Simulación Semestre 2018-I ITAM
30 de marzo de 2018 Jorge de la Vega

Primer Examen Parcial. Simulación

La duración del examen es de 2:30 hrs. Cada problema vale 10 puntos.
No se responden preguntas durante el examen. SI creen que algo está mal, corrijan, si creen que no esta claro lo que se pide interpreten y respondan lo más cercano a su conocimiento.

---

1. Consideren el siguiente modelo:

$$\begin{aligned} Y &\sim \textbf{Bin}\,(n,p) \\ X|Y &= \sum_{i=1}^{Y} X_i \end{aligned}$$

donde $X_i \sim \mathcal{G}\,(\alpha,\beta)$. Entonces X es una suma aleatoria de variables Gamma. Lo que interesa calcular es $P(X > M)$. En una tarea se dejó un problema muy similar a éste.

- ¿Porqué funciona el siguiente algoritmo: Repetir para $j = 1$ hasta $j = N$ veces los pasos (i) a (iii):
  i. Genera $Y \sim \textbf{Bin}\,(n,p)$. Obtener y
  ii. Genera y variables Gamma $\mathcal{G}\,(\alpha,\beta)$ y calcula $X|(Y = y) = \sum_{i=1}^{y} X_i$. (Si $y = 0$, se considera $X = 0$).
  iii. Calcula $I_j = I(X|y > M)$.

  Estimar $P(X > M)$ con $\sum_{j=1}^{N} I_j/N$.
- ¿Difiere si se calcula como usualmente se calcula condicionando $P(X > M) = \sum_{j=0}^{n} P(X > M|Y = j)P(Y = j)$?

Solución.

El algoritmo funciona porque básicamente se está condicionando como se hace en el segundo punto. El segundo punto es la forma en la que se demuestra el algortimo, y el primer punto es su implementación. La única diferencia, es que puede ser que al muestrar los valores de Y no se obtengan todos los puntos de la partición, pero mientras más grande es el número de simulaciones, es más pobrable que se tenga todo el dominio de la función.

$\square$

---

2. Si $X_i = (9X_{i-1} + 3) \mod 2^4$ y $Y_i = (5Y_{i-1} + 3) \mod 2^5$

- Determinar los periodos de X_i y Y_i.
- ¿Cómo se puede construir un generador con periodo mayor que el periodo de cada generador X_i y Y_i?
- Genera 2 números con el nuevo generador.

Solución.

- En los dos casos hay que verificar utilizando el teorema de Hull y Dobell.
  - En el primer glc, $m = 16$, $c = 3$ y $a = 9$. Entonces (i) 16 y 3 son primos relativos, (ii) El único divisor primo de m es $q = 2$ y 2 divide a $a - 1 = 8$ y (iii) 4 a m y divide a 8. Entonces el periodo es 16 .
  - En el segundo glc, $m = 32$, $c = 3$ y $a = 5$. Entonces (i) 32 y 3 son primos relativos, (ii) El único divisor de m es $q = 2$ y 2 divide a $a - 1 = 4$ y (iii) 4 divide a m y a $a - 1$. Tiene periodo completo igual a 32.
- Para construir un generador de periodo mayor se puede utilizar el teorema para el generador con ciclo ampliado, tomando $u_1 + u_2 - [u_1 + u_2]$ que tendrá ciclo de periodo el mínimo común múltiplo de 16 y 32. Para obtener uno mayor, podemos tomar al generador Y_i pero reemplazando m por $2^6 = 64$ para obtener $Z_i = (5Z_{i-1} + 3) \mod 64$. Entonces Z_i o el generador combinado que toma X_i y Z_i tendrá periodo 64.
- Tomado el nuevo generador a Z_i, tomando $Z_0 = 3$, por ejemplo: $Z_1 = (5*3+3) \mod 64 = 18$ y $Z_2 = (5*18+3) \mod 64 = 29$ Entonces $u_1 = 18/64 = 0.28125$ y $u_2 = 29/64 = 0.453125$.

□

---

3. ¿Qué problema(s) tiene el generador RANDU?

Solución.

El generador `RANDU` muestra algunos de los siguientes problemas:

- Falla algunas de las pruebas más comunes de aleatoriedad.
- Los puntos del generador caen sobre exactamente 15 hiperplanos cuando se grafican en el espacio (u_i, u_{i-1}, u_{i-2}) mostrando una fuerte correlación lineal en tres dimensiones, y por lo tanto fallando independencia.

