03061

2

26

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

UNIDAD ACADEMICA DE LOS CICLOS PROFESIONAL Y DE POSGRADO DEL COLEGIO DE CIENCIAS Y HUMANIDADES

LA ESTADISTICA ANDERSON - DARLING PARA MUESTRAS CON CENSURA ALEATORIA

T E S I S

QUE PARA OBTENER EL GRADO DE:

MAESTRO EN ESTADISTICA E INVESTIGACION DE OPERACIONES

P R E S E N T A :

E L A C T U A R I O

JOSE SALVADOR ZAMORA MUÑOZ

MEXICO, D. F. JUNIO 1996

TESIS CON
FALLA DE ORIGEN

AGRADECIMIENTOS

Primeramente, quiero agradecer de manera muy especial a mi directora de tesis, la Dra. Belem Trejo por su paciente y entusiasta dirección. De igual manera, agradezco a todos los profesores de la Maestría por compartir sus conocimientos en las distintas áreas de la Estadística.

Mi más profundo agradecimiento al Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS), por todo el apoyo y todas las facilidades que me otorgaron durante los estudios de Maestría y la realización de esta tesis. Asimismo, agradezco a la UNAM por todo lo que he recibido de ella, y que acrecienta mi deuda con con esta institución.

Quiero agradecer de manera particular al pueblo trabajador de México, ya que gracias a su participación existe la escuela pública que es para muchos de nosotros la única opción.

Agradezco también a Gloria sin cuya presencia esta tesis se hubiera concluido hace mucho tiempo.

Finalmente, un agradecimiento a todos los que me faltaron.

GRACIAS.

INDICE

página

INTRODUCCION i

1. EL PROBLEMA DE BONDAD DE AJUSTE 1

1.1 INTRODUCCION 1

1.2 LA FUNCION DE DISTRIBUCION EMPIRICA (F.D.E.) 2

1.2.1 DEFINICION Y PROPIEDADES 2

1.2.2 LAS ESTADISTICAS DEL SUPREMO 4

1.2.3 LAS ESTADISTICAS CUADRATICAS 4

1.2.4 TRANSFORMACION INTEGRAL DE PROBABILIDAD Y ESTADISTICAS CUADRATICAS 5

1.3 DISTRIBUCION DE LAS ESTADISTICAS DEL TIPO CRAMER-VON MISES 7

1.3.1 DESCOMPOSICION DE PROCESOS ESTOCASTICOS 7

1.3.2 DESCOMPOSICION DE W^2 11

1.3.3 DESCOMPOSICION DE U^2 14

1.3.4 DESCOMPOSICION DE A^2 18

1.4 PRUEBAS DE POTENCIA 22

2. ESTADISTICAS CUADRATICAS CON MUESTRAS CENSURADAS 24

2.1 INTRODUCCION 24

2.2 DISTRIBUCION ASINTOTICA DE LAS ESTADISTICAS 27

2.3 PRUEBAS DE POTENCIA 32

2.4 MUESTRAS CON CENSURA ALEATORIA 32

2.4.1 DISTRIBUCION ASINTOTICA DE ${}_cW_n^2$ 35

2.4.2 PRUEBAS DE POTENCIA 39

3. **LA ESTADISTICA ANDERSON-DARLING PARA MUESTRAS CON CENSURA ALEATORIA** 41

3.1 INTRODUCCION 41

3.2 COMPORTAMIENTO DE A_m^{*2} 45

3.3 SIMULACION DE LA ESTADISTICA A_{cn}^2 50

3.4 PRUEBAS DE POTENCIA 59

4. **CONCLUSIONES** 70

5. **ANEXOS**

ANEXO I 73

ANEXO II 75

BIBLIOGRAFIA 78

INTRODUCCION.

El uso adecuado de las técnicas de análisis de datos así como las inferencias a partir de ellos, se sustentan en la distribución subyacente a los datos, es por ésto que uno de los problemas más importantes en la estadística lo constituye el poder asignar una distribución apropiada a los datos extraidos de una población a través de una muestra.

Los métodos que permiten decidir si la asignación de una distribución a los datos es adecuada se agrupan en el área estadística llamada Bondad de Ajuste. De las propuestas que hay en Bondad de Ajuste, reciben una atención especial las basadas en la función de distribución empírica y, en particular, las correspondientes a las estadísticas del tipo Cramér-von Mises.

