

19  
2ej



**UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO**

FACULTAD DE CIENCIAS  
DIVISION DE ESTUDIOS DE POSGRADO

**“SOBRE PRUEBAS DE BONDAD DE AJUSTE”**

T E S I S

QUE PARA OBTENER EL GRADO ACADEMICO DE

A C T U A R I O

P R E S E N T A :

MARIA DEL ROSARIO ESPINOSA TUFÍÑO

DIRECTOR DE TESIS  
DR. FEDERICO JORGE O' REILLY TOGNO

CIUDAD UNIVERSITARIA, MEXICO D. F.

1993

**TESIS CON FALLA DE ORIGEN**



Universidad Nacional  
Autónoma de México



## **UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso**

### **DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# CONTENIDO

## INTRODUCCION

### 1. GENERALIDADES SOBRE LA INFERENCIA ESTADISTICA

- 1.1 Conceptos Preliminares
- 1.2 Estimación
- 1.3 Pruebas de Hipótesis
- 1.4 Método de Razón de Verosimilitudes Generalizadas

### 2. BONDAD DE AJUSTE

- 2.1 El Problema de Bondad de Ajuste
- 2.2 Clasificación de los Problemas en Bondad de Ajuste
- 2.3 La prueba Ji-cuadrada de Pearson
- 2.4 Bondad de Ajuste Simple (caso continuo)
- 2.5 Bondad de Ajuste en el caso Discreto

### 3. BONDAD DE AJUSTE (Para la Distribución Poisson)

- 3.1 Prueba de Bondad de Ajuste para la Poisson
- 3.2 Divergencia Logarítmica Kullback-Liebler
- 3.3 La estadística Ji-cuadrada de Pearson como la Distribución Límite para la Divergencia

### 4. ANALISIS EXPLORATORIO

- 4.1 Distribuciones Bajo la Hipótesis Nula
- 4.2 Tablas para KL y C (niveles de significancia empíricos)
- 4.3 Comparación de Potencias

## CONCLUSIONES

## APENDICES

- A. Estimadores para un Experimento Multinomial
- B. Entropía y Divergencia Logarítmica
- C. Serie de Taylor
- D. Documentación de Programas y Subrutinas

## BIBLIOGRAFIA

## INTRODUCCION

---

En ocasiones al observar fenómenos aleatorios es deseable contar con un método que ayude a decidir si el fenómeno bajo estudio tiene cierta regularidad en el sentido de que sea posible asociarle alguna distribución de probabilidad conocida.

Uno de los problemas clásicos en inferencia estadística es el de pruebas de hipótesis; en particular aquellos en las que no solo se desconoce uno o varios parámetros de la distribución de probabilidades sino la distribución misma. En este caso se tiene un problema de "Bondad de Ajuste" y es usual tomar a la familia de distribuciones como la familia de todas las distribuciones continuas o de todas las discretas.

Hasta el momento, gran parte de los métodos para probar Bondad de Ajuste han sido desarrollado para el caso en que la familia paramétrica de distribuciones es una subfamilia de las continuas, y en el caso discreto no existen resultados análogos al anterior en igual número.

Los métodos utilizados para Bondad de Ajuste en el caso discreto, son pocos y en muchos casos requieren de agrupamientos de los valores en clases arbitrarias como el caso de la Ji-cuadrada de Pearson. El uso de la Ji-cuadrada "necesita" agrupar los valores del recorrido en celdas para que la aproximación distribucional empleada sea razonable, y esto representa cierta desventaja, aun así ha llegado a ser una de las más utilizadas en la estadística aplicada.

Dentro del caso discreto, se cuenta también con la versión Kolmogorov-Smirnov aportada en Pettit y Stephens (1977) que requiere que el agrupamiento produzca clases equiprobables, además esta prueba no es invariante ante posibles reagrupamientos. Recientemente, se tiene la propuesta aportada en Kocherlakota y Kocherlakota (1986) y extendida en Rueda *et al* (1991) con base en la función generatriz de probabilidad empírica; no obstante el caso discreto esta poco estudiado.

El problema de bondad de ajuste se puede ver formalmente como un problema de prueba de hipótesis, y para este último se cuenta con un criterio de rechazo clásico conocido como método de Razón de Verosimilitudes Generalizadas (Likelihood Ratio). En el caso continuo el uso del cociente de verosimilitudes conduce a una indeterminación (el denominador es infinito, pues, el supremo sobre la alternativa es infinito), pero en el caso discreto el cociente de verosimilitudes sí está definido, aunque ha sido empleado únicamente para una distribución discreta con soporte finito y suponiendo observados todos los valores del recorrido.

En el presente trabajo se analiza el caso discreto y se discute el uso del criterio de Razón de Verosimilitudes Generalizadas. Se proponen dos pruebas estadísticas derivadas de este criterio, la primera es llamada  $KL$ , con base en el Cociente de Verosimilitudes Generalizadas y la Divergencia Logarítmica Kullback-Liebler, y la segunda denotada por  $C$ , que es una combinación vía Bonferroni de  $KL$  y otra estadística llamada  $S$  que es la suma de probabilidades del recorrido observado, dando igual peso a ambas. Además, se comparan los percentiles obtenidos con los reportados previamente en O' Reilly F.J.(1991) para un estudio para el caso Poisson.

El carácter de este análisis es de naturaleza exploratoria, donde la exactitud de los niveles de significancia son reportados por medio de simulaciones Monte Carlo ejemplificado para el caso Poisson únicamente, aunque el procedimiento se puede extender a otras distribuciones discretas.

Para el presente trabajo, fue necesario un equipo de computación (486/33) y el uso del lenguaje de programación FORTRAN 77 ver. 3.2, utilizando ciertas rutinas del paquete estadístico IMSL. Se presenta material derivado de los conceptos matemáticos preliminares y en cierta forma, en la aportación final, los resultados resumidos en tablas de las simulaciones ejecutadas por los programas computacionales.

## 1. GENERALIDADES SOBRE LA INFERENCIA ESTADÍSTICA

### 1.1 Conceptos Preliminares

Parece una preocupación natural del hombre la búsqueda de mejores explicaciones al comportamiento de fenómenos de naturaleza impredecible, procurando reducir esta incertidumbre y sus efectos, y de ser posible, formular predicciones relativamente confiables.

Como una respuesta al modelaje de fenómenos calificados de aleatorios surgieron los modelos probabilísticos; ante un fenómeno de esta naturaleza a lo más que se puede aspirar es a conocer el modelo probabilístico que lo representa.

Se podría decir que la inferencia estadística estudia algunos aspectos de los fenómenos probabilísticos. Si la ley probabilística que representa a cierto fenómeno aleatorio es totalmente conocida, la estadística no tiene cabida alguna; ésta se utiliza sólo cuando la ley probabilística bajo consideración es total o parcialmente desconocida. El grado de desconocimiento de una ley probabilística se va reduciendo mediante un proceso de observación e inferencia que concluye con un modelo más preciso.

Asociado a fenómenos aleatorios existe la noción de variable aleatoria que no es otra cosa que una transformación a la escala real del resultado aleatorio (evento). Dado un conjunto  $\Omega$ , una  $\sigma$ -álgebra  $A$  de subconjuntos de  $\Omega$ , y una medida de probabilidad  $P$  definida sobre  $A$ , llamado espacio de probabilidad y denotado por  $(\Omega, A, P)$ , una variable aleatoria  $X$  es una función con dominio en  $\Omega$  y contradominio en los reales. La función  $X(\cdot)$  debe ser tal que el conjunto  $A_r = \{ \omega | X(\omega) \leq r \}$  pertenece a  $A$  para todo número real.

Una variable aleatoria discreta  $X$  sobre un espacio de probabilidad  $(\Omega, A, P)$  es una función con dominio en  $\Omega$  y con contradominio en un conjunto finito o lo más numerable de números reales. Una variable aleatoria continua  $X$  sobre un espacio de probabilidad  $(\Omega, A, P)$  es una función con dominio en  $\Omega$  y con contradominio igual a uno o más intervalos de la recta real.

Por construcción el valor que toma una variable aleatoria es de antemano impredecible pudiéndose aspirar solamente a conocer las probabilidades con las que tomará ciertos valores.

Observe que una variable aleatoria  $X$  induce un espacio de probabilidad en el conjunto de los números reales. Al trabajar con variables aleatorias se pueda restringir al espacio de probabilidad inducido sin necesidad de referirse al fenómeno aleatorio que dió lugar a la variable aleatoria. Una variable aleatoria queda completamente especificada si se da por una parte, la colección de los valores que puede tomar, y por la otra las probabilidades respectivas. La ley probabilística asociada a una variable aleatoria se conoce como su distribución.

La función de distribución de una variable aleatoria  $X$ , denotada por  $F_X(\cdot)$  es definida como aquella función con dominio en los reales y contradominio en el intervalo  $[0,1]$ , que satisface:  $F_X(x) = P[X \leq x] = P\{\omega \mid X(\omega) \leq x\}$  para todo número real  $x$ .

Desde el punto de vista probabilístico el objeto deja de ser un fenómeno aleatorio y se convierte en la variable aleatoria  $X$  en sí, concentrándose en su distribución. Si  $X$  es una variable aleatoria discreta entonces su correspondiente función de distribución  $F_X(\cdot)$  es discreta ( $F$  escalonada); si  $X$  es una variable aleatoria continua entonces su correspondiente función de distribución  $F_X(\cdot)$  es continua ( $F$  absolutamente continua).

Si  $X$  es una variable aleatoria discreta se llamará a  $f(x) = P[X = x]$  función de probabilidad de  $X$  si satisface :

$$i) f(x) \geq 0 \text{ para todo } x \in \mathfrak{R}$$

$$ii) \sum_x f(x) = 1$$

Si  $X$  es variable aleatoria absolutamente continua, existe una función  $f(x)$  que satisface :

$$i) f(x) \geq 0 \text{ para todo } x \in \mathfrak{R}$$

$$ii) \int_{-\infty}^{\infty} f(x) dx = 1$$

$$iii) P(a \leq x \leq b) = \int_a^b f(x) dx \text{ para cualesquiera } a \text{ y } b.$$

A  $f(x)$  se le conoce como la función de densidad de probabilidad de  $X$ .

El conjunto de posibles valores que puede tomar una variable aleatoria  $X$  es el soporte de  $X$  esto es, el soporte de  $f(x, \theta)$  denotado por  $SP(f)$  es definido como :

$$SP(f) = \{ x \mid f(x; \theta) > 0 \text{ para algún } \theta \in \Theta \}$$

Si se conoce  $F(x) \forall x \in \mathfrak{R}$ , se conoce la ley que gobierna a  $X$ , por lo que en este caso no hay cabida para la estadística así que requerimos de cierto grado de desconocimiento de  $F$ . Suponer que la distribución  $F$  es parcialmente desconocida equivale a decir que pertenece a una cierta clase o familia de funciones de distribución. Tal es el caso de la familia paramétrica  $\{ F(\cdot; \theta) \mid \theta \in \Theta \}$  que equivale a decir que se conoce  $F(\cdot; \theta)$  en cuanto a su forma analítica pero se desconoce el valor del parámetro  $\theta$ .

El problema de inferencia se completa con dos ingredientes más: contar con observaciones (muestra) de  $F$  y el segundo con la pregunta que interesa contestar sobre

$F$ .

Supóngase que de  $F$  pueden obtenerse  $n$  observaciones independientes  $X_1, \dots, X_n$  que se dicen ser variables aleatorias independientes e idénticamente distribuidas según la ley  $F$ , de la cual sólo se sabe que pertenece a una cierta familia de funciones de distribución. Con base en  $X = (X_1, \dots, X_n)$  el problema a resolver puede ser muy variado dependiendo de la pregunta que se haga sobre  $F$ .

Las preguntas típicas que se hacen en un contexto inferencial tienen que ver con estimación, pruebas de hipótesis, y bondad de ajuste, este último es el tema de estudio en el presente trabajo.

## 1.2 Estimación

La estimación proporciona una "aproximación" sobre el valor desconocido de alguna característica  $\tau(F)$  de la distribución  $F$ , de la cual sólo se sabe que  $F \in \mathcal{F}$ , donde  $\mathcal{F}$  denota a alguna familia de distribuciones; misma que puede ser de distribuciones discretas o continuas. Si la familia es paramétrica, la característica  $\tau(F)$  es una función de  $\theta$ , el parámetro desconocido.

Existen dos formas de llevar a cabo la estimación: puntualmente, utilizando la información de la muestra para obtener un sólo punto que estime la característica; y estimación por intervalo que utiliza un intervalo para obtener dos números que se supone acotan a la característica de interés. En cada caso la estimación se hace mediante un estimador, que es una regla que establece cómo calcular el valor estimado con base en los datos de la muestra.

Como los estimadores pueden ser muchos, serán de interés aquellos que sean "los mejores" y como características deseables en un estimador se pide por ejemplo el "insesgamiento" y varianza mínima. Con base en  $X = (X_1, \dots, X_n)$  se obtiene  $\hat{\tau}(F)$  que es una función de la muestra  $X$  y es por tanto una cantidad aleatoria, por lo que está sujeta a una ley probabilística. En el caso del estimador por intervalo, sería conveniente que contara con dos propiedades: contener a la característica "frecuentemente" y que sea relativamente estrecho. Los estimadores por intervalo se llaman comúnmente intervalos de confianza. La probabilidad de que un intervalo de confianza (cuyos extremos son aleatorios; o sea antes de haber observado la muestra) contenga al parámetro, se conoce como coeficiente de confianza  $1-\alpha$ .

