Universidad EAFIT

Maestría en Matemáticas Aplicadas
Practica 1 - Optimización Estocástica
Semestre 2025-2

Profesor: Diego Fonseca

Plazo de entrega: Lunes 8 de Septiembre de 2025 a las 11:59pm
Modalidad: Individual
Importante: Cualquier intento de plagio será llevado a las instancias pertinentes de la universidad.

Instrucciones

- Este examen se debe realizar de manera individual.
- El estudiante debe marcar correctamente su archivo con su trabajo. La primera línea de texto debe contener el nombre y apellidos completos, y código de estudiante (o cédula en caso de no tener código) de todos los integrantes del grupo.
- Puede entregar un archivo de PDF o un Notebook, este debe nombrarse de la siguiente manera: `Practica1-OptSth-Apellido-Nombre`.
- Debe justificar sus procedimientos y soluciones, e interpretar sus resultados de acuerdo con el contexto de la actividad. No basta con llegar a las soluciones; es fundamental explicarlas y argumentarlas.
- Los archivos con sus soluciones se suben al buzón de la actividad de EAFIT Interactiva.

Ruteo Urbano con Tiempos de Viaje Estocásticos en NYC (Enero 2015)

Contexto y propósito

En esta actividad usted abordará un problema de *planificación urbana* con incertidumbre utilizando datos reales de la Ciudad de Nueva York. Una cuadrilla municipal (un único vehículo) debe visitar un conjunto de sitios en Manhattan durante un turno diurno. La **ruta** (orden de visita) se fija al inicio del día, antes de conocer las realizaciones concretas del tráfico. Una vez transcurre el día, los **tiempos de viaje** entre sitios se revelan y se decide cuánto trabajo ejecutar en sitio en cada parada, cuánto *tercerizar* (si no alcanza el tiempo) y cuánta *hora extra* usar.

El objetivo es **minimizar el costo esperado de operación**, equilibrando desplazamientos, trabajo en sitio, tercerización y horas extra, bajo un *horizonte de trabajo* diario.

Esta actividad persigue: (i) que usted **modele matemáticamente** un problema de dos etapas con decisiones binaras en primera etapa y continuas en segunda etapa, (ii) que **construya una muestra SAA** a partir de datos reales de taxis amarillos (enero 2015) para representar la incertidumbre de los tiempos de viaje, y (iii) que **diseñe e implemente** una metodología de solución apropiada para este tipo de problemas (usted debe justificar su elección a partir de la teoría vista en clase).

Conjunto de sitios y coordenadas (reales)

Se trabajará con un depósito y 11 sitios en Manhattan (latitud, longitud en grados decimales). Se sugiere mantener este conjunto para garantizar suficiente densidad de datos en los viajes taxi.

#	Sitio	Latitud	Longitud
0	**Depósito:** Javits Center	40.75750	−74.00250
1	Times Square	40.75800	−73.98550
2	Rockefeller Center	40.75870	−73.97870
3	Grand Central Terminal	40.75278	−73.97722
4	NY Public Library (Main)	40.75306	−73.98194
5	Union Square	40.73590	−73.99110
6	Washington Square Park	40.73083	−73.99750
7	Madison Square Garden	40.75056	−73.99361
8	Columbus Circle	40.76900	−73.98200
9	One World Trade Center	40.71274	−74.01338
10	New York Stock Exchange	40.70693	−74.01125
11	South Street Seaport (Pier 17)	40.70600	−74.00270

Convención. Denote $V = \{0, 1, \ldots, 11\}$ (con 0 el depósito) y $N = \{1, \ldots, 11\}$ los sitios a visitar.

Cómo se construyen los datos: *muestra SAA* de tiempos de viaje

Fuente de datos

Se utilizará el *Yellow Taxi Trip Record* de la NYC TLC para **enero de 2015**. Este conjunto incluye, por viaje: fechas/horas de pickup y dropoff y coordenadas geográficas de ambos eventos. Las columnas de interés (nombres exactos del esquema 2015) son:

- `tpep_pickup_datetime`, `tpep_dropoff_datetime`;
- `pickup_longitude`, `pickup_latitude`;
- `dropoff_longitude`, `dropoff_latitude`.

Usted descargará el archivo de enero de 2015 del portal oficial (*NYC TLC Trip Record Data*) y trabajará únicamente con las columnas mencionadas. El archivo se encuentra en el siguiente link:
`https://www.kaggle.com/datasets/elemento/nyc-yellow-taxi-trip-data?resource=download&select=yellow_tripdata_2015-01.csv`

Filtro temporal y geográfico

El objetivo es caracterizar tiempos de viaje *diurnos en días hábiles*. Aplique los siguientes filtros:

- **Días:** Lunes a viernes del mes de enero de 2015.
- **Franja horaria operativa:** de 09:00 a 17:00 (hora local).

