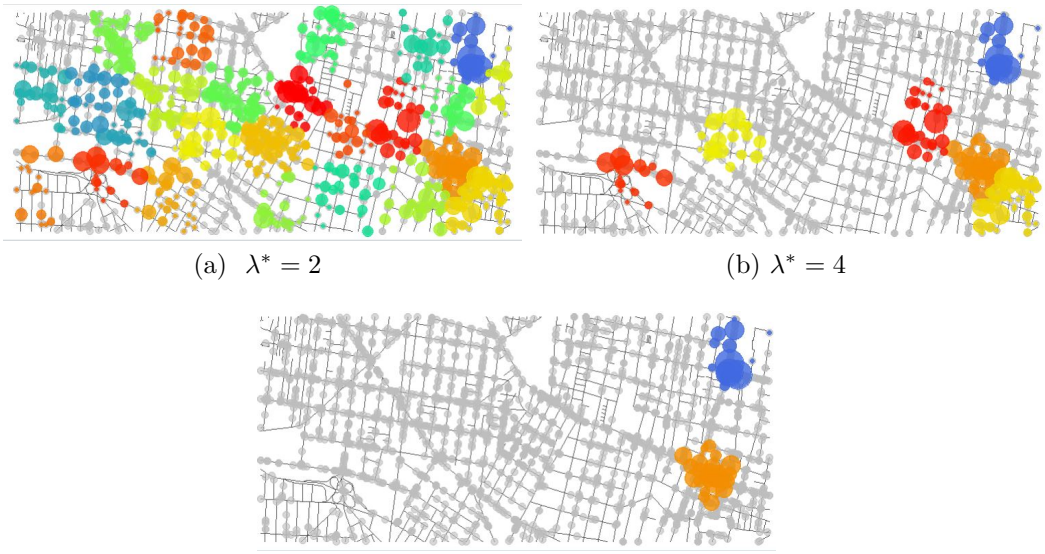


Figura 3.7: Diagrama de caja para el modelo C.



(a) $\lambda^* = 2$ (b) $\lambda^* = 4$

(c) $\lambda^* = 5$

Figura 3.8: Detección de zonas "peligrosas" con $\lambda^* = 2, 4$ y 5.Modelo C.

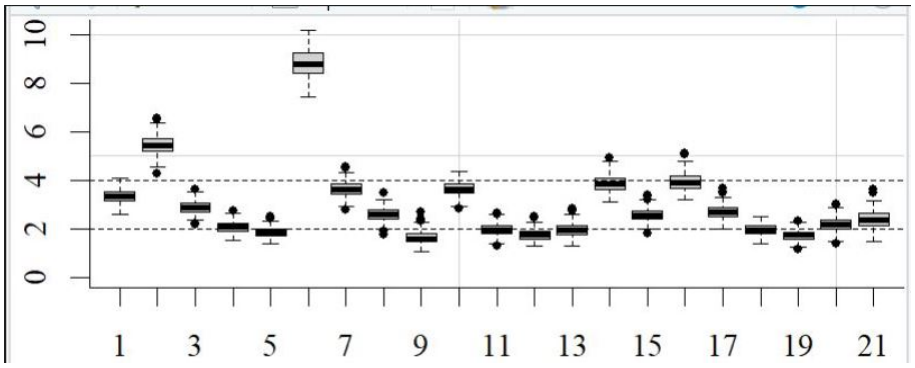
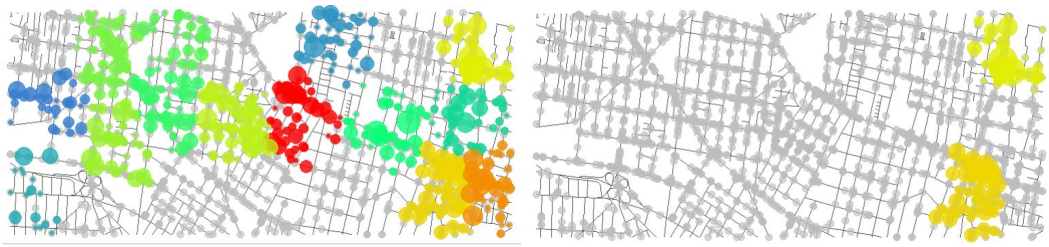


Figura 3.9: Diagrama de caja para el modelo D.