### 1.3 Pruebas de Hipótesis

Otro problema en inferencia es el de pruebas de hipótesis; una hipótesis estadística es una afirmación con respecto a alguna característica desconocida de  $F$ , esta afirmación puede involucrar ya sea algún parámetro o alguna forma funcional de la distribución no conocida de interés  $F$ .

En forma general el problema de prueba de hipótesis consiste en proporcionar con base en  $X = (X_1, \dots, X_n)$  un esquema o regla de decisión para rechazar o no rechazar la aseveración  $F \in F_0$ . Si  $F$  es una familia paramétrica, en cuyo caso

$$F = \{ F(\cdot, \theta) | \theta \in \Theta \}, \quad F_0 = \{ F(\cdot, \theta) | \theta \in \Theta_0 \}$$

$$F - F_0 = \{ F(\cdot, \theta) | \theta \in \Theta - \Theta_0 \}$$

con  $\Theta$  el espacio paramétrico y  $\Theta_0 \subset \Theta$ . A la hipótesis  $H_0: F \in F_0$  se le conoce como hipótesis nula, y a la hipótesis  $H_1: F \in F - F_0$  como la hipótesis alternativa. Dependiendo de las cardinalidades de  $F_0$  y  $F - F_0$  las hipótesis pueden ser simples o compuestas.

Las partes de una prueba estadística son la estadística de prueba (función de la muestra) y la región de rechazo que especifica los valores de la estadística de prueba para los cuales se rechazaría la hipótesis nula. La determinación de una región de rechazo adecuada para una prueba estadística es un problema interesante que requiere de especial atención. Nótese que para cualquier región de rechazo fija se pueden cometer dos tipos de errores al llegar a una decisión: se puede decidir rechazar la hipótesis nula siendo la hipótesis nula la verdadera (llamado error tipo I) o se puede decidir no rechazar la hipótesis nula siendo la hipótesis alternativa la verdadera (error tipo II).

La probabilidad de un error tipo I se denota con  $\alpha$ , la probabilidad de un error tipo II con  $\beta$ , y ambas proporcionan una manera práctica para medir la bondad de una prueba. El número  $\alpha$  suele llamarse nivel de significancia asociado con la prueba, y mide de alguna manera, el riesgo que se corre por un rechazo incorrecto de la hipótesis nula. A pesar de ser recomendados valores pequeños para  $\alpha$ , su elección para aplicarlo en análisis prácticos es algo arbitrario.

Una vez observada la estadística de prueba es posible, de manera alternativa al uso del  $\alpha$ , determinar el valor  $p$  ( $p$  value), es decir el nivel de significancia alcanzado por la prueba para ese valor observado de la estadística. Esta cantidad es una estadística que representa el mínimo valor de  $\alpha$  para el cual se rechaza la hipótesis nula con esa muestra particular. Su uso es muy socorrido en la estadística aplicada y se menciona aquí por completéz pero en lo subsecuente se utiliza la noción del  $\alpha$  y el  $\beta$ .

Un concepto relacionado para evaluar el funcionamiento de una prueba se denomina potencia de la prueba, que es la probabilidad de que la prueba rechace correctamente la hipótesis nula, midiendo la capacidad de ésta para detectar si la hipótesis nula es falsa. La potencia de la prueba y la probabilidad del error tipo II se relacionan entre sí

$$\text{potencia} = 1 - \beta$$

En la mayoría de los casos no existe seguridad sobre si la decisión tomada (rechazar  $H_0$  o no rechazar  $H_0$ ) es acertada, por lo que la teoría busca métodos para elaborar los esquemas de decisión que en promedio sean correctos, uno de los métodos más utilizados es el de Razón de Verosimilitudes Generalizadas.

#### 1.4 Método de Razón de Verosimilitudes Generalizadas

Considere la distribución  $F(\cdot, \theta)$  y  $X_1, \dots, X_n$  una muestra aleatoria de  $F(\cdot, \theta)$  donde  $\theta \in \Theta$ , y sea la función de densidad  $f(x; \theta)$ . La función de verosimilitud de  $\theta$  denotada por  $L(\theta, \underline{x})$  para una muestra observada  $\underline{x} = (x_1, \dots, x_n)$ , se define como la densidad conjunta de  $X$  en  $\underline{x}$ , esto es,

$$L(\theta, \underline{x}) = \prod_{i=1}^n f(x_i, \theta)$$

Para la prueba de hipótesis

$$\begin{array}{l} H_0: \theta \in \Theta_0 \quad (\Theta_0 \subset \Theta) \\ \text{vs.} \\ H_1: \theta \in \Theta_1 \quad (\Theta_1 = \Theta - \Theta_0) \end{array}$$

el criterio de decisión para rechazar  $H_0$  es "rechazar"  $H_0$  si  $X$  es tal que<sup>1</sup>

$$\lambda = \frac{\text{Sup}_{\Theta_1} L(\theta, \underline{x})}{\text{Sup}_{\Theta_0} L(\theta, \underline{x})}$$

es pequeña en comparación con  $K_\alpha$  una cierta constante que depende de  $\alpha$ , (tamaño del error tipo I).

Observe que  $\lambda$  es una función de  $\underline{x}$ ,  $\lambda(x_1, \dots, x_n)$ . Cuando las observaciones son remplazadas por sus correspondientes variables aleatorias  $X_1, \dots, X_n$ , entonces se escribe  $\Lambda$  por  $\lambda$ ; esto es,  $\Lambda = \lambda(X_1, \dots, X_n)$ . En realidad  $\Lambda$  es una estadística, o sea una función de la muestra.

Observe además que: es equivalente tomar en el denominador al Supremo bajo  $H_0$ , ya que para  $\lambda$  que satisface necesariamente que  $0 \leq \lambda \leq 1$ ; los valores "críticos" son cuando es pequeño y en ese caso el denominador puede ser evaluado indistintamente como el

<sup>1</sup>  $\lambda$  denota la Razón de Verosimilitudes Generalizadas.

Supremo sobre  $H_1$  o sobre  $H_0 \cup H_1$ , ya que el denominador no puede ser mas pequeño que el numerador si  $\lambda < 1$ .

En el Cociente de Verosimilitudes para el problema Simple vs. Simple, el numerador de  $\lambda$  es la verosimilitud en  $H_0$  y el denominador de  $\lambda$  es la verosimilitud en  $H_1$ ; de manera que se rechazará  $H_0$  cuando la verosimilitud bajo  $H_0$  es pequeña comparada con la verosimilitud bajo  $H_1$ . Si el problema es Compuesto, el numerador y el denominador son verosimilitudes generalizadas; de hecho el numerador es la verosimilitud máxima bajo  $H_0$  , y si es alcanzable se obtiene cuando  $\theta_0^*$  es el estimador máximo verosímil bajo  $H_0$ .

Para la mayoría de los problemas, este método produce la mejor prueba posible en el sentido de la potencia. Desafortunadamente el método no siempre genera una estadística de prueba con una distribución de probabilidad conocida, sin embargo, si el tamaño de la muestra es grande se puede recurrir a una aproximación Ji-cuadrada para la distribución de  $\Lambda$ .

El criterio anterior pone en evidencia la estructura de la región de rechazo pero queda por cuantificar "que tan pequeño es pequeño". Esto es motivo de estudio en cada situación particular, aunque en lo general existen aproximaciones derivadas de la teoría asintótica.

## 2. BONDAD DE AJUSTE

### 2.1 El Problema de Bondad de Ajuste

En estadística paramétrica, cuando se hace una prueba de hipótesis se conoce de antemano la familia de distribuciones que está asociada a la población, y lo único que se prueba es si los parámetros toman cierto valor específico o varían en un cierto subconjunto del espacio paramétrico. Sin embargo hay situaciones en las cuales no sólo se desconoce uno o varios parámetros de la distribución de probabilidades, sino que se desconoce la distribución misma. El problema de Bondad de Ajuste es formalmente un problema de prueba de hipótesis, solo que en este caso  $F$  es no paramétrica y es usual tomarla como la familia de todas las distribuciones continuas o en otros casos de todas las distribuciones discretas, esto es:

$$H_0: F \in F_0$$

vs.

$$H_1: F \in F - F_0$$

en donde el hecho a subrayar es que  $F$  es una familia grande; de hecho no paramétrica (con  $\theta$  de dimensión finita) y  $F_0$  es paramétrica.

En forma tradicional el tipo de pruebas de Bondad de Ajuste se basa en la comparación de los resultados de una muestra aleatoria con aquellos que se espera observar si la hipótesis nula es correcta, y así comprobar si los resultados de la muestra confirman la distribución hipotética.

## 2.2 Clasificación de los Problemas en Bondad de Ajuste

La única categorización que se hace para distinguir problemas de Bondad de Ajuste, dado que  $F \neq F_0$  es siempre no paramétrica, se basa en la cardinalidad de  $F_0$ :

### *Problema Simple de Bondad de Ajuste.*

Se tendrá un problema simple de Bondad de Ajuste cuando  $F_0$  tenga un solo elemento, es decir,  $F_0 = \{F_0\}$  y se suele escribir como:

$$H_0 : F = F_0$$

vs.

$$H_1 : F \neq F_0$$

dejando implícito el que  $F$  es la de todas las continuas o bien todas las discretas.

### *Problema Compuesto de Bondad de Ajuste.*

Si  $F_0$  tiene más de un elemento y es una familia paramétrica, es decir,  $F_0 = \{F(\cdot, \theta) | \theta \in \Theta\}$  entonces se dice que es un problema compuesto de Bondad de Ajuste, o bien, un problema de Bondad de Ajuste en presencia de parámetros desconocidos.

En todo caso  $F \neq F_0$  sigue siendo muy grande y no parametrizable.

Los procedimientos clásicos<sup>2</sup> para pruebas de hipótesis generalmente no son de utilidad en Bondad de Ajuste. En principio los métodos para contrastar hipótesis fueron desarrollados siendo  $F$  una familia paramétrica, y cuando  $F$  es no paramétrica, estos métodos dejan de tener sentido en el caso continuo.

<sup>2</sup> Procedimientos del tipo Neyman Pearson.

Si en Bondad de Ajuste deseamos probar la hipótesis

$$H_0: F \in F_0$$

vs.

$$H_1: F \in F - F_0$$

siendo  $F$  la familia de todas las distribuciones continuas, utilizando el método de razón de verosimilitudes generalizadas,  $H_1$  sigue siendo no paramétrica y mucho muy grande como familia de distribuciones, entonces la verosimilitud generalizada bajo  $H_1$  es no acotada, por lo que

$$\lambda = \frac{\sup_{\theta_1} L(\theta, \underline{x})}{\sup_{\theta_0} L(\theta, \underline{x})}$$

sería siempre cero. Por ello el criterio de Razón de Verosimilitudes no funciona en Bondad de Ajuste para el caso continuo; y es por ésto que se han desarrollado enfoques específicos.

El argumento anterior, sin embargo, no es aplicable para descalificar al cociente de verosimilitudes  $\lambda$  como criterio para el problema de Bondad de Ajuste si  $F$  es la familia de todas las distribuciones discretas.

### 2.3 La Prueba Ji-cuadrada de Pearson

En general una prueba de Bondad de Ajuste compara los resultados de una muestra aleatoria con aquellos que se espera observar si la hipótesis nula es correcta. La comparación puede hacerse mediante la clasificación de los datos que se observan en cierto número ( $K$ ) de categorías, comparando entonces, las frecuencias observadas con las esperadas por cada categoría.

Una de las estadísticas más utilizadas, la  $X^2$ , fué propuesta en 1900 por Karl Pearson, demostrando que su distribución se puede aproximar por una distribución  $\chi^2$  cuando el tamaño de muestra  $n \rightarrow \infty$  (conservando  $K$  fijo). La estadística propuesta por Pearson es la siguiente

$$X^2 = \sum_i^k \frac{(O_i - E_i)^2}{E_i}$$

en donde  $O_i$  es el número de valores de la muestra observados en la  $i$ -ésima categoría y  $E_i$  los esperados bajo la hipótesis nula.

De acuerdo con lo anterior la estadística es la suma sobre todas las categorías de los cocientes entre el cuadrado de las diferencias entre la frecuencia observada y esperada; y la frecuencia esperada. La estadística anterior recibe el nombre de prueba de bondad de ajuste  $X^2$  de Pearson. Si existe una concordancia perfecta entre las frecuencias que se observan y las que se esperan, la estadística tendrá un valor igual a cero; por otro lado si existe gran discrepancia entre estas frecuencias, la estadística tomará un valor muy grande. Por ello se desprende que para un tamaño dado del error de tipo I, la región crítica es el extremo superior de una distribución  $\chi^2$  con  $[K - 1]$  grados de libertad.

La prueba  $X^2$  se usa cuando los datos pueden clasificarse en un número arbitrario de categorías, además es invariante a los efectos del orden en las categorías. Como

desventaja pierde información a causa de posibles reagrupamientos; esto la hace menos poderosa que la prueba Kolmogorov - Smirnov<sup>3</sup> que utiliza toda la información; aunque si los datos fuesen permutados (por ser etiquetas de clases o categorías) la estadística no resulta invariante.