Construcción de *pools* O-D por arco (i, j)

Para cada par *ordenado* (i, j) con $i \neq j$ donde i y j son sitios en la tabla anterior y que llamaremos nodos en adelante:

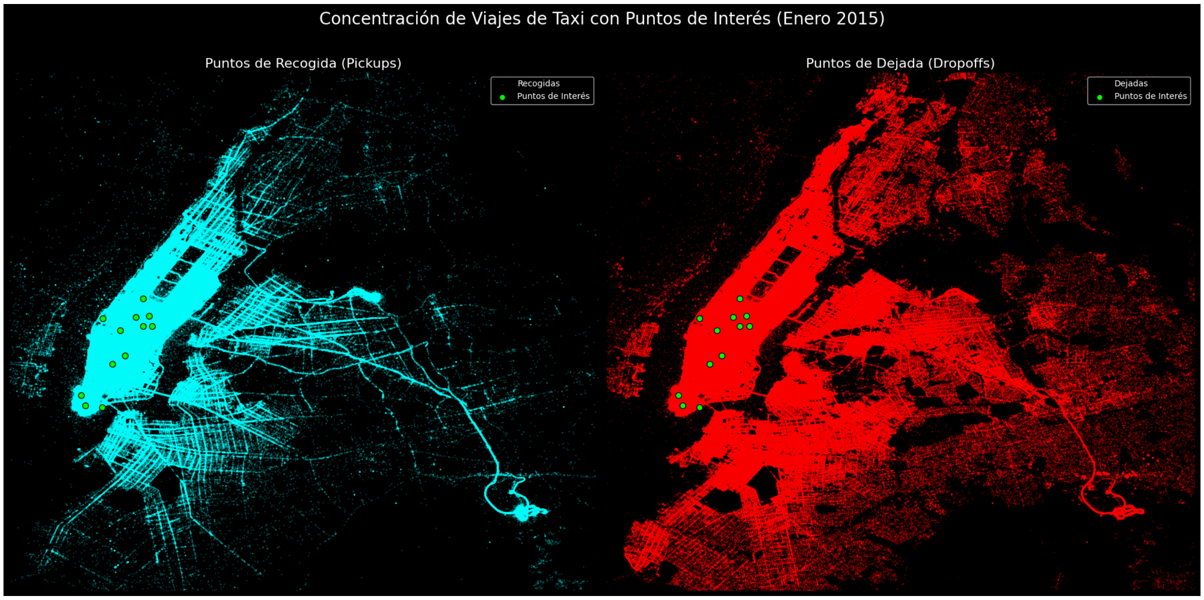

Figure 1: Diagrama de puntos de origen y destino de cada viaje en la base de datos. En verde los sitios que debe visitar la cuadrilla.

1. Fije un **radio inicial** $r = 250\,\text{m}$. Calcule la distancia Euclidiana entre el pickup del viaje y la coordenada del nodo i; y entre el dropoff y el nodo j.
2. Seleccione todos los viajes con *pickup* a una distancia euclidiana menor que r de i y *dropoff* a una distancia menor que r de j.
3. **Limpieza:** Calcule la duración t del viaje como $\texttt{dropoff} - \texttt{pickup}$ en *minutos*. Elimine outliers evidentes: $t < 1$ min o $t > 120$ min; elimine registros con coordenadas faltantes o nulas.
4. Si el **pool resultante** tiene menos de $m_{\min} = 50$ observaciones, aumente r en pasos de $100\,\text{m}$ hasta un máximo de $400\,\text{m}$; como alternativa secundaria, amplíe la ventana horaria a [08:00, 18:00] manteniendo días hábiles.

El resultado es un **pool de duraciones** $\mathcal{P}_{ij}$ para cada arco (i, j) que representa tiempos de viaje *reales* de taxi en condiciones comparables al turno de operación. Documente en un reporte cuántas observaciones quedaron por arco y qué radios finales se usaron.

Muestra SAA de tamaño K

Fije K (recomendación: $K = 50$; si su PC lo permite, $K = 100$).

- Cada **escenario** $s \in \{1, \ldots, K\}$ es una *matriz completa* de tiempos $T^{(s)} = (T_{ij}^{(s)})_{i \neq j}$.
- Para construir $T^{(s)}$: para cada (i, j), tome una **muestra con reemplazo** $T_{ij}^{(s)} \sim \mathcal{P}_{ij}$.
- Trabaje en **minutos** para todos los tiempos (viaje y servicio).

Nota. No se utiliza un único viaje por día; cada escenario sintetiza un *"día posible"* combinando una duración plausible por arco, muestreada de sus pools O-D. Así se obtiene una *muestra SAA* explícita de tamaño K sobre la cual se resolverá el problema.

Parámetros explícitos del experimento

Para dar concreción y mantener tamaños computables, utilice los siguientes valores (podrá explorar sensibilidad en el informe):

- **Horizonte de trabajo**: $H = 480\,\text{min}$ (8 h).
- **Demanda de trabajo en sitio** (minutos) por nodo $i \in N$:

i	1	2	3	4	5	6	7	8	9	10	11
d_i [min]	30	30	35	30	25	25	35	30	40	35	25

  Justificación: inspecciones ligeras de mobiliario/señalética/oservación de seguridad (25–40 min) según afluencia del sitio.

