

Primer Examen Parcial. Simulación grupo Viernes

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 está claro lo que se pide interpreten y respondan lo más cercano a su conocimiento.

1. Consideren el siguiente modelo:

$$Y \sim \mathbf{Bin}(n, p)$$
$$X|Y = \sum_{i=1}^Y X_i$$

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 $Y \sim \mathbf{Bin}(n, p)$. 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 algoritmo, y el primer punto es su implementación. La única diferencia, es que puede ser que al mostrar los valores de Y no se obtengan todos los puntos de la partición, pero mientras más grande es el número de simulaciones, es más probable que se tenga todo el dominio de la función.

□

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

- 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 $Z_i = (5Z_{i-1} + 3) \bmod 64$. 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: $Z_1 = (5 * 3 + 3) \bmod 64 = 18$ y $Z_2 = (5 * 18 + 3) \bmod 64 = 29$ Entonces $u_1 = 18/64 = 0.28125$ y $u_2 = 29/64 = 0.453125$.

□

3. Mencionar dos problemas de 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. Dos dados se lanzan 216 veces, y el número de seises en cada lanzamiento fueron:

Número de seises	0	1	2	Total
Frecuencia	130	76	10	216

- Calculen el valor de la estadística de prueba para la hipótesis de que la probabilidad de un seis es $p = 1/6$.
- Expliquen cómo se puede modificar la prueba si la hipótesis a probar es que la distribución es binomial con parámetro p desconocido.

Solución.

- Si los datos son honestos, y se lanzan independientemente, se tiene $P(0\text{seises}) = 25/36$, $P(1\text{seis}) = 5/18$ y $P(2\text{seises}) = 1/36$. Entonces la estadística de prueba es:

$$x^2 = (150 - 130)^2/150 + (60 - 76)^2/60 + (6 - 10)^2/6 = 9.6$$

con 2 grados de libertad.

- Para hacer la prueba equivalente a una binomial, consideramos un éxito obtener exactamente un seis en un lanzamiento, y fracaso el complemento. Si X es el número de lanzamientos con exactamente un seis, tenemos $X \sim \text{Bin}(216, p)$. Entonces la tabla de resultados la cambiamos a tener 76 éxitos y 140 fracasos observados.

□

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

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

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

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

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

□

7. Supongan que X es una variable aleatoria con $f(x) = 0.125xI_{[0,4]}^{(x)}$. Se quiere generar una muestra x entre los valores 1 y 2 de esa densidad: $x \in [1, 2]$ Obtener la observación si se tiene una uniforme $u = 0.63$.

Solución.

Aquí se requiere aplicar el algoritmo correspondiente a la densidad truncada al intervalo considerado. Primero notamos que $F(x) = x^2/16$ en $(0, 4)$. Entonces

- Tomar $u \sim \mathcal{U}(0, 1)$ ($u = 0.63$ dada)
- $v = F(1) + (F(2) - F(1))u = 1/16 + (1/4 - 1/16)u = 0.180625$
- Evalua $X = F^{-1}(v) = 4\sqrt{v} = 1.7$

□

8. Considerando los siguientes dígitos, determinar el número de gaps del número 3 y sus longitudes, y determinar el valor esperado de obtener esas longitudes:

4, 1, 3, 5, 1, 7, 2, 8, 2, 0, 7, 9, 1, 3, 5, 2, 7, 9, 4, 1, 6, 3,
 3, 9, 6, 3, 4, 8, 2, 3, 1, 9, 4, 4, 6, 8, 4, 1, 3, 8, 9, 5, 5, 7,
 3, 9, 5, 9, 8, 5, 3, 2, 2, 3, 7, 4, 7, 0, 3, 6, 3, 9, 9, 9, 5, 5

Solución.

Los gaps son los siguientes:

Longitud	0	1	2	3	4	5	6	7	8	9	10
Número	1	1	2	1	1	2	0	1	1	0	1

En total hay 11 gaps para el número 3. La probabilidad de un gap de longitud $L = x$ es $0.1(0.9)^x$ Los valores esperados de gaps de la longitud dada es $11 * P(L = x)$ y se calcula en la siguiente tabla:

Longitud	0	1	2	3	4	5	6	7	8	9	10
Número obs	1	1	2	1	1	2	0	1	1	0	1
Número esp	1.1	0.99	0.89	0.80	0.72	0.65	0.58	0.53	0.47	0.43	0.38

Con lo anterior se puede calcular una prueba χ^2 de bondad de ajuste.

□

9. ¿Qué mide y cómo se construye la autocorrelación parcial?

Solución.

La autocorrelación parcial se usa para medir el grado de asociación entre y_t y y_{t-k} , cuando se ha eliminado el efecto de los rezagos intermedios $1, 2, 3, \dots, k-1$.

El coeficiente de correlación parcial de orden k se calcula haciendo la regresión de y_t contra los rezagos y_{t-1}, \dots, y_{t-k} . El coeficiente b_k de y_{t-k} es el coeficiente de autocorrelación de orden k .

□

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 $P(F_n(x) \leq 0.75)$.

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 $F(3) = 1/2$.

□