En los casos en que deban estimarse  $r$  parámetros (porque  $H_0$  especifica una distribución paramétrica con  $\theta$   $r$ -dimensional), la aproximación Ji-cuadrada a la prueba  $X^2$  de Pearson, se modifica dejando  $[K - r - 1]$  grados de libertad, y sustituyendo a  $\theta$  por un estimador adecuado al calcular las  $E_i$ .

Para un tamaño específico del error tipo I, la hipótesis nula será rechazada si existe una diferencia significativa entre las frecuencias observadas y esperadas; la decisión es no rechazar  $H_0$  (mas que aceptarla) si la diferencia existente entre las frecuencias observadas y esperadas es, en forma relativa, pequeña.

Si la prueba se hace a un nivel de significancia  $\alpha$  la estadística  $X^2$  se compara con el percentil  $(1 - \alpha)$  de una distribución  $\chi^2_{(K-r-1)}$ .

---

<sup>3</sup> Ver página 20

## 2.4 Bondad de Ajuste Simple (caso continuo)

Para ilustrar el problema general de Bondad de Ajuste consideremos el caso simple con  $F$  continua, de manera que con base en  $X = (X_1, \dots, X_n)$ , muestra de  $F$ , se desea probar

$$H_0 : F = F_0$$

vs.

$$H_1 : F \neq F_0$$

La idea de todo método para probar  $H_0$ , es contrastar la hipótesis con la evidencia, representados en este caso por  $F_0$  y  $X$ .

Asociada a una muestra, existe una función de distribución llamada "empírica", que permite representar la evidencia muestral en forma de distribución:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{x_i \leq x\}} \quad ^4$$

Observe que  $F_n(x)$  es para cada  $x \in \mathcal{R}$  una variable aleatoria y bajo  $H_0$ ,  $nF_n(x)$  se distribuye como una binomial con parámetros  $n$  y  $F_0(x)$ . Además del resultado distribucional que indica el tipo de comportamiento probabilístico que se espera de  $F_n(x)$ , existen resultados relacionados con el comportamiento asintótico de la función de la distribución empírica.

---

<sup>4</sup>  $\mathbb{I}_A$  es la función indicadora del evento  $A$ .

El lema de Glivenko-Cantelli afirma que, bajo  $H_0$

$$\sup_x |F_n(x) - F_0(x)| \rightarrow 0$$

cuando  $n \rightarrow \infty$  con probabilidad 1; esto es, de ser cierta la hipótesis nula  $H_0$ , las funciones  $F_n$  y  $F_0$  deben ser parecidas. El resultado anterior sugirió utilizar como medida de discrepancia entre evidencia e hipótesis a  $D_n = \sqrt{n} \sup_x |F_n(x) - F_0(x)|$ , la Kolmogorov-Smirnov.

Además de la métrica anterior se pueden obtener variantes de ésta y de otras posibles métricas para sugerir otras medidas como por ejemplo:

$$W_n^2 = n \int_{\mathfrak{R}} [F_n(x) - F_0(x)]^2 dF_0(x),$$

$$A_n^2 = n \int_{\mathfrak{R}} \frac{[F_n(x) - F_0(x)]^2}{F_0(x)[1 - F_0(x)]} dF_0(x)$$

En caso de que se lograra obtener su distribución bajo  $H_0$ , podría entonces evaluarse la probabilidad de que la medida elegida tome un valor "grande" que llevaría a cuestionar si  $H_0$  es cierta.

El objeto de conocer bajo  $H_0$  la probabilidad de que la medida tome un valor "grande" con el que se rechazaría  $H_0$ , es precisamente para controlar la probabilidad de que se rechace  $H_0$  equivocadamente (error tipo I). Sin embargo el cómputo para obtener las distribuciones exactas de una métrica suele ser extraordinariamente difícil, mas aún, si depende del valor del tamaño de la muestra  $n$ .

Tanto  $D_n$  como  $W_n^2$  o la variante  $A_n^2$  pueden verse como funcionales del proceso estocástico  $\xi_n$ , cuya ley límite converge al llamado Browniano "atado".

$$\xi_n(x) = \sqrt{n} \{ F_n(x) - F_0(x) \}, x \in \mathfrak{R} \quad (5)$$

<sup>5</sup>  $\xi_n$  es también llamado Proceso Empírico.

El conocimiento de la distribución asintótica de una tal medida puede deducirse del comportamiento de la distribución asintótica del proceso empírico, esto es, el conocimiento del comportamiento de las fluctuaciones en las trayectorias del proceso empírico se traduce en el conocimiento del comportamiento aleatorio de las cantidades reales que resultan al utilizar la funcional correspondiente.

Notacionalmente la convergencia de  $\xi_n$  se expresa por  $\xi_n(x) \Rightarrow B(t)$  con  $t = F_0(x)$  un cambio de variable "tiempo" y  $B(t)$  un Gaussiano de media cero y función de covarianza  $\rho(s, t) \equiv \text{cov}(B(t), B(s)) = s \wedge t - st$  ( con  $\wedge = \text{mínimo}$ ), conocido como Browniano atado ó también como puente Browniano.

El resultado anterior es la generalización de la convergencia a la normal de  $\xi_n(x)$  para  $x$  fijo (Teorema de de Moivre - Laplace).

Las estadísticas  $D_n$ ,  $A_n^2$ ,  $W_n^2$ , como funcionales de  $\xi_n(x)$ ; son:

- Kolmogorov - Smirnov  $\sqrt{n}D_n = \sup_x |\xi_n(x)|$
- Cramér - Von Mises  $W_n^2 = \int_{\mathfrak{R}} \xi_n^2(x) dF_0(x)$
- Anderson - Darling  $A_n^2 = \int_{\mathfrak{R}} \frac{\xi_n^2(x)}{F_0(x)[1-F_0(x)]} dF_0(x)$

Es importante observar que el cambio de variable "tiempo"  $t = F_0(x)$  trae consigo algo más, la transformación de cada  $X_i$  a una uniforme en el intervalo  $[0, 1]$ .

Si  $X \sim F$ , entonces  $U := F(X) \sim U(0,1)$ , esto se puede ver ya que para  $u \in [0,1]$ ,

$$\begin{aligned} G_u(u) &= P(U \leq u) \\ &= P[F(X) \leq u] \\ &= P[X \leq F^{-1}(u)] \\ &= F(F^{-1}(u)) \end{aligned}$$

$= u$ , con  $F^{-1}$  convenientemente elegida.

Si a cada una de las observaciones  $X_1, X_2, \dots, X_n$  se las transforma con  $F_0$ ,  $U_i = F_0(X_i)$ ,  $i = 1, \dots, n$  entonces la hipótesis nula  $H_0: F = F_0$  del problema original, queda caracterizada por una nueva hipótesis nula  $H'_0$  en el siguiente contexto:

$U_1, U_2, \dots, U_n$  muestra de  $G$  en  $[0,1]$ , pruebe  $H'_0: G(u) = u$ ,  $u \in [0,1]$

En otras palabras, bajo la hipótesis nula  $H_0$  y sólo bajo  $H_0$ , las  $U_i$  son una muestra independiente de la distribución uniforme en el intervalo  $[0,1]$ .

Mas aún, si con las  $U_i$  se construye la correspondiente función de distribución empírica.

$$F_n^*(u) = \frac{1}{n} \sum_{i=1}^n 1_{[U_i \leq u]}$$

se puede exhibir el proceso empírico correspondiente en  $[0,1]$

$$\xi_n^*(u) = \sqrt{n} \{ F_n^*(u) - u \}$$

Si  $\delta_n$  es una funcional de  $\xi_n$  que también se puede expresar como una funcional de  $\xi_n^*$ , entonces, la distribución de  $\delta_n$  será independiente de  $F_0$ , como es el caso de las estadísticas  $D_n$ ,  $A_n^2$ ,  $W_n^2$ , que se expresan en términos del proceso  $\xi_n^*(u)$  como

$$\begin{aligned} &\sup \sqrt{n} |F_n^*(u) - u| \\ &n \int_0^1 [F_n^*(u) - u]^2 du \\ &n \int_0^1 \frac{[F_n^*(u) - u]^2}{u(1-u)} du \end{aligned}$$

que a su vez tienen expresiones computacionales en términos de  $U_1, U_2, \dots, U_n$  o bien de las respectivas estadísticas de orden.

Haber mencionado explícitamente que se está haciendo uso de la transformación de una variable aleatoria continua con su propia función de distribución tiene el objetivo de resaltar la idea de transformar al  $[0, 1]$  y ahí probar uniformidad. Así, después de transformar cada  $X_i$  y obtener las  $U_1, U_2, \dots, U_n$ , se evalúa la estadística elegida y se checa con el valor de tablas correspondiente.

Todo lo presentado en Bondad de Ajuste hasta el momento, está restringido al caso continuo, es decir, la familia paramétrica de distribuciones que se quiere probar es una subfamilia de las continuas, dejando el caso discreto poco estudiado.

## 2.5 Bondad de Ajuste en el caso Discreto

Para probar la Bondad de Ajuste de una distribución discreta, recientemente se ha explorado un método que se basa en la consideración de la función generatriz de probabilidades y su contraparte empírica; de hecho si

$$\Phi(t) = \sum_{k=0}^{\infty} t^k P_k$$

es la generatriz de probabilidades para una distribución discreta conocida sobre los naturales incluyendo el cero con  $P[X = j] = P_j$ ,  $j = 0, 1, 2, \dots$ , se define la función generatriz de probabilidades empírica como

$$\Phi_n(t) = \frac{1}{n} \sum_{i=1}^n t^{x_i}$$

Se encontró bajo ciertas condiciones, la ley límite del proceso empírico asociado a la función generatriz de probabilidades:

$$\sqrt{n}(\Phi_n(t) - \Phi(t))$$

que resulta ser un Gaussiano de media cero y cierta función de covarianza. En Rueda et al (1991) se explora la integral del cuadrado del proceso. Allí se utiliza para el caso de hipótesis nula simple y también para el caso con parámetro desconocido. En ese caso se usa

$$\sqrt{n}(\Phi_n(t) - \hat{\Phi}(t))$$

en el cual

$$\hat{\Phi}(t) = \sum_{k=0}^{\infty} t^k P_k(\hat{\theta})$$

pues las  $P_j(\theta)$  dependen de un parámetro.

Al momento se tiene identificado el proceso asintótico, como un Gaussiano de media cero y covarianza del tipo del límite del proceso empírico con parámetros estimados.

### 3. BONDAD DE AJUSTE (para la distribución Poisson)

#### 3.1 Prueba de Bondad de Ajuste para la Poisson

Considere  $X = (X_1, \dots, X_n)$  una muestra aleatoria de la distribución discreta  $f(x)$  con parámetro  $\theta$  desconocido y  $J = \{0, 1, 2, \dots\}$  el conjunto de valores posibles para cada una de las variables, es decir,  $J$  es el soporte de  $f$ .

Supóngase que se desea probar la hipótesis siguiente :

$$H_0 : f(x) = e^{-\theta} \frac{\theta^x}{x!} ; \theta \text{ desconocida}$$

contra la alternativa de que  $f(x)$  no es Poisson pero es cualquier otra distribución con el mismo soporte  $J$ , es decir,

$$H_1 : f(x) \neq e^{-\theta} \frac{\theta^x}{x!}$$

Ante este planteamiento se puede utilizar el cociente de verosimilitudes generalizadas  $\lambda$ , pues en el caso discreto sí está definido. El empleo del cociente  $\lambda$  presupone para el caso Poisson evaluar el supremo bajo  $H_0$ .

Si los datos los escribimos identificando frecuencias

$$N_0 = \# X_i, s = 0$$

$$N_1 = \# X_i, s = 1$$

$$\vdots$$

$$N_j = \# X_i, s = j \quad \text{con } j \in J$$

$$\dots$$

entonces  $N_0, N_1, \dots, N_j, \dots$  son las frecuencias observadas.

<sup>6</sup>  $N_j$  denota el número de variables aleatorias de la muestra iguales a  $j$ .

La función de verosimilitud está dada por

$$L(f; \underline{x}) = \prod_{i=1}^n f(x_i) = \prod_{j=0}^{\infty} [f(j)]^{N_j}$$

en el caso Poisson se tiene que

$$\begin{aligned} L(f; \underline{x}) &= \prod_{i=1}^n f(x_i) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!} \\ &= \frac{1}{\prod_{i=1}^n x_i!} e^{-n\theta} \theta^{\sum_{i=1}^n x_i} \end{aligned}$$

tomando logaritmo<sup>7</sup> y derivando,

$$\begin{aligned} \log L(f; \underline{x}) &= -n\theta + \sum_{i=1}^n x_i \log(\theta) - \log\left(\prod_{i=1}^n x_i!\right) \\ \frac{\partial}{\partial \theta} \log L(f; \underline{x}) &= \frac{-n\theta + \sum_{i=1}^n x_i}{\theta} = 0 \\ \Rightarrow \hat{\theta}(\underline{x}) &= \frac{\sum_{i=1}^n x_i}{n} \end{aligned}$$

de donde el estimador máximo verosímil para  $\theta$  es  $\hat{\theta} = \bar{X}$ , resultado bien conocido.

