



DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA
DEPARTAMENTO DE MATEMÁTICAS

**Diseño e implementación de un simulador
dinámico del Sistema Metro de
la Ciudad de México**

Tesis que presenta:

María Naturaleza Isaura Cossío Vital

Para obtener el grado de:

Maestra en Ciencias

(Matemáticas Aplicadas e Industriales)

Dirigida por:

Dra. Blanca Rosa Pérez Salvador

Dra. Patricia Saavedra Barrera

Jurado calificador:

Presidente:	Dr. Francisco S. Salem Silva	UV
Secretaria:	Dra. Patricia Saavedra Barrera	UAM-I
Vocal:	Dr. Joaquín Delgado Fernández	UAM-I
Vocal:	Dra. Blanca Rosa Pérez Salvador	UAM-I

*A la memoria de Rufina Santos Gallegos[†]:
quien ablandó mi corazón; forjando mi carácter.*

Por tantos y bellos momentos inolvidables...

¡Gracias viejita!

Agradecimientos

Con profundo agradecimiento a la sociedad mexicana, que con su esfuerzo hace posible la existencia y el desarrollo de la universidad pública. Agradezco a la Universidad Autónoma Metropolitana Unidad Iztapalapa, que me abrió sus puertas ofreciéndome la oportunidad de desarrollarme y estudiar, en un marco de prestigio, calidad y confianza.

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACyT) por brindarme la oportunidad de participar, como becaria, en el Programa Nacional de Posgrados de Calidad (PNPC) durante la realización de mi posgrado.

Así mismo, quiero expresar mi agradecimiento a mis directoras de tesis: agradezco a la Dra. Patricia Saavedra Barrera, por la confianza que depositó en mi persona, por su paciencia y presión ante mi inconsistencia y por todo su apoyo a lo largo de la realización de mi posgrado; agradezco a la Dra. Blanca Rosa Pérez Salvador, por su comprensión a mi persona y por todas sus acertadas sugerencias en esta tesis.

Por su tiempo, pese a su apretada agenda, y por el apoyo que me brindó durante mi posgrado, agradezco a mi profesor y sinodal el Dr. Joaquín Delgado Fernández quien con su experiencia enriqueció notablemente este trabajo. También, agradezco a mi sinodal el Dr. Francisco S. Salem Silva por sus atinadas observaciones en esta tesis.

Agradezco a la señora Nora Hilda Vital Santos: mi madre, cuya vida y esfuerzo es una lección ejemplar inigualable; le agradezco por su incondicional apoyo y cariño que siempre me ha concedido.

Igualmente, agradezco a José María González Lara, por el cariño que me ha brindado y por todo su apoyo y motivación que me ha conferido a lo largo de estos últimos cinco años.

Finalmente, quiero agradecer a todos mis amigos, compañeros y profesores que aquí conocí que, en una ciudad extraña para mí, me acogieron y me hicieron sentir como en casa.

*“... toda persona debe tener en la vida:
un oficio, una pasión y una causa”.*

José María González Lara

*“Ask, and you will receive;
search, and you will find;
knock, and the door will be opened for you”.*

Mateo 7:7

Resumen

En la primera parte de este trabajo presentamos un modelo del comportamiento dinámico de un sistema de transporte urbano tipo Metro, que incluye una función de costo que toma en cuenta el costo operativo más el costo social: el tiempo total de espera en el andén de los pasajeros que ingresan al sistema o transbordan de otras líneas. Para la construcción de éste nos basamos en un modelo de simulación de eventos discretos el cual considera el flujo de trenes y de pasajeros y las interacciones entre líneas a través de las estaciones de correspondencia. En este trabajo aplicamos el modelo para simular el comportamiento de las líneas 1, 2 y 3 del Sistema Metro de la Ciudad de México y así verificar la precisión y eficiencia del modelo propuesto y, en consecuencia, la validación del mismo.

La función de costo permite evaluar el desempeño del sistema, la cual nos permite ir un poco más allá de una simple simulación y pensar en la optimización de la misma. Esto último nos lleva a introducirnos en la optimización estocástica de sistemas de eventos discretos a través de simulación, en donde el principal problema es la estimación del gradiente de una función de desempeño, cuyo valor puede ser obtenido a través de simulación. Por tanto, en la segunda parte de este trabajo nos enfocamos en el estudio de métodos de aproximación estocástica al gradiente de este tipo de funciones: el método de diferencias finitas, el método de perturbación simultánea, de análisis de perturbación infinitesimal (IPA) y el método de razón de verosimilitud (también conocido como el método de la función *score*). Estos estimadores pueden ser posteriormente utilizados en el algoritmo de aproximación estocástica que es; básicamente, la versión estocástica del algoritmo del gradiente proyectado para aproximar la solución de problemas de optimización.

Para ilustrar la precisión de estos métodos consideramos el tiempo medio que un cliente pasa en una cola tipo M/U/1, cuya solución analítica es conocida gracias a la fórmula de Pollaczek-Khinchine; lo cual nos permite comparar los resultados obtenidos a través de simulación. Para el Sistema Metro de la Ciudad de México, el tiempo de espera de los pasajeros que transbordan es calculado a través de simulación y es aquí donde hacemos uso de los métodos de aproximación estocástica para estimar las sensibilidades del tiempo de espera respecto a cada una de las variables de control del sistema.

Contenido

Resumen	IX
Introducción	XIII
Notación	XVII
1. Modelación de un sistema Metro	1
1.1. Flujo de trenes	1
1.2. Flujo de pasajeros	5
1.2.1. Pasajeros en estaciones de correspondencia	7
1.3. Pasajeros en el tren	13
1.3.1. Pasajeros que abordan el tren	14
1.3.2. Pasajeros que abandonan el tren	14
1.4. Función de desempeño	17
1.4.1. Tiempo de espera	18
2. Optimización de la función de desempeño	21
2.1. Método Monte-Carlo	22
2.1.1. Sistematización	22
2.1.2. Algoritmo	23
2.2. Optimización de sistemas de eventos discretos	23
2.3. Números aleatorios comunes	24
2.4. Estimación de sensibilidades	26
2.4.1. Diferencias finitas	26
2.4.2. Perturbación simultánea	28
2.4.3. Análisis de perturbación infinitesimal	31
2.4.4. Método de razón de verosimilitud	34
2.5. Optimización estocástica: algoritmo de aproximación estocástica	37
2.6. Aplicación a un sistema de colas del tipo M/U/1	37

3. Simulación y sensibilidades del Sistema Metro de la Ciudad de México	45
3.1. Matriz origen-destino	45
3.2. Simulación global	48
3.3. Datos del Sistema Metro de la Ciudad de México	48
3.4. Simulación parcial vs. simulación global: resultados comparativos	50
3.4.1. Resultados numéricos simulación Monte-Carlo	53
3.5. Resultados numéricos del tiempo de espera	53
3.5.1. Variación de la longitud de la jornada laboral	54
3.5.2. Variación de la varianza de las variables de fluctuación	54
3.6. Sensibilidades del tiempo de espera	55
3.6.1. Estimadores de FD y SP con tiempos de salida de trenes deterministas	56
3.6.2. Estimadores de FD y SP con tiempos de salidas de trenes no deterministas	57
Conclusiones y perspectivas	59
Apéndice	63
A.1. Construcción de la matriz origen-destino	63
Anexos	69
Bibliografía	79

Introducción

El Sistema de Transporte Colectivo (STC) Metro de la Ciudad de México fue inaugurado el 4 de septiembre de 1969, con el objetivo de resolver los graves problemas de transporte público y congestión de la red vial; principalmente en la zona centro de la ciudad, donde en horas pico del tráfico, la velocidad de circulación era menor a la de una persona caminando.

A la fecha, el Sistema cuenta con 11 líneas (y una más en construcción) a lo largo de las áreas conurbadas del Distrito Federal. Una línea, esquemáticamente, consiste de dos vías: en una vía, el tren corre en cierto sentido y en la otra vía, el tren corre en sentido contrario; por tanto, el origen de una vía es el destino de la otra y viceversa. A su vez, cada línea está constituida por tres tipos de estaciones: las estaciones normales, en las que los pasajeros (usuarios del Sistema Metro), sólo pueden ingresar al sistema o abandonarlo; las estaciones terminales, donde una línea inicia y/o termina su trayecto y las estaciones de correspondencia, que permiten transbordar de una línea a otra sin salir del sistema.

Cabe señalar que el número de estaciones por línea es variable. Por ejemplo: la línea 2 cuenta con 24 estaciones, mientras que la línea A sólo cuenta con 10. Las estaciones de correspondencia son potenciales centros de congestión peatonal, ya que éstas unen de dos a cuatro líneas. Por ejemplo: la estación de correspondencia “La Raza”, une a las líneas 3 y 5; la estación de correspondencia “Chabacano”, une a las líneas 2, 8 y 9; mientras que la estación de correspondencia (y a la vez, estación terminal) “Pantitlán”, une a cuatro líneas: las líneas 1, 5, 9 y A (para una mejor comprensión visual, véase la figura A.2 en la página 70).

Actualmente el Sistema cuenta con 291 trenes con 9 vagones cada uno (con una capacidad de transportar a la vez a 1,530 pasajeros: 360 sentados y 1,170 parados) y 31 trenes con 6 vagones cada uno (cuya capacidad es de 1,020 pasajeros: 240 sentados y 780 parados), formando así, un parque vehicular de 355 trenes en total. Cada línea tiene asignado cierto número de trenes para su funcionamiento:

Línea	1	2	3	4	5	6	7	8	9	A	B
No. de trenes asignados	51	45	58	13	17	18	26	29	33	33	32

Los pasajeros pueden ingresar (por un costo individual de \$ 3.00 pesos¹ en el horario de servicio establecido) o abandonar el sistema en cualquiera de las estaciones.

Justificación.

El control predictivo² es una metodología que se ha utilizado con mucho éxito para modelar los sistemas ferroviario que se rigen a través de un estricto horario en el que los trenes y sus conexiones están sincronizados. En estos tipos de sistemas se asume que los pasajeros llegan al andén con el objetivo común de abordar el tren que sale a determinada hora; cada pasajero tiene asignado un lugar y no hay problemas de sobrecupo. En este caso lo más importante es reducir la diferencia entre el horario asignado de salida y el tiempo real de salida (para más detalles véase Assis & Milani [2004]).

Para los sistemas Metro el contexto es distinto puesto que la frecuencia de servicio de los trenes es mayor y sus conexiones no están sincronizadas, así, los pasajeros no arriban a las estaciones con la intención de abordar un determinado tren, sino que éstos arriban con la intención de abordar el primer tren que llegue y lo hacen de forma independiente. Por ello, es natural considerar que el número de pasajeros que han llegado a un andén a determinada hora, es una variable aleatoria. Asimismo, el tiempo de salida de un tren de una estación es una variable aleatoria, puesto que éste depende del tiempo que transcurre entre la apertura y el cierre de puertas, en el cual los pasajeros descienden o abordan el tren, además de ciertas incidencias que pudieran presentarse.

En la literatura podemos encontrar numerosas investigaciones centradas en todo lo relativo a los movimientos de los trenes y algunas presentan modelos para la simulación de éstos. Sin embargo, la mayoría han dejado de lado la situación de los pasajeros. Lo que impide estudiar cuestiones que mejoren criterios que afecten a los pasajeros y no sólo al retraso de trenes. Una novedad, en este trabajo, respecto a otros modelos encontrados en la literatura es la incorporación de la información de los arribos de pasajeros y sus posteriores movimientos por la red.

Cabe mencionar que la idea de este trabajo surgió como consecuencia de un proyecto del Fondo Mixto ICyTDF-CONACyT, otorgado en el 2010 al Departamento de Matemáticas, de la Universidad Autónoma Metropolitana Unidad Iztapalapa, el cual fue coordinado por el Dr. Joaquín Delgado Fernández (docente e investigador de la Universidad). Este proyecto consiste en diseñar modelos matemáticos para mejorar el funcionamiento del STC Metro. La red de transporte presenta problemáticas complejas, que pueden solucionarse mediante modelos matemáticos basados en algoritmos que son traducidos a programas computacionales (para más detalles véase Boletín UAM [2012, febrero]).

El Metro de la Ciudad de México es uno de los sistemas más grandes del mundo en cuanto al número de pasajeros, como a la longitud del mismo y su funcionamiento es similar al Metro de Madrid, Montreal, París, entre

¹Salvo usuarios mayores de 60 años y usuarios con capacidades diferentes, quienes gozan de acceso gratuito.

²El control predictivo es una metodología que consiste en analizar un sistema como un sistema dinámico discreto que da lugar a un problema de optimización con restricciones y cuya solución permite llevar a cabo la simulación.

otros.

Desde hace algunos años, este sistema cuenta con un software llamado EMME3, que le permite determinar en cada línea, algunas variables importantes en la dinámica del sistema; como la ubicación de los trenes que circulan. Este programa tiene el inconveniente de que analiza línea por línea de forma independiente, sin tomar en cuenta las interacciones de las líneas a través de las estaciones de correspondencia, lo cual impide analizar cómo afecta al sistema en su función conjunta, algunas acciones como: el retiro de un tren de una línea para agregarlo a otra y el incrementar o reducir la frecuencia de salidas de trenes. Esto dificulta la toma de decisiones y en consecuencia la cantidad y calidad del servicio. Por tanto, la seguridad que el sistema ofrece también se ve afectada.

Para los usuarios de un sistema Metro, lo ideal sería que al llegar al andén el siguiente tren pase casi inmediatamente y con espacio suficientes para poder abordarlo; sin embargo, incrementar el número de trenes para satisfacer este propósito, aumentaría prohibitivamente el costo operacional del Metro; una solución intermedia entre los intereses de los usuarios y del sistema Metro, es administrar eficientemente el sistema utilizando la información de cada línea para tomar decisiones en tiempo real, que permitan sincronizar el flujo de pasajeros con el flujo de trenes de manera óptima. Dada la complejidad del fenómeno, no es posible realizar ensayos en el sistema en tiempo real, por lo que una posibilidad es disponer de un sistema computacional que permita simular el comportamiento del sistema, tomando en cuenta el flujo de trenes y de pasajeros y las interacciones a través de las estaciones de correspondencia.

Objetivos.

En el presente trabajo de tesis proponemos un modelo del comportamiento dinámico de un sistema de transporte urbano tipo Metro, basado en la simulación por eventos discretos, que considera tanto el flujo de trenes como el de pasajeros, así como las interacciones entre diferentes líneas a través de las estaciones de correspondencia. De esta manera, el modelo consiste en sólo analizar al sistema en un número finito de tiempos, esto es, cuando ocurren ciertos eventos de interés y que, por supuesto, se definen de antemano. Estos tiempos corresponden a los tiempos de salida de trenes de las estación, los cuales se modelan a través de procesos estocásticos. Para la construcción de este modelo nos basamos en el modelo propuesto en Vázquez-Abad & Zubieta (2005) para un sistema Metro ficticio. En nuestro caso lo aplicamos para simular el comportamiento de las líneas 1, 2 y 3 del Sistema Metro de la Ciudad de México; para esto se contó con datos reales para verificar la precisión y eficiencia del modelo presentado y, en consecuencia, la validación del mismo.

El modelo incluye una función de costo que permite evaluar el desempeño del sistema, la cual nos permite ir un poco más allá de una simple simulación y pensar en la optimización del mismo. Esto último nos lleva a introducirnos en la optimización estocástica de sistemas de eventos discretos a través de simulación, en donde el principal problema es la estimación del gradiente de una función de desempeño, cuyo valor se obtiene a través de simulación, debido a que su expresión analítica es inasequible o complicada de obtener.

Presentación de la tesis.

Este trabajo consta de 3 capítulos. En el capítulo 1 vemos la modelación de un sistema de transporte urbano tipo Metro; la cual comprende el flujo de trenes, el flujo de pasajeros y su interacción a través de las estaciones de correspondencia; además presentamos una función, llamada función de desempeño, que evalúa el desempeño global del sistema. En el capítulo 2 nos enfocamos en el estudio de métodos de aproximación estocástica, para la resolución de problemas de optimización de sistemas de eventos discretos a través de simulación, cuya teoría ilustramos con un ejemplo académico de un sistemas de colas del tipo M/U/1. En el capítulo 3 implementamos computacionalmente el modelo propuesto en el capítulo 1 para simular, bajo algunas restricciones, el comportamiento de las líneas 1, 2 y 3 del Sistema Metro de la Ciudad de México, calculamos las sensibilidades del desempeño de este sistema y presentamos algunos resultados concluyentes. Posteriormente presentamos una sección en la cual resumimos algunas conclusiones obtenidas a lo largo de la realización de este trabajo, así como las limitaciones del modelo propuesto y sus posibles mejoras.

Notación

γ^l	:	Hora de salida del primer tren de $a_1 \in L_l$ en s .
$\delta_j^l(a_q)$:	Variable de fluctuación del tiempo de salida del j -ésimo tren en $a_q \in L_l$.
$\lambda_s^l(a_q)$:	Intensidad de arribos de pasajeros, por ingreso al sistema, en $a_q \in L_l$ durante s .
$\lambda_s^{m,l}(a_o, a_d)$:	Intensidad de arribos de pasajeros en $a_o \in L_m$, con destino $a_d \in L_l$, durante s .
μ_s	:	Vector de control durante s , $\mu_s = (\mu_s^1, \mu_s^2, \dots, \mu_s^{NTL})$.
μ_{seg}	:	Tiempo mínimo permitido entre salidas consecutivas de trenes, $\mu_{seg} \leq \mu_s^l$.
μ_s^l	:	Variable de control (tiempo entre salidas consecutivas de trenes) de L_l durante s .
$\rho_s^{m,l}$:	Tasa de arribos de pasajeros a L_m que demandan transbordo de L_m a L_l en s .
$\rho_s^{n,m,l}$:	Tasa de arribos de pasajeros a L_n que demandan transbordo de L_n a L_m y de L_m a L_l en s .
τ_n	:	Tiempo de ocurrencia del n -ésimo arribo de pasajeros por ingreso al sistema.
a_1	:	Andén inicial.
a_d	:	Andén destino.
a_o	:	Andén origen.
a_{pl}	:	Andén de partida de L_l .
$a_q \in L_l$:	a_q pertenece a L_l .
a_{ql}	:	Andén de llegada de L_l .
a_{NTAl}	:	Andén final de L_l .
c	:	Costo monetario por cada unidad de $W_j^l(a_q)$.
$CO(\mu)$:	Costo operativo.
$CS(\mu)$:	Costo social.
$D_{a_o}(a_{q_m}, a_{p_l})$:	Conjunto de andenes destino a_d tales que, las trayectorias para ir de a_o a a_d , requieren un transbordo de $a_{q_m} \in L_m$ a $a_{p_l} \in L_l$ dado que arribaron en a_o , con $m \neq l$.
$d^l(a_{q-1}, a_q)$:	Distancia entre los andenes a_{q-1} y a_q de L_l .
\mathcal{EC}	:	Conjunto de andenes que pertenecen a estaciones de correspondencia.

$\mathcal{EC}(a_q)$: Conjunto de andenes que confluyen con $a_q \in L_l$ en una misma estación de correspondencia, exceptuando al mismo a_q y al andén que pertenece a la línea opuesta a L_l .
$F(\mu)$: Función de desempeño del Metro.
$I_{\{a_q \in \mathcal{EC}\}}$: Función indicadora de $a_q \in \mathcal{EC}$.
$\mathcal{I}(L_l, a_q)$: Conjunto de líneas que se intersectan con L_l en la estación que contiene a a_q .
J	: Jornada laboral.
k^l	: Costo monetario por cada tren en servicio en L_l .
L_l	: Línea l .
$NP_{a_q}^l(t)$: Proceso de arribos de pasajeros, en $a_q \in L_l$, al tiempo t .
$NP_{a_o, a_d}^{m, l}(t)$: Proceso de arribos de pasajeros, en $a_o \in L_m$ con destino $a_d \in L_l$, al tiempo t .
$NT_{a_q}^l(t)$: Proceso de conteo de trenes, que llegan a $a_q \in L_l$, al tiempo t .
NTA^l	: Número total de andenes de L_l .
NTL	: Número total de líneas del sistema Metro.
NTT_s^l	: Número total de viajes realizados por los trenes de L_l durante s .
$P^l(a_q)$: Tiempo de permanencia, de un tren, en $a_q \in L_l$.
$p^{l, m}(a_o, a_d)$: Probabilidad de que un pasajero que ingresa al sistema en $a_o \in L_l$ lo abandone en $a_d \in L_m$.
$P1_j^l(a_q)$: Número de pasajeros, de primer orden, que abordan el j -ésimo tren en $a_q \in L_l$.
$P2_j^l(a_q)$: Número de pasajeros, de segundo orden, que abordan el j -ésimo tren en $a_q \in L_l$.
$P1_j^{m, l}$: Número de pasajeros, de primer orden, que llegan en el j -ésimo tren de L_m y que transbordarán a L_l .
$PA_j^l(a_q)$: Número de pasajeros que abordarán el j -ésimo tren en $a_q \in L_l$.
$PAI_j^l(a_q)$: Número de pasajeros que abordan, por ingreso al sistema, el j -ésimo tren en $a_q \in L_l$.
$PAI_j^{m, l}$: Número de pasajeros, por ingreso al sistema, que llegan en el j -ésimo tren de L_m y que transbordarán a L_l .
$PAI_k^{m, l}(a_q)$: Número de pasajeros, por ingreso al sistema, que abordarán el k -ésimo tren en $a_q \in L_m$ para transbordar a L_l .
$PAT_j^l(a_q)$: Número de pasajeros que abordan, por transbordo, el j -ésimo tren en $a_q \in L_l$. $PAT_j^l(a_q) = P1_j^l(a_q) + P2_j^l(a_q) + \dots$
$PD_j^l(a_q)$: Número de pasajeros que abandonan el j -ésimo tren en $a_q \in L_l$. $PD_j^l(a_q) = PDA_j^l(a_q) + PDT_j^l(a_q)I_{\{a_q \in \mathcal{EC}\}}$.
$PD1_j^l(a_q)$: Pasajeros que abandonaron el j -ésimo tren, para abandonar el sistema, en $a_q \in L_l$ tales que requirieron transbordar una vez.
$PD2_j^l(a_q)$: Pasajeros que abandonaron el j -ésimo tren, para abandonar el sistema, en $a_q \in L_l$ tales que requirieron transbordar dos veces.
$PDA_j^l(a_q)$: Número de pasajeros que abandonan el j -ésimo tren, para abandonar el sistema, en $a_q \in L_l$. $PDA_j^l(a_q) = PDI_j^l(a_q) + PD1_j^l(a_q) + PD2_j^l(a_q) + \dots$
$PDI_j^l(a_q)$: Pasajeros que abandonan el j -ésimo tren, para abandonar el sistema, en $a_q \in L_l$ tales que no requirieron transbordar.

$PDT_j^l(a_q)$: Número de pasajeros que abandonan del j -ésimo tren, para transbordar, en $a_q \in L_l$.
$PS_j^l(a_q)$: Número de pasajeros en el j -ésimo tren al salir de $a_q \in L_l$.
$PT_k^{m,l}$: Pasajeros en el k -ésimo tren de L_m que transbordarán a L_l . $PT_k^{m,l} = PAI_k^{m,l} + P1_k^{m,l} + \dots$
$R^l(a_q)$: Tiempo de recorrido, de los trenes, entre los andenes consecutivos a_{q-1} y a_q de L_l .
s	: Segmento de J con $s = 1 : S$.
S	: Número de segmentos de J .
$S_j^l(a_q)$: Hora de llegada del j -ésimo tren a $a_q \in L_l$.
t_s	: Hora de finalización del segmento s , $t_s = \sum_{i=1}^s T_i$, con $t_0 = \min\{J\}$.
T_s	: Longitud del segmento s .
$T_j^l(a_q)$: Tiempo entre llegadas consecutivas de los trenes j y $j - 1$ en $a_q \in L_l$.
v^l	: Velocidad promedio con la que viajan los trenes de L_l .
$V_j^l(a_q)$: Hora de salida del j -ésimo tren de $a_q \in L_l$.
$W_j^l(a_q)$: Tiempo de espera asociado al j -ésimo tren en $a_q \in L_l$. $W_j^l(a_q) = WI_j^l(a_q) + WT_j^l(a_q)$.
$WI_j^l(a_q)$: Tiempo de espera, por ingreso al sistema, asociado al j -ésimo tren en $a_q \in L_l$.
$WT_j^l(a_q)$: Tiempo de espera, por transbordo, asociado al j -ésimo tren en $a_q \in L_l$.
$Y_j^l(a_q)$: Tiempo entre salidas consecutivas de los trenes j y $j - 1$ en $a_q \in L_l$.

Capítulo 1

Modelación de un sistema Metro

El Sistema Metro de la Ciudad de México consta de una red de líneas unidas por un conjunto de estaciones, llamadas estaciones de correspondencia. Cada línea está constituida por un conjunto de estaciones y por dos vías (una que corre en sentido opuesto a la otra¹). Cada estación tiene dos andenes, uno para cada vía, y hay tres tipos de estaciones: las estaciones terminales, donde hay un andén inicial y un andén final; las estaciones de correspondencia, donde se intersectan dos líneas o más² y las estaciones normales.

El modelo presentado a continuación es un modelo probabilista y dinámico; y éste está compuesto de dos partes: la primera que considera la modelación del flujo de trenes y la segunda que consta de la modelación del flujo de los pasajeros. Para poder observar el flujo resultante de este modelo conforme transcurre el tiempo, consideramos una simulación por eventos la cual consiste en analizar y/u observar el sistema en un número finito de tiempos; de esta manera mantenemos un seguimiento de las variables y/o eventos de interés: como el arribo de un pasajero a un andén; la llegada o salida de un tren; el número de pasajeros esperando al tren, en el andén; el número de pasajeros en el tren; la salida de un pasajero del sistema, entre otros eventos de interés.

Adicionalmente, en la tercera sección consideramos la modelación de pasajeros que se encuentran transportándose en el tren, cuya modelación incluye: a los pasajeros que abordan y a los pasajeros que abandonan el tren en cada andén. También se incluye en la cuarta sección, una función que nos permite medir el desempeño del sistema Metro; esta función de desempeño se compone de un costo operativo y de un costo social.

1.1. Flujo de trenes

Para la primera parte de la modelación describimos al sistema Metro mediante una gráfica dirigida, en donde cada línea se identifica por medio de una gráfica simple orientada, con un conjunto de nodos que representan las estaciones. Los andenes iniciales, que denotaremos por a_1 , son únicos para cada línea y se encuentran en cada

¹Para simplificar la notación, en el modelo que presentamos, cada vía se considera como una línea independiente.

²En estas estaciones los pasajeros pueden transbordar de una línea a otra sin salir del sistema.

estación terminal; estos andenes se caracterizan por ser ahí donde los trenes inician su recorrido (a lo largo de la línea hasta llegar al andén final, denotado por a_{NTA^l} donde NTA^l denota el número total de andenes de la línea l [véase ³]); cabe hacer notar, que cada andén corresponde a una única dirección; así, para cada línea los andenes siguen la secuencia $a_1 \rightarrow a_2 \rightarrow \dots \rightarrow a_{NTA^l}$. Por consiguiente, la red del Metro consiste en la unión de gráficas de todas las líneas unidas por un conjunto de vertices que representan las estaciones de correspondencia (para una mejor comprensión visual véase la gráfica A.2 en la página 70).

Para la modelación del flujo de trenes desde su salida de a_1 hasta su llegada a a_{NTA^l} , consideramos que los flujos de trenes no están sincronizados entre líneas, es decir, los tiempos de salida de los trenes son independientes entre líneas.

Para comenzar con esta primera parte de la modelación, introducimos dos conceptos de lo que se entenderá por tiempo de recorrido y tiempo de permanencia:

Tiempo de recorrido. El tiempo de recorrido es el tiempo que tarda un tren en trasladarse de un andén a otro (pertenecientes a una misma línea), considerando para ello la velocidad promedio con que se mueven los trenes; así, el tiempo de recorrido entre los andenes consecutivos $q - 1$ y q de L_l está dado por

$$R^l(a_q) = \frac{d^l(a_{q-1}, a_q)}{v^l},$$

donde $d^l(a_{q-1}, a_q)$ es la distancia que hay entre los andenes $q - 1$ y q de L_l y v^l es la velocidad promedio con la que viajan los trenes de L_l .

Tiempo de permanencia. El tiempo de permanencia que denotaremos por $P^l(a_q)$, es el tiempo que el tren permanece en $a_q \in L_l$ (es el tiempo que transcurre entre la hora de llegada al andén y la hora de salida del andén), en el cual, el operador del tren abre y cierra las puertas para permitir a los pasajeros el abandono o abordaje del tren⁴. $P^l(a_q)$ varía según la línea y el andén en el que se encuentre el tren y éstos son determinísticos.

En la Ciudad de México, el horario de una jornada laboral del Sistema Metro es variable según el día de la semana: de lunes a viernes, el horario es de 5 hrs. a 24 hrs.; sábados, de 6 a 24 hrs.; y domingos y días festivos, de 7 a 24 hrs. Es decir, para el Metro de la Ciudad de México existen tres tipos de jornadas laborales que denotaremos de manera arbitraria por J .

En una jornada normal (de lunes a viernes) el Sistema Metro divide su jornada laboral J en cuatro segmentos de tiempo: $[5, 10]$, $(10, 15]$, $(15, 21]$ y $(21, 24]$, donde el primer y tercer segmento corresponden a horas pico (turno matutino y turno vespertino, respectivamente) y el segundo y cuarto segmento corresponden a horas valle⁵. En cada segmento, que denotaremos por s con $s = 1 : S$ donde se asume que cada jornada J tiene S segmentos, los parámetros varían dada la demanda del servicio.

³En adelante “línea l ” se denotará simplemente por L_l y $a_q \in L_l$ indica que “el andén q pertenece a la línea l ”, donde q puede tomar valores enteros desde 1 hasta NTA^l .

⁴En los andenes iniciales a_1 , los pasajeros sólo pueden abordar; y en los andenes finales a_{NTA^l} , los pasajeros sólo pueden abandonar el tren.

⁵Se entiende por “horas pico”, aquellos segmentos de tiempo en los cuales la demanda del servicio es mayor que en los otros segmentos de tiempo que llamaremos “horas valle”, por lo que la frecuencia de salidas de trenes es distinta para cada tipo de segmento.

En lo siguiente obtendremos resultados para un segmento arbitrario s , de una jornada laboral J también arbitraria.

Una vez iniciado un segmento s de longitud T_s (cuyo segmento inicia al tiempo t_{s-1} y termina al tiempo t_s , con $t_s = \sum_{i=1}^s T_i$ y $t_0 = \min\{J\}$, para dar lugar al siguiente segmento), el primer tren sale de $a_1 \in L_l$, con $l = 1 : NTL$ donde NTL es el número total de líneas, a una hora determinada γ^l . Denotemos esta hora de salida por

$$V_1^l(a_1) = \gamma^l,$$

donde γ^l es una constante que varía según la línea y el segmento⁶.

A partir de $V_1^l(a_1)$ cada μ^l minutos sale un tren de a_1 , considerando para esto, un tiempo mínimo de seguridad, denotado por μ_{seg} , el cual es el tiempo mínimo permitido (por cuestiones de seguridad) entre salidas consecutivas de trenes⁷. Así, la variable de control μ^l está acotada inferiormente por μ_{seg} , esto es $\mu^l \geq \mu_{seg}$ (véase ⁸). Por consiguiente, la salida del siguiente tren de a_1 depende de la hora de salida del tren que le antecede; de la variable de control μ^l y de una v.a., que denotaremos por $\delta_{j+1}^l(a_1)$ (asociada a cada línea, a cada andén y a cada tren), a la cual llamaremos variable de fluctuación; cuya relación es la siguiente:

$$V_{j+1}^l(a_1) = V_j^l(a_1) + \mu^l + \delta_{j+1}^l(a_1), \quad (1.1)$$

con $t_{s-1} < V_j^l(a_1) \leq t_s$.

