Estadistica II
Tarea Pruebas de Bondad de Ajuste

Parte Teorica y Computacional

1. **Estadistico de Prueba K.S** En el contexto de la prueba de bondad de ajuste KS pruebe que:

$$D_n = \sup_{x\in R} |F_n(x) - F_X^*(x)| =^{1} \max\left\{ \max_{1\le i\le n} \left\{ \left|F_n\left(x_{(i-1)}\right) - F_X^*(x_{(i)})\right| \right\}, \max_{1\le i\le n} \left\{ \left|F_n\left(x_{(i)}\right) - F_X^*(x_{(i)})\right| \right\} \right\}$$

2. **(Potencia de las pruebas bajo Normalidad)** El presente ejercicio tiene por objetivo verificar las potencias de las pruebas de bondad de ajuste para el caso Normal.

   a) Simule 1 muestra de tamaño $n = 20$ de una distribución t-student con 2 grados de libertad

   b) Tome $\alpha = 0.05$ y lleve acabo la prueba de bondad de Ajuste:

   $$H_0 : X_i \sim N(0,1) \qquad vs \qquad X_i \nsim N(0,1)$$

      1) Utilizando la prueba de bondad de Ajuste χ^2. (Para esta prueba construya 4 clases $K = 4$ cada una con la misma probabilidad de ocurrir bajo H_0, es decir $p_j = 0.25$ con $j \in \{1,2,3,4\}$)
      2) Utilizando la prueba K.S
      3) Utilizando la prueba Anderson-Darling

   c) En cada prueba determine si rechazo correctamente (Recuerda que los datos fueron generados de una distirbución t-student por lo que se espera que las pruebas sean capaces de detectarlo y rechazar H_0)

   d) Repita los pasos a),b), y c) 10,000 veces contando el número de rechazos que se hizo en cada prueba.

[1]Donde $F\left(x_{(0)}\right) := 0$

e) ¿Cuál prueba de bondad de ajuste rechazo de forma correcta mas ocasiones? Luego entonces ¿Cual es la prueba mas potente para detectar normalidad en este caso?

f) Repita este ejercicio pero ahora simulando en punto a) muestras de una Normal(0,1). Observe ahora que en este caso las pruebas no deberían de rechazar tantas ocasiones pues la muestra en efecto viene de una Normal(0,1). ¿Cuál prueba de bondad de ajuste rechazo de forma incorrecta mas ocasiones? (Comente)

3. **Distribución Lilliefors**
Realice un programa en R para verificar que la distribución de Lilliefors **no** depende de los parámetros de la distribución Normal. Para ello realice 500,000 simulaciones de la distribución Lilliefors con distintos valores para μ y σ. Para este ejercicio realice los casos:

   - $\mu = 0$, $\sigma^2 = 2$
   - $\mu = 2$, $\sigma^2 = 1$
   - $\mu = 100$, $\sigma^2 = 10$

   En cada uno de estos casos encuentre los cuantiles 0.90, 0.95 y 0.99 de la distribucion y verifique que son iguales (2 decimales de precisión)

4. **(Estadístico de prueba Cramer Von-Mises)** Demuestre que el estadístico de prueba Cramer Von-Mises (Ponderador $w(x) = 1$)está dado por la siguiente expresión:

$$V^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(\frac{2i-1}{2n} - u_{(i)} \right)^2$$

   Donde $u_{(i)} = F^*(x_{(i)})$ y $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ es la muestra ordenada. Es decir demuestre que en efecto:

$$V^2 = n \int_{-\infty}^{\infty} (F_n(x) - F^*(x))^2 \, dF^* = \frac{1}{12n} + \sum_{i=1}^{n} \left(\frac{2i-1}{2n} - u_{(i)} \right)^2$$

5. **(Simulación de la distribución Cramer Von-Mises)** Escriba un programa en R que simule la distribución Crame Von-Misses con $n = 5, 10, 15, 20$. (Para tener una buena aproximación de la distirbución se recomienda realizar al menos 100,000 simulaciones.) Encuentre los cuantiles del $1-\alpha$ usando la función *quantile* con $\alpha \in \{0.1, 0.05, 0.025, 0.001\}$. (La tabla de contiene los cuantíles asociados para distintas n, compare sus resultados con esta tabla.)

6. **(QEDF Statistics)** Sabemos que las QEDF se basan en hacer la prueba de bondad de ajuste mediante el cálculo de la estadística:

$$Q^2 = n \int_{-\infty}^{\infty} (F_n(x) - F^*(x))^2 \, w(x) dF^*$$

Donde $w(x)$ es un ponderador. Suponga que usted está interesado en desarrollar una prueba estadística de bondad de ajuste que sea capaz de detectar diferencias solo en la cola derecha por lo que se propone el siguiente ponderador,

$$w(x) = \frac{1}{1 - F^*(x)}$$

Llamemos al estadístico de prueba T^2, entonces:

$$T^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F^*(x))^2}{1 - F^*(x)} dF^*$$

Luego entonces haciendo la transformación $U = F^*(X)$ para probar la hipótesis $H_0 : F(X) = F^*(X)$ obtenemos que el estadístico de prueba se transforma en la siguiente integral

$$T^2 = n \int_{0}^{1} \frac{(F_n(u) - u)^2}{1 - u} du$$

Encuentre el estadístico de prueba en términos de $u_{(i)}$ (la muestra ordenada) y por medio de simulación encuentre los cuantiles $(0.90, 095, 0.075, 0.99)$ de la distribución de este estadístico de prueba cuando $n = 10$

Parte Práctica

1. Suponga que tenemos la siguiente muestra:

$$2.90, 2.85, 1.72, 1.60, 1.33, 2.78, 3.78, 0.39, 2.89, 2.31$$

Realice la prueba de hipótesis utilizando $\alpha = 0.05$

$$H_0 : X_i \sim N(\mu, \sigma^2) \quad vs \quad H_0 : X_i \nsim N(\mu, \sigma^2)$$

- a) Realice la prueba lilliefors y concluya
- b) Realice la prueba Anderson Darling (parámetros son desconocidos) y concluya.

2. Una aseguradora quiere ajustar una distribución a la variable aleatoria T definida como el tiempo de vida de un Mexicano medido en años. La aseguradora tomó una muestra en la que observó a 20 individuos y midió los tiempo de muerte de cada uno:

$$67.55, 58.07, 59.58, 71.21, 70.97, 47.66, 78.05, 62.24, 75.24, 65.71$$

$$65.47, 53.57, 61.17, 60.80, 81.36, 70.61, 51.00, 55.27, 54.64, 72.35$$

La aseguradora sospecha que la variable aleatoria T sigue una distribución Gamma de parámetros $\alpha = 42$, $\beta = 2/3$, donde la densidad Gamma está dada por:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{(\alpha-1)} e^{-(\beta x)}$$

Por lo que se plantea la siguiente hipótesis:

$$H_0 : T \sim Gamma(42, 2/3) \qquad vs \qquad H_1 : T \nsim Gamma(42, 2/3)$$

Tomando $\alpha = 0.05$ concluya en cada una de las siguientes pruebas

- Realice la prueba χ^2 de bondad de ajuste. Tome $c = 4$, donde la partición del rango está formada por los cuartiles de la distribución bajo H_0
- Realice la prueba K.S.
- Realice la prueba Anderson-Darling
- Realice la prueba Cramer-Von-Mises

Con la información de las pruebas anteriores, ¿usted justificaría el uso del modelo Gamma con estos parámetros para este fenómeno aleatorio.?