De manera que bajo  $H_0$

$$\text{máx } L(f; \underline{x}) = \prod_{j=0}^{\infty} \left[ e^{-\hat{\theta}} \frac{\hat{\theta}^j}{j!} \right]^{N_j}$$

<sup>7</sup> En lo sucesivo se trabaja con logaritmo natural ( $\log_e$ ).

Se puede ver (Apéndice A) que elegir  $f(0), f(1), \dots$  que maximicen a  $L(f; \underline{x})$  en  $H_0 \cup H_1$ , conducen a tomar  $f(j)$  como la frecuencia observada,

$$f(j) = \frac{N_j}{n}$$

Por lo tanto

$$\begin{aligned} \lambda &= \frac{\text{Sup}_{H_0} L(f; \underline{x})}{\text{Sup}_{H_0 \cup H_1} L(f; \underline{x})} = \frac{\prod_{j=0}^{\infty} \left[ e^{-\hat{\theta}^j} \frac{\hat{\theta}^j}{j!} \right]^{N_j}}{\prod_{j=0}^{\infty} \left[ \frac{N_j}{n} \right]^{N_j}} \\ &= \prod_{j=0}^{\infty} \left[ \frac{e^{-\hat{\theta}^j} \frac{\hat{\theta}^j}{j!}}{\frac{N_j}{n}} \right]^{N_j} \\ &= \prod_{j=0}^{\infty} \left[ \frac{f(j; \hat{\theta})}{\frac{N_j}{n}} \right]^{N_j} \end{aligned}$$

Obsérvese que en el producto es equivalente utilizar sólo los valores de  $j$  donde  $N_j > 0$  ya que los otros términos son iguales a 1.

Si se denota con  $J$  al subconjunto de  $J = \{0, 1, 2, \dots\}$  tal que  $N_j > 0$ , entonces,

$$\lambda = \prod_{j \in J} \left[ \frac{f(j; \hat{\theta})}{\frac{N_j}{n}} \right]^{N_j}$$

es la estadística del cociente de verosimilitudes generalizadas, en dónde  $f(j; \hat{\theta})$  es la densidad en  $j$  bajo la hipótesis nula, habiendo sustituido el parámetro por el estimador máximo verosímil.

Aplicando logaritmo natural y multiplicando por  $-\frac{1}{n}$ , se obtiene :

$$\begin{aligned} -\frac{1}{n} \log(\lambda) &= \sum_{j \in J} -\frac{N_j}{n} \log \left[ \frac{f(j; \hat{\theta})}{\frac{N_j}{n}} \right] \\ &= \sum_{j \in J} -\frac{N_j}{n} \left[ \log(f(j; \hat{\theta})) - \log\left(\frac{N_j}{n}\right) \right] \\ &= \sum_{j \in J} \frac{N_j}{n} \left[ \log\left(\frac{N_j}{n}\right) - \log(f(j; \hat{\theta})) \right] \\ &= \sum_{j \in J} \frac{N_j}{n} \log \left[ \frac{\frac{N_j}{n}}{f(j; \hat{\theta})} \right] \\ &= \bar{D}^* \left[ \left( \frac{N_j}{n} \right)_{j \in J}, (f(j; \hat{\theta}))_{j \in J} \right]. \end{aligned}$$

Con  $\bar{D}^*$ , casi la medida de divergencia logarítmica Kullback-Liebler de las frecuencias empíricas  $\left( \frac{N_j}{n} \right)$  respecto a las frecuencias estimadas  $f(j; \hat{\theta})$ . (Ver Apéndice B).

Se hace la observación de que si  $\theta$  fuera conocido  $-\frac{1}{n} \log \lambda$  sería

$$\bar{D}^* \left[ \left( \frac{N_j}{n} \right), (f(j; \theta))_{j \in J} \right].$$

### 3.2 Divergencia Logarítmica Kullback-Liebler

Para que  $-\frac{1}{n} \log(\lambda)$  fuera una medida de divergencia logarítmica se requeriría no sólo que  $\left(\frac{N_j}{n}\right)$  sumen la unidad a lo largo de  $J$ , (ya que esto sucede), sino que también las  $f(j; \hat{\theta})$  sumen la unidad a lo largo de  $J$ , cosa que no ocurre.

En vista de esta situación, se define la cantidad  $S$  como la probabilidad bajo la hipótesis nula de los valores observados del soporte, por tanto aleatoria, dada por la muestra  $X_1, \dots, X_n$ , esto es :

$$S = \sum_{j \in J} f(j; \hat{\theta}),$$

o bien, 
$$S = \sum_{j \in J} f(j; \theta) \quad \text{si } \theta \text{ es conocido,}$$

y se denota por  $f^*(j; \hat{\theta})$  a los valores de la densidad estandarizada de tal manera que a lo largo del conjunto  $J$  suman la unidad, esto es,

$$f^*(j; \hat{\theta}) = \frac{f(j; \hat{\theta})}{S}$$

y

$$\sum_{j \in J} f^*(j; \hat{\theta}) = 1.$$

En consecuencia  $-\frac{1}{n} \log(\lambda)$  es función monótona de  $\lambda$  y tiene la siguiente expresión :

$$\begin{aligned}
 -\frac{1}{n} \log(\lambda) &= \sum_{j \in J} \frac{N_j}{n} \log \left[ \left( \frac{\frac{N_j}{n}}{f^*(j; \hat{\theta})} \right) \frac{1}{S} \right] \\
 &= \sum_{j \in J} \frac{N_j}{n} \left[ \log \left( \frac{N_j}{n} \right) - \log(S) \right] \\
 &= \sum_{j \in J} \frac{N_j}{n} \log \left[ \frac{N_j}{f^*(j; \hat{\theta})} \right] - \log(S) \\
 &= D \left[ \left( \frac{N_j}{n} \right), f^*(j; \hat{\theta}) \right]_{j \in J} - \log(S)
 \end{aligned}$$

El primer sumando es la medida de la Divergencia Logarítmica Kullbak-Liebler de las frecuencias empíricas  $\left( \frac{N_j}{n} \right)$  respecto a las "probabilidades" estandarizadas  $f^*(j; \hat{\theta})$  que se denotará por  $KL$  y el segundo sumando corresponde al logaritmo natural de  $S$ . A la estadística  $S$  (que llevaría a rechazar para valores chicos) si se le resta su media y se divide por su desviación estándar, se le utilizará mas adelante con una aproximación normal.

### 3.3 La Estadística Ji-cuadrada de Pearson como la Distribución Límite para la Divergencia Logarítmica

Antes de pasar a la distribución bajo la hipótesis nula, cabe mencionar la relación entre  $\bar{D}$  y la estadística  $X^2$  de Pearson (pero con un pequeño cambio).

$$\begin{aligned}
 \bar{D} &= \sum_{j \in \omega} \frac{N_j}{n} \log \left[ \frac{\frac{N_j}{n}}{f^*(j; \hat{\theta})} \right] \\
 &= - \sum_{j \in \omega} \frac{N_j}{n} \log \left[ \frac{f^*(j; \hat{\theta})}{\frac{N_j}{n}} \right] \\
 &= - \sum_{j \in \omega} \frac{N_j}{n} \log \left[ \frac{f^*(j; \hat{\theta}) - \frac{N_j}{n} + \frac{N_j}{n}}{\frac{N_j}{n}} \right] \\
 &= - \sum_{j \in \omega} \frac{N_j}{n} \log \left[ 1 + \frac{\left( f^*(j; \hat{\theta}) - \frac{N_j}{n} \right)}{\frac{N_j}{n}} \right] \\
 &\stackrel{(*)}{=} - \sum_{j \in \omega} \frac{N_j}{n} \left[ \frac{f^*(j; \hat{\theta}) - \frac{N_j}{n}}{\frac{N_j}{n}} \right] + \frac{1}{2} \sum_{j \in \omega} \frac{N_j}{n} \frac{\left( f^*(j; \hat{\theta}) - \frac{N_j}{n} \right)^2}{\left( \frac{N_j}{n} \right)^2} + \dots
 \end{aligned}$$

<sup>\*)</sup> Desarrollado en serie de Taylor (Ver Apéndice C).

$$\begin{aligned} &\approx \frac{1}{2} \sum_{j \in J} \frac{\left( f^*(j; \hat{\theta}) - \frac{N_j}{n} \right)^2}{\frac{N_j}{n}} \\ &= \frac{1}{2} \sum_{j \in J} \frac{1}{n} \frac{\left( n f^*(j; \hat{\theta}) - N_j \right)^2}{N_j} \\ &= \frac{1}{2n} \sum_{j \in J} \frac{\left( n f^*(j; \hat{\theta}) - N_j \right)^2}{N_j} \end{aligned}$$

Por lo que,

$$2n\bar{D} \approx \sum_{j \in J} \frac{\left( n f^*(j; \hat{\theta}) - N_j \right)^2}{N_j}$$

Por lo tanto, con la notación previa,

$$2nKL \approx \sum_{j \in J} \frac{(O_j - E_j)^2}{O_j}$$

Que se espera tenga una distribución  $\chi^2$  con  $[card(J) - 1]$  grados de libertad (ya que se estimó un parámetro). Lo anterior es una justificación heurística de la distribución Ji-cuadrada que se conjetura como la distribución límite para  $2nKL$ . El  $E_j$  en este caso depende de  $\theta$  y se estima, además de haberse estandarizado.

Se hace la observación de que si  $\theta$  fuera conocido, la aproximación a la  $\chi^2$  sería con  $[card(J) - 1]$  grados de libertad solamente.

#### 4. ANALISIS EXPLORATORIO

##### 4.1 Distribución de KL y C bajo la Hipótesis Nula

El criterio es rechazar la hipótesis nula cuando

$$\lambda = \frac{\text{Sup}_{H_0} L(f; \mathbf{x})}{\text{Sup}_{H_0 \cup H_1} L(f; \mathbf{x})}$$

$$= \prod_{j \in J} \left[ \frac{f(j; \hat{\theta})}{\frac{N_j}{n}} \right]^{N_j}$$

es pequeña, pero esto es equivalente a rechazar la hipótesis nula si  $KL$  es grande y/o  $S$  es pequeño ya que

$$-\frac{1}{n} \log(\lambda) = D \left[ \left( \frac{N_j}{n} \right), f^*(j; \hat{\theta}) \right]_{j \in J} - \log(S)$$

$$= KL - \log(S)$$

Encontrar la distribución exacta de la estadística y la de  $S$  se ve como un problema combinatorio complicado aun haciendo uso de teoría asintótica surgida heurísticamente; ya que se conoce la relación existente entre  $KL$  (la Divergencia Logarítmica Kullback-Liebler) y la estadística  $X^2$  de Pearson, se espera que  $2nKL$  tenga una distribución aproximada Ji-cuadrada con  $[card(J) - 2]$  grados de libertad. Esta es la aproximación empírica que se estará usando. Para  $S$ , su media y varianza pueden ser calculadas y se puede usar entonces una aproximación normal, que bajo ciertas condiciones podría ser buena, esto es:

$$2nKL \sim \chi^2$$

$$S \sim Normal$$

Se procederá a la exploración de tales aproximaciones bajo  $H_0$  realizando estudios de simulación utilizando el "Método Monte Carlo". Con base en el número de rechazos de la hipótesis nula para las distribuciones de  $KL$  y  $S$  comparadas con la Ji-cuadrada y la normal respectivamente, se trata de probar la bondad de una distribución Poisson con parámetro conocido, y cuando el parámetro sea desconocido, se utiliza el estimador Máximo Verosímil para este caso. Estos resultados se obtienen por simulación para obtener los niveles de significancia empíricos.

Se denota con  $KL$  la prueba que rechaza para valores grandes de  $KL$  utilizando la aproximación Ji-cuadrada antes mencionada y con  $C$  la prueba que combina vía Bonferroni  $KL$  con  $S$  ya estandarizado utilizando para cada una un nivel de  $\frac{\alpha}{2}$ .

El valor del nivel de significancia  $\alpha$  es por supuesto un valor "teórico" ya que en la práctica la verdadera distribución no es la utilizada. Es común cuando se usa la prueba de ajuste de Pearson, elegir un número arbitrario de celdas y agrupándolas también de manera arbitraria, esperando lograr una buena aproximación, esto es, un  $\alpha$  próximo al verdadero.

Es sabido que el grado de aproximación de una  $X^2$  de Pearson a una distribución  $\chi^2$  depende de si existen muchos conteos o valores esperados pequeños. Cuando los valores esperados en las categorías son pequeños la aproximación es mala; análogamente, cuando los conteos observados son muy pequeños, la aproximación suele ser mala. Por lo anterior y dado que la estadística  $KL$  esencialmente es una estadística  $X^2$  de Pearson, se exploraron dos posibles agrupamientos de los conteos que eviten tener frecuencias observadas iguales a 1.

Un tipo de agrupamiento, llamado "Arbitrario", simplemente agrupa el o los valores observados con frecuencia 1 con el vecino inmediato que tenga la frecuencia más baja. La otra variante de agrupamiento consistió en agrupar aquel valor en el recorrido con

frecuencia igual a 1, con el valor del recorrido de entre los restantes de la frecuencia más baja. A éste lo vamos a referir como agrupamiento "Invariante".