El tiempo de salida del j -ésimo tren de a_{q+1} , depende del tiempo en que éste salió del andén anterior y es igual a

$$V_j^l(a_{q+1}) = V_j^l(a_q) + R^l(a_{q+1}) + P^l(a_{q+1}) + \delta_j^l(a_{q+1}), \quad (1.2)$$

donde $R^l(a_{q+1})$ es el tiempo de recorrido del andén q al andén $q + 1$ y $P^l(a_{q+1})$ es el tiempo de permanencia.

Las fluctuaciones $\delta_j^l(a_q)$ son v.a. que representan el tiempo variable causado por cuestiones humanas: como algún retraso o adelanto en la salida y/o llegada del tren, ocasionado por el operador; algún retraso ocasionado por los pasajeros; alguna cuestión humana; entre otras causas. Dada su naturaleza, $\delta_j^l(a_q)$ se suponen v.a.i.i.d. con $E(\delta_j^l(a_q)) = 0$ para $l = 1 : NTL$, $q = 1 : NTA^l$ y para $j = 2 : NTT^l$, donde NTT^l denota el número total de viajes que se realizaron en L_l durante el segmento s .

Proposición 1.1.1. *Para cualquier $a_q \in L_l$ los tiempos entre salidas consecutivas de trenes, $Y_j^l(a_q) = V_j^l(a_q) - V_{j-1}^l(a_q)$, son v.a. tales que $E(Y_j^l(a_q)) = \mu^l$ y $Var(Y_j^l(a_q)) = (2q - 1)Var(\delta_j^l(a_1))$.*

Demostración. De la ecuación (1.1) tenemos que los tiempos entre salida consecutivas de trenes en a_1 son

$$Y_j^l(a_1) = \mu^l + \delta_j^l(a_1).$$

⁶Como se mencionó anteriormente, los parámetros varían según el segmento, pero éstos permanecen constantes a lo largo del mismo.

⁷ $\mu_{seg} = 90$ segundos para el Metro de la Ciudad de México.

⁸Al vector μ cuyos elementos son $\mu^1, \mu^2, \dots, \mu^{NTL}$ le llamaremos parámetro de control, el cual varía, dados sus elementos, según el segmento s en el que se encuentre el sistema y será de suma importancia en la sección 1.4.

Ahora, de la ecuación (1.2) para a_2 tenemos que

$$Y_j^l(a_2) = \mu^l + \delta_j^l(a_1) + \delta_j^l(a_2) - \delta_{j-1}^l(a_2),$$

y para a_3

$$Y_j^l(a_3) = \mu^l + \delta_j^l(a_1) + \sum_{i=2}^3 [\delta_j^l(a_i) - \delta_{j-1}^l(a_i)].$$

De forma general, para cualquier a_q , los tiempos entre salidas consecutivas de trenes están dados por la siguiente ecuación

$$Y_j^l(a_q) = \mu^l + \delta_j^l(a_1) + \sum_{i=2}^q [\delta_j^l(a_i) - \delta_{j-1}^l(a_i)]. \quad (1.3)$$

Por hipótesis tenemos que $\{\delta_j^l(a_q)\}_{l=1:NTL, q=1:NTA^l, j=2:NTT^l}$ son v.a.i.i.d. con $E(\delta_j^l(a_q)) = 0$, lo cual implica que $E(\delta_j^l(a_q) - \delta_{j-1}^l(a_q)) = 0$. Entonces

$$\begin{aligned} E(Y_j^l(a_q)) &= E\left(\mu^l + \delta_j^l(a_1) + \sum_{i=2}^q [\delta_j^l(a_i) - \delta_{j-1}^l(a_i)]\right) \\ &= \mu^l + E(\delta_j^l(a_1)) + \sum_{i=2}^q E(\delta_j^l(a_i) - \delta_{j-1}^l(a_i)) \\ &= \mu^l, \end{aligned}$$

y

$$\begin{aligned} Var(Y_j^l(a_q)) &= Var\left(\mu^l + \delta_j^l(a_1) + \sum_{i=2}^q [\delta_j^l(a_i) - \delta_{j-1}^l(a_i)]\right) \\ &= Var(\delta_j^l(a_1)) + \sum_{i=2}^q Var(\delta_j^l(a_i) - \delta_{j-1}^l(a_i)) \\ &= Var(\delta_j^l(a_1)) + \sum_{i=2}^q [Var(\delta_j^l(a_i)) + Var(\delta_{j-1}^l(a_i))] \\ &= Var(\delta_j^l(a_1)) + 2 \sum_{i=2}^q Var(\delta_j^l(a_i)) \\ &= Var(\delta_j^l(a_1)) + 2(q-1)Var(\delta_j^l(a_1)) \\ &= (2q-1)Var(\delta_j^l(a_1)). \end{aligned}$$

Por lo tanto, la proposición queda demostrada. ■

Los procesos de salidas de trenes son dependientes a lo largo de las plataformas de la misma línea, como indica la ecuación (1.3), pero independientes de los procesos de salida de trenes de las otras líneas. Esto debido al supuesto implícito de la no sincronización entre líneas.

Proposición 1.1.2. Para cualquier $a_q \in L_l$ los tiempos entre llegadas consecutivas de trenes, $T_j^l(a_q) = S_j^l(a_q) - S_{j-1}^l(a_q)$, son v.a. tales que $E\left(T_j^l(a_q)\right) = \mu^l$ y $Var\left(T_j^l(a_q)\right) = (2q - 3)Var\left(\delta_j^l(a_1)\right)$.

Demostración. Los tiempos de llegadas de trenes a $a_q \in L_l$, denotados por $S_j^l(a_q)$, pueden ser expresados en términos de $Y_j^l(a_q)$ de la siguiente manera:

$$S_j^l(a_q) = V_j^l(a_q) - P^l(a_q) - \delta_j^l(a_q).$$

Entonces

$$\begin{aligned} T_j^l(a_q) &= S_j^l(a_q) - S_{j-1}^l(a_q) \\ &= Y_j^l(a_q) - \left[\delta_j^l(a_q) - \delta_{j-1}^l(a_q)\right] \end{aligned}$$

y, usando la proposición anterior,

$$\begin{aligned} E\left(T_j^l(a_q)\right) &= E\left(Y_j^l(a_q) - \left[\delta_j^l(a_q) - \delta_{j-1}^l(a_q)\right]\right) \\ &= E\left(Y_j^l(a_q)\right) - E\left(\delta_j^l(a_q) - \delta_{j-1}^l(a_q)\right) \\ &= \mu^l, \end{aligned}$$

y

$$\begin{aligned} Var\left(T_j^l(a_q)\right) &= Var\left(Y_j^l(a_q) - \left[\delta_j^l(a_q) - \delta_{j-1}^l(a_q)\right]\right) \\ &= Var\left(\mu^l + \delta_j^l(a_1) + \sum_{i=2}^q \left[\delta_j^l(a_i) - \delta_{j-1}^l(a_i)\right] - \left[\delta_j^l(a_q) - \delta_{j-1}^l(a_q)\right]\right) \\ &= Var\left(\delta_j^l(a_1)\right) + 2(q - 2)Var\left(\delta_j^l(a_1)\right) \\ &= (2q - 3)Var\left(\delta_j^l(a_1)\right). \end{aligned}$$

Por lo que la proposición queda demostrada. ■

1.2. Flujo de pasajeros

En esta segunda parte del modelo consideramos el movimiento de los pasajeros a través del sistema Metro, en el cual, los pasajeros arriban a cualquier andén (distinto a a_{NTA^l}) provenientes de la calle o, si el andén pertenece a una estación de correspondencia, de otras líneas del Metro. Si los pasajeros provienen de la calle, entonces el andén es llamado andén origen, por ser el andén donde el pasajero ingresa al sistema, y lo denotamos por a_o .

Una vez que los pasajeros se encuentran en el andén, éstos abordan el primer tren que llega. Para comenzar con esta parte de la modelación haremos los siguientes supuestos: (a) se asume que el cupo de cada tren es ilimitado, es decir, el tren tiene suficiente espacio para albergar en él a todos los pasajeros que esperan en el

andén; (b) los pasajeros no se regresan, esto es, los pasajeros no transbordarán a la línea que corre en sentido contrario a la línea en la cual vienen; (c) los pasajeros que arriban por ingreso al sistema en $a_o \in L_l$, lo hacen de acuerdo a un proceso de Poisson $NP_{a_o}^l(t)$ con intensidad constante $\lambda_s^l(a_o)$; (d) el parámetro de intensidad $\lambda_s^l(a_o)$ varía para cada andén, para cada línea y para cada segmento s , pero éste permanece constante a lo largo de s (véase ⁹); y (e) los procesos de arribos de pasajeros son independientes entre sí, e independientes del proceso de salida de trenes.

Proposición 1.2.1. *El número de pasajeros que abordan por ingreso al sistema el j -ésimo tren en $a_o \in L_l$, denotado por $PAI_j^l(a_o)$, satisface*

$$E\left(PAI_j^l(a_o)\right) = \mu^l \lambda^l(a_o).$$

Demostración. El término $PAI_j^l(a_o) = NP_{a_o}^l\left(V_j^l(a_o)\right) - NP_{a_o}^l\left(V_{j-1}^l(a_o)\right)$ indica el número de pasajeros que arriban por ingreso al sistema en $a_o \in L_l$ en el intervalo de tiempo $\left(V_{j-1}^l(a_o), V_j^l(a_o)\right]$, y en consecuencia, es el número de pasajeros por ingreso al sistema que abordan el j -ésimo tren.

Como $NP_{a_o}^l(t)$ es un proceso de Poisson, entonces el término $PAI_j^l(a_o)$ dado $V_j^l(a_o) - V_{j-1}^l(a_o)$ se distribuye como una v.a. de Poisson con parámetro $\lambda^l(a_o)\left(V_j^l(a_o) - V_{j-1}^l(a_o)\right)$, por tanto

$$\begin{aligned} E\left(PAI_j^l(a_o)\right) &= E\left(E\left[PAI_j^l(a_o) \mid V_j^l(a_o) - V_{j-1}^l(a_o)\right]\right) \\ &= E\left(E\left[NP_{a_o}^l\left(V_j^l(a_o)\right) - NP_{a_o}^l\left(V_{j-1}^l(a_o)\right) \mid V_j^l(a_o) - V_{j-1}^l(a_o)\right]\right) \\ &= E\left(\lambda^l(a_o)\left[V_j^l(a_o) - V_{j-1}^l(a_o)\right]\right) \\ &= \lambda^l(a_o)E\left(V_j^l(a_o) - V_{j-1}^l(a_o)\right) \\ &= \lambda^l(a_o)E\left(Y_j^l(a_o)\right) \\ &= \mu^l \lambda^l(a_o). \end{aligned}$$

La tercera igualdad se obtiene porque el proceso de arribos de pasajeros $NP_{a_o}^l(t)$ es independiente del proceso de salidas de trenes $V_j^l(a_o)$, y la última igualdad se obtiene de aplicar la proposición 1.1.1. ■

El proceso de arribos de pasajeros por ingreso al sistema $NP_{a_o}^m(t)$, puede descomponerse en la suma de los procesos de Poisson $NP_{a_o, a_d}^{m,l}(t)$, con intensidad $\lambda^{m,l}(a_o, a_d)$, que representan los proceso de arribos de pasajeros, por ingreso al sistema, en $a_o \in L_m$ con destino $a_d \in L_l$, con m y l no necesariamente iguales. Esto es,

$$NP_{a_o}^m(t) = \sum_l \sum_{a_d \in L_l} NP_{a_o, a_d}^{m,l}(t),$$

y

$$\lambda^m(a_o) = \sum_l \sum_{a_d \in L_l} \lambda^{m,l}(a_o, a_d),$$

⁹De forma similar que en la sección anterior, se obtendrán resultados para un segmento dado s , pero arbitrario, por lo que omitiremos su notación.

con $\lambda^{m,l}(a_o, a_d)$ la intensidad de pasajeros que arriban, por ingreso al sistema, en $a_o \in L_m$ con destino $a_d \in L_l$ (véase ¹⁰).

1.2.1. Pasajeros en estaciones de correspondencia

Cuando el andén destino a_d se encuentra en una línea distinta a la que pertenece a_o , esto es, $a_o \in L_m$ y $a_d \in L_l$ tales que $m \neq l$, entonces los pasajeros deberán transbordar cuantas veces sean necesarias, a través de las estaciones de correspondencia hasta llegar a su andén destino.

Anteriormente vimos como se distribuyen los pasajeros que arriban por ingreso al sistema, ahora veremos como se modelan los pasajeros que arriban por transbordo (pasajeros provenientes de otras líneas). Sea $a_{p_l} \in L_l$ un andén en una estación de correspondencia. Al andén a_{p_l} arriban pasajeros por ingreso al sistema y también arriban pasajeros por transbordo. Denotemos por $\mathcal{EC}(a_{p_l})$ al conjunto de andenes que confluyen con a_{p_l} en una misma estación de correspondencia, exceptuando al mismo a_{p_l} y al andén que pertenece a la línea opuesta (línea que corre en sentido contrario) a L_l (véase ¹¹) y denotemos, de forma general, por \mathcal{EC} al conjunto de andenes que pertenecen a estaciones de correspondencia, así, $\mathcal{EC}(a_{p_l}) \subset \mathcal{EC}$ para todo $a_{p_l} \in \mathcal{EC}$.

Los pasajeros que arriban por transbordo a $a_{p_l} \in L_l$ son clasificados como: pasajeros de primer orden, si es su primer transbordo; pasajeros de orden dos, si previamente transbordaron una vez, es decir, si es su segundo transbordo; pasajeros de orden tres, si es su tercer transbordo; y así sucesivamente para pasajeros de orden superior. En otras palabras, los pasajeros de primer orden son aquellos pasajeros que se encuentran en $a_{p_l} \in L_l$ cuyo andén origen $a_o \in L_m$, con $L_m \neq L_l$, y transbordaron a $a_{p_l} \in L_l$ en el andén $a_{q_m} \in L_m$ (véase ¹²), por ende, $a_o < a_{q_m}$, $a_{q_m} \in \mathcal{EC}(a_{p_l})$ y $L_m \cap L_l \neq \emptyset$; los pasajeros de segundo orden son aquellos pasajeros que se encuentran en $a_{p_l} \in L_l$ tales que $a_o \in L_n$, con $L_n \neq L_l$ pero $L_n \cap L_l = \emptyset$, y existe L_m tal que $L_m \cap L_n \neq \emptyset$ y $L_m \cap L_l \neq \emptyset$, por lo que estos pasajeros fueron pasajeros de primer orden de L_n a L_m ; y de forma similar se define a los pasajeros de orden tres o más (véase figura 1.1).

En las estaciones de correspondencia asumimos que los pasajeros que transbordan forman una unidad, es decir, los pasajeros transbordan en conjunto; y también asumimos, que éstos se tele-transportan entre los andenes que confluyen en la misma estación de correspondencia, esto es, no se considera el tiempo que toma a los pasajeros caminar de un andén a otro para cambiar de línea.

¹⁰ a_d representa todos los destinos posibles sobre la red del Metro dado a_o , y se entiende como andén destino, el andén donde el pasajero abandona el sistema. Existen métodos que se enfocan en determinar las trayectorias para ir de cierto lugar de origen a un destino determinado (ya que pueden haber múltiples trayectorias para un mismo par origen-destino; en este caso, la trayectoria a elegir debe ser óptima; y la trayectoria óptima depende de los intereses que se buscan: minimizar la distancia, minimizar el tiempo de recorrido, el costo del traslado o minimizar el número de transbordos, entre otros intereses), estos métodos son conocidos como métodos de asignación de tránsito. En el presente trabajo, para simplificar el modelo, suponemos que para cada par de andenes (a_o, a_d) , la trayectoria a seguir a lo largo de la red del Metro es única y en consecuencia determinística, por lo cual no abordamos los métodos de asignación de tránsito.

¹¹ Note que si a'_{p_l} es el andén que pertenece a la línea opuesta a L_l , entonces, $\mathcal{EC}(a_{p_l}) = \mathcal{EC}(a'_{p_l})$.

¹² Los pasajeros llegan en un tren de L_m a a_{q_m} caminan para trasladarse a a_{p_l} y una vez en a_{p_l} abordan el siguiente tren de L_l .

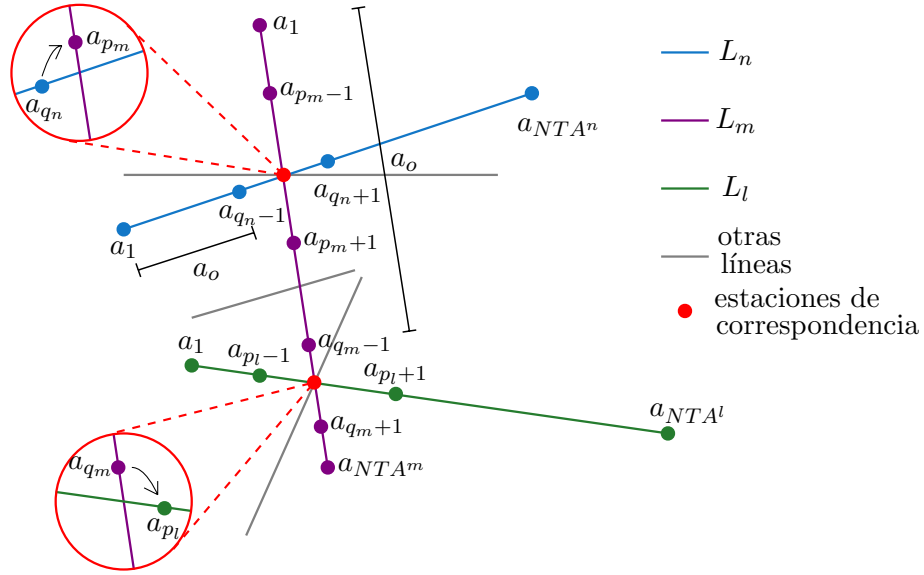


Figura 1.1: Andenes en estaciones de correspondencia.

En esta sección, para los transbordos, usamos el sub-índice q_* para los andenes de llegada y el sub-índice p_* para los andenes de partida, en los cuales el sub-índice “*” indica la línea a la cual pertenecen estos andenes. Del mismo modo, podemos identificar a los pasajeros que transbordan por el tren en que llegan y por el tren que, posteriormente, abordan.

Para modelar el número de pasajeros que arriban por transbordo a $a_{p_l} \in L_l$, tal que $a_{p_l} \in \mathcal{EC}$, consideremos la siguiente notación:

- $\mathcal{I}(L_l, a_{p_l})$: Denota el conjunto de líneas que se intersectan con L_l en la estación a la que pertenece a_{p_l} .
- $D_{a_o}(a_{q_m}, a_{p_l})$: Es el conjunto de andenes destino a_d , tales que las trayectorias para ir de a_o a a_d requieren transbordo de $a_{q_m} \in L_m$ a $a_{p_l} \in L_l$.
- $NT_{a_{p_l}}^l(t)$: Denota el proceso de conteo de trenes que llegan al andén $a_{p_l} \in L_l$ al tiempo t .

Proposición 1.2.2. Sea $a_{q_m} \in L_m$ fijo tal que $a_{q_m} \in \mathcal{EC}$ y sea $a_{p_l} \in L_l$ con $a_{p_l} \in \mathcal{EC}(a_{q_m})$, entonces el número de pasajeros por ingreso al sistema que abordaron el k -ésimo tren y que lo abandonaron en $a_{q_m} \in L_m$ para transbordar a $a_{p_l} \in L_l$, denotado por $PAI_k^{m,l}$, satisface

$$E\left(PAI_k^{m,l}\right) = \mu^m \rho^{m,l},$$

donde

$$\rho^{m,l} = \sum_{\substack{a_o \in L_m, \\ a_o < a_{q_m}}} \sum_{a_d \in D_{a_o}(a_{q_m}, a_{p_l})} \lambda^{m,l}(a_o, a_d).$$

Demostración. Los pasajeros por ingreso al sistema que llegaron a a_{q_m} en el k -ésimo tren, necesariamente lo abordaron en algún $a_o \in \{a_1, a_2, \dots, a_{q_m-1}\}$ a las $V_k^m(a_o)$, por lo tanto, estos pasajeros arribaron a $a_o \in L_m$ en el intervalo de tiempo $(V_{k-1}^m(a_o), V_k^m(a_o)]$, por consiguiente, el valor esperado del número de los pasajeros que abordaron el k -ésimo tren en $a_o \in L_m$ para transbordar de $a_{q_m} \in L_m$ a $a_{p_l} \in L_l$, denotado por $PAI_k^{m,l}(a_o)$, está dado por

$$\begin{aligned}
 E\left(PAI_k^{m,l}(a_o)\right) &= E\left(\sum_{a_d \in D_{a_o}(a_{q_m}, a_{p_l})} \left[NP_{a_o, a_d}^{m,l}(V_k^m(a_o)) - NP_{a_o, a_d}^{m,l}(V_{k-1}^m(a_o)) \right]\right) \\
 &= E\left(E\left[\sum_{a_d \in D_{a_o}(a_{q_m}, a_{p_l})} \left[NP_{a_o, a_d}^{m,l}(V_k^m(a_o)) - NP_{a_o, a_d}^{m,l}(V_{k-1}^m(a_o)) \right] \middle| Y_k^m(a_o)\right]\right) \\
 &= \sum_{a_d \in D_{a_o}(a_{q_m}, a_{p_l})} E\left(\lambda^{m,l}(a_o, a_d) [V_k^m(a_o) - V_{k-1}^m(a_o)]\right) \\
 &= \sum_{a_d \in D_{a_o}(a_{q_m}, a_{p_l})} \lambda^{m,l}(a_o, a_d) E(Y_k^m(a_o)) \\
 &= \mu^m \sum_{a_d \in D_{a_o}(a_{q_m}, a_{p_l})} \lambda^{m,l}(a_o, a_d).
 \end{aligned}$$

Luego, $E\left(PAI_k^{m,l}\right)$ es calculado usando este último resultado para todo $a_o \in L_m$, tal que $a_o \in \{a_1, a_2, \dots, a_{q_m-1}\}$, de la siguiente manera:

$$\begin{aligned}
 E\left(PAI_k^{m,l}\right) &= E\left(\sum_{\substack{a_o \in L_m, \\ a_o < a_{q_m}}} PAI_k^{m,l}(a_o)\right) \\
 &= \sum_{\substack{a_o \in L_m, \\ a_o < a_{q_m}}} E\left(PAI_k^{m,l}(a_o)\right) \\
 &= \sum_{\substack{a_o \in L_m, \\ a_o < a_{q_m}}} \left(\mu^m \sum_{a_d \in D_{a_o}(a_{q_m}, a_{p_l})} \lambda^{m,l}(a_o, a_d)\right) \\
 &= \mu^m \sum_{\substack{a_o \in L_m, \\ a_o < a_{q_m}}} \sum_{a_d \in D_{a_o}(a_{q_m}, a_{p_l})} \lambda^{m,l}(a_o, a_d) \\
 &= \mu^m \rho^{m,l}
 \end{aligned}$$

con

$$\rho^{m,l} = \sum_{\substack{a_o \in L_m, \\ a_o < a_{q_m}}} \sum_{a_d \in D_{a_o}(a_{q_m}, a_{p_l})} \lambda^{m,l}(a_o, a_d).$$

■

$\rho^{m,l}$ es la intensidad de arribos de pasajeros, por ingreso al sistema, a L_m que demandan transbordar (de L_m) a L_l .

En esta proposición vimos como llegan a $a_{p_m} \in L_m$ los pasajeros que van a transbordar por primera vez. En la siguiente proposición veremos como estos pasajeros, una vez que ya transbordaron, son agrupados en $a_{p_l} \in L_l$ para tomar el siguiente tren.

Proposición 1.2.3. Sea $P1_j^l(a_{p_l})$ el número de pasajeros (por transbordo) de primer orden que arriban a $a_{p_l} \in L_l$

(fijo) para abordar el j -ésimo tren, con a_{p_l} tal que $a_{p_l} \in \mathcal{EC}$, entonces

$$E\left(P1_j^l(a_{p_l})\right) = \sum_{L_m \in \mathcal{I}(L_l, a_{p_l})} \mu^m \rho^{m,l} E\left(NT_{a_{q_m}}^m\left(V_j^l(a_{p_l})\right) - NT_{a_{q_m}}^m\left(V_{j-1}^l(a_{p_l})\right)\right),$$

con $\rho^{m,l}$ como en la proposición anterior.

Demostración. Consideremos $a_{q_m} \in L_m$ tal que $a_{q_m} \in \mathcal{EC}(a_{p_l})$. El j -ésimo tren sale de $a_{p_l} \in L_l$ a las $V_j^l(a_{p_l})$ entonces los pasajeros de L_m que abordarán el j -ésimo tren en $a_{p_l} \in L_l$ tuvieron que llegar a a_{q_m} durante el intervalo de tiempo $\left[V_{j-1}^l(a_{p_l}), V_j^l(a_{p_l})\right]$. Por consiguiente, $NT_{a_{q_m}}^m\left(V_{j-1}^l(a_{p_l})\right) + 1$ y $NT_{a_{q_m}}^m\left(V_j^l(a_{p_l})\right)$ son el primer y el último tren, respectivamente, de L_m , en los cuales llegaron pasajeros a a_{q_m} que tomarán el j -ésimo tren en a_{p_l} . Por tanto, el valor esperado del número de pasajeros que vienen de L_m y que tomarán el j -ésimo tren de L_l , convirtiéndose así en pasajeros de primer orden y denotado por $P1_j^{l \leftarrow m}$, esta dado por

$$\begin{aligned} E\left(P1_j^{l \leftarrow m}\right) &= E\left(\sum_{k=NT_{a_{q_m}}^m\left(V_{j-1}^l(a_{p_l})\right)+1}^{NT_{a_{q_m}}^m\left(V_j^l(a_{p_l})\right)} PAI_k^{m,l}\right) \\ &= E\left(\sum_{k=1}^{NT_{a_{q_m}}^m\left(V_j^l(a_{p_l})\right)} PAI_k^{m,l}\right) - E\left(\sum_{k=1}^{NT_{a_{q_m}}^m\left(V_{j-1}^l(a_{p_l})\right)} PAI_k^{m,l}\right) \\ &= E\left(E\left[\sum_{k=1}^{NT_{a_{q_m}}^m\left(V_j^l(a_{p_l})\right)} PAI_k^{m,l} \middle| NT_{a_{q_m}}^m\left(V_j^l(a_{p_l})\right)\right]\right) + \\ &\quad - E\left(E\left[\sum_{k=1}^{NT_{a_{q_m}}^m\left(V_{j-1}^l(a_{p_l})\right)} PAI_k^{m,l} \middle| NT_{a_{q_m}}^m\left(V_{j-1}^l(a_{p_l})\right)\right]\right) \\ &= E\left(NT_{a_{q_m}}^m\left(V_j^l(a_{p_l})\right) E\left(PAI_k^{m,l}\right)\right) - E\left(NT_{a_{q_m}}^m\left(V_{j-1}^l(a_{p_l})\right) E\left(PAI_k^{m,l}\right)\right) \\ &= E\left(PAI_k^{m,l}\right) E\left(NT_{a_{q_m}}^m\left(V_j^l(a_{p_l})\right) - NT_{a_{q_m}}^m\left(V_{j-1}^l(a_{p_l})\right)\right) \\ &= \mu^m \rho^{m,l} E\left(NT_{a_{q_m}}^m\left(V_j^l(a_{p_l})\right) - NT_{a_{q_m}}^m\left(V_{j-1}^l(a_{p_l})\right)\right). \end{aligned}$$

Esto último por la proposición anterior.

Ahora, para obtener $E\left(P1_j^l(a_{p_l})\right)$ consideramos todas las líneas L_m tales que $L_m \in \mathcal{I}(L_l, a_{p_l})$, así

$$\begin{aligned} E\left(P1_j^l(a_{p_l})\right) &= E\left(\sum_{L_m \in \mathcal{I}(L_l, a_{p_l})} P1_j^{l \leftarrow m}\right) \\ &= \sum_{L_m \in \mathcal{I}(L_l, a_{p_l})} E\left(P1_j^{l \leftarrow m}\right) \\ &= \sum_{L_m \in \mathcal{I}(L_l, a_{p_l})} \mu^m \rho^{m,l} E\left(NT_{a_{q_m}}^m\left(V_j^l(a_{p_l})\right) - NT_{a_{q_m}}^m\left(V_{j-1}^l(a_{p_l})\right)\right). \end{aligned}$$

Por lo que la proposición queda demostrada. ■

Los pasajeros de segundo orden en $a_{p_l} \in L_l$ tienen su anden origen en L_n : línea que no tiene estación de correspondencia con L_l , pero si tiene correspondencia con L_m en la estación donde confluyen los andenes $a_{q_n} \in L_n$ y $a_{p_m} \in L_m$. Por otro lado, L_m tiene correspondencia con L_l en la estación donde confluyen los andenes $a_{q_m} \in L_m$ y $a_{p_l} \in L_l$. Entonces, los pasajeros de segundo orden ingresaron al sistema en algún $a_o \in L_n$ con $a_o < a_{q_n}$ y son pasajeros de primer orden de L_n a L_m y son pasajeros de segundo orden de L_m a L_l (véase figura 1.1 en la página 8).

Proposición 1.2.4. *Sea $a_{q_m} \in L_m$ fijo tal que $a_{q_m} \in \mathcal{EC}$ y sea $a_{p_l} \in L_l$ con $a_{p_l} \in \mathcal{EC}(a_{q_m})$, entonces el número de pasajeros de primer orden que abordaron k -ésimo tren y que lo abandonaran en $a_{q_m} \in L_m$ para transbordar a $a_{p_l} \in L_l$, denotado por $P1_k^{m,l}$, satisface*

$$E\left(P1_k^{m,l}\right) = \sum_{\substack{a_{p_m} \in L_m \cap \mathcal{EC}, \\ a_{p_m} < a_{q_m}}} \sum_{L_n \in \mathcal{I}(L_m, a_{p_m})} \mu^n \rho^{n,m,l} E\left(NT_{a_{q_n}}^n\left(V_k^m(a_{p_m})\right) - NT_{a_{q_n}}^n\left(V_{k-1}^m(a_{p_m})\right)\right),$$

donde

$$\rho^{n,m,l} = \sum_{\substack{a_o \in L_n, \\ a_o < a_{q_n}}} \sum_{a_d \in D_{a_o}(a_{q_n}, a_{p_m}) \cap D_{a_o}(a_{q_m}, a_{p_l})} \lambda^{n,l}(a_o, a_d),$$

con $a_{q_n} \in \mathcal{EC}(a_{p_m})$, $a_{q_n} \in L_n$ y $a_{p_m}, a_{q_m} \in L_m$.

Demostración. De la proposición anterior se puede deducir que el valor esperado del número de pasajeros de primer orden, que tomarán el k -ésimo tren en $a_{p_m} \in L_m$ para después transbordar (por segunda vez) de $a_{q_m} \in L_m$ a $a_{p_l} \in L_l$, cuyo valor denotaremos por $P1_k^{m,l}(a_{p_m})$, satisface:

$$E\left(P1_k^{m,l}(a_{p_m})\right) = \sum_{L_n \in \mathcal{I}(L_m, a_{p_m})} \mu^n \rho^{n,m,l} E\left(NT_{a_{q_n}}^n\left(V_k^m(a_{p_m})\right) - NT_{a_{q_n}}^n\left(V_{k-1}^m(a_{p_m})\right)\right),$$

con

$$\rho^{n,m,l} = \sum_{\substack{a_o \in L_n, \\ a_o < a_{q_n}}} \sum_{a_d \in D_{a_o}(a_{q_n}, a_{p_m}) \cap D_{a_o}(a_{q_m}, a_{p_l})} \lambda^{n,l}(a_o, a_d).$$

Por lo que la proposición queda demostrada, al considerar todos los posibles andenes en estaciones de correspondencia de L_m , anteriores a a_{q_m} , que reciben pasajeros de primer orden. ■

En esta proposición vimos como llegan a $a_{p_m} \in L_m$ los pasajeros que van a transbordar por segunda vez. En la siguiente proposición veremos como estos pasajeros, una vez que ya transbordaron, son agrupados en $a_{p_l} \in L_l$ para tomar el siguiente tren.

