

PROCESOS DE POISSON NO HOMOGÉNEOS APLICADOS A LA EXTRACCIÓN DE DINERO EN UNA RED DE CAJEROS AUTOMÁTICOS

Informe de Pasantía

Licenciatura en Estadística

Facultad de Ciencias Económicas y de Administración



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Julio de 2018

Luciana Haussmann
Ana Carolina Romero

Tutor: Dr. Marco Scavino

Resumen

El objetivo de este informe consiste en mostrar la aplicación de técnicas no paramétricas al estudio de procesos de Poisson no homogéneos (PPNH) en presencia de datos agrupados.

Se trabajó con información correspondiente a una red de cajeros automáticos ubicados en Buenos Aires, Argentina. El estudio se centró en el análisis de los retiros de dinero de estos cajeros, con la particularidad de que en lugar de tener la hora exacta de cada extracción -evento del PPNH- se contó con el total de retiros efectuados en tramos de diez minutos. Por este motivo se debió recurrir a técnicas particulares dentro de la estadística no paramétrica. En este estudio se llevó a cabo entonces la modelización, mediante PPNH, de este tipo de eventos, estimando la función de intensidad y la función de intensidad acumulada que caracterizan el proceso según el tramo horario.

Dentro del análisis del fenómeno, se exploraron los diversos comportamientos del proceso en diferentes momentos del tiempo. Para ello, se estimaron funciones de intensidad acumulada según los diferentes días de la semana y momentos del mes. Esto originó que se lograra estimar, en el caso del cajero en el cual se centró el análisis, una función de intensidad unificada para dos días de la semana; y dos funciones de intensidad acumulada diferenciadas para la primera y la segunda quincena del mes.

La motivación inicial fue analizar los períodos de inactividad genuinos de los cajeros para poder diferenciarlos de eventos externos que impidan a los clientes concretar la extracción de dinero que deseen realizar. Por lo tanto, una vez obtenidas las estimaciones de las funciones de intensidad y de intensidad acumulada, se calculó la probabilidad de que no ocurran eventos a lo largo del tiempo, es decir, la probabilidad de inactividad intrínseca del proceso.

Estos resultados se obtuvieron tomando como base un cajero en particular, a los efectos de probar la metodología desarrollada. Posteriormente, se analizaron los comportamientos de otros cajeros para compararlos entre sí y mostrar que la metodología de estimación que hemos implementado es flexible y aplicable a los diferentes casos.

Palabras claves: Datos agrupados; Estadística no paramétrica; Extracciones de efectivo; Procesos de Poisson no homogéneos; Red de cajeros automáticos.

Índice general

Índice general	1
Índice de cuadros	4
Índice de figuras	5
1. Introducción	8
1.1. Motivación	9
1.2. Objetivos	9
1.3. Información disponible y otros aspectos	9
1.4. Análisis Descriptivo	10
2. Marco Teórico	16
2.1. Procesos puntuales y de conteo	16
2.2. Procesos de Poisson no homogéneos	19
2.2.1. Propiedades del PPNH	21
2.2.2. Ejemplos de PPNH	22
3. Metodología	23
3.1. Estimaciones paramétricas y no paramétricas para PPNH	23
3.2. Caracterización del PPNH en presencia de datos agrupados	24
3.2.1. Antecedentes	25
3.3. Estimaciones de la función de intensidad y función de intensidad acumulada	27

3.3.1. Estimación propuesta por Henderson	28
3.3.2. Estimación propuesta por Leemis	31
3.4. Intervalos de confianza	32
3.5. Detección de períodos de inactividad	33
4. Resultados	34
4.1. Estimación de la función de intensidad	34
4.2. Estimación de la función de intensidad acumulada	39
4.2.1. Estimación de la función de intensidad acumulada por quincena	44
4.3. Inactividad del cajero	45
4.4. Análisis comparativo de diferentes cajeros	48
5. Conclusiones	58
5.1. Caracterización de los retiros en la red de cajeros	58
5.2. Detección de inactividad	59
5.3. Posibles extensiones del trabajo	60
Bibliografía	62
Apéndices	63
A. Glosario	63
B. Demostraciones	64
B.1. Demostración de la Proposición 2.2.4	64
B.2. Demostración de la Proposición 2.2.5	67
B.3. Demostración de la Proposición 3.3.3	68
C. Simulación de un PPNH	71
D. Tablas y gráficos	72

Índice de cuadros

1.1. Medidas de Resumen del cajero b309	11
1.2. Medidas de Resumen para la variable “Cantidad de Retiros” por día de la semana para el cajero b309	11
4.1. Medidas de posición, tendencia central y dispersión de la función de intensidad estimada para el cajero b309	38
D.1. Probabilidad estimada de inactividad del cajero b309 por hora para miércoles y jueves	81
D.2. Probabilidad de Inactividad estimada para cada cajero por hora	81

Índice de figuras

1.1. Diagrama de caja del total de extracciones en intervalos de diez minutos para el cajero b309	12
1.2. Diagrama de barras del total de extracciones	13
1.3. Diagrama de barras para el total de extracciones del cajero b309 en tramos de a 10 minutos	14
1.4. Total de retiros promedio del cajero b309 en tramos de una hora a lo largo del día	15
2.1. Recorrido típico de un muestreo de un proceso de conteo	17
2.2. Incrementos de las extracciones del cajero b309 durante el día 08/03/2010	18
2.3. Incrementos por hora de las extracciones del cajero b309 durante el día 08/03/2010	19
4.1. Estimación de la Función de Intensidad del cajero b309 en tramos de diez minutos	35
4.2. Estimación de la Función de Intensidad del cajero b309 en tramos de una hora	36
4.3. Estimación de la Función de Intensidad del cajero b309 para cada día de la semana	37
4.4. Diagramas de caja de la función de intensidad estimada por día de la semana para el cajero b309	38
4.5. Estimación de la FIA del cajero b309	40
4.6. Estimación de la FIA estimada del cajero b309 por día	41
4.7. Estimación de la FIA del cajero b309 para miércoles y jueves	42

4.8. Estimación de la FIA del cajero b309 para miércoles y jueves vs domingo .	43
4.9. Estimación de la FIA para miércoles y jueves vs estimación de la FIA para algunos días en particular del cajero b309	44
4.10. Estimación de la FIA para miércoles y jueves vs estimación de la FIA por quincena del cajero b309	45
4.11. Estimación de la probabilidad de inactividad en la siguiente hora y siguiente media hora por intervalo de a diez minutos	47
4.12. Total de extracciones cada diez minutos para los cajeros de muestra	49
4.13. Retiros promedio por hora para los cajeros de muestra	50
4.14. Estimación de la función de intensidad para los cajeros de muestra	51
4.15. Estimación de la función de intensidad de cada cajero	52
4.16. Diagramas de caja para los cajeros de muestra por día de la semana	53
4.17. Estimación de la función de intensidad acumulada por día de la semana para los cajeros de muestra	54
4.18. Estimación de la función de intensidad de cada cajero	55
4.19. Estimación de la función de intensidad acumulada para los cajeros de muestras en ciertos días	56
D.1. Estimación de la función de intensidad del domingo para el cajero b309 . .	72
D.2. Estimación de la función de intensidad del lunes para el cajero b309	73
D.3. Estimación de la Función de Intensidad del martes para el cajero b309 . .	73
D.4. Estimación de la función de intensidad del miércoles para el cajero b309 . .	74
D.5. Estimación de la función de intensidad del jueves para el cajero b309 . . .	74
D.6. Estimación de la función de intensidad del viernes para el cajero b309 . . .	75
D.7. Estimación de la función de intensidad del sábado para el cajero b309 . . .	75
D.8. Estimación de la función de intensidad acumulada para para el domingo para el cajero b309	76
D.9. Estimación de la función de intensidad acumulada para para el lunes para el cajero b309	76

D.10.Estimación de la función de intensidad acumulada para para el martes para el cajero b309	77
D.11.Estimación de la función de intensidad acumulada para para el miércoles para el cajero b309	77
D.12.Estimación de la función de intensidad acumulada para para el jueves para el cajero b309	78
D.13.Estimación de la función de intensidad acumulada para para los viernes para el cajero b309	78
D.14.Estimación de la función de intensidad acumulada para para el sábado para el cajero b309	79
D.15.Estimación de la función de intensidad acumulada del cajero b397 para miércoles y jueves	79
D.16.Comparación de la estimación de la FIA para diferentes días según los cajeros de muestra	80

Capítulo 1

Introducción

En el marco de los servicios financieros clásicos, los bancos no solo se limitan a tener sucursales abiertas con un horario determinado, sino que las entidades ponen a disposición de sus clientes otros medios para realizar sus operaciones, como por ejemplo los cajeros automáticos. Para brindar este servicio, las entidades bancarias deben adherirse a empresas que poseen una red de cajeros, a través de la cual los clientes pueden operar desde cualquier punto de esta red y en cualquier horario del día. A su vez estas redes tienen conexiones entre sí, aunque generalmente operar en otra red implica un costo para el cliente.

Estas redes tienen la información en tiempo real acerca del funcionamiento de sus cajeros, se sabe la cantidad de dinero que queda disponible y si el cajero está en correcto funcionamiento o no. Por ejemplo, si hay un corte de energía, la central tiene la información de que un cajero se ha desconectado de la red. Sin embargo, puede ocurrir que algún evento externo le impida a los clientes acceder al recinto donde se ubica el cajero. Esto podría ser, alguna persona u objeto bloqueando la entrada, una rotura en la puerta o cualquier otra eventualidad.

A raíz de esto, surge el presente trabajo con el propósito de diferenciar los períodos donde se registra inactividad debido al comportamiento habitual, de los períodos donde la falta de actividad responde a factores externos y por lo tanto se debe enviar a una persona a solucionarlo. Para ello, se cuenta con información de 20 cajeros automáticos provenientes de una red de cajeros de Buenos Aires, Argentina. Se miden las extracciones realizadas en cada uno de ellos en tramos de a 10 minutos, por lo tanto se tienen 144 mediciones por día para cada cajero.

Los datos están comprendidos entre el 1° de agosto de 2010 y el 15 de octubre de 2012, lo cual resulta en un total de 116.064 observaciones por cajero. El conjunto de datos consiste

en las siguientes variables: fecha, hora, cantidad de extracciones, monto total extraído y segundos funcionando dentro de esos diez minutos. Estos datos fueron proporcionados por el Prof. Juan José Goyeneche y provienen de un trabajo privado para optimizar la carga de efectivo.

1.1. Motivación

El problema planteado que motivó nuestro trabajo fue detectar los períodos de inactividad anómalos de los cajeros. Problemas como falta de energía eléctrica o falta de billetes son detectados por el cajero madre. Sin embargo, pueden ocurrir problemas externos que la central no detecta y que deriven en períodos de inactividad que no sean producto del comportamiento habitual de las extracciones, por ejemplo la rotura de la puerta de entrada o cualquier eventualidad que obstruya el ingreso de personas, como se mencionó anteriormente.

1.2. Objetivos

El objetivo general del presente trabajo consiste en caracterizar las extracciones de los cajeros automáticos aplicando una metodología basada en procesos estocásticos.

Los objetivos específicos son estimar las llamadas función de intensidad y función de intensidad acumulada del proceso que modela las extracciones de un cajero. A su vez, se procura evaluar la posible generalización o particularidad de cada día durante los diferentes días de la semana y momentos del mes.

Finalmente, se explora una metodología que permita detectar los períodos de inactividad genuinos y diferenciarlos de los extraordinarios, para alertar en caso de fallas o anomalías en los cajeros automáticos que no sean detectables por la central.

1.3. Información disponible y otros aspectos

Los datos disponibles son datos agregados o datos de conteo, es decir que no se cuenta con el registro de cada extracción. Por lo tanto, no se tiene el instante preciso del tiempo, sino que se cuenta con la cantidad de extracciones realizadas durante cada tramo de diez minutos y el monto total que suman dichas extracciones.

La información disponible mide una dimensión de la realidad, por lo tanto existen algunos supuestos que se deben hacer a la hora de analizarla.

En primer lugar, no hay información sobre las personas que llegan al cajero y forman fila para ingresar a operar. En segundo lugar, los clientes que ingresan no necesariamente realizan una extracción de dinero, sino que puede ser que formen fila y ocupen parte o la totalidad del tiempo en realizar otro tipo de operación dentro de las posibles en el sistema de cajeros. O sea que no se tendrán en consideración estos aspectos de la realidad y solo se trabajará asumiendo la llegada de personas a extraer dinero.

Otro aspecto a tener en cuenta es que no necesariamente una extracción equivale a la llegada de una persona, ya que podría darse el caso en el que una sola persona realice más de un retiro de dinero, aspecto en el cual no nos adentraremos a analizar.

Además, si en vez de un solo cajero en ese lugar hubiera más, tal vez más clientes hubiesen realizado una extracción. Ya que si por ejemplo ocurre que una persona llega con el objetivo de retirar efectivo y se encuentra frente a una larga fila, puede decidir no hacerlo, mientras que si hubiera más cajeros en el mismo sitio, no se formaría una fila tan extensa.

También puede ocurrir que en un período de diez minutos lleguen cinco personas y realicen cinco extracciones y en otro tramo lleguen ocho personas pero también realicen cinco extracciones debido a los tiempos en la operación, y las restantes tres personas realicen la extracción en el siguiente tramo. Esta situación tampoco será tomada en cuenta en el análisis. De todos modos, influyen en el comportamiento de los clientes pero no es de relevancia a los efectos de detectar los tiempos de inactividad del cajero.

En definitiva, el análisis se enfocará en el modelado de las extracciones propiamente dichas y no del comportamiento humano, ya que sería un abordaje más complejo para el cual se carece de información.

1.4. Análisis Descriptivo

En esta sección se mostrará la descripción de datos de un cajero en particular, el b309. Si bien se han analizado la totalidad de los datos, el objetivo aquí es mostrar el camino recorrido, que es análogo para todos los cajeros. Para una primera aproximación a los datos se toma el resumen de los mismos sin ningún tipo de manipulación previa.

Una perspectiva es analizar los datos en su versión inicial, es decir, tal y como fueron recibidos, con mediciones cada diez minutos. A continuación se presenta un resumen con todas las variables para el cajero b309.

Cuadro 1.1: Medidas de Resumen del cajero b309

	Monto	Cant. Retiros	Seg. Funcionando
Mínimo	0	0	600
Primer Cuartil	0	0	600
Mediana	0	0	600
Media	1010	-	600
Tercer Cuartil	1500	2	600
Máximo	16200	12	600

Las medidas de resumen de la variable Tiempo, medida en segundos funcionando, no nos aportan información de relevancia, por lo tanto no será tomada en cuenta de aquí en más, ya que como se puede apreciar en la tabla, este cajero estuvo siempre en correcto funcionamiento.

Tanto la cantidad de extracciones como el monto son contabilizados cada diez minutos, lo que hace que durante las altas horas de la noche y la madrugada se contabilicen muchos ceros. Es por eso que no solo el valor mínimo es cero sino que el primer cuartil y la mediana también lo son. Se puede deducir que existe una importante asimetría en la distribución.

Para la variable cantidad de retiros, la mediana es 0 y el máximo 12. Como se verá más adelante, para este caso particular, en la mayoría de los intervalos de 10 minutos no se realizan extracciones, es por eso que la mediana es cero.

Para el caso del monto, la mediana es cero, lo cual es de esperar dada la mediana de cantidad de retiros, la media es \$1010 y el máximo \$16200. Es decir que en promedio la media de los tramos es que cada persona retira \$798 (asumiendo una extracción por persona), y como máximo hay un tramo en el que cada persona retira en promedio \$1350.

A partir de ahora, se estudiará la variable cantidad de extracciones y se dividirá por día de la semana para un análisis más detallado.

Cuadro 1.2: Medidas de Resumen para la variable “Cantidad de Retiros” por día de la semana para el cajero b309

	Domingo	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
Mínimo	0	0	0	0	0	0	0
Primer Cuartil	0	0	0	0	0	0	0
Mediana	0	0	0	0	0	1	1
Tercer Cuartil	1	2	2	2	2	3	2
Máximo	10	12	12	12	12	12	11

Como se puede ver en el cuadro anterior, la mediana es cero en cada día excepto por el viernes y sábado que se ubica en 1. Estos valores son bajos porque toman en cuenta las horas de la noche cuando no hay prácticamente extracciones.

Por último se observa que el máximo de extracción en el intervalo de diez minutos es 12 para todos los días a excepción del fin de semana donde es menor. Se puede pensar que es muy poco probable que se efectúen más de 12 extracciones en 10 minutos ya que es físicamente imposible realizar una gran cantidad de extracciones en un tiempo tan breve.

Si bien en este trabajo se pone el foco en los retiros de dinero, el fenómeno real implica que entre una persona al cajero, ingrese la tarjeta de débito, su código, etc. Si se quisiera modelar la llegada de personas, tendrían que tomarse en cuenta las filas que se forman, pero en este caso no se dispone de esa información y por lo tanto el estudio se centra en la cantidad de extracciones.

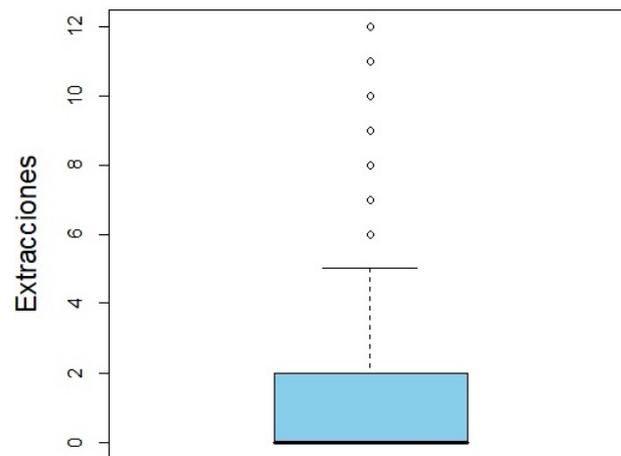


Figura 1.1: Diagrama de caja del total de extracciones en intervalos de diez minutos para el cajero b309

Como se puede observar, la distribución es asimétrica, la mediana se ubica en el cero, es decir que para el caso de este cajero, el 50% de los datos son cero. Esto probablemente ocurra debido a que en las horas de la noche casi nadie acude a retirar dinero, pero se verá más adelante.

En el diagrama de caja de la Figura 1.1 se puede ver a simple vista que hay 7 valores atípicos, se trata de 5.053 observaciones que representan un 4% del total. El estudio se basará en el total de las observaciones.

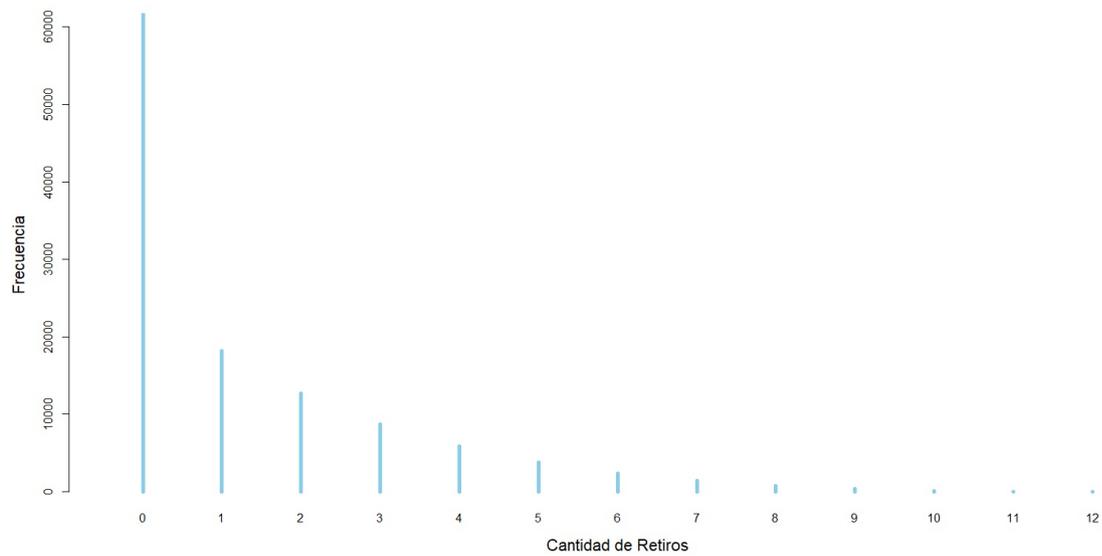


Figura 1.2: Diagrama de barras del total de extracciones

En el diagrama de barras de la Figura 1.2 se puede visualizar que en este cajero la relación entre la frecuencia y la cantidad de extracciones es decreciente. Hay un gran salto entre la cantidad de tramos que tienen nula cantidad de retiros con los que tienen una o más cantidad, porque como se dijo anteriormente, durante la noche hay muchos tramos sin extracciones.

Para visualizar mejor la cantidad de extracciones que se suceden a lo largo del día, se presenta en la figura 1.3 un gráfico de barras con el total de extracciones diarias. A los efectos de disponer de la información completa para todos los días, se deja fuera la información parcial del día 15/10/2012, por lo tanto a partir de aquí se sigue trabajando con 115 días enteros para cada día de la semana.

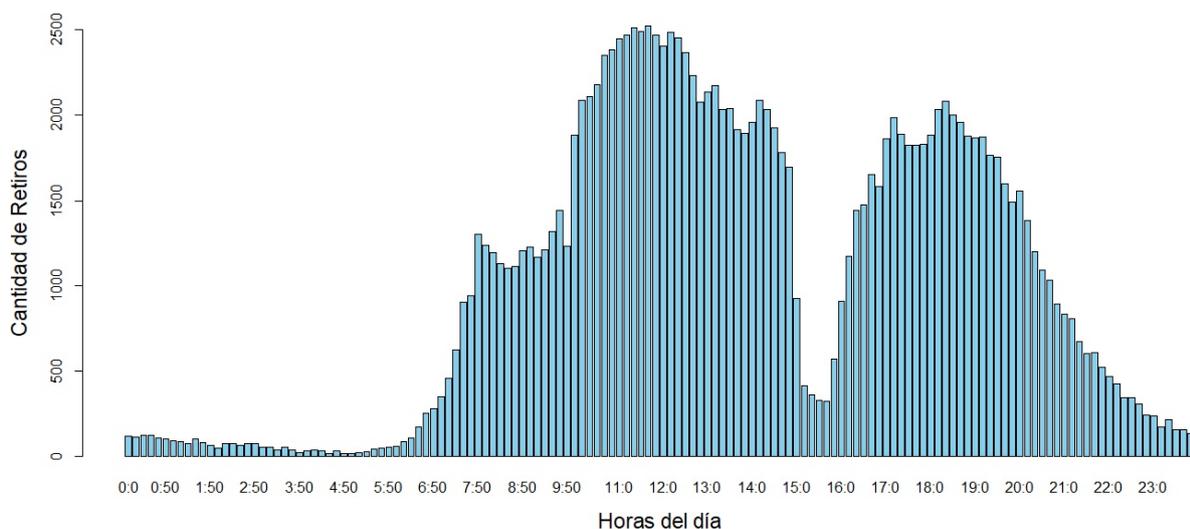


Figura 1.3: Diagrama de barras para el total de extracciones del cajero b309 en tramos de a 10 minutos

En el gráfico anterior se puede apreciar que durante la madrugada es cuando menos cantidad de extracciones se realizan. A partir de las seis de la mañana aproximadamente se visualiza un crecimiento exponencial hasta las ocho donde comienza una meseta hasta las diez de la mañana cuando comienza a verse nuevamente un crecimiento importante, llegando a los máximos niveles rondando el mediodía.

Luego del mediodía la actividad es intensa y alrededor de las cuatro de la tarde se observa una abrupta disminución de las extracciones. Esto llama mucho la atención pero tiene una clara explicación, como se mencionó anteriormente, estos datos pertenecen a una de las redes de cajeros de la República Argentina, donde el horario bancario de atención al público es de lunes a viernes entre las 10:00 y las 15:00 hs. Al cerrar las sucursales se realizan los arqueos de caja y conciliaciones a nivel general. Posteriormente, a las 16 hs, la red de cajeros efectúa un balance de operaciones y por lo tanto permanece inactiva durante media hora aproximadamente. Por este motivo se ven afectados todos los cajeros pertenecientes a la red, no solo los que se ubican en las sucursales de los bancos.

Inmediatamente después de esta caída, se presenta un crecimiento potencial hasta llegar a un pico cercano a las siete de la tarde, aunque no tan grande como el del mediodía. Finalmente la cantidad retiros vuelve a decaer hasta finalizar el día.

Es importante tener en cuenta que esta es una descripción del comportamiento de la variable a grandes rasgos, ya que es el total de los retiros de dinero realizados durante

los 805 días de medición. Como se verá más adelante, estos días tienen comportamientos diferentes entre sí.

A continuación se presenta el mismo gráfico que el de la figura 1.3 pero con el promedio de extracciones agrupados por hora. Este agrupamiento elimina tramos sin extracciones y suaviza los máximos relativos.

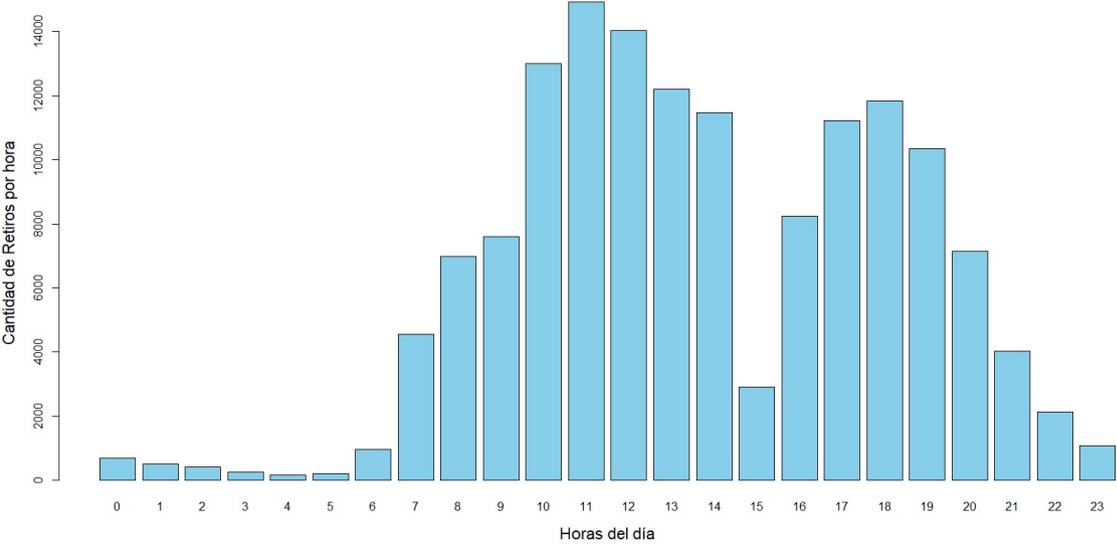


Figura 1.4: Total de retiros promedio del cajero b309 en tramos de una hora a lo largo del día

A partir de aquí se trabajará con la cantidad de extracciones agrupadas tanto de a diez minutos como por hora según se entienda conveniente.

Capítulo 2

Marco Teórico

En virtud de la descripción de los datos y en función de los objetivos expuestos anteriormente, en este capítulo se buscará describir el fenómeno de interés a partir de un marco probabilístico que permita modelarlo. Para ello se sistematizan definiciones y proposiciones de la bibliografía de referencia, en particular de ([Denis Bosq, 1996](#)) y ([Lefebvre, 2007](#)).

Como fue expuesto en el capítulo anterior, se cuenta con un grupo de datos agregados, pues no se tienen los tiempos individuales de las extracciones de dinero sino que se tiene el dato agrupado de a diez minutos.

2.1. Procesos puntuales y de conteo

A partir de la extracciones que se suceden en el cajero se construye un proceso puntual.

Un **Proceso Puntual** es una secuencia de eventos que ocurren de modo aleatorio en algún espacio subyacente, que puede ser por ejemplo el tiempo o el espacio.

Asociado a este proceso puntual se puede definir un proceso de conteo según las definiciones vertidas por Bosq y Nguyen en ([Denis Bosq, 1996](#)) p.79. Se considera a las extracciones de dinero del cajero como una secuencia de eventos que ocurren en instantes aleatorios T_1, T_2, \dots, T_n y $\lim_{n \rightarrow \infty} T_n = \infty$ con probabilidad uno.

Esto implica que el registro de los eventos comienza en el tiempo cero, que dos eventos no pueden ocurrir al mismo tiempo y que los fenómenos observados tienen lugar durante un largo período.

Tener en cuenta que el tiempo cero no es un tiempo de llegada, es decir que no puede

ocurrir un evento en el instante cero, pues se asume que la distribución de T_n es continua y $T_1 > 0$.

En este caso los retiros de dinero ocurren a lo largo del tiempo en instantes aleatorios pero se desconocen los momentos exactos donde acontecen la extracciones, se tiene la cantidad de eventos que ocurren cada diez minutos. Es decir que se trata de un proceso puntual con datos agrupados.

Por lo tanto para describir al fenómeno de las extracciones del cajero (T_n) un método conveniente es considerar al **proceso de conteo** ($N_t, t \geq 0$) asociado al proceso puntual descrito anteriormente, donde N_t representa a la cantidad de eventos que ocurren en el intervalo $[0, t]$.

$(N_t, t \geq 0)$ y $(T_n, n = 1, 2, \dots)$ contienen la misma información ya que con probabilidad uno se cumple que

$$N_t = \sup\{n/n = 0, 1, 2, \dots; T_n \leq t\}, t \geq 0 \quad (2.1)$$

con el supuesto de que $T_0 = 0$, mientras que

$$T_n = \inf\{t/t \geq 0, N_t \geq n\}; n = 0, 1, \dots$$

Estas relaciones se pueden apreciar en la Figura 2.1 que muestra el recorrido de una muestra de un proceso de conteo (N_t).

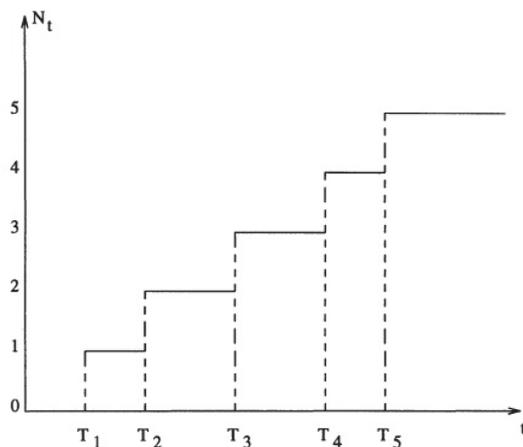


Figura 2.1: Recorrido típico de un muestreo de un proceso de conteo

En la Figura 2.1 se puede visualizar claramente como cada T_n marca el instante donde ocurre cada arribo y por lo tanto cada incremento de N_t es constante e igual a uno.