(a) $\lambda^* = 2$ (b) $\lambda^* = 4$

Figura 3.10: Detección de zonas "peligrosas" dependiendo $\lambda^* = 2$ y 4.Modelo D.

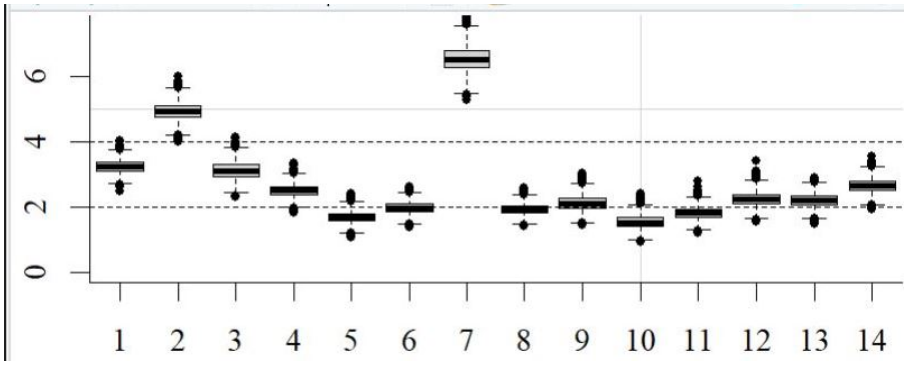
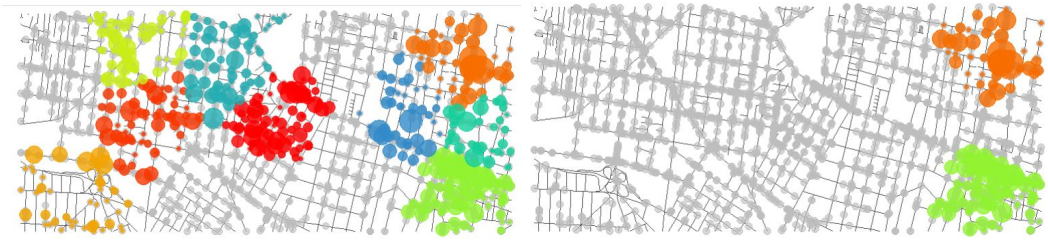


Figura 3.11: Diagrama de caja para el modelo E.



(a) $\lambda^* = 2$ (b) $\lambda^* = 4$

Figura 3.12: Detección de zonas "peligrosas" con $\lambda^* = 2$ y 4.Modelo E.

redes para cada uno de los modelos realizados, esto se debe a que la cantidad de delitos en esa zona específica de la red es mayor a la de los demás grupos y eso se puede notar en los diagramas de cajas de todos los modelos, por lo que sin importar los cambios que realicemos esa zona siempre estará ubicada como peligrosa (son puntos rojos sobre esta zona).

Gracias al parámetro λ en el modelo propuesto podemos controlar el número de grupos que se generan sobre la red lineal, por consecuencia con un grupo podremos abarcar más regiones sobre la red y así podremos saber que tan alta es la incidencia de eventos sobre la red. Esto ayuda a detectar las calles las cuales son mas peligrosas o con un gran índice delictivo que es precisamente el objetivo principal del problema, por lo que al final observamos que el modelo es útil.

Por lo que podemos concluir que el modelo estadístico propuesto funcionó exitosamente ya que pudimos generar agrupaciones sobre datos espaciales (que son los delitos) y gracias a esto se pudo encontrar las áreas con una gran incidencia delictiva en esta red lineal que es el objetivo principal de la aplicación.

Capítulo 4

Conclusiones

La primera parte de este trabajo de tesis fue introducir al lector a la perspectiva de la estadística conocida como bayesiana partiendo del Teorema de Bayes para eventos y posteriormente introducir las funciones de densidad que son las que nos interesaban y mostrando cada uno de los elementos que la conformaban. Culminamos esta primera parte mostrando un modelo, así como los elementos como lo fueron los modelos jerárquicos, métodos de simulación (MCMC) y proceso de restaurante chino todos estos usados posteriormente.