Obsérvese que si por alguna razón los valores del recorrido fueran permutados, el primer tipo de agrupamiento no necesariamente daría un valor para la estadística invariante a esta permutación, es decir, depende del orden en que se presenten las frecuencias; sin embargo el segundo tipo de agrupamiento produciría un valor de la estadística invariante ante una posible permutación del recorrido. Por lo anterior se consideró de interés explorar el comportamiento ante estos dos posibles agrupamientos (Apéndice D).

En lo sucesivo, se reportan los resultados de 1,000 corridas para diferentes valores del parámetro, los valores empíricos para  $\alpha$  serán reportados para las estadísticas  $KL$  al 10% y 5%, y para  $C$  al 10%, efectuando entonces la comparación entre el  $\alpha$  empírico de las aproximaciones contra el  $\alpha$  teórico (únicamente para la distribución Poisson).

#### 4.2 Tablas para KL y C (niveles de significancia empíricos)

Se presenta a continuación el resumen en tablas de los resultados del número de rechazos en 1,000 simulaciones para la hipótesis nula según las estadísticas *KL* y *C* para una distribución Poisson con parámetro  $\theta = \{1.0, 3.0, 7.0, 10.0\}$  para diferentes tamaños de muestra  $n = \{10, 20, 30, 40, 50\}$ , con agrupamientos Arbitrario e Invariante, y contemplando en ambas variantes los casos en que el parámetro sea conocido y desconocido: en este último caso se comparan los valores de la estadística correspondiente contra los cuantiles de la distribución Ji-cuadrada restando un grado de libertad adicional.

Se puede observar que la exactitud en la reconstrucción de los percentiles utilizando la aproximación Ji-cuadrada es buena salvo para muestras pequeñas y valores grandes del parámetro, se deja ver también la similitud entre éstas y las tablas aportadas previamente para la misma estadística en O' Reilly (1991).

En el análisis de las estadísticas utilizando el agrupamiento Invariante se observa que la convergencia es mas lenta y mejora para muestras de mayor tamaño; pero presenta cierta ventaja teórica con respecto al agrupamiento Arbitrario.

VALORES EMPIRICOS DEL  $\alpha$  PARA LA PRUEBA DE POISSON ( $\theta$ )  
 EN LA TABLA APARECEN LOS RECHAZOS CON BASE EN 1000 CORRIDAS  
 ESTADISTICA "KL" CON PARAMETRO  $\theta$  CONOCIDO Y TAMAÑO DE MUESTRA  $n$

AGRUPAMIENTO ARBITRARIO

$\theta \backslash n$	10	20	30	40	50
1.0	86 (32)	79 (47)	110 (54)	81 (53)	114 (46)
3.0	71 (27)	73 (49)	86 (38)	83 (47)	94 (49)
7.0	44 (26)	79 (38)	79 (35)	83 (49)	81 (46)
10.0	40 (24)	60 (31)	85 (32)	89 (44)	76 (51)

O'REILLY (1991) AGRUPAMIENTO ARBITRARIO

$\theta \backslash n$	10	20	30	40	50
1.0	96 (43)	94 (49)	115 (54)	121 (44)	113 (57)
3.0	61 (27)	74 (44)	97 (49)	86 (49)	88 (39)
7.0	47 (23)	64 (33)	72 (40)	62 (40)	97 (54)
10.0	44 (13)	72 (31)	81 (31)	89 (46)	86 (35)

AGRUPAMIENTO INVARIANTE

$\theta \backslash n$	10	20	30	40	50
1.0	101 (39)	78 (45)	107 (61)	82 (54)	112 (47)
3.0	55 (17)	69 (43)	82 (37)	83 (43)	97 (51)
7.0	25 (19)	66 (22)	64 (27)	76 (44)	76 (43)
10.0	26 (11)	40 (14)	64 (23)	77 (31)	66 (36)

VALORES EMPIRICOS DEL  $\alpha$  PARA LA PRUEBA DE POISSON( $\theta$ )  
 EN LA TABLA APARECEN LOS RECHAZOS CON BASE EN 1000 CORRIDAS  
 ESTADISTICA "C" CON PARAMETRO  $\theta$  CONOCIDO Y TAMAÑO DE MUESTRA  $n$

AGRUPAMIENTO ARBITRARIO

$\theta \backslash n$	10	20	30	40	50
1.0	93	67	216	132	76
3.0	93	111	97	145	153
7.0	75	110	109	119	103
10.0	83	108	98	111	131

O'REILLY (1991) AGRUPAMIENTO ARBITRARIO

$\theta \backslash n$	10	20	30	40	50
1.0	116	68	202	106	89
3.0	97	102	87	128	114
7.0	76	85	106	112	127
10.0	73	90	100	106	111

AGRUPAMIENTO INVARIANTE

$\theta \backslash n$	10	20	30	40	50
1.0	99	65	213	133	77
3.0	83	108	95	142	155
7.0	68	96	101	116	100
10.0	70	92	88	101	116

VALORES EMPÍRICOS DEL  $\alpha$  PARA LA PRUEBA DE POISSON( $\theta$ )  
 EN LA TABLA APARECEN LOS RECHAZOS CON BASE EN 1000 CORRIDAS  
 ESTADÍSTICA "KL" CON PARAMETRO  $\theta$  DESCONOCIDO Y TAMAÑO DE MUESTRA  $n$

AGRUPAMIENTO ARBITRARIO

$\theta \backslash n$	10	20	30	40	50
1.0	36 (8)	105 (60)	117 (75)	108 (63)	128 (48)
3.0	99 (37)	80 (48)	94 (43)	97 (51)	97 (55)
7.0	77 (32)	96 (39)	87 (44)	92 (58)	80 (47)
10.0	66 (23)	72 (35)	99 (30)	109 (53)	86 (61)

O'REILLY (1991) AGRUPAMIENTO ARBITRARIO

$\theta \backslash n$	10	20	30	40	50
1.0	34 (8)	105 (51)	119 (57)	100 (55)	94 (45)
3.0	92 (41)	73 (36)	96 (48)	90 (41)	77 (37)
7.0	58 (28)	91 (41)	97 (39)	87 (41)	77 (33)
10.0	64 (26)	74 (36)	70 (37)	102 (47)	91 (44)

AGRUPAMIENTO INVARIANTE

$\theta \backslash n$	10	20	30	40	50
1.0	41 (8)	100 (64)	121 (79)	104 (63)	126 (48)
3.0	81 (28)	77 (43)	98 (37)	93 (53)	108 (63)
7.0	40 (22)	75 (25)	77 (27)	82 (51)	76 (47)
10.0	41 (14)	50 (18)	76 (22)	95 (41)	77 (45)

VALORES EMPIRICOS DEL  $\alpha$  PARA LA PRUEBA DE POISSON ( $\theta$ )  
 EN LA TABLA APARECEN LOS RECHAZOS CON BASE EN 1000 CORRIDAS  
 ESTADISTICA "C" CON PARAMETRO  $\theta$  DESCONOCIDO Y TAMAÑO DE MUESTRA  $n$

AGRUPAMIENTO ARBITRARIO

$\theta \backslash n$	10	20	30	40	50
1.0	42	93	113	119	80
3.0	83	109	107	129	125
7.0	72	100	111	113	91
10.0	68	102	92	115	135

O'REILLY (1991) AGRUPAMIENTO ARBITRARIO

$\theta \backslash n$	10	20	30	40	50
1.0	46	69	107	102	78
3.0	96	89	103	89	99
7.0	64	94	98	89	96
10.0	46	96	103	112	97

AGRUPAMIENTO INVARIANTE

$\theta \backslash n$	10	20	30	40	50
1.0	42	97	116	119	80
3.0	74	106	102	132	133
7.0	63	88	95	107	91
10.0	57	84	81	103	121

### 4.3 Comparación de Potencias

Se presentan dos pruebas  $KL$  y  $C$  para probar la distribución Poisson, con parámetro conocido. Se concluye haciendo una comparación de estas pruebas con la  $X^2$  de Pearson, que puede ayudar al experimentador a seleccionar la más adecuada.

COMPARACION DE PRUEBAS  $X^2$ ,  $KL$  y  $C$  al 10% PARA LA POISSON  
CON  $\theta$  CONOCIDO. LA ALTERNATIVA ES BINOMIAL NEGATIVA( $r,p$ )  
RESULTADO DEL NUMERO DE RECHAZOS EN 1000 SIMULACIONES  
AGRUPAMIENTO ARBITRARIO

	10			30			50		
( $r$ )= $\theta$	$X^2$	$KL$	$C$	$X^2$	$KL$	$C$	$X^2$	$KL$	$C$
$\theta = 1.0$									
null	88	86	93	83	110	216	88	114	76
1	225	236	294	371	593	535	709	764	709
5	97	137	144	126	172	201	141	202	151
10	112	108	133	92	130	178	102	125	93
$\theta = 3.0$									
null	131	71	93	103	86	97	87	94	153
1	306	622	659	875	990	983	992	999	999
5	141	211	229	238	370	324	384	578	449
10	111	124	150	122	192	172	188	271	213
$\theta = 7.0$									
null	41	44	75	102	79	109	140	81	103
1	185	913	937	993	1000	1000	1000	1000	1000
5	72	364	421	513	789	779	700	950	924
10	69	201	209	234	464	408	334	618	539
$\theta = 10.0$									
null	49	40	83	85	85	98	97	78	131
1	204	974	974	1000	1000	1000	1000	1000	1000
5	93	520	551	776	931	903	897	996	989
10	61	277	309	450	628	570	572	823	784

COMPARACION DE PRUEBAS  $X^2$ ,  $KL$  y  $C$  al 10%  
 CON  $\theta$  CONOCIDO. LA ALTERNATIVA ES BINOMIAL NEGATIVA( $r,p$ )  
 RESULTADO DEL NUMERO DE RECHAZOS EN 1000 SIMULACIONES  
 AGRUPAMIENTO ARBITRARIO  
 O' REILLY (1991)

	10			30			50		
( $r$ )= $\theta$	$X^2$	$KL$	$C$	$X^2$	$KL$	$C$	$X^2$	$KL$	$C$
$\theta = 1.0$									
null	88	88	116	83	115	202	88	113	89
1	225	260	269	371	606	558	709	779	704
5	97	126	134	126	162	218	141	220	154
10	112	86	151	92	181	248	102	169	177
$\theta = 3.0$									
null	131	61	97	103	97	87	87	88	114
1	306	601	608	875	984	976	992	1000	1000
5	141	198	191	238	391	316	384	534	441
10	111	138	122	122	194	153	188	261	181
$\theta = 7.0$									
null	41	47	76	102	72	106	140	97	127
1	185	926	927	993	1000	1000	1000	1000	1000
5	72	338	352	513	811	758	700	947	911
10	69	167	201	234	431	399	334	626	514
$\theta = 10.0$									
null	49	44	73	85	81	100	97	86	111
1	204	968	970	1000	1000	1000	1000	1000	1000
5	93	496	543	776	943	902	897	996	988
10	61	253	276	450	644	581	572	829	751

COMPARACION DE PRUEBAS  $X^2$ ,  $KL$  y  $C$  al 10%  
 CON  $\theta$  CONOCIDO. LA ALTERNATIVA ES BINOMIAL NEGATIVA(r,p)  
 RESULTADO DEL NUMERO DE RECHAZOS EN 1000 SIMULACIONES  
 AGRUPAMIENTO INVARIANTE

(r)= $\theta$	10			30			50		
	$X^2$	$KL$	$C$	$X^2$	$KL$	$C$	$X^2$	$KL$	$C$
$\theta = 1.0$									
null	88	101	99	83	107	213	88	112	77
1	236	203	306	371	543	520	709	810	897
5	137	129	123	126	146	196	141	198	140
10	108	92	135	92	136	197	102	163	116
$\theta = 3.0$									
null	131	55	97	103	82	87	87	97	114
1	306	331	595	875	975	969	992	999	1000
5	141	133	154	238	347	287	384	507	441
10	111	72	142	122	186	132	188	235	201
$\theta = 7.0$									
null	41	25	76	102	64	106	140	76	127
1	185	896	919	993	1000	1000	1000	1000	1000
5	72	298	342	513	764	654	700	929	889
10	69	125	181	234	395	333	334	567	461
$\theta = 10.0$									
null	49	26	73	85	64	98	97	66	131
1	204	958	973	1000	1000	1000	1000	1000	1000
5	93	424	493	776	895	865	897	988	978
10	61	199	277	450	537	491	572	779	701

## CONCLUSIONES

---

En el presente trabajo, de caracter exploratorio, se reportaron los niveles de significancia ( $\alpha$  empíricos) para dos estadísticas propuestas; denotadas como  $KL$  y  $C$ , para probar la Bondad de Ajuste en el caso discreto. Aunque el procedimiento es aplicable a otras distribuciones discretas se ejemplificó solamente para el caso Poisson.

En este estudio se observó que la convergencia es muy parecida en ambas variantes del agrupamiento (el Arbitrario y el Invariante ante posibles reagrupaciones del recorrido observado) en comparación con las reportadas previamente para el mismo problema por O'Reilly F.J.(1991). En cualquier caso los resultados se aproximan a las anteriores exploraciones publicadas.