Aplicando al caso de los cajeros, estas mismas relaciones se pueden apreciar en la Figura 2.2, donde se muestran los incrementos asociados al proceso de conteo (N_t) de extracciones de un cajero, en este caso correspondientes a las extracciones observadas en un día concreto.

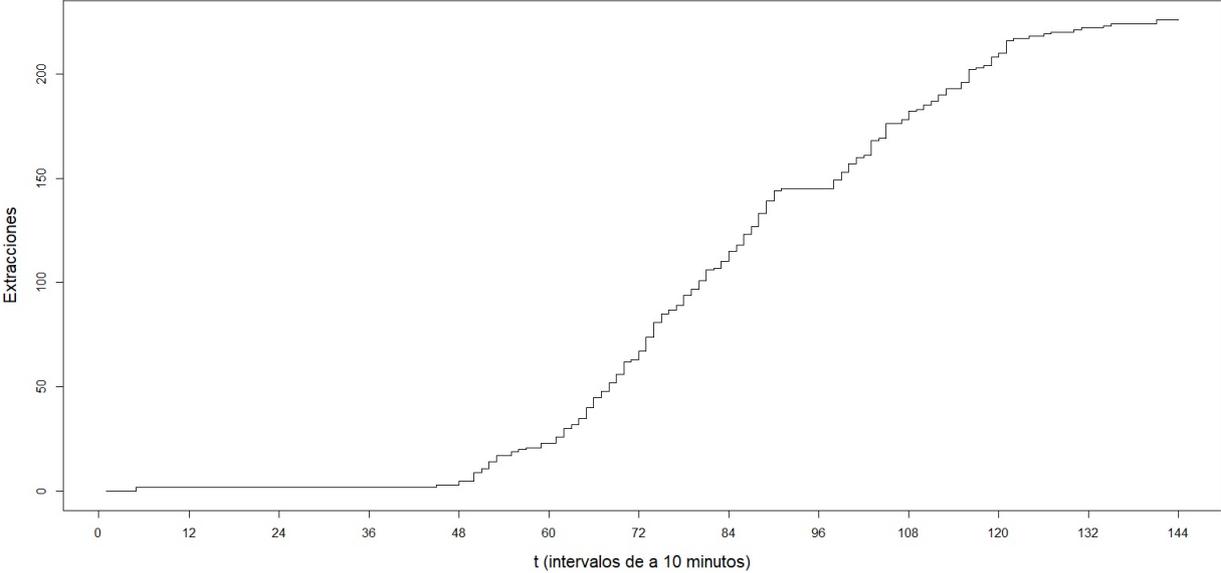


Figura 2.2: Incrementos de las extracciones del cajero b309 durante el día 08/03/2010

Como se puede ver en la Figura 2.2, en lugar de tener un incremento unitario en T_n cuando ocurre un evento como en la Figura anterior 2.1, los incrementos se dan cada diez minutos o cada múltiplos de diez y tienen tamaño variable. Esto es debido a que los datos con los que se cuentan están agrupados en tramos de a diez minutos y por lo tanto en cada uno de esos tramos pueden ocurrir varios eventos.

En la Figura 2.3 se pueden ver los incrementos por hora para el cajero b309, donde es más sencillo apreciar las diferencias de los incrementos cada sesenta minutos.

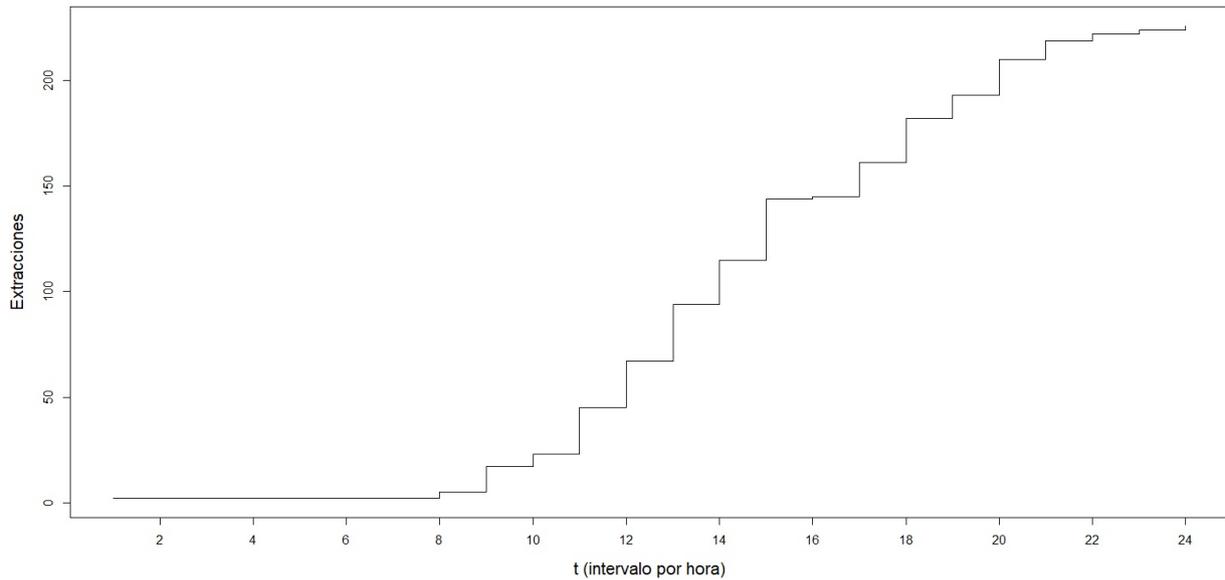


Figura 2.3: Incrementos por hora de las extracciones del cajero b309 durante el día 08/03/2010

Se está en presencia entonces de un proceso de conteo con datos agregados y con incrementos no estacionarios y que se asumen independientes dada la naturaleza del fenómeno.

En algunos casos no se cumple la independencia, por ejemplo dos extracciones correspondientes a una pareja que acudió junta al cajero. Sin embargo, estos casos se supone que ocurren excepcionalmente y como se cuenta con una gran cantidad de extracciones se asume que los eventos son independientes.

2.2. Procesos de Poisson no homogéneos

Un proceso de conteo puede modelarse como un Proceso de Poisson (PP), que consiste en contar la cantidad de arribos o realizaciones de cierto evento a lo largo del tiempo. El PP también es un proceso de nacimiento puro, y es el más usado para modelar filas de espera. Junto con el Proceso de Wiener son los procesos estocásticos más importantes para aplicaciones.

El PP puede ser homogéneo cuando la tasa de arribos es constante en el tiempo y puede ser no homogéneo cuando la intensidad de arribos varía en el tiempo y en lugar de

representarse con una tasa de intensidad, se representa con una función con el tiempo como variable.

Un proceso de conteo con incrementos independientes y no estacionarios puede verse como un **Proceso de Poisson No Homogéneo** (PPNH). Bosq y Nguyen en su trabajo (Denis Bosq, 1996) pp. 81, 84 y 91 introducen al PPNH mediante la siguiente definición.

Definición 2.2.1. Un proceso de conteo $(N_t, t \geq 0)$ es un PPNH si se cumplen los siguientes axiomas:

- $0 < T_1 < T_2 < \dots < T_n < \dots$ y $\lim_{n \rightarrow \infty} T_n = \infty$ con probabilidad uno.
- $(N_t, t \geq 0)$ es un proceso con incrementos independientes. Para cualquier $k \geq 2$ y $0 \leq t_0 < t_1 < \dots < t_k$ las variables aleatorias $N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, \dots, N_{t_k} - N_{t_{k-1}}$ son independientes.

- Para cualquier $t \geq 0$,

$$\lim_{h \rightarrow 0} \frac{P(N_{t+h} - N_t \geq 2)}{P(N_{t+h} - N_t = 1)} = 0$$

- Existe una función positiva localmente integrable $\lambda(\cdot)$ definida sobre \mathbb{R}^+ de modo que

$$\lim_{h \rightarrow 0^+} \frac{1 - P(N_{t+h} - N_t = 0)}{h} = \lambda(t), \quad t \geq 0$$

Bajo las condiciones anteriores se cumple que $N_t \sim Poisson(\Lambda(t))$ donde $\Lambda(t) = \int_0^t \lambda(s) ds$. Esta es una característica muy relevante, pues indica que la tasa del proceso varía en el tiempo, por lo que se trabaja con la tasa promedio.

A continuación se presenta la definición propuesta por Henderson (Henderson, 2003) de PPNH a partir de la Función de Intensidad del Proceso.

Definición 2.2.2. Sea $N = (N(t), t \geq 0)$ un proceso no decreciente de valor entero con $N(0) = 0$. Se dice que N es un **Proceso de Poisson No Homogéneo** con función de intensidad: $\lambda = (\lambda(t), t \geq 0)$ si

- el proceso N tiene incrementos independientes
- $\forall t \geq 0, P(N(t+h) - N(t) \geq 2) = o(h)$
- $\forall t \geq 0, P(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$

Nota: $f(h) = o(g(h))$ si $\frac{f(h)}{g(h)} \xrightarrow{h \rightarrow 0^+} 0$ y $f(h) = O(g(h))$ si $\frac{f(h)}{g(h)}$ está acotada en h

Definición 2.2.3. Asociada a la función de intensidad de un PPNH se define la función de intensidad acumulada como

$$\Lambda = (\Lambda(t), t \geq 0) \quad (2.2)$$

donde

$$\Lambda(t) = \int_0^t \lambda(s) ds \quad \forall t \geq 0 \quad (2.3)$$

2.2.1. Propiedades del PPNH

La segunda condición de la definición 2.2.2 implica que el proceso $\{N(t), t \geq 0\}$ no tiene incrementos estacionarios a menos que $\lambda(t) = \lambda > 0$. En este caso, $\{N(t), t \geq 0\}$ es un proceso de Poisson homogéneo, con tasa λ .

Al igual que en el caso particular en el que la tasa promedio de ocurrencia de eventos es constante, la cantidad de eventos que suceden en un intervalo dado tiene una distribución Poisson.

Esta propiedad es de fundamental importancia para el presente trabajo, en el siguiente capítulo se desarrolla su aplicación metodológica y la demostración se puede consultar en el apéndice B.1

Proposición 2.2.4. *Sea $\{N(t), t > 0\}$ un PPNH con función de intensidad $\lambda(t)$, se cumple que*

$$N(t+h) - N(t) \sim \text{Poisson}(\Lambda(t+h) - \Lambda(t)) \quad \forall t, h \geq 0 \quad (2.4)$$

donde

$$\Lambda(t) = \int_0^t \lambda(\tau) d\tau \quad (2.5)$$

Sea T_1 la variable aleatoria que representa el tiempo de llegada del primer evento del proceso $\{N(t), t > 0\}$. Se calcula la distribución de T_1 dado que $N(t) = 1$.

Proposición 2.2.5. *Sea $\{N(t), t > 0\}$ un PPNH con función de intensidad $\lambda(t)$. La función de densidad de la probabilidad de la variable aleatoria $T_1 | \{N(t) = 1\}$ está dada por*

$$f_{T_1}(s) = \frac{\lambda(s)}{m(t)} \quad \text{para } 0 < s \leq t \quad (2.6)$$

2.2.2. Ejemplos de PPNH

En el artículo (Leemis, 2004), Leemis menciona las aplicaciones más comunes. La más conocida es la *Teoría de colas*, que estudia los arribos a una fila de espera. Estos pueden ser personas que llegan a una caja en un supermercado, llamadas que recibe un call center o cualquier otro arribo aleatorio que no sea a tasa constante en el tiempo, ya que en ese caso sería homogéneo.

Otra aplicación es la “Teoría de la confiabilidad”, en la cual se estudian los tiempos de falla en un sistema reparable. Por ejemplo, las computadoras fabricadas con fallas dentro de una misma serie.

Como éstos, existen incontables ejemplos de aplicaciones en diversos campos, tanto de la ciencia como del mundo empresarial y en el campo gubernamental. Siempre y cuando la función de intensidad sea variable en el tiempo, estaremos hablando de PPNH, en los casos en los que la función de intensidad no varíe en el tiempo, será posible representarla con una tasa de intensidad y estaremos hablando de PP homogéneos.

Capítulo 3

Metodología

Teniendo en cuenta el marco probabilístico expuesto en el capítulo anterior y con el propósito de caracterizar el proceso de interés, a continuación se verá la metodología de inferencia estadística que se ajusta al caso de estudio.

Con el propósito de caracterizar el PPNH se busca estimar la función de intensidad y la función de intensidad acumulada para un intervalo finito que cubre todo un día, pues los cajeros automáticos están abiertos las veinticuatro horas y se cuenta con la información.

3.1. Estimaciones paramétricas y no paramétricas para PPNH

La inferencia estadística clásica consta de técnicas basadas en supuestos específicos sobre la distribución de las variables observables de interés en una población. En el enfoque clásico de la estadística paramétrica se indica la forma funcional de la distribución de probabilidad de las variables de interés, la cual depende de un vector de parámetros desconocidos.

Para modelar los PPNH se han desarrollado varios modelos paramétricos a lo largo de la historia vinculados a diferentes ramas de actividad, algunos por ejemplo son derivados de estudios de la rama actuarial que modelan la tasa de riesgo.

McCollin, en su artículo ([McCollin, 2014](#)), menciona que los modelos más simples son los que consideran constante al número esperado de eventos, es decir que la función de intensidad no depende del tiempo y por lo tanto es una tasa de intensidad. Estos casos en realidad son procesos de Poisson homogéneos. Los dos mejores modelos que se

conocen son el *Duane plot* que se aproxima con la distribución de Weibull y el modelo lineal logarítmico que utiliza una aproximación a partir de mínimos cuadrados no lineales. También se han planteado funciones de intensidad de PPNHs asociadas a funciones de densidad de probabilidad conocidas tales como normal estándar, gamma, Erlang, Wald, Burr, Logística, Pareto, exponencial inversa y normal inversa. Otras formas de estimar la función de intensidad ha sido mediante el desarrollo de modelos logarítmicos.

Para el caso del presente trabajo no se aplica ninguna de las estimaciones paramétricas anteriormente mencionadas, ya que, como se vio en la descripción de los datos del primer capítulo, la intensidad con la que se suceden las extracciones no se asemeja a ninguna familia paramétrica conocida. Tampoco se encontraron antecedentes que modelen el fenómeno de las extracciones de cajeros automáticos.

Por lo tanto, para modelar el PPNH que caracteriza a las extracciones del cajero dada la forma en que se dispone de los datos, se entiende que un enfoque no paramétrico es el más adecuado para los objetivos de este trabajo, permite cierta flexibilidad pues no impone caracterizar a priori la forma funcional de distribución del fenómeno de interés.

En un procedimiento no paramétrico se realiza una inferencia sin una función de distribución de referencia. Ya sea para testear o para estimar, los métodos no paramétricos se basan en funciones construidas a partir de los datos observados de la muestra.

3.2. Caracterización del PPNH en presencia de datos agrupados

Como se ha mencionado a lo largo de este informe, los datos disponibles son agrupados puesto que la información que se tiene es la cantidad de extracciones realizadas en tramos de a diez minutos. No se conocen los momentos t en los que se sucede el evento, los cuales representarían los tiempos de llegada de los clientes al cajero o bien los momentos en los que se realizan las extracciones.

Si se tuvieran los datos individuales de los tiempos en los que se efectúan los retiros de dinero, la estimación sería completamente diferente. Pero la presencia de datos agrupados obliga a pensar el problema a partir del conteo de las extracciones por intervalos predefinidos de tiempo.

Para la construcción del modelo se se busca estimar la función de intensidad $\lambda(t)$ y la función de intensidad acumulada $\Lambda(t)$ asociadas al proceso N para el intervalo $[0, T)$. Esto sirve para caracterizar el proceso y brinda la posibilidad de generar realizaciones en una simulación de eventos discretos.

En este caso, si se consideraran los tramos de a diez minutos, entonces $T = 144$ pues en cada hora hay seis tramos de diez minutos.

Pero como se vio en el análisis exploratorio, al ser tramos pequeños la mayoría cuentan con cero eventos, sobre todo en las horas de la noche. Por este motivo y sobre todo a efectos prácticos del trabajo, se opta por unificar los seis tramos de cada hora y trabajar con 24 subintervalos de 60 minutos cada uno, es decir, subintervalos de una hora de duración para la estimación de $\lambda(t)$, ya que se podría caer en una subestimación. Para el caso de la estimación de $\Lambda(t)$ se vuelve a trabajar con los tramos originales, ya que lo importante en ese caso es la acumulación de la ocurrencia de eventos.

Por lo tanto el problema consiste en estimar $\lambda(t)$ para el intervalo $[0, 24)$ y $\Lambda(t)$ para el intervalo $[0, 144)$ para cada día a partir de $n = 805$ realizaciones del proceso, que corresponden a los 805 días observados.

Dicho intervalo de interés se subdivide en 24 subintervalos $[a, b)$ de una hora de duración. Al medir en horas en lugar de minutos, la notación se simplifica, siendo la cantidad de subintervalos $k = 24$, el largo de cada intervalo $\delta = 1$ y por lo tanto $T = k\delta = 24 \times 1 = 24$.

A la cantidad de eventos observados en el subintervalo $[a, b)$ durante la i -ésima realización se denomina $N_i(a, b)$ y en este caso entonces se cuenta con $N_1(0, 1); \dots; N_1(23, 24); \dots; N_{805}(0, 1); \dots; N_{805}(23, 24)$.

Se puede visualizar la notación resumida en el glosario que se encuentra en el apéndice [A](#).

3.2.1. Antecedentes

Existen varias propuestas de estimadores para la función de intensidad de un PPNH. Las formas paramétricas suelen elegirse para incorporar información previa sobre la función de intensidad. Tienen la desventaja que no se conocen resultados generales sobre su comportamiento asintótico.

Además, los estimadores no van a converger a la verdadera función de intensidad si ésta no se encuentra dentro de la clase paramétrica asumida. Otra desventaja, más importante es que las estimaciones mediante máxima verosimilitud y otras técnicas implican tareas computacionales complejas.

Aunque la mayoría de estos estimadores se basan en un número fijo de parámetros, entonces el espacio de memoria que se precisa no aumenta a medida que aumenta la cantidad de realizaciones.

Algunas de estas estimaciones paramétricas fueron desarrolladas por Kao y Chang en

1988, Lee y Wilson en 1991, Kuhl, Wilson y Johnson en 1997 y Kuhl y Wilson en 2000.([Henderson, 2003](#))

Por otro lado, en 1976 Lewis y Shedler desarrollaron un estimador no paramétrico basado en las técnicas de estimación de Kernel, dicho estimador es para casos donde se cuenta con los tiempos exactos de los eventos, no para datos agrupados.([Lewis and Shedler, 1979](#))

En 1977, Kuhl y Wilson desarrollaron un estimador híbrido paramétrico/no paramétrico utilizando oscilaciones.([Kuhl and Wilson, 2001](#))

En 1991 Leemis presentó un estimador no paramétrico para $\Lambda(t)$ considerando realizaciones independientes del proceso estocástico N en $[0, T)$. Establece un comportamiento asintótico del estimador a través de la Ley Fuerte de los Grandes Números y el Teorema Central del Límite.([Leemis, 1991](#)) Este trabajo es extendido por Arkin y Leemis en el año 2000 aceptando realizaciones parciales del proceso N del tramo $[0, T)$, es decir que no es necesario que todas las realizaciones cubran la totalidad del tramo $[0, T)$.([Arkin and Leemis, 2000](#))

Según Henderson en ([Henderson, 2003](#)), en estos dos casos -Leemis 1991 y Arkin y Leemis 2000- el estimador se calcula fácilmente a partir de los datos y existen resultados de su comportamiento asintótico. Además, se pueden generar simulaciones a partir de la función de intensidad estimada. Una desventaja importante es que -a diferencia de los estimadores paramétricos con una cantidad fija de parámetros- requieren que se guarden en la memoria todos los tiempos de los eventos observados en cada realización para permitir la generación de simulaciones del proceso N .

Leemis considera la división del intervalo de interés en subintervalos de ancho δ_n , es decir que en cada realización n , el ancho puede variar. En su artículo ([Henderson, 2003](#)), Henderson opta por trabajar con el estimador de la función de intensidad propuesto por Law y Kelton en el año 2000 en ([Law and Kelton, 2013](#)), donde los subintervalos se asumen de igual tamaño, es decir $\delta_n = \delta$. Elige esta metodología porque permite trabajar con información disponible de forma desagregada o agregada, lo cual resulta de gran interés a los efectos del presente trabajo pues se cuenta con datos agregados. Además, no requiere procesar todas las observaciones sino que alcanza con el total por intervalo, por lo tanto es posible utilizarlo en este caso de extracciones de cajeros, donde se cuenta con datos agrupados.

Se plantea una disyuntiva al elegir el ancho del intervalo δ_n , ya que dicha cantidad regula el grado de suavización de la estimación funcional que se quiere llevar a cabo. Si es muy grande, el comportamiento de los tiempos de llegada se homogeneiza. Si es muy chico, puede haber muchos intervalos sin ningún arribo.

Henderson propone dos métodos para generar Procesos de Poisson No Homogéneos con n realizaciones del proceso. Para ello asume que la función de intensidad se divide en intervalos de igual tamaño. Por esto el elegir el tamaño de los tramos o intervalos es una decisión subjetiva que debe considerar que el tamaño no puede ser ni muy grande ni muy pequeño.

Si es muy grande, en los tramos más heterogéneos la función de la tasa de intensidad se “suaviza”, es decir, que se pierden grandes intensidades o pequeñas intensidades en un promedio que no representa la realidad. A su vez, si el intervalo es muy pequeño puede ocurrir un sobreajuste de los datos y de este modo se generarían muchos tramos con tasa igual a cero.

Los métodos propuestos por Henderson en (Henderson, 2003) son los siguientes:

- El Procedimiento por inversión, que se basa en la estimación de la función de intensidad Λ . El sesgo al usar intervalos de ancho fijo δ es del orden de δ^2 . Cuando $\delta = \delta_n$ varía apropiadamente con el número de realizaciones observadas n , el estimador Λ_n converge a Λ a una tasa $n^{1/2}$ cuando $n \rightarrow \infty$, que es la misma velocidad a la que converge el estimador propuesto por Leemis. Para cada t fijo, $\Lambda_n(t)$ cumple el Teorema del Límite Central (TLC) al igual que el estimador de Leemis, con la diferencia que el propuesto por Henderson requiere menos espacio de almacenamiento.
- El segundo método es el mismo que introdujeron Lewis y Shedler que se basa en estimar la función de la tasa λ por afinamiento (“*thinning*”) y consiste en una estimación de la función de intensidad λ . La velocidad de convergencia más lenta surge porque Λ_n se puede expresar como el número esperado de eventos en $[0, t)$ y por lo tanto, se puede estimar mediante un promedio muestral, sin embargo esto no aplica para $\lambda(t)$

3.3. Estimaciones de la función de intensidad y función de intensidad acumulada

Como se menciona al principio del presente capítulo, al intervalo de interés se lo divide en subintervalos. Como la tasa por unidad de tiempo de los eventos que ocurren en cada intervalo es el promedio de la Función de Intensidad del intervalo, el estimador máximo verosímil es el número de eventos promedio que ocurren en cada intervalo.

La función de intensidad acumulada asociada, es una función lineal a tramos que acumula el total de arribos estimados para cada intervalo.

3.3.1. Estimación propuesta por Henderson

En el artículo (Henderson, 2003), Henderson estima la función λ para un intervalo finito $[0, T)$ siendo T una constante fija, con el propósito de generar realizaciones en una simulación de eventos discretos. Si bien se utilizan intervalos $[a, b)$ por definición, es lo mismo considerar o no los límites del intervalo ya que la probabilidad de que ocurra un evento en un instante preciso del tiempo es cero.

Se asume que se conocen varias observaciones de realizaciones *iid* de N para el intervalo $[0, T)$, aunque algunos resultados se basan en versiones agregadas de estos datos.

Henderson considera en (Henderson, 2003) el problema de estimar la función de intensidad acumulada Λ en el intervalo $[0, T)$. Para eso, divide al intervalo en subintervalos de ancho fijo $\delta > 0$ y considera n realizaciones del PPNH N que crece sin límite.

De esta forma, es irrelevante si los datos están de forma agregada o no, ya que lo que nos interesa es tener para cada realización el número de eventos ocurrido en cada intervalo $[(k-1)\delta, k\delta]$ para $k = 1, 2, \dots, \lceil T/\delta \rceil$ donde $\lceil x \rceil$ indica el primer número entero mayor o igual que x .

A continuación se define $\widehat{\Lambda}_n(t)$, el estimador de $\Lambda(t)$ basado en n realizaciones independientes del proceso N en el intervalo $[0, T)$.

Definición 3.3.1. Sea $N_i(a, b)$ el número de eventos dentro del intervalo $[a, b)$ en la i -ésima observación independiente de N . Para $t \geq 0$ sea $\ell(t) = \lfloor \frac{t}{\delta} \rfloor \delta$ de modo que t pertenece al intervalo $[\ell(t), \ell(t) + \delta]$, donde $\lfloor x \rfloor$ indica el número entero menor o igual que x .

Entonces para $t \in [0, T)$:

$$\widehat{\Lambda}_n(t) = \frac{1}{n} \sum_{i=1}^n N_i(0, \ell(t)) + \frac{t - \ell(t)}{\delta} \frac{1}{n} \sum_{i=1}^n N_i(\ell(t), \ell(t) + \delta) \quad (3.1)$$

La expresión $\lfloor \frac{t}{\delta} \rfloor$ representa a la parte entera del cociente entre t y δ . Donde t es el tiempo y δ el ancho fijo de los subintervalos. Por lo tanto cuando t es el comienzo de un intervalo o punto de quiebre, se cumple que $\ell(t) = t$.

El hecho de trabajar con datos agrupados hace que se deba tomar una metodología estadística para poder estimar la función de intensidad acumulada a lo largo de todo $[0, T)$ y lo que hacen Leemis y Henderson es en cada extremo de los subintervalos promediar la cantidad de eventos observados en las n realizaciones y realizar una aproximación lineal a tramos de Λ interpolando para los valores $t \neq \ell(t)$.

Definición 3.3.2. Para $t \in [0, T)$ se define la aproximación lineal a tramos de Λ a la siguiente expresión:

$$\boxed{\widehat{\Lambda}(t) = \Lambda(\ell(t)) + \frac{t - \ell(t)}{\delta} \times [\Lambda(\ell(t) + \delta) - \Lambda(\ell(t))]} \quad (3.2)$$

Esto es un aspecto fundamental en el desarrollo del presente trabajo. Concretamente, $\widehat{\Lambda}(t)$ es igual a $\Lambda(t)$ en los puntos de quiebre $t : t = \ell(t)$ y se realiza una interpolación lineal entre cada uno de estos puntos de quiebre.

Se indica con $X_n \Rightarrow X$ a la convergencia en distribución de una sucesión de variables aleatoria $\{X_n\}_{n \geq 1}$ a la variable aleatoria X ; y se indica con $N(0, 1)$ a una variable aleatoria distribuida normalmente con media μ y varianza σ^2 .

Proposición 3.3.3. *Según las definiciones vertidas anteriormente se cumple que:*

1. $\sup_{t \in [0, T)} |\widehat{\Lambda}_n(t) - \widehat{\Lambda}(t)| \rightarrow 0$ casi seguramente cuando $n \rightarrow \infty$
2. $\forall t \in [0, T), \sqrt{n}(\widehat{\Lambda}_n(t) - \widehat{\Lambda}(t)) \Rightarrow \sigma(t)N(0, 1)$ cuando $n \rightarrow \infty$, donde

$$\sigma^2(t) = \Lambda(\ell(t)) + \left(\frac{t - \ell(t)}{\delta} \right)^2 \times [\Lambda(\ell(t) + \delta) - \Lambda(\ell(t))]$$

3. $\forall n \geq 1, E\widehat{\Lambda}_n(t) = \widehat{\Lambda}(t)$. Si λ es continuamente diferenciable en un intervalo abierto que contiene $[\ell(t), \ell(t) + \delta]$ entonces

$$|E\widehat{\Lambda}_n(t) - \Lambda(t)| \leq |\lambda'(\xi)|\delta^2$$

para algún $\xi \in [\ell(t), \ell(t) + \delta]$

La proposición 3.3.3 -cuya demostración se presenta en el Apéndice B.3- muestra el buen rendimiento del estimador propuesto en (Law and Kelton, 2013) por Law y Kelton con subintervalos de ancho fijo. De todas formas, se aclara que presenta un buen comportamiento solo si están disponibles los datos agregados.

El estimador $\widehat{\Lambda}_n(t)$ converge al valor real de la función de intensidad acumulada $\Lambda(t)$ en puntos de corte, y en general converge a $\widehat{\Lambda}(t)$ a una tasa $n^{1/2}$. Además, su sesgo es independiente de n , típicamente de orden δ^2 y pequeño cuando $|\lambda'|$ es pequeña.

Se considera entonces si es posible o no obtener un estimador consistente de $\Lambda(t)$ si se elige un intervalo de ancho $\delta = \delta_n$ para ser una función de n , el número de realizaciones observadas de N .

Henderson considera la estimación de la función de intensidad λ para el intervalo $[0, T)$. Esta función se puede utilizar para generar realizaciones simuladas del PPNH mediante un procedimiento de afinamiento propuesto por Lewis y Shedler en (Lewis and Shedler, 1979). La naturaleza del estimador de la función de intensidad es de tal modo que el afinamiento proporciona un procedimiento de generación rápido.

Para ver porqué, Henderson propone definir a λ_n el estimador de la función de intensidad. El método de Lewis y Shedler primero genera un candidato a tiempo de evento T^* y luego acepta el tiempo del evento con probabilidad $\lambda_n(T^*)/\lambda^*$, donde λ^* es un límite superior en λ_n . Por lo tanto este método requiere un cálculo rápido de $\lambda_n(\cdot)$.

Al definirse λ_n por partes en intervalos de igual tamaño, el intervalo que contiene un tiempo dado T^* se puede computar en $O(l)$ veces a medida que n crece y por lo tanto $\lambda_n(T^*)$ se puede calcular rápidamente. Esta observación es igual a la mencionada por Leemis en (Leemis, 1991) relacionada con la eficiencia de generar el proceso N basado en el estimador de la función de intensidad acumulativa no paramétrica.

Nuevamente, se considera primero el caso en el que se fija la longitud δ del subintervalo.