Estos métodos tienen un desarrollo limitado en el área de Análisis de Supervivencia, además de que la difusión de las propuestas relacionadas con las estadísticas del tipo Cramér-von Mises y en general de Bondad de Ajuste ya existentes es escasa. Es por lo anterior que en este trabajo se presenta un panorama general del desarrollo, primeramente, de algunos métodos de Bondad de Ajuste de uso común en la estadística para, en seguida, dar una visión de los métodos que se han desarrollado en Análisis de Supervivencia y finalmente hacer una propuesta para realizar una prueba de Bondad de Ajuste en esta área usando una aproximación a la estadística Anderson-Darling.

Para lograr dicho objetivo el trabajo se divide en cuatro capítulos de la siguiente manera:

En el primer capítulo se hace la introducción del problema de Bondad de Ajuste y se presentan algunos desarrollos de las estadísticas del tipo Cramér-von Mises. En el siguiente capítulo se introduce el tipo de estudios característicos del Análisis de Supervivencia, además de algunas de las estadísticas que se han desarrollado para el trabajo con muestras que presentan el tipo de censura catalogada como censura controlada por el investigador y se presenta la propuesta que para trabajar con una muestra con censura aleatoria desarrollan Koziol y Green, quienes introducen la estadística Cramér-von Mises bajo un modelo particular para la distribución de la censura. En el tercer capítulo se introduce la propuesta para trabajar con una muestra con censura aleatoria utilizando una aproximación a la estadística de Anderson Darling. Esta propuesta se basa en considerar el mismo modelo que proponen Koziol y Green para la censura. Además se analiza el comportamiento general de esta estadística, se encuentra su distribución asintótica y se realizan algunas pruebas para determinar la potencia de la misma. Finalmente, en el último capítulo se hacen las conclusiones del trabajo y se presenta un panorama de posibilidades de investigación relacionadas con este tema.

CAPITULO I

EL PROBLEMA DE BONDAD DE AJUSTE

1.1 INTRODUCCION.

Una de las preguntas más comunes que uno puede hacerse en cualquier área de trabajo donde los métodos estadísticos tengan un uso habitual, consiste en saber qué tan seguros estamos de asociar alguna distribución a los datos extraídos de una población. Esta pregunta no es, en lo absoluto, ociosa, ya que muchos de los modelos estadísticos requieren, por lo general, algunos supuestos distribucionales sobre los datos, con la finalidad de que el uso de un determinado modelo sea el más adecuado. Además, violaciones sobre estos supuestos en un modelo, pueden conducir a conclusiones erróneas y por lo tanto a toma de decisiones incorrectas.

El estudio y desarrollo de los métodos estadísticos para verificar los supuestos distribucionales recibe el nombre de Bondad de Ajuste y, por supuesto, cuando se requiere comprobar que un conjunto de datos pertenecen a una distribución particular, se está ante un problema de Bondad de Ajuste.

El planteamiento exacto de este problema es el siguiente. Se tiene una muestra aleatoria $X_1,\ldots,X_n$, de una población con función de distribución $F(x)$, con $F \in \mathbb{F}$ una familia de distribuciones. Nos interesa probar la hipótesis

$$H_0: F \in \mathbb{F}_0,$$

con $\mathbb{F}_0$ un subconjunto de $\mathbb{F}$ indexado paramétricamente, es decir,

$$\mathbb{F}_0 = \left\{ F \in \mathbb{F} \,|\, F(x) = F(x;\theta) : \theta \in \Theta \subset \mathbb{R}^k \right\}.$$

Un enfoque que es usado para atacar este problema consiste en el uso de técnicas gráficas, que pueden usarse además como técnicas exploratorias para descubrir características de la distribución subyacente a los datos, tales como simetría, colas pesadas, etc.. Otro enfoque, que puede usarse conjuntamente con el anterior, lo constituye el uso de técnicas numéricas, cuyo objetivo es probar formalmente alguna hipótesis preconcebida sobre la distribución de los datos, misma que puede ser sugerida por los métodos gráficos. Dentro de las múltiples propuestas que hay en técnicas numéricas, son de especial interés aquellas basadas en la función de distribución empírica (F.D.E.).

Las estadísticas basadas en la F.D.E. (en adelante estadísticas F.D.E.) son medidas de la discrepancia entre la F.D.E. y alguna distribución dada como hipótesis. Dichas estadísticas se usan para probar el ajuste de la distribución hipotética (que puede estar completamente especificada o contener algunos parámetros desconocidos) a la muestra.