- **Productividad**: $\alpha_i = 1$ (un minuto de recurso por minuto de trabajo), para todo i.
- **Costos de segunda etapa**:
  - *Tercerización* $c_i^{\text{out}} = \$2.00/\text{min}$ (equivale a 120 \$/h); constante para todo i.
  - *Hora extra* $c^{\text{OT}} = \$1.00/\text{min}$ (equivale a 60 \$/h).

  Ambos valores son plausibles para trabajos contratados y sobrecarga horaria.

- **Costos de primera etapa**: Para internalizar un costo operativo realista sin mezclarlo con la aleatoriedad de los tiempos, definimos el costo de viaje por arco como

$$c_{ij} \;=\; \kappa\, d_{ij}^{\text{road}}, \qquad i \neq j,$$

  donde d_{ij}^{road} es la *distancia vial más corta (en km)* entre los sitios i y j sobre la red de calles (ruta en carro más corta), y κ es un *costo operativo por kilómetro* que resume combustible, mantenimiento y depreciación del vehículo. En la actividad fijaremos

$$\kappa \;=\; \$1.00 \;/\; \text{km}.$$

  ¿Cómo obtener d_{ij}^{road}?: Estas distancias se encuentran el archivo `matriz_distancias_conduccion.csv` que se encuentra en el buzón de la actividad.

  Justificación de escala. Con $\kappa = \$1.00/\text{km}$, un salto típico intra–Manhattan de $d_{ij}^{\text{road}} \in [1, 5]\,\text{km}$ aporta \$1–\$5 al objetivo de primera etapa, incentivando rutas compactas sin eclipsar los costos de recourse (tercerización a \$2/min y hora extra a \$1/min). Nótese que c_{ij} es totalmente *determinista* y *exógeno* a la muestra SAA de tiempos; su rol es reflejar el costo físico de mover el vehículo, mientras que la segunda etapa captura el impacto de la *incertidumbre en tiempos* sobre tercerización y sobretiempo.

Formulación que usted debe construir

Se espera que usted:

1. Modele una **primera etapa binaria** que decida la *ruta* sobre V (partiendo y regresando al depósito, visitando N), controlando subtours con el mecanismo que considere adecuado.
2. Modele una **segunda etapa continua** (por escenario) que, dados los tiempos $T^{(s)}$, asigne para cada sitio i el trabajo realizado en sitio $u_i^{(s)} \in [0, d_i]$ y el trabajo tercerizado $r_i^{(s)} \geq 0$, además de la hora extra $o^{(s)} \geq 0$, respetando restricciones que usted debe modelar, restricciones como:

- el trabajo realizado y el tercerizado debe suplir la demanda.
- se debe satisfacer que el trabajador trabaje a lo mas las horas establecidas por ley mas las horas extras si estas son necesarias.
- coherencia entre visita y trabajo: si no se visita i, no puede haber trabajo en ese sitio.

3. Implemente el algoritmo Integer L-Shaped para dar solución a esta actividad.

Productos esperados (entregables)

1. **Memoria técnica (PDF)**:
   - Motivación, hipótesis operativas y *claridad* en la construcción de la **muestra SAA**: radios usados por arco, tamaño de los pools $\mathcal{P}_{ij}$, filtros temporales, control de outliers; incluya histogramas o boxplots de duraciones por arco y un pequeño cuadro con el número de observaciones por pool.
   - Formulación matemática: conjuntos, parámetros, variables, funciones objetivo, restricciones de primera y segunda etapa; *no* basta con listar ecuaciones: describa su rol e intuición.
   - Descripción de la **metodología**: arquitectura del procedimiento, cómo y cuándo se generan cortes, criterio de parada.
   - Experimentos: tamaño de muestra K y variaciones de radios.
2. **Resultados y visualizaciones**:
   - **Mapa** con la *ruta óptima* sobre Manhattan (coordenadas dadas), marcando depósito e índices de visita.
   - **Evolución del algoritmo**: gráfico de cota inferior y mejor incumbente por iteración/nodo; número acumulado de cortes; tiempo de cómputo por nodo o por iteración.
   - **Distribución de tiempos** de algunos arcos clave (boxplots) y **contribución de costos esperados** (porcentaje de tercerización vs. hora extra).
3. **Interpretación**: explique la solución en el contexto urbano:
   - ¿Qué patrones tiene la ruta óptima? ¿Evita desplazamientos largos?
   - ¿En qué sitios se realiza la mayor parte del trabajo en sitio y dónde se terceriza más? ¿Por qué?
   - ¿Qué implican los resultados para la *planificación operativa real*? Proponga dos **extensiones** posibles (p. ej., ventanas de tiempo suaves, múltiples cuadrillas, selección de subconjunto de sitios, días con distinta franja horaria, demanda aleatoria).
4. **Repositorio reproducible**: Sea ordenado al momento de entregar sus resultados, el script debe estar completo y ejecutado (con parámetros de radios, filtros), incluyendo generador de la muestra SAA $T^{(1:K)}$, implementaciones del maestro y subproblemas, y un README con instrucciones para replicar o con comentarios claros.