Definición 3.3.4. Se define a la función de intensidad agregada como:

$$\lambda(t) = \frac{1}{\delta} \int_{\ell(t)}^{\ell(t)+\delta} \lambda(s) ds \quad (3.3)$$

que es constante en cada intervalo de la forma $[(k-1)\delta, k\delta)$ para $k \geq 1$.

Definición 3.3.5. se define el estimador de la función de intensidad como:

$$\widehat{\lambda}(t) = \frac{1}{n\delta} \sum_{i=1}^n N_i(\ell(t), \ell(t) + \delta) \quad (3.4)$$

A continuación se presenta la proposición 3.3.6, la cual refleja la proposición 3.3.3 en cuanto a que describe el comportamiento de $\widehat{\lambda}_n$ para un intervalo de ancho fijo δ .

Proposición 3.3.6. Para todo $t \in [0, T)$ se cumple que:

1. $\widehat{\lambda}_n(t) \rightarrow \widehat{\lambda}(t)$ casi seguramente cuando $n \rightarrow \infty$ y

$$n^{1/2}(\widehat{\lambda}_n(t) - \widehat{\lambda}(t)) \Rightarrow \delta^{-1}N(0, \Lambda(\ell(t) + \delta) - \Lambda(\ell(t)))$$

cuando $n \rightarrow \infty$

2. Para todo $n \geq 1$, $E\widehat{\lambda}_n(t) = \widehat{\lambda}(t)$. Si λ es continua en un intervalo abierto que contiene $[\ell(t), \ell(t) + \delta]$, entonces $E\widehat{\lambda}_n(t) = \lambda(\zeta)$ para algún ζ en el intervalo $[\ell(t), \ell(t) + \delta]$.

El sesgo en $\widehat{\lambda}_n$ depende del comportamiento local de $\lambda(\cdot)$ en un entorno de t . Se puede encontrar más evidencia en el trabajo (Henderson, 2003) de Henderson.

3.3.2. Estimación propuesta por Leemis

Finalmente se opta por trabajar con la estimación del modelo planteada por Leemis en (Leemis, 2004). Para presentar su metodología tomaremos las definiciones vertidas en dicho artículo pero con la notación similar a la de Henderson para evitar confusiones.

Leemis asume en (Leemis, 2004) k realizaciones independientes de un PPNH con función de intensidad acumulada $\Lambda(t)$ -lo que para Henderson son n realizaciones- tomadas dentro del intervalo $(0, T]$ donde T es un real fijo, en este estudio al considerarse el análisis diario, el intervalo de tiempo refiere al día.

Las técnicas de estimación y los algoritmos para la simulación de un PPNH con datos agregados presentados por Leemis, requieren que las unidades de tiempo sean consistentes con los datos. Por ejemplo, si el intervalo de interés es de 08:00 a 14:30 hs, entonces el intervalo medido en horas será $(0; 6, 5]$.

El intervalo de tiempo $(0, T]$ se particiona en k subintervalos:

$$(a_0, a_1], (a_1, a_2], \dots, (a_{k-1}, a_k]$$

siendo $a_0 = 0$ y $a_k = T$. El estimador que plantea Leemis no requiere que los subintervalos sean de igual ancho, pero en este estudio los anchos de los intervalos son siempre iguales.

Sean N_1, N_2, \dots, N_k los números totales de eventos observados en cada subintervalo para las n realizaciones. Se asume que la población PPNH tiene una función de intensidad $\lambda(t)$ que es constante a tramos en cada subintervalo de la partición $(a_0, a_1], (a_1, a_2], \dots, (a_{k-1}, a_k]$.

Como la función de intensidad promedio en el intervalo $(a_{i-1}, a_i]$ es la tasa por unidad de tiempo de los eventos que ocurren en ese intervalo, el estimador de máxima verosimilitud

es el número de eventos promedio que ocurren en dicho intervalo normalizado por el largo de cada intervalo.

$$\widehat{\lambda}(t) = \frac{N_i}{n(a_i - a_{i-1})} \quad (3.5)$$

siendo $a_{i-1} < t \leq a_i$, para $i = 1, 2, \dots, k$.

La función de intensidad acumulada asociada es una función lineal a tramos continua en $(0, T]$:

$$\widehat{\Lambda}(t) = \left(\sum_{j=1}^{i-1} \frac{N_j}{N} \right) + \frac{n_i(t - a_{i-1})}{N(a_i - a_{i-1})} \quad (3.6)$$

siendo $a_{i-1} < t \leq a_i$, para $i = 1, 2, \dots, k$.

Si no hay eventos que se hayan observado en el intervalo i , es decir que $N_i = 0$, entonces la función de intensidad estimada es cero en el intervalo i y la función de intensidad acumulada permanece constante.

Las propiedades asintóticas de este estimador para cuando el ancho de los intervalos son siempre iguales y fijos son las mismas que se vieron en la sección anterior planteadas por Henderson en (Henderson, 2003).

En el Apéndice C puede encontrarse un algoritmo propuesto en (Leemis, 2004) que simula las trayectorias de un PPNH con datos agrupados, lo cual es útil para modelar un intervalo donde pueden o no ocurrir eventos.

3.4. Intervalos de confianza

En el artículo (Leemis, 2004), Leemis propone un intervalo de confianza para el estimador $\widehat{\Lambda}(t)$ que son asintóticamente exactos en los extremos de los subintervalos.

La función de intensidad poblacional $\lambda(t)$ no es constante a tramos en cada subintervalo $(a_{i-1}, a_i]$ de la partición arbitraria de $(0, T]$ en la mayoría de las aplicaciones. Como plantea Leemis, para el caso general $\widehat{\Lambda}(t)$ solo es consistente en los puntos extremos de cada intervalo a medida que $n \rightarrow \infty$ siendo n la cantidad de realizaciones observadas.

$$\widehat{\Lambda}(t) - z_{\alpha/2} \sqrt{\frac{\widehat{\Lambda}(t)}{n}} < \Lambda(t) < \widehat{\Lambda}(t) + z_{\alpha/2} \sqrt{\frac{\widehat{\Lambda}(t)}{n}} \quad (3.7)$$

para $0 < t \leq T$, donde $z_{\alpha/2}$ es el percentil $1 - \alpha/2$ de la distribución normal estándar, es solo asintóticamente exacta en los extremos de los intervalos

$$\lim_{n \rightarrow \infty} P \left(\widehat{\Lambda}(a_i) - z_{\alpha/2} \sqrt{\frac{\widehat{\Lambda}(a_i)}{n}} < \Lambda(a_i) < \widehat{\Lambda}(a_i) + z_{\alpha/2} \sqrt{\frac{\widehat{\Lambda}(a_i)}{n}} \right) = 1 - \alpha \quad (3.8)$$

para $i = 0, 1, \dots, k$.

3.5. Detección de períodos de inactividad

Una de las motivaciones iniciales de este trabajo, además de caracterizar el proceso que modela las extracciones que se efectúan durante el día fue la detección de períodos de inactividad de los cajeros.

Según la proposición 2.2.4, la distribución del número de eventos en el intervalo $[t, t + h)$ es la de una variable de Poisson con parámetro $\Lambda(t + h) - \Lambda(t)$, para todo $t, h \geq 0$.

$$N(t + h) - N(t) \sim \text{Poisson}(\Lambda(t + h) - \Lambda(t)), \quad \forall t, h \geq 0$$

Por lo tanto la probabilidad de inactividad del cajero en el intervalo $[t, t + h)$ está dada por

$$P(N(t + h) - N(t) = 0) = e^{-(\Lambda(t+h) - \Lambda(t))}.$$

A partir de este resultado, la estimación de la probabilidad de inactividad del cajero en el intervalo $[t, t + h)$ está dada por

$$\boxed{\widehat{P}(N(t + h) - N(t) = 0) = e^{-(\widehat{\Lambda}(t+h) - \widehat{\Lambda}(t))}} \quad (3.9)$$

Capítulo 4

Resultados

Según la metodología de estimación planteada en el capítulo anterior, se presentan los resultados obtenidos y las decisiones que, de acuerdo a éstos, se fueron tomando a lo largo del trabajo.

En principio se verán las estimaciones de la función de intensidad y de la función de intensidad acumulada del PPNH. Luego se analizará si es posible realizar una sola estimación para todos los días de la semana o si se deben realizar estimaciones individuales para cada día. A su vez se evaluará la posibilidad de estimar el comportamiento de las extracciones en diferentes momentos del mes.

Finalmente se intentará dar respuesta a la interrogante sobre los períodos de inactividad genuinos del cajero.

4.1. Estimación de la función de intensidad

Como se vio en el capítulo anterior, para caracterizar el PPNH se busca estimar la función de intensidad del proceso. Se tomará la estimación desarrollada por Leemis en su artículo (Leemis, 2004) p. 1156, donde se asume que la población PPNH tiene una función de intensidad $\lambda(t)$ que es constante a tramos en cada subintervalo de la partición $(a_0, a_1], (a_1, a_2], \dots, (a_{m-1}, a_m]$.

$$\hat{\lambda}(t) = \frac{N_i}{n(a_i - a_{i-1})}, \quad \text{para } t \in (a_{i-1}, a_i], \quad i = 1, 2, \dots, m.$$

Esta estimación se ve representada en el siguiente gráfico, donde se puede ver la distribución que tiene la variable Cantidad de Retiros a lo largo de las veinticuatro horas

del día en tramos de diez minutos. Para esta estimación se tuvo en cuenta al total de extracciones durante los dos años y medio para los que se dispone información, que son 805 días en total.

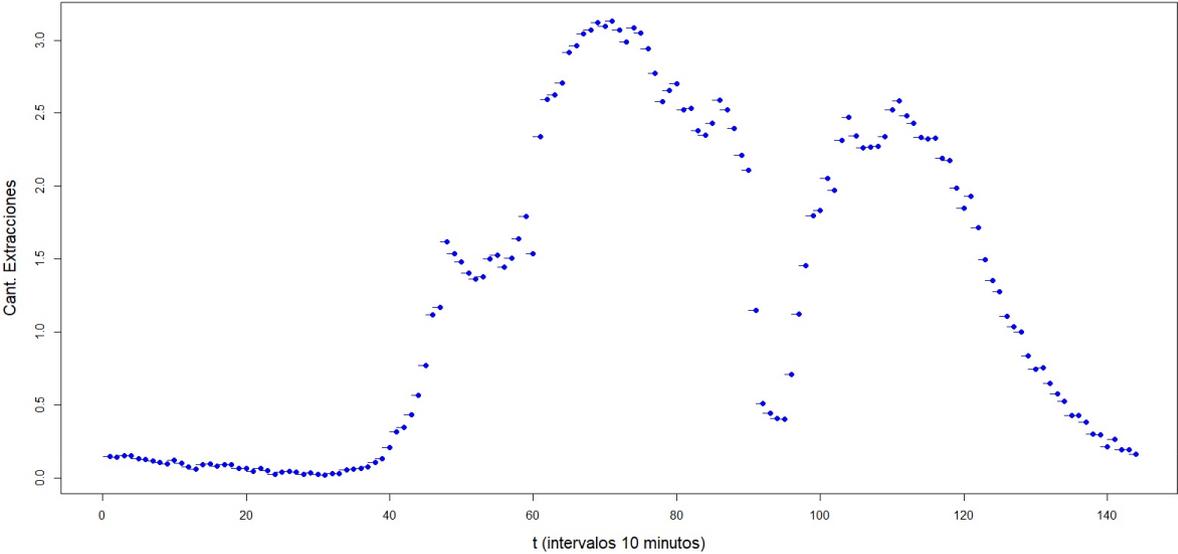


Figura 4.1: Estimación de la Función de Intensidad del cajero b309 en tramos de diez minutos

Al visualizar el gráfico de la función de intensidad, se aprecia que durante la madrugada las cantidad de extracciones es prácticamente nula, comenzando a despegar lentamente luego del intervalo 35, es decir a las seis de la mañana, donde se observa una mayor cantidad de extracciones, y ya con más velocidad a partir del intervalo 40, cercano a las siete de la mañana.

Esto ocurre hasta el intervalo 48, las ocho de la mañana, donde durante cuatro intervalos consecutivos la cantidad de extracciones disminuye, continuando posteriormente con un crecimiento importante hasta alcanzar el máximo cercano al mediodía. En el entorno del intervalo 95 se aprecia una caída muy importante, o sea alrededor de las cuatro de la tarde, donde el cajero casi llega a no tener extracciones¹ y vuelve a subir, generando una segunda cresta con pico a las siete de la tarde. A partir de allí las extracciones decrecen hasta finalizar el día prácticamente sin extracciones.

A continuación se presenta la misma estimación pero realizada a partir del conteo de extracciones en tramos de una hora.

¹El motivo de esta drástica caída se explica en la sección 1.4

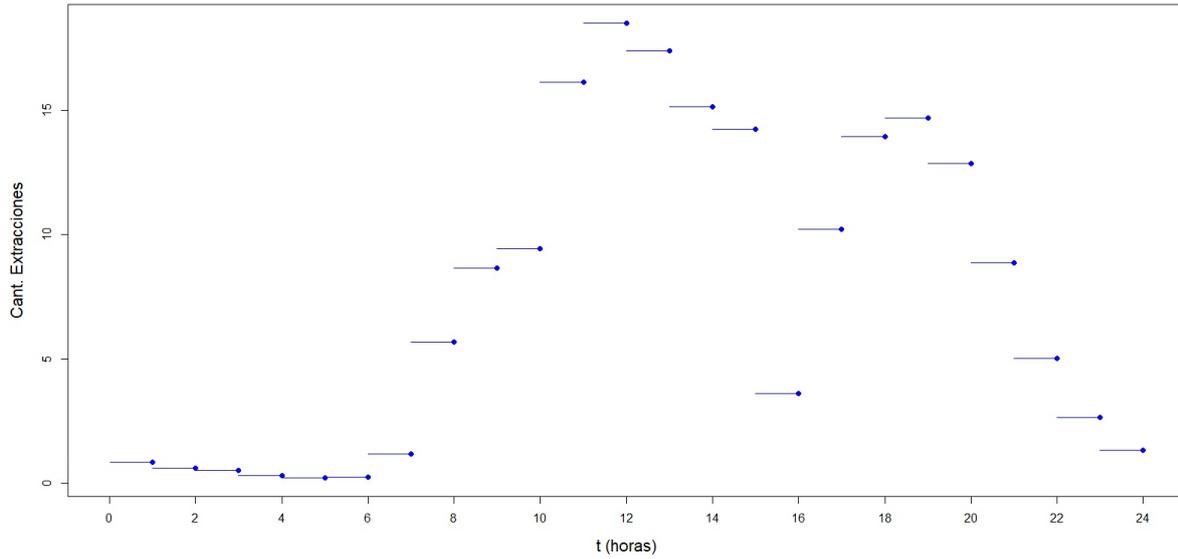


Figura 4.2: Estimación de la Función de Intensidad del cajero b309 en tramos de una hora

Como se puede ver en la figura 4.2, el gráfico de la función de intensidad estimada a partir de tramos de una hora, permite visualizar con mayor claridad que la cantidad estimada de extracciones es constante en cada intervalo. Sin embargo, se pierde continuidad respecto a la visualización del gráfico en la figura 4.1, y a su vez, si el tramo del intervalo es muy largo se pierde variación dentro del mismo, mientras que si el intervalo es muy pequeño se pueden tener varios de ellos sin la ocurrencia de eventos.

Como se había visto en la sección 1.4, se dividen los datos según los días de la semana, teniendo 115 observaciones para cada día. Por lo tanto, la función de intensidad separada por día de la semana se estima de la siguiente manera, a modo de ejemplo para el martes:

$$\hat{\lambda}_{martes}(t) = \frac{N_i}{115(a_i - a_{i-1})} \quad (4.1)$$

siendo $a_{i-1} < t \leq a_i$, para $i = 1, 2, \dots, 24$.

De la misma manera que para el martes, se estima la función de intensidad para cada día de la semana, ahora en lugar de tomar los tramos originales de a diez minutos, se consideran tramos de una hora de duración, es decir que se tienen 24 intervalos a lo largo del día.

A los efectos de una mejor visualización, a partir de ahora se graficará la función de intensidad estimada de manera continua, a partir de una interpolación lineal entre los

puntos finales de cada intervalo, ya que estos puntos son los que tienen una estimación consistente como se vio en el capítulo anterior.

A continuación, se presentan las estimaciones diarias superpuestas, la línea punteada representa la estimación de la función de intensidad genérica, para todos los días de la semana. Con diferentes colores se ve la estimación de la función de intensidad para cada día en particular.

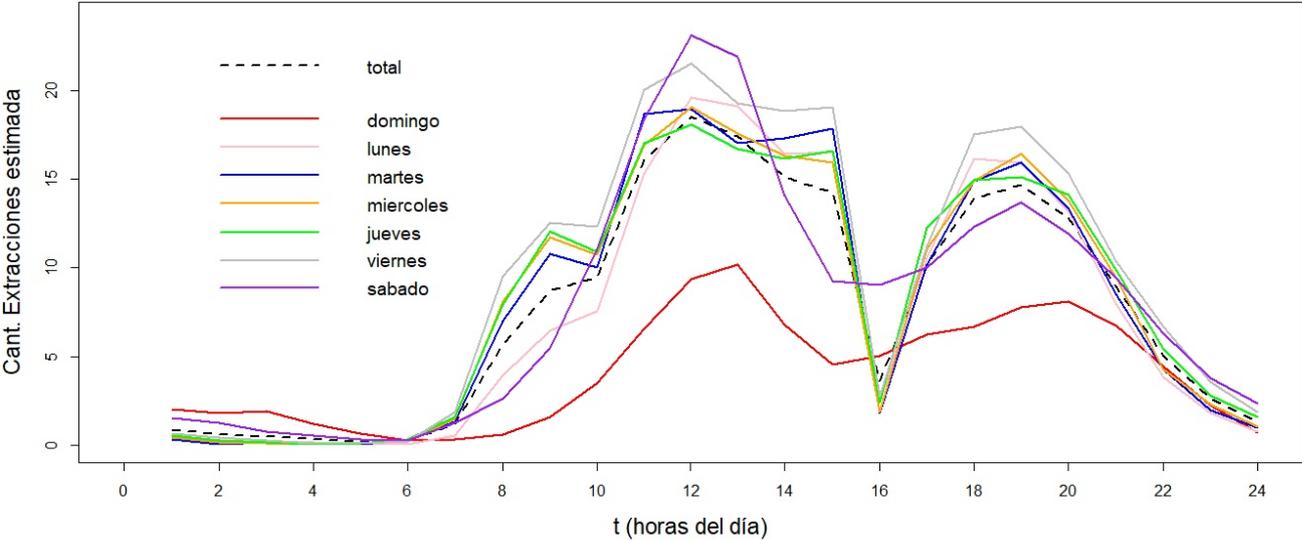


Figura 4.3: Estimación de la Función de Intensidad del cajero b309 para cada día de la semana

Según se visualiza en el gráfico anterior, las estimaciones de los días sábados y domingos son las más extremas, el domingo es cuando hay menor actividad, mientras que el sábado presenta un pico de actividad hacia el mediodía, cayendo por la tarde. En ambos casos, no ocurre el pico a la baja alrededor de las 16 horas que se da entre semana debido al balance de operaciones de la red. Sin embargo sí se observa una caída moderada.

En contraposición, las estimaciones para el resto de los días de la semana se estarían comportando de manera similar. Esto sugiere que podría tratarse de una misma función de intensidad y no sería necesario estimar una para cada día. Pero aún es muy pronto para tomar decisiones al respecto.

Esto también se puede ver de forma resumida a través de diagramas de caja para cada una de las estimaciones de la función de intensidad.

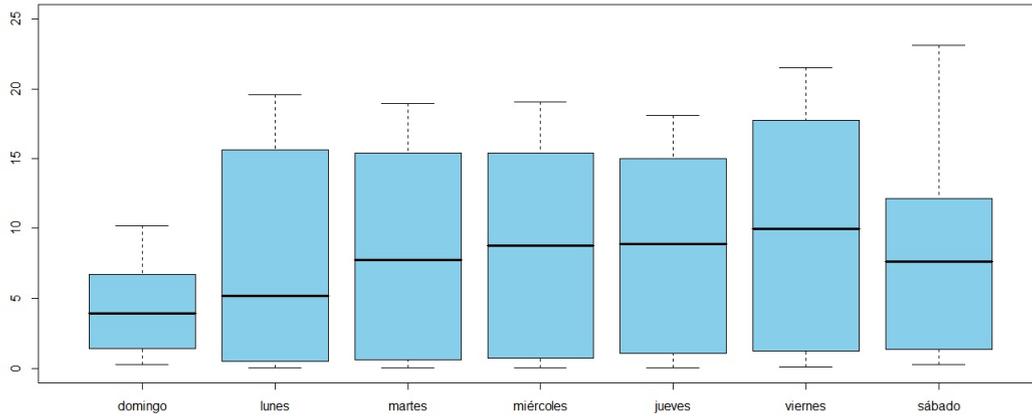


Figura 4.4: Diagramas de caja de la función de intensidad estimada por día de la semana para el cajero b309

Como se puede apreciar en la Figura anterior, el domingo es el día que presenta una menor mediana y a su vez un rango intercuartílico más compacto, de hecho toda la distribución es más compacta puesto que el máximo para el domingo se ubica cercano a las 10 extracciones por hora. Por otro lado, el sábado tiene la distribución más dispersa ya que posee un máximo superior a 20 retiradas por hora. De lunes a viernes el rango intercuartílico tiene una amplitud similar, sin embargo la mediana va en ascenso a medida que avanza la semana. Esto se verifica al observar las medidas de resumen presentadas en el cuadro 4.1.

Para ver estas cifras exactas se presenta a continuación una tabla con las medidas de posición, de tendencia central y dispersión.

Cuadro 4.1: Medidas de posición, tendencia central y dispersión de la función de intensidad estimada para el cajero b309

		Estimación						
		Domingo	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
Medidas de Posición	Mínimo	0,270	0,052	0,043	0,052	0,035	0,104	0,287
	Primer Cuartil	1,489	0,537	0,804	0,904	1,315	1,550	1,441
	Mediana	3,952	5,174	7,739	8,761	8,891	9,939	7,643
	Tercer Cuartil	16,69	15,48	15,18	15,19	15,00	17,66	12,05
	Máximo	10,19	19,61	18,95	19,09	18,09	21,50	23,12
Medida de Tendencia Central	Media	4,134	7,483	7,997	8,097	8,216	9,342	7,945
Medidas de Dispersión	Desvío Estándar	3,119	7,198	7,256	7,082	6,886	7,898	6,921
	Varianza	9,733	51,81	52,65	50,16	47,42	62,38	47,90

4.2. Estimación de la función de intensidad acumulada

Como se vio en el capítulo anterior, la estimación de la función de intensidad acumulada (FIA) asociada al PPNH tiene la siguiente forma:

$$\widehat{\Lambda}(t) = \left(\sum_{j=1}^{i-1} \frac{N_j}{n} \right) + \frac{N_i(t - a_{i-1})}{n(a_i - a_{i-1})}$$

con $a_{i-1} < t \leq a_i$, para $i = 1, 2, \dots, k$.

Siendo N_i la cantidad de eventos observados en el i -ésimo intervalo, n la cantidad de realizaciones del proceso y k la cantidad de subintervalos.

Para el cálculo de esta estimación se trabajó con los datos originales, es decir, con la cantidad de extracciones medidas cada diez minutos, por lo tanto se tienen 144 intervalos para cada día y como cada día se considera una realización del proceso, se tienen 805 realizaciones, pues se consideran los 7 días a lo largo de las 115 semanas. Por lo tanto la FIA se estima de la siguiente manera:

$$\widehat{\Lambda}(t) = \left(\sum_{j=1}^{i-1} \frac{N_j}{805} \right) + \frac{N_i(t - a_{i-1})}{805(a_i - a_{i-1})}$$

con $a_{i-1} < t \leq a_i$, para $i = 1, 2, \dots, 144$.

A continuación se presenta la estimación de la FIA sin discriminar por día de la semana con intervalos de 90 % y 99 % de confianza.

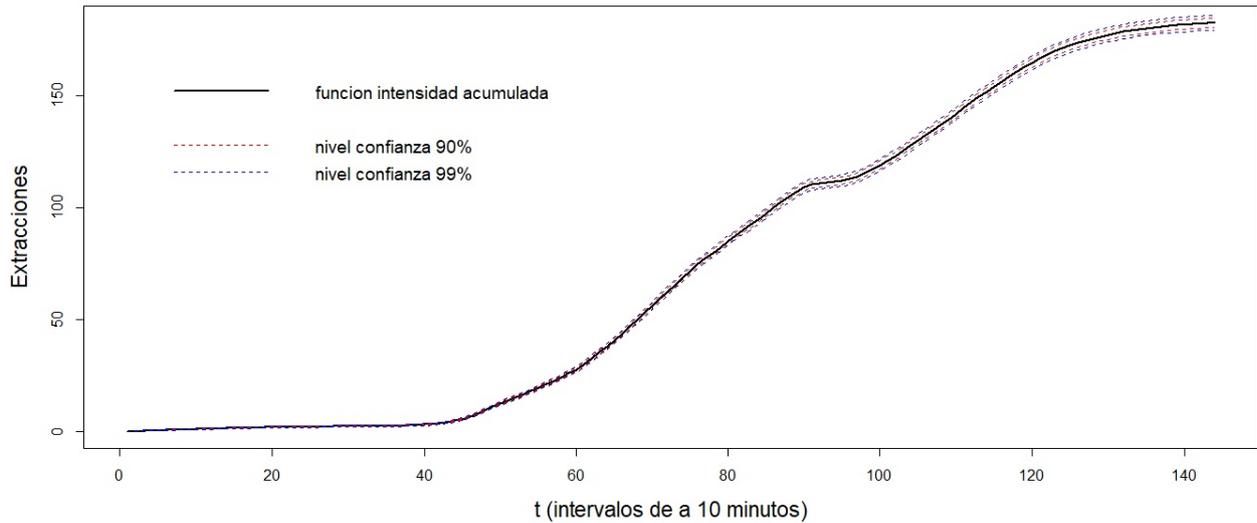


Figura 4.5: Estimación de la FIA del cajero b309

El recorrido de la FIA estimada para este cajero, comienza sin un crecimiento perceptible, son las primeras horas de la madrugada, en donde raramente ocurren extracciones. A partir del subintervalo 45 aproximadamente se visualiza un crecimiento sostenido, que corresponde a minutos previos a las 8 de la mañana. A partir del subintervalo 60, es decir las 10 de la mañana, se observa un incremento en la pendiente de la gráfica. Entre los subintervalos 90 y 100 aproximadamente, se puede ver una cuasimeseta que corresponde con el período cercano a las 16 hs. al que ya se ha hecho referencia.

Para estimar la FIA por día de la semana, se consideran las 115 realizaciones del proceso para cada día, y por lo tanto la fórmula para dicha estimación es la siguiente, tomando como ejemplo los días lunes.

$$\hat{\Lambda}_{lunes}(t) = \left(\sum_{j=1}^{i-1} \frac{N_j}{115} \right) + \frac{N_i(t - a_{i-1})}{115(a_i - a_{i-1})}$$

siendo $a_{i-1} < t \leq a_i$, para $i = 1, 2, \dots, 144$.

Análogamente se calcula la estimación para los siete días de la semana. En el siguiente gráfico se representan todas las FIAs estimadas.

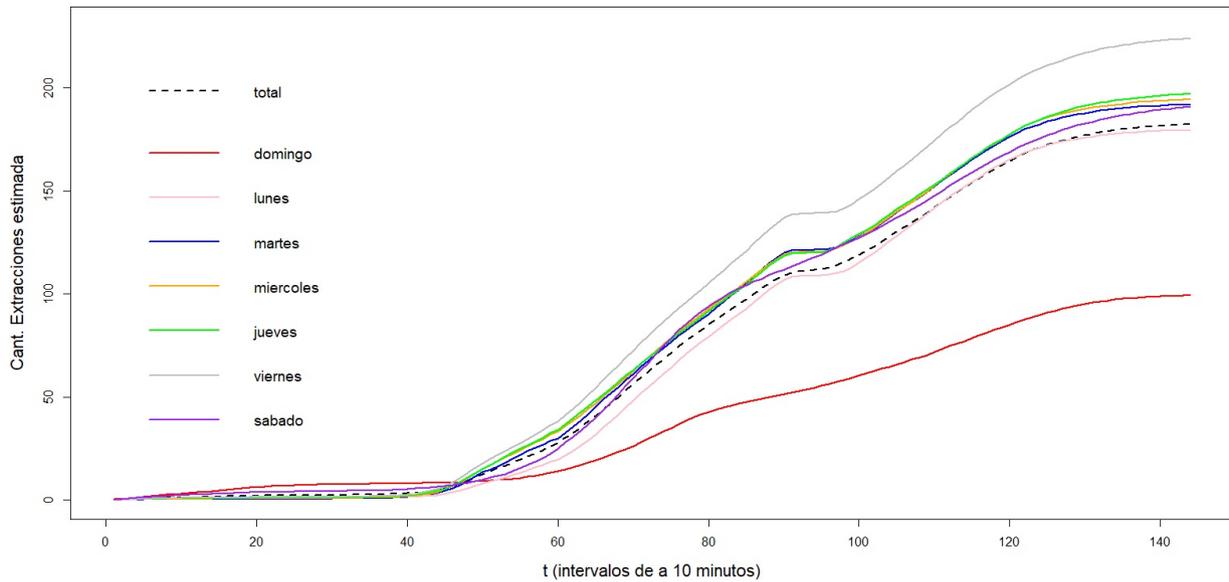


Figura 4.6: Estimación de la FIA estimada del cajero b309 por día

Visualizando cada día en forma separada, se puede ver que los días viernes y domingo son los que presentan funciones de intensidad estimadas que se diferencian notoriamente del resto. El viernes es el día en donde más extracciones se realizan mientras que el domingo es el día con menor actividad.

De lunes a jueves y el sábado, al finalizar el día se alcanza un total de extracciones de entre 150 y 200, mientras que el viernes se superan las 200 y el domingo no alcanza las 100 extracciones. También se puede observar que el sábado y el domingo no presentan la meseta previa al subintervalo número 100 (alrededor de las 16 hs) como el resto de los días.

En cuanto al comportamiento general, se puede distinguir que de lunes a sábado la forma del gráfico es similar, mientras que el domingo presenta una actividad superior durante la madrugada y una pendiente prácticamente constante entre las 10 y las 23 hs.

A partir de las observaciones de los gráficos presentados en la Figura 4.6 se realizaron varias comparaciones entre los días similares con el fin de agrupar sus estimaciones de las FIA.

