Métodos Estadísticos

Tarea 9

Rubén Pérez Palacios Lic. Computación Matemática
Profesor: Dr. Rogelio Ramos Quiroga

31 de mayo de 2022

Considere un generador de dígitos aleatorios. Suponga que, en una secuencia dada, se producen 51 ceros, de modo que hay 50 pares de ceros concecutivos. La siguiente tabla da el número de espacios entre cada par de ceros.

1	1	6	8	10	22	12	15	0	0
2	26	1	20	4	2	0	10	4	19
2	3	0	5	2	8	1	6	14	2
2	2	21	4	3	0	0	7	2	4
4	7	16	18	2	13	22	7	3	5

1. Describa un modelo probabilístico apropiado para estos datos, bajo el supuesto de que realmente los dígitos son generados aleatoriamente.

   Solución 1 – *Como los datos son generados aleatoriamente entonces la probablidad de generar un digíto entre los* 10 *posibles sería igual además de ser independiente uno de otro. Ahora puesto que los datos recabados solo nos dicen cuantos dígitos distintos de* 0 *hay entre cada par concecutivos de estos, entonces proponemos* X *la variable aleatoria que representa la distancia entre dos ceros concecutivos. La probabilidad de que haya* x *dígitos distintos de cero entre dos ceros concecutivos es de* $\left(\frac{9}{10}^x\right)\left(\frac{1}{10}\right)$ *donde el primero es la probablidad de los primeros* x *dígitos distintos de cero y el segundo de el* $x+1$ *dígito sea cero. Es decir nuestro modelo propuesto es* $X \sim Geo(1/10)$.

2. Construya una tabla de frecuencias y pruebe la bondad de ajuste del modelo

La tabla de frecuencias agrupadas de nuestros datos es

I	Frecuencia
0	6
1 a 2	13
3 a 4	8
5 a 7	7
8 a 12	5
13 a 18	5
19 o mas	6

Por el modelo propuesto nuestra hipotesis nula es

$$H_0 : p_i = \left(\frac{9}{10}^i\right)\left(\frac{1}{10}\right), i \in I.$$

Si f_i es la cantidad de veces que la distancia entre dos ceros fue i, y p_i la probabilidad de que la distancia entre dos ceros sea i, entonces la logverosmilitud de los datos es

$$l(\vec{p}) = \sum_{i=1}^{k} f_i \log(p_i).$$

Ahora recordemos que los resultados de un experimento tienen una distirbución multinomial con probabilidades $\vec{p}$ y que su emv es $\hat{\vec{p}} = \left(\frac{f_1}{n}, \cdots, \frac{f_n}{n}\right)$. En concreto para nuestros datos tenemos que

$$l(\hat{\vec{p}}) = -94.4044$$

En cambio para el modelo propuesto tenemos que

$$l(\hat{\vec{p}}) = -96.7446.$$

Por lo que el cociente de verosimilitus observado es

$$D_{obs} = 4.680509$$

Ahora como tenemos 7 clases el modelo multinomial tiene 6 parámetros libres, además el modelo geométrico no tiene ninguno. Entonces D se distribuye asintóticamente como Ji cuadrada con 6 grados de libertad. Por lo tanto

$$p = \mathbb{P}\left[D \geq D_{obs}\right] = 0.5853.$$

Puesto que $p > 0.05$ concluimos que no hay evidencia suficiente para rechazar la hipotesis nula.