Proposición 1.2.5. *Sea $P2_j^l(a_{p_l})$ el número de pasajeros (por transbordo) de segundo orden que arriban a $a_{p_l} \in L_l$ (fijo) para abordar el j -ésimo tren, con a_{p_l} tal que $a_{p_l} \in \mathcal{EC}$, entonces*

$$E\left(P2_j^l(a_{p_l})\right) = \sum_{L_m \in \mathcal{I}(L_l, a_{p_l})} \sum_{\substack{a_{p_m} \in L_m \cap \mathcal{EC}, \\ a_{p_m} < a_{q_m}}} \sum_{L_n \in \mathcal{I}(L_m, a_{p_m})} \mu^n \rho^{n,m,l} E\left(NT_{a_{q_n}}^n\left(V_{N_j}^m(a_{p_m})\right) + \right. \\ \left. - NT_{a_{q_n}}^n\left(V_{N_j-1}^m(a_{p_m})\right)\right),$$

donde $N_j = NT_{a_{qm}}^m (V_j^l(a_{pl}))$, $N_{j-1} = NT_{a_{qm}}^m (V_{j-1}^l(a_{pl}))$ y $\rho^{n,m,l}$ como en la proposición anterior.

Demostración. Para conocer el número de pasajeros de segundo orden que abordan el j -ésimo tren en $a_{pl} \in L_l$, basta ver, cuantos de estos pasajeros como pasajeros de primer orden abordaron algún tren k en L_m con la condición que $NT_{a_m}^m(V_{j-1}^l(a_{pl})) < k \leq NT_{a_m}^m(V_j^l(a_{pl}))$. Además, debemos considerar todas las líneas L_m que hacen correspondencia con L_l en la misma estación que contiene a a_{pl} , es decir, se suma sobre todas la líneas L_m tales que $L_m \in \mathcal{I}(L_l, a_{pl})$, entonces (por la proposición anterior) $P2_j^l(a_{pl})$ se puede expresar como

$$P2_j^l(a_{pl}) = \sum_{L_m \in \mathcal{I}(L_l, a_{pl})} \sum_{k=NT_{a_m}^m(V_{j-1}^l(a_{pl}))+1}^{NT_{a_m}^m(V_j^l(a_{pl}))} P1_k^{m,l}.$$

Para calcular $E(P2_j^l(a_{pl}))$ consideremos lo siguiente:

$$\begin{aligned} E \left(\sum_{k=1}^{NT_{a_{qm}}^m(V_j^l(a_{pl}))} P1_k^{m,l} \right) &= E \left(E \left[\sum_{k=1}^{NT_{a_{qm}}^m(V_j^l(a_{pl}))} P1_k^{m,l} \middle| NT_{a_{qm}}^m(V_j^l(a_{pl})) \right] \right) \\ &= E \left(\sum_{k=1}^{NT_{a_{qm}}^m(V_j^l(a_{pl}))} E(P1_k^{m,l}) \right) \\ &= E \left(\sum_{\substack{a_{p_m} \in L_m \cap \mathcal{EC}, \\ a_{p_m} < a_{q_m}}} \sum_{L_n \in \mathcal{I}(L_m, a_{p_m})} \mu^n \rho^{n,m,l} E \left(NT_{a_{q_n}}^n (V_{N_j}^m(a_{p_m})) + \right. \right. \\ &\quad \left. \left. - NT_{a_{q_n}}^n (V_0^m(a_{p_m})) \right) \right) \\ &= \sum_{\substack{a_{p_m} \in L_m \cap \mathcal{EC}, \\ a_{p_m} < a_{q_m}}} \sum_{L_n \in \mathcal{I}(L_m, a_{p_m})} \mu^n \rho^{n,m,l} E \left(NT_{a_{q_n}}^n (V_{N_j}^m(a_{p_m})) + \right. \\ &\quad \left. - NT_{a_{q_n}}^n (V_0^m(a_{p_m})) \right), \end{aligned}$$

con $N_j = NT_{a_{qm}}^m (V_j^l(a_{pl}))$. De igual manera obtenemos que

$$\begin{aligned} E \left(\sum_{k=1}^{NT_{a_{qm}}^m(V_{j-1}^l(a_{pl}))} P1_k^{m,l} \right) &= \sum_{\substack{a_{p_m} \in L_m \cap \mathcal{EC}, \\ a_{p_m} < a_{q_m}}} \sum_{L_n \in \mathcal{I}(L_m, a_{p_m})} \mu^n \rho^{n,m,l} E \left(NT_{a_{q_n}}^n (V_{N_{j-1}}^m(a_{p_m})) + \right. \\ &\quad \left. - NT_{a_{q_n}}^n (V_0^m(a_{p_m})) \right), \end{aligned}$$

con $N_{j-1} = NT_{a_{qm}}^m (V_{j-1}^l(a_{pl}))$. Entonces

$$E(P2_j^l(a_{pl})) = \sum_{L_m \in \mathcal{I}(L_l, a_{pl})} E \left(\sum_{k=NT_{a_{qm}}^m(V_{j-1}^l(a_{pl}))+1}^{NT_{a_{qm}}^m(V_j^l(a_{pl}))} P1_k^{m,l} \right)$$

$$\begin{aligned}
 &= \sum_{L_m \in \mathcal{I}(L_l, a_{p_l})} \left[E \left(\sum_{k=1}^{NT_{a_{q_m}}^m(V_j^l(a_{p_l}))} P1_k^{m,l} \right) - E \left(\sum_{k=1}^{NT_{a_{q_m}}^m(V_{j-1}^l(a_{p_l}))} P1_k^{m,l} \right) \right] \\
 &= \sum_{L_m \in \mathcal{I}(L_l, a_{p_l})} \sum_{\substack{a_{p_m} \in L_m \cap \mathcal{EC}, \\ a_{p_m} < a_{q_m}}} \sum_{L_n \in \mathcal{I}(L_m, a_{p_m})} \mu^n \rho^{n,m,l} E \left(NT_{a_{q_n}}^n \left(V_{N_j}^m(a_{p_m}) \right) + \right. \\
 &\quad \left. - NT_{a_{q_n}}^n \left(V_{N_{j-1}}^m(a_{p_m}) \right) \right).
 \end{aligned}$$

Por lo que la proposición queda demostrada. ■

Para este modelo, los pasajeros pueden transbordar a lo más $NTL/2 - 1$ veces, dado que el sistema Metro cuenta con NTL líneas, por tanto, el orden máximo de los pasajeros puede ser $NTL/2 - 1$ o menos¹³. En este trabajo presentamos resultados para pasajeros de primer y segundo orden; proposiciones similares pueden ser demostradas para pasajeros de orden superior. Cabe señalar que a medida que los pasajeros elevan su orden, el modelo se vuelve más complejo.

1.3. Pasajeros en el tren

Previamente presentamos el modelo matemático que permite replicar el comportamiento de trenes y de pasajeros, y su interacción a través de las estaciones de correspondencia. Ahora modelaremos el número de pasajeros que están en el tren al salir éste de cada andén, para esto, consideremos que al llegar un tren a cierto andén, una cantidad de pasajeros que vienen en el tren, lo abandonan y el grupo de pasajeros que esperaban en el andén, lo abordan.

Sean $PA_j^l(a_q)$ el número de pasajeros que abordan el j -ésimo tren en $a_q \in L_l$, $PD_j^l(a_q)$ el número de pasajeros que lo abandonan y $PS_j^l(a_q)$ el número de pasajeros que permanecen en el tren al salir éste del andén a_q . Estos términos se relacionan a través de las siguientes ecuaciones:

$$PS_j^l(a_1) = PA_j^l(a_1), \quad (1.4)$$

$$PS_j^l(a_{q+1}) = PS_j^l(a_q) - PD_j^l(a_{q+1}) + PA_j^l(a_{q+1}), \quad (1.5)$$

y como es de esperarse

$$PS_j^l(a_{NTA^l}) = PS_j^l(a_{NTA^l-1}) - PD_j^l(a_{NTA^l}) = 0. \quad (1.6)$$

¹³El orden de los pasajeros está acotado superiormente por $NTL/2 - 1$, pero no siempre existen trayectorias (a_o, a_d) , en la red del sistema Metro, que requieran $NTL/2 - 1$ transbordos, es decir, basta una simple inspección visual a la geografía de la red del Metro, para determinar todas las trayectorias posibles (a_o, a_d) y por ende el máximo número de transbordos requeridos para una sistema Metro dado.

1.3.1. Pasajeros que abordan el tren

Podemos ver que $PA_j^l(a_q)$ es igual al número de pasajeros que abordan por ingreso al sistema, $PAI_j^l(a_q)$; más, si es que $a_q \in \mathcal{EC}$, el número de pasajeros que abordan por transbordo, denotado por $PAT_j^l(a_q)$; es decir,

$$PA_j^l(a_q) = PAI_j^l(a_q) + PAT_j^l(a_q)I_{\{a_q \in \mathcal{EC}\}} \quad (1.7)$$

donde $I_{\{a_q \in \mathcal{EC}\}}$ representa la función indicadora.

El valor esperado de $PAI_j^l(a_q)$ es calculado como se vio en la proposición 1.2.1 y el valor esperado de $PAT_j^l(a_q)$ es calculado como se vio en las proposiciones 1.2.3 y 1.2.5, ya que $PAT_j^l(a_q) = P1_j^l(a_q) + P2_j^l(a_q) + \dots$. Por consiguiente se obtiene $E\left(PA_j^l(a_q)\right)$.

1.3.2. Pasajeros que abandonan el tren

Ahora, para analizar $PD_j^l(a_q)$ observemos que si el andén se encuentra en una estación normal, implica que a_q es el andén destino para los pasajeros que abandonan el tren (cuyos pasajeros pudieron haber transbordado previamente o no), denotemos a estos pasajeros por $PDA_j^l(a_q)$; si el andén se encuentra en una estación de correspondencia, entonces a los pasajeros que abandonan el tren (en dicho andén) los podemos distinguir en dos tipos: los pasajeros que abandona el tren porque han llegado a su andén destino (al igual que en las estaciones normales), y los pasajeros que abandonan el tren porque requieren transbordar en esa estación (lo cual difiere de las estaciones normales), denotemos a estos últimos por $PDT_j^l(a_q)$. Es decir,

$$PD_j^l(a_q) = PDA_j^l(a_q) + PDT_j^l(a_q)I_{\{a_q \in \mathcal{EC}\}}. \quad (1.8)$$

Observemos que

$$PDT_j^l(a_q) = \sum_{L_i \in \mathcal{I}(L_l, a_q)} \left(PAI_j^{l,i} + P1_j^{l,i} + P2_j^{l,i} + \dots \right),$$

por lo que $E\left(PDT_j^l(a_q)\right)$ es calculado como se vio en las proposiciones 1.2.2 y 1.2.4.

Entre los pasajeros que abandonan el tren porque han llegado a su andén destino, $PDA_j^l(a_q)$, encontramos a los pasajeros que abandonan el sistema sin haber realizado transbordo alguno; a los pasajeros que abandonan el sistema, que requirieron transbordar una vez; también encontramos a los pasajeros que abandonan el sistema, que requirieron transbordar dos veces; y así sucesivamente. Para calcular $E\left(PDA_j^l(a_q)\right)$ consideremos las siguientes proposiciones.

Proposición 1.3.1. *Sea $PDI_j^l(a_q)$ el número de pasajeros que abandonan el j -ésimo tren para abandonar el sistema en $a_q \in L_l$ tales que $a_o \in L_l$ (es decir, estos pasajeros no requirieron transbordar para llegar a su destino), entonces*

$$E\left(PDI_j^l(a_q)\right) = \mu^l \sum_{\substack{a_o \in L_l, \\ a_o < a_q}} \lambda^{l,l}(a_o, a_q).$$

Demostración. Es fácil ver que

$$\begin{aligned}
E\left(PD1_j^l(a_q)\right) &= E\left(\sum_{\substack{a_o \in L_l, \\ a_o < a_q}} NP_{a_o, a_q}^{l,l}\left(V_j^l(a_o)\right) - NP_{a_o, a_q}^{l,l}\left(V_{j-1}^l(a_o)\right)\right) \\
&= \sum_{\substack{a_o \in L_l, \\ a_o < a_q}} E\left(E\left[NP_{a_o, a_q}^{l,l}\left(V_j^l(a_o)\right) - NP_{a_o, a_q}^{l,l}\left(V_{j-1}^l(a_o)\right) \middle| Y_j^l(a_o)\right]\right) \\
&= \sum_{\substack{a_o \in L_l, \\ a_o < a_q}} E\left(\lambda^{l,l}(a_o, a_q) Y_j^l(a_o)\right) \\
&= \sum_{\substack{a_o \in L_l, \\ a_o < a_q}} \lambda^{l,l}(a_o, a_q) E\left(Y_j^l(a_o)\right) \\
&= \mu^l \sum_{\substack{a_o \in L_l, \\ a_o < a_q}} \lambda^{l,l}(a_o, a_q).
\end{aligned}$$

Lo cual prueba la proposición. ■

Proposición 1.3.2. Sea $PD1_j^l(a_q)$ el número de pasajeros que abandonan el j -ésimo tren para abandonar el sistema en $a_q \in L_l$ tales que requirieron transbordar una vez, entonces

$$E\left(PD1_j^l(a_q)\right) = \sum_{\substack{a_{p_l} \in L_l \cap \mathcal{EC}, \\ a_{p_l} < a_q}} \sum_{L_m \in \mathcal{I}(L_l, a_{p_l})} \mu^m \omega^{m,l},$$

donde

$$\omega^{m,l} = E\left(NT_{a_{q_m}}^m\left(V_j^l(a_{p_l})\right) - NT_{a_{q_m}}^m\left(V_{j-1}^l(a_{p_l})\right)\right) \sum_{\substack{a_o \in L_m, \\ a_o < a_{q_m}}} \lambda^{m,l}(a_o, a_q).$$

Demostración. $PD1_j^l(a_q)$ satisface la siguiente igualdad:

$$PD1_j^l(a_q) = \sum_{\substack{a_{p_l} \in L_l \cap \mathcal{EC}, \\ a_{p_l} < a_q}} \sum_{L_m \in \mathcal{I}(L_l, a_{p_l})} \omega_*^{m,l},$$

donde

$$\omega_*^{m,l} = \sum_{\substack{a_o \in L_m, \\ a_o < a_{q_m}}} \sum_{k=NT_{a_{q_m}}^m(V_j^l(a_{p_l}))}^{NT_{a_{q_m}}^m(V_j^l(a_{p_l}))} \left[NP_{a_o, a_q}^{m,l}\left(V_k^m(a_o)\right) - NP_{a_o, a_q}^{m,l}\left(V_{k-1}^m(a_o)\right) \right].$$

Por lo que

$$E\left(PD1_j^l(a_q)\right) = \sum_{\substack{a_{p_l} \in L_l \cap \mathcal{EC}, \\ a_{p_l} < a_q}} \sum_{L_m \in \mathcal{I}(L_l, a_{p_l})} E(\omega_*^{m,l}).$$

Luego

$$\begin{aligned}
 E\left(\omega_*^{m,l}\right) &= \sum_{\substack{a_o \in L_m, \\ a_o < a_{q_m}}} E\left(\sum_{k=NT_{a_{q_m}}^m(V_j^l(a_{p_l}))+1}^{NT_{a_{q_m}}^m(V_j^l(a_{p_l}))} \left[NP_{a_o, a_q}^{m,l}(V_k^m(a_o)) - NP_{a_o, a_q}^{m,l}(V_{k-1}^m(a_o)) \right] \right) \\
 &= \sum_{\substack{a_o \in L_m, \\ a_o < a_{q_m}}} \lambda^{m,l}(a_o, a_q) \mu^m E\left(NT_{a_{q_m}}^m(V_j^l(a_{p_l})) - NT_{a_{q_m}}^m(V_{j-1}^l(a_{p_l})) \right) \\
 &= \mu^m E\left(NT_{a_{q_m}}^m(V_j^l(a_{p_l})) - NT_{a_{q_m}}^m(V_{j-1}^l(a_{p_l})) \right) \sum_{\substack{a_o \in L_m, \\ a_o < a_{q_m}}} \lambda^{m,l}(a_o, a_q) \\
 &= \mu^m \omega^{m,l},
 \end{aligned}$$

donde

$$\omega^{m,l} = E\left(NT_{a_{q_m}}^m(V_j^l(a_{p_l})) - NT_{a_{q_m}}^m(V_{j-1}^l(a_{p_l})) \right) \sum_{\substack{a_o \in L_m, \\ a_o < a_{q_m}}} \lambda^{m,l}(a_o, a_q).$$

Lo cual prueba la proposición. ■

Proposición 1.3.3. *Sea $PD2_j^l(a_q)$ el número de pasajeros que abandonan el j -ésimo tren para abandonar el sistema en $a_q \in L_l$ tales que requirieron transbordar dos veces, entonces*

$$E\left(PD2_j^l(a_q)\right) = \sum_{\substack{a_{p_l} \in L_l \cap \mathcal{E}\mathcal{C}, \\ a_{p_l} < a_q}} \sum_{L_m \in \mathcal{I}(L_l, a_{p_l})} \sum_{\substack{a_{p_m} \in L_m \cap \mathcal{E}\mathcal{C}, \\ a_{p_m} < a_{q_m}}} \sum_{L_n \in \mathcal{I}(L_m, a_{p_m})} \mu^n \omega^{n,m,l},$$

donde

$$\omega^{n,m,l} = E\left(NT_{a_{q_n}}^n(V_{N_j}^m(a_{p_m})) - NT_{a_{q_n}}^n(V_{N_{j-1}}^m(a_{p_m})) \right) \sum_{\substack{a_o \in L_n, \\ a_o < a_{q_n}}} \lambda^{n,l}(a_o, a_q),$$

con $N_j = NT_{a_{q_m}}^m(V_j^l(a_{p_l}))$ y $N_{j-1} = NT_{a_{q_m}}^m(V_{j-1}^l(a_{p_l}))$.

Demostración. $PD2_j^l(a_q)$ satisface la siguiente igualdad:

$$PD2_j^l(a_q) = \sum_{\substack{a_{p_l} \in L_l \cap \mathcal{E}\mathcal{C}, \\ a_{p_l} < a_q}} \sum_{L_m \in \mathcal{I}(L_l, a_{p_l})} \sum_{\substack{a_{p_m} \in L_m \cap \mathcal{E}\mathcal{C}, \\ a_{p_m} < a_{q_m}}} \sum_{L_n \in \mathcal{I}(L_m, a_{p_m})} \omega_*^{n,m,l},$$

donde

$$\omega_*^{n,m,l} = \sum_{\substack{a_o \in L_n, \\ a_o < a_{q_n}}} \sum_{k=NT_{a_{q_m}}^m(V_{j-1}^l(a_{p_l}))+1}^{NT_{a_{q_m}}^m(V_j^l(a_{p_l}))} \sum_{i=NT_{a_{q_n}}^n(V_{k-1}^m(a_{p_m}))}^{NT_{a_{q_n}}^n(V_k^m(a_{p_m}))} \left[NP_{a_o, a_q}^{n,l}(V_i^n(a_o)) - NP_{a_o, a_q}^{n,l}(V_{i-1}^n(a_o)) \right].$$

Entonces para calcular $E\left(PD1_j^l(a_q)\right)$ basta calcular $E\left(\omega_*^{n,m,l}\right)$. Para esto, procedemos de la misma forma que en la demostración de la proposición anterior y obtenemos que

$$E\left(\omega_*^{n,m,l}\right) = \mu^n E\left(\sum_{k=NT_{a_{q_m}}^m(V_{j-1}^l(a_{p_l}))+1}^{NT_{a_{q_m}}^m(V_j^l(a_{p_l}))} \left(NT_{a_{q_n}}^n(V_k^m(a_{p_m})) - NT_{a_{q_n}}^n(V_{k-1}^m(a_{p_m})) \right) \right) \sum_{\substack{a_o \in L_n, \\ a_o < a_{q_n}}} \lambda^{n,l}(a_o, a_q)$$

$$\begin{aligned}
 &= \mu^n E \left(NT_{a_{q_n}}^n (V_{N_j}^m(a_{p_m})) - NT_{a_{q_n}}^n (V_{N_{j-1}}^m(a_{p_m})) \right) \sum_{\substack{a_o \in L_n, \\ a_o < a_{q_n}}} \lambda^{n,l}(a_o, a_q) \\
 &= \mu^n \omega^{n,m,l},
 \end{aligned}$$

donde

$$\omega^{n,m,l} = E \left(NT_{a_{q_n}}^n (V_{N_j}^m(a_{p_m})) - NT_{a_{q_n}}^n (V_{N_{j-1}}^m(a_{p_m})) \right) \sum_{\substack{a_o \in L_n, \\ a_o < a_{q_n}}} \lambda^{n,l}(a_o, a_q),$$

con $N_j = NT_{a_{q_m}}^m (V_j^l(a_{p_l}))$ y $N_{j-1} = NT_{a_{q_m}}^m (V_{j-1}^l(a_{p_l}))$.

Lo cual prueba la proposición. ■

Por consiguiente, $PDA_j^l(a_q) = PDI_j^l(a_q) + PD1_j^l(a_q) + PD2_j^l(a_q) + \dots$ y su valor esperado es calculado usando las proposiciones anteriores.

El mismo razonamiento utilizado en estas proposiciones, se utiliza para analizar el número de pasajeros que abandonan el tren, para abandonar el sistema, y que requirieron hacer tres transbordos o más.

1.4. Función de desempeño

El vector de control $\mu = (\mu_s^1, \mu_s^2, \dots, \mu_s^{NTL})$ es el parámetro que usamos en el presente trabajo para proponer un mecanismo que evalúe el desempeño global del sistema Metro. Este desempeño global, denotado por $F(\mu)$, lo definimos como el costo monetario por cada jornada laboral J del sistema Metro, y es medido en términos de dos factores: el costo operativo y el costo social, denotados por $CO(\mu)$ y $CS(\mu)$, respectivamente. Cuyos términos se relacionan como

$$F(\mu) = CO(\mu) + CS(\mu).$$

$CO(\mu)$ es considerado como el costo diario global que el sistema Metro debe solventar para poder ofrecer el servicio: gastos como salarios de los trabajadores (personal de limpieza, operarios, secretarias, vigilantes, coordinadores, directivos, etc.), servicios básicos (pagos de agua, electricidad, etc.), combustibles, papelería, entre otros. Estos gastos se asocian a un costo general por cada línea y por cada tren, denotado por k^l , de tal manera que el costo operativo es lineal respecto al número de viajes hechos, así,

$$CO(\mu) = \sum_{l=1}^{NTL} k^l \sum_{s=1}^S \frac{T_s}{\mu_s^l}, \quad (1.9)$$

donde T_s/μ_s^l es el número de viajes hechos por la línea l (desde el andén inicial a_1 hasta el andén final a_{NTA^l}) en el segmento s .

$CS(\mu)$ considera el tiempo productivo que pierden, en el andén, los pasajeros mientras esperan al tren, cuyo tiempo llamaremos tiempo de espera¹⁴. Este tiempo de espera es asociado a cada tren y a cada andén de cada línea.

Sea $W_j^l(a_q)$ el tiempo de espera asociado al j -ésimo tren en $a_q \in L_l$. A cada unidad de $W_j^l(a_q)$ se le asigna un costo monetario c , por tanto, el costo social es igual a

$$CS(\mu) = c \sum_{s=1}^S \sum_{l=1}^{NLT} \sum_{j=1}^{NTT_s^l} \sum_{q=1}^{NTA^l} E\left(W_j^l(a_q)\right).$$

1.4.1. Tiempo de espera

Si el andén está en una estación de correspondencia, entonces los pasajeros que esperan en el andén son pasajeros que arribaron tanto por ingreso al sistema como por transbordo; con la diferencia de que el valor esperado del tiempo de espera por ingreso al sistema, es analíticamente asequible; mientras que el del tiempo de espera por transbordo, es calculado a través de simulación.

Supongamos que $a_q \in L_l$ se encuentra en una estación normal, entonces todos los pasajeros que arriban a a_q provienen de la calle, por tanto, son pasajeros que arriban por ingreso al sistema y, como vimos en la sección 1.2, éstos arriban de acuerdo a un proceso de Poisson $NP_{a_q}^l(t)$ con intensidad $\lambda^l(a_q)$.

Sea $WI_j^l(a_q)$ el tiempo de espera de los pasajeros por ingreso al sistema. Los pasajeros por ingreso al sistema que abordaron el j -ésimo tren, arribaron a $a_q \in L_l$ entre $\left(V_{j-1}^l(a_q), V_j^l(a_q)\right]$, por lo que su tiempo de espera termina al tiempo $V_j^l(a_q)$, es decir,

$$WI_j^l(a_q) = \sum_{n=NP_{a_q}^l(V_{j-1}^l(a_q))+1}^{NP_{a_q}^l(V_j^l(a_q))} \left(V_j^l(a_q) - \tau_n\right), \quad (1.10)$$

donde τ_n es el tiempo en que ocurrió el n -ésimo arribo por ingreso al sistema.

Sin pérdida de generalidad, puesto que tanto la línea como el andén están fijos, para una notación más ligera, en lo siguiente omitiremos la notación a_q y l para el proceso $NP_{a_q}^l(t)$ y para los tiempos $V_j^l(a_q)$.

Proposición 1.4.1. *Sea $WI_j^l(a_q)$ el tiempo de espera de los pasajeros por ingreso al sistema que abordaron el j -ésimo tren en $a_q \in L_l$, entonces*

$$E\left(WI_j^l(a_q)\right) = \frac{\lambda^l(a_q)}{2} E\left(Y_j^l(a_q)^2\right).$$

Demostración. De la ecuación (1.10) obtenemos que

$$E\left(WI_j^l(a_q)\right) = E\left(\sum_{n=NP(V_{j-1})+1}^{NP(V_j)} (V_j - \tau_n)\right)$$

¹⁴Algunos autores llaman a éste “tiempo de ocio”.

$$\begin{aligned}
 &= E \left(E \left[\sum_{n=NP(V_{j-1})+1}^{NP(V_j)} (V_j - \tau_n) \middle| V_j - V_{j-1} \right] \right) \\
 &= E (E (V_j - \tau_n) E (NP(V_j) - NP(V_{j-1}))) \\
 &= E \left(\frac{V_j - V_{j-1}}{2} \lambda^l(a_q) (V_j - V_{j-1}) \right) \\
 &= \frac{\lambda^l(a_q)}{2} E \left((V_j - V_{j-1})^2 \right) \\
 &= \frac{\lambda^l(a_q)}{2} E \left(Y_j^l(a_q)^2 \right).
 \end{aligned}$$

En esta demostración consideramos que dado que han arribado $NP(V_j) - NP(V_{j-1})$ pasajeros en el intervalo de tiempo $(V_{j-1}, V_j]$, los tiempos τ_n se distribuyen de manera uniforme sobre el intervalo (V_{j-1}, V_j) . Además, usamos el hecho de que el proceso de llegadas de trenes V_j es independiente del proceso de arribos de pasajeros $NP(t)$. ■

Ahora supongamos que $a_q \in L_l$ es tal que $a_q \in \mathcal{EC}$, entonces los pasajeros arriban a a_q por ingreso al sistema o por transbordo (véase figura 1.2).

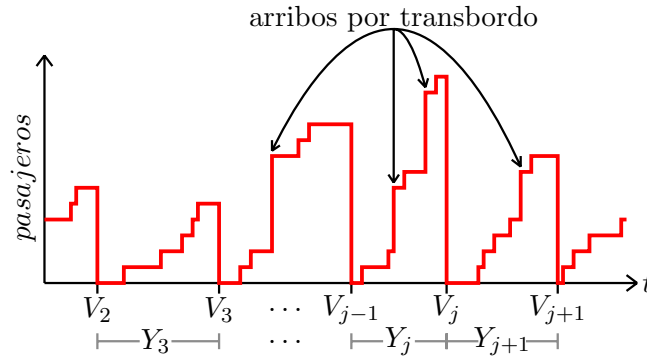


Figura 1.2: Pasajeros esperando en $a_q \in L_l$ cuando $a_q \in \mathcal{EC}$.

Sea $WT_j^l(a_q)$ el tiempo de espera de los pasajeros que arriban por transbordo a $a_q \in L_l$ asociado al j -ésimo tren, entonces

$$WT_j^l(a_q) = \sum_{L_m \in \mathcal{I}(L_l, a_q)} \sum_{k=NT_{a_{qm}}^m(V_{j-1})+1}^{NT_{a_{qm}}^m(V_j)} PT_k^{m,l} [V_j - S_k^m(a_{qm})],$$

donde $S_k^m(a_{qm})$ es la hora de llegada del k -ésimo tren a $a_{qm} \in L_m$ (véase ¹⁵) y $PT_k^{m,l}$ es el número de pasajeros que transbordan de L_m a L_l y que llegaron en el k -ésimo tren¹⁶ (para más detalles véanse Heidergott & Vázquez-Abad [2009] y Vázquez-Abad & Zubietta [2000]). $E(WT_j^l(a_q))$ es calculado a través de simulación.

De modo general, para cualquier $a_q \in L_l$ podemos expresar el tiempo de espera $W_j^l(a_q)$ por

$$W_j^l(a_q) = WI_j^l(a_q) + WT_j^l(a_q)I_{\{a_q \in \mathcal{EC}\}}. \quad (1.11)$$

¹⁵Como se vio en la demostración de la proposición 1.1.2 $S_k^m(a_{qm}) = V_k^m(a_{qm}) - P^m(a_{qm}) - \delta_k^m(a_{qm})$.

¹⁶Se entiende que $PT_k^{m,l}$ es la suma de los pasajeros que transbordan de L_m a L_l , es decir, $PT_k^{m,l} = PAI_k^{m,l} + P1_k^{m,l} + P2_k^{m,l} + \dots$

Luego, $E\left(W_j^l(a_q)\right)$ es calculado usando la proposición 1.4.1 para $E\left(WI_j^l(a_q)\right)$ y una simulación para calcular $E\left(WT_j^l(a_q)\right)$.

En este capítulo presentamos el modelo matemático que permite replicar el flujo de pasajeros y el flujo de trenes, así como su interacción a través de las estaciones de correspondencia¹⁷. También analizamos la composición y evaluación de la función de desempeño del sistema Metro, la cual depende del vector de control μ .

Es importante mencionar que la simulación de un sistema Metro utilizando los valores esperados vistos en este capítulo la llamaremos simulación parcial; en el capítulo 3 veremos una simulación alternativa, que llamaremos simulación global, la cual no utiliza valores esperados sino que genera cada uno de los arribos de pasajeros.

En el siguiente capítulo nos enfocaremos en el problema de optimización de la función de desempeño $F(\mu)$; básicamente, en el estudio de métodos para la optimización de sistemas de eventos discretos a través de simulación.

¹⁷Para un estudio más detallado de la teoría que fundamenta las demostraciones presentadas en este capítulo véanse Freund et al (2000), Hernández (1981), Mood y Graybill (1970), Nelson (2002), Quevedo y Pérez (2008) y Ross (2007).

Capítulo 2

Optimización de la función de desempeño

Optimizar recursos, minimizar costos y maximizar beneficios son cuestiones que siempre están presentes en el desarrollo de nuevas tecnologías y los sistemas Metro de transporte urbano no están exentos de ello: para los pasajeros lo ideal sería que al arribar al andén el siguiente tren pase casi inmediatamente, sin embargo, incrementar el número de trenes para satisfacer este propósito, aumentaría prohibitivamente el costo operativo del Metro; una solución intermedia entre los intereses de los pasajeros y el Metro, es administrar eficientemente el sistema utilizando la información de cada línea para tomar decisiones en tiempo real, que permita sincronizar el flujo de trenes con el flujo de pasajeros de manera óptima, esto es, que el costo sea mínimo. Como vimos en la sección 1.4, éstos intereses pueden ser resumidos a través de la función de desempeño; por lo que administrar el sistema de manera eficientemente, se reduce a optimizar $F(\mu)$.