Este fue un trabajo exploratorio del problema de Bondad de Ajuste en el caso discreto que hasta hoy ha quedado poco estudiado.

Estas estadísticas para el caso discreto son sólo algunas propuestas, y quedaría la labor de explorar a otras variantes de  $KL$  y  $C$  en las que se podría reconocer una mejor aproximación.

Por simples que parezcan ciertos problemas en Bondad de Ajuste su solución no es tan sencilla. La Bondad de Ajuste para otros casos, como lo son distribuciones multivariadas conocidas, o con parámetros desconocidos es un reto en la actualidad.

APENDICES

Apéndice A (Estimadores para un experimento multinomial)

---

Considérese que un experimento E se divide en  $k + 1$  eventos  $A_0, A_1, \dots, A_k$  mutuamente excluyentes con probabilidades  $P_0, P_1, \dots, P_k$ , esto es,  $P_i = P(A_i)$  para  $i = 0, 1, \dots, k$

Se consideran  $n$  repeticiones independientes de E con  $n_0, n_1, \dots, n_k$ , el número de veces que ocurren los eventos  $A_0, A_1, \dots, A_k$ , respectivamente, donde  $n_i \geq 0$  para  $i = 0, 1, \dots, k$  y tales que  $n_0 + \dots + n_k = n$ . Obsérvese que  $n_k$  está determinado automáticamente por  $n_0, n_1, \dots, n_{k-1}$  y  $n$ .

Por tanto se tiene solamente  $k$  variables aleatorias independientes que son  $n_0, n_1, \dots, n_{k-1}$ ; del mismo modo  $P_k$  está determinada automáticamente ya que  $\sum_{i=0}^k P_i = 1$ .

Para  $n_0, n_1, \dots, n_k$  enteros fijos, la función de verosimilitud está dada por:

$$L(\underline{P}, \underline{n}) = P_0^{n_0} P_1^{n_1} \dots P_k^{n_k}$$

y su logaritmo natural por:

$$\log L(\underline{P}, \underline{n}) = n_0 \log(P_0) + \dots + n_k \log(P_k)$$

$$= \sum_{i=0}^k n_i \log(P_i)$$

Para encontrar el vector probabilidades  $[P_0, P_1, \dots, P_k]$  que maximice  $\log L(\underline{P}, n)$

restringido a la condición de que  $\sum_{i=0}^k P_i = 1$  se utiliza el método de Lagrange para maximizar directamente

$$\frac{\partial}{\partial P_j} \left[ \log L(\underline{P}, n) - \eta \sum_{i=0}^k P_i \right] = \frac{n_j}{P_j} - \eta$$

Las derivadas se anulan para  $i = 0, 1, \dots, k$  sólo si  $\frac{n_j}{P_j} = \eta$  para todo  $j$ , de donde se obtiene que el estimador máximo verosímil es la frecuencia relativa del evento del tipo  $j$ ,

ya que  $\sum_{i=0}^k P_i = 1$ , es decir  $\hat{P}_j = \frac{n_j}{n}$

Por lo tanto  $L(\underline{P}, n)$  se maximiza eligiendo  $\hat{P}_0 = \frac{n_0}{n}$ ,  $\hat{P}_1 = \frac{n_1}{n}$ , ...,  $\hat{P}_k = \frac{n_k}{n}$

Observe que el vector aleatorio de frecuencias  $\underline{n} = [n_0, n_1, \dots, n_k]$ , es una estadística suficiente minimal y tiene una distribución multinomial con parámetros

$\underline{P} = [P_0, P_1, \dots, P_k]$ , donde la probabilidad de tener  $n_0, n_1, \dots, n_k$  ocurrencias en  $n$  ensayos será :

$$f(n_0, n_1, \dots, n_k) = \frac{n!}{n_0! n_1! \dots n_k!} P_0^{n_0} P_1^{n_1} \dots P_k^{n_k}$$

## Apéndice B (Entropía y Divergencia Logarítmica)

---

En la teoría de la probabilidad un sistema completo de eventos  $A_1, A_2, \dots, A_n$  es un conjunto de eventos mutuamente excluyentes y exhaustivo, esto es, uno y sólo uno de ellos debe ocurrir en cada ensayo.

Si se dan eventos  $A_1, A_2, \dots, A_n$  de un sistema completo con probabilidades  $P_1, P_2, \dots, P_n$  estas satisfacen que  $P_i \geq 0$  y  $\sum_{i=1}^n P_i = 1$  entonces se dice que se tiene esquema finito

$$A = \begin{bmatrix} A_1 & A_2 & A_n \\ P_1 & P_2 & P_n \end{bmatrix}.$$

Todo esquema finito describe un estado de incertidumbre. En un experimento el resultado debe ser uno de los eventos  $A_1, A_2, \dots, A_n$  de quienes únicamente conocemos las probabilidades de su posible ocurrencia. Parece obvio que el grado de incertidumbre es diferente en esquemas diferentes.

Se define la cantidad  $H(P_1, P_2, \dots) = -\sum_{k=1}^n P_k \log P_k$  como la Entropía del esquema finito  $A$ .

La entropía es una medida conveniente de la cantidad de incertidumbre asociada a un esquema finito dado.

Definimos  $P_k \log P_k = 0$  si  $P_k = 0$ , y  $H(P_1, \dots, P_k) = 0$  si uno de los  $P_i$  es uno (y todos los demás son cero). En otro caso la entropía es positiva.

La distribución de probabilidad para un vector de frecuencias  $n = (n_1, n_2, \dots, n_k)$  con ciertas probabilidades  $P = (P_1, \dots, P_k)$  está dado por

$$W = \frac{n!}{n_1! n_2! \dots n_k!} P_1^{n_1} P_2^{n_2} \dots P_k^{n_k} \quad \text{donde } n = n_1 + \dots + n_k. \quad \text{Usando la relación}$$

$= n \log n - n + o(n)$  se tiene que  $\log W = nB(r; p) + o(n)$  donde  $B(r; p)$  está definido

como  $B(r; p) = -\sum_{k=1}^n r_k \log \left( \frac{r_k}{p_k} \right)$  donde los  $r_k$  están definidos como las frecuencias empíricas

$$r_k = \frac{n_k}{n}.$$

Si en realidad el vector de frecuencias  $n = (n_1, n_2, \dots, n_k)$  tiene probabilidad  $q = (q_1, \dots, q_k)$ ;  $B(r; p)$  converge a la entropía de la verdadera distribución  $q$  con respecto a  $p$  definido por:

$$B(q; p) = -\sum_{k=1}^n q_k \log \left( \frac{q_k}{p_k} \right)$$

La presente definición de entropía conserva la interpretación original de Boltzman como el logaritmo de la verdadera distribución  $q$  respecto a la supuesta distribución  $p$ , obteniendo una medida natural de la desviación de la distribución  $q$  respecto de  $p$ .

Se tiene que  $-\infty \leq B(q; p) \leq 0$  y  $B(q; p) = 0$  sólo si  $p = q$ .

Cuando las distribuciones de probabilidad  $p$  y  $q$  están definidas respectivamente por las funciones de densidad  $q(x)$  y  $p(x)$ , continuas respecto a la medida  $dx$ , la entropía está definida por

$$B(q; p) = - \int q(x) \log \left[ \frac{q(x)}{p(x)} \right] dx .$$

Si  $p = \langle p_i \rangle_{i=1}^k$  y  $q = \langle q_i \rangle_{i=1}^k$  son densidades discretas, la entropía de la

distribución  $p$  respecto a  $q$  es definida por  $B(q;p) = - \sum_{i=1}^k q_i \log \left[ \frac{q_i}{p_i} \right]$

La cantidad definida por  $\bar{D}(q;p) = -B(q;p)$  es llamada medida de divergencia logarítmica Kullback-Leibler entre las distribuciones  $p$  y  $q$ . De hecho  $\bar{D}(q;p)$  no es por supuesto una distancia o una métrica en sentido general, ya que  $\bar{D}(q;p)$  no es una función simétrica de  $p$  y  $q$ . No obstante  $\bar{D}(q;p)$  caracteriza en realidad la desviación de  $p$  respecto de  $q$ .

Se deduce que  $\bar{D}(q;p) \geq 0$  y que  $\bar{D}(q;p) = 0$  si y sólo si  $p = q$ .

Para el caso de la familia multinomial donde  $n = (n_1, n_2, \dots, n_k)$  denotan las frecuencias de la  $i$ -ésima categoría en  $n$  repeticiones con parámetro  $\theta = (\theta_1, \dots, \theta_k)$  donde

$0 < \theta_i < 1$  y  $\sum_{i=1}^k \theta_i = 1$  se observa que:

$$\begin{aligned} \log \left( \frac{r_i}{\theta_i} \right) &= \log \left( \frac{\theta_i + r_i - \theta_i}{\theta_i} \right) \\ &= \log \left( 1 + \frac{(r_i - \theta_i)}{\theta_i} \right) \\ &= \frac{r_i - \theta_i}{\theta_i} - \frac{1}{2} \frac{(r_i - \theta_i)^2}{\theta_i^2} + Op \left( \frac{1}{n} \right) \end{aligned}$$

donde  $Op(f(n))$  denota un término cuyo orden es estocásticamente menor que  $f(n)$ .

$$B(r; \theta) = - \sum_{i=1}^k r_i \log \left( \frac{r_i}{\theta_i} \right)$$

$$-B(r; \theta) = \sum_{i=1}^k r_i \log \left( \frac{r_i}{\theta_i} \right)$$

$$= \sum_{i=1}^k r_i \left[ \frac{\theta_i - r_i}{r_i} - \frac{1}{2} \frac{(\theta_i - r_i)^2}{r_i^2} \right] + Op \left( \frac{1}{n} \right)$$

$$= \sum_{i=1}^k r_i \left( \frac{\theta_i - r_i}{r_i} \right) - \frac{1}{2} \sum_{i=1}^k r_i \frac{(\theta_i - r_i)^2}{r_i^2} + Op \left( \frac{1}{n} \right)$$

$$= - \frac{1}{2} \sum_{i=1}^k \frac{(\theta_i - r_i)^2}{r_i} + Op \left( \frac{1}{n} \right)$$

$$\Rightarrow 2B(r; \theta) = \sum_{i=1}^k \frac{(\theta_i - r_i)^2}{r_i} + Op \left( \frac{1}{n} \right)$$

$$\Rightarrow 2nB(r; \theta) = n \sum_{i=1}^k \frac{(\theta_i - r_i)^2}{r_i} + Op \left( \frac{1}{n} \right).$$

Este es el resultado clásico que demuestra que la estadística de la prueba Ji-cuadrada de Pearson es asintóticamente equivalente a  $-2nB(r; \theta)$ , el logaritmo del cociente de verosimilitud bajo la hipótesis nula  $\theta = \theta_0$ , donde  $\theta_0$  denota la verdadera distribución.

La prueba  $X^2$  de Pearson para la bondad de ajuste se puede interpretar como una medida de la entropía de la verdadera distribución con respecto a la distribución supuesta.

Apéndice C (Serie de Taylor)

---

Si  $f$  es una función con derivadas de todos los órdenes en  $a$ , entonces la serie de Taylor para  $f(x)$  en  $x = a$  se define como :

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)(x-a)^2}{2!} + \dots + \frac{f^{(n)}(a)(x-a)^n}{n!} + R_n$$

$$\text{dónde } f^{(i)}(a) = \left. \frac{d^i f(x)}{dx^i} \right|_{x=a} ;$$

$$R_n = \frac{f^{(n)}(c)(x-a)^n}{(n+1)!} \text{ y } a \leq c \leq x$$

El residuo es una medida de qué tanto difiere  $f$  de alguna aproximación polinómica de grado  $n$ , y podremos representar una función  $f$  por su serie de Taylor cuando ésta tenga derivadas continuas hasta de orden  $n+1$ . Además, si una función  $f$  tiene derivadas de todos los órdenes en un intervalo que contiene a  $a$  y  $x$  se puede representar por su serie de Taylor siempre que

$$\lim_{n \rightarrow \infty} R_n = 0$$

Supongamos  $f(x) = \log_e(1+x)$  y  $a = 0$ ; entonces,

$$f^{(1)}(x) = \frac{1}{1+x} \quad f^{(1)}(0) = 1$$

$$f^{(2)}(x) = -\frac{1}{(1+x)^2} \quad f^{(2)}(0) = -1$$

$$f^{(3)}(x) = \frac{1}{2(1+x)^3} \quad f^{(3)}(0) = 2$$

de manera que  $\log_e(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$  para  $-1 < x \leq 1$ .

