Proyecto Final

Fecha de entrega: 11 de diciembre de 2017

Debe entregar un informe donde, para cada punto del proyecto, debe incluir:

1. Una introducción del problema.
2. Código y todos los detalles de implementación (valores de tamaño de las muestras, número de iteraciones, valores de parámetros, etc.).
3. Resultados de las simulaciones.
4. Discusión de los resultados y conclusiones.

Puede usar todas las funciones disponibles en R para hacer las pruebas y/o procedimientos necesarios. La fecha de sustentación del proyecto será asignada más adelante.

Problema 1

Este problema busca establecer puebas de hipótesis unilaterales para el problema de localización de de dos muestras para distribuciones con colas pesadas. Recuerde que para este problema se tienen dos muestras independientes $X_i = \theta + \varepsilon_i$ $i = 1, \ldots, n$ y $Y_j = \theta + \Delta + \varepsilon_j$, $j = 1, \ldots, m$, donde las variables ε son i.i.d con media 0. Las hipótesis son $H_0 : \Delta = 0$ y $H_1 : \Delta > 0$.

1. Realice una prueba basada en remuestreo con el 5 % de significancia con $\Delta = 0$ y $\Delta = 0,3$ (puede tomar $\theta = 0$) donde las variables ε tienen las siguientes distribuciones:

   a) Laplace

   b) t_r, con $r = 2, 10, 50$.

   En ambos casos tome $n = m = 60$.

2. Realice las mismas pruebas anteriores usando el estadístico de Wilcoxon para dos muestras.
3. Realice las mismas pruebas anteriores usando el estadístico t-combinado $T = \dfrac{\overline{X} - \overline{Y}}{S_p\sqrt{1/n + 1/m}}$, donde $S_p^2 = \dfrac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$. Asuma que T tiene una distribución t_{n+m-2}.
4. Realice las pruebas anteriores usando los retornos diarios de las acciones de Nike durantes los meses abril, mayo y junio de 2016 y los mismos meses del 2017. ¿Cómo decide qué año corresponde a los datos X?

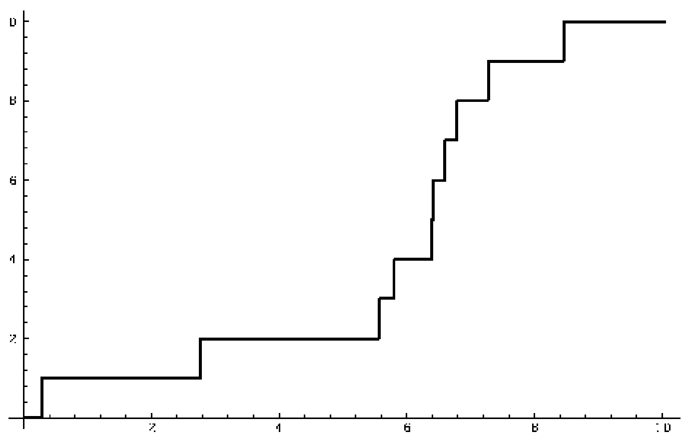
Problema 2

Este problema busca establecer puebas de hiṕotesis sobre la homogeneidad o no de un proceso de Poisson, y para el caso de procesos homogéneos, una prueba sobre igualdad o no de sus intensidades. Para esto vamos a dar las definiciones y los teoremas necesarios.

Definición 1 *Se dice que el proceso $\{N(t)\}_{t\geq 0}$ es un proceso de Poisson homogéneo con intensidad $\lambda > 0$, o $PPH(\lambda)$, si:*

- $N(0) = 0$.
- *Los intervalos $N(t) - N(s)$, para $s < t$ son independientes y estacionarios; es decir que si $0 = t_0 < t_1 \ldots < t_n \leq T$, entonces las variables aletorias $N(t_i) - N(t_{i-1})$ con $i = 1 \ldots n$ son independientes y su distribución depende únicamente del tamaño del intervalo.*
- *El número de eventos en un intervalo de longitud T sigue una distribución de Poisson con media $T\lambda$.*

Un proceso de Poisson se usa por ejemplo para modelar la ocurrencia de enventos como por ejemplo las llamadas a un *call center* o terremotos en algúna región determinada.



Unas propiedades importante de los $PPH(\lambda)$ están dadas por el siguiente teorema.

Teorema 2 *Sea $\{N(t)\}_{t\geq 0}$ un $PPH(\lambda)$ y sea $T_1, T_2, \ldots$ los tiempos de llegada del proceso. Entonces*

(i) Las variables S_i, $i = 1\ldots$, con $S_i = T_i - T_{i-1}$ y $T_0 = 0$, son independientes y siguen una distribución exponencial con media $1/\lambda$.

(ii) Bajo la condición $N(T) = n$, los tiempos de llegada $T_1, T_2, \ldots, T_n$ tienen la misma distribución que los estadśticos de orden correspondientes a n variables aleatorias uniformemente distribuidas en el intervalo $(0, T)$.

El teorema anterior permite simular los tiemspo de llegada de un $PPH(\lambda)$ de dos formas. Una, a apartir de una muestra aleatoria de variables exponenciales con media $1/\lambda$. La otra, fijando un tiempo T, generar una variable Poisson con media $T\lambda$, llamada N, y luego generar N variables uniformes $(0, T)$ y ordenarlas de manera ascendente.