Finalmente se constató que el día miércoles y jueves presentan una FIA estimada tan similar que puede considerarse que corresponden al mismo PPNH y por lo tanto presentar una misma FIA. Esto se puede visualizar en el gráfico de la Figura 4.7 donde se representa

a la FIA estimada del día miércoles con su intervalo de confianza al 99% y a la FIA estimada del día jueves, la cual se encuentra durante todo su recorrido dentro del intervalo correspondiente al día miércoles.

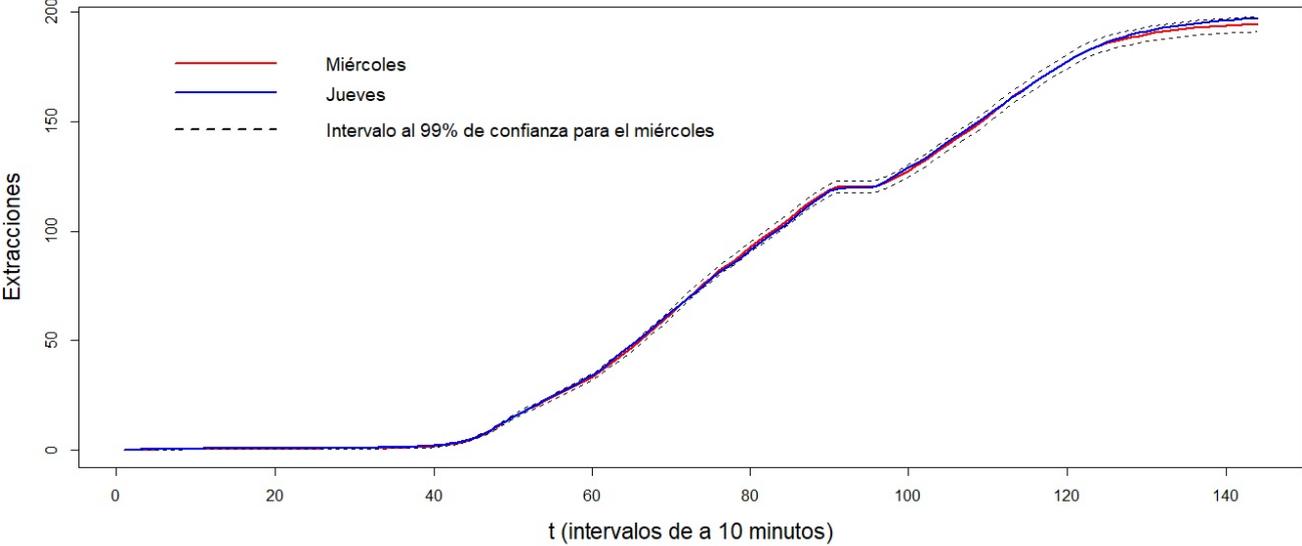


Figura 4.7: Estimación de la FIA del cajero b309 para miércoles y jueves

Como puede observarse en el gráfico anterior, la FIA estimada para el día jueves cae dentro del intervalo de confianza de la FIA estimada para el miércoles, por lo tanto se decide unificar las observaciones de estos dos días y estimar una sola FIA común para el miércoles y el jueves. A continuación se presenta un gráfico que incluye dicha estimación y se la compara con la estimación de la FIA para el domingo, cada una con su intervalo de confianza correspondiente.

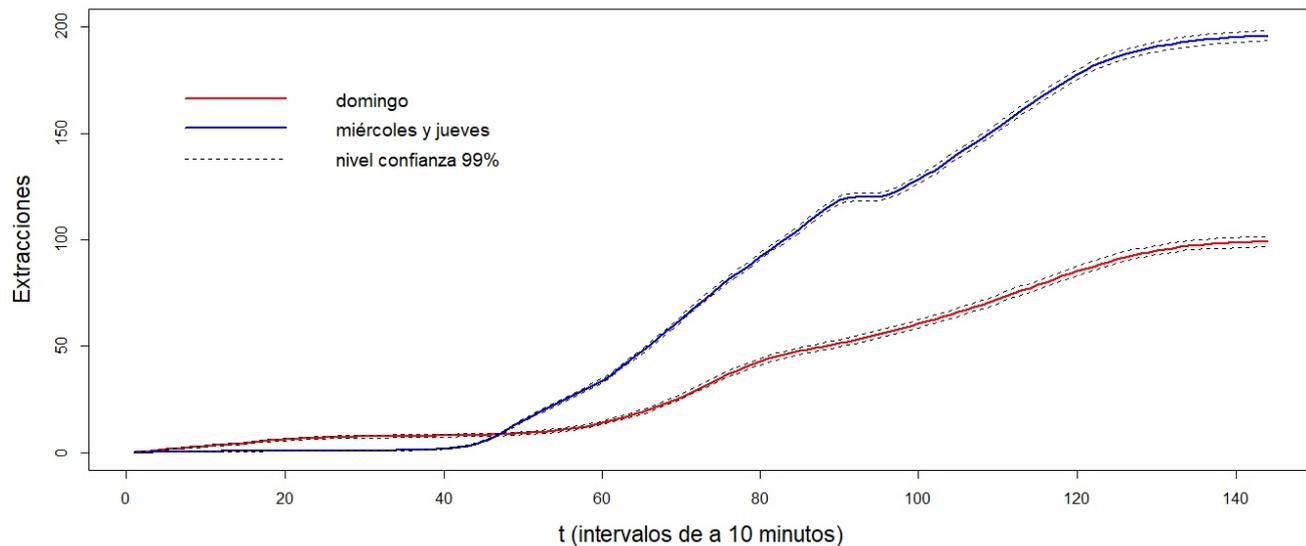


Figura 4.8: Estimación de la FIA del cajero b309 para miércoles y jueves vs domingo

Claramente se está ante la presencia de dos PPNH muy diferentes. Para empezar, el nivel de extracciones alcanzado en la estimación conjunta para miércoles y jueves es de unos 200 eventos, mientras que el domingo llega solo a la mitad.

A su vez, mientras que la FIA estimada del miércoles y jueves no comienza a crecer hasta el subintervalo 45 (aproximadamente las 7:30), el domingo presenta ocurrencia de eventos desde el comienzo del día. También se puede observar que las pendientes que presentan ambos gráficos a lo largo del intervalo de estudio son diferentes, a excepción de los últimos subintervalos.

El miércoles y jueves, en el entorno de las 7:30 hs la actividad es intensa hasta el impasse de la tarde, luego se retoma la ocurrencia de eventos con una intensidad levemente menor, aunque casi imperceptible. Pasadas las 20 hs la intensidad comienza a disminuir ya que se puede ver una pendiente menor en la FIA estimada y se llega a una cuasi meseta al finalizar el día.

Por su parte, recién pasadas las 10 hs del domingo, comienzan a registrarse una ocurrencia significativa de eventos aunque con menor intensidad que en la conjunta para miércoles y jueves, ya que la pendiente es visiblemente menor. Este día no presenta ningún impasse durante la tarde, lo cual es de esperar porque el domingo no se realiza el corte de la red para el balance de operaciones, y durante las últimas dos horas del día casi no se registran extracciones, al igual que los días miércoles y jueves.

4.2.1. Estimación de la función de intensidad acumulada por quincena

Resulta natural pensar que durante la primera parte de cada mes puede haber más cantidad de extracciones ya que la mayoría de las personas reciben la acreditación de su salario mensual en el entorno del comienzo del mes. Esto motiva realizar un análisis comparativo entre la estimación de la FIA para miércoles y jueves vista en la sección anterior y algunos miércoles y jueves del mes de setiembre del año 2010 a modo de ejemplo para ver si existen tales diferencias.

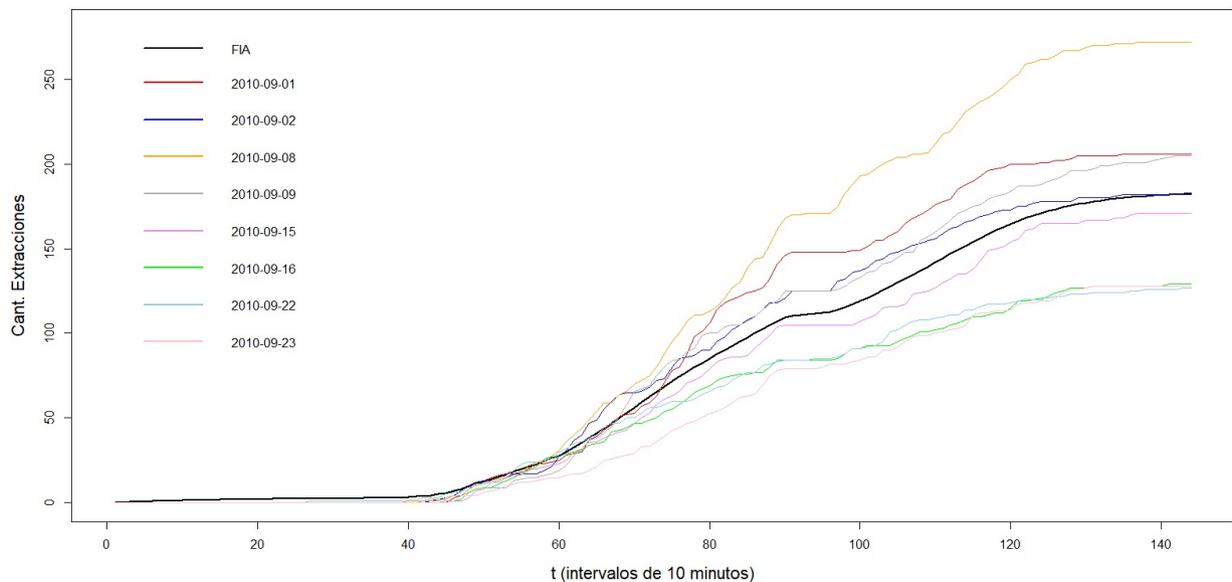


Figura 4.9: Estimación de la FIA para miércoles y jueves vs estimación de la FIA para algunos días en particular del cajero b309

Como se puede apreciar, hay diferencias entre los días observados y la estimación de la FIA calculada a partir de todos los miércoles y jueves. El recorrido de los días 01/09/2010, 02/09/2010, 08/09/2010 y 09/09/2010 se encuentra por encima de la FIA estimada, mientras que los días 15/09/2010, 16/09/2010, 22/09/2010 y 23/09/2010 se encuentran por debajo.

Las diferencias más notorias son para el 08/09/2010 que acumula más de 250 extracciones a lo largo del día y para los días 16/09/2010, 22/09/2010 y 23/09/2010 que no llegan a acumular 150 extracciones. Por su parte los días 01/09/2010, 02/09/2010, 09/09/2010 y 15/09/2010 presentan recorridos más próximos a la FIA estimada genérica.

A partir de este análisis visual, se procede a estimar una FIA para la primera quincena del mes y otra para la segunda quincena, lo cual se puede visualizar en la figura 4.10. Si bien la comparación que se hizo fue para un mes en particular, se aplica para todo el período con el que se está trabajando.

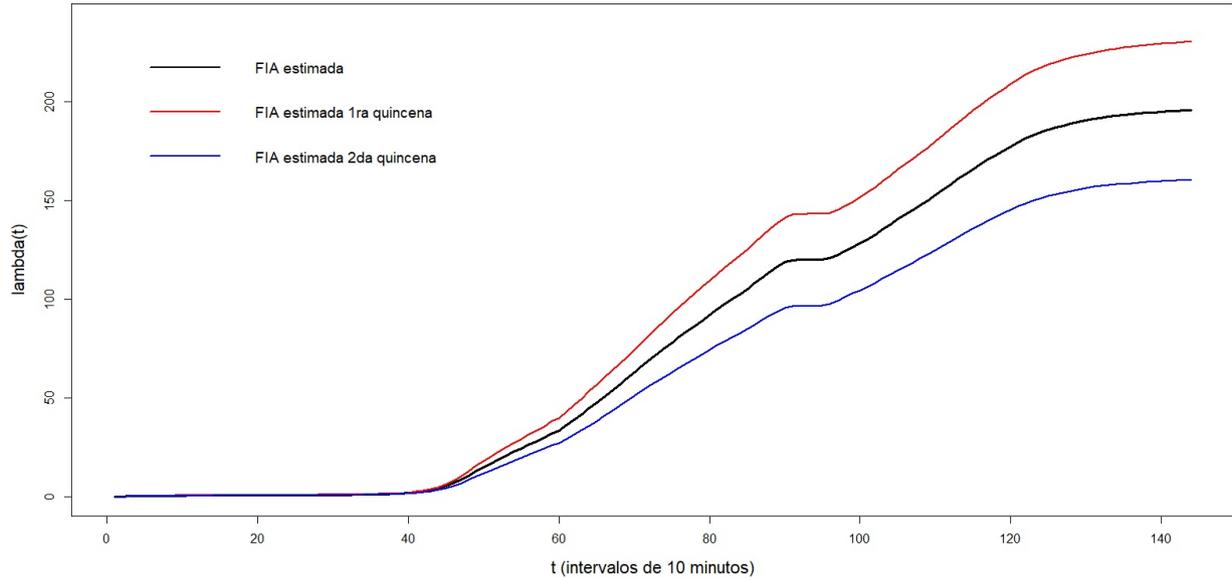


Figura 4.10: Estimación de la FIA para miércoles y jueves vs estimación de la FIA por quincena del cajero b309

Se puede distinguir claramente que la estimación para la primera quincena supera las 200 extracciones mientras que la estimación para la segunda quincena apenas supera las 150. Visualmente es notorio que la estimación de la FIA general se encuentra en el medio de las estimaciones por quincena, lo cual respalda la decisión de dividir la estimación por quincena, la cual si bien surge a partir de un supuesto y del análisis visual realizado a partir del gráfico 4.9, no deja de ser una decisión arbitraria.

4.3. Inactividad del cajero

Según se vio en el capítulo anterior, la distribución del número de eventos en el intervalo $[t, t+h)$ es la de una variable de Poisson con parámetro $\Lambda(t+h) - \Lambda(t)$, para todo $t, h \geq 0$.

$$N(t+h) - N(t) \sim \text{Poisson}(\Lambda(t+h) - \Lambda(t)) \quad \forall t, h \geq 0.$$

Por lo tanto la probabilidad de que no ocurran eventos en el intervalo $[t, t + h)$ está dada por:

$$P(N(t + h) - N(t) = 0) = e^{-(\Lambda(t+h) - \Lambda(t))}$$

Al contar con la estimación de la función de intensidad acumulada, es posible obtener la estimación de la probabilidad de que no ocurran eventos en el intervalo $[t, t + h)$

$$\widehat{P}(N(t + h) - N(t) = 0) = e^{-(\widehat{\Lambda}(t+h) - \widehat{\Lambda}(t))}. \quad (4.2)$$

Por ejemplo, la probabilidad estimada de que haya cero extracciones entre las 21:00 hs y las 22:00 hs es:

$$\widehat{P}(N(22 : 00hs) - N(21 : 00hs) = 0) = e^{-(\widehat{\Lambda}(22:00 \text{ hs}) - \widehat{\Lambda}(21:00 \text{ hs}))} = e^{\widehat{\Lambda}(21:00 \text{ hs}) - \widehat{\Lambda}(22:00 \text{ hs})}.$$

Según el objetivo planteado inicialmente, la expresión 4.2 permite estimar la probabilidad de inactividad en el cajero durante un período de tiempo a elección para contrastarlo con la realidad y así poder tomar decisiones.

A modo de ejemplo, en el anexo D se presenta un cuadro con las estimaciones de la probabilidad de que no haya extracciones durante 60 minutos para cada hora del día.

Los resultados presentados en esa tabla son interesantes pero sobre todo lo es la expresión 4.2, que permite una flexibilidad mayor a las probabilidades expuestas en la tabla y tiene un valor estadístico que podría servir para contrastar con la realidad y así poder tomar decisiones a nivel de la red de cajeros.

Para una visualización de estos resultados, se presenta a continuación una figura conteniendo los gráficos de las estimaciones de la probabilidad de inactividad por hora y por media hora para el límite superior de cada intervalo de diez minutos.

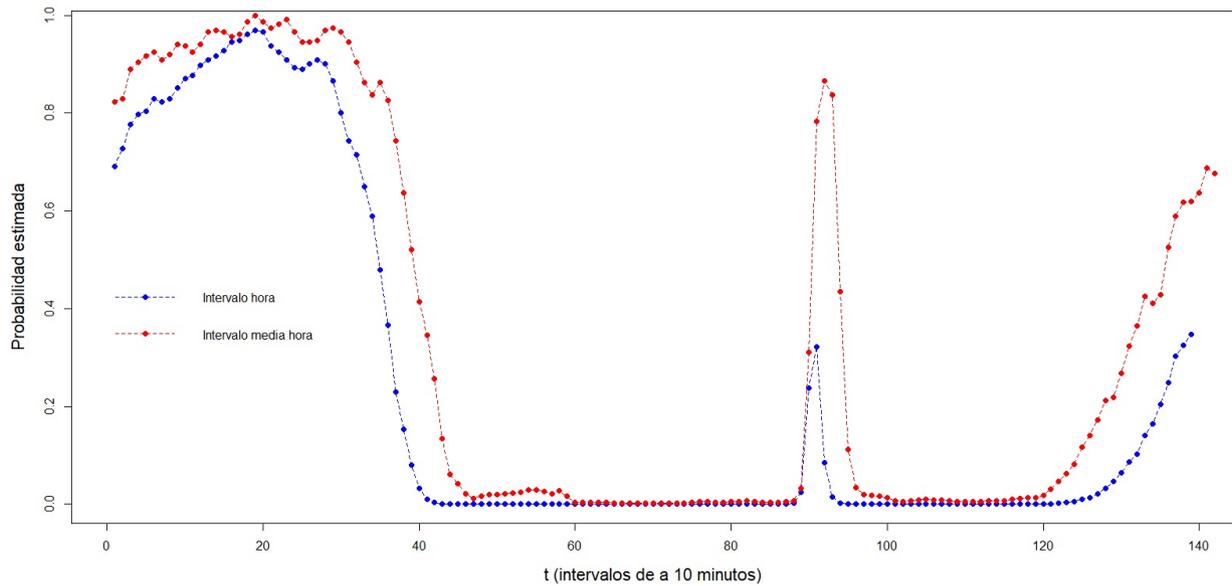


Figura 4.11: Estimación de la probabilidad de inactividad en la siguiente hora y siguiente media hora por intervalo de a diez minutos

Como se puede ver en la figura anterior, el gráfico rojo, que representa la probabilidad estimada de inactividad para los siguientes treinta minutos, siempre está a igual nivel o superior que el gráfico azul, que representa la probabilidad estimada de inactividad para la siguiente hora. Esto ocurre porque en tramos más largos es menos probable que haya inactividad.

Las horas en las que se estima que es más probable la ausencia de realizaciones del proceso, es decir, las horas en las que es más probable que el cajero esté inactivo, es durante la madrugada y en el momento de la tarde en donde no se les permite realizar extracciones a los usuarios, éste es el pico que se observa en la figura 4.11 luego del intervalo 90.

A efectos prácticos, si la empresa propietaria de la red de cajeros realiza un monitoreo permanente, se puede contrastar la probabilidad de inactividad en la última media hora contra la actividad real del cajero, y como la estimación es móvil cada diez minutos, este contraste se puede realizar casi en tiempo real.

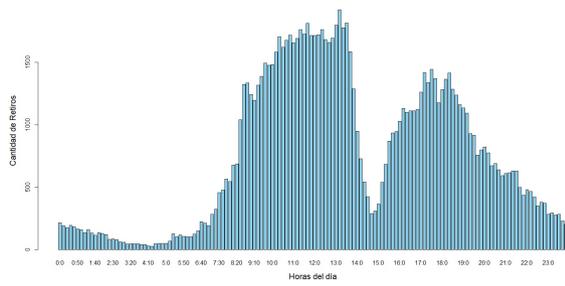
4.4. Análisis comparativo de diferentes cajeros

El trabajo realizado para el cajero b309 presentado en las secciones [1.4](#), [4.1](#), [4.2](#), y [4.3](#) se realizó para el resto de los cajeros. En esta sección se presentan los resultados más relevantes para seis de los cajeros a modo de ejemplo. A saber, los cajeros b397, b435, b834, m389, m404 y m437.

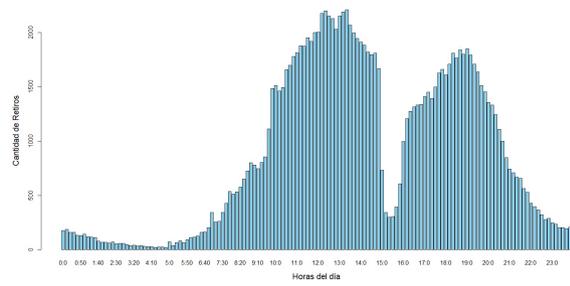
El objetivo es realizar un breve análisis comparativo entre los diferentes PPNH correspondientes a seis de los 19 cajeros restantes, ver que algunas características son comunes a varios o todos los cajeros mientras algunas otras son particulares.

Es posible realizar estas comparaciones de forma bastante ágil ya que la metodología desarrollada para la estimación de la función de intensidad y función de intensidad acumulada del PPNH correspondiente al cajero b309 se puede replicar para el resto de los cajeros, esa es la gran ventaja de haber implementado técnicas de estadística no paramétrica.

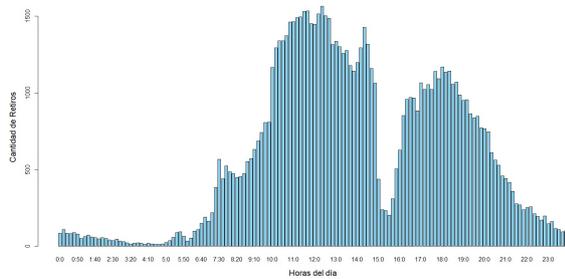
A continuación, en la figura [4.12](#) se presentan los diagramas de barra con el total de extracciones en tramos de diez minutos para los seis cajeros de muestra.



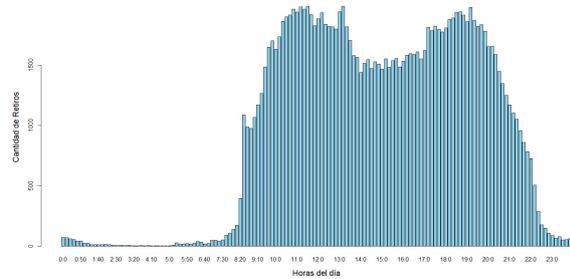
(a) Cajero b397



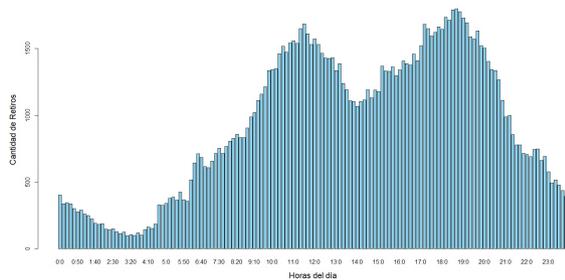
(b) Cajero b435



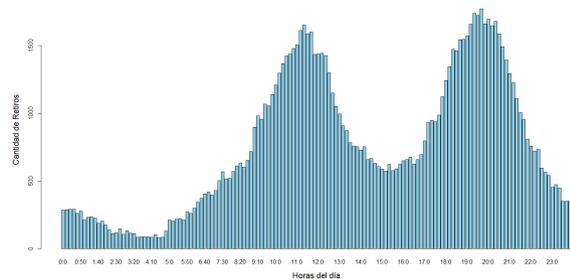
(c) Cajero b834



(d) Cajero m389



(e) Cajero m404



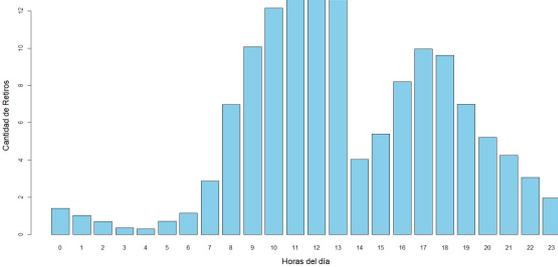
(f) Cajero m437

Figura 4.12: Total de extracciones cada diez minutos para los cajeros de muestra

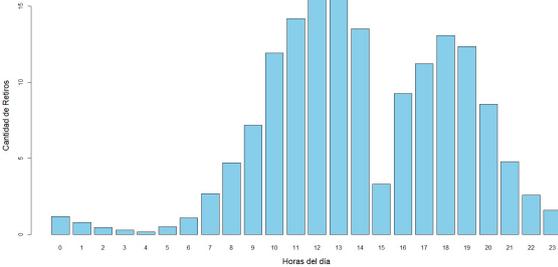
Como se puede apreciar en la figura 4.12, los gráficos correspondientes a los diferentes cajeros presentan formas diversas, es decir que en cada cajero el comportamiento es diferente. Por ejemplo, los cajeros m389 y m404 no tienen la forma bimodal tan acentuada como el resto.

El cajero m389 no presenta extracciones en varios tramos de la madrugada, mientras que los cajeros m404 y m437 son los que más actividad presentan durante las primeras horas del día en comparación con el resto.

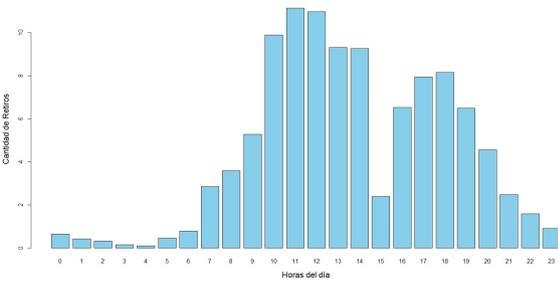
En la figura 4.13 se presentan diagramas de barras para cada cajero con el promedio de retiros agrupados por hora. La forma de cada uno de estos gráficos es similar a la de la figura 4.12, pero al ser el promedio de las extracciones en tramos de una hora, dicha forma es más suave.



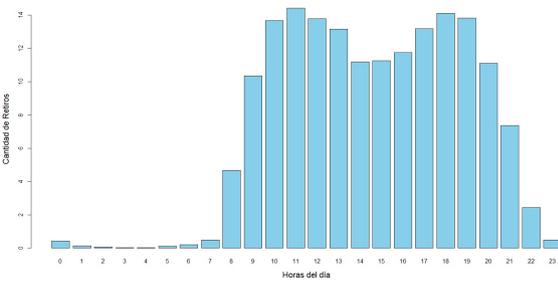
(a) Cajero b397



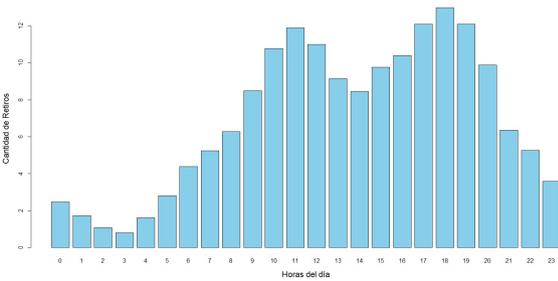
(b) Cajero b435



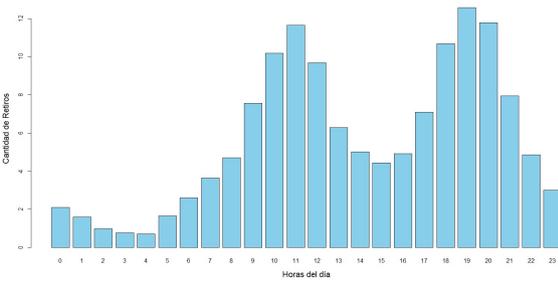
(c) Cajero b834



(d) Cajero m389



(e) Cajero m404



(f) Cajero m437

Figura 4.13: Retiros promedio por hora para los cajeros de muestra

A continuación, en la figura 4.14 se presentan los gráficos con la estimación de la función de intensidad por día de la semana para cada uno de los cajeros.

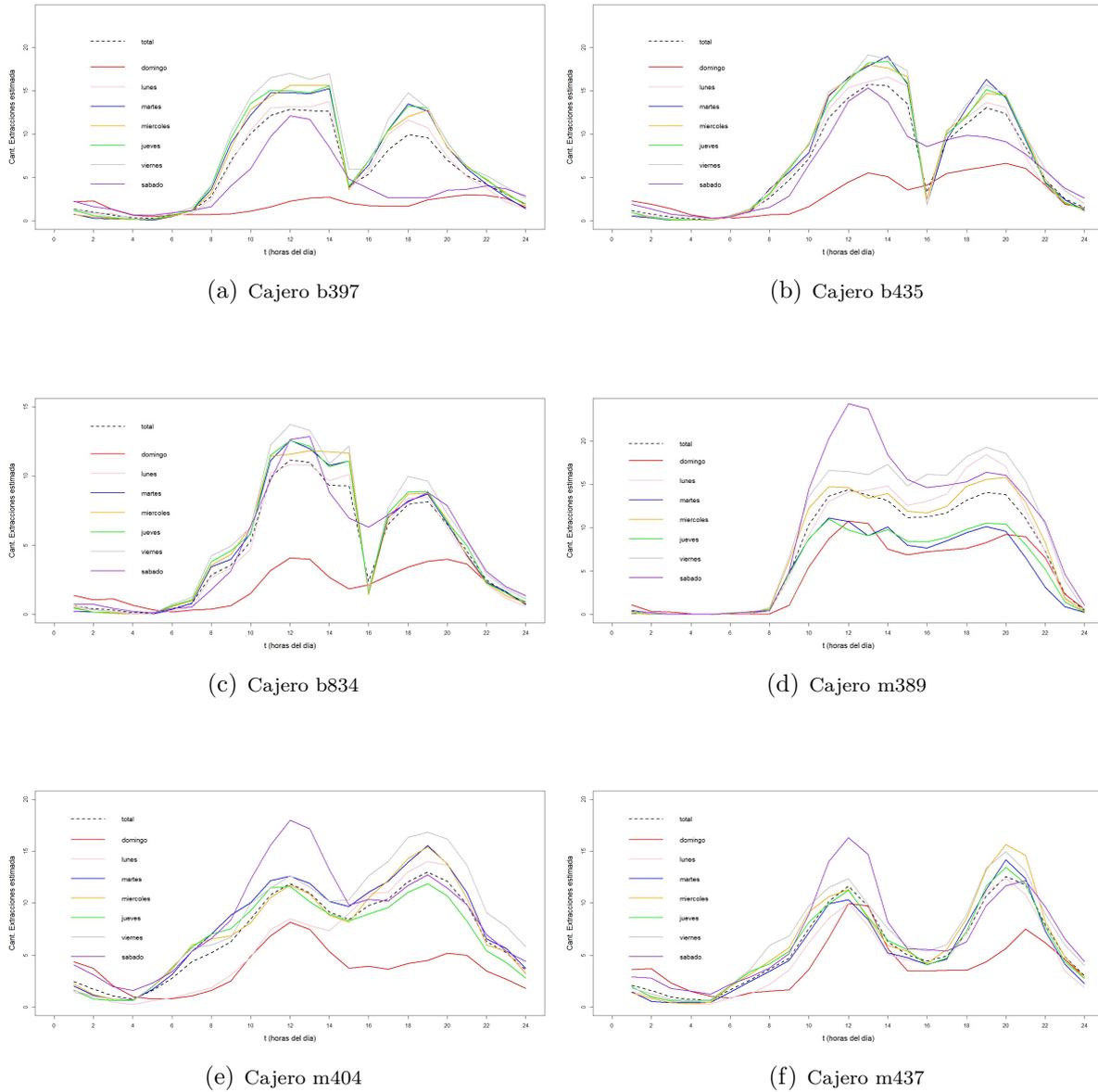


Figura 4.14: Estimación de la función de intensidad para los cajeros de muestra

En todos los cajeros se distingue que el domingo presenta menor actividad, sin embargo los cajeros m389 y m437 son los que menos diferencia tienen entre el domingo y el resto de los días de la semana. Por otro lado, el cajero m389 es el que presenta menor bimodalidad en la forma de las estimaciones de sus funciones de intensidad para los diferentes días de la semana, mientras que en el resto de los cajeros se ve claramente la diferencia entre las estimaciones para los días de la semana y las estimaciones del sábado y del domingo.