□

---

4. Con 10,000 números que se suponen aleatorios, se obtuvo la siguiente tabla de frecuencias:

Intervalo	[0,0.1)	[0.1,0.2)	[0.2,0.3)	[0.3,0.4)	[0.4,0.5)	[0.5,0.6)	[0.6,0.7)	[0.7,0.8)	[0.8,0.9)	[0.9,1]
Frecuencia	1023	1104	994	993	1072	930	1104	969	961	850

Usando el hecho de que $\sqrt{2\chi^2_{(m)}} \to \mathcal{N}\left(\sqrt{2m-1}, 1\right)$ obtener una conclusión para la prueba de hipótesis H_0 de que la muestra proviene es una muestra aleatoria $\mathcal{U}(0,1)$ vs. la alternativa H_a : de que la muestra no es uniforme.

Solución.

La estadística es

$$\begin{aligned} x^2 &= \frac{(1023-1000)^2}{1000} + \frac{(1104-1000)^2}{1000} + \frac{(994-1000)^2}{1000} + \cdots + \frac{(961-1000)^2}{1000} + \frac{(850-1000)^2}{1000} \\ &= 57.312 \end{aligned}$$

x^2 tiene una distribución ji-cuadrada con $10-1=9=m$ grados de libertad. De acuerdo a la aproximación normal $\sqrt{2\times 57.312}=\sqrt{114.624}=10.70626\sim\mathcal{N}(4.12,1)$ o $z=10.70626-4.12=6.58626\sim\mathcal{N}(0,1)$ El p-value de este valor tan grande es 0. Por lo tanto, se rechaza la hipótesis nula de uniformidad.

□

---

5. Usando el Teorema de la transformación inversa, obtener una muestra de tamaño 2 de una variable aleatoria Weibull(2,2) dada por:

$$X\sim Weibull(\alpha,\lambda) \text{ si } f(x)=\alpha\lambda^{\alpha}x^{\alpha-1}e^{-(\lambda x)^{\alpha}}I_{(0,\infty)}^{(x)}$$

Solución.

La función de distribución es directa, $F(x)=1-e^{-(\lambda x)^{\alpha}}$, por lo que resolviendo la ecuación $u=1-e^{-(\lambda x)^{\alpha}}$ para x: $x=\frac{1}{\lambda}(-\log(1-u))^{1/\alpha}$. Entonces, tomando cualesquiera dos números uniformes (por ejemplo tomados de la calculadora con la tecla `RND`: $u_1=0.6173115$ y $u_2=0.4169844$, así que : $x_1=0.4900342$ y $x_2=0.3672674$.

□

---

6. Consideren la siguiente sucesión de números:

$$0.134,\ 0.279,\ 0.886,\ 0.197,\ 0.011,\ 0.923,\ 0.990,\ 0.876$$

La media de los números es 0.537.

- ¿Cuántas rachas crecientes hay y de qué longitud?
- ¿Cuántas rachas decrecientes hay y de qué longitud?
- ¿Cuántas rachas de números arriba de la media hay y de qué longitud?
- ¿Cuál es la probabilidad de que haya 2 o más rachas de números arriba de la media?

Solución.

- (0.134 0.279 0.886) (0.197) (0.011 0.923 0.990) (0.876). Hay 4 rachas crecientes de longitudes 3,1,3,1.
- (0.134) (0.279) (0.886 0.197 0.011) (0.923) (0.990 0.876). Hay 5 rachas, 3 de longitud 1, una de longitud 3 y una de longitud 2.
- (0.134 0.279) (0.886) (0.197 0.011) (0.923 0.990 0.876). Hay 2 rachas arriba de la media, una de longitud 1 y una de longitud 3. Las otras dos particiones corresponden a rachas por debajo de la media, 2 de longitud 2.
- De acuerdo a lo que vimos en clase, si R_1 son las rachas por arriba de la media (y R_2 las que están por debajo), entonces

$$f_{R_1}(x)=\frac{\binom{3}{x-1}\binom{5}{x}}{\binom{8}{4}} \text{ para } x=1,2,3,4$$

Entonces $P(R_1\geq 2)=1-P(R_1=1)=1-\frac{\binom{3}{0}\binom{5}{1}}{\binom{8}{4}}=1-5/70=0.9285714$.

