Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Departamento de Ingeniería Industrial y de Sistemas
ICS2123 – Modelos Estocásticos
Profesores Maximiliano González, Nicolás Vargas y Andrés Navarro
Ayudante Jefe Francisco Tamarín
Segundo Semestre 2024

Tarea 2:
Simulación y CMTD

Instrucciones

- Esta tarea debe ser realizada **individualmente** o **en parejas** (pueden ser de diferentes secciones dentro de las que son coordinadas). Deben indicar en la primera plana del PDF los nombres de los/as integrantes, junto con las secciones respectivas de c/u.
- Tienen plazo hasta el **Martes 01 de octubre a las 23:59 hrs** para entregar sus soluciones.
- Deben entregar un archivo reporte, **en formato PDF**. En el archivo PDF deben entregar sus respuestas de los 3 problemas. Los códigos implementados deben adjuntarse como un archivo .zip aparte en el buzón. En el PDF deberá estar el desarrollo, resumen de resultados y correspondiente análisis. Se recomienda fuertemente utilizar LaTeX. Sin embargo, estas también pueden ser escritas en algún otro editor de texto o a mano y escaneadas, pero **debe estar ordenado y legible**, de lo contrario, se descontarán 3 décimas a la nota final de la tarea. Cualquier código que utilicen de internet debe ser debidamente citado insertando el *link* asociado.
- Si realizan la tarea completa en LaTeX, incluyendo todas las fórmulas y ecuaciones, tendrán **3 décimas extras**.
- La tarea tiene un total de 60 puntos. La nota final será calculada como la suma de las tres preguntas mediante la fórmula
$$Nota = \frac{P_1 + P_2 + P_3}{60} \cdot 6 + 1$$
Donde P_i corresponde al puntaje de la pregunta i. Se aproxima a dos decimales.
- Las preguntas se harán a través del foro de discusión asociado a la Tarea 1 en Canvas, por lo que les pedimos que **NO** manden dudas por mail sobre la tarea. De esta manera, todos tienen acceso a las respuestas.
- Seremos estrictos en que **no se aceptarán envíos por mail**. Evite problemas de internet, del sistema Canvas, de la calidad de la foto o cualquier otro que pudiese ocurrir al entregar sobre la hora. El/La alumno/a que envíe su tarea después de la hora límite, tendrá **nota 1.0**.

Problema 1 (20 puntos)

Parte 1 (10 puntos)

a) **(4 puntos)** Imagina que se está analizando el rendimiento de un nuevo hospital que ha comenzado a operar en la región. Este hospital ha estado en funcionamiento durante un período de ajuste para optimizar sus operaciones y recursos. Se cuenta con una muestra de la cantidad de personas en el hospital en un cierto horizonte de tiempo, registradas en el archivo *muestra1.xlsx*.

i) **(2 puntos)** Determine la media y varianza de la cantidad de gente promedio en el hospital considerando todo el horizonte de tiempo.

ii) **(2 puntos)** Considere ahora un periodo de "warm-up"(calentamiento) del sistema del hospital en el contexto de una simulación en estado de régimen. A partir de eso, determine la cantidad de gente promedio en el hospital, junto con la varianza. Analice con respecto al caso anterior. ¿Cómo puede afectar la inclusión del período de calentamiento a la interpretación de los resultados finales en un análisis de rendimiento del hospital?

b) **(6 puntos)** Suponga que un connotado analista financiero se encuentra trabajando en una empresa de gestión de activos. La empresa le ha proporcionado una serie de datos reales que representan el rendimiento diario del fondo de inversión durante un día, donde se registraron observaciones por minuto de un cierto activo riesgoso. El registro se encuentra en el archivo *muestra2.xlsx*.

iii) **(3 puntos)** Divida los datos en batches, definiendo adecuadamente el tamaño de c/u y el número de grupos a considerar. Determine el promedio de la media de los batches y la varianza de estos. Justifique su elección y cada uno de sus pasos.

iv) **(3 puntos)** A partir de las medias de los batches, elabore un intervalo de confianza del 95 % para el rendimiento promedio del activo riesgoso. Interprete el intervalo de confianza obtenido en el contexto de la estabilidad del rendimiento del fondo. ¿Qué implica este intervalo para los inversores en términos de riesgo y rentabilidad?