1.2 LA FUNCION DE DISTRIBUCION EMPIRICA(F.D.E.)

1.2.1 Definición y propiedades

Supongamos que tenemos $X_1, X_2, \ldots, X_n$, una muestra aleatoria de variables aleatorias independientes e idénticamente distribuidas. Sean $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ las correspondientes estadísticas de orden; supongamos además que la función de distribución de X está dada por $F(x)$. La función de distribución empírica $F_n(x)$ se define como :

$$F_n(x) = \frac{\text{número de observaciones} \leq x}{n} \qquad -\infty < x < \infty,$$

de manera más precisa , si no hay empates en la muestra, se tiene

$$F_n(x) = \begin{cases} 0 & \text{si} \quad -\infty < x < X_{(1)} \\ \frac{i}{n} & \text{si} \quad X_{(i)} \leq x < X_{(i+1)} \qquad i=1,\ldots,n-1 \\ 1 & \text{si} \quad X_{(n)} \leq x < \infty . \end{cases}$$

Esto es, $F_n(x)$ es una función escalonada no decreciente, calculada de la muestra, que estima a $F(x)$. De hecho es un estimador consistente de $F(x)$, es decir

$$\sup_x |F_n(x)-F(x)| \xrightarrow[c.s.]{} 0 \quad \text{cuando } n \longrightarrow \infty$$

El uso de la F.D.E. ofrece algunas ventajas sobre otros procesos estadísticos, como:

1.- Es invariante bajo transformaciones monótonas crecientes de los datos.

2.- Proporciona información sobre la forma de la distribución subyaciente. (e.g.simetría).

3.- No tiene las dificultades por agrupamiento de los datos que surgen, por ejemplo,en un histograma.

Como medidas de la diferencia entre $F_n(x)$ y $F(x)$ las estadísticas F.D.E. se pueden dividir en dos clases, las correspondientes a las estadísticas del supremo y las estadísticas cuadráticas.

1.2.2 LAS ESTADISTICAS DEL SUPREMO.

Esta clase está compuesta principalmente por las estadísticas

$$D^{+} = \sup_{x}\left\{F_n(x)-F(x)\right\} ,$$

$$D^{-} = \sup_{x}\left\{F(x)-F_n(x)\right\}$$

y

$$D = \sup_{x}|F_n(x)-F(x)| = \text{máx}(D^{+},D^{-}),$$

D^{+} es la diferencia vertical más grande entre $F_n(x)$ y $F(x)$ cuando $F_n(x)$ es mayor que $F(x)$. De manera semejante, D^{-} es la diferencia vertical más grande entre $F_n(x)$ y $F(x)$ cuando $F(x)$ sea mayor que $F_n(x)$. Finalmente D, que es el máximo entre las dos estadísticas anteriores, es la estadística de esta clase más conocida, introducida por Kolmogorov(1933).

1.2.3 LAS ESTADISTICAS CUADRATICAS.

Esta segunda clase la constituyen las estadísticas del tipo Cramér-von Mises, a saber:

$$Q = n\int_{-\infty}^{\infty} \left\{F_n(x)-F(x)\right\}^2 \psi(x)dF(x),$$

donde $\psi(x)$ es una función que asigna pesos a las diferencias cuadráticas $\left\{F_n(x)-F(x)\right\}^2$.

Cuando $\psi(x)=1$ tenemos la estadística Cramér-von Mises, denotada por W^2. Cuando $\psi(x)=\left\{F(x)(1-F(x))\right\}^{-1}$ tenemos la estadística Anderson-Darling, que se denota por A^2. Una modificación a W^2, es

la estadística U^2 de Watson definida como:

$$U^2 = n\int_{-\infty}^{\infty}\left\{ F_n(x)-F(x) - \int_{-\infty}^{\infty}(F_n(u)-F(u))dF(u) \right\}^2 dF(x)$$

De estas dos clases, nuestro interés se centrará en las estadísticas cuadráticas.

1.2.4 TRANSFORMADA INTEGRAL DE PROBABILIDAD Y ESTADISTICAS CUADRATICAS.

Para estas dos clases de estadísticas, es posible encontrar procesos de cálculo que sean fáciles de implementar, dichos procesos están estrechamente relacionados al uso de la transformada integral de probabilidad.

