Simulación ITAM
2 de noviembre de 2021 2021-II

Tarea 5.

La fecha de entrega es el **17 de noviembre de 2021**.

Lecturas

- Casella y Robert: Capítulos 3 y 4.
- Brandimarte: Capítulos 2 y 8.
- Dagpunar: Capítulos 5 y 6.
- Voss: Capítulo 3.

Problemas

1. Cinco elementos, numerados del 1 al 5 se acomodan inicialmente en un orden aleatorio (esto es, el orden inicial es una permutación aleatoria de los números {1,2,3,4,5}) En cada estado del proceso, uno de los elementos es seleccionado y puesto en el frente de la lista. Por ejemplo, si el orden presente es {2,3,4,1,5} y el elemento 1 se elige, entonces el nuevo orden es {1,2,3,4,5}. Supongan que cada selección es obtenida independientemente y se toma el elemento i con probabilidad p_i, donde las $p_i's$ son $(\frac{1}{15}, \frac{2}{15}, \frac{3}{15}, \frac{4}{15}, \frac{5}{15})$. Sea L_j la variable que denota la posición del j-ésimo elemento seleccionado, y sea $L = \sum_{j=1}^{100} L_j$. Queremos usar simulación para estimar $E(L)$.

   (a) Expliquen cómo utilizarían la simulación para estimar el valor de $E(L)$ y llévenlo a cabo.

   (b) Calculen $E(N_i)$ donde N_i es el número de veces que el elemento i es elegido en 100 selecciones.

   (c) Sea $Y = \sum_{i=1}^{5} iN_i$ ¿Cómo se correlaciona la variable Y con la variable L?

   (d) Estimar L usando a Y como variable de control.

2. Sean X y Y dos independientes exponenciales con medias 1 y 2 respectivamente y supongan que queremos estimar $P(X+Y > 4)$. ¿Cómo utilizarían condicionamiento para reducir la varianza del estimador? Digan si considerarían condicionar en X o en Y y porqué.

3. Supongan que queremos estimar $\theta = \int_0^1 e^{x^2}\,dx$. Muestren que generar un número aleatorio u y usar el estimador $e^{u^2}(1+e^{1-2u})/2$ es mejor que generar dos números aleatorios u_1 y u_2 y usar $(e^{u_1^2}+e^{u_2^2})/2$.

4. Explicar cómo se pueden usar variables antitéticas en la simulación de la integral

$$\theta = \int_0^1 \int_0^1 e^{(x+y)^2}\,dx\,dy$$

   ¿Es claro en este caso que usando variables antitéticas es más eficiente que generando nuevos pares de variables aleatorias? Dar una justificación a su respuesta.

5. En ciertas situaciones una variable aleatoria X con media conocida, se simula para obtener una estimación de $P(X \le a)$ para alguna constante dada a. El estimador simple de una simulación para una corrida es $I = I(X \le a)$.

   - Verificar que I y X están correlacionadas negativamente.
   - Por el inciso anterior, un intento natural de reducir la varianza es usar X como variable de control (esto es, usar $Y_c = I + c[X - E(X)]$). En este caso, determinar el porcentaje de reducción de varianza de Y_c sobre I es es posible (usando la mejor c si X es $\mathcal{U}(0,1)$.
   - Repetir el inciso anterior si $X \sim \exp(1)$.

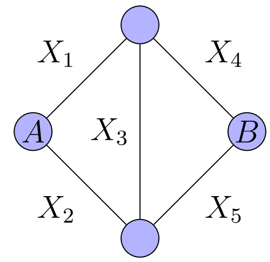
6. El número de reclamos en una aseguradora que se harán la próxima semana depende de un factor ambiental U. Si el valor de este factor es $U = u$, entonces el número de reclamos tendrá distribución Poisson con media $\frac{15}{0.5+u}$. Suponiendo que $U \sim \mathcal{U}(0,1)$, sea p la probabilidad de que habrá al menos 20 reclamos la siguiente semana.

   - Explicar como obtener una simulación cruda de p.
   - Desarrollar un estimador de simulación eficiente usando esperanza condicional junto con una variable de control
   - Desarrollar un estimador de simulación eficiente usando esperanza condicional y variables antitéticas.
   - Escriban un programa para determinar las varianzas de los incisos anteriores.

7. Consideren la siguiente gráfica, representando una red puente:

   Supongan que queremos estimar la longitud esperada l de la ruta más corta entre los nodos A y B, donde las longitudes de los arcos son variables aleatorias $X_1, \ldots, X_5$. Entonces tenemos que $l = \mathrm{E}[H(\mathbf{X})]$, donde

$$H(\mathbf{X}) = \min\{X_1 + X_4, X_1 + X_3 + X_5, X_2 + X_3 + X_4, X_2 + X_5\}$$



Noten que $H(\mathbf{x})$ es no decreciente en cada componente del vector $\mathbf{x}$. Supongan que las longitudes son independientes y $X_i \sim \mathcal{U}\,(0, a_i)$, con $\mathbf{a} = (1, 2, 3, 1, 2)$. Escribiendo $X_i = a_i U_i$ se puede restablecer el problema como la estimación de $l = \mathrm{E}[h(\mathbf{U})]$ con $h(\mathbf{U}) = H(a_1 U_1, \ldots, a_5 U_5)$.

- Obtener un estimador crudo de MonteCarlo para l.
- Obtener un estimador usando variabes antitéticas
- Obtener un estimador usando variables de control.
- Obtener un estimador usando condicionamiento.

En todos los casos anteriores, calcular la reducción de varianza obtenida y determinar el mejor método.

8. Sea S la suma de los resultados de lanzar 100 veces un dado honesto. Usen la desigualdad de Chebyshev para acotar el valor de $P(S \geq 380)$.

9. Supongan que $Y|\theta \sim \mathcal{G}\,(1, \theta)$ y que $\theta \sim IG(\alpha, \beta)$.

   - Encuentren la distribución posterior de θ.
   - Encuentren la media y varianza posterior de θ.
   - Encuentren la moda posterior de θ.
   - Escriban dos ecuaciones integrales que se pueden resolver para encontrar el intervalo de 95 % de colas simétricas para θ

10. Los siguientes datos corresponden a las horas adicionales de sueño de 10 pacientes tratados con un somnífero B comparado con un somnífero A:

$$1.2, 2.4, 1.3, 1.3, 0, 1, 1.8, 0.8, 4.6, 1.4$$

    Lleven a cabo un análisis bayesiano de estos datos y extraigan conclusiones, asumiendo cada componente de la verosimilitud que sea:

   - normal
   - $t_{(3)}$

- $t_{(1)}$
- Bernoulli (de alguna manera que se les ocurra)

En este ejercicio, escriban un código para manejar cualquier integración necesaria y cálculo de probabilidades marginales posteriores.