Optimizar la función de desempeño, $F(\mu) : \mathbb{R}^{NTL} \mapsto \mathbb{R}$, equivale a resolver el siguiente problema de optimización con restricciones:

$$\begin{aligned} \text{mín} \quad & F(\mu) = CO(\mu) + CS(\mu) \\ \text{s.a.} \quad & \mu \in \Omega, \end{aligned}$$

con $\Omega = \{\mu \in \mathbb{R}^{NTL} \mid \mu = (\mu^1, \mu^2, \dots, \mu^{NLT}) \text{ con } \mu^i \geq \mu_{seg}, \text{ para } i = 1 : NTL\}$.

Es fácil ver, intuitivamente, que un incremento en μ indudablemente reduciría el costo operativo; pero a la vez, incrementaría el costo social cuando el resto de los parámetros permanecen constantes. El problema para optimizar $F(\mu)$ es que ésta no puede diferenciarse analíticamente, ya que para los andenes en estaciones de correspondencia no contamos con una expresión analítica de $E(WT_j^l(a_q))$ (véase ¹). Por lo que el problema radica en encontrar las sensibilidades de $E(WT_j^l(a_q))$; lo cual implica encontrar el gradiente de una función que es estimada a través de simulación.

Particularmente, veremos cuatro métodos de aproximación para estimar el gradiente de una función estocástica a través de simulación: el método de diferencias finitas, el método de perturbación simultánea, análisis

¹Vázquez-Abad & Zubieta (2005), bajo ciertos supuestos adicionales, proponen una expresión analítica para $E(WT_j^l(a_q))$; en el presente trabajo no abordamos esta expresión, sino que obtenemos valores de $E(WT_j^l(a_q))$ a través de simulación.

de perturbación infinitesimal (IPA) y el método de razón de verosimilitud (también conocido como el método de la función *score*). Estos métodos proporcionan estimadores eficientes, aunque no todos pueden emplearse de manera general, al gradiente y pueden ser utilizados en el algoritmo del gradiente proyectado en el contexto de optimización estocástica² para la optimización de sistemas de eventos discretos.

2.1. Método Monte-Carlo

El método Monte-Carlo o simulación Monte-Carlo es un método estadístico para aproximar el valor esperado de una función. Este método proporciona soluciones aproximadas a una amplia variedad de problemas estocásticos, ya sean dinámicos o estáticos, posibilitando la realización de experimentos con muestreos de números pseudo-aleatorios.

Por su clara analogía con los juegos de ruleta, como un generador simple de números “aleatorios”, el método Monte-Carlo fue llamado así en referencia al Casino de Monte-Carlo, inaugurado en 1861 en Mónaco: “*La capital del juego de azar*”. Sin embargo, el nombre y el desarrollo sistemático del método datan aproximadamente de 1944 con el desarrollo de la computadora electrónica.

El uso de este método como una herramienta de investigación, proviene del trabajo realizado por John von Neumann y Stanislaw Ulam en el desarrollo de la bomba atómica durante la Segunda Guerra Mundial en E.U., el cual involucraba la simulación de problemas probabilísticos concernientes a la difusión de neutrones aleatorios en material nuclear. No obstante, fue hasta 1948 que Harris y Herman Kahn desarrollaron de manera sistemática este método. En el mismo año, Enrico Fermi, Nicholas Metropolis y Ulam, obtuvieron estimadores para los valores característicos de la ecuación de Schrödinger para la captura de neutrones a nivel nuclear usando este método.

2.1.1. Sistematización

Las bases teóricas de la simulación Monte-Carlo se tienen en la ley de los grandes números y en el Teorema Central del Límite. Estos dos resultados se enuncian de la siguiente manera: sea X una v.a. con distribución F y sean X_1, X_2, \dots, X_R una muestra de R v.a.i.i.d. con distribución F y con media y varianza finitas. La ley fuerte de los grandes números establece que

$$\mathcal{P} \left(\lim_{R \rightarrow \infty} \bar{X} = E(X) \right) = 1,$$

donde $\bar{X} = 1/R \sum_{i=1}^R X_i$. Es decir, con toda certeza, a largo plazo, el promedio de una sucesión de v.a.i.i.d. convergerá a su media. Y el Teorema Central del Límite asegura que

$$\lim_{R \rightarrow \infty} \frac{\bar{X} - E(X)}{\sigma/\sqrt{R}} = z,$$

²Cuando la función objetivo y/o restricciones del problema entrañan v.a., el tipo de optimización a realizar es una optimización estocástica, en la cual, la estimación de las derivadas es de vital importancia para la implementación computacional de sus algoritmos.

donde z es una v.a. tal que $z \sim normal(0, 1)$ y σ^2 es la varianza.

Luego, sea g una función continua tal que $E(g(X)) = \theta$ y $Var(g(X)) = \sigma^2$ entonces $\hat{\theta}_R = 1/R \sum_{r=1}^R g(X_r)$ es una v.a. y cuando R es grande satisface dos propiedades:

- i. θ y $\hat{\theta}_R$ están cerca con una probabilidad alta, y
- ii. $(\hat{\theta}_R - \theta) / (\sigma/\sqrt{R})$ se distribuye como una v.a. normal con media cero y varianza uno.

De lo cual se sigue que, con una confianza del $(1 - \alpha)100\%$,

$$\theta \in \left[\hat{\theta}_R - \frac{z_{1-\alpha/2}\sqrt{s^2}}{\sqrt{R}}, \hat{\theta}_R + \frac{z_{1-\alpha/2}\sqrt{s^2}}{\sqrt{R}} \right],$$

donde s^2 denota la varianza muestral y $z_{1-\alpha/2}$ es el cuantil $(1 - \alpha/2)$ de la distribución normal estándar. Por ejemplo: para $\alpha = 0.05$ tenemos $z_{1-\alpha/2} = z_{0.975} = 1.96$ (para más detalles véanse Saavedra e Ibarra [2008] y Rubinstein & Kroese [2008]).

2.1.2. Algoritmo

El algoritmo consta de R realizaciones en el que, en cada realización r , con $r = 1 : R$, se simula el comportamiento de un sistema cuyo desempeño es una función estocástica $f(x)$. Luego, se obtiene una sucesión de R v.a. para calcular, de manera recursiva, la media y la varianza muestrales, para posteriormente calcular el intervalo al 95% de confianza para $f(x)$.

Algoritmo del método Monte-Carlo.

Paso 1. Se inicializan las variables

$$\theta = 0 \text{ y } s = 0.$$

Paso 2. Para $r = 1 : R$ se calcula de forma iterativa

$$\theta = \theta + f(x) \text{ y } s = s + f(x)^2.$$

Paso 3. Se calcula la media y la varianza muestrales,

$$\hat{\theta}_R = \frac{\theta}{R} \text{ y } s^2 = \frac{s - \hat{\theta}_R^2 R}{R - 1}, \text{ respectivamente.}$$

Paso 4. Se determina el intervalo de confianza para $f(x)$ al 95%:

$$IC_{F(x)} = \left[\hat{\theta}_R - \frac{1.96\sqrt{s^2}}{\sqrt{R}}, \hat{\theta}_R + \frac{1.96\sqrt{s^2}}{\sqrt{R}} \right].$$

2.2. Optimización de sistemas de eventos discretos

De manera general, consideremos un sistema estocástico y supongamos que su desempeño puede ser resumido en un número real z , el cual, puede ser expresado como $E(Z)$ para alguna v.a. Z . Por ejemplo: en un cola, z puede

ser el tiempo promedio que un cliente pasa en cierto sistema de servicio; en un modelo de riesgo, z puede ser la probabilidad de que el monto de las reclamaciones ejercidas C , en un lapso de un año, exceda un valor determinado; o en un sistema de transporte urbano Metro, z puede ser la función de desempeño $f(\mu)$.

Obviamente z depende de ciertos parámetros. Por ejemplo: en una cola, z depende de la intensidad de arribos de clientes y de la distribución de los tiempos de servicio; en un modelo de riesgo, z depende de los parámetros de la distribución que tengan los reclamos; en un sistema Metro, z depende del parámetro de control μ . Luego, uno podría estar interesado no sólo en el desempeño del sistema $z = E(Z)$ sino también en sus sensibilidades, es decir, la tasa de cambio de la función objetivo con respecto al cambio en los parámetros o variables de control. El problema ahora es cómo estimar el gradiente de un sistema de eventos discretos, cuyo desempeño es calculado a través de simulación.

A continuación veremos un método de reducción de varianza que emplearemos, posteriormente, en los métodos de diferencias finitas y perturbación simultánea.

2.3. Números aleatorios comunes

Los métodos de reducción de varianza permiten reducir el tamaño del intervalo de confianza de la media de una v.a. que se estima por Monte-Carlo, sin incrementar el número de realizaciones, es decir, estos métodos mejoran la precisión de la estimación sin incrementar el costo computacional.

El método de números aleatorios comunes³ es uno de los métodos de reducción de varianza más usuales: este método consiste en realizar simulaciones para parámetros distintos, con el mismo conjunto de números aleatorios. Así, las diferencias en los desempeños simulados del sistema, se deben a los distintos parámetros y no a las condiciones experimentales.

Para mostrar la ventaja del uso de números aleatorios comunes en simulación, consideremos el siguiente ejemplo: sean z^* y z^+ las medias de las v.a.i. Z^* y Z^+ , respectivamente, que dependen de las mismas v.a. de entrada con $Var(Z^*) = \sigma^2$ y $Var(Z^+) = \sigma^2$. Supongamos que queremos estimar la diferencia

$$z = z^* - z^+$$

a través de simulación, donde $z^* = E(Z^*)$ y $z^+ = E(Z^+)$ con $Z^* = Z^*(\omega^*)$ y $Z^+ = Z^+(\omega^+)$ donde ω^* y ω^+ representan los conjuntos de números aleatorios utilizados para generar Z^* y Z^+ , respectivamente.

Entonces, podemos estimar $z = E(Z)$ a través de la media aritmética

$$\hat{z} = \frac{1}{R} \sum_{r=1}^R Z_r$$

donde cada Z_r es una realización independiente de

$$Z = Z^*(\omega^*) - Z^+(\omega^+)$$

³También conocido como muestreo correlacionado, simulación comparativa o cadenas sincronizadas.

y R es el número de realizaciones generadas de Z . El estimador \hat{z} , así definido, es llamado el estimador Crudo Monte-Carlo (CMC) (para más detalles véase Glasserman [2004]).

Si las muestras para las dos simulaciones se realizaran de manera independiente, esto es, con números aleatorios diferentes: Z^* y Z^+ serán independientes y su covarianza será igual a cero, por tanto, la varianza muestral de \hat{z} está dada por

$$\begin{aligned} \text{Var}(\hat{z}) &= \frac{1}{R} \text{Var}(Z) \\ &= \frac{1}{R} \text{Var}(Z^* - Z^+) \\ &= \frac{1}{R} [\text{Var}(Z^*) + \text{Var}(Z^+)] \\ &= \frac{2}{R} \sigma^2. \end{aligned}$$

Por el contrario, si las muestras están positivamente correlacionadas, la covarianza será positiva y la varianza de \hat{z} será menor, esto es, consideremos ahora

$$\omega^* = \omega^+. \quad (2.1)$$

Entonces Z^* y Z^+ ya no serán independientes puesto que utilizan el mismo conjunto de números aleatorios. Suponga ρ , con $\rho > 0$, como su coeficiente de correlación; entonces $\text{Cov}(Z^*, Z^+) = \rho\sigma^2$. Si calculamos nuevamente $\text{Var}(\hat{z})$, considerando esta dependencia, obtenemos que

$$\begin{aligned} \text{Var}(\hat{z}) &= \frac{1}{R} \text{Var}(Z^* - Z^+) \\ &= \frac{1}{R} [\text{Var}(Z^*) + \text{Var}(Z^+) - 2\text{Cov}(Z^*, Z^+)] \\ &= \frac{1}{R} [\sigma^2 + \sigma^2 - 2\rho\sigma^2] \\ &= \frac{2}{R} \sigma^2 (1 - \rho). \end{aligned}$$

Si comparamos estas varianzas podemos ver que la reducción de la varianza es, en efecto, obtenida cuando se utilizan números aleatorios comunes, ya que

$$(1 - \rho) < 1$$

bajo el supuesto de una correlación positiva, lo cual siempre se cumple por la relación (2.1).

En resumen, este método trata de inducir una correlación positiva entre los conjuntos de números aleatorios implicados en las dos simulaciones. Por otro lado, la implementación correcta de este método requiere la sincronización de los números aleatorios para las diferentes simulaciones. Es decir, un número aleatorio utilizado con un fin en una simulación deberá ser utilizado exactamente para el mismo fin en la otra simulación (para más detalles véanse Ross [1999] y Assmussen & Glynn [2007]).

2.4. Estimación de sensibilidades

Existen numerosas razones para estar interesados en las sensibilidades de una función. Por ejemplo: para identificar los parámetros que más impactan en el desempeño de un sistema; para evaluar los efectos de un cambio pequeño en alguna de sus variables independientes; en optimización estocástica para optimizar (maximizar o minimizar) el desempeño de un sistema.

Desde la perspectiva metodológica, existe una diferencia muy pequeña entre los problemas uni-dimensionales y los multi-dimensionales. Por lo tanto, para simplificar la notación, en ocasiones, nos enfocaremos en un contexto uni-dimensional, con θ definido en un conjunto Θ , el cual, está contenido en \mathbb{R} o en \mathbb{R}^p dependiendo del contexto dimensional.

2.4.1. Diferencias finitas

Sabemos por el método convencional de diferencias finitas, que el error de aproximación por diferencia central es menor que el error de aproximación por diferencias progresivas o regresivas; ya que el error de aproximación por diferencia central, es del orden h^2 y los errores de aproximación por diferencias progresiva y regresiva, son del orden h (donde h es el incremento utilizado). Por lo que es preferible, para diferenciación numérica, utilizar aproximaciones por diferencia central, independientemente de que los errores de aproximación disminuyen a medida que h disminuye su valor. Por esta razón, en lo posterior sólo consideraremos diferencias centrales cuando hablemos de diferencias finitas.

Ahora veremos el método de diferencias finitas (FD por sus siglas en inglés *finite difference*) en el contexto de simulación. Supongamos que para cada $\theta \in \mathbb{R}$ se puede generar una v.a. $Z(\theta)$ tal que $z(\theta) = E(Z(\theta))$. Se desea obtener un estimador simulado de $z'(\theta)$.

La idea principal del método de FD es usar

$$f'(x) = \lim_{h \rightarrow 0} \frac{f\left(x + \frac{h}{2}\right) - f\left(x - \frac{h}{2}\right)}{h}$$

para la derivada de una función determinista $f(x)$. En el contexto de simulación, este método sugiere calcular un estimador CMC usando

$$D(\theta) = \frac{Z\left(\theta + \frac{h}{2}\right) - Z\left(\theta - \frac{h}{2}\right)}{h}$$

(al cual llamaremos estimador FD) para alguna h pequeña, donde, en un principio, se asume que $Z(\theta + h/2)$ y $Z(\theta - h/2)$ son independientes. Entonces, siguiendo el método CMC, la media aritmética

$$\hat{z}'(\theta) = \frac{1}{R} \sum_{r=1}^R D^r(\theta)$$

es un estimador de $z'(\theta)$, donde cada $D^r(\theta)$ es una realización independiente de $D(\theta)$. Cuando $R \rightarrow \infty$ necesariamente $h \rightarrow 0$, pero a medida que h disminuye su valor: la varianza de la aproximación se incrementa (Asmussen & Glynn [2007]). La siguiente proposición nos dice que, para una aproximación óptima, h debe ser del orden $R^{-1/6}$.

Proposición 2.4.1. *Considere el estimador FD*

$$D(\theta) = \frac{Z\left(\theta + \frac{h}{2}\right) - Z\left(\theta - \frac{h}{2}\right)}{h}$$

con $Z(\theta + h/2)$ y $Z(\theta - h/2)$ independientes. Entonces, la raíz del error cuadrático medio

$$[E([\hat{z}'(\theta) - z'(\theta)]^2)]^{1/2}$$

es minimizado cuando

$$h = h(R) = \frac{1}{R^{1/6}} \frac{[576 \text{Var}(Z(\theta))]^{1/6}}{|z'''(\theta)|^{1/3}}. \quad (2.2)$$

Demostración. Claramente $\text{Var}(Z(\theta \pm h/2)) \sim \text{Var}(Z(\theta))$ (véase ⁴) y usando la independencia de $Z(\theta + h/2)$ y $Z(\theta - h/2)$ obtenemos que $\text{Var}(D(\theta)) \sim 2\text{Var}(Z(\theta))/h^2$. De esto último y considerando el hecho de que las realizaciones de $D(\theta)$ son independientes entre sí, obtenemos que

$$\text{Var}(\hat{z}'(\theta)) \sim \frac{2}{Rh^2} \text{Var}(Z(\theta)) \quad (2.3)$$

y

$$\begin{aligned} E(\hat{z}'(\theta)) &= E\left(\frac{Z\left(\theta + \frac{h}{2}\right) - Z\left(\theta - \frac{h}{2}\right)}{h}\right) \\ &\approx E\left(Z'(\theta) + \frac{h^2}{24}Z'''(\theta)\right) \\ &\approx z'(\theta) + \frac{h^2}{24}z'''(\theta) \end{aligned} \quad (2.4)$$

con $z'''(\theta) = E(Z'''(\theta))$.

Por otro lado, dado x una v.a. y y una constante, tenemos que

$$\begin{aligned} E([x - y]^2) &= E(x^2) - E(x)^2 + E(x)^2 - 2yE(x) + y^2 \\ &= \text{Var}(x) + [E(x) - y]^2. \end{aligned}$$

Si tomamos $x = \hat{z}'(\theta)$ y $y = z'(\theta)$ y usamos (2.3) y (2.4) obtenemos que

$$\begin{aligned} E([\hat{z}'(\theta) - z'(\theta)]^2) &= \text{Var}(\hat{z}'(\theta)) + [E(\hat{z}'(\theta)) - z'(\theta)]^2 \\ &\sim \left(\frac{z'''(\theta)}{24}\right)^2 h^4 + \frac{2}{Rh^2} \text{Var}(Z(\theta)). \end{aligned}$$

⁴En este contexto \sim significa que $\text{Var}(Z(\theta \pm h/2))/\text{Var}(Z(\theta)) \rightarrow 1$ cuando $h \rightarrow 0$; en otros casos, \sim denota la distribución de una v.a.

Consideremos ahora la función $g(x) = bx^2 + ax^{-1}$ con $b > 0$, la cual alcanza su valor mínimo cuando $x = [a/(2b)]^{1/3}$. Si tomamos $x = h^2$, $b = (z'''(\theta)/24)^2$ y $a = 2\text{Var}(Z(\theta))/R$, entonces $g(h^2) = E([\hat{z}'(\theta) - z'(\theta)]^2)$. Por tanto, la raíz del error cuadrático medio alcanza su mínimo valor cuando

$$h = \frac{1}{R^{1/6}} \frac{[576\text{Var}(Z(\theta))]^{1/6}}{|z'''(\theta)|^{1/3}}.$$

Lo cual prueba la proposición. ■

Esta proposición es, en cierto sentido, bastante académica, ya que el valor óptimo de h es desconocido puesto que si $z'(\theta)$ es desconocido es de esperarse que $z'''(\theta)$ lo sea aún más. Sin embargo, una característica importante es que el error cuadrático medio disminuye cuando el número de realizaciones R aumenta, esto es, la tasa de convergencia de $\hat{z}'(\theta)$ a $z'(\theta)$, en la norma euclidiana, es del orden $R^{-1/3}$ (Asmussen & Glynn [2007]) y cuando R aumenta, dada su dependencia en (2.2), h se vuelve más pequeña aún, a una tasa $R^{-1/6}$.

Los métodos de aproximación estocástica pueden desempeñar una mejora considerable si se utilizan números aleatorios comunes, lo cual puede provocar que la tasa de convergencia sea del orden $R^{-1/2}$ (cuya tasa es conocida como tasa canónica y es la mejor tasa de convergencia que se pueda obtener).

Para generalizar el método de aproximación estocástica por FD, utilizando números aleatorios comunes, supongamos el caso en el que $\theta \in \mathbb{R}^p$ con $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ entonces

$$\nabla z(\theta) = \left(\frac{\partial}{\partial \theta_1} z(\theta), \frac{\partial}{\partial \theta_2} z(\theta), \dots, \frac{\partial}{\partial \theta_p} z(\theta) \right)$$

y el estimador $D(\theta)$ es un vector de dimensión p cuyas componente $D_i(\theta)$, con $i = 1 : p$, son de la forma

$$D_i(\theta) = \frac{Z\left(\theta + \frac{h}{2} e_i; \omega_i\right) - Z\left(\theta - \frac{h}{2} e_i; \omega_i\right)}{h}, \quad (2.5)$$

donde e_i es el i -ésimo vector canónico de dimensión p y ω_i denota el conjunto de números aleatorios utilizado para ambas simulaciones, es decir, $Z(\theta + h/2 e_i)$ y $Z(\theta - h/2 e_i)$ son generadas utilizando números aleatorios comunes. Luego,

$$\frac{\partial}{\partial \theta_i} z(\theta) = \frac{\partial}{\partial \theta_i} E(Z(\theta)) \approx \frac{1}{R} \sum_{r=1}^R D_i^r(\theta)$$

donde cada $D_i^r(\theta)$ es una realización de $D_i(\theta)$.

2.4.2. Perturbación simultánea

La aproximación estocástica por FD y la aproximación estocástica por perturbación simultánea (SP por sus siglas en inglés *simultaneous perturbation*) utilizan sólo estimaciones de la función de desempeño, esto es, no requieren un conocimiento detallado del sistema dinámico, ni de la distribución de sus entradas; por lo que se pueden aplicar a cualquier sistema que pueda ser simulado.

Una de las ventajas de SP, como veremos a continuación, sobre FD es que FD requiere de $2p$ simulaciones para la estimación del gradiente, mientras que SP requiere sólo de dos simulaciones para estimar el gradiente, independientemente de la dimensión del parámetro θ y por tanto requiere, significativamente, menos simulaciones que FD para estimar gradientes en gran dimensión.

Cuando p es grande; FD puede llegar a ser prohibitivamente caro, computacionalmente, mientras que SP permite estimar el gradiente con sólo dos simulaciones, en lugar de $2p$ como FD.

Pensando en un contexto iterativo de un algoritmo de optimización, la aproximación estocástica por SP consiste en lo siguiente: sea $\{\Delta_n\}_{n=1,2,\dots}$ una sucesión de vectores de perturbación de dimensión p i.i.d., cuyas componentes $\{\Delta_{n_i}\}_{i=1:p}$ son v.a.i.i.d. con función de densidad simétrica, $E(\Delta_{n_i}) = 0$ y $E(\Delta_{n_i}^{-1})$ uniformemente acotada y sea $\{c_n\}_{n=1,2,\dots}$ una sucesión positiva que converge a cero. La i -ésima componente del estimador $D(\theta_n)$, utilizando SP, está dada por:

$$D_i(\theta_n) = \frac{Z_n^+ - Z_n^-}{2c_n \Delta_{n_i}}, \quad (2.6)$$

con

$$Z_n^+ = Z(\theta_n + c_n \Delta_n; \omega_n^+)$$

y

$$Z_n^- = Z(\theta_n - c_n \Delta_n; \omega_n^-),$$

donde ω_n^+ y ω_n^- denotan los conjuntos de números aleatorios implicados en la n -ésima iteración del algoritmo, y Z_n^+ y Z_n^- son los desempeños del sistema evaluados en el valor de θ_n perturbado simultáneamente en todas sus componentes: de aquí el nombre del método. De igual manera que en FD, si se emplea el método de números aleatorios comunes entonces $\omega_n^+ = \omega_n^-$ y lo denotamos simplemente por ω_n .

Es importante remarcar que cada simulación $Z(\theta_n; \omega)$ es computacionalmente caro en comparación con la generación de las Δ 's. Observemos que el estimador de FD (ecuación (2.5)) requiere, en el numerador, un par de simulaciones para cada componente del gradiente a estimar, por tanto, FD requiere de $2p$ simulaciones; mientras que el estimador SP (ecuación (2.6)) usa el mismo par de simulaciones, en el numerador, para cada componente del gradiente a estimar y en su lugar sólo cambia el denominador. Así, SP requiere sólo dos simulaciones para estimar el gradiente.

Nuevamente, en este método se considera un estimador CMC de $z'(\theta_n)$ usando el estimador de SP definido en (2.6) y el método de números aleatorios comunes. Esto es,

$$\frac{\partial}{\partial \theta_i} z(\theta) = \frac{\partial}{\partial \theta_i} E(Z(\theta)) \approx \frac{1}{R} \sum_{r=1}^R D_i^r(\theta)$$

donde cada $D_i^r(\theta)$ es una realización de $D_i(\theta)$.

Consideremos ahora \hat{g}_{n_i} como la i -ésima componente de \hat{g}_n (el estimador de $\nabla E(Z(\theta_n))$), esto es, $\hat{g}_n = \hat{\nabla} z(\theta_n)$ y consideremos

$$b_n = E(\hat{g}_n) - g_n,$$

el sesgo del gradiente estimado; donde g_n denota el gradiente de $E(Z(\theta_n))$ y supongamos $E(Z)$ continuamente diferenciable.

Lema 2.4.1. Sean $\{\Delta_{n_i}\}_{i=1:p}$ mutuamente independientes con media cero, segundos momentos acotados y $E(\Delta_{n_i}^{-1})$ uniformemente acotado. Entonces $b_n \rightarrow 0$ con probabilidad 1.

Demostración. Si consideramos la definición del producto punto obtenemos que

$$\begin{aligned}
 E\left(\frac{g_n^t \Delta_n}{\Delta_{n_i}}\right) &= E\left(\Delta_{n_i}^{-1} \sum_{j=1}^p g_{n_j} \Delta_{n_j}\right) \\
 &= E\left(\Delta_{n_i}^{-1} \left[g_{n_i} \Delta_{n_i} + \sum_{j=1}^{i-1} g_{n_j} \Delta_{n_j} + \sum_{j=i+1}^p g_{n_j} \Delta_{n_j} \right]\right) \\
 &= E\left(g_{n_i} + \Delta_{n_i}^{-1} \left[\sum_{j=1}^{i-1} g_{n_j} \Delta_{n_j} + \sum_{j=i+1}^p g_{n_j} \Delta_{n_j} \right]\right) \\
 &= g_{n_i} + E(\Delta_{n_i}^{-1}) \left[\sum_{j=1}^{i-1} g_{n_j} E(\Delta_{n_j}) + \sum_{j=i+1}^p g_{n_j} E(\Delta_{n_j}) \right] \\
 &= g_{n_i}.
 \end{aligned} \tag{2.7}$$

En la penúltima igualdad utilizamos la linealidad de la esperanza, el hecho de que g_{n_i} es una constante y la independencia de $\{\Delta_{n_i}\}_{i=1:p}$ y para obtener la última igualdad consideramos las hipótesis de que $E(\Delta_{n_j}) = 0$ y que $E((\Delta_{n_i})^{-1})$ es uniformemente acotado, lo cual nos permite eliminar el segundo sumando.

Por otro lado, tenemos que

$$\begin{aligned}
 E\left(\frac{\Delta_n^2}{\Delta_{n_i}}\right) &= E\left(\Delta_{n_i}^{-1} \sum_{j=1}^p \Delta_{n_j}^2\right) \\
 &= E(\Delta_{n_i}) + E(\Delta_{n_i}^{-1}) \left[\sum_{j=1}^{i-1} E(\Delta_{n_j}^2) + \sum_{j=i+1}^p E(\Delta_{n_j}^2) \right] \\
 &= E(\Delta_{n_i}^{-1}) \left[\sum_{j=1}^{i-1} E(\Delta_{n_j}^2) + \sum_{j=i+1}^p E(\Delta_{n_j}^2) \right].
 \end{aligned} \tag{2.8}$$

Como $E(\Delta_{n_i}^{-1})$ es uniformemente acotado y Δ_{n_j} tienen segundos momentos finitos, entonces $E(\Delta_n^2/\Delta_{n_i})$ está acotado superiormente.

Ahora, si utilizamos el estimador de SP y aplicamos la expansión de $E(Z_n^+)$ y $E(Z_n^-)$ en su serie de Taylor obtenemos que

$$\begin{aligned}
 E(\hat{g}_{n_i}) &= E\left(\frac{Z_n^+ - Z_n^-}{2c_n \Delta_{n_i}}\right) \\
 &= E\left(E\left[\frac{Z_n^+ - Z_n^-}{2c_n \Delta_{n_i}} \middle| \Delta_n\right]\right) \\
 &= E\left(\frac{1}{2c_n \Delta_{n_i}} \{E[Z_n^+ | \Delta_n] - E[Z_n^- | \Delta_n]\}\right)
 \end{aligned}$$

$$\begin{aligned}
 &= E \left(\frac{1}{2c_n \Delta_{n_i}} [E(Z_n^+) - E(Z_n^-)] \right) \\
 &= E \left(\frac{1}{2c_n \Delta_{n_i}} [E(Z(\theta_n + c_n \Delta_n; \omega)) - E(Z(\theta_n - c_n \Delta_n; \omega))] \right) \\
 &= E \left(\frac{1}{2c_n \Delta_{n_i}} [E(Z(\theta_n; \omega)) + g_n^t c_n \Delta_n + O^+(c_n^2 \Delta_n^2) - \{E(L(\theta_n, \omega)) - g_n^t c_n \Delta_n + O^-(c_n^2 \Delta_n^2)\}] \right) \\
 &= E \left(\frac{1}{2c_n \Delta_{n_i}} [2g_n^t c_n \Delta_n + O(c_n^2 \Delta_n^2)] \right) \\
 &= E \left(\frac{g_n^t \Delta_n}{\Delta_{n_i}} \right) + E \left(\frac{O(c_n \Delta_n^2)}{\Delta_{n_i}} \right) \\
 &= E \left(\frac{g_n^t \Delta_n}{\Delta_{n_i}} \right) + O \left(E \left(\frac{c_n \Delta_n^2}{\Delta_{n_i}} \right) \right) \\
 &= E \left(\frac{g_n^t \Delta_n}{\Delta_{n_i}} \right) + O \left(c_n E \left(\frac{\Delta_n^2}{\Delta_{n_i}} \right) \right) \\
 &= g_{n_i} + O(c_n).
 \end{aligned}$$

La última igualdad la obtenemos de aplicar (2.7) y (2.8). Por tanto

$$\begin{aligned}
 b_n &= E(\hat{g}_n) - g_n \\
 &= g_n + O(c_n) - g_n \\
 &= O(c_n).
 \end{aligned}$$

Como $c_n \rightarrow 0$ cuando n aumenta, ya que $\{c_n\}_{n=1,2,\dots}$ es una sucesión positiva que converge a cero, entonces la sucesión $\{O(c_n)\}_{n=1,2,\dots}$ converge a cero, por lo que $b_n \rightarrow 0$ con probabilidad 1. ■

El método de aproximación estocástica por SP posee tres ventajas significativas: (a) generalidad, su ventaja sobre otros métodos como IPA y LR (cuyos métodos se verán en las siguientes secciones) es que éste puede aplicarse a cualquier sistema que pueda ser simulado, ya que no requiere conocimiento detallado del sistema ni de la distribución de sus entradas; (b) eficiencia, su ventaja sobre la aproximación estocástica por FD es su ahorro computacional para problemas de gran dimensión y (c) facilidad para usar, este método es tan fácil de implementar como el método de FD.

2.4.3. Análisis de perturbación infinitesimal

El método de análisis de perturbación infinitesimal (IPA por sus siglas en inglés *infinitesimal perturbation analysis*), también conocido como el método de la derivada trayectorial (*pathwise derivative*), es un método alternativo de aproximación estocástica para estimar derivadas. La idea de este método es la diferenciación de una trayectoria muestral. Como vimos en el método de FD, el error cuadrático medio decrementa cuando el parámetro h disminuye su valor. Esto sugiere que debemos decrementar h a cero y estimar la derivada de $z(\theta) = E(Z(\theta))$ a

través de la v.a.

$$D(\theta) = \frac{d}{d\theta}Z(\theta) = Z'(\theta) = \lim_{h \rightarrow 0} \frac{Z(\theta + h) - Z(\theta)}{h}, \quad (2.9)$$

(suponemos, por supuesto, que la derivada existe con probabilidad uno). Esta v.a. tiene esperanza $E(Z'(\theta))$ la cual es un estimador insesgado de $z'(\theta)$ siempre que

$$E\left(\frac{d}{d\theta}Z(\theta)\right) = \frac{d}{d\theta}E(Z(\theta)), \quad (2.10)$$

es decir, si es posible el intercambio de los operadores diferencial e integral.