Consideremos X una variable aleatoria continua con función de distribución F(x), definimos la variable aleatoria T=F(X), a esta transformación se le conoce como la transformada integral de probabilidad (T.I.P). Esta nueva variable aleatoria se distribuye uniforme en (0,1). Es decir

$$P(T\le t) = \begin{cases} 0 & t \le 0 \\ t & t \in (0,1) \\ 1 & t \ge 1 \end{cases} .$$

Ahora mostraremos el uso de esta transformación en el cálculo de las estadísticas del tipo Cramér-von Mises. Supongamos que tenemos una muestra aleatoria $X_1,\ldots,X_n$ con función de distribución F(x). Usando la T.I.P., las estadísticas se calculan de la siguiente manera

Cramér-von Mises por

$$W_n^2 = n\int_{-\infty}^{\infty} \left\{F_n(x)-F(x)\right\}^2 dF(x) = n\sum_{i=1}^{n} \int_{z_{(i-1)}}^{z_{(i)}} \left\{ G_n(z)-z \right\}^2 dz$$

con $G_n(\cdot)$ la función de distribución empírica de una muestra $Z_{(1)},\ldots,Z_{(n)}$ de una uniforme (0,1) y $Z_{(0)}=0$. De aqui obtenemos

$$W_n^2 = \sum_{i=1}^{n} \left\{z_{(i)}-\frac{(2i-1)}{2n}\right\} + \frac{1}{12n} .$$

De manera semejante para la Watson, tenemos que

$$U_n^2 = n\int_{-\infty}^{\infty} \left\{F_n(x)-F(x) -\int_{-\infty}^{\infty}(F_n(u)-F(u))dF(u) \right\}^2 dF(x),$$

se convierte en

$$U_n^2 = W_n^2 - (\bar{z} - 0.5)^2 \quad \text{con } \bar{z} = \sum_{i=1}^{n} \frac{z_{(i)}}{n}.$$

Finalmente, la Anderson-Darling que está dada por

$$A_n^2 = n\int_{-\infty}^{\infty} \frac{\left\{F_n(x)-F(x)\right\}^2}{F(x)(1-F(x))} dF(x),$$

se reduce a

$$A_n^2= -n - \frac{1}{n} \left\{ \sum_{i=1}^{n}(2i-1)\log z_{(i)}+(2n-2i+1)\log(1-z_{(i)}) \right\}$$

En cada caso $Z_{(i)} = F(X_{(i)})$.

Dadas estas expresiones para las estadísticas cuadráticas, podemos realizar la prueba de hipótesis sobre si la muestra aleatoria $X_1,\ldots,X_n$ provieve de una población con función de distribución $F_0(x;\theta)$. En cualquier caso se rechaza la hipótesis nula si el valor de la estadística de prueba es mayor que el valor de la tabla correspondiente, al nivel de significancia α.

En general, pueden considerarse dos casos para realizar esta prueba. En el primero de ellos, llamado caso simple de Bondad de Ajuste, $F_0(x;\theta)$ está totalmente especificada, es decir, se conoce la familia a la que pertenece además de los valores de su(s) parámetro(s). En el segundo caso, se conoce la familia a la que pertenece F_0 pero θ se desconoce total o parcialmente. En este caso se pueden estimar, generalmente por máxima verosimilitud, los parámetros desconocidos y utilizar las expresiones de cálculo anteriores tomando $Z_{(i)}=F(X_{(i)};\hat{\theta})$ con $\hat{\theta}$ el estimador de θ. Cuando los parámetros se desconocen la distribución de estas estadísticas puede depender de la distribución a probar e incluso de los verdaderos parámetros (D'Agostino-Stephens(1986)).

1.3 DISTRIBUCIONES DE LAS ESTADISTICAS DEL TIPO CRAMER-VON MISES.

1.3.1 Descomposición de procesos estocásticos.

Dado que ya tenemos una expresión de cálculo para las estadísticas del tipo Cramér-von Mises para pruebas de Bondad de Ajuste, el siguiente paso es encontrar la distribución de dichas estadísticas. Para lograr este objetivo y debido a la relación que guardan estas estadísticas con los procesos estocásticos, primeramente introduciremos algunas ideas generales de procesos estocásticos, y en particular aquéllas relacionadas con los procesos Gaussianos.

En la teoría de procesos estocásticos, los procesos Gaussianos tienen un papel predominante ya que muchos de los procesos de

interés pueden aproximarse por un proceso Gaussiano, inclusive algunos de los procesos más conocidos (e.g. Movimiento Browniano) son casos especiales de procesos Gaussianos.