Apéndice D (Documentación de Programas y Subrutinas)

```

CCCCCCCCC PROGRAMA POISARB1
CCCCCCCCC Agrupamiento ARBITRARIO con parámetro CONOCIDO
CCCCCCCCC Este programa calcula el número de veces que las estadísticas
CCCCCCCCC KL, S, C y KLA conducen al rechazo de la hipótesis nula, con
CCCCCCCCC una confianza del 90 y 95 %.
CCCCCCCCC ENTRADA:
CCCCCCCCC ALAM      Parámetro de la distribución Poisson (1,15)
CCCCCCCCC N         Tamaño de muestra a generar (<=50)
CCCCCCCCC NUMSIM    Numero de simulaciones
CCCCCCCCC DSEED     Semilla para la generar números aleatorios
CCCCCCCCC ALPHA     Nivel de significancia (10% O 5%)
CCCCCCCCC SALIDA:
CCCCCCCCC IT1       Número de Rechazos segun KL
CCCCCCCCC IT2       Número de Rechazos segun S
CCCCCCCCC IT3       Número de Rechazos segun C
CCCCCCCCC IT4       Número de Rechazos segun KLA
CCCCCCCCC
$debug
DOUBLE PRECISION DSEED
DIMENSION NX(50),NVAL(50),NFREQ(50),PROB(50),REJ1(30),
1          QN(50),Q(50,50),FREQ(50),PPROB(50),
2          ALAM1(4),ALPHA1(2),N1(5)
INTEGER TG1,TG2,TG3,TG4,G1,G2,G3,G4,FREQ
REAL KL,KLA
DATA ALAM1/1,0,3,0,7,0,10,0/,ALPHA1/0,10,0,05/,N1/10,20,30,40,50/
CCCC ARCHIVOS DE SALIDA
OPEN(9,FILE='KL.SAL',STATUS='NEW')
CCCC Continúa la generación de NUMSIM vectores aleatorios de tamaño
CCCC N de una POISSON de parámetro ALAM (NX). Calcula el soporte NVAL.
CCCC y las frecuencias observadas sobre el soporte NFREQ.
DSEED = 84225113
NUMSIM = 1000
DO 888 II = 1,2
ALPHA = ALPHA1(II)
WRITE(*,8888)ALPHA
8888 FORMAT(1X,' ALPHA ',F8.4)
DO 777 JJ = 10,50,10
N = JJ
WRITE(*,7777)N
7777 FORMAT(1X,' N ',I2)
DO 666 KKK = 1,4
ALAM = ALAM1(KKK)
WRITE(*,6666)ALAM
6666 FORMAT(1X,' ALAM ',F5.2)
TG1 = 0
TG2 = 0
TG3 = 0
TG4 = 0
DO 3 K=1,NUMSIM
WRITE(*,53)K,NUMSIM
53  FORMAT(1H+', SIMULACION No. ',I4,' DE ',I4)

```

```

CALL GGPON(ALAM,DSEED,N,NX,IER)
CALL * BIN(DSEED1,DSEED2,DSEED3,N,NRB,PBIN,NX)
DO 2 I=1,N
  NFREQ(I) = 0
  PROB(I) = 0.0
  NVAL(I) = 0
2 CONTINUE
C LLAMADA DE SUBRUTINAS
CALL ORDENA(N,NX)
CALL CLASIFIC(N,NX,NVAL,NFREQ,J)
MM = J
CALL NPROB(NVAL,ALAM,MM,PROB)
CALL NS(PROB,MM,SUM)
CALL MEDVAR(ALAM,N,SUM,SUMSTAR)
CALL ASOCIA(N,NFREQ,PROB,MM,FREQ,PPOB,MN)
CALL * INVAR(NFREQ,PROB,MM,FREQ,PPOB,MMM)
MMM = MN
CALL NKL(K,FREQ,PPOB,MMM,N,SUM,KL,KLA)
IF(ALPHA .EQ.0.1)
ICALL DECISION(K,SUMSTAR,MMM,KL,KLA,IT1,IT2,IT3,IT4)
IF(ALPHA .EQ.0.05)
ICALL CHOICE(SUMSTAR,MMM,KL,KLA,IT1,IT2,IT3,IT4)
TG1 = TG1 + IT1
TG2 = TG2 + IT2
TG3 = TG3 + IT3
TG4 = TG4 + IT4
3 CONTINUE
WRITE(*,1003) NUMSIM,N,ALAM,ALPHA, TG1, TG2, TG3, TG4
WRITE(9,1003) NUMSIM,N,ALAM,ALPHA,TG1, TG2, TG3, TG4
1003 FORMAT(1X,'HICIMOS ',4,' SIMULACIONES DE TAMAÑO N = ',14,/,
1 IX,'CON PARAMETRO LAMDA = ',F8.4,' Y ALPHA = ',F5.2,/,
1 IX,'RECHAZOS ACUMULADOS PARA LAS ESTADISTICAS :/,/,
2 IX,' KL = ',14,/,
3 IX,' S* = ',14,/,
4 IX,' C = ',14,/,
5 IX,' KLA = ',14)
666 CONTINUE
777 CONTINUE
888 CONTINUE
CLOSE(9)
STOP
END

```

```

C***** SUBROUTINA ORDENA *****C
C OBJETIVO : C
C La subrutina ORDENA los valores POISSON del vector NX C
C en forma ascendente, regresando el mismo vector NX ya ordenado C
C ENTRADA : NX Vector de Observaciones Poisson C
C N Tamaño del Vector C
C SALIDA : NX Vector Ordenado Ascendentemente C
C*****C

```

```

SUBROUTINE ORDENA(N,NX)
DIMENSION NX(50)
NMI = N - 1
DO 20 I=1,NMI
JJ = I + 1
DO 20 J=JJ,N
IF ( NX(I) -NX(J) ) 20,20,5
5 NTEMP = NX(I)
NX(I) = NX(J)
NX(J) = NTEMP
20 CONTINUE
RETURN
END

```

```

C*****SUBROUTINA CLASIFIC *****C
C OBJETIVO: C
C Agrupa y contabiliza la frecuencia con la que se presenta C
C cada valor de NVAL y los guarda en el vector NFREQ. C
C Da una relación 1-1 entre NVAL y NFREQ. C
C ENTRADA : NX Vector de Observaciones Poisson Ordenado C
C N Tamaño del Vector C
C SALIDA : NVAL Vector de Valores Ordenados>0 (Soporte) C
C NFREQ Vector de Frecuencias Observadas en NX C
C para cada elemento de NVAL. C
C J Dimensión de NVAL y NFREQ. C
C*****C

```

```

SUBROUTINE CLASIFIC (N,NX,NVAL,NFREQ,J)
DIMENSION NX(50),NVAL(50),NFREQ(50)
J = 0
ICLASE = NX(1)
DO 15 I=1,N
IF( ICLASE .EQ. NX(I) .AND. I .EQ. N ) GO TO 14
IF( ICLASE .EQ. NX(I) ) GO TO 15
J = J + 1
NVAL(J) = ICLASE
ICLASE = NX(I)
IF( I .EQ. N ) GO TO 14
GO TO 15
14 J = J + 1
NVAL(J) = NX(I)
15 CONTINUE
I = 1
NVAL(I) = NX(1)
DO 40 JJ = 1,N
35 IF( NVAL(I) .EQ. NX(JJ) ) GO TO 30
I = I + 1
GO TO 35
30 NFREQ(I) = NFREQ(I) + 1
40 CONTINUE
RETURN
END

```

```

C..... SUBROUTINA NPROB .....C
C OBJETIVO : C
C Evalua cada una de la probabilidades Poisson sobre el C
C vector NVAL (soporte) de dimension MM. C
C ENTRADA : NVAL Vector de valores C
C ALAM Parámetro lamda C
C MM J C
C SALIDA : PROB Vector con las probabilidades Poisson C
C.....C
SUBROUTINE NPROB(NVAL,ALAM,MM,PROB)
DIMENSION NX(50),NVAL(50),PROB(50)
EX = EXP(-ALAM)
DO 20 I=1,MM
IF( NVAL(I) .EQ. 0)GO TO 14
ITOP = NVAL(I)
FAC = NVAL(I)
DALAM = ALAM ** NVAL(I)
DO 21 K=1,ITOP
NI = NVAL(I) - K
IF(NI .EQ. 0) GO TO 21
FAC = FAC * NI
21 CONTINUE
COC = DALAM / FAC
PROB(I) = EX * COC
GO TO 20
14 EX = EXP(-ALAM)
PROB(I) = EX
20 CONTINUE
RETURN
END
C..... SUBROUTINA PROBLAM .....C
C OBJETIVO : C
C Evaluar cada una de las probabilidades Poisson con C
C parámetro desconocido sobre NVAL, estimando el parámetro C
C por máxima verosimilitud con base en la muestra. C
C ENTRADA : NVAL Vector de valores C
C NX Valores observados Poisson C
C MM J C
C N Tamaño de muestra C
C SALIDA : PROB Vector de probabilidades; parámetro estimado C
C.....C
SUBROUTINE PROBLAM(NX,NVAL,MM,N,PROB)
DIMENSION NVAL(50),PROB(50),NX(50)
KON = 0
DO 10 I=1, N
KON = KON + NX(I)
10 CONTINUE
ALAM = FLOAT(KON) / FLOAT(N)
EX = EXP(-ALAM)
DO 20 I = 1, MM
IF( NVAL(I) .EQ. 0)GO TO 14
ITOP = NVAL(I)
FAC = NVAL(I)
DALAM = ALAM ** NVAL(I)
DO 21 K=1,ITOP
NI = NVAL(I) - K
IF(NI .EQ. 0) GO TO 21
FAC = FAC * NI

```

```

21 CONTINUE
   COC = DALAM / FAC
   PROB(I) = EX * COC
   GO TO 20
14 EX = EXP(-ALAM)
   PROB(I) = EX
20 CONTINUE
   RETURN
   END

```

```

C***** SUBROUTINA MEDVAR *****C
C OBJETIVO :
C Calcula la estadística "S" estandarizada. Calcula la media
C y la varianza de la distribución.
C ENTRADAS: ALAM Parámetro de la Poisson
C N Tamaño de la muestra
C SUM Estadística "S"
C SALIDA : SUMSTAR Estadística "S" Estandarizada.
C*****C

```

```

SUBROUTINE MEDVAR(ALAM,N,SUM,SUMSTAR)
DIMENSION QN(50),Q(50,50),PJ(50),MM(15),ARREG(30),NX(50)
INTEGER ARREG
REAL MEDIA
DATA MM/6,8,10,12,14,16,17,19,20,22,23,25,26,28,29/
MEDIA=0.0
LAM = ALAM
IF(LAM.GE.1) GO TO 20
MMS= MM(1)
GO TO 21
20 MMS= MM(LAM)
21 DO 22 I=1,MMS
   ARREG(I)=I-1
22 CONTINUE
   CALL NPROB(ARREG,ALAM,MMS,PJ)
   DO 30 J=1,MMS
      PJM=1.0-PJ(J)
      QN(J)=PJM**FLOAT(N)
      MEDIA=MEDIA+ QN(J)*PJ(J)
30 CONTINUE
   DO 33 I=1,MMS
      DO 33 J=1,MMS
         IF (I-J)31,32,31
31 Q(I,J)=QN(I)*QN(J)+(1.0-PJ(I)-PJ(J))**FLOAT(N)
      GO TO 33
32 Q(I,J)=QN(I)*(1.0- QN(I))
33 CONTINUE
   VAR=0.0
   MEDIA=1.0-MEDIA
   DO 40 I=1,MMS
      DO 40 J=1,MMS
         VAR=VAR+PJ(I)*PJ(J)*Q(I,J)
40 CONTINUE
   DEN=SQRT(VAR)
   SUMSTAR=(SUM-MEDIA)/DEN
   RETURN
   END

```

```

C***** SUBROUTINA NS *****C
C OBJETIVO : C
C Calcular estadística "S". Checa que la probabilidades C
C estandarizadas por S sumen la unidad. C
C ENTRADAS : PROB Vector de probabilidades C
C MM Dimension del vector NVAL C
C SALIDA : SUM Estadística "S" C
C*****C
SUBROUTINE NS(PROB,MM,SUM)
DIMENSION PROB(50),PROBS(50)
SPROBS=0.0
SUM = 0.0
DO 10 K = 1, MM
10 SUM = SUM + PROB(K)
DO 11 K = 1, MM
PROBS(K) = PROB(K) / SUM
SPROBS = SPROBS + PROBS(K)
11 CONTINUE
RETURN
END
C***** SUBROUTINA ASOCIA *****C
C OBJETIVO : C
C Agrupa las frecuencias con valor I con la frecuencia vecina C
C más pequeña para no tener celdas con un solo elemento. C
C ENTRADAS : PROB Probabilidades del vector NVAL C
C NFREQ Frecuencias del vector NVAL C
C MM Dimension de los vectores NFREQ y NPROB C
C SALIDAS : FREQ Vector de frecuencias agrupadas C
C PPROB Vector de Probabilidades agrupadas C
C MN Dimension de los vectores FREQ y PPROB C
C*****C
SUBROUTINE ASOCIA(N, NFREQ, PROB, MM, FREQ, PPROB, MN)
DIMENSION PROB(50),NFREQ(50),FREQ(50),PPROB(50)
INTEGER FREQ
DO 1 I = 1,N
FREQ(I) = 0
PPROB(I) = 0.0
1 CONTINUE
C***PREGUNTA POR EL ULTIMO ELEMENTO IGUAL A "1", Y LO SUMA AL PENULTIMO
IF ( NFREQ(MM) .EQ. 1 ) GO TO 5
GO TO 10
5 MMI = MM - 1
NFREQ(MM1) = NFREQ(MM1) + NFREQ(MM)
PROB(MM1) = PROB(MM1) + PROB(MM)
MM = MM1
C***PREGUNTA POR EL PRIMER ELEMENTO IGUAL A "1", Y LO SUMA AL SEGUNDO
10 IF ( NFREQ(1) .EQ. 1 ) GO TO 15
GO TO 25
15 NFREQ(1) = NFREQ(1) + NFREQ(2)
PROB(1) = PROB(1) + PROB(2)
DO 20 I = 3, MM
IM = I - 1
NFREQ(IM) = NFREQ(I)
PROB(IM) = PROB(I)
20 CONTINUE
MM = MM - 1
C*** HACE COMPARACIONES CON LOS INTERMEDIOS
25 MM1 = MM - 1