El método sugiere entonces, calcular un estimador CMC. Esto es, realizar R realizaciones independientes de $D(\theta)$ y usar la media aritmética

$$\hat{z}'(\theta) = \frac{1}{R} \sum_{r=1}^R D^r(\theta)$$

como estimador de $z'(\theta)$. Así, $\hat{z}'(\theta)$ se mantiene consistente cuando $R \rightarrow \infty$ siempre que $E(D(\theta)) = z'(\theta)$, es decir, cuando (2.10) se cumple. Esto garantiza que el estimador $\hat{z}'(\theta)$ converge a $z'(\theta)$ con la tasa canónica $R^{-1/2}$ (Asmussen & Glynn [2007]).

Para precisar (2.9) observemos que $Z(\theta + h)$ y $Z(\theta)$ son dependientes cuando utilizamos números aleatorios comunes. Luego, consideremos $\{Z(\theta)\}_{\theta \in \Theta}$ una sucesión de funciones aleatorias definidas sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathcal{P})$ con $\Theta \subseteq \mathbb{R}$. En otras palabras, $Z(\theta)$ es un proceso estocástico indexado por θ . Podemos fijar $\omega \in \Omega$ y pensar en el mapeo $\theta \mapsto Z(\theta; \omega)$ como una función aleatoria sobre Θ . Entonces podemos interpretar a $Z'(\theta) = Z'(\theta; \omega)$ como la derivada de la función aleatoria Z con respecto a θ y con ω fijo. Entonces llamamos a $Z'(\theta)$ la derivada trayectorial de Z en θ : de aquí el nombre del método.

Usualmente interpretamos cada ω como una realización de la salida de un conjunto de números aleatorios generado. Cada $Z(\theta; \omega)$ es entonces, la salida de un algoritmo de simulación evaluado en θ con la secuencia de números aleatorios ω , y cada $Z'(\theta; \omega)$ es la derivada del resultado obtenido de una simulación evaluada en θ y con los números aleatorios fijos ω .

Un impedimento para aplicar el método de IPA es la no satisfacción de (2.10), el cual es un problema teórico-práctico. Ahora veremos las condiciones que garantizan la validación de (2.10), la insesgidez del estimador IPA y así, la validez del método.

Como vimos anteriormente, podemos considerar $\{Z(\theta)\}_{\theta \in \Theta}$ como una sucesión de funciones aleatorias que representa, cada $Z(\theta)$, el resultado (o la salida) de un algoritmo de simulación como una función de θ y con el conjunto de números aleatorios ω fijos, es decir, se utilizan números aleatorios comunes. Dada la existencia de $Z'(\theta)$ (ya que estamos trabajando bajo el contexto en el cual la derivada $Z'(\theta)$ existe con probabilidad uno, al menos en $\theta \in \Theta$), la cuestión ahora es la validación de (2.10), lo cual es un asunto de intercambiar un límite y una esperanza para asegurar que

$$E\left(\lim_{h \rightarrow 0} \frac{Z(\theta + h) - Z(\theta)}{h}\right) = \lim_{h \rightarrow 0} E\left(\frac{Z(\theta + h) - Z(\theta)}{h}\right).$$

Una condición necesaria y suficiente para que esto se cumpla es la integrabilidad uniforme⁵ del cociente diferencial $(Z(\theta + h) - Z(\theta))/h$. El objetivo es probar las condiciones suficientes que, en la práctica, sean fácilmente verificables. La condición estándar para la validez del IPA se enuncia en la siguiente proposición.

Proposición 2.4.2. *Suponga que $Z(\theta)$ es diferenciable c.s. en θ_0 y que $Z(\theta)$ satisface c.s. la condición de Lipschitz, esto es, existe $M(\omega) \geq 0$ tal que*

$$|Z(\theta_1) - Z(\theta_2)| \leq M(\omega)|\theta_1 - \theta_2|$$

para θ_1 y θ_2 en una vecindad no aleatoria de θ_0 , con $E(M(\omega)) < \infty$. Entonces

$$E\left(\frac{d}{d\theta}Z(\theta)\right) = \frac{d}{d\theta}E(Z(\theta))$$

se cumple en $\theta = \theta_0$.

Demostración. Como

$$\lim_{h \rightarrow 0} \frac{Z(\theta_0 + h) - Z(\theta_0)}{h} = Z'(\theta_0)$$

y el cociente $(Z(\theta_0 + h) - Z(\theta_0))/h$ es dominado por la función integrable $M(\omega)$, ya que $Z(\theta)$ satisface la condición de Lipschitz, esto es,

$$\left| \frac{Z(\theta_0 + h) - Z(\theta_0)}{h} \right| \leq M(\omega)$$

y $E(M(\omega)) < \infty$. Entonces por el Teorema de la Convergencia Dominada podemos afirmar que $Z'(\theta_0) \leq M(\omega)$, $E(Z'(\theta_0)) < \infty$ y además

$$\lim_{h \rightarrow 0} E\left(\frac{Z(\theta_0 + h) - Z(\theta_0)}{h}\right) = E(Z'(\theta_0)).$$

Lo cual prueba la proposición. ■

Supongamos que en efecto (2.10) se satisface (es decir, supongamos que el IPA se puede aplicar) y que $M(\omega)$ tiene segundos momentos finitos. Entonces

$$E([Z(\theta + h) - Z(\theta)]^2) \leq h^2 E(M(\omega)^2)$$

de lo cual se sigue que $Var(Z(\theta + h) - Z(\theta)) = O(h^2)$. Además, como $Z(\theta + h) \rightarrow Z(\theta)$ c.s. cuando $h \rightarrow 0$ (puesto que estamos utilizando números aleatorios comunes) entonces $Var(Z(\theta + h) - Z(\theta)) \rightarrow 0$ con una tasa h^2 .

Por tanto, si el IPA es válido entonces

$$\frac{Z(\theta + h) - Z(\theta)}{h} \rightarrow z'(\theta)$$

c.s. cuando $h \rightarrow 0$ y su esperanza también converge, ya que

$$\lim_{h \rightarrow 0} E\left(\frac{Z(\theta + h) - Z(\theta)}{h}\right) = \lim_{h \rightarrow 0} \frac{z(\theta + h) - z(\theta)}{h} = z'(\theta).$$

⁵La condición de integrabilidad uniforme es equivalente a que exista una función integrable, en el sentido de Lebesgue, dominando; esto es, una función integrable que acote superiormente.

Así, el alcance del método de IPA es, esencialmente, el mismo que el alcance del método de FD. Salvo que el IPA, a diferencia de FD, nos proporciona un estimador insesgado pero la función de desempeño requiere la condición de Lipschitz para su aplicación.

2.4.4. Método de razón de verosimilitud

Como vimos en la sección anterior, el alcance del método de IPA es limitado debido a sus requerimientos para su aplicación. El método de razón de verosimilitud (LR por sus siglas en inglés *likelihood ratio*), también conocido como el método de la función *score*, proporciona una aproximación alterna para estimar derivadas, cuyo método no requiere de la suavidad de la función de desempeño; pero sí requiere del conocimiento de las distribuciones de sus v.a. de entrada. Esta aproximación se logra a través de la diferenciación de sus respectivas funciones de densidad.

En el método de LR expresamos a Z como una función del vector aleatorio ω y suponemos que ω tiene una función de densidad f_θ que depende del parámetro θ y para enfatizar que una esperanza es calculada con respecto a f_θ la denotaremos por E_θ . En este contexto, la esperanza de Z está dada por

$$z(\theta) = E_\theta(Z(\omega)) = \int Z(\omega) f_\theta(\omega) d\omega.$$

Para obtener un estimador de la derivada de $z(\theta)$, suponemos que el orden de la diferenciación e integración puede ser intercambiable para obtener

$$z'(\theta) = \frac{d}{d\theta} E_\theta(Z(\omega)) = \int Z(\omega) \frac{d}{d\theta} f_\theta(\omega) d\omega. \quad (2.11)$$

Si esto se satisface, entonces multiplicando y dividiendo el lado derecho de la ecuación por $f_\theta(\omega)$ obtenemos que

$$\begin{aligned} z'(\theta) &= \int Z(\omega) \frac{f'_\theta(\omega)}{f_\theta(\omega)} f_\theta(\omega) d\omega \\ &= E_\theta \left(Z(\omega) \frac{f'_\theta(\omega)}{f_\theta(\omega)} \right) \end{aligned}$$

con $f'_\theta(\omega) = df_\theta(\omega)/d\theta$. De lo cual se sigue que el estimador CMC

$$\hat{z}'(\theta) = \frac{1}{R} \sum_{r=1}^R D^r(\theta)$$

es un estimador insesgado de la derivada de $E_\theta(Z(\omega))$, con

$$D^r(\theta) = Z(\omega_r) \frac{f'_\theta(\omega_r)}{f_\theta(\omega_r)}.$$

D^r , así definido, es conocido como el estimador LR.

La expresión $\nabla_\theta f_\theta(\omega)/f_\theta(\omega)$ ó $\nabla_\theta \log f_\theta(\omega)$ se conoce como función *score* y suele denotarse por $S(\theta, \omega)$, $S_\omega(\theta)$ ó simplemente S_ω , donde $\theta \in \mathbb{R}^p$ y ∇_θ denota el gradiente con respecto al parámetro θ .

Para un vector aleatorio $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ cuyas componentes $\omega_1, \omega_2, \dots, \omega_n$ son v.a.i. con funciones de densidad $f_\theta^{(1)}(\omega_1), f_\theta^{(2)}(\omega_2), \dots, f_\theta^{(n)}(\omega_n)$, respectivamente. La función de densidad conjunta de las componentes de ω está dada por $f_\theta(\omega) = \prod_{i=1}^n f_\theta^{(i)}(\omega_i)$ y su función *score* es

$$\begin{aligned} S(\theta, \omega) &= \nabla_\theta \log f_\theta(\omega) \\ &= \nabla_\theta \log \left(\prod_{i=1}^n f_\theta^{(i)}(\omega_i) \right) \\ &= \sum_{i=1}^n \nabla_\theta \log f_\theta^{(i)}(\omega_i) \\ &= \sum_{i=1}^n S_{\omega_i}(\theta, \omega_i). \end{aligned}$$

Esta propiedad se conoce como la propiedad aditiva de la función *score*. Otra propiedad fundamental está dada en la siguiente proposición.

Proposición 2.4.3. $E_\theta(S(\theta, \omega)) = 0$.

Demostración. Sabemos que $\int f_\theta(\omega) d\omega = 1$, entonces

$$\begin{aligned} E_\theta(S(\theta, \omega)) &= E_\theta(\nabla_\theta \log f_\theta(\omega)) \\ &= E_\theta\left(\frac{\nabla_\theta f_\theta(\omega)}{f_\theta(\omega)}\right) \\ &= \int \nabla_\theta f_\theta(\omega) d\omega \\ &= \nabla_\theta \int f_\theta(\omega) d\omega \\ &= 0. \end{aligned}$$

■

Para que los operadores diferencial e integral se puedan intercambiar en (2.11) se requiere que la función *score* sea continua. En la siguiente proposición veremos, para el caso uni-dimensional, las condiciones que debe satisfacer $f_\theta(\omega)$ para que el intercambio de estos operadores sea válido.

Proposición 2.4.4. Sea $\{f_\theta(\omega)\}_{\theta \in \Theta}$ una familia de densidades en \mathbb{R} tales que, para ω casi en todas partes, $f_\theta(\omega)$ es continuamente diferenciable en θ con derivada $f'_\theta(\omega)$. Entonces

$$\frac{d}{d\theta} \int Z(\omega) f_\theta(\omega) d\omega = \int Z(\omega) f'_\theta(\omega) d\omega$$

para toda θ en un sub-intervalo abierto Θ_0 de Θ siempre que existan p y q exponentes conjugados, es decir, con $1/p + 1/q = 1$ tales que $Z \in L^p$ y $|f'_\theta(\omega)| \leq M(\omega)$ para alguna $M \in L^q$, para todo $\theta \in \Theta_0$, y ω casi en todas partes.

Demostración. Suponga $(\theta - \varepsilon, \theta + \varepsilon) \subseteq \Theta_0$. Para $|h| < \varepsilon$ tenemos que

$$\frac{d}{d\theta} Z(\omega) f_\theta(\omega) = \lim_{h \rightarrow 0} \frac{Z(\omega) f_{\theta+h}(\omega) - Z(\omega) f_\theta(\omega)}{h}$$

$$\begin{aligned}
 &= \lim_{h \rightarrow 0} Z(\omega) \frac{f_{\theta+h}(\omega) - f_{\theta}(\omega)}{h} \\
 &= Z(\omega) \lim_{h \rightarrow 0} f'_{\theta+h^*}(\omega) \\
 &= Z(\omega) f'_{\theta}(\omega) \\
 &\leq |Z(\omega)M(\omega)|.
 \end{aligned}$$

Para obtener la tercera igualdad aplicamos el Teorema del Valor Medio, para derivadas, para alguna $h^* \in (-h, h)$, así garantizamos que $\theta + h^* \in (\theta - \varepsilon, \theta + \varepsilon)$. Por otro lado, como $Z \in L^p$ y $M \in L^q$ y p y q son exponentes conjugados entonces, por la desigualdad de Hölder: $Z(\omega)M(\omega) \in L^1$, esto es

$$\|Z(\omega)M(\omega)\|_1 \leq \|Z(\omega)\|_p \|M(\omega)\|_q$$

o lo que es equivalente

$$\int |Z(\omega)M(\omega)| \, d\omega \leq \left[\int |Z(\omega)|^p \, d\omega \right]^{1/p} \left[\int |M(\omega)|^q \, d\omega \right]^{1/q}.$$

Luego, como $Z(\omega)M(\omega) \in L^1$ entonces

$$\int |Z(\omega)M(\omega)| \, d\omega < \infty$$

por tanto $Z(\omega)M(\omega)$ es integrable en el sentido de Lebesgue (véase Kreyszig [1978]). Entonces tenemos las condiciones necesarias para aplicar el Teorema de la Convergencia Dominada y concluir que

$$\lim_{h \rightarrow 0} \int \frac{Z(\omega)f_{\theta+h}(\omega) - Z(\omega)f_{\theta}(\omega)}{h} \, d\omega = \int Z(\omega)f'_{\theta}(\omega) \, d\omega,$$

o lo que es lo mismo

$$\frac{d}{d\theta} \int Z(\omega)f_{\theta}(\omega) \, d\omega = \int Z(\omega)f'_{\theta}(\omega) \, d\omega.$$

■

El método de LR proporciona un estimador insesgado cuando (2.11) se satisface; esto es, cuando la integral (sobre ω) del límite (cuando $h \rightarrow 0$) de la función

$$Z(\omega) \frac{1}{h} \left(\frac{f_{\theta+h}(\omega)}{f_{\theta}(\omega)} - 1 \right) f_{\theta}(\omega)$$

equivale al límite de su integral.

Tanto el método de LR como en el método de IPA, la validación de la aproximación radica en el intercambio de la diferenciación e integración. Sin embargo, el intercambio en (2.11) es relativamente más asequible en comparación con el intercambio en (2.10), ya que las funciones de densidad tienden a ser funciones suaves. Sin embargo, en la práctica, la aplicación del método de LR es, a menudo, limitado ya sea por la necesidad de conocer explícitamente la función de densidad f , o por la obtención de una gran varianza debido al incumplimiento de (2.11).

Es importante considerar que existen ejemplos no patológicos en los cuales el intercambio de operadores requerido por el estimador LR falla; ya que existen casos en los cuales algún elemento del parámetro θ , importante para el desempeño de un sistema estocástico, aparece de cierta manera en la función de densidad $f_\theta(\omega)$ que al momento de calcular la función *score*, ésta se anula en al menos uno de sus elementos. Por ejemplo: una cola $GI/D/1$ cuyo tiempo de servicio es determinista o una cola $D/G/1$ en la cual, los tiempos entre arribos son deterministas. Para algunos de estos casos se puede aplicar el método *Push-out*, iniciado por Rubinstein, el cual es una extensión del método *LR* (para más detalles véase Asmussen & Glynn [2007]).

2.5. Optimización estocástica: algoritmo de aproximación estocástica

Sea $Z(\theta; \omega)$ la función de desempeño del sistema, donde $\theta = (\theta_1, \theta_2, \dots, \theta_p) \in \Theta \subset \mathbb{R}^p$ corresponde a los parámetro del sistema y ω denota los efectos estocásticos implicados, definamos la función objetivo del sistema

$$J(\theta) = E(Z(\theta; \omega)).$$

El objetivo es encontrar $\theta^* \in \Theta$ que minimice $J(\theta)$.

Para resolver la ecuación $\nabla J = 0$ se utiliza un algoritmo iterativo que es, básicamente, la versión estocástica del algoritmo del gradiente proyectado:

$$\theta_{n+1} = \mathbf{\Pi}_\Theta(\theta_n - \alpha_n \hat{g}_n),$$

donde $\theta_n = (\theta_{n_1}, \theta_{n_2}, \dots, \theta_{n_p})$ representa el parámetro, obtenido, de la n -ésima iteración, \hat{g}_n denota un estimador de $\nabla J(\theta_n)$, $\{\alpha_n\}_{n=1,2,\dots}$ es una sucesión positiva que converge a cero y $\mathbf{\Pi}_\Theta$ denota una proyección sobre el conjunto de restricciones Θ (para más detalles véase Asmussen & Glynn [2007]).

La aplicación de los métodos de aproximación estocástica (SA por sus siglas en inglés *stochastic approximation*) a la optimización de sistemas de eventos discretos, ha sido desarrollado recientemente dando lugar a la aparición de una completa línea de investigación dedicada a métodos de estimación de gradientes basados en simulación. Empíricamente, desde su aparición a finales de los 80's y principios de los 90's, estos métodos han proporcionado buenos resultados de convergencia y han sido una herramienta esencial, en las aplicaciones (para más detalles véase Fu & Hill [1997]).

2.6. Aplicación a un sistema de colas del tipo M/U/1

En esta sección presentamos la optimización, a través de simulación, de un sistema de eventos discretos. Para ilustrar la precisión de los métodos de SA vistos anteriormente consideramos como medida de desempeño el tiempo medio que un cliente pasa en una cola M/U/1, cuya solución analítica es conocida gracias a la fórmula de Pollaczek-Khinchine (P-K); lo cual nos permite comparar los resultados obtenidos a través de simulación.

Sea una cola con un servidor, en donde los tiempos de servicio se distribuyen de manera uniforme y los arribos de clientes, al sistema, ocurren de acuerdo a un proceso de Poisson. Supongamos que deseamos minimizar el tiempo medio que un cliente pasa en el sistema (tiempo que pasa en la cola más el tiempo de servicio), denotado por T , y consideremos K : un costo para la mejora de los tiempos de servicio. Específicamente, deseamos determinar los valores, de los dos parámetros, de la distribución uniforme de los tiempos de servicio $Y \sim \text{uniforme}(\theta_1 - \theta_2, \theta_1 + \theta_2)$ que minimicen la función objetivo

$$J(\theta) = E(T(\theta)) - K^t \theta,$$

con $\theta = (\theta_1, \theta_2) \in \Theta \subset \mathbb{R}^2$, $K = (k_1, k_2) \in \mathbb{R}^2$ donde k_1 y k_2 representan los costos para reducir la media y la “variabilidad”, respectivamente, de los tiempos de servicio, $\Theta = \{\theta \in \mathbb{R}^2 \mid 0 < \theta_{\text{mín}} \leq \theta_2 \leq \theta_1 \leq \theta_{\text{máx}} < 1/\lambda\}$ es el conjunto de restricciones que aseguran la estabilidad del sistema (véase figura 2.1) y λ es la intensidad de arribos.

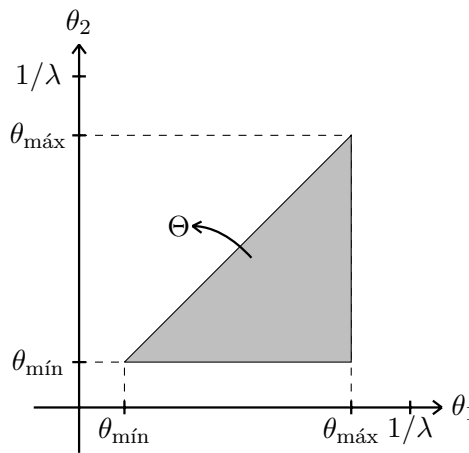


Figura 2.1: Región factible Θ .

El gradiente estimado, en la n -ésima iteración del algoritmo de optimización, es de la forma

$$\hat{g}_n = \hat{\nabla} E(T(\theta_n)) - K,$$

donde $\hat{\nabla} E(T(\theta_n))$ denota un estimador de $\nabla E(T(\theta_n))$ evaluado en θ_n .

Solución analítica.

De manera general, para una cola M/G/1 con arribos Poisson con intensidad λ , sea Y la v.a. de los tiempos de servicio con función de distribución $F(y; \theta)$. La fórmula de P-K nos proporciona la expresión analítica de $E(T)$ y es de la forma

$$E(T(\theta)) = E(Y) + \frac{\lambda E(Y^2)}{2(1 - \lambda E(Y))},$$

con $\lambda E(Y) < 1$ para asegurar que $E(T(\theta)) < \infty$ (para más detalles véase Sury & Zazani [1988]).

Para nuestro caso, como $Y \sim \text{uniforme}(\theta_1 - \theta_2, \theta_1 + \theta_2)$ entonces $E(Y) = \theta_1$ y $E(Y^2) = \theta_1^2 + \theta_2^2/3$. Por tanto

$$E(T(\theta)) = \frac{\lambda \theta_2^2 - 3\lambda \theta_1^2 + 6\theta_1}{6(1 - \lambda \theta_1)}$$

y

$$\nabla E(T(\theta)) = \left(\frac{1}{2} + \frac{\lambda^2 \theta_2^2 + 3}{6(1 - \lambda \theta_1)^2}, \frac{\lambda \theta_2}{3(1 - \lambda \theta_1)} \right).$$

Ahora veremos la estimación de este gradiente utilizando los métodos de SA. A diferencia de los métodos de IPA y LR, los métodos de FD y SP no requieren de un conocimiento detallado de la función de desempeño, por lo que su aplicación es directa; en cambio IPA y LR sí requieren de un análisis previo de la función para su aplicación.

Para LR la función *score* es de la forma

$$S(\theta, \omega) = \left(0, -\frac{n}{\theta_2} \right)$$

con n el número de clientes observados. Por tanto, para este ejemplo en particular no es posible aplicar el método de LR para estimar sus sensibilidades, ya que el primer elemento de $S(\theta, \omega)$ es nulo.

Solución a través de IPA.

El método de IPA nos dice que, bajo determinadas condiciones, es posible intercambiar los operadores diferencial e integral. Esto es

$$\nabla E(T(\theta)) = E(\nabla T(\theta)).$$

Por lo que estimar $\nabla E(T(\theta))$ es equivalente a calcular $\nabla T(\theta)$ y, posteriormente, usar este término para calcular un estimador CMC. Ahora veremos como calculamos $\nabla T(\theta)$.

Sea T_i , con $i = 1, 2, \dots$, el tiempo que el i -ésimo cliente pasa en el sistema y sea A_i el tiempo transcurrido entre los arribos de los clientes $(i - 1)$ e i (como los arribos ocurren de acuerdo a un proceso de Poisson con intensidad λ entonces $A_i \sim \text{exponencial}(\lambda)$), entonces el proceso del tiempo de espera, a través de la fórmula de Lindley (Rubinstein & Kroese [2008]), está dado por

$$T_i = \text{máx} \{T_{i-1} - A_i, 0\} + Y_i$$

con $i = 1, 2, \dots$ y con $T_0 = 0$ (véase figura 2.2).

Si consideramos un periodo ocupado⁶ entonces $T_i = T_{i-1} - A_i + Y_i$ y $\nabla T_i = \nabla T_{i-1} + \nabla Y_i$. Por lo tanto, calcular $\nabla T(\theta)$ se reduce a calcular $\nabla Y = (\partial y / \partial \theta_1, \partial y / \partial \theta_2)$.

Para una realización del sistema: sea n el número de clientes observados que generan M periodos ocupados en el sistema. Puesto que la función de distribución de los tiempos de servicio está dada por la forma

$$F(y; \theta) = \frac{1}{2} + \frac{y - \theta_1}{2\theta_2}$$

⁶Un periodo ocupado es un lapso de tiempo en el cual hay uno o más clientes en el sistema; por el contrario, un periodo libre u ocioso es un lapso de tiempo en el cual no hay cliente alguno en el sistema, por lo que se dice que el sistema está libre.

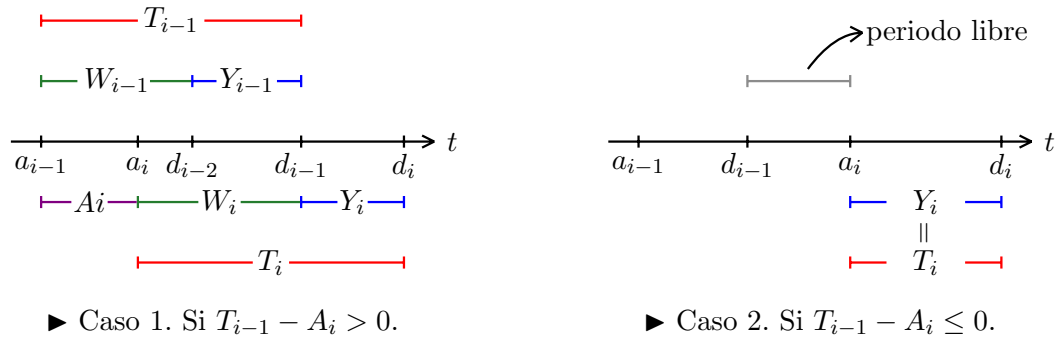


Figura 2.2: Tiempo que el i -ésimo cliente pasa en el sistema.

si $\theta_1 - \theta_2 < y < \theta_1 + \theta_2$. De Suri & Zazanis [1988] obtenemos que

$$\nabla Y = -\frac{\nabla F(y; \theta)}{\partial F(y; \theta) / \partial y} = \left(1, \frac{y - \theta_1}{\theta_2} \right)$$

y por tanto

$$\nabla T(\theta) = \frac{1}{n} \sum_{m=1}^M \sum_{i=1}^{n_m} \sum_{j=1}^i \nabla Y_{k_m+j},$$

donde n_m denota la cantidad de clientes servidos en el m -ésimo periodo ocupado del sistema, esto es, $n = \sum_{m=1}^M n_m$ y Y_{k_m+1} es el tiempo de servicio del primer cliente atendido en el m -ésimo periodo ocupado.

Resultados.

Realizamos experimentos numéricos usando el algoritmo de optimización estocástica con los métodos de FD, SP e IPA para la estimación del gradiente. La implementación de SP requiere de tres sucesiones: la sucesión del tamaño de paso $\{\alpha_n\}_{n=1,2,\dots}$; la sucesión del incremento $\{c_n\}_{n=1,2,\dots}$ para la estimación del gradiente y la sucesión de vectores de perturbación simultánea $\{\Delta_n\}_{n=1,2,\dots}$. Las dos primeras sucesiones deben ser sucesiones positivas y que convergen a cero con una tasa apropiada (para más detalles véase Fu & Hill [1997]); en nuestros experimentos tomamos $a_n = a/n^\alpha$ y $c_n = c/n^\beta$, con α, β, a y c constantes fijas sujetas a $\alpha \leq 1$ y $\alpha - \beta > 1/2$. Para la sucesión $\{\Delta_n\}_{n=1,2,\dots}$ consideramos v.a.i.i.d. Bernoulli simétricas, esto es, $P(\Delta_n = 1) = P(\Delta_n = -1) = 1/2$. Para crear un marco más apropiado de comparación entre los distintos métodos de SA, consideramos las mismas sucesiones para la implementación del resto de los métodos, según lo requiera su estimador. De este modo, para la implementación de FD consideramos el tamaño del incremento h diferente para cada iteración, específicamente, tomamos $h_n = 2c_n$.

Consideramos el costo de reducción de los tiempos de servicio $K = (1.28125, 0.00125)$, $\alpha = 0.9$, $\beta = 0.3$, $a = 1$ y $c = 0.001$. El resto de los valores para la implementación son los siguientes: $\lambda = 1$, tomamos $\theta_{\min} = 0.001$ y $\theta_{\max} = 1/\lambda - \theta_{\min}$, el punto inicial $\theta_0 = (0.5, 0.3) \in \Theta$, observamos a 100 clientes por cada realización del sistema y consideramos 1,000 realizaciones para cada estimador. Para los criterios de paro, consideramos un número máximo de iteraciones $N = 100$ y una tolerancia $tol = 10^{-4}$.

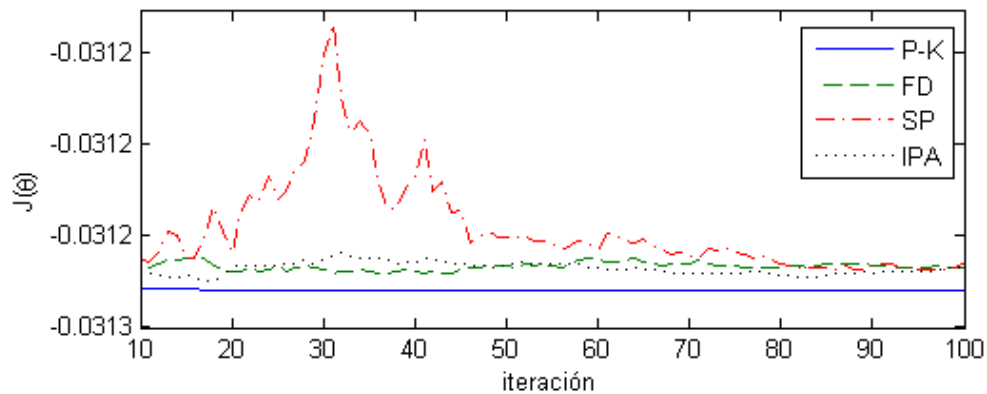
La tabla 2.1 muestra los resultados experimentales obtenidos de $\nabla E(T(\theta))$, evaluado en θ_0 , basado en 1,000

Método		$\nabla E(T(\theta_0))$		CPU (seg.)
		$\partial E(T(\theta_0))/\partial\theta_{0_1}$	$\partial E(T(\theta_0))/\partial\theta_{0_2}$	
Teórico	Fórmula P-K	2.56	0.20	---
Aproximado	FD	2.5279 ± 0.0438	0.1970 ± 0.0149	1.60
	SP	2.4748 ± 0.0476	0.1530 ± 0.1604	0.88
	IPA	2.4490 ± 0.0427	0.1923 ± 0.0133	2.67

Tabla 2.1: Estimación de $\nabla E(T(\theta_0))$.

realizaciones independientes de 100 clientes observados por cada realización. En la última columna se muestra el tiempo computacional requerido para realizar cada una de las estimaciones presentadas. También, para fines de comparación, se muestra el resultado analítico obtenido a través de la fórmula de P-K.

La figura 2.3 muestra las trayectoria del proceso de optimización de $J(\theta)$, a través de los distintos métodos de SA; cuyas trayectorias muestran que, en efecto, los métodos convergen a su valor óptimo. Cabe mencionar que los tiempos computacionales requeridos por el algoritmo de optimización fue: a través de P-K, 0.01 segundos y satisfizo los criterios de paro en la iteración 66; FD, requirió 164.68 segundos; SP, 90.51 segundos; e IPA, requirió 321.36 segundos. La figura 2.4 muestra la trayectoria de cada componente del parámetro $\theta = (\theta_1, \theta_2)$.

Figura 2.3: Optimización estocástica de $J(\theta)$.