Parte 2 (10 puntos)

c) **(6 puntos)** Sean $X_1, X_2, \ldots$ variables aleatorias independientes tal que $\mathbb{P}(X_i = j) = \alpha_j$, con $j \geq 0$. Se dice que ocurre un evento notable en el instante n si $X_n > \max(X_1, \ldots, X_{n-1})$, donde $X_0 = -\infty$, y si ocurre un evento notable en el instante n, sea X_n el valor del evento notable. Sea además Z_i el i-ésimo valor de evento notable.

Entregue un argumento sobre si $\{Z_i : i \geq 1\}$ es una Cadena de Markov en Tiempo Discreto y, en caso de serlo, calcule sus probabilidades de transición.

d) **(4 puntos)** Sea $\{X_n; n \geq 0\}$ una CMTD con espacio de estados $S = \{1, 2, \ldots, N\}$, con matriz de probabilidades de transición P y distribución inicial $a = [a_1, \ldots, a_N]$. Sea $N_j(n)$ el número de veces que la CMTD visita el estado j durante el intervalo de tiempo $\{0, 1, \ldots, n\}$, y sea

$$m_{i,j}(n) = \mathbb{E}[N_j(n) \mid X_0 = i].$$

La cantidad $m_{i,j}(n)$ se denomina tiempo de ocupación hasta n del estado j comenzando desde el estado i. Definamos la matriz de tiempos de ocupación como

$$M(n) = \begin{bmatrix} m_{1,1}(n) & m_{1,2}(n) & \ldots & m_{1,N}(n) \\ m_{2,1}(n) & m_{2,2}(n) & \ldots & m_{2,N}(n) \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1}(n) & m_{N,2}(n) & \ldots & m_{N,N}(n) \end{bmatrix}$$

Demuestre que la matriz de tiempos de ocupación se puede representar como

$$M(n) = \sum_{r=0}^{n} P^r$$

Problema 2 (20 puntos)

Horace Nebbercracker, el anciano de la película *Monster house - La casa de los sustos*, ha decidido cambiar su estilo de vida sombrío y amargado. Para este próximo *Halloween* cuando los niños salgan a pedir dulces, ha decidido entregar dos tipos de golosinas: bolsas de Gomitas Ácidas, y paquetes de Huevitos de Pascua (¿acaso solo se pueden entregar en Pascua?). Bien, cada 60 minutos, Horace abre su puerta y suple la demanda de ambas dulces por parte de los niños (si es que no tiene suficientes dulces, Horace entrega los dulces que tiene en el momento). Para esto consideremos que la cantidad de bolsas de gomitas ácidas que le demandan a Horace en cada hora, se comporta como una variable aleatoria $Binomial(n, p)$, donde n es la cantidad máxima de Gomitas ácidas que le pueden pedir y p una probabilidad predefinida. Por otro lado, la cantidad de paquetes de Huevitos de Pascua que le piden en cada hora distribuye $\mathcal{U}(0, m)$, donde m es la cantidad máxima de paquetes de Huevitos de Pascua que le pueden demandar. Ambas demandas son completamente independientes entre si. Por otro lado, Horace tiene una política de inventario $(\{s_x, s_y\}, \{S_x, S_y\})$, donde si las bolsas de Gomitas Ácidas disminuyen de s_x, entonces procede a reponerlas hasta S_x. Lo mismo para los paquetes de Huevitos de Pascua, pero con s_y y S_y. Estas reposiciones son independientes entre si.

Por sus conocimientos en CMTD, usted identifica que esto podría tratarse de un problema de inventario bidimensional. A partir de lo anterior, se le solicita:

a) **(3 puntos)** Identifique el espacio de estados de la CMTD y escriba una fórmula que relacione la cantidad de bolsas de Gomitas Ácidas y paquetes de Huevitos de Pascua en la hora $t-1$ y t (es decir, definir (X_t, Y_t) en función de (X_{t-1}, Y_{t-1}), donde (X, Y) son las cantidades de Gomitas y Huevitos, respectivamente). Argumente por qué esta situación puede modelarse como una CMTD.

b) **(5 puntos)** Explicite las probabilidades de transición $\mathbb{P}_{(i,j)(k,l)} = \mathbb{P}((X_t, Y_t) = (k, l)|(X_{t-1}, Y_{t-1}) = (i, j))$.

Hint: Podría servirle crear funciones auxiliares $p_z(i)$ y $P_z(i)$, para representar $\mathbb{P}(D_z = i)$ y $\mathbb{P}(D_z \geq i)$, respectivamente, con z la V.A asociada. Le recomendamos analizar por casos.

c) **(2 puntos)** Identifique las clases de la cadena, y clasifíquelas según corresponda.