Para la segunda etapa procedimos a la elaboración del modelo sin dejar de lado el objetivo principal este trabajo el cual fue trabajar con datos espaciales. Este problema lo abordamos de tal manera que logramos hacer que la partición aleatoria dependa de la función de penalización $w(e, u \mid \tau)$, la cual nos ayuda a ver cuáles son los grupos que tienen una alta incidencia sobre la red lineal.

Durante el desarrollo del modelo que se propuso se tuvo un problema topológico debido al lugar en el que se estuvo trabajando (que fueron sobre redes lineales) pero se pudo resolver gracias a que se anexaron dos variables más (los centroides e para los conteos y_i y u que ayudara a generar las agrupaciones ayudándonos del proceso del restaurante Chino) el cual ataca este problema de la ubicación espacial enfocándonos en modelar la ocurrencia de eventos en cada borde de la red lineal. De esta manera se pudo agrupar las aristas, preservando la ubicación espacial de los datos para así finalizar con el modelo que se muestra en la ecuación (2.5).

Finalmente, en la última etapa una vez obtenido el modelo de particiones aleatorias para observaciones espaciales se procedió a ponerlo en práctica para un problema de delincuencia el cual fue exitoso pues el objetivo principal

del trabajo se realizó adecuadamente ya que al generar las agrupaciones de manera correcta sobre la red propuesta ayudo a observar los lugares donde existía un gran índice delictivo.

Si bien el modelo propuesto en este trabajo es bueno, cabe señalar que para el desarrollo de este se tuvieron que revisar una gran variedad de temas como el enfoque bayesiano de la estadística, así como la creación de redes espaciales, métodos de simulación estocástica, agrupaciones de datos, el análisis de los datos espaciales. Este último en conjunto con la estructura que tiene la red es lo que se llevó más tiempo hablando computacionalmente, pero a su vez hace que el trabajo obtenido sea una herramienta atractiva y también fácil de manipular.

El trabajo realizado se puede extender de varias maneras. Una de ellas es cambiar la distribución de Poisson usada para la verosimilitud por una distribución Binomial negativa pues ambas son usadas para conteos.

Incluso podríamos proponer otro problema que se acomode al modelo propuesto en este trabajo, por ejemplo, pensemos en una enfermedad en base a el trabajo realizado podríamos encontrar los lugares donde existe una mayor infección de alguna ciudad. O saber cuales son los lugares con un alto índice de accidentes automovilísticos.

Apéndice A

Apéndice A

A.1. Elaboración de la red y carga de datos espaciales

En este apéndice daremos las herramientas necesarias para elaborar y crear redes lineales, así como la obtención de una muestra de datos espaciales (primero realizaremos una muestra simulada y posteriormente procederemos a cargar datos reales) para trabajar posteriormente.

Cabe mencionar que para cualquier tipo de muestra (simulados o datos reales) existirán datos los cuales no estarán sobre las aristas de la red lineal L pero este problema se resolverá fácilmente con ayuda de la proyección ortogonal las cual ayudará precisamente a colocar los datos que estén fuera de las aristas en estas y de esta forma poder trabajar adecuadamente con la muestra.

A.2. Elaboración de la red

Para la elaboración de la red usamos el lenguaje R.

1. Con ayuda de las librerías *Maptpool* la cual nos ayudará a cargar cartogramas de alguna ciudad, *Mapview* que nos permite incorporar datos geoespaciales así como la visualización de los mismo sobre la red, así como las librerías $rgdal, sf, sp$ entre otras que serán usadas para la lectura y manipulación de los datos espaciales.

Ahora procederemos a leer el cartograma de alguna ciudad previamente guardado y usando el comando $readORG$, posteriormente se transformará usando el comando $spTransform$, este permitirá convertirlo a un archivo $WGS84$ el cual es un sistema de coordenadas el cual evitará confusiones ya que es algo estándar para este tipo de datos espaciales. Finalmente podremos visualizar el cartograma que elegimos en un principio esto gracias al comando $mapview$.

```
mx_nw <- readOGR("C:/Users/Luis_/OneDrive/Escritorio/
Proyecto_de_Investigacion/Programas/mapa/bbox/data/North1_Nw.shp")

mx_proj <- spTransform(mx_nw,CRS("+proj=longlat
+ellps=WGS84 +datum=WGS84"))

mapview(mx_proj)
```