La tabla 2.2 muestra los valores óptimos obtenidos y su correspondiente valor de la función de desempeño, para distintos valores de K .

En las gráficas y tablas presentadas podemos observar que los resultados experimentales obtenidos fueron prometedores, en el sentido de que validan de una forma práctica la aplicación de métodos de SA a la optimización de sistemas de eventos discretos. Es importante mencionar que el método de SP ofrece el mayor ahorro computacional, seguido del método de FD; mientras que IPA requiere un tiempo computacional mayor.

Es importante mencionar que todas las simulaciones presentadas en este trabajo de tesis fueron ejecutadas en un ordenador Intel[®] Core[™]2 Duo T6600, 2.20 GHz y 4GB RAM.

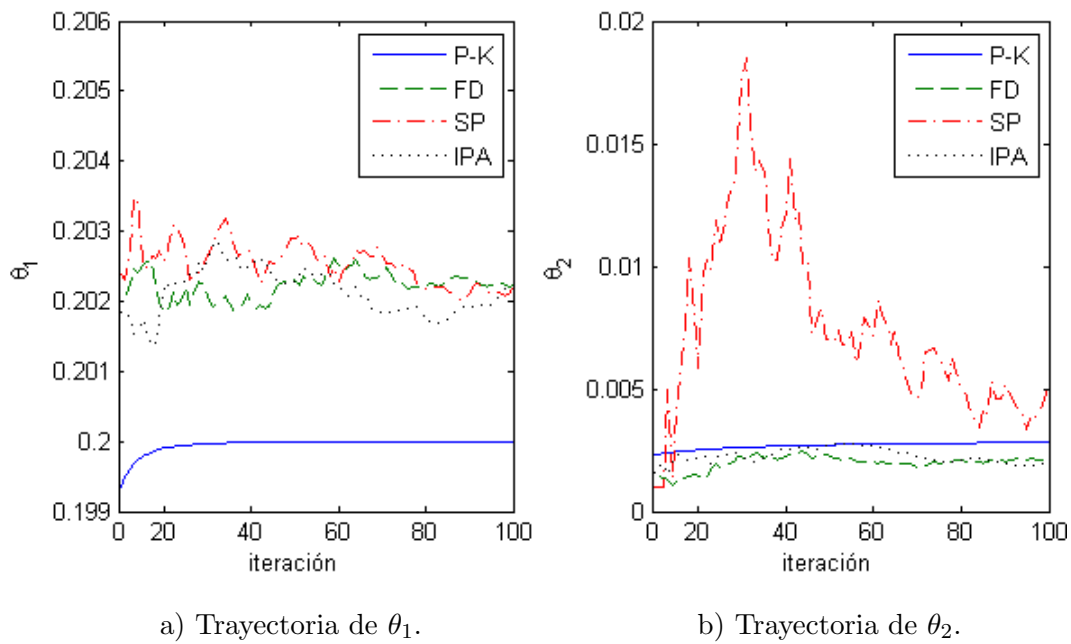


Figura 2.4: Trayectoria de $\theta = (\theta_1, \theta_2)$.

Caso	K		Método	θ^*		$J(\theta^*)$	CPU (seg.)
	k_1	k_2		θ_1^*	θ_2^*		
1	1.28125	0.00125	P-K	0.20000	0.00276	-0.03125	0.01
			FD	0.20186	0.00394	-0.03125	164.68
			SP	0.20258	0.00857	-0.03124	90.51
			IPA	0.20215	0.00333	-0.03125	321.36
2	2.5	0.002	P-K	0.50000	0.00285	-0.50000	0.01
			FD	0.50860	0.00359	-0.49970	172.12
			SP	0.50803	0.01008	-0.49972	84.54
			IPA	0.50837	0.00474	-0.49972	266.40
3	2.6536	0.32	P-K	0.50232	0.45013	-0.65332	0.02
			FD	0.51266	0.47379	-0.65293	176.10
			SP	0.51115	0.48817	-0.65297	105.96
			IPA	0.51093	0.47309	-0.65311	311.05
4	13	0.005	P-K	0.77598	0.00271	-7.96782	0.02
			FD	0.85278	0.01008	-7.76344	190.82
			SP	0.85620	0.05650	-7.72203	106.96
			IPA	0.86393	0.00100	-7.62450	282.50

Tabla 2.2: Óptimos de $J(\theta)$ para distintos valores de K .

En este capítulo vimos los métodos de SA para la optimización de sistemas de eventos discretos a través de simulación, a lo largo del cual, presentamos las ventajas y desventajas que cada método ofrece. Esta teoría la asentamos con el desarrollo de experimentos de una cola M/U/1, a manera de ejemplo; cuyos resultados simulados validan, aunque no de una manera rigurosa, la precisión de la aplicación de métodos de SA a la optimización de sistemas de eventos discretos.

Retomando lo visto en el capítulo 1, en el siguiente capítulo veremos una simulación alternativa de un sistema Metro, que llamaremos simulación global. Presentamos la implementación computacional, de las simulaciones parcial y global, para el Sistema Metro de la Ciudad de México, para cuestiones de comparación entre ambas simulaciones. Además presentamos la estimación de sensibilidades, usando los métodos de SA vistos en este capítulo, de la función de desempeño del sistema Metro.

Capítulo 3

Simulación y sensibilidades del Sistema Metro de la Ciudad de México

En este capítulo veremos la implementación computacional para simular el movimiento de los trenes y de los arribos de pasajeros del Sistema Metro de la Ciudad de México, mediante dos sistemas de simulación, uno conocido como simulación parcial y otro como simulación global. En ambas simulaciones los tiempos de salidas y llegadas de trenes se obtiene mediante el uso de números pseudo-aleatorios. En la simulación global, los arribos de pasajeros se obtienen por simulación de números pseudo-aleatorios; mientras que en la simulación parcial, éstos se obtiene a través de sus valores esperados.

Para realizar la comparación, entre ambas simulaciones, consideramos un número determinado de líneas del Sistema Metro de la Ciudad de México; en particular consideramos tres líneas. Las simulaciones se alimentaron de una serie de datos proporcionados por el personal del STC Metro de la Ciudad de México. Nos facilitaron información sobre los tiempos de permanencia de los trenes en cada andén, así como la distancia entre andenes consecutivos y la velocidad promedio con la que viajan los trenes. También nos entregaron los datos de un sondeo que se realiza cada diez años, en el que se pregunta a los usuarios cuál es su andén origen y cuál es su andén destino, esto da como resultando la matriz origen-destino; los últimos datos de los que se tienen registro datan del año 2007. Además, nos dieron los datos de la cantidad de usuarios que ingresan al Sistema, en cada una de las estaciones del Metro, durante un día de servicio. De todos estos datos sólo la matriz origen-destino requiere de un análisis más detallado.

3.1. Matriz origen-destino

La matriz origen-destino representa información fundamental en la modelación del sistema Metro; ya que a partir de ella se calculan las probabilidades de que una persona que ingresa al sistema en determinada estación, lo abandone en cierta estación. Es decir, nos permite conocer la probabilidad de que un pasajero que ingresa al sistema en $a_o \in L_o$ lo abandone en $a_d \in L_d$, con L_o y L_d (que denotan la línea origen y la línea destino,

respectivamente) no necesariamente iguales. Esta matriz corresponde a una matriz por bloques como se muestra en la siguiente tabla¹.

$L_o \backslash L_d$	L_1	L_2	\dots	L_{NTL-1}	L_{NTL}
L_1	$\mathcal{MP}^{1,1}$	$\mathcal{MP}^{1,2}$	\dots	$\mathcal{MP}^{1,NTL-1}$	$\mathcal{MP}^{1,NTL}$
L_2	$\mathcal{MP}^{2,1}$	$\mathcal{MP}^{2,2}$	\dots	$\mathcal{MP}^{2,NTL-1}$	$\mathcal{MP}^{2,NTL}$
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
L_{NTL-1}	$\mathcal{MP}^{NTL-1,1}$	$\mathcal{MP}^{NTL-1,2}$	\dots	$\mathcal{MP}^{NTL-1,NTL-1}$	$\mathcal{MP}^{NTL-1,NTL}$
L_{NTL}	$\mathcal{MP}^{NTL,1}$	$\mathcal{MP}^{NTL,2}$	\dots	$\mathcal{MP}^{NTL,NTL-1}$	$\mathcal{MP}^{NTL,NTL}$

Cada sub-matriz $\mathcal{MP}^{m,l}$, con $m, l = 1, 2, \dots, NTL$, es una matriz de dimensión $\mathbb{R}^{NTA^m \times NTA^l}$ que contiene las probabilidades de ir de L_m a L_l . Entre estas sub-matrices distinguimos tres tipos: (a) cuando $L_m = L_l$, (b) cuando L_m es la línea opuesta a L_l y (c) las restantes.

Cuando $L_m = L_l$, la matriz $\mathcal{MP}^{l,l}$ es una matriz cuadrada triangular superior de dimensión NTA^l , cuyos elementos en la diagonal son todos iguales a cero. Así, cada fila de $\mathcal{MP}^{l,l}$ corresponde a un andén origen, $a_o \in L_l$, y cada columna corresponde a un andén destino, $a_d \in L_l$. De esta manera, dado que para cada línea los andenes siguen la secuencia $a_1 \rightarrow a_2 \rightarrow \dots \rightarrow a_{NTA}$, la probabilidad de que un pasajero que sube en a_i y baje en a_j , con $j \leq i$, es nula; esto es

$$\mathcal{MP}^{l,l} = \begin{pmatrix} 0 & p^{l,l}(a_1, a_2) & \dots & p^{l,l}(a_1, a_{NTA^l-1}) & p^{l,l}(a_1, a_{NTA^l}) \\ 0 & 0 & \dots & p^{l,l}(a_2, a_{NTA^l-1}) & p^{l,l}(a_2, a_{NTA^l}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & p^{l,l}(a_{NTA^l-1}, a_{NTA^l}) \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

donde $p^{l,l}(a_o, a_d)$ denota la probabilidad de que un pasajero que ingresó al sistema en $a_o \in L_l$, lo abandone en $a_d \in L_l$, con $o = 1, 2, \dots, NTA^l$ y $d = o + 1, o + 2, \dots, NTA^l$.

Cuando L_m es la línea opuesta a L_l , dado el supuesto de que los pasajeros no se regresan, la matriz $\mathcal{MP}^{m,l}$ es una matriz de ceros y la denotamos por $\mathcal{O}^{m,l}$.

$$\mathcal{O}^{m,l} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

¹La matriz origen-destino (original) proporcionada por el STC Metro de la Ciudad de México fue trabajada previamente para obtener la matriz que presentamos en este trabajo. Para más detalles de la construcción de ésta véase la sección A.1 en la página 63

Para el resto de las matrices $\mathcal{MP}^{m,l}$, éstas pueden verse, a su vez, como una matriz por bloques

$$\mathcal{MP}^{m,l} = \begin{pmatrix} \mathcal{O}_1^{m,l} & \mathcal{A}^{m,l} \\ \mathcal{O}_2^{m,l} & \mathcal{O}_3^{m,l} \end{pmatrix},$$

donde $\mathcal{O}_1^{m,l}$, $\mathcal{O}_2^{m,l}$ y $\mathcal{O}_3^{m,l}$ son matrices de ceros de dimensiones $\mathbb{R}^{(q_m-1) \times (p_l)}$, $\mathbb{R}^{(NTA^m - q_m + 1) \times (p_l)}$ y $\mathbb{R}^{(NTA^m - q_m + 1) \times (NTA^l - p_l)}$, respectivamente, con $a_{q_m} \in L_m \cap \mathcal{EC}$ y $a_{p_l} \in L_l \cap \mathcal{EC}$ y $\mathcal{A}^{m,l}$ es una matriz densa de dimensión $\mathbb{R}^{(q_m-1) \times (NTA^l - p_l)}$ tal que

$$\mathcal{A}^{m,l} = \begin{pmatrix} p^{m,l}(a_1, a_{p_l+1}) & p^{m,l}(a_1, a_{p_l+2}) & \cdots & p^{m,l}(a_1, a_{NTA^l}) \\ p^{m,l}(a_2, a_{p_l+1}) & p^{m,l}(a_2, a_{p_l+2}) & \cdots & p^{m,l}(a_2, a_{NTA^l}) \\ \vdots & \vdots & \ddots & \vdots \\ p^{m,l}(a_{q_m-2}, a_{p_l+1}) & p^{m,l}(a_{q_m-2}, a_{p_l+2}) & \cdots & p^{m,l}(a_{q_m-2}, a_{NTA^l}) \\ p^{m,l}(a_{q_m-1}, a_{p_l+1}) & p^{m,l}(a_{q_m-1}, a_{p_l+2}) & \cdots & p^{m,l}(a_{q_m-1}, a_{NTA^l}) \end{pmatrix}.$$

a_{q_m} es el andén donde los pasajeros que ingresaron en $a_o \in L_m$ deben transbordar, por primera vez, si desean llegar a $a_d \in L_l$ y a_{p_l} es el andén que recibe a los pasajeros que vienen de L_m y que transbordaron por última vez para llegar a L_l . Así, $p^{m,l}(a_o, a_d)$ es la probabilidad de que un pasajero que ingresó al sistema en $a_o \in L_m$, lo abandone en $a_d \in L_l$.

Por consiguiente, la matriz origen-destino, denotada por $\mathcal{M}_{\mathcal{O}-\mathcal{D}}$, es una matriz cuadrada de dimensión $n = \sum_{l=1}^{NTL} NTA^l$ y tiene la forma

$$\mathcal{M}_{\mathcal{O}-\mathcal{D}} = \begin{pmatrix} \mathcal{MP}^{1,1} & \mathcal{O}^{1,2} & \cdots & \mathcal{MP}^{1,NTL-1} & \mathcal{MP}^{1,NTL} \\ \mathcal{O}^{2,1} & \mathcal{MP}^{2,2} & \cdots & \mathcal{MP}^{2,NTL-1} & \mathcal{MP}^{2,NTL} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{MP}^{NTL-1,1} & \mathcal{MP}^{NTL-1,2} & \cdots & \mathcal{MP}^{NTL-1,NTL-1} & \mathcal{O}^{NTL-1,NTL} \\ \mathcal{MP}^{NTL,1} & \mathcal{MP}^{NTL,2} & \cdots & \mathcal{O}^{NTL,NTL-1} & \mathcal{MP}^{NTL,NTL} \end{pmatrix}.$$

Las sub-matrices $\mathcal{O}^{2k-1,2k}$, con $k = 1 : NTL/2$, son matrices cuyos elementos son todos iguales a cero con dimensión $\mathbb{R}^{NTA^{2k-1} \times NTA^{2k}}$. $\mathcal{O}^{2k-1,2k}$ es igual a la matriz transpuesta de $\mathcal{O}^{2k,2k-1}$. Cabe hacer notar que cada elemento de $\mathcal{M}_{\mathcal{O}-\mathcal{D}}$ satisface que

$$0 \leq p^{m,l}(a_o, a_d) \leq 1, \quad (3.1)$$

y, salvo algunas excepciones, la suma por filas de $\mathcal{M}_{\mathcal{O}-\mathcal{D}}$ es igual a uno, esto es, para todo $a_o \in L_m$ fijo tal que $a_o \neq a_{NTA^m}$, con $m = 1 : NTL$, se cumple que

$$\sum_{l=1}^{NTL} \sum_{\substack{a_d = a_1, \\ a_d \in L_l}}^{a_{NTA^l}} p^{m,l}(a_o, a_d) = 1. \quad (3.2)$$

Es importante valorar, para aplicaciones más realistas, que la matriz origen-destino no es global: puesto que resulta natural considerar una matriz origen-destino, para el horario matutino y otra distinta para el vespertino

(ya que en las mañanas, las personas acuden, de sus casas, a sus trabajos y/o escuelas y en las tardes, estas personas regresan). Además de tener en cuenta el crecimiento económico de la ciudad, esto es, la construcción de nuevos espacios que representen nodos atractores de urbanización (como centros comerciales, escuelas, fabricas, etc.); los cuales generarían la necesidad de modificar y/o ajustar la matriz origen-destino.

Un sondeo de campo para obtener dichas matrices representaría, además de una forma poco práctica, un costo muy elevado. Afortunadamente, existen maneras de estimar estas matrices a través de métodos de balance. Sin embargo, en el presente trabajo consideramos a $\mathcal{M}_{\mathcal{O}-\mathcal{D}}$ de manera determinista.

3.2. Simulación global

Hay dos formas de simular el comportamiento de un sistema Metro: la simulación parcial y la simulación global. En ambas simulaciones el flujo de trenes se obtiene como en la sección 1.1. En la simulación parcial, la cantidad de arribos se obtiene a través de sus valores esperados (como en la sección 1.2); mientras que en la simulación global, los arribos de pasajeros se obtienen a través de la generación de números pseudo-aleatorios, esto es, para cada simulación siempre que ocurre un arribo por ingreso al sistema, el proceso de conteo $NP_{a_o}^l(t)$ se actualiza y la hora en que ocurrió el arribo se registra para calcular su respectivo tiempo de espera.

Así, cuando el j -ésimo tren llega a $a_o \in L_l$, reunimos los datos importantes para determinar el valor de $PAI_j^l(a_o)$ y el respectivo tiempo de espera por ingreso al sistema asociado al j -ésimo tren, $WI_j^l(a_o)$. Para determinar $PDA_j^l(a_q)$ y, en el caso de las estaciones de correspondencia, $PDT_j^l(a_q)$ hacemos uso de $\mathcal{M}_{\mathcal{O}-\mathcal{D}}$. De esta manera determinamos el número de pasajeros que transbordan, $PT_k^{m,l}$, y su respectivo tiempo de espera, $WT_j^l(a_q)$. Por tanto, la función de desempeño $F(\mu)$ es calculada de forma exacta, para cada realización del sistema Metro,

$$F(\mu) = \sum_{l=1}^{NTL} k^l \sum_{s=1}^S \frac{T_s}{\mu_s^l} + c \sum_{s=1}^S \sum_{l=1}^{NTL} \sum_{j=1}^{NTT_s^l} \sum_{q=1}^{NTA^l} W_j^l(a_q),$$

con

$$W_j^l(a_q) = WI_j^l(a_q) + WT_j^l(a_q)I_{\{a_q \in \mathcal{EC}\}}.$$

3.3. Datos del Sistema Metro de la Ciudad de México

Para probar el modelo presentado en el capítulo 1 consideramos sólo tres líneas del Sistema Metro de la Ciudad de México, ignorando las demás. Trabajamos con las líneas: 1 (rosa), 2 (azul) y 3 (verde), véase la figura A.2 en la página 70. Conforme al modelo propuesto, cada línea se conforma de dos vías y cada vía es considerada como una línea independiente a las otras, de esta manera, trabajamos con seis líneas:

- L_1 corresponde a la línea rosa que corre de Pantitlán a Observatorio, de este a oeste;

- L_2 corresponde a la línea rosa que corre de Observatorio a Pantitlán, de oeste a este;
- L_3 corresponde a la línea azul que corre de Cuatro Caminos a Tasqueña, de oeste a sur;
- L_4 corresponde a la línea azul que corre de Tasqueña a Cuatro Caminos, de sur a oeste;
- L_5 corresponde a la línea verde que corre de Indios Verdes a Universidad, de norte a sur y
- L_6 corresponde a la línea verde que corre de Universidad a Indios Verdes, de sur a norte.

Por consiguiente, $N TL = 6$, $N TA_1 = N TA_2 = 20$, $N TA_3 = N TA_4 = 24$ y $N TA_5 = N TA_6 = 21$ (véanse figura A.3 y tabla A.1 en las páginas 71 y 72, respectivamente).

Consideremos un día normal en donde la jornada de trabajo J inicia a las 5 hrs. y termina a las 24 hrs. Esta jornada es dividida en cuatro segmentos de tiempo: $[5, 10]$, $(10, 15]$, $(15, 21]$ y $(21, 24]$. La salida del primer tren de a_1 , para cada línea, es a las 5 hrs. Por convención tomaremos $\gamma^l = 0$ para $l = 1 : 6$.

Como se mencionó anteriormente, el STC Metro de la Ciudad de México nos proporcionó la siguiente información:

- El vector de control, en minutos, utilizado a la fecha

$$\mu = \left(1.9167, 1.9167, 2.1667, 2.1667, 2.0833, 2.0833 \right).$$

Para el Metro de la Ciudad de México $\mu_{seg} = 1.5$ minutos.

- La distancia entre andenes consecutivos $d^l(a_{q-1}, a_q)$, véase tabla A.2 en la página 73 y la velocidad promedio, en km/hr , con la que viajan los trenes de cada línea

$$v = \left(42.5, 42.5, 42.5, 42.5, 42.5, 42.5 \right).$$

Con lo cual obtenemos el tiempo de recorrido entre andenes consecutivos $R^l(a_q)$, véase tabla A.3 en la página 74.

- El tiempo de permanencia del tren en cada andén, $P^l(a_q)$, véase tabla A.4 en la página 75.
- La matriz origen-destino² (de dimensión $\mathbb{R}^{130 \times 130}$),

$$\mathcal{M}_{\mathcal{O}-\mathcal{D}} = \begin{pmatrix} \mathcal{MP}^{1,1} & \mathcal{O}^{1,2} & \mathcal{MP}^{1,3} & \mathcal{MP}^{1,4} & \mathcal{MP}^{1,5} & \mathcal{MP}^{1,6} \\ \mathcal{O}^{2,1} & \mathcal{MP}^{2,2} & \mathcal{MP}^{2,3} & \mathcal{MP}^{2,4} & \mathcal{MP}^{2,5} & \mathcal{MP}^{2,6} \\ \mathcal{MP}^{3,1} & \mathcal{MP}^{3,2} & \mathcal{MP}^{3,3} & \mathcal{O}^{3,4} & \mathcal{MP}^{3,5} & \mathcal{MP}^{3,6} \\ \mathcal{MP}^{4,1} & \mathcal{MP}^{4,2} & \mathcal{O}^{4,3} & \mathcal{MP}^{4,4} & \mathcal{MP}^{4,5} & \mathcal{MP}^{4,6} \\ \mathcal{MP}^{5,1} & \mathcal{MP}^{5,2} & \mathcal{MP}^{5,3} & \mathcal{MP}^{5,4} & \mathcal{MP}^{5,5} & \mathcal{O}^{5,6} \\ \mathcal{MP}^{6,1} & \mathcal{MP}^{6,2} & \mathcal{MP}^{6,3} & \mathcal{MP}^{6,4} & \mathcal{O}^{6,5} & \mathcal{MP}^{6,6} \end{pmatrix}.$$

Véase tabla A.6 en la página 77.

²Para más detalles de la construcción de ésta véase la sección A.1 en la página 63.

- La intensidad de arribos de pasajeros a cada andén, $\lambda^l(a_q)$, véase tabla A.5 en la página 76. Los datos originales proporcionados por el STC Metro de la Ciudad de México fueron trabajados previamente para obtener los datos presentados en la tabla.

Cabe mencionar que el STC Metro de la Ciudad de México no cuenta con datos que proporcionen la cantidad de arribos de pasajeros por segmentos; sino que éstos nos fueron dados por día, sin especificar cuantos entraron en horas pico y cuantos en horas valle. Por esta razón, para las simulaciones, consideramos sólo un segmento de tiempo para un jornada laboral, lo cual provocará que en algunos intervalos de tiempo de la jornada los valores de la simulación estén sobreestimados y en otros subestimados.

Por otro lado, es importante hacer notar que para la simulación parcial

$$E\left(NP_{a_o, a_d}^{m,l}(t)\right) = E\left(NP_{a_o}^m(t)\right) p^{m,l}(a_o, a_d),$$

y

$$\lambda^{m,l}(a_o, a_d) = \lambda^m(a_o) p^{m,l}(a_o, a_d).$$

3.4. Simulación parcial vs. simulación global: resultados comparativos

Con el objetivo de mostrar la precisión y la eficiencia del modelo propuesto, a través de la simulación parcial; en comparación con la simulación global, consideramos una jornada laboral, del Sistema Metro de la Ciudad de México, con una duración de 60 minutos. Para cuestiones de comparación entre las dos simulaciones, consideramos los tiempos de salida de los trenes de manera determinista, lo cual se logra al tomar a la varianza de las variables de fluctuación igual a cero³. Por ende, el costo operativo $CO(\mu)$ será irrelevante, dado que éste es lineal con respecto al número de viajes hechos; por lo cual consideramos $k^l = 0$ para $l = 1 : 6$, además de que no contamos con datos reales para estimar el costo operativo. Por tanto, enfatizaremos el tiempo de espera tomando el costo monetario por cada minuto de espera igual a uno, $c = 1$. Esto es, la función de desempeño obtenida será, numéricamente, idéntica al tiempo de espera total.

En las siguientes gráficas, para L_1 , se muestran los resultados de una simulación del Sistema Metro; se muestran los resultados arrojados por la simulación parcial, en color azul (gráfica punteada), y por la simulación global, en color rojo (gráfica solida), para un mismo tren⁴.

³Su distribución es irrelevante para tiempos de salida deterministas; para tiempos de salida no deterministas, se puede considerar $\delta_j^l(a_q)$ como una v.a. uniforme o normal.

⁴Para una mejor apreciación visual, pese a que las gráficas presentadas a continuación son gráficas discretas, se añadió una gráfica de línea.

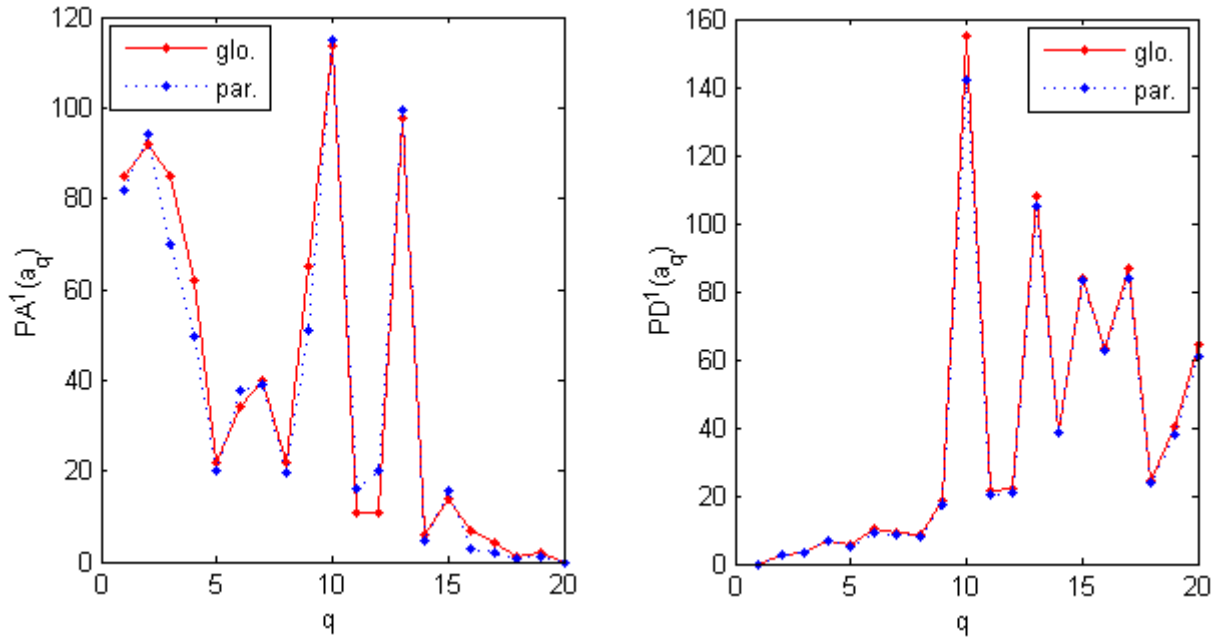


Figura 3.1: Pasajeros que abordan (véase ecuación (1.7)) y pasajeros que abandonan el tren en cada andén (véase ecuación (1.8)), respectivamente.

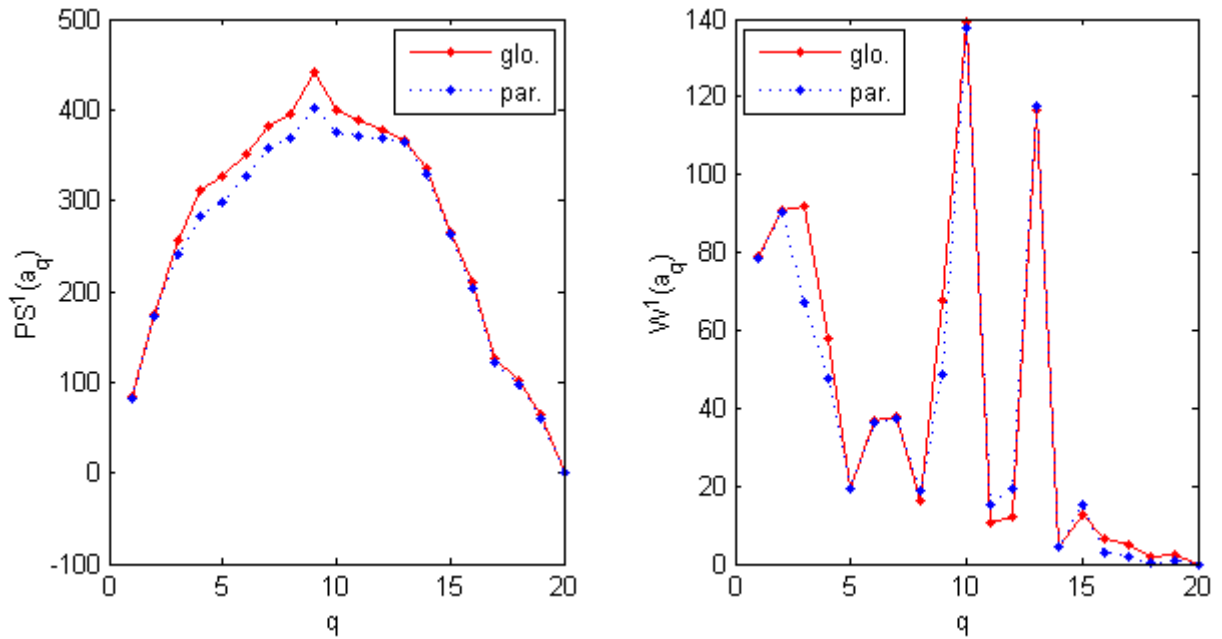


Figura 3.2: Pasajeros en el tren al salir de cada andén (véanse ecuaciones (1.4), (1.5) y (1.6)) y tiempo de espera (en minutos) asociado a cada andén (véase ecuación (1.11)), respectivamente.

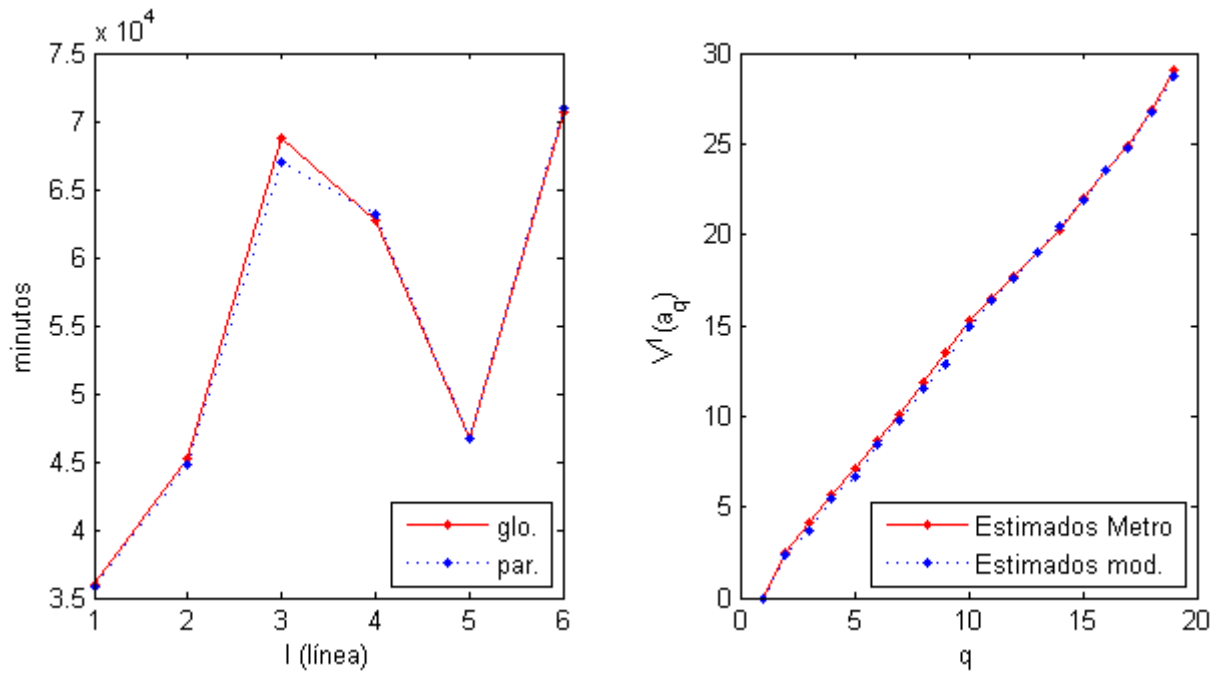


Figura 3.3: Tiempo de espera por línea ($\sum_{j=1}^{NTT^l} \sum_{q=1}^{NTA^l} E(W_j^l(a_q))$ para la simulación parcial y $\sum_{j=1}^{NTT^l} \sum_{q=1}^{NTA^l} W_j^l(a_q)$ para la simulación global) y tiempo acumulado, en minutos, de salida del tren en cada andén (en color azul se muestran los resultados arrojados por la simulación y en color rojo se muestran las estimaciones que el propio Sistema Metro tiene en su manual de “Circulación de Trenes”); respectivamente.