```

```

DO 40 I = 2, MMNI
IF ( NFREQ(I) .EQ. 1 ) GO TO 27
GO TO 40
27  IMNI = I - 1
    IMI = I + 1
    IF ( NFREQ(IMNI) - NFREQ(IMI) ) 30,30,35
30  NFREQ(IMNI) = NFREQ(IMNI) + NFREQ(I)
    PROB(IMNI) = PROB(IMNI) + PROB(I)
    GO TO 40
35  NFREQ(IMI) = NFREQ(IMI) + NFREQ(I)
    PROB(IMI) = PROB(IMI) + PROB(I)
40  CONTINUE
C***RESECUENCIA EL VECTOR "NFREQ" OMITIENDO LOS VALORES CON "1" YA QUE
C***ANTERIORMENTE SE SUMARON AL VECINO MAS PEQUEÑO
MN = 0
DO 50 I = 1, MM
IF ( NFREQ(I) .EQ. 1 ) GO TO 50
MN = MN + 1
FREQ(MN) = NFREQ(I)
PPROB(MN) = PROB(I)
50  CONTINUE
RETURN
END
C..... SUBROUTINA INVAR .....C
C  OBJETIVO:
C      Agrupa las frecuencias observadas para no tener          C
C      conteos de uno                                           C
C  ENTRADA : NFREQ  Vector de Frecuencias                        C
C            PROB   Vector de Probabilidades                    C
C            MM     Tamaño de los Vectores                      C
C  SALIDA  : FREQ   Vector de Frecuencias Agrupadas             C
C            PPROB  Vector de Probabilidades Agrupadas          C
C            MMM    Tamaño de los nuevos Vectores              C
C.....C
SUBROUTINE INVAR(NFREQ,PROB,MM,FREQ,PPROB,MMM)
DIMENSION NFREQ(50),PROB(50),FREQ(50),PPROB(50)
INTEGER FREQ
FREQ(1)=NFREQ(1)+NFREQ(2)
PPROB(1)=PROB(1)+PROB(2)
DO 10 I=2,MM
FREQ(I)=NFREQ(I+1)
PPROB(I)=PROB(I+1)
10  CONTINUE
MMM=MM-1
RETURN
END

```

```

C..... SUBROUTINA NKL .....C
C OBJETIVO : C
C Calcula las estadísticas "KL" y "KLA". C
C ENTRADAS : PROB Vector de probabilidades C
C MM Dimension del vector NVAL C
C NFREQ Vector de frecuencias C
C N Tamaño de muestra C
C SUM Estadística "S" C
C SALIDA : KL Estadística KULLBACK-LIEBLER C
C KLA Estadística KULLBACK-LIEBLER con LOG(S) C
C.....C

```

```

SUBROUTINE NKL(K,FREQ,PPROB,MMM,N,SUM,KL,KLA)

```

```

DIMENSION PPROB(50),FREQ(50)

```

```

INTEGER FREQ

```

```

REAL KL,KLA,KL1

```

```

SA1=0.0

```

```

KL1=0.0

```

```

DO 50 J = 1, MMM

```

```

ARG = FLOAT( FREQ(J) )

```

```

A1 = ARG / FLOAT(N)

```

```

SA1 = SA1 + A1

```

```

A2 = PPROB(J) / SUM

```

```

COCIEN = ALOG(A1/A2)

```

```

A3 = A1 * COCIEN

```

```

KL1 = KL1 + A3

```

```

50 CONTINUE

```

```

SLOG = ALOG(SUM)

```

```

KL = 2.0 * FLOAT(N) * KL1

```

```

KLA = KL - 2.0 * FLOAT(N) * SLOG

```

```

RETURN

```

```

END

```

```

C..... SUBROUTINA DECISION .....C

```

```

C OBJETIVO : C

```

```

C Proporcionar el criterio para rechazar la Hipótesis Nula C

```

```

C comparando el valor de las Estadísticas "KL", "S", "C", "KLA" C

```

```

C con los respectivos cuantiles en tablas. Con 90 % de conf. C

```

```

C ENTRADAS : SUMSTAR Estadística "S" Estandarizada. C

```

```

C KL Estadística Kullback-Liebler C

```

```

C KLA Estadística Kullback-Liebler con Log(S) C

```

```

C MM Dimension del vector NVAL C

```

```

C SALIDA : IT1 Número de rechazos según "KL" C

```

```

C IT2 Número de rechazos según "S" Estandarizada C

```

```

C IT3 Número de rechazos según "C" C

```

```

C IT4 Número de rechazos según "KLA" C

```

```

C.....C

```

```

SUBROUTINE DECISION(K,SUMSTAR,MMM,KL,KLA,IT1,IT2,IT3,IT4)

```

```

DIMENSION REJ1(30)

```

```

INTEGER GL

```

```

REAL KL,KLA

```

```

DATA REJ1/2.706,4.605,6.251,7.779,9.236,10.64,12.02,13.36,

```

```

114.68,15.99,17.28,18.55,19.81,21.06,22.31,23.54,24.77,

```

```

125.99,27.20,28.41,29.62,30.81,32.01,33.20,34.38,35.56,

```

```

136.74,37.92,39.09,40.26/

```

```

REJ2=-1.281

```

```

IT1 = 0

```

```

IT2 = 0

```

```

IT3 = 0

```

```

IT4 = 0

```

ESTA TESTS NO DEBE  
 SER DE SU RESPONSABILIDAD

```

C  TRBAJAREMOS CON ALPHA = 0.1*****
IF ( SUMSTAR .LT. REJ2 ) IT2 = IT2 + 1
IF ( MMM .EQ. 1 ) GO TO 10
GL = MMM - 1
IF ( KL .GT. REJ1(GL))IT1 = IT1 + 1
IF ( KLA .GT. REJ1(GL))IT4 = IT4 + 1
IF ( KL .GT. REJ1(GL) .OR. SUMSTAR .LT. REJ2 ) IT3 = IT3 + 1
GO TO 11
10 IF ( SUMSTAR .LT. REJ2 ) IT3 = IT3 + 1
11 CONTINUE
RETURN
END

```

```

C*****          SUBROUTINA CHOICE *****C
C OBJETIVO :                               C
C Proporcionar el criterio para rechazar la Hipótesis Nula C
C comparando el valor de las Estadísticas "KL", "S", "C", "KLA" C
C con los respectivos cuantiles en tablas. Con 95 % de conf. C
C Cuenta el número de veces que se rechaza la Hip. Nula C
C ENTRADAS : SUMSTAR Estadística "S" Estandarizada. C
C           KL Estadística Kullback-Liebler C
C           KLA Estadística Kullback-Liebler con Log(S) C
C           MM Dimensión del vector NVAL C
C SALIDA : IT1 Número de rechazos según "KL" C
C           IT2 Número de rechazos según "S" Estandarizada C
C           IT3 Número de rechazos según "C" C
C           IT4 Número de rechazos según "KLA" C
C*****          *****C

```

```

SUBROUTINE CHOICE(SUMSTAR,MMM,KL,KLA,IT1,IT2,IT3,IT4)
DIMENSION REJ1(30)
INTEGER GL
REAL KL,KLA

```

```

C  TRBAJAREMOS CON ALPHA = 0.05 *****
DATA REJ1/3.841,5.991,7.815,9.488,11.07,12.59,14.07,15.51,
116.92,18.31,19.68,21.03,22.36,23.68,25.00,26.30,27.59,
128.87,30.14,31.41,32.67,33.92,35.17,36.42,37.65,38.89,
140.11,41.34,42.56,43.77/
REJ2=-1.645
IT1 = 0
IT2 = 0
IT3 = 0
IT4 = 0
IF ( SUMSTAR .LT. REJ2 ) IT2 = IT2 + 1
IF ( MMM .EQ. 1 ) GO TO 10
GL = MMM - 1
IF ( KL .GT. REJ1(GL))IT1 = IT1 + 1
IF ( KLA .GT. REJ1(GL))IT4 = IT4 + 1
IF ( KL .GT. REJ1(GL) .OR. SUMSTAR .LT. REJ2 ) IT3 = IT3 + 1
GO TO 11
10 IF ( SUMSTAR .LT. REJ2 ) IT3 = IT3 + 1
11 CONTINUE
RETURN
END

```

```

C***** SUBROUTINA BIN *****C
C OBJETIVO : C
C Generador de variables aleatorias BINOMIAL NEGATIVA con C
C Par metro (NR,P) C
C ENTRADAS: IX,IY,IZ Semillas Iniciales C
C N Tamaño de muestra C
C NR,P Parámetros de la Distribución C
C SALIDA: NB Vector de variables BINOMIAL NEGATIVA C
C NOTA: C
C ARREGLOS DE TRABAJO: WK VECTOR DE DIMENSION N C
C SUBROUTINAS QUE USA: DUN(IX,IY,IZ,U) GENERADOR DE V.A. C
C UNIFORMES EN EL (0,1) C
C*****C

```

```

SUBROUTINE BIN(IX,IY,IZ,N,NR,P,NB)

```

```

DIMENSION NB(50),WK(50)

```

```

DATA EPS/0.000001/

```

```

PI = 1.0-P

```

```

RI = P**NR

```

```

DO 1 I=1,N

```

```

CALL DUN(IX,IY,IZ,U)

```

```

1 WK(I)=U

```

```

DO 15 I = 1,N

```

```

NB(I) = NR

```

```

R = RI

```

```

T = WK(I)

```

```

IF (T .LE. R) GO TO 15

```

```

SUM = R

```

```

XI = FLOAT(NR)

```

```

XD = 1.0

```

```

KK = NR

```

```

5 KK = KK+1

```

```

R = XI*PI*R/XD

```

```

IF (R .LT. EPS*SUM) GO TO 10

```

```

SUM = SUM+R

```

```

XI = XI+1.0

```

```

XD = XD+1.0

```

```

IF (T .GT. SUM) GO TO 5

```

```

NB(I) = KK

```

```

GO TO 15

```

```

10 CALL DUN(IX,IY,IZ,U)

```

```

WK(I)=U

```

```

NB(I) = KK+1+9*WK(I)

```

```

15 NB(I) = NB(I) - NR

```

```

RETURN

```

```

END

```

```

C***** SUBROUTINA DUN *****C
C OBJETIVO : C
C Generador de Variables Aleatorias Uniformes en el intervalo C
C (0,1), usando el algoritmo AS 183 de Applied Statistics(1982) C
C vol 31. C
C C
C ENTRADA: IX,IY,IZ Semillas Iniciales C
C C
C SALIDA: U Número Aleatorio en (0,1) C
C*****C
SUBROUTINE DUN(IX,IY,IZ,U)
IX=171*MOD(IX,177)-2*(IX/177)
IY=172*MOD(IY,176)-3*(IY/176)
IZ=170*MOD(IZ,178)-63*(IZ/178)
IF(IX.LE.0)IX=IX+30269
IF(IY.LE.0)IY=IY+30307
IF(IZ.LE.0)IZ=IZ+30323
U=AMOD(FLOAT(IX)/30269.0+FLOAT(IY)/30307.0+FLOAT(IZ)
*/30323.0,FLOAT(1))
IF(U.GE.1.0)U=0.999999
IF(U.LE.0.0)U=0.000001
RETURN
END
C

```

**BIBLIOGRAFIA**

---

**AKAIKE, H. (1983)**

Statistical Inference and Measurement of Entropy in Scientific Inference, Data Analysis and Robustes.

Academic Press Inc.

**DEVROYE, L. (1986)**

Non-uniform Random Variate Generation.

Springer-Verlang, Inc., New York.

**D'AGOSTINO, R & STEPHENS (1986)**

Goodness of Fit Techniques

M. Dekker, New York

**KHINCHIN, A.I. (1957)**

Information Theory.

Dover Publications, Inc., New York.

**KOCHERLAKOTA, S. & KOCHERLAKOTA, K. (1986)**

Goodness of Fit for Discrete Distributions

Comm. Statist. Theo. and Meth., 15, 815-829

**MOOD A.M. GRAYBILL F.A. & BOES D.C.(1974)**

Introduction to the Theory of Statistics 3a. Ed.

Mc Graw Hill, New York

**O'REILLY, F.J. (1991)**

Goodness of Fit for Discrete Distributions Based on Likelihood Ratio

Contribución Sesión 48 del ISI

**PETIT, A. N. & STEPHENS, M. A. (1977)**

The Kolmogorov Smirnov Goodness of Fit Statistic with Discrete and Agruped Data

Technometrics 19, 205-210

**RUEDA, R., PEREZ-ABREU, V. & O'REILLY, F. J. (1991)**

Goodness of Fit for the Poisson Distribution Based on the Probability Generating Function.

Comm. Statist. Theo. and Meth. 20(10), pp 3093-3110.