A continuación se presenta un gráfico comparativo de la estimación de la función de intensidad del cajero b309 y de los seis cajeros analizados en la presente sección.

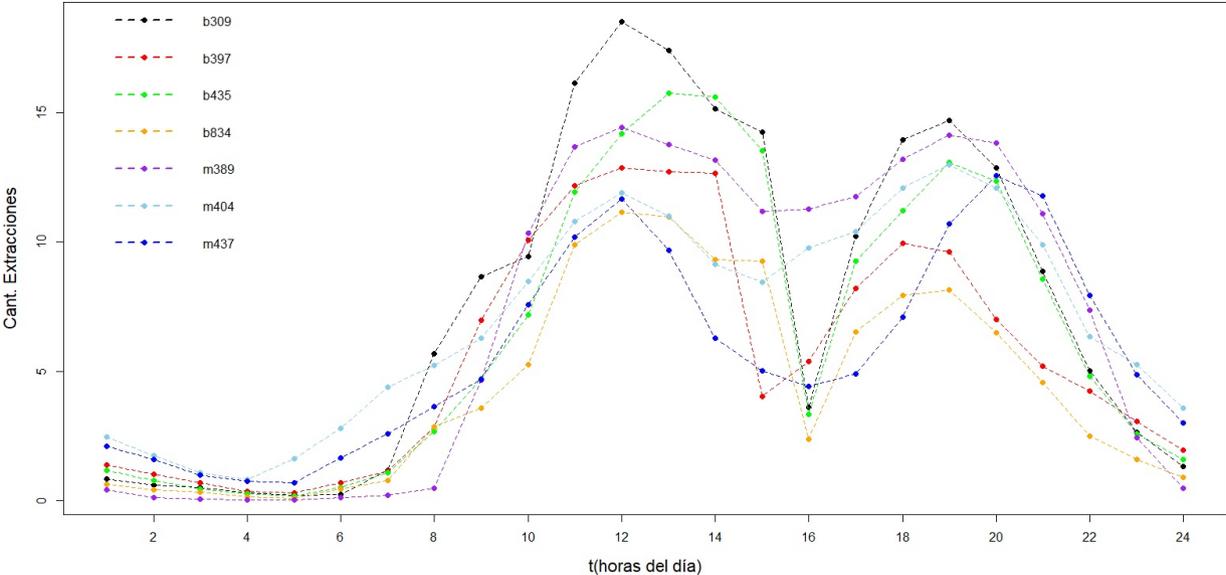
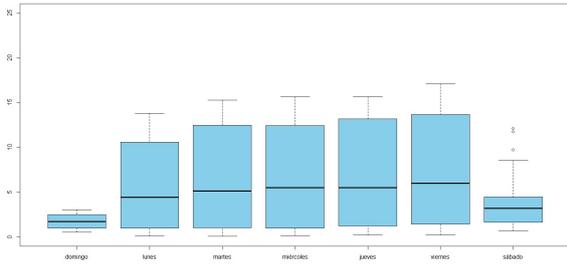


Figura 4.15: Estimación de la función de intensidad de cada cajero

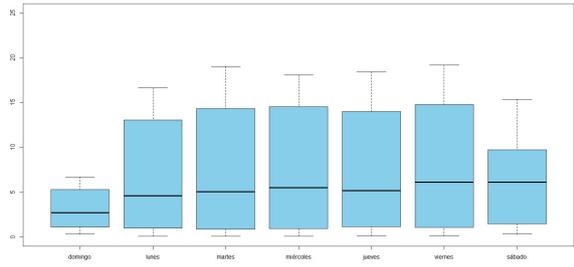
Si bien en términos generales se ven algunas similitudes, se puede apreciar también que en algunos casos la bimodalidad es más marcada que en otros casos. Esto probablemente se deba a que en ciertos cajeros la cantidad de extracciones durante los fines de semana es de tal magnitud que incide en esta bimodalidad, ya que como se mencionó anteriormente, el balance de operaciones que realiza la red de cajeros a las 16 hs no ocurre sábados y domingos.

Puede verse también que el cajero b309 es el que presenta una función de intensidad estimada de mayor magnitud y con un máximo mayor que en el resto de los casos ubicado en el entorno del mediodía.

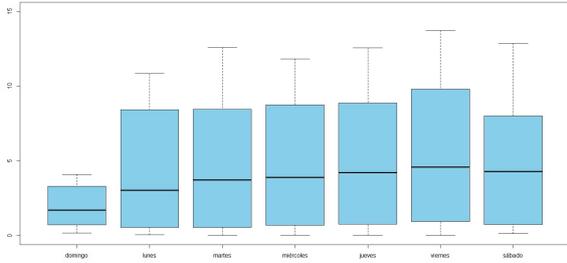
En la figura 4.16 que se presenta a continuación, se pueden visualizar los diagramas de caja de la función de intensidad estimada por día para cada uno de los seis cajeros de muestra.



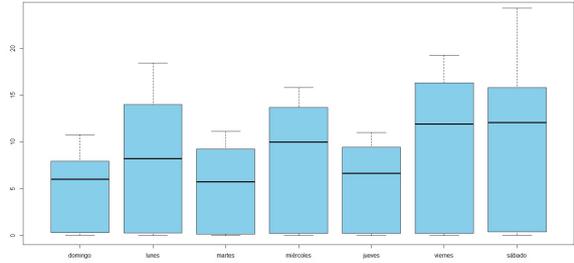
(a) Cajero b397



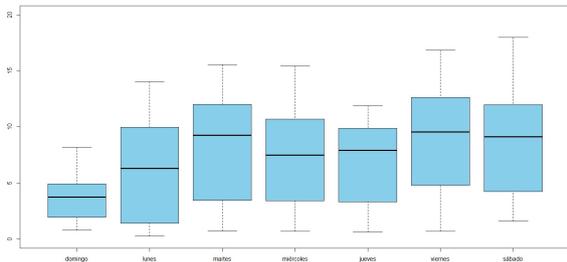
(b) Cajero b435



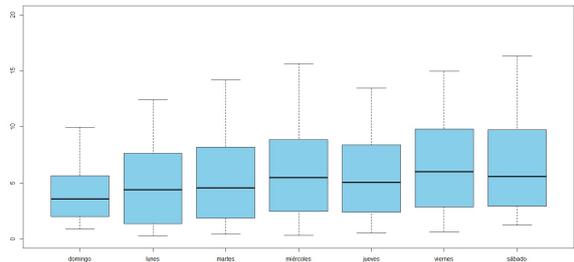
(c) Cajero b834



(d) Cajero m389



(e) Cajero m404



(f) Cajero m437

Figura 4.16: Diagramas de caja para los cajeros de muestra por día de la semana

Cada cajero presenta particularidades en la distribución de las diferentes FI estimadas. Por ejemplo, el b397 es el que tiene sábado y domingo más compactos. Mientras que el m389 es el que posee mayor actividad el día sábado y a su vez la distribución del día sábado para este cajero es el que posee una distribución más amplia entre todas las estimaciones de las FI para estos cajeros. A su vez, el m389 los días sábado y domingo son los que tienen mayores medianas, es decir, mayor actividad.

En la figura 4.17 a continuación, se muestran los gráficos correspondientes a cada cajero con la estimación de la FIA para cada día de la semana.

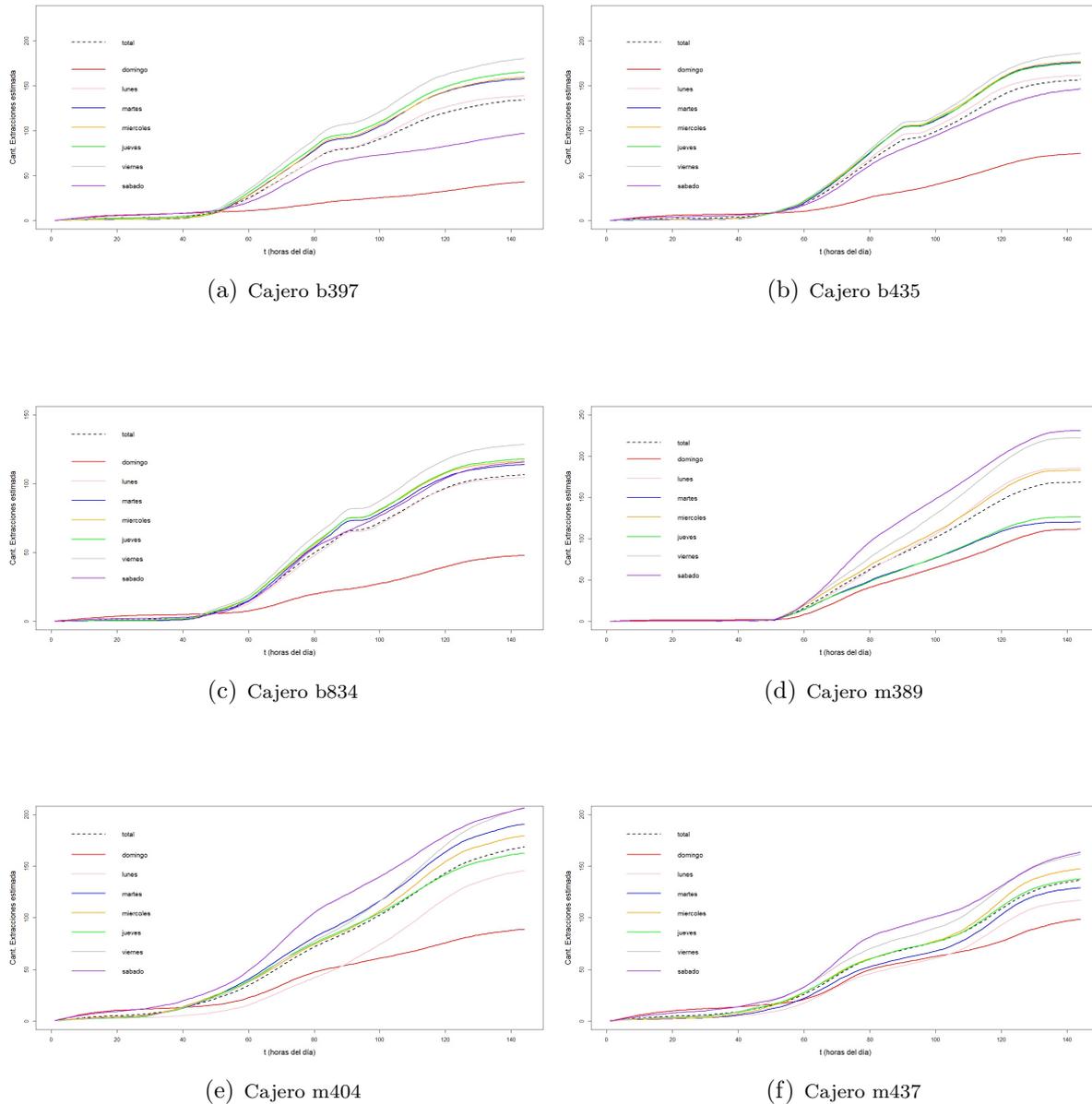


Figura 4.17: Estimación de la función de intensidad acumulada por día de la semana para los cajeros de muestra

Se observa a simple vista que los cajeros b397 y m389 presentan estimaciones de la FIA más disímiles para los diferentes días. En contraposición, los cajeros b435 y b834 tienen estimaciones de la FIA muy similares de lunes a viernes.

Para poder visualizar las FIAs estimadas de modo comparativo, se presenta a continuación una figura con las las gráficas para los cajeros elegidos en simultáneo.

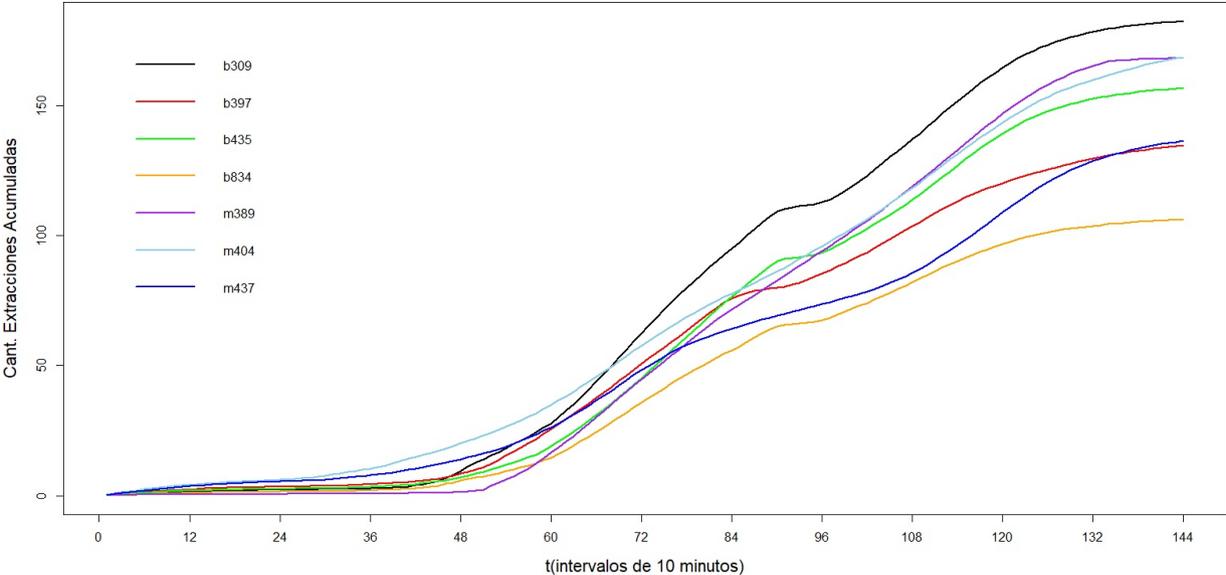


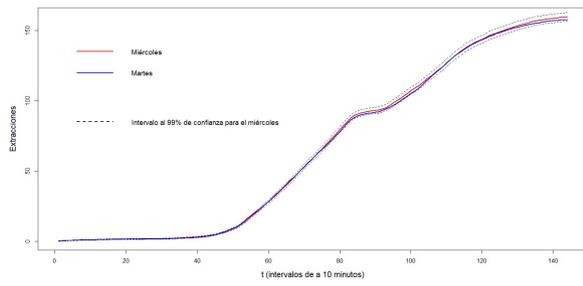
Figura 4.18: Estimación de la función de intensidad de cada cajero

Al comparar las FIA de cada cajero, se puede ver que algunos presentan formas similares, como el b309, el b435, el b397 y el b834, aunque difieren en los niveles de extracciones. El b309 es el que tiene mayor cantidad de extracciones a lo largo del día, llegando a acumular más de 170 extracciones.

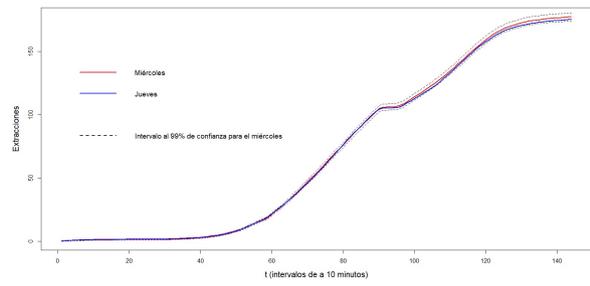
Por otro lado, los cajeros m389, m404 y m437 tienen formas disímiles, aunque al final del día todos acumulan entre 130 y 160 extracciones.

En la figura 4.19 se muestra la comparación entre las estimaciones de la FIA para diferentes días de la semana. Para cada cajero se analizó la posibilidad de agrupar dos días de la semana en una única estimación de la FIA como se hizo para el cajero b309 en la sección 4.2. Como se puede apreciar a continuación, en algunos casos la FIA estimada para ciertos días queda contenida en el intervalo de confianza de otro día, mientras que en otros casos no.

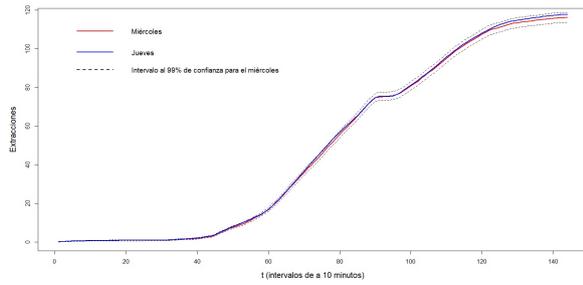
En el caso del cajero b397 se decide estimar la FIA para martes y miércoles de manera conjunta, en el caso de los cajeros b435 y b834 se estima para miércoles y jueves, finalmente para los cajeros m389, m404 y m437 al no presentar ninguna estimación de la FIA de un día que quede contenida en el intervalo de confianza de la estimación de otro día, se decide



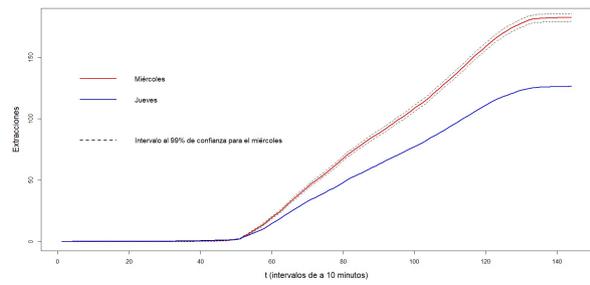
(a) Cajero b397 - Martes y Miércoles



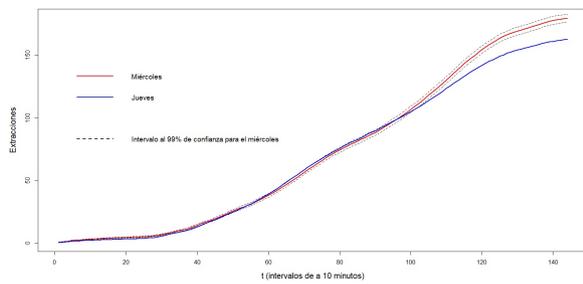
(b) Cajero b435 - Miércoles y Jueves



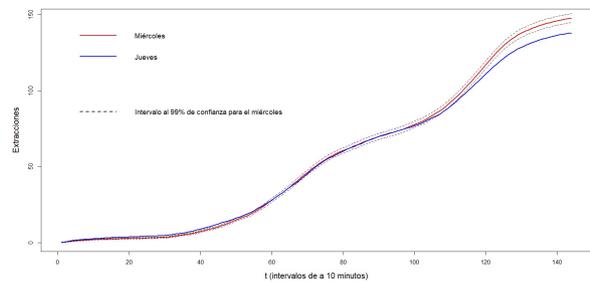
(c) Cajero b834 - Miércoles y Jueves



(d) Cajero m389 - Miércoles y Jueves



(e) Cajero m404 - Miércoles y Jueves



(f) Cajero m437 - Miércoles y Jueves

Figura 4.19: Estimación de la función de intensidad acumulada para los cajeros de muestras en ciertos días

mantener la estimación para el día miércoles únicamente.

En la figura D.16 del anexo D se muestra para cada uno de los seis cajeros comparaciones de la estimación de la FIA para diferentes días, según haya sido adecuado un agrupamiento o no.

Con la misma metodología que se vio en la sección 4.3 donde se calculó la probabilidad de inactividad por hora para el cajero b309, se puede calcular para los cajeros analizados en la presente sección. En la figura D.2 del anexo D se presenta una tabla con la probabilidad de inactividad por hora para cada cajero. Según los agrupamientos mencionados en la página anterior es que se calculan las diferentes probabilidades. En el caso del b397 es para martes y miércoles, en el caso del b435 y del b834 se calcula para el miércoles y jueves, y en el caso de los m389, m404 y m437 es para el miércoles.

Capítulo 5

Conclusiones

El objetivo general de este trabajo ha sido caracterizar el proceso que modela el fenómeno de las extracciones de los cajeros a lo largo del día, a partir de datos disponibles en el período 01/08/2010 - 15/10/2012, para un conjunto de 20 cajeros. El objetivo específico ha sido hallar una metodología (ver sección 3.5) que permitiera detectar los períodos de inactividad naturales de los cajeros.

Se logró desarrollar una metodología flexible para describir el comportamiento de los diferentes cajeros en diferentes días de la semana y en diferentes momentos del mes gracias a la utilización de técnicas de estadística no paramétrica.

En función de los objetivos y según los resultados presentados en el capítulo anterior, se presentan algunos comentarios para concluir el informe.

5.1. Caracterización de los retiros en la red de cajeros

En este trabajo se concibió al fenómeno de retiro de dinero del cajero automático que se produce a lo largo del día, como un proceso puntual con un proceso de conteo asociado. Al ser un fenómeno con una tasa no homogénea, se recurrió a los PPNH para su modelización. Como no se tenían los tiempos en los que ocurren las extracciones, pues se dispone de datos agrupados, se recurrió a una metodología estadística no paramétrica específica presentada en el capítulo 2.

En el marco de los PPNH se logró estimar su función de intensidad y su función de intensidad acumulada para los diferentes días de la semana y diferentes cajeros. A su vez, se pudo ver las diferencias y similitudes entre los diferentes días y momentos del mes. Según estas similitudes, en algunos casos se logró obtener una estimación de la función

de intensidad acumulada unificada para dos días cuando fue oportuno.

En el caso particular del cajero b309 tomado como ejemplo para el análisis, se pudo detectar que las extracciones de dinero según el día de la semana y del mes presentan diferentes comportamientos, por lo tanto responden a diferentes PPNH asociados. Como se puede visualizar en el gráfico 4.6, el día domingo es el que se diferencia notoriamente del resto de los días de la semana, y el viernes es el que tiene el mayor nivel de retiros durante todo el día. Este cajero tiene la particularidad de presentar un mismo PPNH para los días miércoles y jueves, esto se concluye a partir de la figura 4.7, en donde se observa que la estimación de la FIA correspondiente al jueves está contenida en los intervalos de confianza al 99% para la estimación de la FIA del miércoles. Es por esto que se decidió unificar la estimación para ambos días. En el gráfico 4.8 se puede observar esta FIA estimada y su contraste con la FIA estimada para el día domingo, la cual no solo es de un nivel considerablemente menor en cuanto al número de ocurrencia de eventos sino que a lo largo del día presenta una forma diferente.

A su vez, se pudo comprobar que el PPNH que caracteriza a las extracciones de los miércoles y jueves, varía según el momento del mes. En el gráfico 4.9 se puede observar claramente como la estimación para la FIA mantiene la misma forma durante todo el día sin importar el momento del mes, y sin embargo el nivel de extracciones difiere para la primera y segunda quincena, siendo para la primera considerablemente mayor que para la segunda (gráfico 4.10). Esta división fue una decisión arbitraria pero resulta interesante ver como la FIA estimada general mantiene el promedio de retiros para la primera y segunda quincena a lo largo de todo el día, ya que como se puede ver en el gráfico 4.10, la curva de la FIA estimada general se encuentra en el medio de las curvas de las FIAs estimadas por quincena.

En la sección 4.4 se puede observar el análisis comparativo realizado para seis cajeros. Según se puede ver en los diferentes gráficos con las estimaciones de las funciones de intensidad y de las FIAs, cada cajero presenta su PPNH único, aunque algunos de ellos presentan FIA con forma similar a lo largo del día. En algunos casos fue posible unificar la estimación de la FIA para dos días de la semana, como se hizo para el cajero b309, mientras que en otros no fue posible dada la variabilidad que presentan a lo largo de la semana.

5.2. Detección de inactividad

La motivación principal para este trabajo fue explorar una metodología que permitiera detectar anomalías en la actividad de los cajeros. Para ello y según el modelado del proceso

que mide las extracciones de dinero, se encontró una forma de calcular la probabilidad de que el cajero carezca de actividad para diferentes períodos de tiempo, es decir la probabilidad de inactividad *natural*, para poder contrastarlo con la realidad y en función de eso predecir si existe una anomalía o no.

Como los datos se encuentran agrupados de a diez minutos, el intervalo de tiempo más pequeño para el que se puede estimar la probabilidad de inactividad es diez minutos y el máximo veinticuatro horas. En particular, se puede calcular esta probabilidad para cualquier tramo del día a partir de un intervalo dado de diez minutos, por ejemplo en un rango de veinte minutos o de dos horas.

En el gráfico 4.11 del capítulo anterior, se muestran las curvas con las estimaciones de la probabilidad de inactividad en la siguiente hora y media hora. Las estimaciones se han calculado para el límite superior de cada intervalo de diez minutos en el caso del PPNH correspondiente a miércoles y jueves del cajero b309.

Estas estimaciones son coherentes con el análisis descriptivo de los datos que se realizó en el primer capítulo, donde se ve que durante las últimas horas del día y sobre todo durante la madrugada la actividad del cajero es muy baja y que luego de las 16 horas se produce una caída drástica.

Se cuenta por lo tanto con una metodología con respaldo estadístico que permite estimar el comportamiento *natural* del proceso con gran flexibilidad, ya que se puede estimar para intervalos de longitudes múltiples de diez minutos, ya sea un tramo corto de veinte minutos, de una hora o de cuatro horas por ejemplo, según se entienda necesario.

5.3. Posibles extensiones del trabajo

Para seguir profundizando el análisis y obtener más resultados se puede continuar con el estudio pormenorizado de los cajeros, agrupando aquellos que puedan presentar similares estimaciones de la FIA e individualizando a los que presenten comportamientos particulares. A su vez, pueden realizarse estimaciones más específicas si se separan los datos por momentos del mes, como se vio en la sección 4.2.1, y del año.

En cuanto a la detección de períodos de inactividad anómalos, se puede por ejemplo, medir cada diez minutos la actividad real de un cajero durante la última media hora y si no hay actividad alguna -siendo que el cajero está en condiciones normales- ver si la probabilidad estimada de que esto sea así es menor a cierto valor y enviar a un colaborador a inspeccionar el lugar. Para determinar la frecuencia del monitoreo, que dispare este accionar, debe realizarse un análisis que tenga en cuenta la perspectiva del negocio.

A su vez, las herramientas desarrolladas permiten calcular la probabilidad de que dado que hubo cierta cantidad de retiros en determinado período de tiempo, haya inactividad en el período siguiente.

Desde el punto de vista del estudio de los PPNH con datos agrupados y su estimación no paramétrica, pueden considerarse otros enfoques para la estimación y desarrollarse contrastes de hipótesis estadísticas.

Bibliografía

- Arkin, B. L. and Leemis, L. M. (2000). Nonparametric estimation of the cumulative intensity function for a nonhomogeneous poisson process from overlapping realizations. *Management Science*, 46(7):989–998.
- Denis Bosq, H. T. N. a. (1996). *A Course in Stochastic Processes: Stochastic Models and Statistical Inference*. Theory and Decision Library 34. Springer Netherlands, 1 edition.
- Henderson, S. G. (2003). Estimation for nonhomogeneous poisson processes from aggregated data. *Operations Research Letters*, 31(5):375 – 382.
- Kuhl, M. E. and Wilson, J. R. (2001). Modeling and simulating poisson processes having trends or nontrigonometric cyclic effects. *European Journal of Operational Research*, 133(3):566 – 582.
- Law, A. and Kelton, W. D. (2013). *Simulation modeling and analysis*. McGraw-Hill Education, Dubuque.
- Leemis, L. M. (1991). Nonparametric estimation of the cumulative intensity function for a nonhomogeneous poisson process. *Management Science*, 37(7):886–900.
- Leemis, L. M. (2004). Technical note: Nonparametric estimation and variate generation for a nonhomogeneous poisson process from event count data. *IIE Transactions*, 36(12):1155–1160.
- Lefebvre, M. (2007). *Applied Stochastic Processes*. Universitext. Springer-Verlag New York, first edition.
- Lewis, P. A. W. and Shedler, G. S. (1979). Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413.
- McCollin, C. (2014). *Intensity Functions for Nonhomogeneous Poisson Processes*. American Cancer Society.

Apéndice A

Glosario

N : Proceso de interés, PPNH.

t : Valores que toma la variable tiempo.

$N(t)$: Valor que toma el proceso N en el momento t .

$\lambda(t)$: Función de intensidad asociada al proceso N .

$\Lambda(t)$: Función de intensidad acumulada asociada al proceso N .

$\hat{\lambda}(t)$: Función de intensidad estimada.

$\hat{\Lambda}(t)$: Función de intensidad acumulada estimada.

$[0, T)$: Intervalo de interés.

k : Cantidad de particiones del intervalo, subintervalos.

δ : Ancho del subintervalo.

n : Cantidad de realizaciones del proceso N .

$l(t)$: Inicio del subintervalo al que pertenece cada t , $l(t) = \lfloor \frac{t}{\delta} \rfloor \delta$. De modo que los subintervalos son $[l(t), l(t) + \delta)$.

Esto asocia a un momento t al intervalo correspondiente, la parte entera del tiempo t dividido el ancho del subintervalo será el comienzo del intervalo al que corresponde, y por lo tanto dicho límite más el ancho del intervalo será el final.

$N_i[a, b)$: Cantidad de eventos que ocurren en la i -ésima observación durante el subintervalo $[a, b)$.