2. Una vez que tenemos el mapa procederemos a seleccionar la sección con la que vamos a trabajar, para esto damos las coordenadas específicas y las guardamos, después se convierten y anexan en la zona que se usara, de esta manera al volver a generar la gráfica de la zona obtendremos la delimitación de la zona con la cual estaremos trabajando (esta es la que denotamos como red lineal).

```
min_lon <- -99.19
min_lat <- 19.435
max_lon <- min_lon + 0.04
max_lat <- min_lat + 0.015

bbx <- as(raster::extent(min_lon, max_lon, min_lat, max_lat),
"SpatialPolygons")

mapview(bbx)
```

3. Finalmente convertiremos sólo la red lineal (y no todo el cartograma) en un objeto sf y se guarda para una futura manipulación. Observemos que hasta este punto solo hemos seleccionado la red lineal con la que vamos a trabajar.

```
ss <- st_as_sf(crop_poly)

st_write(ss, dsn=paste(path,"/map",sep=""), layer="medtoy",
driver="ESRI Shapefile")
```

De esta manera para estas coordenadas se genera la Figura A.1. Para una investigación más profunda sobre la redes lineales el lector puede dirigirse a [18, Cap 7].

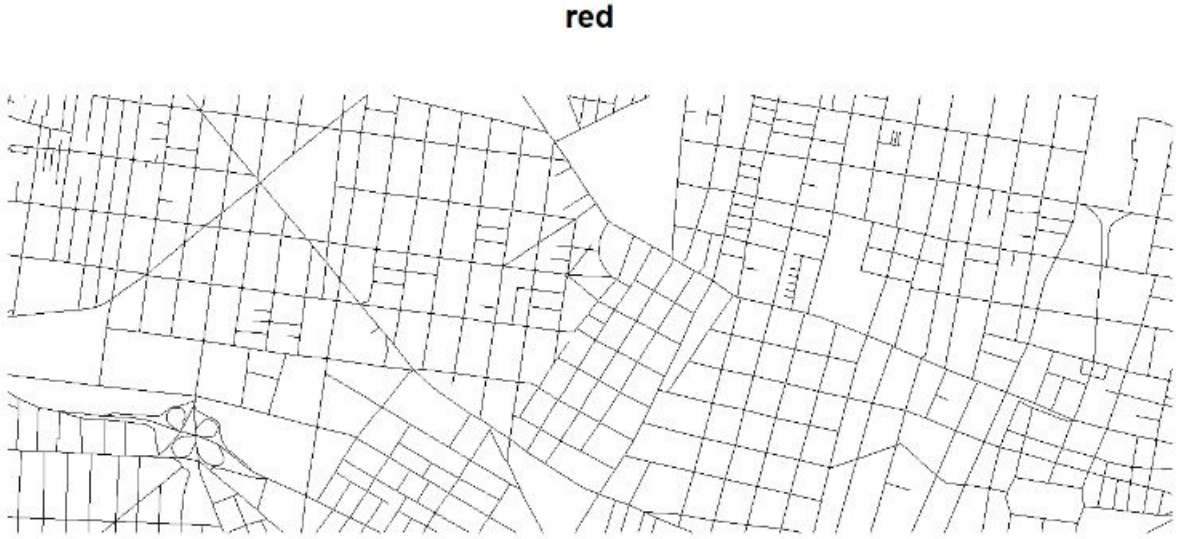


Figura A.1: Red lineal generada.

Una vez que se generó la red lineal podremos incorporar los datos sobre esta, primero mostraremos como se hace para datos reales y posteriormente para datos simulados cabe mencionar que para cada uno de los datos debido a que existen casos donde los datos no están sobre las aristas de la red será necesario hacer uso de la proyección ortogonal para que los datos queden sobre las aristas. El concepto de proyección ortogonal se mostrará a continuación.

A.3. Proyección ortogonal

Se llama proyección ortogonal del punto P sobre la recta l al punto P' que se obtiene de la intersección de la recta l con el plano que contiene al punto P y es perpendicular a ella como se ilustra en la Figura A.2.