Para estas simulaciones (parcial y global) presentadas parcialmente en las gráficas anteriores, obtuvimos los siguientes resultados:

Tiempo de espera (min.)		Simulación	
		Parcial	Global
Por pasajero		2.4819	2.4784
Por transbordo		53,619.24	53,827.35
Por línea	L_1	35,853.07	35,930.48
	L_2	44,802.54	45,202.72
	L_3	66,975.68	68,807.82
	L_4	63,128.84	62,667.13
	L_5	46,621.92	46,737.66
	L_6	70,910.32	70,726.72
Total		328,292.37	330,072.53
CPU (seg.)		0.28	1.51

Como podemos ver en los resultados numéricos obtenidos, la simulación global requirió un tiempo computacional de 1.51 segundos; mientras que la simulación parcial requirió sólo de 0.28 segundos. Sin embargo, los resultados arrojados por ambas simulaciones son muy similares e incluso, en ocasiones, iguales. Por lo que

podemos presumir de la precisión y rapidez, de la simulación parcial, en el sentido de que proporciona datos confiables del comportamiento del Sistema Metro con un ahorro computacional considerable.

3.4.1. Resultados numéricos simulación Monte-Carlo

Realizamos una simulación Monte-Carlo, utilizando la simulación global, para estimar un intervalo de confianza al 95 % para el tiempo de espera total y para el tiempo de espera por transbordo. Para esto, consideramos 1,000 realizaciones del Sistema Metro y obtuvimos los siguientes resultados:

```

+-----+
|          SIMULACIÓN MONTE-CARLO PARA 1,000 REALIZACIONES DEL          |
|          SISTEMA METRO (SIMULACIÓN GLOBAL)                             |
+-----+
Intervalo de confianza al 95%
para el tiempo de espera por transbordo:  [ 53,608.46,  53,634.26]
Longitud del intervalo:                    25.80490
Longitud relativa:                          0.00048
Error relativo:                             0.00012

Intervalo de confianza al 95%
para el tiempo de espera total:             [328,180.50, 328,380.77]
Longitud del intervalo:                    200.26534
Longitud relativa:                          0.00061
Error relativo:                             0.00016

Tiempo requerido para el cálculo:  807.39 segundos (13.45 minutos)
+-----+
|          FIN SIMULACIÓN (GLOBAL) MONTE-CARLO                          |
+-----+

```

(véase ⁵). De estos últimos resultados y de los resultados obtenidos en la tabla anterior, podemos ver que los tiempos de espera total y por transbordo de la simulación parcial se encuentran dentro de los intervalos de confianza obtenidos por la simulación Monte-Carlo, cuya simulación requirió de un tiempo computacional de 13.45 minutos para 1,000 realizaciones; mientras que una sólo simulación parcial nos proporciona en menos tiempo, datos confiables del comportamiento del Sistema Metro.

3.5. Resultados numéricos del tiempo de espera

Los resultados que veremos en esta sección fueron obtenidos mediante diversos experimentos en los cuales utilizamos sólo simulación parcial. Los datos utilizados para tales simulaciones son los mismos que en la sección

⁵Considerando la notación vista en la sección 2.1.1, para el cálculo de la longitud del intervalo (l_i), de la longitud relativa (l_r) y del error relativo (e_r) utilizamos las ecuaciones $l_i = 2z_{1-\alpha/2}\sqrt{s^2}/\sqrt{R}$, $l_r = l_i/\hat{\theta}_R = (2z_{1-\alpha/2}\sqrt{s^2}) / (\hat{\theta}_R\sqrt{R})$ y $e_r = l_r / (2z_{1-\alpha/2}) = \sqrt{s^2} / (\hat{\theta}_R\sqrt{R})$, respectivamente (para más detalles véase Rubinstein & Kroese [2008]).

anterior; salvo en algunos casos en los que se especifica la variación de algunos datos.

3.5.1. Variación de la longitud de la jornada laboral

En este experimento incrementamos la longitud de la jornada laboral y en los resultados obtenidos observamos cierta tendencia en el tiempo medio de espera por pasajero. Así, el tiempo medio de espera por pasajero es: para L_1 , 1.26 minutos; para L_2 , 1.31 minutos; para L_3 , 1.37; para L_4 , 1.49; para L_5 , 1.35 y para L_6 1.41 minutos. La figura 3.4 muestra estos resultados en donde podemos ver que: L_4 tiene el mayor tiempo medio de espera seguido por L_6 , L_3 , L_5 y L_2 ; L_1 tiene el menor tiempo medio de espera por pasajero. Mientras que, la figura 3.5 muestra que el tiempo medio de espera por pasajero del Sistema Metro de la Ciudad de México es de 1.37 minutos.

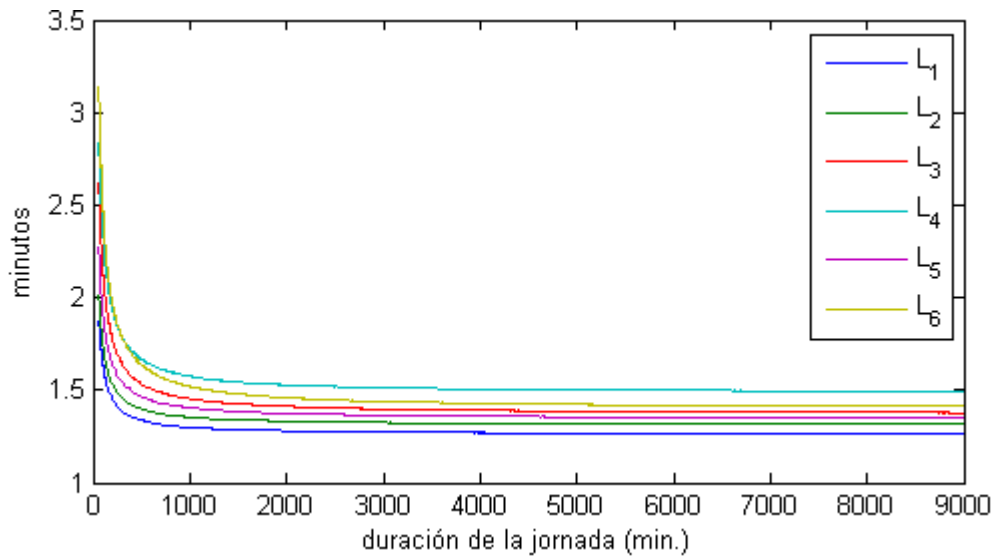


Figura 3.4: Tiempo medio de espera por pasajero (por línea).

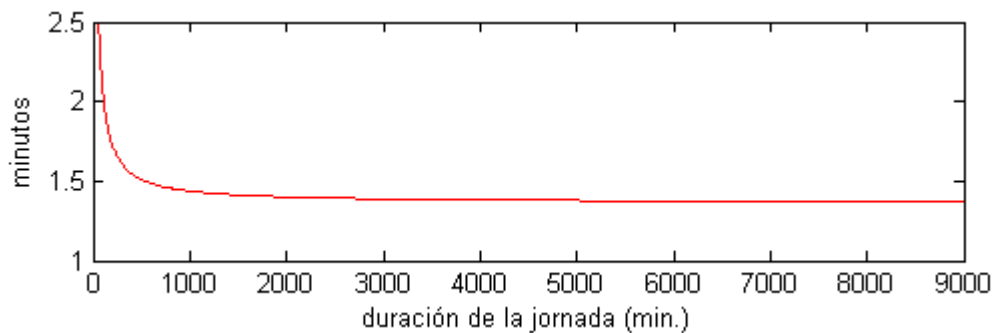


Figura 3.5: Tiempo medio de espera por pasajero.

3.5.2. Variación de la varianza de las variables de fluctuación

La tabla 3.1 muestra los resultados de experimentos para distintos valores de la varianza de la variable de fluctuación, $\sigma^2 = Var(\delta)$. La tabla muestra, para tiempos de salida de trenes no deterministas ($\sigma^2 > 0$), un

intervalo de confianza al 95 % para 100 realizaciones del sistema.

σ^2	Tiempo de espera	$\delta \sim \text{uniforme}(-\sqrt{3}\sigma, \sqrt{3}\sigma)$	$\delta \sim \text{normal}(0, \sigma^2)$
0	total	328,292.3731	
	por pasajero	2.4819	
0.001	total	$329,042.3096 \pm 145.2162$	$328,859.3185 \pm 149.3122$
	por pasajero	2.4869 ± 0.0011	2.4861 ± 0.0011
0.005	total	$331,018.8145 \pm 263.1432$	$331,329.7507 \pm 234.8920$
	por pasajero	2.5001 ± 0.0019	2.5030 ± 0.0018
0.01	total	$334,424.9031 \pm 330.4265$	$334,370.0440 \pm 307.1189$
	por pasajero	2.5209 ± 0.0024	2.5223 ± 0.0023
0.05	total	$353,026.9580 \pm 539.2542$	$352,246.2060 \pm 644.7068$
	por pasajero	2.6613 ± 0.0048	2.6519 ± 0.0047
0.1	total	$369,064.3647 \pm 885.9089$	$368,483.6399 \pm 893.5747$
	por pasajero	2.7796 ± 0.0070	2.7763 ± 0.0066

Tabla 3.1: Tiempo de espera.

Podemos ver que, en promedio, el tiempo de espera total y por persona aumenta a medida que σ^2 aumenta, así mismo, la longitud de los intervalos de confianza se vuelven más grandes. Por otro lado, podemos ver que no hay mucha variación en los resultados obtenidos si consideramos la distribución de las δ 's de manera uniforme o de manera normal⁶; para tiempos de salidas deterministas el uso de una u otra distribución, es irrelevante.

3.6. Sensibilidades del tiempo de espera

Puesto que el problema para optimizar la función de desempeño $F(\mu)$ radica en estimar las sensibilidades del valor esperado del tiempo de espera por transbordo, $WT(\mu)$; ya que el costo operativo $CO(\mu)$ y el valor esperado del tiempo de espera por ingreso al sistema, $WI(\mu)$, pueden ser calculados analíticamente (véanse ecuación (1.9) y proposición 1.4.1). Así, para un segmento fijo

$$\frac{\partial}{\partial \mu^l} CO(\mu) = -k^l \frac{T}{\mu^{l^2}}$$

y

$$\frac{\partial}{\partial \mu^l} E(WI(\mu)) = \sum_{j=1}^{NTT^l} \sum_{q=1}^{NTA^l} \lambda^l(a_q) \mu^l = T \sum_{q=1}^{NTA^l} \lambda^l(a_q), \quad (3.3)$$

donde T es la duración del segmento actual. Por tanto,

$$\frac{\partial}{\partial \mu^l} F(\mu) = \frac{\partial}{\partial \mu^l} CO(\mu) + c \left(\frac{\partial}{\partial \mu^l} E(WI(\mu)) + \frac{\partial}{\partial \mu^l} E(WT(\mu)) \right)$$

⁶La ventaja de considerar una distribución uniforme es que ésta nos permite acotar los valores de las δ 's mientras que una distribución normal no; sin embargo considerar esta última distribución nos permite asegurar que los tiempos entre salidas consecutivas de trenes, Y 's, son v.a. normales, véase ecuación (1.3).

$$= -k^l \frac{T}{\mu^{l^2}} + c \left(T \sum_{q=1}^{NTA^l} \lambda^l(a_q) + \frac{\partial}{\partial \mu^l} E(WT(\mu)) \right),$$

con $E(WT(\mu)) = \sum_{l=1}^{NTL} \sum_{j=1}^{NTT^l} \sum_{q=1}^{NTA^l} E(WT_j^l(a_q))$.

En lo subsecuente usaremos los métodos de SA para la estimación de $\nabla E(WT(\mu))$ y para calcular $\nabla E(WI(\mu))$ hacemos uso de su expresión analítica (3.3). Luego,

$$\nabla E(WI(\mu)) = (17, 131.94, 19, 564.92, 22, 324.55, 19, 166.13, 18, 639.28, 19, 528.66).$$

3.6.1. Estimadores de FD y SP con tiempos de salida de trenes deterministas

Tanto el método de FD como el de SP pueden ser aplicados directamente sin problema alguno; mientras que el uso del método de IPA requiere que $E(WT(\mu))$ satisfaga la condición de Lipschitz (proposición 2.4.2), lo cual no se cumple en este caso: puesto que pequeñas perturbaciones en μ pueden tener como consecuencia que un tren llegue o no dentro de cierto intervalo de tiempo, lo que impactaría considerablemente en el tiempo de espera por transbordo (para más detalles véase Vázquez-Abad & Heidegott [2004]); razón por la cual prescindiremos del uso de este método.

La tabla 3.2 ilustra la precisión del estimador de $\nabla E(WT(\mu))$ obtenido por SP comparado con el obtenido por FD. La última columna muestra el intervalo de confianza al 95 % para el estimador SP. Para obtener estos resultados consideramos: para FD, $h = 10^{-4}$ y sólo una realización y para SP, tomamos $c_0 = 10^{-5}$, consideramos 10^5 realizaciones y $\mathcal{P}(\Delta_0 = 1) = \mathcal{P}(\Delta_0 = -1) = 1/2$ para cada realización.

$\nabla E(WT(\mu))$ \backslash SA	FD	SP
$\frac{\partial}{\partial \mu^1} E(WT(\mu))$	-2,244.97	-2,183.93 ± 252.13
$\frac{\partial}{\partial \mu^2} E(WT(\mu))$	30,938.58	30,822.03 ± 165.09
$\frac{\partial}{\partial \mu^3} E(WT(\mu))$	-12,351.11	-12,165.90 ± 240.97
$\frac{\partial}{\partial \mu^4} E(WT(\mu))$	16,915.15	16,862.61 ± 229.84
$\frac{\partial}{\partial \mu^5} E(WT(\mu))$	-13,446.57	-13,325.39 ± 238.60
$\frac{\partial}{\partial \mu^6} E(WT(\mu))$	9,291.54	9,289.16 ± 245.84

Tabla 3.2: Estimadores de $\nabla E(WT(\mu))$ con tiempos de salida de trenes deterministas.

Un resultado importante de este experimento es que el sistema es sumamente sensible a pequeñas perturbaciones de μ . Por esta razón realizamos multiples experimentos, para FD, para distintos valores de h

cada vez más pequeños, en los cuales observamos que a partir de $h = 10^{-4}$ los valores obtenidos de $\nabla E(WT(\mu))$ se mantienen estables para pequeñas perturbaciones de μ . Esto se debe a que para determinado valor de h , la función de desempeño evaluada en $\mu \pm h/2$ genera el mismo número de salidas de trenes para cada línea y lo mismo ocurre para SP donde $F(\mu \pm c_0)$ genera también el mismo número de salidas de trenes.

3.6.2. Estimadores de FD y SP con tiempos de salidas de trenes no deterministas

Para obtener los resultados numéricos, que muestra la tabla 3.3, simulamos 10,000 realizaciones de un intervalo laboral con una duración de 60 minutos para cada método de SA empleado. Consideramos: $\delta \sim normal(0, 0.001)$, para FD tomamos $h = 10^{-10}$ y para SP tomamos $c_0 = 10^{-9}$.

$\nabla ET(W(\mu))$ \backslash SA	FD	SP
$\frac{\partial}{\partial \mu^1} E(WT(\mu))$	$-3,014.05 \pm 18.96$	$-3,034.37 \pm 819.92$
$\frac{\partial}{\partial \mu^2} E(WT(\mu))$	$31,737.58 \pm 18.17$	$31,885.06 \pm 534.06$
$\frac{\partial}{\partial \mu^3} E(WT(\mu))$	$-12,375.71 \pm 3.71$	$-12,481.61 \pm 784.83$
$\frac{\partial}{\partial \mu^4} E(WT(\mu))$	$17,280.81 \pm 25.31$	$17,071.35 \pm 750.90$
$\frac{\partial}{\partial \mu^5} E(WT(\mu))$	$-13,689.36 \pm 10.36$	$-14,101.73 \pm 774.22$
$\frac{\partial}{\partial \mu^6} E(WT(\mu))$	$9,329.39 \pm 3.59$	$9,698.61 \pm 799.80$

Tabla 3.3: Estimadores de $\nabla E(WT(\mu))$ con tiempos de salida de trenes no deterministas.

Para este experimento, como en el anterior, realizamos multiples experimentos para FD para distintos valores de h cada vez más pequeños, en los cuales observamos que a partir de $h = 10^{-10}$ el valor de la media de $\nabla E(WT(\mu))$ se mantiene estable para pequeñas perturbaciones de μ , además de obtener una longitud del intervalo pequeña y lo mismo ocurre para SP en donde la media del gradiente se mantiene estable a partir de $c_0 = 10^{-9}$ aunque este método nos proporciona una longitud del intervalo mayor. El tiempo computacional requerido por FD fue de 37.55 minutos mientras que SP requirió 6.07 minutos, por lo que SP proporciona un ahorro computacional considerable pero esto se refleja en la obtención de una gran varianza, lo cual podemos ver en la longitud de los intervalos de confianza, por el contrario, FD nos proporciona una varianza menor pero también un costo computacional mayor.

En este capítulo presentamos la implementación computacional, del modelo propuesto en el capítulo 1, para las líneas 1, 2 y 3 del Sistema Metro de la Ciudad de México. De manera adicional, vimos una simulación alternativa, llamada simulación global, para fines de comparación, en cuestiones de precision y rapidez, con la simulación parcial. Lo cual asienta y valida, de una manera práctica, la teoría desarrollada en el capítulo 1.

Realizamos diversos experimentos, en los cuales utilizamos sólo simulación parcial, y observamos el comportamiento del tiempo medio de espera por pasajero para distintos parámetros como: el tamaño de la longitud de un intervalo laboral y la varianza de las variables de fluctuación. También realizamos experimentos para aproximar las sensibilidades del valor esperado del tiempo de espera por transbordo, consideramos tanto el caso determinista ($Var(\delta) = 0$) como el caso no determinista ($Var(\delta) > 0$). Para esto aplicamos los métodos de SA vistos en el capítulo 2.

En la siguiente sección presentamos algunas conclusiones obtenidas a lo largo de la realización de este trabajo, así como algunas perspectivas del mismo y con las cuales, damos por terminada esta tesis.

Conclusiones y perspectivas

En el capítulo 1 presentamos un modelo dinámico del comportamiento de un sistema de transporte urbano tipo Metro, con algunas limitaciones, a través de la metodología de eventos discretos. La implementación computacional de este modelo, a través de la simulación parcial, ahorra un tiempo computacional considerable, a diferencia de la simulación global. También podemos asegurar la precisión de la simulación parcial, en el sentido de que proporciona datos confiables del comportamiento del sistema como: los tiempos de salida y/o llegada de trenes; la cantidad de pasajeros que descenden y/o abordan cada tren; el número de pasajeros que se encuentran viajando en los trenes; el tiempo de espera de los pasajeros y la función de desempeño del sistema.

La implementación computacional del simulador, mediante la simulación parcial, nos proporciona una herramienta útil y rápida, que nos permite “seguir” el comportamiento del sistema conforme avanza el tiempo. Así, usando la simulación parcial, realizamos diversos experimentos en los cuales nos enfocamos en observar el tiempo de espera total y el tiempo medio de espera por pasajero. Se sabe que para tiempos de salidas de trenes deterministas, si sólo se considera una línea, el tiempo medio de espera por persona tiende a $\mu^l/2$. En uno de los experimentos realizados, incrementamos la duración de una jornada laboral para sólo un tipo de segmento de tiempo, y observamos que el tiempo medio de espera por persona tiende aproximadamente a 1.37 minutos; mientras que el mayor tiempo medio de espera por persona corresponde a L_4 (que va de la estación Tasqueña a la estación Cuatro Caminos) con 1.49 minutos y el menor corresponde a L_1 (que va de la estación Pantitlán a la estación Observatorio) con 1.26 minutos. De este modo podemos observar el impacto que tiene el tiempo de espera por transbordo, esto es, si sólo consideramos L_4 entonces el tiempo medio de espera por persona asociado a esta línea tendería a 1.08 minutos y si sólo consideramos L_2 entonces su tiempo medio de espera por persona tendería a 0.96 minutos.

En otro experimento utilizamos simulación Monte-Carlo para obtener un intervalo de confianza de los tiempos de espera, antes mencionados, para distintos valores de $Var(\delta)$. Consideramos a las v.a. δ 's tanto uniformes como normales con media cero y observamos que los intervalos de confianza obtenidos son muy similares para las distintas distribuciones. También observamos que las medias de los tiempos de espera, así como la longitud de los intervalos de confianza aumentan si $Var(\delta)$ aumenta. Por ejemplo: para $Var(\delta) = 0.001$ la media del tiempo

medio de espera por persona es de 2.49 minutos y la longitud de su intervalo de confianza es de 0.0011 minutos y para $Var(\delta) = 0.1$ la media es 2.78 minutos y la longitud de su intervalo es de 0.0070, esto para un intervalo laboral con una duración de 60 minutos.

La teoría desarrollada para la simulación parcial propicia un mejor acercamiento a la optimización estocástica de la función de desempeño de un sistema Metro, a través del uso de métodos de aproximación estocástica que nos permiten estimar el gradiente de la función de desempeño $F(\mu)$ a través de simulación. El problema de optimizar $F(\mu)$ puede ser visto como un problema de minimización con restricciones en donde el vector de control μ está acotado inferiormente.

De manera intuitiva podemos ver que el costo social alcanza su valor mínimo cuando cada una de las componentes del parámetro de control μ es igual a su cota inferior μ_{seg} pero a la vez, esto provoca que el costo operativo se incremente. En la realización de este trabajo no se contó con los datos reales del costo operativo del Sistema Metro de la Ciudad de México. Sin embargo, se da la pauta teórica para la optimización de $F(\mu)$ (que es básicamente la versión estocástica del algoritmo del gradiente proyectado) aunque, dada la ausencia de datos relevantes para su optimización, ésta no la llevamos prácticamente a cabo. No obstante, nos enfocamos simplemente en estimar las sensibilidades del tiempo de espera a través de métodos de aproximación estocástica.

En este trabajo vimos los métodos de aproximación estocástica: diferencias finitas (FD), perturbación simultánea (SP), análisis de perturbación infinitesimal (IPA) y razón de verosimilitud (LR) los cuales permiten estimar el gradiente de una función, para la optimización de sistemas de eventos discretos, a través de simulación. El método de SP ofrece el mayor ahorro computacional, seguido por el método de FD; IPA requiere un tiempo computacional mayor que los otros. Los métodos de FD y SP pueden ser aplicados directamente a cualquier sistema que pueda ser simulado, puesto que no requieren de algún estudio de la función de desempeño ni de la distribución de sus entradas, pero estos proporcionan estimadores sesgados del gradiente de la función de desempeño del sistema simulado. Una de las ventajas de SP, sobre FD, es que FD requiere de $2p$ simulaciones para la estimación del gradiente; mientras que SP sólo requiere de 2 simulaciones, independientemente de la dimensión, p , del problema. Por el contrario, IPA y LR requieren un estudio detallado de la función de desempeño, por lo que no siempre estos métodos pueden ser aplicados; aunque, cuando estos pueden ser empleados, los estimadores que proporcionan son estimadores insesgados.

Con el objetivo de ilustrar y profundizar un poco más en la implementación de estos métodos de aproximación estocástica, aplicamos los métodos de FD, SP e IPA para la optimización de un sistema de colas, del tipo M/U/1, a través de simulación. Realizamos experimentos para mostrar la viabilidad de estos métodos de aproximación. Los resultados numéricos simulados fueron muy prometedores, en el sentido de que validan de una forma práctica la aplicación de métodos de SA a la optimización de sistemas de eventos discretos a través de simulación.

Posteriormente, aplicamos los métodos de FD y SP para estimar las sensibilidades del tiempo de espera del Sistema Metro de la Ciudad de México. Un resultado importante de estos experimentos es que el gradiente del tiempo de espera es sumamente sensible a pequeñas perturbaciones de μ aunque a partir de cierto tamaño de

perturbación el valor del gradiente se comporta de manera estable. Por otra parte, los métodos de FD y SP arrojan resultados numéricos similares, por lo que uno pensaría que estos resultados son fiables⁷.

El modelo propuesto en este trabajo puede, por lo tanto, ser la base de un sistema de adaptación inteligente que, como vimos en el capítulo 3, mejora los métodos de optimización usados hoy en día en los sistemas Metro. Sin embargo, este modelo es un primer acercamiento a la realidad del sistema, ya que éste es perfectible en cuanto a sus limitaciones:

- Controlar el tiempo entre salidas consecutivas de trenes, en los andenes iniciales, y mantener una velocidad promedio a lo largo de una línea: garantiza que estos mantendrán, al menos, la distancia mínima permitida por cuestiones de seguridad a lo largo del trayecto (y así evitar un choque por alcance). Una mejora que se le podría agregar al modelo que presentamos: es considerar que el Sistema Metro de la Ciudad de México cuenta con semáforos ubicados entre andenes consecutivos, cuya función es alertar al operador del tren: si el semáforo se encuentra en verde entonces puede continuar con el viaje y si está en rojo (significa que el tren se ubica muy cerca del tren que le antecede) entonces el operador debe detener el tren hasta que el semáforo cambie a verde. Esto permitiría que la velocidad con la que viajan los trenes sea variable a lo largo del trayecto, como sucede en la realidad.
- En el modelo presentado se asume que el cupo de cada tren es ilimitado. Una mejora del modelo podría ser: considerar el cupo máximo de cada tren y trabajar con los pasajeros rezagados. Esto es, considerar que no todos los pasajeros que esperan en el andén, pueden abordar el primer tren que llega; sino que éstos deben dejar pasar, tal vez uno, dos o más trenes: antes de poder abordar uno.
- Otra mejora al modelo es considerar, en las estaciones de correspondencia, el tiempo promedio que toma a los pasajeros, que transbordan, caminar de un andén a otro para cambiar de línea. Esto repercutiría en un decremento en el tiempo de espera por transbordo.

Estos puntos son sólo algunas sugerencias de las mejoras que se podrían aplicar al modelo; pensando en un trabajo a futuro. Por otro lado, una manera de mejorar los resultados de las simulaciones, para una jornada laboral completa, es considerar la naturaleza de los datos que la alimentan. Esto es:

- El STC Metro de la Ciudad de México no cuenta con datos que proporcionen las intensidades de arribos de pasajeros tanto en horas pico como en horas valle, sino que estos datos fueron recolectados por día; lo cual provoca que en algunos intervalos de tiempo los resultados de la simulación estén sobreestimados y en otros subestimados. Esto representa una área de oportunidad para mejorar los resultados simulados.
- La matriz origen-destino no es estática: resulta natural considerar una matriz origen-destino para el horario matutino y otra distinta para el vespertino; puesto que en las mañanas, las personas acuden de sus casas

⁷Los códigos fuentes de los programas utilizados a lo largo de este trabajo, fueron puestos a disposición del Departamento de Matemáticas de la División de Ciencias Básicas e Ingeniería de la Universidad Autónoma Metropolitana Unidad Iztapalapa, mismos que están disponibles al lector.

a sus trabajos y/o escuelas y en las tardes, estas personas regresan. El Sistema Metro cuenta con éstas matrices origen-destino (tanto para horas pico como para horas valle), que obtiene y actualiza cada 10 años a través de un sondeo de campo que realiza. Sin embargo, es importante estimar estas matrices entre un sonde y otro, debido al crecimiento económico de la ciudad, esto es, la construcción de nuevos espacios que representan nodos atractores de urbanización (como centros comerciales, escuelas, fabricas, etc.); lo cual generarían la necesidad de modificar y/o ajustar las matrices. Existen maneras de estimar éstas a través de métodos conocidos como métodos de balance.

Es importante mencionar que, dado que la estación inicial de una línea es la estación final de su línea opuesta, los trenes que terminan su recorrido son cambiados de línea e inician un nuevo recorrido como trenes de la línea opuesta. Las frecuencias de salidas de trenes de una línea y su línea opuesta no siempre son iguales, lo cual implica: una acumulación de trenes (en espera de iniciar su recorrido) en la estación inicial de la línea con menor frecuencia y una escasez de trenes en la estación terminal de la línea con mayor frecuencia. Para el primer caso: el Sistema Metro cuenta con patios, en cada estación terminal, diseñados para posibilitar el cambio de un tren entre líneas (opuestas); además, en estos se almacenan los trenes: en espera de iniciar el recorrido y/o al finalizar cada jornada laboral. Estos patios tienen una capacidad limitada de almacenaje, la cual no consideramos en la realización de este trabajo.

El parque vehicular asignado a cada línea no siempre satisface la demanda de salidas de trenes, cuando su parámetro de control y el parámetro de control de su línea opuesta son los suficientemente diferentes. Para evitar este problema, se requiere que los valores de los parámetros de control, de estas líneas, sean lo suficientemente cercanos como para evitar la acumulación de trenes estacionados en el patio de una línea; y la escasez de estos en su línea opuesta.

Apéndice

En esta sección presentamos la construcción de la matriz origen-destino, que utilizamos para alimentar las simulaciones cuyos resultados presentamos en el capítulo 3.

A.1. Construcción de la matriz origen-destino

El STC Metro de la Ciudad de México nos proporcionó las matrices origen-destino para los distintos segmentos de tiempo: para horas pico y horas valle matutinas, de 5 hrs. a 10 hrs. y de 10 hrs. a 15 hrs., respectivamente y para horas pico y horas valle vespertinas, de 15 hrs. a 21 hrs. y de 21 hrs. a 24 hrs., respectivamente. Para fines de este trabajo se consideró la matriz origen-destino diaria acumulada, esto es, sin distinguir los segmentos (matutino o vespertino, pico o valle).