Apéndice B

Demostraciones

Las demostraciones incluidas en este apéndice son extraídas de libros y artículos citados en la bibliografía.

B.1. Demostración de la Proposición 2.2.4

A continuación se reitera la proposición 2.2.4 enunciada por Lefebvre y luego se presenta su demostración.

Sea $\{N(t), t > 0\}$ un PPNH con función de intensidad $\lambda(t)$, se cumple que

$$N(t+h) - N(t) \sim \text{Poisson}(\Lambda(t+h) - \Lambda(t)) \quad \forall t, h \geq 0 \quad (\text{B.1})$$

donde

$$\Lambda(t) = \int_0^t \lambda(\tau) d\tau \quad (\text{B.2})$$

Demostración

Sea $p_n(t, h) = P[N(t+h) - N(t) = n]$ para $n = 0, 1, 2, \dots$

Sabiendo que los incrementos del proceso $N(h), h > 0$ son independientes, y usando las dos condiciones en la definición anterior, se puede escribir, para $n = 1, 2, \dots$ que

$$\begin{aligned}
p_n(t, h + \delta) &= \\
&= P[N(t+h) - N(t) = n, N(t+h+\delta) - N(t+h) = 0] + \\
&P[N(t+h) - N(t) = n-1, N(t+h+\delta) - N(t+h) = 1] + o(\delta) \\
&= p_n(t, h)[1 - \lambda(t+h)\delta + o(\delta)] + p_{n-1}(t, h)[\lambda(t+h)\delta + o(\delta)] + o(\delta)
\end{aligned} \tag{B.3}$$

de lo cual se deduce que

$$p_n(t, h + \delta) - p_n(t, h) = \lambda(t+h)\delta[p_{n-1}(t, h) - p_n(t, h)] + o(\delta)$$

Dividiendo ambos lados por δ y tomando el límite con δ cercana a cero, se obtiene

$$\frac{\partial}{\partial t} p_n(t, h) = \lambda(t+h)[p_{n-1}(t, h) - p_n(t, h)] \tag{B.4}$$

Cuando $n = 0$ la ecuación anterior queda

$$\frac{\partial}{\partial t} p_0(t, h) = -\lambda(t+h)p_0(t, h).$$

La variable h se puede considerar como una constante. Por lo tanto, esta ecuación es una ecuación diferencial ordinaria homogénea lineal de primer orden. Su solución general está dada por

$$p_0(t, h) = c_0 \exp \left\{ - \int_t^{t+h} \lambda(\tau) d\tau \right\},$$

donde c_0 es una constante. Haciendo uso de la condición de borde $p_0(t, 0) = 1$, se obtiene que $c_0 = 1$, y por lo tanto

$$p_0(t, h) = \exp \left\{ - \int_t^{t+h} \lambda(\tau) d\tau \right\} = e^{\Lambda(t) - \Lambda(t+h)},$$

para $t, h \geq 0$.

Sustituyendo esta solución en [B.4](#) se tiene que

$$\frac{\partial}{\partial t} p_1(t, h) = \lambda(t+h) [e^{\Lambda(t)-\Lambda(t+h)} - p_1(t, h)].$$

Se puede reescribir esta ecuación de la siguiente manera

$$\frac{\partial}{\partial t} p_1(t, h) = [e^{\Lambda(t)-\Lambda(t+h)} - p_1(t, h)] \frac{\delta}{\delta t} [\Lambda(t+h) - \Lambda(t)].$$

Se puede comprobar fácilmente que la solución de esta ecuación diferencial no homogénea, que satisface la condición de borde $p_1(t, 0) = 0$ es

$$p_1(t, h) = e^{\Lambda(t)-\Lambda(t+h)} [\Lambda(t+h) - \Lambda(t)],$$

para $t, h \geq 0$.

Finalmente, se puede mostrar mediante inducción que

$$p_n(t, h) = e^{\Lambda(t)-\Lambda(t+h)} \frac{[\Lambda(t+h) - \Lambda(t)]^n}{n!},$$

$\forall t, h \geq 0$ y $n = 0, 1, \dots$

△

B.2. Demostración de la Proposición 2.2.5

A continuación se reitera la proposición 2.2.5 enunciada por Lefebvre y posteriormente se presenta su demostración.

Sea $\{N(t), t > 0\}$ un PPNH con función de intensidad $\lambda(t)$. La probabilidad de la función de densidad de probabilidad de la variable aleatoria $S = T_1 \mid \{N(t) = 1\}$ está dada por

$$f_S(s) = \frac{\lambda(s)}{m(t)} \quad \text{para } 0 < s \leq t$$

Demostración

Si $0 < s \leq t$ y por independencia se puede decir que

$$\begin{aligned} P[T_1 \leq s \mid N(t) = 1] &= \\ &= \frac{P[N(s) = 1, N(t) - N(s) = 0]}{P[N(t) = 1]} \\ &= \frac{(m(s)e^{-m(s)}) e^{-m(t)+m(s)}}{m(t)e^{-m(t)}} = \frac{m(s)}{m(t)} \end{aligned}$$

Ya que

$$\frac{d}{ds}m(s) = \frac{d}{ds} \int_0^s \lambda(u)du = \lambda(s)$$

Se obtiene que

$$f_S(s) = \frac{\lambda(s)}{m(t)} \quad \text{para } 0 < s \leq t$$

△

B.3. Demostración de la Proposición 3.3.3

A continuación se reitera la proposición 3.3.3 enunciada por Henderson y posteriormente se incluye su demostración.

Según las definiciones vertidas en la sección 3.3.1 se cumple que:

1. $\sup_{t \in [0, T]} |\widehat{\Lambda}_n(t) - \widehat{\Lambda}(t)| \rightarrow 0$ casi seguramente cuando $n \rightarrow \infty$

2. $\forall t \in [0, T], \sqrt{n}(\widehat{\Lambda}_n(t) - \widehat{\Lambda}(t)) \Rightarrow \sigma(t)N(0, 1)$ cuando $n \rightarrow \infty$, donde

$$\sigma^2(t) = \Lambda(\ell(t)) + \left(\frac{t - \ell(t)}{\delta} \right)^2 \times [\Lambda(\ell(t) + \delta)] - \Lambda(\ell(t))$$

3. $\forall n \geq 1, E\widehat{\Lambda}_n(t) = \widehat{\Lambda}(t)$. Si λ es continuamente diferenciable en un intervalo abierto que contiene $[\ell(t), \ell(t) + \delta]$ entonces

$$|E\widehat{\Lambda}_n(t) - \Lambda(t)| \leq |\lambda'(\xi)|\delta^2$$

para algún $\xi \in [\ell(t), \ell(t) + \delta]$

Demostración

$N_i(a, b)$ es Poisson con media y varianza $\Lambda(b) - \Lambda(a) \forall 0 \leq a < b$ y $\forall i$. Aplicando la ley fuerte de los grandes números a cada uno de los promedios en la ecuación 3.1 se cumple la ley fuerte para cada t . Queda pendiente establecer la parte uniforme del resultado.

$\widehat{\Lambda}$ es una función continua no decreciente, por lo tanto para todo ϵ existen $m(\epsilon) < \infty$ y puntos $u_0 = 0, u_1, \dots, u_{m(\epsilon)} = T$ de modo que $\widehat{\Lambda}(u_i) - \widehat{\Lambda}(u_{i-1}) \leq \epsilon$ para todo $i = 1, \dots, m(\epsilon)$. Para $t \in [0, T]$, sea $a(t) = a_\epsilon(t)$, i tal que $t \in [u_i, u_{i+1})$.

Entonces

$$\begin{aligned}
|\widehat{\Lambda}_n(t) - \widehat{\Lambda}(t)| &= \max\{\widehat{\Lambda}_n(t) - \widehat{\Lambda}(t), \widehat{\Lambda}(t) - \widehat{\Lambda}_n(t)\} \\
&\leq \max\{\widehat{\Lambda}_n(u_{a(t)+1}) - \widehat{\Lambda}(u_{a(t)}), \widehat{\Lambda}(u_{a(t)}) - \widehat{\Lambda}_n(u_{a(t)+1})\} \\
&\leq \max\{|\widehat{\Lambda}_n(u_{a(t)+1}) - \widehat{\Lambda}(u_{a(t)+1})| + \widehat{\Lambda}(u_{a(t)+1}) - \widehat{\Lambda}(u_{a(t)}), \\
&\quad \widehat{\Lambda}(u_{a(t)}) - \widehat{\Lambda}(u_{a(t)+1}) + \widehat{\Lambda}(u_{a(t)+1}) - \widehat{\Lambda}_n(u_{a(t)+1})|\} \\
&\leq \varepsilon + \max_{i=1, \dots, m(\varepsilon)} |\widehat{\Lambda}_n(u_i) - \widehat{\Lambda}(u_i)| \tag{B.5}
\end{aligned}$$

donde se usa la monotonía de $\widehat{\Lambda}_n$ and $\widehat{\Lambda}$. También la cota B.5 se cumple para todo $t \in [0, T)$. Tomando el límite supremo en la expresión anterior, se obtiene que casi seguramente $\lim_{n \rightarrow \infty} \sup_{t \in [0, T)} |\widehat{\Lambda}_n(t) - \widehat{\Lambda}(t)| \leq \varepsilon$

Como ε es arbitrario, esto completa la prueba de la uniformidad en sentido fuerte.

Yendo al segundo resultado, el teorema del límite central estándar permite concluir que

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n N_i(a, b) - (\Lambda(b) - \Lambda(a)) \right) \Rightarrow \mathcal{N}(0, \Lambda(b) - \Lambda(a))$$

cuando $n \rightarrow \infty$.

Del Teorema del límite central para $\widehat{\Lambda}_n(t)$ se desprende esta observación y la propiedad de incrementos independientes de N .

El hecho de que $E\widehat{\Lambda}_n(t) = \widehat{\Lambda}(t)$ es inmediato a partir de las definiciones de $\widehat{\Lambda}_n(t)$ y $\widehat{\Lambda}(t)$, pues $EN_i(0, T) = \Lambda(t)$ para todo $t \geq 0$ y todo i .

Finalmente

$$\begin{aligned}
E\widehat{\Lambda}_n(t) - \Lambda(t) &= \widehat{\Lambda}(t) - \Lambda(t) \\
&= [\Lambda(\ell(t)) - \Lambda(t)] + \frac{t - \ell(t)}{\delta} [\Lambda(\ell(t) + \delta) - \Lambda(\ell(t))] \\
&= -\lambda(\xi)(t - \ell(t)) + \frac{t - \ell(t)}{\delta} \lambda(\theta)\delta \tag{B.6}
\end{aligned}$$

Para algún $\xi \in [\ell(t), t]$ y $\theta \in [\ell(t), \ell(t) + \delta]$ por el teorema de la media.

Entonces

$$| E\widehat{\Lambda}_n(t) - \Lambda(t) | = (t - \ell(t)) | \lambda(\theta) - \lambda(\xi) | \leq \delta | \lambda'(\xi) | \delta$$

,

donde $\xi \in [\ell(t), \ell(t) + \delta]$.

\triangle

Apéndice C

Simulación de un PPNH

A continuación se muestra el pseudocódigo para generar tiempos de eventos de un PPNH a partir de los tiempos $a_0, a_1, a_2, \dots, a_m$; la cantidad de eventos n_1, n_2, \dots, n_m y la cantidad de intervalos k .

- Se calcula el límite superior para el PPH
- Se inicia el contador del intervalo (i)
- Se inicia el contador de la variable (j)
- Se inicia la acumulación de intensidad
- Se genera el primer número aleatorio a partir de una distribución uniforme $(0, 1)$
- Se genera el primer evento del HPP en el tiempo

Mientras haya más eventos para generar y está en el intervalo incorrecto

- Incrementar el contador del intervalo
- Incrementar la intensidad acumulada
- Generar el tiempo del evento j -ésimo
- Incrementar el contador de la variable
- Generar un j -ésimo número aleatorio
- Generar el siguiente tiempo de los eventos HPP
- Devuelve los tiempos de los eventos NHPP

Apéndice D

Tablas y gráficos

En este apéndice se presentan tablas y gráficos que a pesar de haber sido excluidos del informe, sirven de apoyo visual para el entendimiento del problema y permiten visualizar detalles de posible interés.

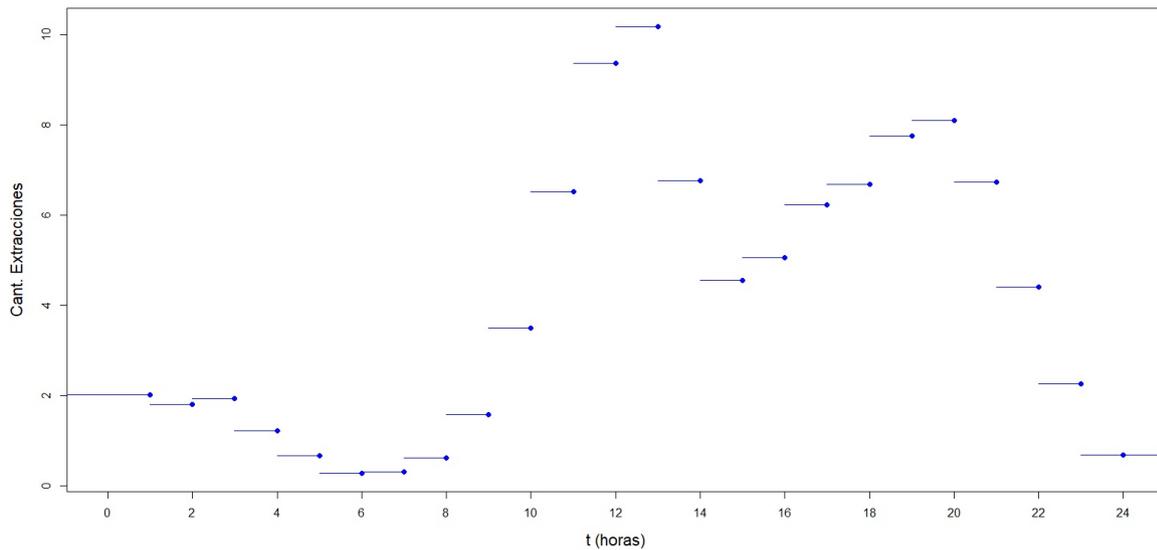


Figura D.1: Estimación de la función de intensidad del domingo para el cajero b309

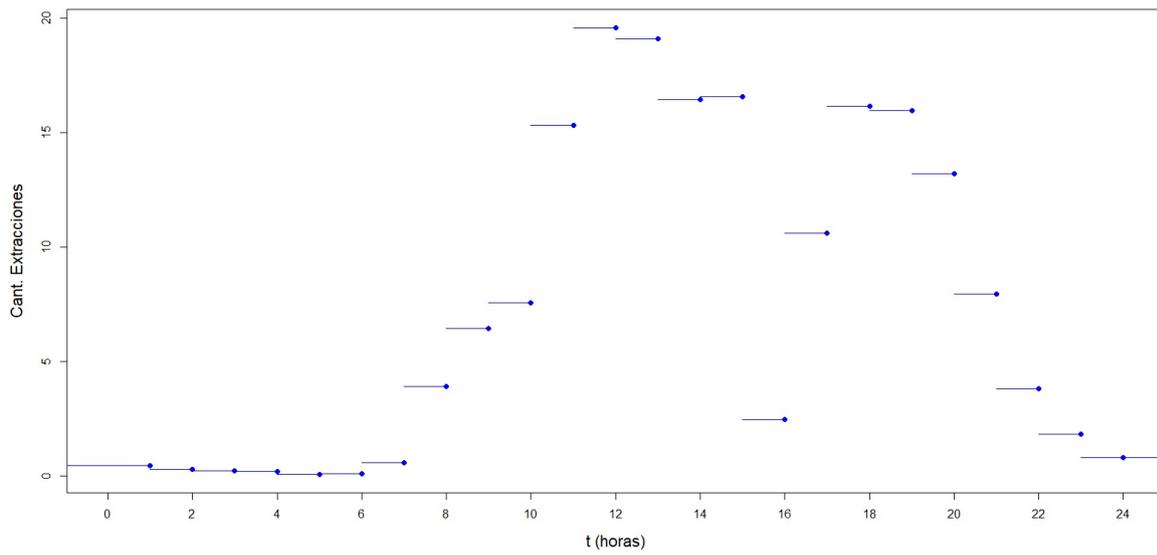


Figura D.2: Estimación de la función de intensidad del lunes para el cajero b309

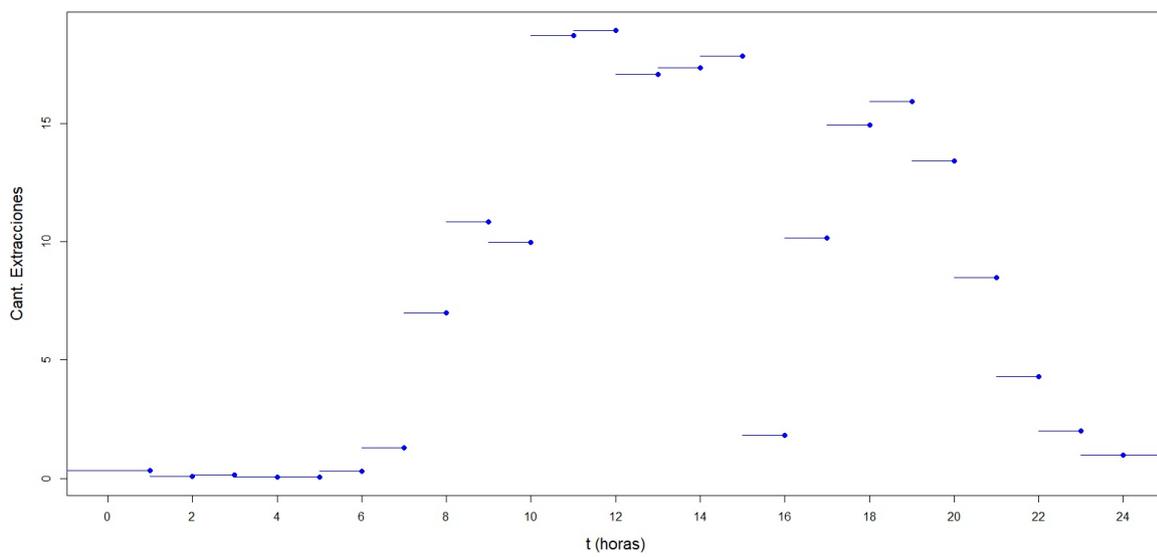


Figura D.3: Estimación de la Función de Intensidad del martes para el cajero b309

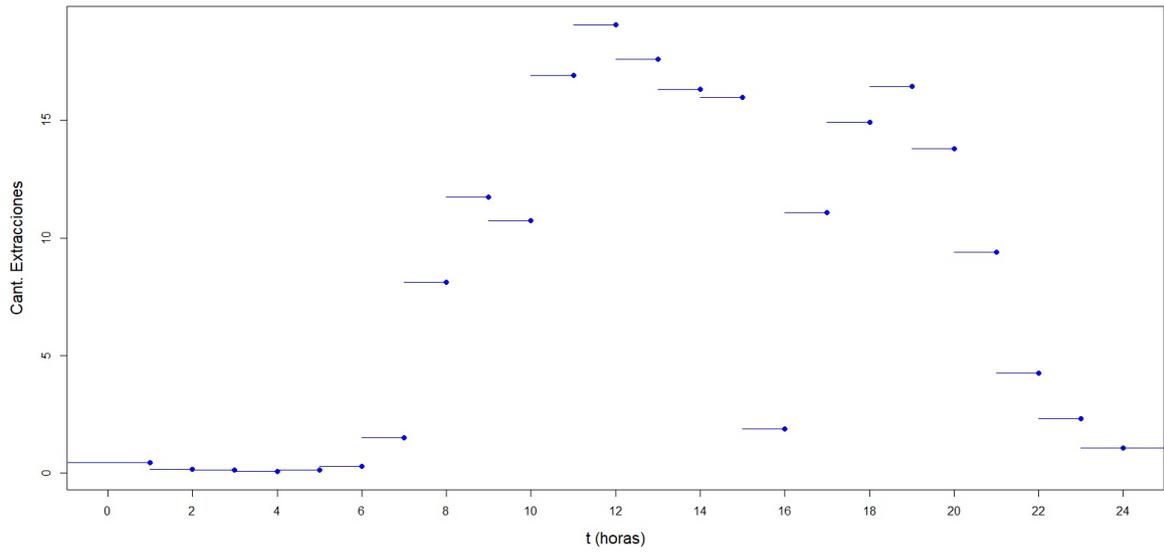


Figura D.4: Estimación de la función de intensidad del miércoles para el cajero b309

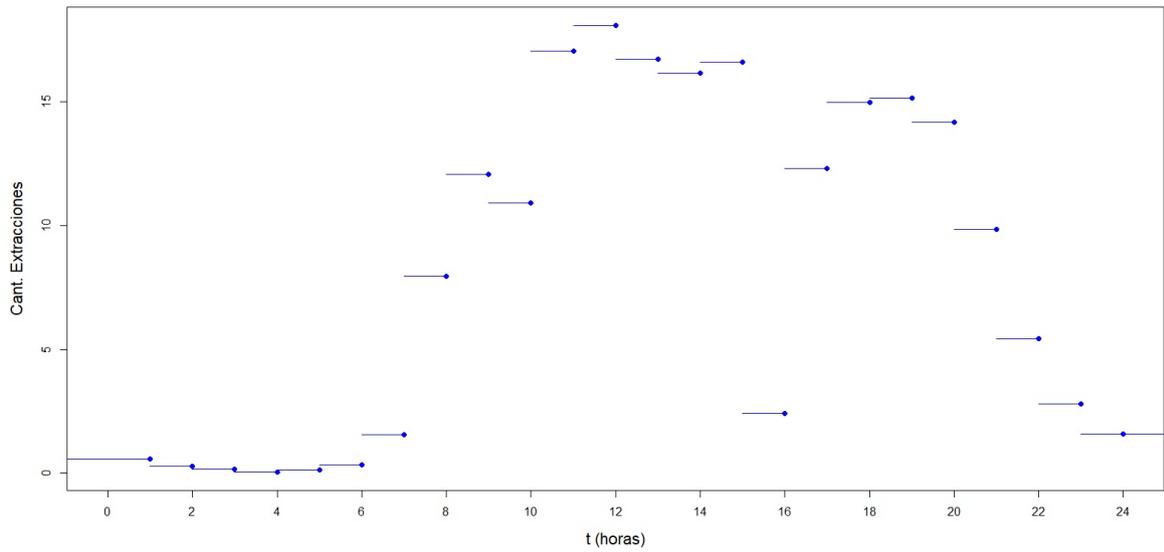


Figura D.5: Estimación de la función de intensidad del jueves para el cajero b309

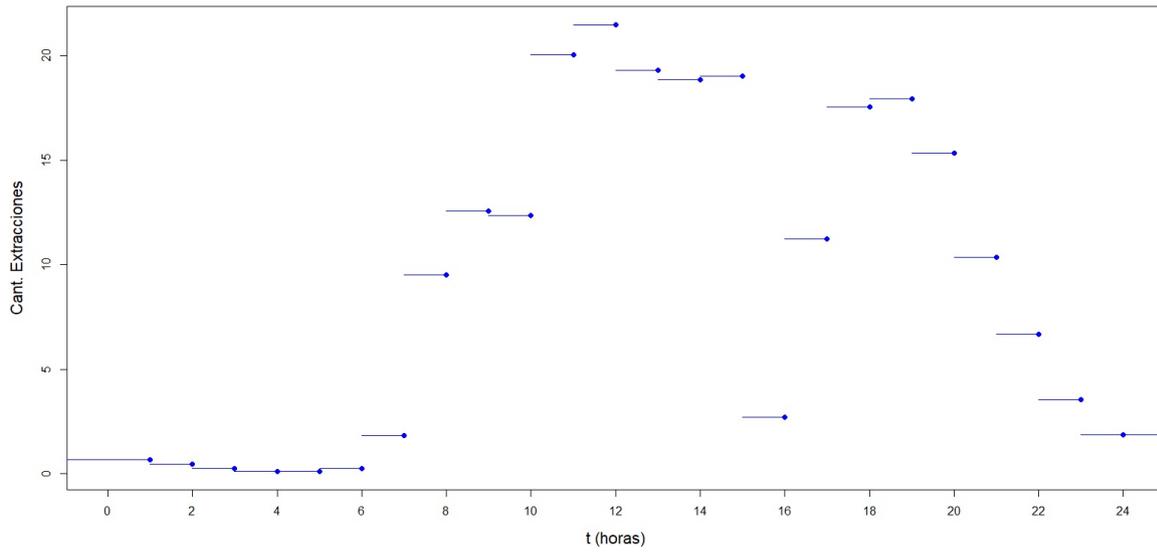


Figura D.6: Estimación de la función de intensidad del viernes para el cajero b309

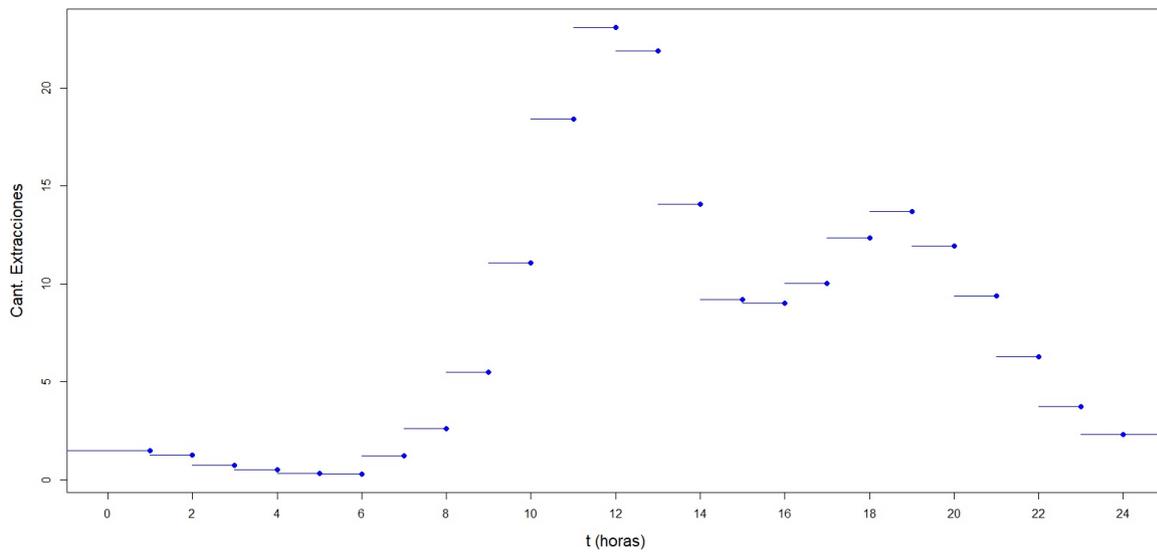


Figura D.7: Estimación de la función de intensidad del sábado para el cajero b309

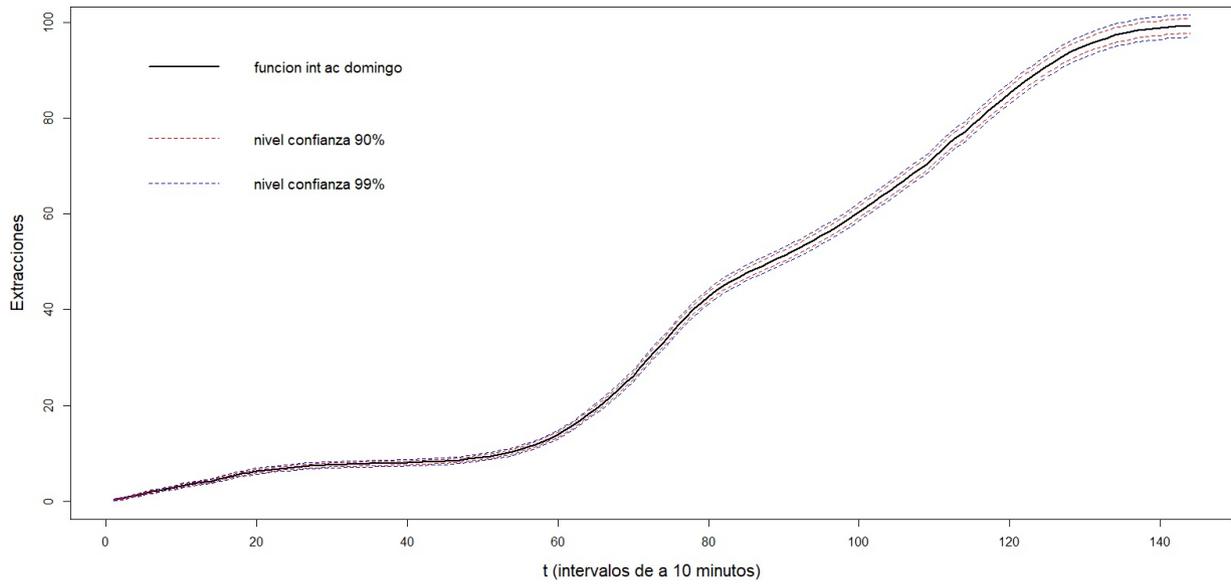


Figura D.8: Estimación de la función de intensidad acumulada para para el domingo para el cajero b309

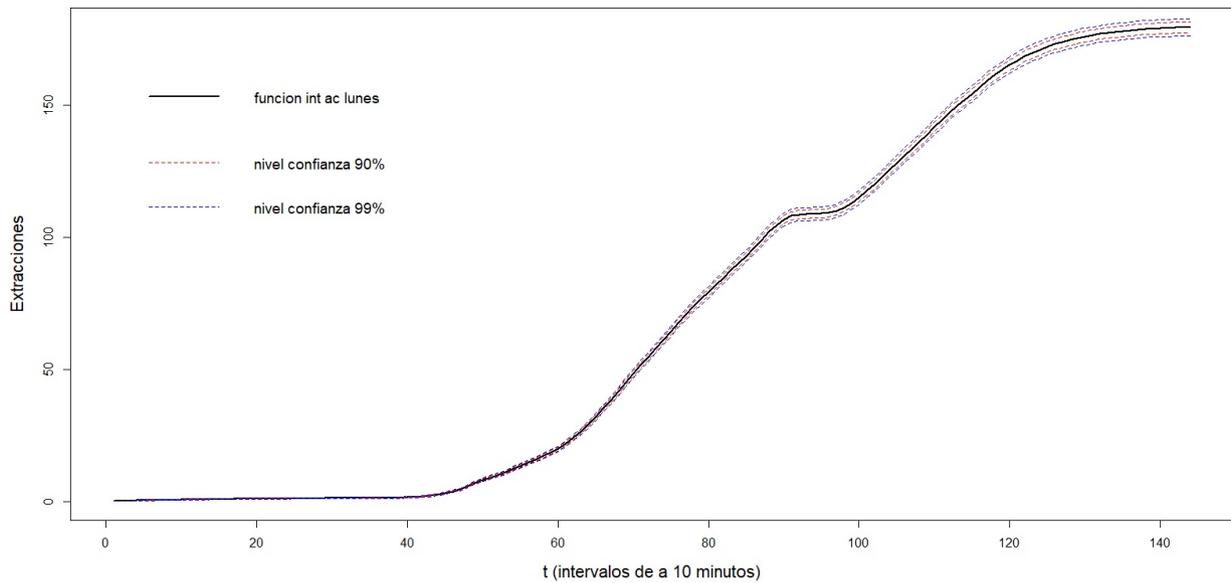


Figura D.9: Estimación de la función de intensidad acumulada para para el lunes para el cajero b309