Se sabe que el conocimiento de la función del valor medio y la función de covarianza de un proceso Gaussiano determina de manera única su distribución.

Es un hecho conocido dentro de la literatura de procesos estocásticos, que el proceso empírico

$$Z_n(t)=\sqrt{n}\left\{G_n(t)-t\right\} \qquad 0\le t\le 1.$$

con $G_n(\cdot)$ como se definió anteriormente, converge débilmente a un proceso Gaussiano $Z(t)$ de media cero y función de covarianza

$$K(s,t)=\min(s,t)-st \qquad 0\le s,t\le 1$$

y cuando $n\to\infty$, se tienen los siguientes resultados de convergencia

$$W_n^2 = \int_0^1 Z_n^2(t)\,dt \longrightarrow_d W^2 = \int_0^1 Z^2(t)\,dt$$

$$U_n^2 = \int_0^1 \left\{ Z_n(t)-\int_0^1 Z_n(u)\,du \right\}^2 dt \longrightarrow_d U^2 = \int_0^1 \left\{ Z(t)-\int_0^1 Z(u)\,du \right\}^2 dt$$

$$A_n^2 = \int_0^1 \frac{Z_n^2(t)}{t(1-t)}\,dt \longrightarrow_d A^2 = \int_0^1 \frac{Z^2(t)}{t(1-t)}\,dt.$$

Utilizaremos la convergencia del proceso empírico a un proceso Gaussiano y los resultados anteriores para encontrar la distribución de las estadísticas del tipo Cramér-von Mises.

Los pasos que se siguen para encontrar la distribución de las

estadísticas del tipo Cramér-von Mises están relacionados con la descomposición en componentes principales de un proceso estocástico. A su vez, esta descomposición es similar a la que se hace en estadística de un vector aleatorio en sus correspondientes componentes principales. De esta manera, los pasos que resultan familiares en la descomposición de un vector aleatorio; componentes principales del vector, descomposición de la matriz de varianza-covarianza y la distribución de la suma de cuadrados, resultarán análogos a los propios de la descomposición de un proceso estocástico; componentes principales del proceso, descomposición de la función de covarianza y la distribución de la integral del cuadrado del proceso.

La descomposición en componentes principales de un proceso estocástico de media cero $\{X(t);\ 0\le t\le 1\}$ se puede realizar siguiendo esencialmente los pasos descritos en el anexo I para la descomposición de un vector aleatorio.

Consideremos un proceso estocástico $X(t)$, de media cero y función de covarianza $K(s,t)$. Las componentes principales de $X(t)$ son

$$Z_j = \int_0^1 X(t)f_j(t)dt \qquad j=1,2,\ldots,$$

con $f_j(t)$ la correspondiente eigenfunción (ortonormal) asociada al eigenvalor λ_j, que se obtiene al resolver la ecuación integral

$$\lambda f(t) = \int_0^1 K(s,t)f(s)ds.$$

También tenemos la descomposición de la función de covarianza

$$K(s,t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \qquad 0\le s,t\le 1.$$

La descomposición de $X(t)$ en sus componentes principales es

$$X(t) = \sum_{j=1}^{\infty} Z_j f_j,$$

o bien

$$X(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\, Z_j^* f_j \quad \text{con } Z_j^* = Z_j/\sqrt{\lambda_j},$$

con lo cual obtenemos

$$\int_0^1 X^2(t)dt = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2}. \tag{1}$$

Si $X(t)$ es un proceso Gaussiano, entonces (1) se distribuye como una suma ponderada infinita de variables aleatorias χ^2 independientes con 1 g.l.; las ponderaciones λ_j son los eigenvalores de la función de covarianza y las Z_j son las proyecciones de X hacia las eigenfunciones asociadas. Como puede observarse los dos procesos de descomposición, el de un vector aleatorio y el de un proceso estocástico, son totalmente análogos.

La descomposición anterior se realizó para un proceso arbitrario X de media cero. Ahora realizaremos dicha descomposición para los procesos relacionados con las estadísticas del tipo Cramér-von Mises.

Mostraremos como se realiza la descomposición correspondiente para W^2, U^2 y A^2 y daremos una descomposición similar de sus correspondientes procesos empíricos W_n^2, U_n^2 y A_n^2.