Paso 1. De la matriz origen-destino diaria acumulada que nos proporcionó el STC Metro de la Ciudad de México, consideramos sólo los datos correspondientes a las líneas 1 (rosa), 2 (azul) y 3 (verde). Así, trabajamos con una matriz cuadrada de dimensión 65, cuyos elementos en la diagonal son todos iguales a cero, puesto que no hay pasajeros cuyos andén origen y andén destino sean el mismo, y ésta tiene la siguiente forma:

o \ d		Línea rosa			Línea azul			Línea verde		
		PAN	...	OBS	CCA	...	TAS	INV	...	UNI
Línea rosa	PAN	0	...	205	329	...	104	88	...	74

Línea azul	OBS	433	...	0	55	...	149	352	...	121

Línea verde	CCA	329	...	43	0	...	830	626	...	302

Línea verde	TAS	172	...	166	859	...	0	470	...	38

Línea verde	INV	133	...	244	778	...	518	0	...	1211

Línea verde	UNI	69	...	68	201	...	39	749	...	0

donde sus elementos indican número de personas. En este trabajo se toma esta matriz para construir una matriz por bloques, que muestre la información para 6 líneas (L_1, L_2, \dots, L_6) en lugar de 3 (línea rosa, línea azul y línea verde) véase ⁸. Cuya matriz tiene la forma:

$$M_{O-D} = \begin{pmatrix} MP^{1,1} & O^{1,2} & MP^{1,3} & MP^{1,4} & MP^{1,5} & MP^{1,6} \\ O^{2,1} & MP^{2,2} & MP^{2,3} & MP^{2,4} & MP^{2,5} & MP^{2,6} \\ MP^{3,1} & MP^{3,2} & MP^{3,3} & O^{3,4} & MP^{3,5} & MP^{3,6} \\ MP^{4,1} & MP^{4,2} & O^{4,3} & MP^{4,4} & MP^{4,5} & MP^{4,6} \\ MP^{5,1} & MP^{5,2} & MP^{5,3} & MP^{5,4} & MP^{5,5} & O^{5,6} \\ MP^{6,1} & MP^{6,2} & MP^{6,3} & MP^{6,4} & O^{6,5} & MP^{6,6} \end{pmatrix},$$

la cual es cuadrada y tiene dimensión 130.

Paso 2. Para obtener las sub-matrices $MP^{l,l}$, con $l = 1 : 6$, (véase ⁹) procedemos de la siguiente manera:

Para obtener $MP^{1,1}$ y $MP^{2,2}$ tomamos la sección correspondiente a la línea rosa:

o \ d		Línea rosa								
		PAN	ZAR	GOM	BOU	...	CHP	JNA	TCY	OBS
Línea rosa	PAN	0	241	113	224	...	600	125	188	205
	ZAR	85	0	123	166	...	418	81	157	147
	GOM	116	32	0	64	...	256	62	81	151
	BOU	91	55	16	0	...	197	62	98	102

	CHP	840	438	138	370	...	0	14	45	69
	JNA	72	45	13	28	...	47	0	9	24
	TCY	190	139	52	86	...	435	134	0	55
	OBS	433	265	75	160	...	458	112	187	0

⁸Recordemos que L_1 y L_2 pertenecen a la línea rosa, L_3 y L_4 pertenecen a la azul y L_5 y L_6 pertenecen a la línea verde.

⁹Recordemos que éstas son las matrices cuya línea origen, de los pasajeros, es igual a la línea destino.

La parte superior de la diagonal¹⁰ corresponde a L_1 , que va de Pantitlán a Observatorio, y la parte inferior corresponde a L_2 , que va de Observatorio a Pantitlán.

Ahora, partimos la matriz anterior en dos matrices: una triangular superior y otra triangular inferior. Luego las reordenamos, en orden ascendente según el número de andén, para obtener las matrices $\mathcal{MP}^{1,1}$ y $\mathcal{MP}^{2,2}$.

			L_1								
		L_d	PAN	ZAR	GOM	BOU	...	CHP	JNA	TCY	OBS
L_o											
	a_d	a_o	a_1	a_2	a_3	a_4	...	a_{17}	a_{18}	a_{19}	a_{20}
L_1	PAN	a_1	0	241	113	224	...	600	125	188	205
	ZAR	a_2	0	0	123	166	...	418	81	157	147
	GOM	a_3	0	0	0	64	...	256	62	81	151
	BOU	a_4	0	0	0	0	...	197	62	98	102
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	CHP	a_{17}	0	0	0	0	...	0	14	45	69
	JNA	a_{18}	0	0	0	0	...	0	0	9	24
	TCY	a_{19}	0	0	0	0	...	0	0	0	55
	OBS	a_{20}	0	0	0	0	...	0	0	0	0

			L_2								
		L_d	OBS	TCY	JNA	CHP	...	BOU	GOM	ZAR	PAN
L_o											
	a_d	a_o	a_1	a_2	a_3	a_4	...	a_{17}	a_{18}	a_{19}	a_{20}
L_2	OBS	a_1	0	187	112	458	...	160	75	265	433
	TCY	a_2	0	0	134	435	...	86	52	139	190
	JNA	a_3	0	0	0	47	...	28	13	45	72
	CHP	a_4	0	0	0	0	...	370	138	438	840
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	BOU	a_{17}	0	0	0	0	...	0	16	55	91
	GOM	a_{18}	0	0	0	0	...	0	0	32	116
	ZAR	a_{19}	0	0	0	0	...	0	0	0	85
	PAN	a_{20}	0	0	0	0	...	0	0	0	0

El mismo procedimiento se realiza para obtener las matrices $\mathcal{MP}^{3,3}$ y $\mathcal{MP}^{4,4}$ ($\mathcal{MP}^{5,5}$ y $\mathcal{MP}^{6,6}$), salvo que ahora se trabaja con la sección correspondiente a la línea azul (línea verde).

Paso 3. Como vimos en la sección 3.1, cuando L_m es la línea opuesta a L_l la matriz $\mathcal{MP}^{m,l}$ es una matriz de ceros de dimensión $\mathbb{R}^{NTA^m \times NTA^l}$ y la denotamos por $\mathcal{O}^{m,l}$. Aquí, L_1 es opuesta a L_2 , L_3 es opuesta a L_4

¹⁰Observemos que los elementos de la diagonal son todos iguales a cero.

y L_5 es opuesta a L_6 , y viceversa. Esto es, $\mathcal{O}^{2k-1,2k}$ y $\mathcal{O}^{2k,2k-1}$, con $k = 1 : 3$, son matrices de ceros de dimensión $\mathbb{R}^{(2k-1) \times 2k}$ y $\mathbb{R}^{2k \times (2k-1)}$, respectivamente.

Paso 4. Para el resto de las matrices (cuando la línea origen y la línea destino son distintas, pero no son líneas opuestas) procedemos de la siguiente manera:

- i. Consideremos la sección correspondiente a la línea rosa (como línea origen) y a la línea azul (como línea destino).

		Línea azul								
		CCA	PTN	...	ZOC	PIN	SAN	CHB	...	TAS
Línea rosa	PAN	329	30	...	307	0	29	18	...	104
	ZAR	285	17	...	143	0	30	22	...	97
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	MER	176	1	...	45	0	15	20	...	82
	PIN	0	0	...	0	0	0	0	...	0
	ISA	39	1	...	8	0	10	9	...	56
	SAL	15	1	...	13	0	6	4	...	23
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	OBS	55	6	...	174	0	30	16	...	149

Los elementos de la columna y de la fila correspondientes a la estación Pino Suárez (PIN) son todos iguales a cero; esto se debe a que la línea rosa y la línea azul tienen correspondencia en esta estación. Por tanto, en esta matriz, no se consideran a los pasajeros que ingresan en Pino Suárez por la línea rosa para bajar en alguna estación de la línea azul (ya que en este caso se considera que estos pasajeros ingresaron al sistema en Pino Suárez pero por la línea azul, así, estos pasajeros son contabilizados en las matrices correspondientes a la línea azul como línea origen y destino, esto es: en las matrices $\mathcal{MP}^{3,3}$ y $\mathcal{MP}^{4,4}$). Lo mismo ocurre cuando los pasajeros ingresan en cualquier estación de la línea rosa para abandonar el sistema en Pino Suárez por la línea azul (ya que en este caso se considera que estos pasajeros abandonan el sistema en Pino Suárez pero por la línea rosa, así, estos pasajeros son contabilizados en las matrices correspondientes a la línea rosa como línea origen y destino, esto es: en las matrices $\mathcal{MP}^{1,1}$ y $\mathcal{MP}^{2,2}$).

En la matriz anterior podemos ver que, si tomamos como ejes de referencia a la columna y a la fila correspondientes a la estación de correspondencia Pino Suárez, la parte superior izquierda corresponde a la matriz de las personas que ingresan al sistema en L_1 y bajan en L_4 , la parte superior derecha corresponde a la matriz de las personas que ingresan al sistema en L_1 y bajan en L_3 , la parte inferior izquierda corresponde a las personas que ingresan al sistema en L_2 y bajan en L_4 y la parte inferior derecha corresponde a la matriz de las personas que ingresan al sistema en L_2 y bajan en L_3 . Por tanto, si descomponemos la matriz en cuatro, la reordenamos y agregamos algunos ceros: obtenemos las matrices $\mathcal{MP}^{1,4}$, $\mathcal{MP}^{1,3}$, $\mathcal{MP}^{2,4}$ y $\mathcal{MP}^{2,3}$.

		L_d	L_4							
			TAS	...	PIN	ZOC	ALL	...	PTN	CCA
L_o		a_o	a_d							
			a_1	...	a_{11}	a_{12}	a_{13}	...	a_{23}	a_{24}
L_1	PAN	a_1	0	...	0	307	95	...	30	329
	ZAR	a_2	0	...	0	143	62	...	17	285
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	CAN	a_8	0	...	0	19	5	...	1	31
	MER	a_9	0	...	0	45	14	...	1	176
	PIN	a_{10}	0	...	0	0	0	...	0	0
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	OBS	a_{20}	0	...	0	0	0	...	0	0

		L_d	L_3							
			CCA	...	PIN	SAN	CHB	...	GEN	TAS
L_o		a_o	a_d							
			a_1	...	a_{14}	a_{15}	a_{16}	...	a_{23}	a_{24}
L_1	PAN	a_1	0	...	0	29	18	...	25	104
	ZAR	a_2	0	...	0	30	22	...	28	97
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	CAN	a_8	0	...	0	9	6	...	7	28
	MER	a_9	0	...	0	15	20	...	29	82
	PIN	a_{10}	0	...	0	0	0	...	0	0
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	OBS	a_{20}	0	...	0	0	0	...	0	0

		L_d	L_4							
			TAS	...	PIN	ZOC	ALL	...	PTN	CCA
L_o		a_o	a_d							
			a_1	...	a_{11}	a_{12}	a_{13}	...	a_{23}	a_{24}
L_2	OBS	a_1	0	...	0	174	20	...	6	55
	TCY	a_2	0	...	0	91	13	...	1	19
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	SAL	a_9	0	...	0	13	1	...	1	15
	ISA	a_{10}	0	...	0	8	2	...	1	39
	PIN	a_{11}	0	...	0	0	0	...	0	0
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	PAN	a_{20}	0	...	0	0	0	...	0	0

		L_d	L_3							
			CCA	...	PIN	SAN	CHB	...	GEN	TAS
L_o		a_d	a_1	...	a_{14}	a_{15}	a_{16}	...	a_{23}	a_{24}
		a_o								
L_2	OBS	a_1	0	...	0	30	16	...	16	149
	TCY	a_2	0	...	0	8	3	...	6	38
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	SAL	a_9	0	...	0	6	4	...	6	23
	ISA	a_{10}	0	...	0	10	9	...	5	56
	PIN	a_{11}	0	...	0	0	0	...	0	0
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	PAN	a_{20}	0	...	0	0	0	...	0	0

- ii. Para obtener las matrices $MP^{1,5}$, $MP^{1,6}$, $MP^{2,5}$ y $MP^{2,6}$ repetimos el proceso descrito en i. considerando ahora la sección correspondiente a la línea rosa, como línea origen, y a la línea verde, como línea destino.
- iii. Luego, para construir el resto de las matrices se repiten los procesos descritos en i. y ii. considerando las secciones correspondiente a la línea azul, como línea origen, y posteriormente a la línea verde, también como línea origen.

Paso 5. Ahora tenemos 36 matrices ($MP^{m,l}$, con $m, l = 1 : 6$) que fueron construidas como se indica en los pasos anteriores y las cuales forman la matriz origen-destino

$$M_{O-D} = \begin{pmatrix} MP^{1,1} & O^{1,2} & MP^{1,3} & MP^{1,4} & MP^{1,5} & MP^{1,6} \\ O^{2,1} & MP^{2,2} & MP^{2,3} & MP^{2,4} & MP^{2,5} & MP^{2,6} \\ MP^{3,1} & MP^{3,2} & MP^{3,3} & O^{3,4} & MP^{3,5} & MP^{3,6} \\ MP^{4,1} & MP^{4,2} & O^{4,3} & MP^{4,4} & MP^{4,5} & MP^{4,6} \\ MP^{5,1} & MP^{5,2} & MP^{5,3} & MP^{5,4} & MP^{5,5} & O^{5,6} \\ MP^{6,1} & MP^{6,2} & MP^{6,3} & MP^{6,4} & O^{6,5} & MP^{6,6} \end{pmatrix}.$$

Pero ésta aún no es una matriz de probabilidades, ya que cada elemento de la matriz representa un número de personas. Por tanto, para convertirla en una matriz de probabilidades debemos dividir cada uno de sus elementos entre la suma de los elementos de su fila. Así, obtenemos una matriz tal que la suma por filas nos da uno, a excepción de las filas correspondientes a los últimos andenes de cada línea: en las cuales su suma es igual a cero (véanse ecuaciones 3.1 y 3.2). Véase la matriz final en la tabla A.6 en la página 77.



Anexos



Figura A.1: Unidad de servicio del Sistema Metro de la Ciudad de México. (Ref: http://es.wikipedia.org/wiki/Archivo:Metro_Mexico_DF_MP68_R93_01.jpg)

SISTEMA DE TRANSPORTE COLECTIVO

Red del Metro



Figura A.2: Mapa del STC Red del Metro de la Ciudad de México. (Ref: <http://www.metro.df.gob.mx/>)

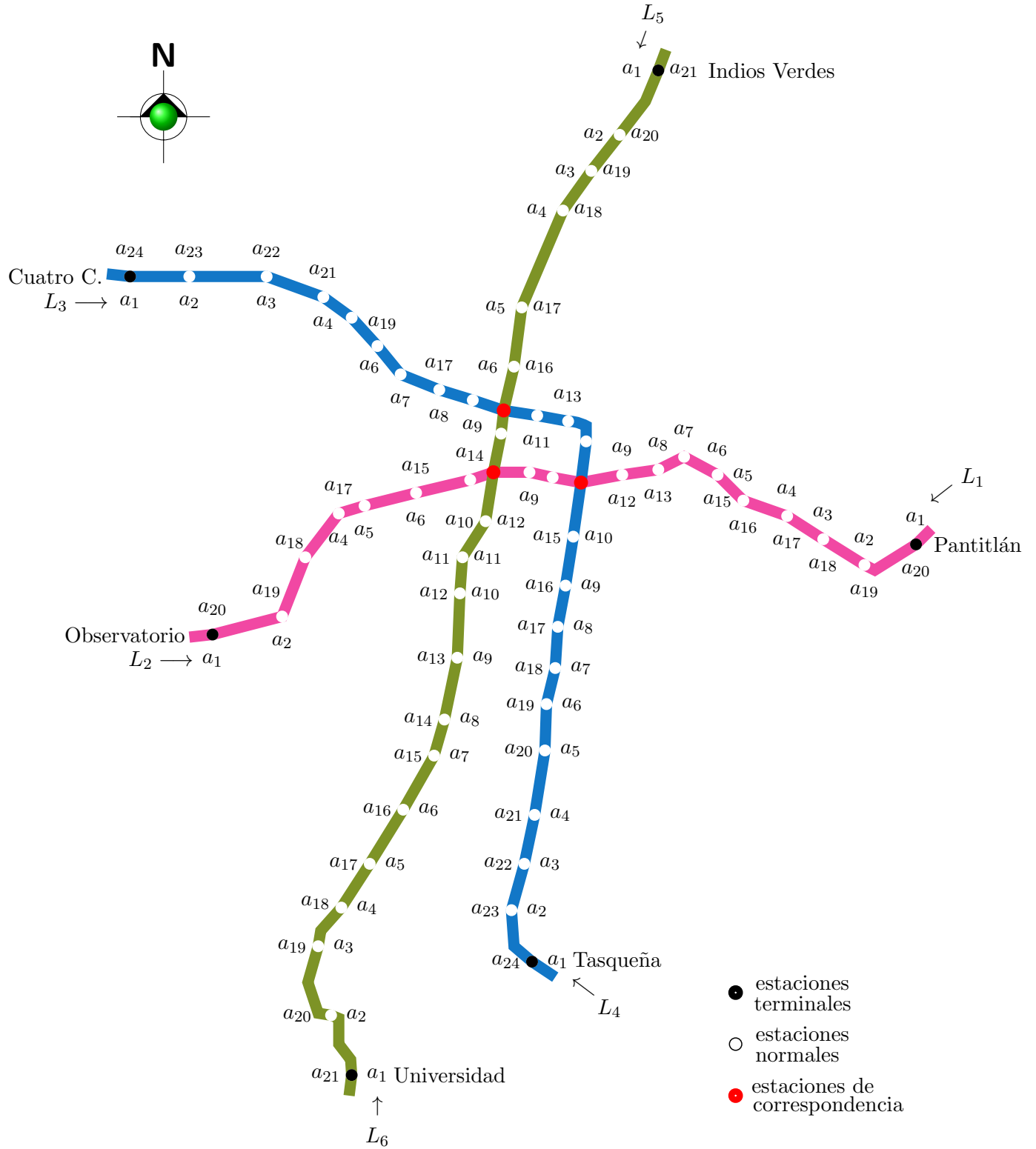


Figura A.3: Líneas 1, 2 y 3 del Sistema Metro de la Ciudad de México. (Ref: elaboración propia con información de <http://www.metro.df.gob.mx/>)

Línea rosa			Línea azul			Línea verde		
Línea	L_1	L_2	Línea	L_3	L_4	Línea	L_5	L_6
Estación	Andén		Estación	Andén		Estación	Andén	
Pantitlán	a_1	a_{20}	Cuatro C.	a_1	a_{24}	Indios Verdes	a_1	a_{21}
Zaragoza	a_2	a_{19}	Panteones	a_2	a_{23}	D. 18 de M.	a_2	a_{20}
Gómez Farias	a_3	a_{18}	Tacuba	a_3	a_{22}	Potrero	a_3	a_{19}
Blvd. Pto. A.	a_4	a_{17}	Cuitlahuac	a_4	a_{21}	La Raza	a_4	a_{18}
Balbuena	a_5	a_{16}	Popotla	a_5	a_{20}	Tlatelolco	a_5	a_{17}
Moctezuma	a_6	a_{15}	Colegio M.	a_6	a_{19}	Guerrero	a_6	a_{16}
San Lázaro	a_7	a_{14}	Normal	a_7	a_{18}	Hidalgo	a_7	a_{15}
Candelaria	a_8	a_{13}	San Cosme	a_8	a_{17}	Juárez	a_8	a_{14}
Merced	a_9	a_{12}	Revolución	a_9	a_{16}	Balderas	a_9	a_{13}
Pino Suárez	a_{10}	a_{11}	Hidalgo	a_{10}	a_{15}	Niños Héroe	a_{10}	a_{12}
I. la Católica	a_{11}	a_{10}	Bellas Artes	a_{11}	a_{14}	Hospital Gral.	a_{11}	a_{11}
Salto del Agua	a_{12}	a_9	Allende	a_{12}	a_{13}	Centro Médico	a_{12}	a_{10}
Balderas	a_{13}	a_8	Zócalo	a_{13}	a_{12}	Etiopía	a_{13}	a_9
Cuauhtémoc	a_{14}	a_7	Pino Suárez	a_{14}	a_{11}	Eugenia	a_{14}	a_8
Insurgentes	a_{15}	a_6	S. Antonio A.	a_{15}	a_{10}	Div. del Norte	a_{15}	a_7
Sevilla	a_{16}	a_5	Chabacano	a_{16}	a_9	Zapata	a_{16}	a_6
Chapultepec	a_{17}	a_4	Viaducto	a_{17}	a_8	Coyoacán	a_{17}	a_5
Juanacatlán	a_{18}	a_3	Xola	a_{18}	a_7	Viveros	a_{18}	a_4
Tacubaya	a_{19}	a_2	V. de Cortés	a_{19}	a_6	M. A. de Q.	a_{19}	a_3
Observatorio	a_{20}	a_1	Nativitas	a_{20}	a_5	Copilco	a_{20}	a_2
			Portales	a_{21}	a_4	Universidad	a_{21}	a_1
			Ermita	a_{22}	a_3			
			Gral. Anaya	a_{23}	a_2			
			Tasqueña	a_{24}	a_1			

Tabla A.1: Andenes y estaciones por línea.

Andén \ Línea	Línea					
	L_1	L_2	L_3	L_4	L_5	L_6
a_1	—	—	—	—	—	—
a_2	1470	1412	1,789	1,480	1,316	1,456
a_3	912	1308	1,566	988	1,116	1,445
a_4	761	1123	787	898	1,256	974
a_5	745	651	770	1,074	1,595	1,058
a_6	853	795	612	900	1,192	1,303
a_7	628	943	666	848	852	944
a_8	1016	559	807	640	401	865
a_9	848	608	687	924	809	1,100
a_{10}	895	595	737	792	815	1,269
a_{11}	532	532	597	967	709	803
a_{12}	595	895	537	895	803	709
a_{13}	608	848	752	752	1,269	815
a_{14}	559	1016	895	537	1,100	809
a_{15}	943	628	967	597	865	401
a_{16}	795	853	792	737	944	852
a_{17}	651	745	924	687	1,303	1,192
a_{18}	1123	761	640	807	1,058	1,595
a_{19}	1308	912	848	666	974	1,256
a_{20}	1412	1470	900	612	1,445	1,116
a_{21}	—	—	1,074	770	1,456	1,316
a_{22}	—	—	898	787	—	—
a_{23}	—	—	988	1,566	—	—
a_{24}	—	—	1,480	1,789	—	—

Tabla A.2: Distancia (en metros) entre andenes consecutivos.

Andén \ Línea	Línea					
	L_1	L_2	L_3	L_4	L_5	L_6
a_1	—	—	—	—	—	—
a_2	34.5882	33.2235	42.0941	34.8235	30.9647	34.2588
a_3	21.4588	30.7765	36.8471	23.2471	26.2588	34.0000
a_4	17.9059	26.4235	18.5176	21.1294	29.5529	22.9176
a_5	17.5294	15.3176	18.1176	25.2706	37.5294	24.8941
a_6	20.0706	18.7059	14.4000	21.1765	28.0471	30.6588
a_7	14.7765	22.1882	15.6706	19.9529	20.0471	22.2118
a_8	23.9059	13.1529	18.9882	15.0588	9.4353	20.3529
a_9	19.9529	14.3059	16.1647	21.7412	19.0353	25.8824
a_{10}	21.0588	14.0000	17.3412	18.6353	19.1765	29.8588
a_{11}	12.5176	12.5176	14.0471	22.7529	16.6824	18.8941
a_{12}	14.0000	21.0588	12.6353	21.0588	18.8941	16.6824
a_{13}	14.3059	19.9529	17.6941	17.6941	29.8588	19.1765
a_{14}	13.1529	23.9059	21.0588	12.6353	25.8824	19.0353
a_{15}	22.1882	14.7765	22.7529	14.0471	20.3529	9.4353
a_{16}	18.7059	20.0706	18.6353	17.3412	22.2118	20.0471
a_{17}	15.3176	17.5294	21.7412	16.1647	30.6588	28.0471
a_{18}	26.4235	17.9059	15.0588	18.9882	24.8941	37.5294
a_{19}	30.7765	21.4588	19.9529	15.6706	22.9176	29.5529
a_{20}	33.2235	34.5882	21.1765	14.4000	34.0000	26.2588
a_{21}	—	—	25.2706	18.1176	34.2588	30.9647
a_{22}	—	—	21.1294	18.5176	—	—
a_{23}	—	—	23.2471	36.8471	—	—
a_{24}	—	—	34.8235	42.0941	—	—

Tabla A.3: Tiempo de recorrido (en segundos) entre andenes consecutivos.

Andén \ Línea	Línea					
	L_1	L_2	L_3	L_4	L_5	L_6
a_1	–	–	–	–	–	–
a_2	24	24	16	20	35	30
a_3	24	24	30	20	30	30
a_4	24	24	18	20	35	25
a_5	24	24	18	18	30	25
a_6	24	24	20	18	30	30
a_7	30	24	18	18	35	30
a_8	24	30	18	18	30	25
a_9	24	26	20	18	35	30
a_{10}	33	26	26	18	25	35
a_{11}	24	33	20	30	30	30
a_{12}	24	24	20	30	35	25
a_{13}	30	24	30	20	30	35
a_{14}	24	30	30	20	25	30
a_{15}	24	26	18	26	30	35
a_{16}	24	24	18	20	30	30
a_{17}	24	24	18	18	25	30
a_{18}	24	24	18	18	25	35
a_{19}	24	24	18	18	30	30
a_{20}	–	–	18	18	30	35
a_{21}	–	–	20	20	–	–
a_{22}	–	–	20	30	–	–
a_{23}	–	–	20	16	–	–
a_{24}	–	–	–	–	–	–

Tabla A.4: Tiempo de permanencia (en segundos) de los trenes en cada andén.

Andén \ Línea	Línea					
	L_1	L_2	L_3	L_4	L_5	L_6
a_1	42.7132	56.1355	116.5546	75.5748	114.0039	60.1373
a_2	49.1675	29.0111	5.8621	22.6595	35.7887	37.8856
a_3	36.4846	5.8639	28.0656	14.0189	17.9187	24.8138
a_4	25.8508	51.9742	14.8642	10.4449	34.3653	22.8076
a_5	10.3754	24.5936	6.1556	14.0541	20.1037	21.9503
a_6	19.6639	44.5300	15.1663	9.3364	8.0905	27.7101
a_7	20.2881	17.6980	32.8704	11.0159	12.4427	17.0107
a_8	10.3330	16.0789	15.8848	12.2029	10.6171	12.9472
a_9	26.5839	10.9959	19.5469	16.0397	2.3545	24.6735
a_{10}	8.7594	12.7414	20.3422	14.1837	2.8712	8.3941
a_{11}	8.3974	16.9699	4.7355	8.5843	10.0424	15.6254
a_{12}	10.4761	14.8357	8.3296	39.0867	13.7156	14.3445
a_{13}	2.3024	9.2997	31.5839	19.9444	6.2235	5.7253
a_{14}	2.3804	3.9959	4.4479	14.8541	5.0206	8.7096
a_{15}	8.2507	5.0444	5.2750	8.5214	4.0200	5.9437
a_{16}	1.5951	2.6197	8.2299	13.2534	9.6763	3.7563
a_{17}	1.1187	1.4196	7.4848	2.9336	1.8257	1.8626
a_{18}	0.2615	1.5198	6.3349	3.5875	0.3447	3.5797
a_{19}	0.5305	0.7549	4.3818	1.2463	1.0660	3.6080
a_{20}	—	—	5.1625	0.5178	0.1635	3.9924
a_{21}	—	—	4.2827	1.9226	—	—
a_{22}	—	—	4.3747	5.3665	—	—
a_{23}	—	—	2.1397	0.0861	—	—
a_{24}	—	—	—	—	—	—

Tabla A.5: Intensidad de arribos de pasajeros por minuto.

Tabla A.6: Matriz origen-destino, $\mathcal{M}_{\mathcal{O}-\mathcal{D}}$, de las líneas 1, 2 y 3 del Sistema Metro de la Ciudad de México. (Ref: elaboración propia con información proporcionada por el STC Metro de la Ciudad de México.)

Bibliografía

- [1] ASMUSSEN, SØREN & GLYNN, PETER (2007). *“Stochastic Simulation: Algorithms and Analysis”*. Springer Science+Business Media, LLC publisher. New York, USA.
- [2] ASSIS, WÂNDERSON & MILANI, BASÍLIO (2004). *“Generation of Optimal Schedules for Metro Lines Using Model Predictive Control”*. Automatica (Journal of IFAC), Vol. 40, No. 8, pp. 1397-1404.
- [3] BOLETÍN UAM. UNIVERSIDAD AUTÓNOMA METROPOLITANA. (2012, febrero). No. 036. *“Diseñan investigadores de la UAM modelos para mejora del transporte”*. Recuperado de <http://www.uamero.uam.mx/UAMeros/insides/newsb.aspx?pid=1684&returnUrl=&../main.aspx&repositoryUrl=noticiasindice.aspx>
- [4] ENCICLOPEDIA LIBRE WIKIPEDIA. *“Metro de la Ciudad de México”*. Recuperado el 6 de julio del 2012, de http://es.wikipedia.org/wiki/Archivo:Metro_Mexico_DF_MP68_R93_01.jpg
- [5] FREUND, JOHN; MILLER, IRWIN Y MILLER, MARYLEES (2000). *“Estadística matemática con aplicaciones”* (6ª ed.). Editorial Pearson Prentice Hall. México.
- [6] FU, MICHAEL & HILL, STACY (1997). *“Optimization of Discrete Event Systems via Simultaneous Perturbation Stochastic Approximation”*. IIE Transactions, Vol. 29, No. 3, pp. 233-243.
- [7] GLASSERMAN, PAUL (2004). *“Monte Carlo Methods in Financial Engineering”*. Springer Science+Business Media, Inc. publisher. New York, USA.
- [8] HEIDERGOTT, BERND & VÁZQUEZ-ABAD, FELISA (2009). *“Gradient Estimation for a Class of Systems with Bulk Services: A Problem in Public Transportation”*. ACM Transactions on Modeling and Computer Simulation, Vol. 19, No. 3, Article 13.
- [9] HERNÁNDEZ, ONÉSIMO (1981). *“Procesos estocásticos: Introducción a la teoría de colas”*. Depto. de Matemáticas, CINVESTAV-IPN, II Coloquio. México.

- [10] KREYSZIG, ERWIN (1978). *“Introductory Functional Analysis with Applications”*. John Wiley & Sons, Inc. publisher. Toronto, Canada.
- [11] METRO DE LA CIUDAD DE MÉXICO. *“La red del Sistema de Transporte Colectivo”*. Recuperado el 4 de julio del 2011, de <http://www.metro.df.gob.mx/>
- [12] MOOD, ALEXANDER Y GRAYBILL, FRANKLIN (1970). *“Introducción a la teoría de la estadística”* (4^a ed.). Editorial Aguilar. Madrid, España.
- [13] NELSON, BARRY (2002). *“Stochastic Modeling: Analysis & Simulation”*. Dover Publications, Inc. publisher. New York, USA.
- [14] QUEVEDO, HÉCTOR Y PÉREZ, BLANCA (2008). *“Estadística para ingeniería y ciencias”*. Editorial Patria. México.
- [15] ROSS, SHELDON (1999). *“Simulación”* (2^a ed.). Editorial Pearson Prentice Hall. México.
- [16] ROSS, SHELDON (2007). *“Introduction to Probability Models”* (9th ed.). Academic Press, Elsevier, Inc. publisher. California, USA.
- [17] RUBINSTEIN, REUVEN & KROESE, DIRK (2008). *“Simulation and the Monte Carlo Method”* (2nd ed.). John Wiley & Sons, Inc. publisher. New York, USA.
- [18] SAAVEDRA, PATRICIA E IBARRA, VICTOR (2008). *“El método Monte-Carlo y su aplicación a finanzas”*. Depto. de Matemáticas, Universidad Autónoma Metropolitana Unidad Iztapalapa y Escuela de Actuaría, Universidad Anáhuac. México.
- [19] SURI, RAJAN & ZAZANIS, MICHAEL (1988). *“Perturbation Analysis Gives Strongly Consistent Sensitivity Estimates for the M/G/1 Queue”*. Management Science, Vol. 34, No. 1, pp. 39-64.
- [20] VÁZQUEZ-ABAD, FELISA & HEIDERGOTT, BERND (2004). *“Gradient Estimation for a Problem in Public Transportation: A comparison of SPA, SF and MVD”*. Proceedings of WODES'04, IFAC., Reims, France, pp. 241-246.
- [21] VÁZQUEZ-ABAD, FELISA & ZUBIETA, LOURDES (2000). *“Distributed Stochastic Approximation for Adaptive Frequency Allocation in Subway Networks”*. Proceedings of the 39th IEEE Conference on Decision and Control, pp. 4729-4734.
- [22] VÁZQUEZ-ABAD, FELISA & ZUBIETA, LOURDES (2005). *“Ghost Simulation Model for the Optimization of an Urban Subway System”*. Discrete Event Dynamic Systems: Theory and Applications, Vol. 15, No. 3, pp. 207-235.