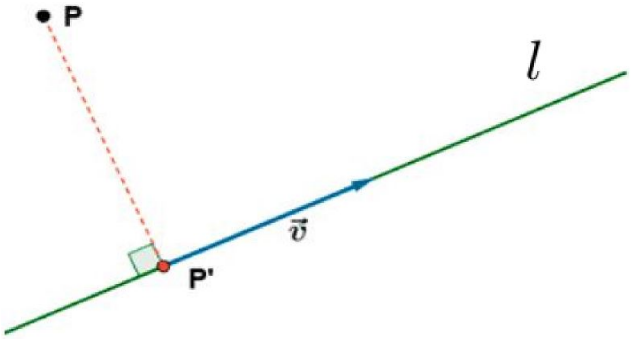


Figura A.2: Proyección Ortogonal de P sobre la recta l.

Para poder calcular la proyección P' de P sobre l se necesita:

1. Calcular $l^{\perp}$ que contiene a P.

2. Hallar el punto P' que es la intersección de ambas rectas, para esto se necesita resolver el sistema planteado por ambas ecuaciones.

Adicionalmente también se puede calcular la distancia del punto $P(x, y)$ a la recta l. Para que la longitud de ese segmento sea la mínima, el segmento y la recta deben de ser perpendiculares. Conociendo las coordenadas del punto P y la ecuación general de la recta l la distancia de ese punto a la recta está dada por:

$$d = \left| \frac{Ax + By + C}{\sqrt{A^2 + B^2}} \right|$$

Ejemplo 5. *Hallar la proyección del punto* $P(-8, 12)$ *sobre la recta* $l : 4x + 7y = -13$ *y su distancia.*

Empecemos por calcular d*, dado que* l *esta en su forma general y teniendo* P *bastara con usar la formula para calcular la distancia, entonces tenemos que:*

$$d = \left| \frac{4(-8) + 7(12) + 13}{\sqrt{(4)^2 + (7)^2}} \right|$$

$$d = \left| \frac{65}{\sqrt{65}} \right|$$

Por lo que $d = \sqrt{65}$

Ahora procederemos a calcular la proyección ortogonal, es decir, calcular lar coordenadas de P' *usando los pasos anteriormente mencionados.*

1. *Primero calculemos* $l^{\perp}$ *que contiene a* P*.*

   Dado dos puntos $A(2, -3)$ *y* $B(-5, 1)$ *que pertenecen a* l *podemos obtener su vector dirección* $\vec{u} = (-7, 4)$ *y al buscar* $l^{\perp}$ *calculamos el vector ortogonal a* $\vec{u}$ *el cuál esta dado por* $(4, 7)$ *que será vector director de* $l^{\perp}$*. Como deseamos que la recta pase por* P *se debe cumplir que* $P \in l^{\perp}$

   De esta manera $l^{\perp}$ *la obtenemos en su forma continua como:*

   $\frac{x+8}{4} = \frac{y-12}{7}$

   Convertida en su forma genera es:

   $l^{\perp} : 7x - 4y + 104$

2. *Una vez obtenida la recta procederemos a hallar* P' *que es el punto de intersección de ambas rectas.*

   $$\begin{cases} 4x + 7y + 13 \\ 7x - 4y + 104 \end{cases}$$

   Al resolver este sistema obtenemos que el punto $P'(-12, 5)$

Realizando lo anterior obtenemos que la proyección ortogonal del punto P *sobre la recta* l *es* $P'(-12, 5)$ *como se muestra en la Figura A.3.*

Una vez entendido como funciona la proyección ortogonal procederemos a obtener la muestra para un red lineal, primero realizaremos una simulación de la muestra y en esta sección se explicara detalladamente como se usa la proyección ortogonal y finalmente se explicara como es que se obtienen los datos reales para su uso, para este último caso la proyección es la misma que para los datos simulados.