1.3.2 DESCOMPOSICION DE W^2.

La expresión para esta estadística es

$$W^2 = \int_0^1 Z^2(t)dt,$$

con $Z(t)$ un proceso Gaussiano de media cero y función de covarianza

$$K(s,t) = \min(s,t)-st \quad 0\le t,s\le 1.$$

Para realizar la descomposición, debemos resolver la ecuación integral

$$\lambda f(t) = \int_0^1 f(s)(\min(s,t)-st)ds$$

$$= \int_0^t f(s)s(1-t)ds + \int_t^1 f(s)t(1-s)ds,$$

derivando con respecto a t, tenemos

$$\lambda f'(t) = \int_t^1 f(s)ds - \int_0^1 sf(s)ds,$$

de donde obtenemos la siguiente ecuación diferencial

$$\lambda f''(t) + f(t) = 0 \qquad (2)$$

con las condiciones iniciales $f(0) = f(1) = 0$.

Como las f's son ortonormales, se cumple

$$\int_0^1 f^2(t)dt = 1.$$

La solución para (2) se puede encontrar por métodos usuales de ecuaciones diferenciales lineales. En este caso el polinomio asociado es $\lambda y^2 + 1 = 0$, con soluciones

$$y_1 = \lambda^{-1/2} i \ , \ y_2 = - \lambda^{-1/2} i,$$

entonces, la solución general de la ecuación diferencial es

$$f(t) = a \cos(\lambda^{-1/2} t) + b \operatorname{sen}(\lambda^{-1/2} t) \qquad 0 \le t \le 1,$$

usando las condiciones iniciales y de ortonormalidad tenemos que

$$a = 0, \ b = \sqrt{2} \qquad \lambda_j = \frac{1}{(j\pi)^2} \qquad j = 1, 2, \ldots ,$$

finalmente obtenemos

$$f_j(t) = \sqrt{2} \operatorname{sen}(j\pi t) \qquad 0 \le t \le 1 \qquad \text{con } \lambda_j = \frac{1}{(j\pi)^2}.$$

Con ésto tenemos que la descomposición del proceso $Z(t)$ queda de la siguiente forma

$$Z = \sum_{j=1}^{\infty} \frac{Z_j^{\bullet} f_j}{j\pi},$$

con $Z_j^{\bullet} = \lambda^{-1/2} Z_j$ y $f_j = \sqrt{2} \operatorname{sen}(j\pi t)$, y además $Z_1^{\bullet}, Z_2^{\bullet}, \ldots$ son v.a.i.i.d. $N(0,1)$, por que $Z(t)$ es un proceso Gaussiano.

Finalmente

$$W^2 = \int_0^1 Z^2(t)\,dt = \sum_{j=1}^{\infty} \frac{Z_j^{*2}}{(j\pi)^2}.$$

La descomposición anterior está basada en el teorema de Kac-Siegert (ver anexo II).

Para la descomposición del correspondiente proceso empírico W_n^2 usaremos el teorema de Durbin-Knott(ver anexo II). Los pasos correspondientes a esta descomposición son análogos a los del proceso teórico. Ahora tenemos

$$W_n^2 = \int_0^1 Z_n^2(t)\,dt$$

con $Z_n(t)=\sqrt{n}\left\{G_n(t)-t\right\}$. Las componentes principales(normalizadas) de W_n^2 son

$$Z_{nj}^* = \lambda^{-1/2}\int_0^1 f_j(t)Z_n(t)\,dt = n^{-1/2}\sum_{k=1}^{n} \sqrt{2}\,\cos(j\pi\xi_k),$$

y la correspondiente descomposición del proceso empírico es

$$Z_n(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j}Z_{nj}^* f_j = \sum_{j=1}^{\infty} \frac{Z_{nj}^* f_j}{j\pi}$$

y por lo tanto

$$W_n^2 = \int_0^1 Z_n^2(t)\,dt = \sum_{j=1}^{\infty} \frac{Z_{nj}^{*2}}{(j\pi)^2}.$$

Finalmente tenemos lo siguiente

$$W_n^2 \longrightarrow_d W^2 = \sum_{j=1}^{\infty} \frac{Z_j^{*2}}{(j\pi)^2} .$$

1.3.3 DESCOMPOSICION DE U^2.

Las pruebas de Bondad de Ajuste se refieren, en general, a casos donde el espacio muestral es la recta real. Sin embargo, un problema especial surge cuando las observaciones de la variable aleatoria X se miden sobre puntos en una circunferencia. En este caso la circunferencia reemplaza a la recta como región de integración. Un problema importante que surge aquí es que el valor de $F(x;\theta)$ varía con la elección del punto de origen, y las estadísticas W^2 y A^2 toman diferentes valores para cada elección de dicho origen. Este problema se supera cuando se considera la estadística U^2 que es invariante ante distintos puntos de origen.