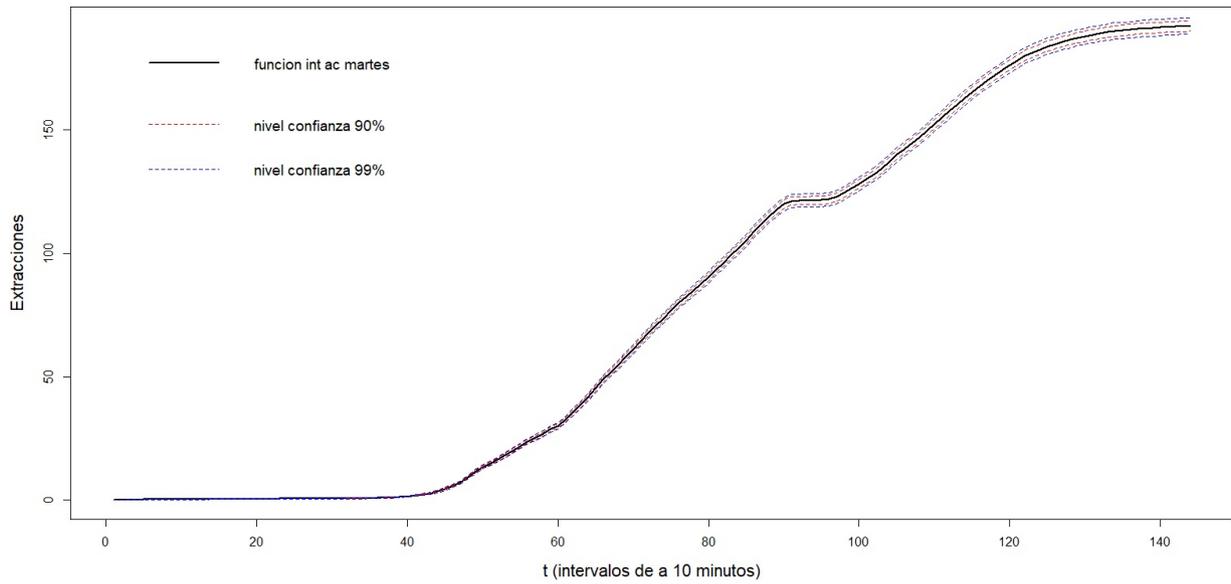


Figura D.10: Estimación de la función de intensidad acumulada para para el martes para el cajero b309

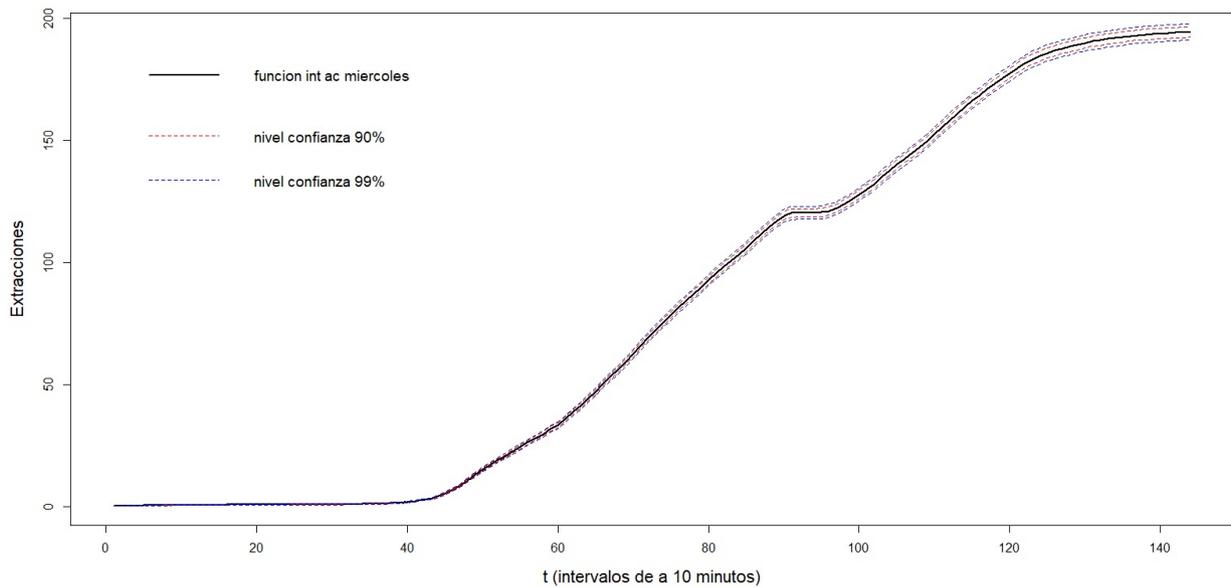


Figura D.11: Estimación de la función de intensidad acumulada para para el miércoles para el cajero b309

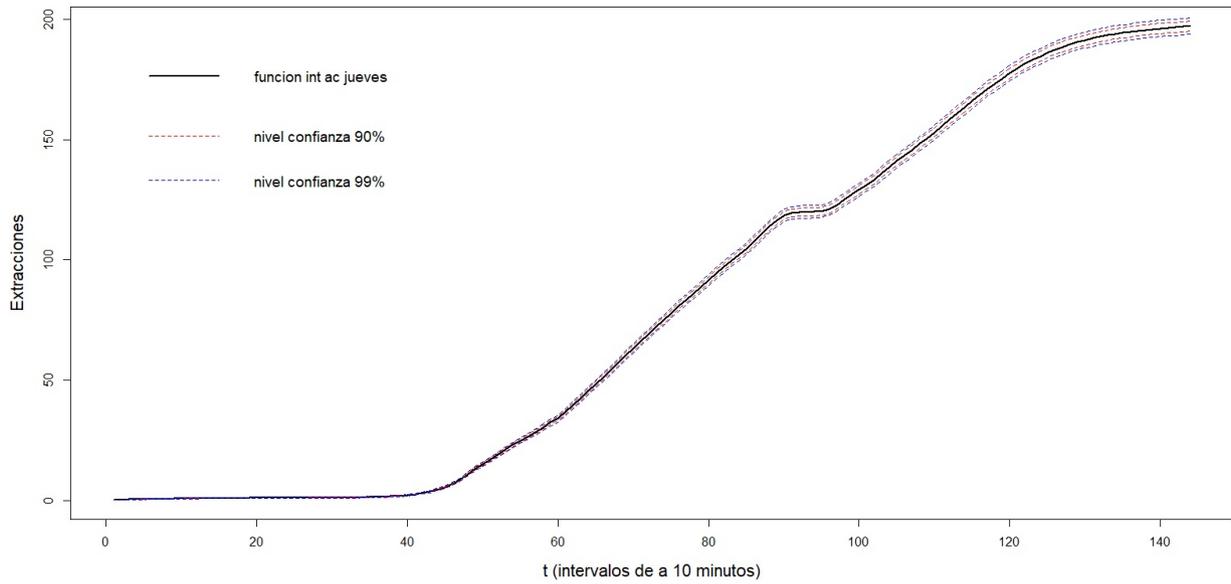


Figura D.12: Estimación de la función de intensidad acumulada para para el jueves para el cajero b309

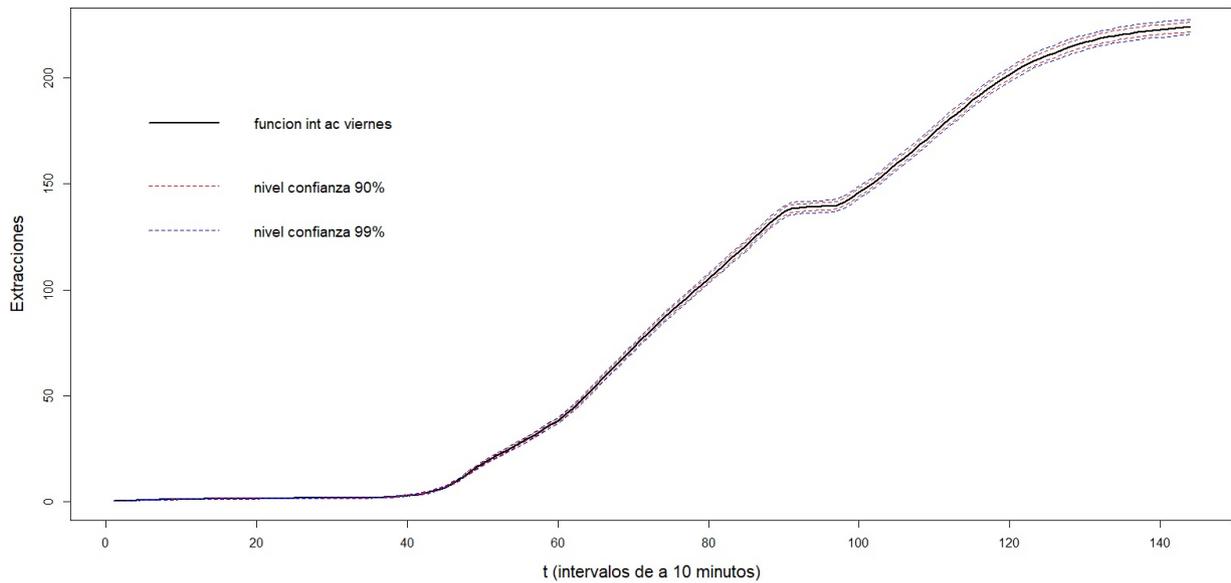


Figura D.13: Estimación de la función de intensidad acumulada para para los viernes para el cajero b309

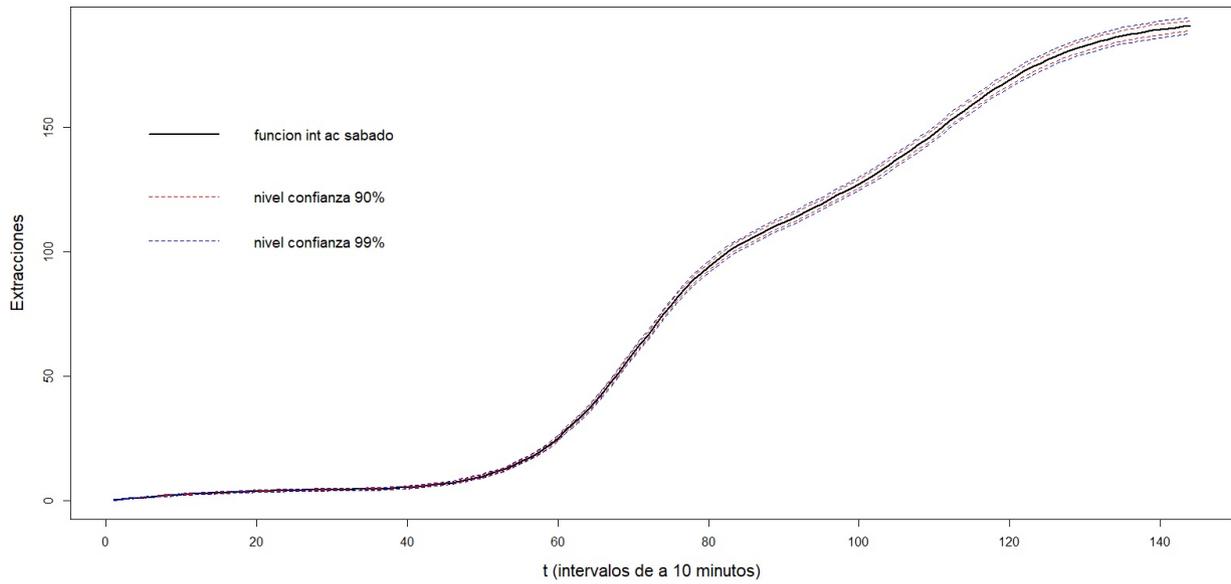


Figura D.14: Estimación de la función de intensidad acumulada para para el sábado para el cajero b309

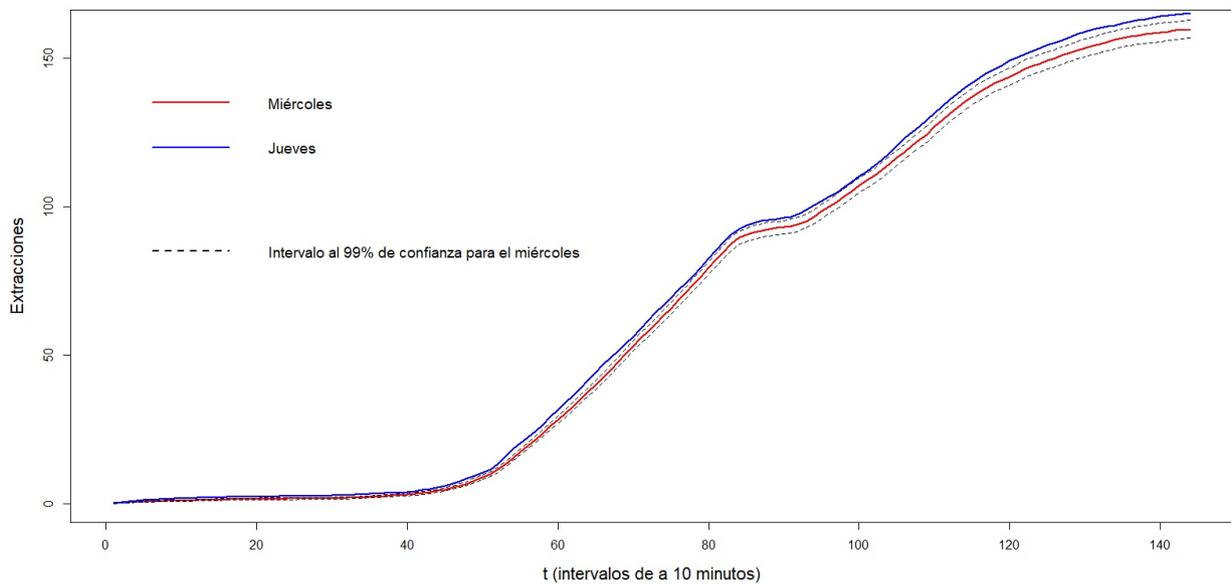
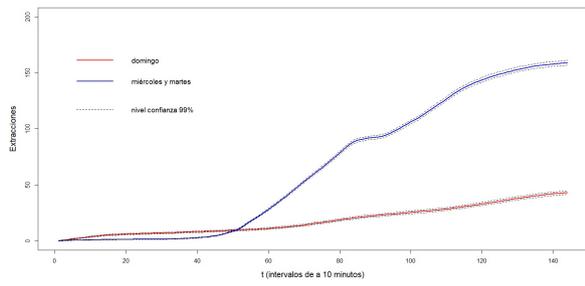
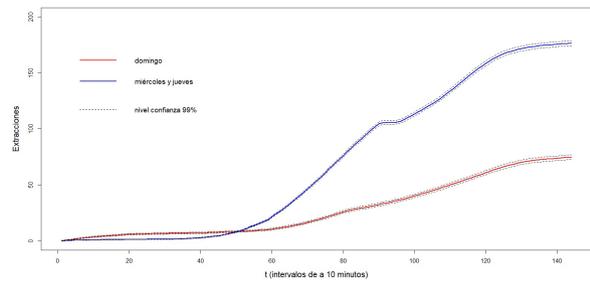


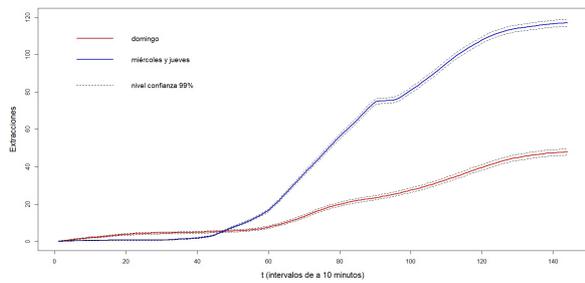
Figura D.15: Estimación de la función de intensidad acumulada del cajero b397 para miércoles y jueves



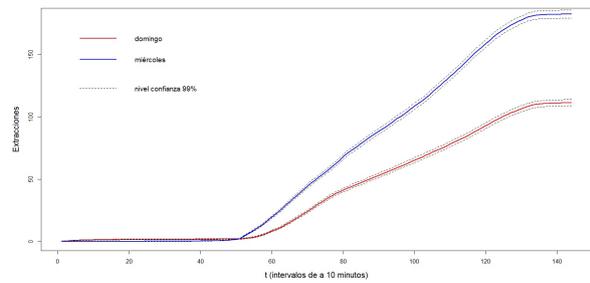
(a) Cajero b397 - Domingo vs. Miércoles y Martes



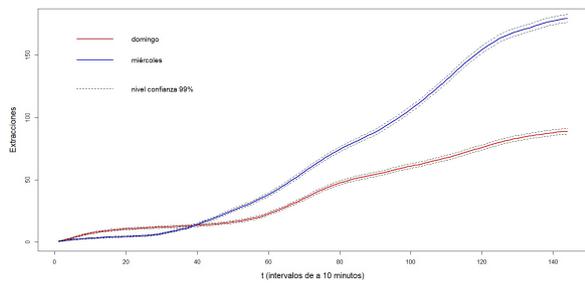
(b) Cajero b435 - Domingo vs. Miércoles y Jueves



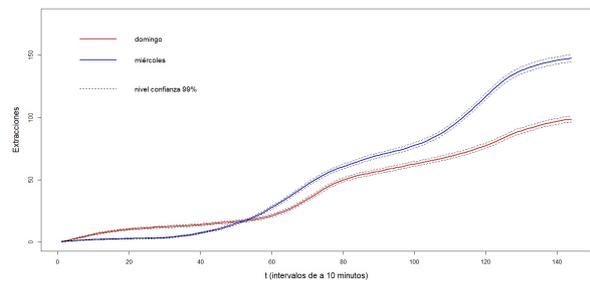
(c) Cajero b834 - Domingo vs. Miércoles y Jueves



(d) Cajero m389 - Domingo vs. Miércoles



(e) Cajero m404 - Domingo vs. Miércoles



(f) Cajero m437 - Domingo vs. Miércoles

Figura D.16: Comparación de la estimación de la FIA para diferentes días según los cajeros de muestra

Cuadro D.1: Probabilidad estimada de inactividad del cajero b309 por hora para miércoles y jueves

Tramo horario	00:00-01:00 hs	01:00-02:00 hs	02:00-03:00 hs	03:00-04:00 hs	04:00-05:00 hs	05:00-06:00 hs
$\hat{P}(Inactividad)$	0,691034	0,822298	0,908780	0,889237	0,744046	0,230025
Tramo horario	06:00-07:00 hs	07:00-08:00 hs	08:00-09:00 hs	09:00-10:00 hs	10:00-11:00 hs	11:00-12:00 hs
$\hat{P}(Inactividad)$	0,230025	0,000558	0,000058	0,000122	$5,430378e^{-7}$	$1,879751e^{-1}$
Tramo horario	12:00-13:00 hs	13:00-14:00 hs	14:00-15:00 hs	15:00-16:00 hs	16:00-17:00 hs	17:00-18:00 hs
$\hat{P}(Inactividad)$	$6,518325e^{-7}$	$1,582473e^{-6}$	$1,324093e^{-6}$	0,321492	$3,668864e^{-5}$	$4,965267e^{-6}$
Tramo horario	18:00-19:00 hs	19:00-20:00 hs	20:00-21:00 hs	21:00-22:00 hs	22:00-23:00 hs	23:00-24:00 hs
$\hat{P}(Inactividad)$	$1,555190e^{-6}$	$1,067265e^{-5}$	0,000593	0,021700	0,140125	0,476630

Cuadro D.2: Probabilidad de Inactividad estimada para cada cajero por hora

Tramo horario	$\hat{P}(Inac_{b397})$	$\hat{P}(Inac_{b435})$	$\hat{P}(Inac_{b834})$	$\hat{P}(Inac_{m389})$	$\hat{P}(Inac_{m404})$	$\hat{P}(Inac_{m437})$
00:00 - 01:00 hs	0,5370	0,5209	0,6559	0,9288	0,4454	0,5536
01:00 - 02:00 hs	0,7473	0,7249	0,8663	0,9913	0,6198	0,7217
02:00 - 03:00 hs	0,8331	0,9009	0,9088	1,0000	0,7376	0,8367
03:00 - 04:00 hs	0,8626	0,8970	0,9828	0,9957	0,7408	0,8892
04:00 - 05:00 hs	0,8892	0,9207	0,9492	0,9913	0,4155	0,8116
05:00 - 06:00 hs	0,5833	0,5782	0,5732	0,9616	0,2118	0,4155
06:00 - 07:00 hs	0,3402	0,3187	0,3743	0,8626	0,0652	0,2382
07:00 - 08:00 hs	0,0406	0,0730	0,0318	0,7538	0,0572	0,1445
08:00 - 09:00 hs	0,0004	0,0047	0,0264	0,0507	0,0587	0,0931
09:00 - 10:00 hs	$2,778e^{-5}$	0,0005	0,0053	0,0055	0,0287	0,0193
10:00 - 11:00 hs	$4,512e^{-6}$	$7,636e^{-6}$	$5,792e^{-5}$	0,0017	0,0112	0,0117
11:00 - 12:00 hs	$2,773e^{-6}$	$1,047e^{-6}$	$4,423e^{-5}$	0,0024	0,0071	0,0090
12:00 - 13:00 hs	$2,986e^{-6}$	$2,189e^{-7}$	$4,309e^{-5}$	0,0035	0,0104	0,0314
13:00 - 14:00 hs	$3,051e^{-6}$	$2,767e^{-7}$	0,0001	0,0036	0,0267	0,0939
14:00 - 15:00 hs	0,0753	$1,324e^{-6}$	$8,059e^{-5}$	0,0075	0,0307	0,1188
15:00 - 16:00 hs	0,0020	0,2632	0,3229	0,0071	0,0115	0,1719
16:00 - 17:00 hs	0,0001	0,0001	0,0019	0,0050	0,0058	0,0927
17:00 - 18:00 hs	$2,085e^{-5}$	$3,528e^{-5}$	0,0007	0,0017	0,0023	0,0254
18:00 - 19:00 hs	$2,274e^{-5}$	$3,506e^{-6}$	0,0008	0,0014	0,0015	0,0037
19:00 - 20:00 hs	0,0011	$7,976e^{-6}$	0,0038	0,0015	0,0036	0,0013
20:00 - 21:00 hs	0,0074	0,0004	0,0287	0,0049	0,0145	0,0027
21:00 - 22:00 hs	0,0267	0,02286	0,1787	0,0367	0,08838	0,0399
22:00 - 23:00 hs	0,0952	0,1933	0,2973	0,5757	0,1007	0,1497
23:00 - 24:00 hs	0,2834	0,3462	0,5075	0,8663	0,2834	0,3038

Apéndice E

Código para R

```
## se lee la base (cajero 309) y se preparan los datos para trabajarlos a posteriori
b309<- read.table("b309.txt",sep=";",header=FALSE,
  col.names=c("Fecha","Fecha2","Monto","Cant.Retiros","Seg.Funcionando","No_se","No_se2"))
b309<- b309[,c(1,3:5)]
dim(b309);head(b309)

## problem:el 1er día pone NA: solución sumar 1 a la column fecha
b309$Fecha<- b309$Fecha+1
b309$Fecha<- strptime(b309$Fecha,"%Y%m%d %H%M%S")
class(b309$Fecha)
names(unclass(b309$Fecha))

## elem 7 es el weekday (0=dom), se divide la base por día de la semana
b309_tot<- b309
b309_sun<- subset(b309, b309$Fecha[[7]]==0)
b309_mon<- subset(b309, b309$Fecha[[7]]==1)
b309_tue<- subset(b309, b309$Fecha[[7]]==2)
b309_wed<- subset(b309, b309$Fecha[[7]]==3)
b309_thu<- subset(b309, b309$Fecha[[7]]==4)
b309_fri<- subset(b309, b309$Fecha[[7]]==5)
b309_sat<- subset(b309, b309$Fecha[[7]]==6)

int = 10 # 10 minutos
obs = 60/int * 24 # hora (60 min)/ intervalo por 24 horas del día
sun = length(b309_sun$Fecha)/obs
mon = length(b309_mon$Fecha)/obs
tue = length(b309_tue$Fecha)/obs
wed = length(b309_wed$Fecha)/obs
thu = length(b309_thu$Fecha)/obs
fri = length(b309_fri$Fecha)/obs
sat = length(b309_sat$Fecha)/obs
min_obs = min(sun,mon,tue,wed,thu,fri,sat)
# cant mínima que hay x día de la semana (115 dom, 115 lun,...)

## trunca en 16560 observaciones=115días
b309_tot<- b309_tot[1:((obs*min_obs)*7),]
b309_sun<- b309_sun[1:(obs*min_obs),]
b309_mon<- b309_mon[1:(obs*min_obs),]
```

```

b309_tue<- b309_tue[1:(obs*min_obs),]
b309_wed<- b309_wed[1:(obs*min_obs),]
b309_thu<- b309_thu[1:(obs*min_obs),]
b309_fri<- b309_fri[1:(obs*min_obs),]
b309_sat<- b309_sat[1:(obs*min_obs),]

## separa el data frame por día (se crea column día)
b309_tot$dia<- as.Date(b309_tot$Fecha)
b309_sun$dia<- as.Date(b309_sun$Fecha)
b309_mon$dia<- as.Date(b309_mon$Fecha)
b309_tue$dia<- as.Date(b309_tue$Fecha)
b309_wed$dia<- as.Date(b309_wed$Fecha)
b309_thu$dia<- as.Date(b309_thu$Fecha)
b309_fri$dia<- as.Date(b309_fri$Fecha)
b309_sat$dia<- as.Date(b309_sat$Fecha)

## suma las extracciones por hora y se hace promedio de las mismas
b309_tot$Horas<- b309_tot$Fecha[[3]]
b309_sun$Horas<- b309_sun$Fecha[[3]]
b309_mon$Horas<- b309_mon$Fecha[[3]]
b309_tue$Horas<- b309_tue$Fecha[[3]]
b309_wed$Horas<- b309_wed$Fecha[[3]]
b309_thu$Horas<- b309_thu$Fecha[[3]]
b309_fri$Horas<- b309_fri$Fecha[[3]]
b309_sat$Horas<- b309_sat$Fecha[[3]]

atot<- by(b309_tot$Cant.Retiros, b309_tot$Horas, sum);atot=atot/(min_obs*7)
## suma las extracciones por hora y promedia por día
a0<- by(b309_sun$Cant.Retiros, b309_sun$Horas, sum);a0=a0/min_obs
a1<- by(b309_mon$Cant.Retiros, b309_mon$Horas, sum);a1=a1/min_obs
a2<- by(b309_tue$Cant.Retiros, b309_tue$Horas, sum);a2=a2/min_obs
a3<- by(b309_wed$Cant.Retiros, b309_wed$Horas, sum);a3=a3/min_obs
a4<- by(b309_thu$Cant.Retiros, b309_thu$Horas, sum);a4=a4/min_obs
a5<- by(b309_fri$Cant.Retiros, b309_fri$Horas, sum);a5=a5/min_obs
a6<- by(b309_sat$Cant.Retiros, b309_sat$Horas, sum);a6=a6/min_obs

##### ANÁLISIS DESCRIPTIVO

## Diagrama de caja y de barras de la cantidad de extracciones
boxplot(b309_tot$Cant.Retiros, col="skyblue", ylab="Extracciones", cex.lab=1.4)
tn<-table(b309_tot$Cant.Retiros)
r<-barplot(tn, col="white", border="white", ylab="Frecuencia",xlab="Cantidad de Retiros",
          cex.lab=1.4)
lines(r, tn, type="h", col="skyblue", lwd=5)

b309_tot$Minutos<- b309_tot$Fecha[[2]] #creo variable minutos
b309_tot$Hr_Min<- paste0(b309_tot$Horas,":",b309_tot$Minutos) #creo variable hr_min
btot<- by(b309_tot$Cant.Retiros, b309_tot$Hr_Min, sum) #suma extracciones por 10 min
#ordeno
b1tot=btot [1:12];b2tot=btot [13:72];b3tot=btot [73:78];b4tot=btot [79:102];b5tot=btot [103:144]
bbtot=c(b1tot,b3tot,b5tot,b2tot,b4tot)
bbprom<-bbtot/805
barplot(bbtot, col="Skyblue", cex.lab=1.4, xlab="Horas del día", ylab="Cantidad de Retiros")
barplot(bbprom, col="Skyblue",cex.lab=1.4, xlab="Horas del día", ylab="Cantidad de Retiros")

btott<- by(b309_tot$Cant.Retiros, b309_tot$Horas, sum) #suma extracciones por hr
bbpromm<-btott/805
barplot(btott, col="Skyblue", cex.lab=1.4,xlab="Horas del día",ylab="Cantidad de Retiros")