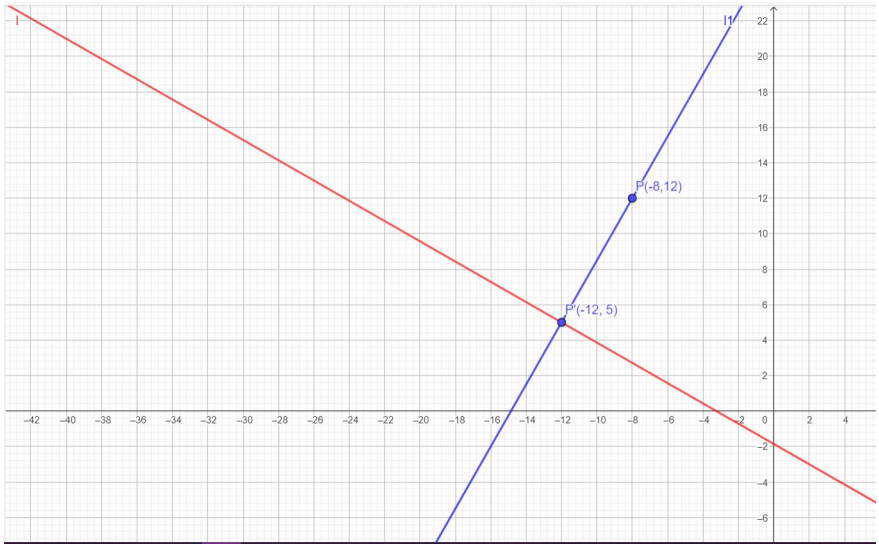


Figura A.3: Proyección Ortogonal de P sobre la recta l.

A.4. Datos simulados

La red lineal que se tomará como ejemplo para esta sección será la mostrada en la Figura A.4. Se tomo esta red L ya que ayudará a observar el proceso de ortogonalización de datos que se simularon.

Figura A.4: Red lineal L.

Una vez que se tiene cargada la red L sobre la cual estaremos trabajando vamos a simular puntos sobre la misma. Para que se pueda realizar el análisis sobre el modelo existen algunos problemas pues recordemos que la red L pertenece a una región U tal que $U \subset \mathbb{R}^2$, es decir, la red está en $\mathbb{R}^2$ por esto para poder realizar la simulación de puntos sobre esta red se usaron dis-

tribuciones normales bivariadas, estas se generaron sobre U, cabe mencionar que no necesariamente estos puntos estaban sobre L como se muestra con los puntos negros en la Figura A.5.

Veamos que existe un problema pues estos puntos están sobre U y no necesariamente sobre L y el modelo propuesto esta dirigido para las observaciones sobre redes lineales por lo que para solucionar este problema se usó la proyección ortogonal (explicada en el apéndice A.2) sobre L.

El procedimiento para calcular las proyecciones ortogonales de algún punto sobre una recta se elaboró para cada uno de los puntos simulados esto para todas las calles (segmentos de recta) posteriormente se calcula la mínima de las proyecciones para saber en cual de todas estas calles será dirigido cada punto una vez realizado este procedimiento se obtuvieron los puntos rojos de la Figura A.5.

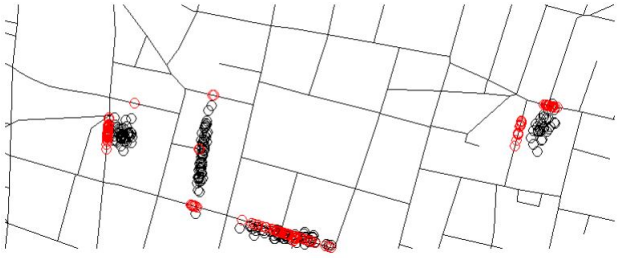


Figura A.5: Datos proyectados sobre L.

Observemos que los círculos están sobre U mientras que los puntos en rojo ya están sobre L, es decir, ya son las proyecciones mínimas dadas para cada uno de los segmentos de L. Por lo que estos serán los datos usados posteriormente para nuestro modelo. Este ultimo proceso de la proyección ortogonal también se realizó sobre los datos reales, pero no se notaban tanto pues en su gran mayoría están sobre los segmentos de la red.

A.5. Cómo cargar los datos reales

Primero obtendremos los datos (los cuales serán delitos de cualquier tipo) sobre algunas regiones de la CDMX, para obtener estos datos nos dirigiremos a la siguiente liga:

https://datos.cdmx.gob.mx/dataset/carpetas-de-investigacion-pgj-cdmx.

Posteriormente procederemos a descargar un archivo que contiene todos los datos sobre delitos de la CDMX.Como se muetsra en la Figura A.6.