Para las siguientes preguntas, considere la siguiente asignación de valores: $(\{s_x, s_y\}, \{S_x, S_y\}) = (\{1, 2\}, \{3, 5\})$ y $(n, m, p) = (3, 4, 0.5)$. A partir de esta asignación, responda:

d) **(4 puntos)** Obtenga explícitamente la matriz de probabilidades de transición de la CMTD asociada.

e) **(3 puntos)** Si en una hora en concreto, Horace tiene 3 bolsas de Gomitas Ácidas y 4 paquetes de Huevitos de Pascua ¿Cuál es la probabilidad de que después de 6 horas vuelva a tener dicha cantidad de dulces (sin antes haberla tenido)?

f) **(3 puntos)** Si en un instante concreto Horace tiene 2 paquetes de Huevitos de Pascua y 2 bolsas de Gomitas Ácidas, ¿Cuál es el valor esperado de horas que transcurrirán para que Horace tenga 3 bolsas de Gómitas Ácidas y 4 paquetes de Huevitos de Pascua?

Problema 3 (20 puntos)

Considere un local de venta de sopaipillas, administrado por *Arturo Vidal* en la playa del Quisco, donde atiende solo una persona de 08:00 a 00:00. Debido a que es un local en la playa, el afluente de personas no es constante durante el día. Los tiempos entre llegadas de personas son *iid* y están definidos a partir de la hora en cesutión, en concreto:

- ◇ Entre las 08:00 y las 12:00, los tiempos entre llegadas se comportan según una $Uniforme(2, 4)$ en minutos.
- ◇ Entre las 12:00 y 17:00 los tiempos entre llegadas se comportan según una *Exponencial* con media 3 minutos.
- ◇ Y desde las 17:00 hasta las 00:00, los tiempos entre llegadas son según una distribución *Rayleigh*(3) en minutos.

Los tiempos de atención son mediante una variable aleatoria exponencial con media 4 minutos. Además, cada persona puede requerir una cantidad $max(1, Binomial(n, p))$ de sopaipillas, donde n y p dependen de la hora:

- ◇ Entre las 08:00 y 11:00, es $n = 4$ y $p = 0.1$.
- ◇ Entre las 11:00 y 15:00, es $n = 7$ y $p = 0.2$.
- ◇ Entre las 15:00 y 19:00 es $n = 5$ y $p = 0.3$.
- ◇ Después de las 19:00, es $n = 8$ y $p = 0.5$.

El local adapta una política de inventario (s, S) con revisión periódica y no continua. En este caso, quien atiende revisa cada τ minutos el stock de inventario. Si este es menor a s, entonces repone hasta S. Esto implica que las revisiones no se hacen cada vez que se hace una venta. Existe un costo fijo por emitir una orden de reposición de α y un costo por unidad repuesta \$50 (por lo tanto, al realizar una orden de l unidades, se paga $\alpha + 50l$). También, se paga $\$\beta$, por cada unidad en inventario. Estos costos dependen de la frecuencia de revisión τ y se obtienen como:

$$\alpha(\tau) = \$ \left\lceil \frac{600}{\tau} \right\rceil + 35, \quad \beta(\tau) = \$ \left\lceil \frac{\tau}{20} \right\rceil + 20$$

También, en esta revisión cada τ minutos, el arrendatario debe hacer el pago de inventario por cada unidad que tenga **antes** de realizar la reposición respectiva.

El arrendatario paga por jornada fija, un valor por uso del local y operario $\$\zeta$, asociado a la capacidad del local, máximo de inventario y diferencia relativa entre mínimo y máximo de niveles de stock (hay diferentes tipos, por eso diferentes costos de arriendo por jornada):

$$\zeta(K, S) = \$700 \cdot K + 200 \cdot S$$

Donde K es la capacidad del sistema (recuerde que como se tiene un solo servidor, la capacidad de personas en cola es, entonces $K - 1$) y S el valor de la política (s, S). Por otro lado, existen costos de oportunidad por demanda perdida. Por ejemplo, cuando un cliente requiere de c sopaipillas y hay $h < c$, entonces se realiza la venta con los h existentes, pero existe un costo de oportunidad de demanda pérdida unitario de \$50, incurrido por cada uno de los $c - h$ que no se pudieron vender. También puede ocurrir que llegue gente al local cuando este lleno. Este costo de oportunidad por local lleno, tiene un valor unitario de \$300 por cada unidad que el cliente iba a querer comprar (suponga que usted conoce dicha información).