La expresión para esta estadística es

$$U^2 = \int_0^1 \left\{ Z(t) - \int_0^1 Z(u)du \right\}^2 dt.$$

Si definimos

$$Y(t) = Z(t) - \int_0^1 Z(t)dt \qquad 0 \le t \le 1$$

entonces

$$U^2 = \int_0^1 Y^2(t)dt$$

Para obtener la función de covarianza obsérvese que

$$E\left\{Z(t)\int_0^1 Z(u)du\right\} = \int_0^1 E(Z(t)Z(u))du = \int_0^t u(1-t)du + \int_t^1 t(1-u)du$$

$$= t/2 - t^2/2.$$

De forma análoga se obtiene

$$E\left\{Z(u)\int_0^1 Z(t)dt\right\} = u/2 - u^2/2$$

y además

$$E\left\{\int_0^1 Z(t)dt\int_0^1 Z(u)du\right\} = \int_0^1\int_0^1 E(Z(t)Z(u))dtdu$$

$$= \int_0^1\int_0^u t(1-u)dtdu + \int_0^1\int_0^t u(1-t)dudt = 1/12$$

de donde

$$\mathrm{cov}(Y(t),Y(s)) = E\left\{[Z(t) - \int_0^1 Z(t)dt]\ [\ Z(s) - \int_0^1 Z(s)ds]\right\}$$

$$= \min(s,t) - st - s/2 + s^2/2 - t/2 + t^2/2 + 1/12$$

$$= \min(s,t) - (s+t)/2 + (s-t)^2/2 + 1/12.$$

Obsérvese que $\mathrm{var}(Y(t))=1/12$, por lo que es conveniente tomar

$$W(t) = \sqrt{12}\ Y(t),$$

por lo tanto, para $-1\le\tau\le1$ se tiene

$$\mathrm{cov}(W(t),W(t+\tau)) = 1-6|\tau|(1-|\tau|) \qquad 0\le t\le 1 \tag{3}$$

y entonces

$$12U^2 = \int_0^1 W^2(t)dt,$$

con W(t) un proceso Gaussiano de media cero y función de covarianza dada por (3) que llamaremos $\rho(\tau)$.

La distribución asintótica de $12U^2$ se obtiene a través del método de Kac y Siegert(ver anexo II). Para ésto, necesitamos resolver la siguiente ecuación integral

$$\lambda f(t) = \int_0^1 \rho(t-s)f(s)ds \qquad 0 \le t \le 1,$$

podemos observar que

$$\lambda f(0) = \int_0^1 (-s/2+s^2/2+1/12)f(s)ds = \lambda f(1),$$

de donde tenemos la condición $f(0)=f(1)$.

Por un procedimiento semejante al anterior, tenemos que la ecuación integral se transforma en la siguiente ecuación diferencial

$$\lambda f''(t) + 12f(t) = 0$$

cuyas soluciones son

$$f_{2j}(t) = \sqrt{2}\cos(j\pi t)$$

$$f_{2j-1}(t) = \sqrt{2}\,\mathrm{sen}(j\pi t)$$

con $\lambda_{2j} = \lambda_{2j-1} = 3/j^2\pi^2 \quad j = 1,2,\ldots$.

Finalmente, la distribución asintótica de esta estadística es

$$12U^2 \sim \sum_{j=1}^{\infty} \frac{3}{(j\pi)^2} (A_j^2 + B_j^2)$$

donde A_j y B_j son sucesiones de v.a.i. normales estándar, entonces

$$U^2 \sim \sum_{j=1}^{\infty} \frac{Z_j}{4(j\pi)^2} \quad \text{con } Z_j \text{ v.a.i.i.d. } \chi^2_{(2)} \text{ o equivalentemente}$$

$$U^2 \sim \sum_{j=1}^{\infty} \frac{E_j}{2(j\pi)^2} \quad \text{con } E_j \text{ v.a.i.i.d. Exp(1).}$$