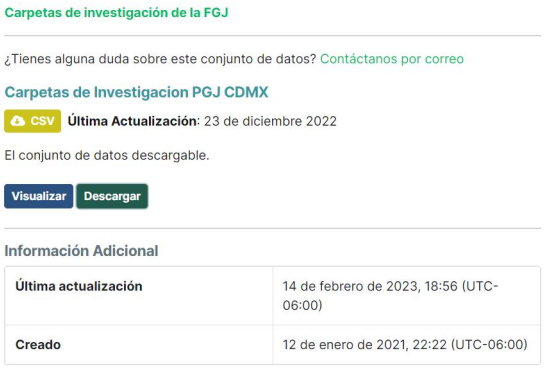


Figura A.6: Archivo para usar.

Descargado el archivo lo abriremos y podremos observar toda la información como el tipo de delito, alcaldía, municipio, fecha, etc., como se muestra en el Cuadro A.1.

Columnas que contiene el archivo
Fecha de los hechos
Hora de los hechos
Delito
Fiscalía
Fecha de los hechos
Alcaldía
Colonia
Longitud
Latitud

Cuadro A.1: Tabla con datos de la PJG.

Ya que el archivo que se descargó, observamos que contiene los delitos de todas las alcaldías de la CDMX, la información es demasiada por lo que se tendrá que seleccionar solo una parte de toda la información contenida en el archivo. Para poder realizar un buen manejo de los datos tomaremos varios factores en cuenta:

1. Para fines prácticos consideraremos indistintos el tipo de delito.

2. Sólo tomaremos en cuenta los delitos que se ejecutaron en el año 2019.

3. Ya que en el apéndice A.1 se generó la red lineal tomaremos los datos que correspondan a esta misma (por lo que debemos tener en cuenta que las dos variables que nos interesan en realidad son la longitud y latitud) los cuales son las alcaldías Cuauhtémoc y Miguel Hidalgo. De esta manera la selección de los datos se redujo a 3458 (ya que en un principio eran más de 500000 datos).

Ahora para poder cargar los datos a R se hará uso de la librería *readxl* el cual ayudará a leer este tipo de archivo y seguiremos los siguientes pasos para poder cargar los datos sobre la red lineal:

1. Cargamos la red lineal que se usará (siguiendo los pasos del apéndice A.1).

```
red <- maptools::readShapeSpatial('C:/Users/Luis_/OneDrive/Escritorio/
Proyecto_de_Investigacion/Programas/map/medtoy')

red <- maptools::as.linnet.SpatialLines(red,fuse=TRUE,trace=TRUE)
```

2. Usando el comando *read_excel* incorporado en la librería anterior leemos los datos y los guardamos en una variable para nuestro caso la llamamos "dts".

```
datos1 <-read_excel("C:/Users/Luis_/OneDrive/Escritorio/
Proyecto_de_Investigacion/Programas/Carga de datos/datos1.xlsx")

dts<-datos1
```

3. Seleccionamos los datos de longitud y latitud de la variable dts la cual contienen las coordenadas para anexarlas en la red lineal.

```
dts<-dts%>%dplyr::select(longitud,latitud)
```

4. Cambiamos los datos seleccionados para que tengan una representación espacial así como asignarle un sistema de referencia (esta es la ubicación espacial que tiene en este caso México).

```
coordinates(dts)<-~longitud+latitud

crs(dts)<- '+proj=longlat +ellps=WGS84 +datum=WGS84'
```

5. Se grafica primero la red lineal y posteriormente los datos generados. Obteniendo así la Figura A.7.

```
plot(red)
points(dts)
```

Figura A.7: Datos reales sobre la red lineal.

Bibliografía

[1] Dale, Andrew I y Andrew I Dale: *Thomas Bayes: a biographical sketch*. A History of Inverse Probability: From Thomas Bayes to Karl Pearson, 1991.

[2] Correa Morales, Juan Carlos y Carlos Javier Barrera Causil: *Introducción a la estadística Bayesiana*. Textos Académicos, 2018.

[3] Gelman, Andrew, John B Carlin, Hal S Stern y Donald B Rubin: *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

[4] Hoel, Paul G, Sidney C Port y Charles J Stone: *Introduction to stochastic processes*. Waveland Press, 1986.

[5] Robert, Christian P, George Casella y George Casella: *Introducing monte carlo methods with r*, volumen 18. Springer, 2010.

[6] Núñez Andrés, María Amparo, Maria José Iniesto *y cols.*: *Introducción a las infraestructuras de datos espaciales*. Centro Nacional de Información Geográfica, 2014.