1. Escriba un programa en R que simule un $PPP(\lambda)$ usando la segunda parte del teorema anteriores, para diferentes valores de λ y T.

2. Con el fin de hacer una prueba de hipótesis sobre la homogeneidad o no de un proceso de Poisson, haga una prueba de bondad de ajuste para una distribución uniforme, con una significancia del 5 %, usando los siguientes estadísticos:

   a) El estadístico χ^2 de Pearson.

   b) El estadístico de Cramér- Von Mises definido como $CVM_n := n \int_{-\infty}^{\infty} (\hat{F}_n(x) - F(x))^2 f(x)dx$, donde F y f son las funciones de distribución y densidad de la hipótesis nula, y $\hat{F}_n$ la distribución empírica. Puesto que la distribución de CVM_n no se conoce, debe estimar el percentil 0,95 usando Monte Carlo.

   Use los valores $\lambda = 1$ y $T = 20, 50, 100$, para las pruebas anteriores.

Con el fin de probar la efectividad de las anteriores pruebas, vamos a definir procesos de Poisson no homogéneos.

Definición 3 *Sea $\lambda(t), t \geq 0$ una función y $m(t) = \int_0^t \lambda(s)ds$. Un $PPNH$ con intensidad $\lambda(t)$ es un proceso $\{N(t)\}_{t\geq 0}$ tal que*

- $N(0) = 0$.
- *Los intervalos $N(t) - N(s)$, para $s < t$ son independientes.*
- *La variable aleatoria $N(t) - N(s)$, para $s < t$, tiene una distribución de Poisson con media $m(t) - m(s)$.*

Note que si la función λ es constante tenemos un PPH. Una forma de simular un $PPNH$ con intensidad $\lambda(t)$ es la siguiente:

I. Escoja un valor $T > 0$ y defina como λ^* el valor máximo de la función de intesidad en el intervalo $[0, T]$.

II. Genere una variable Poisson con media $T\lambda^*$, llamada N.

III. Genere una muestra aleatoria $(u_1, \ldots, u_N)$ con distribución uniforme $(0, T)$.

IV. Genere una muestra aleatoria $(h_1, \ldots, h_N)$ con distribución uniforme $(0, \lambda^*)$.

V. Si $h_i < \lambda(u_i)$ seleccione u_i, de lo contrario descártelo.

VI. Ordene los datos seleccionados de manera ascendente.

Use el procedimento anterior para responder lo que sigue.

3. Escriba un programa en R que simule $PPNH$ hasta $T = 30, 50, 100$, con intensidades

   a) $\lambda(t) = 1 + 0,02t$

   b) $\lambda(t) = \begin{cases} 1 & \text{si } t \in [0, 20) \cup [40, 60) \cup [80, 100] \\ 1{,}2 & \text{si } t \in [20, 40) \cup [60, 80) \end{cases}$

4. Haga las pruebas del punto 2. y calcule la potencia de cada una usando 3000 simulaciones de los procesos no homogéneos.

5. Los datos sobre tiempos de temblores en Colombia están disponibles en `http://200.119.88.135/RSNC/index.php/consultas/consulgen`. Escoja 4 conjuntos de datos y haga las pruebas de homogeneidad descritas arriba. Ejemplo: Temblores entre 1998 y 2000 en el departamento del Huila con magnitud mayor a 4. Procure depurar los datos para no contar las réplicas de los temblores. Asegúrese de que cada conjunto de datos tenga al menos 100 datos.

Consideremos ahora el problema de dos muestras para PPH. Sean $X_1, \ldots, X_n$ y $Y_1, \ldots, Y_m$ los tiempos interllegada de los procesos homogéneos $N^1(t)$ y $N^2(t)$ independientes con intensidades λ_1 y λ_2 respectivamente. Queremos plantear una prueba para las hipótesis $H_0 : \lambda_1 = \lambda_2$ y $H_1 : \lambda_1 \neq \lambda_2$. Sean $\overline{X}$ y $\overline{Y}$ las medias muestrales de los tiempos interllegada.

6. Usando el primer punto del Teorema 2 podemos plantear una prueba para el problema de dos muestras con ditribución exponencial. Sea $F = \frac{\overline{X}}{\overline{Y}}$, muestre que bajo H_0 este estadístico tiene una distrbución F y encuentre los grados. Así, podemos definir la prueba: Rechazar H_0 si $F \leq c_1$ o $F \geq c_2$. Haga la prueba anterior con PPH simulados con intensidades $\lambda_1 = \lambda_2 = 1$ y $\lambda_1 = 1, \lambda_2 = 1,2$ para un nivel de significancia del $5\,\%$. Para el segundo caso calcule la potencia de la prueba.

7. Usando los datos de temblores em Colombia, encuentre dos conjuntos de datos que acepten la hipótesis de homogeneidad y realice la prueba de dos muestras anterior.

Problema 3

En en el enlace `https://matematicas.uniandes.edu.co/~mjunca/datos.xlsx` encuentra datos corresondientes a estudiantes que tomaron el curso Cálculo 3 para Administradores y Economistas. Usando los métodos de regresión lineal y ANOVA analice los datos y plantee preguntas, hipótesis, conjeturas, etc. Concluya usando estimaciones, intervalos de confianza, prediccines, etc. Sea muy preciso en sus planteamientos.

Mauricio Junca