□

---

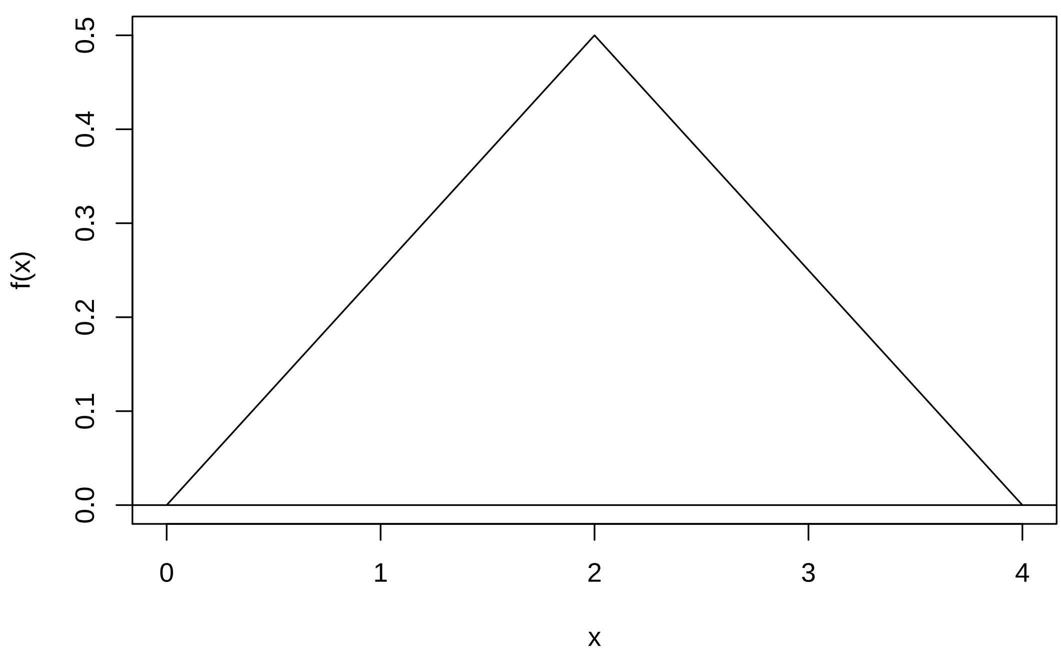
7. Supongan que X tiene densidad en $[0,4]$ dada por:

$$f(x) = kxI_{[0,2]}^{(x)} + k(4-x)I_{(2,4]}^{(x)}$$

- ¿Cuál es el valor de k?
- Describan en detalle un algoritmo de Aceptación-Rechazo para generar una muestra de X, dado un generador de números uniformes.

Solución.

- Se puede resolver la integral o ver en la gráfica cuáles son las áreas de los dos triángulos que se forman. En términos de los triángulos, area= $\frac{(2k)*2}{2} + \frac{(2k)*2}{2} = 1$ o $k = 1/4$.



- Se puede mayorizar la densidad con una uniforme en el intervalo $(0,4)$. Entonces el algoritmo sería:
  a) Generar $u_1, u_2 \sim \mathcal{U}(0,1)$
  b) Calcular $(x,y) = (4u_1, u_2/4)$.
  c) Si $u_2/4 \leq f(4u_1)$, aceptar $x = 4u_1$. Si no, repetir el paso (a)

□

---

8. Consideren números uniformes de 5 dígitos. ¿Cuál es la probabilidad de obtener un número con 3 dígitos iguales a un número y 2 dígitos iguales a otro número?

Solución.

Se pide la probabilidad de un 'full' en una mano similar a la de poquer. Se puede calcular muy fácil razonando de la siguiente manera: El número total de 'manos'de 5 dígitos del total de 10 dígitos es de $10^5 = 100,000$. Los casos posibles suponen que el primer dígito se puede escoger de 10 formas (del 0 al 9) y una vez fijo, dos lugares de la mano quedan fijos con ese mismo número. El siguiente número se puede elegir de los 9 restantes, y luego queda otro lugar que debe ser igual a este número. Por último, se pueden permutar estos dos números diferentes de $\binom{5}{2} = 10$ formas. Entonces

$$P(\text{full}) = \frac{900}{100000} = 0.009$$

□

---

9. Mencionar 3 de las 5 propiedades que deben cumplir los números pseudoaleatorios para tener calidad suficiente para propósitos de simulación.

Solución.

Aquí listaré las 5 propiedades para poder ver las diferentes combinaciones de respuestas:

a) Al menos, perecer distribuirse con una distribución dada y mantenerse de manera estable en esa distribución
b) ser muy rápidos y eficientes de generar
c) ser capaces de reproducir secuencias más de una vez (reproducibilidad)
d) poder generar más de una secuencia
e) tener periodo largo

□

---

10. Dados los siguientes números normales, con $\mu = 3$ y $\sigma = 1$:

$$4.59,\ 4.153,\ 2.460,\ 2.732,\ 2.973$$

Calcular la función de distribución empírica $F_n(x)$ y calcular $P(F_n(3) \leq 2/5)$.

Solución.

La función de distribución empírica tiene saltos $i/5$ en los puntos dados ordenados de menor a mayor $X_{(i)}$:

x	$F_n(x)$
< 2.46	0
2.46	1/5
2.73	2/5
2.973	3/5
4.153	4/5
≥ 4.53	1

Entonces $P(F_n(3) \leq 2/5) = P(F_n(3) = 1/5) + P(F_n(3) = 2/5) = \binom{5}{1} F(3)(1 - F(3))^4 + \binom{5}{2} F(3)^2(1 - F(3))^3 = \frac{1}{2^5}\left[\binom{5}{1} + \binom{5}{2}\right] = 15/32$, ya que $F(3) = 1/2$.

$\square$