Inferencia Estadística

Tarea 2

26/08/2020

1. Cuando una máquina no se ajusta adecuadamente tiene una probabilidad 0.15 de producir un artículo defectuoso. Diariamente, la máquina trabaja hasta que se producen 3 artículos defectuosos. Se detiene la máquina y se revisa para ajustarla. ¿Cuál es la probabilidad de que una máquina mal ajustada produzca 5 o más artículos antes de que sea detenida? ¿Cuál es el número promedio de artículos que la máquina producirá antes de ser detenida?

2. Los empleados de una compañía de aislantes son sometidos a pruebas para detectar residuos de asbesto en sus pulmones. Se le ha pedido a la compañía que envíe a tres empleados, cuyas pruebas resulten positivas, a un centro médico para realizarles más análisis. Si se sospecha que el 40 % de los empleados tienen residuos de asbesto en sus pulmones, encuentre la probabalidad de que deban ser analizados 10 trabajadores para poder encontrar a 3 con resultado positivo.

3. Para el siguiente ejercicio es necesario usar `R`.

   a) Considere una moneda desequilibrada que tiene probabilidad p de obtener águila. Usando el comando `sample`, escriba una función que simule N veces lanzamientos de esta moneda hasta obtener un águila. La función deberá recibir como parámetros a la probabilidad p de obtener águila y al número N de veces que se repite el experimento; y tendrá que regresar un vector de longitud N que contenga el número de lanzamientos hasta obtener un águila en cada uno de los N experimentos.

   b) Usando la función anterior simule $N = 10^4$ veces una variable aleatoria $\text{Geom}(p)$ para $p = 0.5, 0.1, 0.01$. Grafique las frecuencias normalizadas en color azul. Sobre está última figura empalme en rojo la gráfica de la función de masa correspondiente. ¿Qué observa?

   c) Repita el inciso anterior para $N = 10^6$. Además calcule el promedio y la desviación estándar de las simulaciones que realizó ¿Qué observa?

4. Usando las ideas del inciso anterior escriba una función en `R` que simule N veces los lanzamientos de moneda hasta obtener r águilas. La función deberá recibir como parámetros a la probabilidad p de obtener águila, al número r de águilas a observar antes de detener el experimento y al número N de veces que se repite el experimento; y tendrá que regresar un vector de longitud N que contenga el número de lanzamientos hasta obtener las r águilas en cada uno de los N experimentos. Grafique las frecuencias normalizadas de los experimentos para $N = 10^6$, $p = 0.2, 0.1$ y $r = 2, 7$ y compárelos contra la función de masa de la distribución más adecuada para modelar este tipo de experimentos.

5. Considera X una v.a. con función de distribución F y función de densidad f, y sea A un intervalo de la línea real $\mathbb{R}$. Definimos la función indicadora $1_A(x)$:

$$1_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{en otro caso} \end{cases}$$

Sea $Y = 1_A(X)$. Encuentre una expresión para la distribución acumulada y el valor esperado de Y.

6. Las calificaciones de un estudiante de primer semestre en un examen de química se describen por la densidad de probabilidad

$$f_y(y) = 6y(1-y) \qquad 0 \leq y \leq 1.$$

donde y representa la proporción de preguntas que el estudiante contesta correctamente. Cualquier calificación menor a 0.4 es reprobatoria. Responda lo siguiente:

   a) ¿Cuál es la probabilidad de que un estudiante repruebe?

   b) Si 6 estudiantes toman el examen, ¿cuál es la probabilidad de exactamente 2 reprueben?

7. Escriba una función en `R` que simule una aproximación al proceso Poisson a partir de las 5 hipótesis que usamos en clase para construir tal proceso. Usando esta función, simule tres trayectorias de un proceso Poisson $\lambda = 2$ sobre el intervalo $[0, 10]$ y grafíquelas. Además simule 10^4 veces un proceso de Poisson N con $\lambda = 1/2$ y hasta el tiempo $t = 1$. Haga un histograma de $N(1)$ en su simulación anterior y compare contra la distribución de Poisson correspondiente.
Hint: Considere el intervalo $[0, T]$ *y un número real positivo* dt *que sea mucho más pequeño que la longitud de* $[0, T]$ *y que divida dicha longitud, digamos* $T/dt = 1000$ *veces. Divida el intervalo* $[0, T]$ *en intervalitos de longitud* dt *que tengan la forma* $(k \cdot dt, (k+1) \cdot dt]$, $k = 0, 1, 2, \ldots, (T/dt - 1)$. *Para cada uno de estos intervalitos simule una v.a. Bernoulli*$(\lambda \cdot dt + 10^{-6})$ *y guarde su resultado en un vector del tamaño adecuado.*

8. En una oficina de correo los paquetes llegan según un proceso de Poisson de intensidad λ. Hay un costo de almacenamiento de c pesos por paquete y por unidad de tiempo. Los paquetes se acumulan en el local y se despachan en grupos cada T unidades de tiempo (es decir, se despachan en T, $2T$, $3T$, ...). Hay un costo por despacho fijo de K pesos (es decir, el costo es independiente del número de paquetes que se despachen). (a) ¿Cuál es el costo promedio por paquete por almacenamiento en el primer ciclo $[0, T]$? (b) ¿Cuál es el costo promedio por paquete por almacenamiento y despacho en el primer ciclo? (c) ¿Cuál es el valor de T que minimiza este costo promedio?

9. Considere la siguiente función.

$$F(x) = \begin{cases} 0 & \text{para } x < 0 \\ 0.1 & \text{para } x = 0 \\ 0.1 + 0.8x & \text{para } 0 < x < 1 \\ 1 & \text{para } 3/4 \leq x \end{cases}$$

¿Es una función de distribución? Si es una función de distribución, ¿corresponde a una variable aleatoria discreta o continua?

10. **Este es un problema al que se recurirá en el futuro**; su intención es que empiencen a jugar con datos reales. El archivo `Delitos.csv` contiene información sobre los delitos denunciados en la ciudad de Aguascalientes, para el período comprendido entre enero de 2011 a junio del 2016. Dicho archivo contiene 5 columnas: la primera columna contiene la fecha

de denuncia del delito; la columna TIPO muestra una descripción del tipo de delito; la columna CONCATENAD presenta un descripción más amplia del delito; la columna SEMANA contiene la semana del año a la que corresponde la fecha de denuncia; y la columna SEMANA_COMPLETAS indica la semana a lo largo del estudio en la cual se presentó la denuncia. A través de métodos gráficos (e.g. boxplots) traten de determinar el comportamiento semanal de los delitos y discutan alternativas de modelos para describir los delitos cometidos en forma relativamente apropiada.

Honour problems (no es obligatorio entregarlos, pero dan crédito extra)

1. Cambiando las hipótesis 2 y 3 que se usaron para contruir los procesos de Poisson homogéneos a la forma indicada en la diapositiva 135, deduzca la distribución del número de eventos que ocurren durante el intervalo $[t_1, t_2]$.

2. Sea $N(t)$, $t \geq 0$ un proceso de Poisson con parámetro $\lambda > 0$. Para $0 < u < t$ y $0 \leq k \leq n$, calcule la probabilidad $P(N(u) = k|N(t) = n)$. Interpreta los resultados.

Entrega: 08/09/2020.