[7] Chicaiza, Elena: *Análisis espacial con R: Usa R como un Sistema de Información Geográfica*. Revista cartográfica, (100), 2020.

[8] Montane, M.: ***Datos espaciales en R.***, 2020, May 11. https://bookdown.org/martinmontaneb/CienciaDeDatosParaCuriosos/datos-espaciales-en-r.html.

[9] *Model-based Clustering and Classification for Data Science, Charles Bouveyron, Gilles Celeux, T. Brendan Murphy and Adrian E. Raftery, Cambridge University Press, 2019,.*

[10] Martínez, Asael Fabian: *Clustering via Nonsymmetric Partition Distributions*. En *Selected Contributions on Statistics and Data Science in Latin America: 33 FNE and 13 CLATSE, 2018, Guadalajara, Mexico, October 1- 5*, páginas 69–80. Springer, 2019.

[11] Blackwell, David y James B MacQueen: *Ferguson distributions via Pólya urn schemes*. The annals of statistics, 1(2), 1973.

[12] Sethuraman, Jayaram: *A constructive definition of Dirichlet priors*. Statistica sinica, 1994.

[13] Pacheco Góngora, Rafael Eduardo: *Mezcla infinita de Gaussianas*. Tesis de Maestría, CIMAT, Diciembre 2010.

[14] Parra Ardila, Wilson y Ricardo Andrés Cárdenas Izquierdo: *Estudio de la métrica de Manhattan. Segmentos, rectas, rayos, circunferencias y algunos lugares geométricos en la geometría del taxista.* Tesis de licenciatura, Universidad Pedagógica Nacional, 2013.

[15] Escobar, Michael D y Mike West: *Bayesian density estimation and inference using mixtures*. Journal of the american statistical association, 90(430):577–588, 1995.

[16] Pansters, Wil y Héctor Castillo Berthier: *Violencia e inseguridad en la ciudad de México: entre la fragmentación y la politización*. Foro Internacional, páginas 577–615, 2007.

[17] Jiménez, René: *La cifra negra de la delincuencia en México: sistema de encuestas sobre victimización*. Proyectos legislativos y otros temas penales: segundas jornadas sobre justicia penal, páginas 167–190, 2003.

[18] Baddeley, Adrian, Ege Rubak y Rolf Turner: *Spatial point patterns: methodology and applications with R*. CRC press, 2015.

Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA

ACTA DE EXAMEN DE GRADO

No. 00236
Matrícula: 2212801666

Un modelo de particiones aleatorias para observaciones espaciales.

Luis Enrique
LUIS ENRIQUE LARA PEREZ
ALUMNO

REVISÓ

MTRA. ROSALIA SERRANO DE LA PAZ
DIRECTORA DE SISTEMAS ESCOLARES

En la Ciudad de México, se presentaron a las 15:00 horas del día 26 del mes de marzo del año 2024 en la Unidad Iztapalapa de la Universidad Autónoma Metropolitana, los suscritos miembros del jurado:

DR. ANDREY NOVIKOV
DRA. BLANCA ROSA PEREZ SALVADOR
DR. JOSE ANTONIO PERUSQUIA CORTES
DR. ASAEL FABIAN MARTINEZ MARTINEZ

Bajo la Presidencia del primero y con carácter de Secretario el último, se reunieron para proceder al Examen de Grado cuya denominación aparece al margen, para la obtención del grado de:

MAESTRO EN CIENCIAS (MATEMÁTICAS APLICADAS E INDUSTRIALES)

DE: LUIS ENRIQUE LARA PEREZ

y de acuerdo con el artículo 78 fracción III del Reglamento de Estudios Superiores de la Universidad Autónoma Metropolitana, los miembros del jurado resolvieron:

Aprobar

Acto continuo, el presidente del jurado comunicó al interesado el resultado de la evaluación y, en caso aprobatorio, le fue tomada la protesta.

DIRECTOR DE LA DIVISIÓN DE CBI

Román Linares Romero
DR. ROMAN LINARES ROMERO

PRESIDENTE

DR. ANDREY NOVIKOV

VOCAL

DRA. BLANCA ROSA PEREZ SALVADOR

VOCAL

DR. JOSE ANTONIO PERUSQUIA CORTES

SECRETARIO

DR. ASAEL FABIAN MARTINEZ MARTINEZ