Simulación ITAM
16 de abril de 2018 Jorge de la Vega
2018-I

Tarea 4.

La fecha de entrega es el **30 de abril de 2018**.

Lectura

- Salmon, F. (2009) Recipe for Disaster: The Formula That Killed Wall Street, Wired
- Arturo Ederly: Cópulas y variables aleatorias, una introducción
- Robert & Casella Caps. 3 y 4.
- Dagpunar Cap.5

Problemas

1. Obtener una muestra aleatoria de 5,000 observaciones del vector $X = (X_1, X_2, X_3, X_4)$ donde $X_1 \sim \mathcal{N}(4, 9)$, $X_2 \sim$ **Bernoulli** (0.6), $X_3 \sim Beta(2, 3)$ y $X_4 \sim Gamma(3, 2)$. Considerar la siguiente estructura de correlación entre las variables: $cor(X_1, X_3) = -0.7$, $cor(X_2, X_4) = 0.4$, $cor(X_1, X_4) = 0.5$ y $cor(X_2, X_3) = 0.2$. En otro caso consideramos que las variables son independientes. Hacer los histogramas de las funciones marginales y corroborar que tienen la distribución considerada.

2. La τ de Kendall entre X y Y es 0.55. Tanto X como Y son positivas. ¿Cuál es la τ entre X y $1/Y$? ¿Cuál es la τ de $1/X$ y $1/Y$?

3. Mostrar que cuando $\theta \to \infty$, $C^{Fr}(u_1, u_2) \to min\{u_1, u_2\}$, donde C^{Fr} es la cópula de Frank.

4. - Construyan un vector de 100 números crecientes y espaciados regularmente entre 0.1 y 20. Llámenlo `SIG2` . Ahora construyan otro vector de longitud 21 empezando en -1 y terminando en 1. Llámenlo `RHO`.
   - Para cada entrada σ^2 de `SIG2` y cada entrada de `RHO`:
     - Generar una muestra de tamaño $N = 500$ de una distribución bivariada normal $Z = (X, Y)$ donde $X \sim \mathcal{N}(0, 1)$ y $Y \sim \mathcal{N}(0, \sigma^2)$ y el coeficiente de correlación de X y Y es ρ. Z es una matriz de dimensiones 500×2.
     - Crear una matriz de 500×2, llámenlo `EXPZ`, con las exponenciales de las entradas de Z. ¿Qué distribución tienen estas variables transformadas?

- Calculen el coeficiente de correlación, $\tilde{\rho}$ de las columnas de `EXPZ`. Grafiquen los puntos $(\sigma^2, \tilde{\rho})$ y comenten sobre lo que obtuvieron.

5. Consideren la cópula de Clayton. Mostrar que converge a la cópula de comonotonicidad cuando $\theta \to \infty$. [Hint: usen la regla de l'Hópital considerando que la cópula de Clayton se puede escribir como $exp\{log(u_1^{-\theta} + u_2^{-\theta} - 1)/\theta\}$ para θ positivo.]

6. Supongan que tienen dos vectores de datos $(x_1, \ldots, x_n)$ y $(y_1, \ldots, y_n)$. Entonces la cópula empírica es la función $C : [0,1] \times [0,1] \to [0,1]$ definida por

$$C(u,v) = \frac{1}{n} \sum_{j=1}^{n} I\left(\frac{r_j}{n+1} \leq u, \frac{s_j}{n+1} \leq v \right)$$

   donde $(r_1, \ldots, r_n)$ y $(s_1, \ldots, s_n)$ denotan los vectores de rangos de x y y respectivamente.

   Escriban una función *vectorizada* (esto es, que se pueda vauar en vectores) llamada `empCopula` que tome cuatro argumentos `u, v, xVec` y `yVec`. Pueden suponer que los valores `u,v` están en $[0,1]$ y que `xVec` y `yVec` son vectores numéricos que tienen la misma longitud (no vacios).

7. la cópula Farlie-Gumbel-Morgenstern es $C(u,v) = uv[1+\alpha(1-u)(1-v)]$ para $|\alpha| \leq 1$. Mostrar que la densidad conjunta correspondiente $\partial^2 C(u,v)/\partial u \partial v$ es no negativa. Mostrar que C tiene marginales uniformes en $(0,1)$. Encontrar el coeficiente de correlación de Spearman y la tau de Kendall.

8. Este es un ejercicio de calibración de las cópulas utilizando correlaciones de rangos. Supongan que una muestra produce un estimado de la τ de Kendall de 0.2. ¿Qué parámetro debe usarse para

   i. la cópula normal,

   ii. la cópula de Gumbel,

   iii. la cópula de Calyton?

9. Usen la función `normalCopula` del paquete `copula` para crear una cópula gaussiana bidimensional con un parámetro de 0.9. Luego creen otra cópula gaussiana con parámetro de 0.2 y describan la estructura de ambas cópulas (diferencias y semejanzas).

10. Usen la función `rCopula` del paquete `copula` para generar muestras de 500 puntos cuya distribución son las cópulas del ejercicio 8 anterior. Hagan una gráfica de las dos muestras. Teniendo en mente que una cópula determina la estructura de dependencia de una distribución multivariada conjunta, mirando estas gráficas, ¿pueden decir cuál de estas dos cópulas debe usarse para simular una distribución con una fuerte dependencia entre las marginales?

11. 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, independientemente, 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 simulación para estimar $E(L)$.

    (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 Y con L?

    (d) Desarrollen un estudio para estimar L usando Y como variable de control.

12. 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.

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

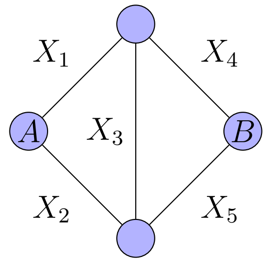
    - Verificar que I y X están correlacionadas negativamente.
    - Por el inciso anterior, un intento natural de reducir la varianza es usar X como variable de control (esto es usar $Y_c = I + c(X - E(X))$. En este caso, determinar el porcentaje de reducción de varianza de Y_c sobre I es es posible (usando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

14. 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.

15. 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 $l = E(H(\mathbf{X}))$, 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 $l = E[h(\mathbf{U})]$ con $h(\mathbf{U}) = H(a_1 U_1, \ldots, a_5 U_5)$.

- Obtener un estimador crudo de Monte Carlo 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.