```

```

barplot(bbpromm,col="Skyblue",cex.lab=1.4,xlab="Horas del día",ylab="Cantidad de Retiros")

## gráfico de un martes cualquiera: verifica incrementos no estacionarios
tue_08_03<-subset(b309_tue, b309_tue$dia=="2010-08-03")
plot(cumsum(tue_08_03$Cant.Retiros), type="s", ylab="Cantidad de extracciones",
      xlab="t (intervalos de 10 minutos)", cex.lab=1.4, xaxp=c(0,144,12))

tue_08_03_hs<- by(tue_08_03$Cant.Retiros, tue_08_03$Horas,sum)
plot(cumsum(tue_08_03_hs), type="s", ylab="Cantidad de extracciones", cex.lab=1.4,
      xlab="t (horas)", xaxp=c(0,24,12))

##### FUNCIÓN DE INTENSIDAD (FI) (según paper de Leemis)

## estimación de la intensidad por hora
est_tot=NULL
for (i in 2:25){
  j<- i-1
  est_tot[j]<- (atot[i-1]/(i-(i-1)))
}

## gráfico de la estimación constante a tramos (hora y 10 min)
xx=c(1:23)
plot(stepfun(xx, est_tot, right=TRUE), verticals=FALSE, xlim=c(0,24), pch=16, main=NULL,
      xlab="t (horas)", col="blue", ylab="Cant. Extracciones", cex.lab=1.4, xaxp=c(0,24,12),
      xval=c(1:24))

yy=c(1:143)
plot(stepfun(yy,bbtot, right=TRUE), verticals=FALSE, xlim=c(0,144),col="blue", pch=16,
      xlab="t (intervalos 10 minutos)", ylab=("Cant. Extracciones"),main=NULL, cex.lab=1.4,
      xval=c(1:144))

## FI domingos por hora
est0=NULL
for (i in 2:25){
  j<- i-1
  est0[j]<- (a0[i-1]/(i-(i-1)))
}

plot(stepfun(xx, est0, right=TRUE), verticals=FALSE, xlim=c(0,24), pch=16, main=NULL,
      xlab="t (horas)", col="blue", ylab="Cant. Extracciones", cex.lab=1.4, xaxp=c(0,24,12),
      xval=c(1:24))

## FI lunes por hora
est1=NULL
for (i in 2:25){
  j<- i-1
  est1[j]<- (a1[i-1]/(i-(i-1)))
}

plot(stepfun(xx, est1, right=TRUE), verticals=FALSE, xlim=c(0,24), pch=16, main=NULL,
      xlab="t (horas)", col="blue", ylab="Cant. Extracciones", cex.lab=1.4, xaxp=c(0,24,12),
      xval=c(1:24))

## FI martes por hora
est2=NULL
for (i in 2:25){
  j<- i-1
  est2[j]<- (a2[i-1]/(i-(i-1)))
}

```

```

}

plot(stepfun(xx, est2, right=TRUE), verticals=FALSE, xlim=c(0,24), pch=16, main=NULL,
      xlab="t (horas)", col="blue", ylab="Cant. Extracciones", cex.lab=1.4, xaxp=c(0,24,12),
      xval=c(1:24))

## FI miércoles por hora
est3=NULL
for (i in 2:25){
  j<- i-1
  est3[j]<- (a3[i-1]/(i-(i-1)))
}

plot(stepfun(xx, est3, right=TRUE), verticals=FALSE, xlim=c(0,24), pch=16, main=NULL,
      xlab="t (horas)", col="blue", ylab="Cant. Extracciones", cex.lab=1.4, xaxp=c(0,24,12),
      xval=c(1:24))

## FI jueves por hora
est4=NULL
for (i in 2:25){
  j<- i-1
  est4[j]<- (a4[i-1]/(i-(i-1)))
}

plot(stepfun(xx, est4, right=TRUE), verticals=FALSE, xlim=c(0,24), pch=16, main=NULL,
      xlab="t (horas)", col="blue", ylab="Cant. Extracciones", cex.lab=1.4, xaxp=c(0,24,12),
      xval=c(1:24))

## FI viernes por hora
est5=NULL
for (i in 2:25){
  j<- i-1
  est5[j]<- (a5[i-1]/(i-(i-1)))
}

plot(stepfun(xx, est5, right=TRUE), verticals=FALSE, xlim=c(0,24), pch=16, main=NULL,
      xlab="t (horas)", col="blue", ylab="Cant. Extracciones", cex.lab=1.4, xaxp=c(0,24,12),
      xval=c(1:24))

## FI sábados por hora
est6=NULL
for (i in 2:25){
  j<- i-1
  est6[j]<- (a6[i-1]/(i-(i-1)))
}

plot(stepfun(xx, est6, right=TRUE), verticals=FALSE, xlim=c(0,24), pch=16, main=NULL,
      xlab="t (horas)", col="blue", ylab="Cant. Extracciones", cex.lab=1.4, xaxp=c(0,24,12),
      xval=c(1:24))

## gráfico de las intensidades por día para visualizar posibles similitudes
plot(est_tot, type="l", col="black", lwd=2, lty=2.4, xlab="t (horas del día)", xaxp=c(0,24,12),
      ylab="Cant. Extracciones estimada", cex.lab=1.4, ylim=c(0,24), xlim=c(0,24))
lines(est0,col="red", lwd=2)
lines(est1,col="pink", lwd=2)
lines(est2,col="blue", lwd=2)
lines(est3,col="orange",lwd=2)
lines(est4,col="green", lwd=2)

```

```

lines(est5,col="grey", lwd=2)
lines(est6,col="purple", lwd=2)
legend(1,20,bty="n",c("domingo","lunes", "martes", "miercoles", "jueves ", "viernes", "sabado")
      ,cex=1.2,col=c("red","pink","blue","orange", "green", "grey", "purple"), lty=1, lwd=2)
legend(1,23, bty="n", c("total") ,cex=1.2,col=c("black"), lty=2, lwd=2)

## medida de posición:mediana; medida de tendencia central=media;
## medidas de dispersión=max, min, desvío, varianza
summary(est_tot);sd(est_tot);var(est_tot)
summary(est0);sd(est0);var(est0) # dom
summary(est1);sd(est1);var(est1) # lunes
summary(est2);sd(est2);var(est2) # martes
summary(est3);sd(est3);var(est3) # mierc
summary(est4);sd(est4);var(est4) # juev
summary(est5);sd(est5);var(est5) # viern
summary(est6);sd(est6);var(est6) # sab

## boxplot (muestra los cuartiles de las extracciones por día de la semana)
boxplot(est0,est1,est2,est3,est4,est5,est6,ylim=c(0,25), col="skyblue",cex.names=4,
        names=c("domingo","lunes","martes","miércoles","jueves","viernes","sábado"))

##### FUNCIÓN DE INTENSIDAD ACUMULADA (FIA)

## FIA total
k_tot=805
est_total=NULL
for (i in 2:145){
  j<- i-1
  est_total[j]<- sum(bbtot[1:j-1])/k_tot + (bbtot[i-1]*(i-(i-1)))/k_tot
} #btot<- by(b309_tot$Cant.Retiros, b309_tot$Hr_Min, sum) #suma retiros c/10 min
#bbtot=c(b1tot,b3tot,b5tot,b2tot,b4tot)

plot(est_total, xlab="t (intervalos de 10 minutos)", ylab="Cant. Extracciones",
     type="l", cex.lab=1.4, lwd=2)

## FIA domingos
b309_sun$Minutos<- b309_sun$Fecha[[2]]
b309_sun$Hr_Min<- paste0(b309_sun$Horas,":",b309_sun$Minutos)
bsun<- by(b309_sun$Cant.Retiros, b309_sun$Hr_Min, sum)
b1su=bsun[1:12];b2su=bsun[13:72];b3su=bsun[73:78];b4su=bsun[79:102];b5su=bsun[103:144]
bbsun=c(b1su,b3su,b5su,b2su,b4su)

k=115 #(son 115 domingos)
est_sun=NULL
for (i in 2:145){
  j<- i-1
  est_sun[j]<- sum(bbsun[1:j-1])/k + (bbsun[i-1]*(i-(i-1)))/k
}

plot(est_sun, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", type="l",
     cex.lab=1.4, lwd=2)

## FIA lunes
b309_mon$Minutos<- b309_mon$Fecha[[2]]
b309_mon$Hr_Min<- paste0(b309_mon$Horas,":",b309_mon$Minutos)
bmon<- by(b309_mon$Cant.Retiros, b309_mon$Hr_Min, sum)
b1mo=bmon[1:12];b2mo=bmon[13:72];b3mo=bmon[73:78];b4mo=bmon[79:102];b5mo=bmon[103:144]
bbmon=c(b1mo,b3mo,b5mo,b2mo,b4mo)

```

```

est_mon=NULL
for (i in 2:145){
  j<- i-1
  est_mon[j]<- sum(bbmon[1:j-1])/k + (bbmon[i-1]*(i-(i-1)))/k
}

plot(est_mon, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", type="l",
      cex.lab=1.4, lwd=2)

## FIA martes
b309_tue$Minutos<- b309_tue$Fecha[[2]]
b309_tue$Hr_Min<- paste0(b309_tue$Horas,":",b309_tue$Minutos)
btue<- by(b309_tue$Cant.Retiros, b309_tue$Hr_Min, sum)
b1tu=btue[1:12];b2tu=btue[13:72];b3tu=btue[73:78];b4tu=btue[79:102];b5tu=btue[103:144]
bbtue=c(b1tu,b3tu,b5tu,b2tu,b4tu)

est_tue=NULL
for (i in 2:145){
  j<- i-1
  est_tue[j]<- sum(bbtue[1:j-1])/k + (bbtue[i-1]*(i-(i-1)))/k
}

plot(est_tue, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", type="l",
      cex.lab=1.4, lwd=2)

## FIA miércoles
b309_wed$Minutos<- b309_wed$Fecha[[2]]
b309_wed$Hr_Min<- paste0(b309_wed$Horas,":",b309_wed$Minutos)
bwed<- by(b309_wed$Cant.Retiros, b309_wed$Hr_Min, sum)
b1w=bwed[1:12];b2w=bwed[13:72];b3w=bwed[73:78];b4w=bwed[79:102];b5w=bwed[103:144]
bb_wed=c(b1w,b3w,b5w,b2w,b4w)

est_wed=NULL
for (i in 2:145){
  j<- i-1
  est_wed[j]<- sum(bb_wed[1:j-1])/k + (bb_wed[i-1]*(i-(i-1)))/k
}

plot(est_wed, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", type="l",
      cex.lab=1.4, lwd=2)

## FIA jueves
b309_thu$Minutos<- b309_thu$Fecha[[2]]
b309_thu$Hr_Min<- paste0(b309_thu$Horas,":",b309_thu$Minutos)
bthu<- by(b309_thu$Cant.Retiros, b309_thu$Hr_Min, sum)
b1th=bthu[1:12];b2th=bthu[13:72];b3th=bthu[73:78];b4th=bthu[79:102];b5th=bthu[103:144]
bb_thu=c(b1th,b3th,b5th,b2th,b4th)

est_thu=NULL
for (i in 2:145){
  j<- i-1
  est_thu[j]<- sum(bb_thu[1:j-1])/k + (bb_thu[i-1]*(i-(i-1)))/k
}

plot(est_thu, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", type="l",
      cex.lab=1.4, lwd=2)

```

```

## FIA viernes
b309_fri$Minutos<- b309_fri$Fecha[[2]]
b309_fri$Hr_Min<- paste0(b309_fri$Horas,":",b309_fri$Minutos)
bfri<- by(b309_fri$Cant.Retiros, b309_fri$Hr_Min, sum)
b1fr=bfri[1:12];b2fr=bfri[13:72];b3fr=bfri[73:78];b4fr=bfri[79:102];b5fr=bfri[103:144]
bb_fri=c(b1fr,b3fr,b5fr,b2fr,b4fr)

est_fri=NULL
for (i in 2:145){
  j<- i-1
  est_fri[j]<- sum(bb_fri[1:j-1])/k + (bb_fri[i-1]*(i-(i-1)))/k
}

plot(est_fri, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", type="l",
      cex.lab=1.4, lwd=2)

## FIA sábados
b309_sat$Minutos<- b309_sat$Fecha[[2]]
b309_sat$Hr_Min<- paste0(b309_sat$Horas,":",b309_sat$Minutos)
bsat<- by(b309_sat$Cant.Retiros, b309_sat$Hr_Min, sum)
b1sa=bsat[1:12];b2sa=bsat[13:72];b3sa=bsat[73:78];b4sa=bsat[79:102];b5sa=bsat[103:144]
bb_sat=c(b1sa,b3sa,b5sa,b2sa,b4sa)

est_sat=NULL
for (i in 2:145){
  j<- i-1
  est_sat[j]<- sum(bb_sat[1:j-1])/k + (bb_sat[i-1]*(i-(i-1)))/k
}

plot(est_sat, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", type="l",
      cex.lab=1.4, lwd=2)

## gráfico FIAs por día
plot(est_total, type="l", col="black", lwd=2, lty=2, xlab="t (intervalos de 10 minutos)",
      ylim=c(0,230), ylab="Cant. Extracciones estimada", cex.lab=1.4)
lines(est_sun,col="red", lwd=2)
lines(est_mon,col="pink", lwd=2)
lines(est_tue,col="blue", lwd=2)
lines(est_wed,col="orange",lwd=2)
lines(est_thu,col="green", lwd=2)
lines(est_fri,col="grey", lwd=2)
lines(est_sat,col="purple", lwd=2)
legend(1,190, bty="n", c("domingo","lunes", "martes", "miercoles", "jueves", "viernes", "sabado")
      ,cex=1.2,col=c("red","pink","blue","orange", "green", "grey", "purple"), lty=1, lwd=2)
legend(1,220, bty="n", c("total") ,cex=1.2,col=c("black"), lty=2, lwd=2)

##### INTERVALOS DE CONFIANZA

## total
plot(est_total, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", lwd=2
      , type="l", cex.lab=1.4)

inf_to<- est_total-qnorm(0.95)*(sqrt(est_total/115)) #90% de confianza
lines(inf_to, col="red", lty=2)
sup_to<- est_total+qnorm(0.95)*(sqrt(est_total/115))
lines(sup_to, col="red", lty=2)

inf_to2<- est_total-qnorm(0.995)*(sqrt(est_total/115)) #99% de confianza

```

```

lines(inf_to2, col="blue", lty=2)
sup_to2<- est_total+qnorm(0.995)*(sqrt(est_total/115))
lines(sup_to2, col="blue", lty=2)

legend(1,140, bty="n", c("nivel confianza 90%","nivel confianza 99%")
      ,cex=1.2,col=c("red","blue"), lty=2, lwd=1)
legend(1,164, bty="n", c("FIA") ,cex=1.2,col=c("black"), lty=1, lwd=2)

## domingo
plot(est_sun, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", lwd=2,
     type="l", cex.lab=1.4)

inf_su<- est_sun-qnorm(0.95)*(sqrt(est_sun/115)) #90% de confianza
lines(inf_su, col="red", lty=2)
sup_su<- est_sun+qnorm(0.95)*(sqrt(est_sun/115))
lines(sup_su, col="red", lty=2)

inf_su2<- est_sun-qnorm(0.995)*(sqrt(est_sun/115)) #99% de confianza
lines(inf_su2, col="blue", lty=2)
sup_su2<- est_sun+qnorm(0.995)*(sqrt(est_sun/115))
lines(sup_su2, col="blue", lty=2)

legend(1,85, c("nivel confianza 90%","nivel confianza 99%")
      ,cex=1.2,col=c("red","blue"), lty=2, lwd=1, bty="n")
legend(1,100, c("FIA domingo") ,cex=1.2,col=c("black"), lty=1, lwd=2, bty="n")

## lunes
plot(est_mon, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", lwd=2,
     type="l", cex.lab=1.4)

inf_mo<- est_mon-qnorm(0.95)*(sqrt(est_mon/115)) #90% de confianza
lines(inf_mo, col="red", lty=2)
sup_mo<- est_mon+qnorm(0.95)*(sqrt(est_mon/115))
lines(sup_mo, col="red", lty=2)

inf_mo2<- est_mon-qnorm(0.995)*(sqrt(est_mon/115)) #99% de confianza
lines(inf_mo2, col="blue", lty=2)
sup_mo2<- est_mon+qnorm(0.995)*(sqrt(est_mon/115))
lines(sup_mo2, col="blue", lty=2)

legend(1,145, c("nivel confianza 90%","nivel confianza 99%")
      ,cex=1.2,col=c("red","blue"), lty=2, lwd=1, bty="n")
legend(1,170, c("FIA lunes") ,cex=1.2,col=c("black"), lty=1, lwd=2, bty="n")

## martes
plot(est_tue, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", lwd=2,
     type="l", cex.lab=1.4)

inf_tu<- est_tue-qnorm(0.95)*(sqrt(est_tue/115))
lines(inf_tu, col="red", lty=2)
sup_tu<- est_tue+qnorm(0.95)*(sqrt(est_tue/115))
lines(sup_tu, col="red", lty=2)

inf_tu2<- est_tue-qnorm(0.995)*(sqrt(est_tue/115))
lines(inf_tu2, col="blue", lty=2)
sup_tu2<- est_tue+qnorm(0.995)*(sqrt(est_tue/115))
lines(sup_tu2, col="blue", lty=2)

```

```

legend(1,170, c("nivel confianza 90% ", "nivel confianza 99% ")
      ,cex=1.2,col=c("red","blue"), lty=2, lwd=1, bty="n")
legend(1,195, c("FIA martes") ,cex=1.2,col=c("black"), lty=1, lwd=2, bty="n")

## miercoles
plot(est_wed, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", lwd=2,
     cex.lab=1.4,type="l")

inf_we<- est_wed-qnorm(0.95)*(sqrt(est_wed/115))
lines(inf_we, col="red", lty=2)
sup_we<- est_wed+qnorm(0.95)*(sqrt(est_wed/115))
lines(sup_we, col="red", lty=2)

inf_we2<- est_wed-qnorm(0.995)*(sqrt(est_wed/115))
lines(inf_we2, col="blue", lty=2)
sup_we2<- est_wed+qnorm(0.995)*(sqrt(est_wed/115))
lines(sup_we2, col="blue", lty=2)

legend(1,170, c("nivel confianza 90%","nivel confianza 99%")
      ,cex=1.2,bty="n",col=c("red","blue"), lty=2, lwd=1)
legend(1,195, c("FIA miércoles") ,cex=1.2,bty="n",col=c("black"), lty=1, lwd=2)

## jueves
plot(est_thu, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", lwd=2,
     cex.lab=1.4, type="l")

inf_th<- est_thu-qnorm(0.95)*(sqrt(est_thu/115))
lines(inf_th, col="red", lty=2)
sup_th<- est_thu+qnorm(0.95)*(sqrt(est_thu/115))
lines(sup_th, col="red", lty=2)

inf_th2<- est_thu-qnorm(0.995)*(sqrt(est_thu/115))
lines(inf_th2, col="blue", lty=2)
sup_th2<- est_thu+qnorm(0.995)*(sqrt(est_thu/115))
lines(sup_th2, col="blue", lty=2)

legend(1,170, c("nivel confianza 90%","nivel confianza 99%")
      ,cex=1.2,col=c("red","blue"), lty=2, lwd=1, bty="n")
legend(1,195, c("FIA jueves") ,bty="n",cex=1.2,col=c("black"), lty=1, lwd=2)

## miércoles y jueves conjunta:juev cae dentro de la conf del mierc?
plot(est_wed, xlab="t (intervalos de a 10 minutos)", ylab="Extracciones", lwd=2, cex.main=0.9,
     type="l", cex.lab=1.4, col="red")
lines(inf_we2, col="black", lty=2)
lines(sup_we2, col="black", lty=2)
lines(est_thu, col="blue", lwd=2)
legend(1,190, bty="n", c("Miércoles","Jueves") ,cex=1.2,col=c("red", "blue"), lty=1, lwd=2)
legend(1,150, bty="n", c("Intervalo al 99% de confianza para el miércoles")
      ,cex=1.2,col=c("black"), lty=2, lwd=2)

## viernes
plot(est_fri, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", lwd=2,
     cex.lab=1.4,type="l")

inf_fr<- est_fri-qnorm(0.95)*(sqrt(est_fri/115))
lines(inf_fr, col="red", lty=2)
sup_fr<- est_fri+qnorm(0.95)*(sqrt(est_fri/115))
lines(sup_fr, col="red", lty=2)

```

```

inf_fr2<- est_fri-qnorm(0.995)*(sqrt(est_fri/115))
lines(inf_fr2, col="blue", lty=2)
sup_fr2<- est_fri+qnorm(0.995)*(sqrt(est_fri/115))
lines(sup_fr2, col="blue", lty=2)

legend(1,170, c("nivel confianza 90%", "nivel confianza 99%")
      ,cex=1.2,col=c("red", "blue"), lty=2, lwd=1, bty="n")
legend(1,200, c("FIA viernes") ,cex=1.2,col=c("black"), lty=1, lwd=2, bty="n")

## sábado
plot(est_sat, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", lwd=2,
      cex.lab=1.4, type="l")

inf_sa<- est_sat-qnorm(0.95)*(sqrt(est_sat/115))
lines(inf_sa, col="red", lty=2)
sup_sa<- est_sat+qnorm(0.95)*(sqrt(est_sat/115))
lines(sup_sa, col="red", lty=2)

inf_sa2<- est_sat-qnorm(0.995)*(sqrt(est_sat/115))
lines(inf_sa2, col="blue", lty=2)
sup_sa2<- est_sat+qnorm(0.995)*(sqrt(est_sat/115))
lines(sup_sa2, col="blue", lty=2)

legend(1,140, c("nivel confianza 90% ", "nivel confianza 99%")
      ,cex=1.2,col=c("red", "blue"), lty=2, lwd=1, bty="n")
legend(1,165, c("FIA sábado") ,cex=1.2,col=c("black"), lty=1, lwd=2, bty="n")

## cálculo FIA conjunta miércoles y jueves
b309_wt<- rbind(b309_wed,b309_thu) # sacamos del dataframe los días elegidos (33120*8)
b309_wt$Minutos<- b309_wt$Fecha[[2]]
b309_wt$Hr_Min<- paste0(b309_wt$Horas,":",b309_wt$Minutos)
bwt<- by(b309_wt$Cant.Retiros, b309_wt$Hr_Min, sum) #se suman las extracciones por 10 min
b1wt=bwt[1:12];b2wt=bwt[13:72];b3wt=bwt[73:78];b4wt=bwt[79:102];b5wt=bwt[103:144]
bb_wt=c(b1wt,b3wt,b5wt,b2wt,b4wt)

est_wt=NULL # k realizaciones de los mierc+juev (115*2)
for (i in 2:145){
  j<- i-1
  est_wt[j]<- sum(bb_wt[1:j-1])/(k*2) + (bb_wt[i-1]*(i-(i-1)))/(k*2)
}

## gráfica de la FIA conjunta vs FIA domingos e intervalos de confianza
plot(est_sun, xlab="t (intervalos de 10 minutos)", ylab="Extracciones", cex.lab=1.4,
      lwd=2, ylim=c(0,200), col="red", type="l")
lines(inf_su2, col="black", lty=2) #99% de confianza
lines(sup_su2, col="black", lty=2)

lines(est_wt, lwd=2, col="blue")
inf_wt2<- est_wt-qnorm(0.995)*(sqrt(est_wt/230)) #99% de confianza
lines(inf_wt2, col="black", lty=2)
sup_wt2<- est_wt+qnorm(0.995)*(sqrt(est_wt/230))
lines(sup_wt2, col="black", lty=2)
legend(1,180, bty="n", c("domingo", "miércoles y jueves") ,cex=1.2,
      col=c("red", "blue"), lty=1, lwd=2)
legend(1,140, bty="n", c("nivel confianza 99%"),cex=1.2,col=c("black"), lty=2, lwd=1)

## gráfico (datos reales vs FIA total): se puede dividir por quincena?

```

```

plot(est_total, type="l", xlab="t (intervalos de 10 minutos)", ylab="Cant. Extracciones",
     ylim=c(0,280), lwd=2, cex.lab=1.4)
m1<-cumsum(subset(b309_wed$Cant.Retiros,b309_wed$dia=="2010-09-01"))
m2<-cumsum(subset(b309_thu$Cant.Retiros,b309_thu$dia=="2010-09-02"))
m3<-cumsum(subset(b309_wed$Cant.Retiros,b309_wed$dia=="2010-09-08"))
m4<-cumsum(subset(b309_thu$Cant.Retiros,b309_thu$dia=="2010-09-09"))
m5<-cumsum(subset(b309_wed$Cant.Retiros,b309_wed$dia=="2010-09-15"))
m6<-cumsum(subset(b309_thu$Cant.Retiros,b309_thu$dia=="2010-09-16"))
m7<-cumsum(subset(b309_wed$Cant.Retiros,b309_wed$dia=="2010-09-22"))
m8<-cumsum(subset(b309_thu$Cant.Retiros,b309_thu$dia=="2010-09-23"))

lines(m1, col="red")
lines(m2, col="blue")
lines(m3, col="orange" )
lines(m4, col="dark grey")
lines(m5, col="violet")
lines(m6, col="green")
lines(m7, col="skyblue")
lines(m8, col="pink")
legend(1,290, c("FIA"),cex=1,col="black", lty=1, lwd=2, bty="n")
legend(1,270, c("2010-09-01", "2010-09-02", "2010-09-08", "2010-09-09", "2010-09-15",
              "2010-09-16", "2010-09-22", "2010-09-23"),cex=1,
      col=c("red", "blue", "orange", "dark grey", "violet", "green", "skyblue", "pink"),
      lty=1, bty="n", lwd=2)

library(lubridate)
h1=rep(1:144,230)
b = ymd(b309_wed$dia) #dias del mes que cayeron mierc
c = ymd(b309_thu$dia) #dias del mes que cayeron jueves
db = day(b)
dc = day(c)
dd = c(db,dc)
cant.Retiros_tot= c(b309_wed$Cant.Retiros, b309_thu$Cant.Retiros) #retiros días elegidos
ext_dia = cbind(dd,h1,cant.Retiros_tot)
quin1 = subset(ext_dia,ext_dia[,1]<=15) #quincenas (116 días)
quin2 = subset(ext_dia,ext_dia[,1] > 15) #(114 días)
ext_hora = by(quin1[,3], quin1[,2], sum)/116 # cant mierc y juev 1era quincena
ext_hora2 = by(quin2[,3], quin2[,2], sum)/114 # cant mierc y juev 2da quincena

plot(est_wt, type="l", lwd=3, xlab="t (intervalos de 10 minutos)", ylab="lambda(t)",
     ylim=c(0,240), cex.lab=1.4)
lines(cumsum(ext_hora), col="red", lwd=2)
lines(cumsum(ext_hora2), col="blue", lwd=2)
legend(1,240, c("FIA estimada","FIA estimada 1ra quincena", "FIA estimada 2da quincena"),
      cex=1.2,bty="n", col=c("black","red","blue"), lty=1, lwd=2)

## probabilidad de inactividad del cajero por hora para miércoles y jueves
Horas<- c("00:00_01:00","01:00_02:00", "02:00_03:00", "03:00_04:00", "04:00_05:00",
          "05:00_06:00", "06:00_07:00", "07:00_08:00", "08:00_09:00", "09:00_10:00",
          "10:00_11:00", "11:00_12:00", "12:00_13:00", "13:00_14:00", "14:00_15:00",
          "15:00_16:00", "16:00_17:00", "17:00_18:00", "18:00_19:00", "19:00_20:00",
          "20:00_21:00", "21:00_22:00", "22:00_23:00", "23:00_24:00")

Prob_Inact_hr <- 0
i=1
j=1
while (i < 140){
  Prob_Inact_hr[j] <- exp(est_wt[i]-est_wt[i+5])

```

```

    i=i+6
    j=j+1
}
Prob_Inact_Hora<- data.frame(cbind(Horas,Prob_Inact_hr))
plot(Prob_Inact_hr, type="o", pch=16, col="blue", lty=2, cex.lab=1.4,
     xlab="t(horas)", ylab="probabilidad inactividad")

## Probabilidad de inactividad por hora para cada tramo de diez minutos
Prob_Inact_hr_10min <- 0
i=1
j=1
while (i < 140){
  Prob_Inact_hr_10min[j] <- exp(est_wt[i]-est_wt[i+5])
  i=i+1
  j=j+1
}

plot(Prob_Inact_hr_10min, col="blue", pch=16, type="o", lty=2,
     ylab="Probabilidad estimada", xlab="Intervalos de 10 minutos", cex.lab=1.4)

## Probabilidad de inactividad por media hora para cada tramo de diez minutos
Prob_Inact_media_hr <- 0
i=1
j=1
while (i < 143){
  Prob_Inact_media_hr[j] <- exp(est_wt[i]-est_wt[i+2])
  i=i+1
  j=j+1
}

plot(Prob_Inact_media_hr, col="blue", pch=16, type="o", lty=2, cex.lab=1.4,
     ylab="Probabilidad estimada", xlab="Intervalos de 10 minutos")

plot(Prob_Inact_hr_10min, col="blue", pch=16, type="o", lty=2, cex.lab=1.4,
     ylab="Probabilidad estimada", xlab="t (intervalos de a 10 minutos)")
lines(Prob_Inact_media_hr, col="red", pch=16, lty=2, type="o")
legend(0,0.5, bty="n", c("Intervalo hora", "Intervalo media hora")
      ,cex=1,col=c("blue", "red"), lty=2, pch=16)

### COMPARACION CON OTROS CAJEROS

## gráfica conjunta FIs de diferentes cajeros
plot(est_tot, col="black", pch=16, type="o", lty=2, xlab="t(horas del día)",
     ylab="Cant. Extracciones", cex.lab=1.4, xaxp=c(0,24,12)) #b309
lines(est_tot2, col="red", pch=16, type="o", lty=2) #b397
lines(est_tot3, col="green", pch=16, type="o", lty=2) #b435
lines(est_tot5, col="orange", pch=16, type="o", lty=2) #b834
lines(est_tot7, col="purple", pch=16, type="o", lty=2) #m389
lines(est_tot8, col="skyblue", pch=16, type="o", lty=2) #m404
lines(est_tot9, col="blue", pch=16, type="o", lty=2) #m437
legend(1,20, bty="n", c("b309","b397", "b435", "b834", "m389","m404", "m437"),
     col=c("black", "red", "green", "orange", "purple", "skyblue", "blue"),
     lty=2, pch=16,lwd=2)

## gráfica conjunta FIAs de diferentes cajeros
plot(est_total, col="black", type="l", xlab="t(intervalos de 10 minutos)",
     ylab="Cant. Extracciones Acumuladas",cex.lab=1.4, lwd=2, xaxp=c(0,144,12)) #b309
lines(est_total2, col="red", lwd=2) #b397

```

```
lines(est_total3, col="green", lwd=2) #b435
lines(est_total5, col="orange", lwd=2) #b834
lines(est_total7, col="purple", lwd=2) #m389
lines(est_total8, col="skyblue", lwd=2) #m404
lines(est_total9, col="blue", lwd=2) #m437
legend(1,180, bty="n", c("b309", "b397", "b435", "b834","m389","m404", "m437"),
      col=c("black", "red", "green", "orange", "purple", "skyblue", "blue"),lty=1, lwd=2)

## tabla de inactividad de los diferentes cajeros
Tabla<- data.frame(cbind(Horas,Prob_Inact_b397, Prob_Inact_b435, Prob_Inact_b834,
                        Prob_Inact_m389, Prob_Inact_m404, Prob_Inact_m437))
```