En caso de que queden personas en el local después de la hora de termino, el dueño presenta la siguiente política: se cierran las cortinas, por lo que no se admiten más clientes que lleguen, pero se atienden a todos los clientes que ya estaban dentro del local después de la hora de termino. Sin embargo, por cada minuto extra desde la hora de cierre, el arrendatorio deberá pagar \$500 pesos.

El precio de venta de un sopaipilla es de \$300. Considere que el local no cuenta con más ingresos que la venta misma de sopaipillas. Como referencia, tenga en cuenta la siguiente definición de Ganancias:

Ganancias = IngresosTotales - CostoInventario - CostoReposicion - CostoFaltante - CostoLocalLleno - CostoArriendo - CostoTiempoExtra

A partir de lo anterior, se le solicita:

a) (8 puntos) Implemente un modelo computacional que permita simular el experimento. Setee solo para esta parte la semilla 2123. Para la política $(s, S, \tau, K) = (10, 50, 30, 30)$, ejecute una simulación de una jornada y registre los resultados principales, junto con los siguientes gráficos:

(a) Cantidad de Clientes en el sistema a través del tiempo.

(b) Nivel de inventario a través del tiempo.

(c) Demanda perdida acumulada a través del tiempo.

Además, incluya un análisis del desglose de los costos obtenidos. Comente sus resultados y analice detallada y exhaustivamente en función del contexto del problema.

b) (5 puntos) *Arturo Vidal* le dice que su política actual consiste en $(s, S, \tau, K) = (30, 45, 30, 15)$. El dueño la apuesta uno de sus caballos con el desafío de encontrar una política diferente que genere mayores ganancias en promedio. Si S_1 es la política actual y S_2 la política nueva, se debe cumplir que $||S_1 - S_2||_1 \geq 10$ (es decir, debe haber un cambio de al menos 10 unidades en total entre los parámetros a perturbar de la política). Para esto realice el siguiente procedimiento (nótese que son pasos para guiar, no ítems de cosas que se le pide incluir como incisos):

Paso 1: Una réplica de la simulación va a estar compuesta por al menos, 100 simulaciones o corridas diarias del local (independientes entre si). Obtenga el promedio de ganancias diarias. Eso es el valor de la réplica.

Paso 2: Debe generar muchas réplicas, tanto para la política base, como también para la alternativa propuesta por usted. Registre los resultados de cada réplica o iteración.

Paso 3: Deberá generar tantas réplicas como sea posible para asegurar con un 95 % de confianza, que su propuesta es mejor. Para eso debe armar un intervalo de confianza entre con las diferencias de las medias de cada replica. Utilice la siguiente fórmula para el intervalo de confianza:

$$\bar{X}(n) \pm t_{n-1,1-\alpha/2}\sqrt{S^2(n)/n}$$

Indicación: Tenga cuidado con las semillas a setear. Para comparar las políticas, en cada réplica deben compartir la misma semilla, para que el análisis sea justo y valido para comparar.

Explicite la política propuesta y muestre los siguientes intervalos de confianza: i) Para el de la política base ii) Para su política propuesta iii) Para la diferencia entre las medias de cada réplica entre las políticas. Realice un análisis comparativo por tipo de costo. ¿Son sus réplicas las suficientes para asegurar que su política es mejor a ese nivel de confianza? Analice, y de forma cualitativa entregue recomendaciones al dueño del local, considerando los parámetros fijos del problema y la estructura de costos asociados.

c) (4 puntos) Tome su política propuesta de la parte **b)** y genere 1000 réplicas (de 100 simulaciones c/u). Muestre, con un gráfico, como evoluciona la anchura del intervalo de confianza de las ganancias de dicha política a medida que aumentan las réplicas.

¿Qué puede decir respecto a la evolución de los intervalos? ¿Cómo se relaciona esto con los contenidos vistos en el capítulo de Simulación? Comente exhaustivamente.

d) (3 puntos) Imagine que el dueño le indica que los resultados obtenidos y procedimientos realizados pueden estar sesgados. Indique de forma cualitativa que cambios debería realizar a su implementación en el contexto de análisis de resultados de un modelo de simulación para disminuir el sesgo. Comente detalladamente sobre las implicancias que puede haber al considerar resultados sesgados o poco representativos de un modelo de simulación.