Ahora, para descomponer U_n consideremos las siguientes componentes principales

$$Z_{nj} = \int_0^1 Z_n(t)\,\mathrm{sen}(j\pi t)\,dt = \sqrt{\frac{2}{n}} \sum_{k=1}^{n} \cos(j\pi\xi_k)$$

y

$$Z^*_{nj} = \int_0^1 Z_n(t)\cos(j\pi t)\,dt = -\sqrt{\frac{2}{n}} \sum_{k=1}^{n} \mathrm{sen}(j\pi\xi_k)$$

de aquí se concluye que

$$U_n^2 = \sum_{j=1}^{\infty} \frac{(Z_{nj}^2 + Z_{nj}^{*2})}{4(j\pi)^2},$$

y finalmente

$$U_n^2 \longrightarrow_d U^2 \sim \sum_{j=1}^{\infty} \frac{Z_j}{4(j\pi)^2}.$$

1.3.4 DESCOMPOSICION DE A^2.

En muchas ocasiones un estadístico puede decidir que la función de peso $\psi(t)=1$ no es conveniente en el problema que le ocupa. Por ejemplo, puede considerar que esta función no proporciona suficiente peso a las colas de la distribución. Al respecto, Anderson y Darling (1952), tomando en cuenta que la varianza de $Z_n(t)=\sqrt{n}\left\{G_n(t)-t\right\}$ es $t(1-t)$, proponen para $Z_n^2(t)$ la función de peso $\psi(t)=(t(1-t))^{-1}$ que es el recíproco de la varianza asintótica. Esta función asigna mayor peso a las colas de la distribución que el que se les da con la Cramér-von Mises.

La forma de la estadística de Anderson y Darling es la siguiente

$$A^2 = \int_0^1 \frac{Z^2(t)}{t(1-t)}\, dt = \int_0^1 Y^2(t)\, dt,$$

con $Y(t) = \dfrac{Z(t)}{\sqrt{t(1-t)}} \quad 0 \le t \le 1.$

La función de covarianza del proceso $Y(t)$ está dada por

$$K(s,t) = Cov(Y(t),Y(s)) = \frac{Cov(Z(t),Z(s))}{\sqrt{t(1-t)s(1-s)}} = \frac{\min(s,t)-st}{\sqrt{t(1-t)s(1-s)}} \quad 0 \le s,t \le 1.$$

La descomposición de esta función de covarianza vía el Teorema de Mercer(ver anexo II) no es posible, ya que K no es continua en el cuadrado unitario. Sin embargo, la descomposición de K puede realizarse usando el teorema de Hammerstein(ver anexo II).

Ahora debemos resolver la ecuación integral

$$\lambda f(t) = \int_0^1 \frac{\min(s,t)-st}{\sqrt{t(1-t)s(1-s)}} f(s)ds.$$

Se puede demostrar que si $f(t)$ satisface la ecuación anterior para alguna λ, entonces $h(t)=f(t)\sqrt{t(1-t)}$ satisface

$$\lambda h''(t) + \psi(t)h(t) = 0 \quad \text{con } \psi(t) = (t(1-t))^{-1},$$

además se tiene que $h(0)=h(1)=0$. La solución de la ecuación diferencial es

$$f_j(t) = 2\sqrt{\frac{2j+1}{j(j+1)}}\sqrt{t(1-t)}\, p_j'(2t-1) \quad y \quad \lambda_j = (j(j+1))^{-1} \quad j=1,2,\ldots,$$

con $f_j(t)$ normalizada y $p_j(\cdot)$ el polinomio de Legendre de grado j definido como

$$p_j(x) = \left\{\frac{1}{j!\,2^j}\right\} \frac{d^j}{dx^j}(x^2-1)^j,$$

o equivalentemente, como el coeficiente de h^j en la expansión en potencias de h de $(1-2xh+h^2)^{-1/2}$. Siguiendo la teoría general de Kac y Siegert(ver anexo II), tenemos que las componentes principales son

$Z_j = \int_0^1 Y(t)f_j(t)dt$ que son v.a. (0,1) no correlacionadas $j=1,2,\ldots,$ y

$$Y \sim \sum_{j=1}^{\infty} \sqrt{\lambda_j}\, Z_j^{*} f_j \quad \text{con } Z_j^{*} = Z_j/\sqrt{\lambda_j}\ .$$

Finalmente

$$A^2 = \int_0^1 Y^2(t)dt \sim \sum_{j=1}^{\infty} \frac{Z_j^{*2}}{j(j+1)}\ ,$$

donde las Z_j^{*}'s son v.a.i.i.d. N(0,1).

En este caso para la descomposición de A_n^2, tenemos que las correspondientes componentes principales normalizadas Z_{nj}^{*} de Y_n son

$$Z_{nj}^{*} = \lambda^{-1/2} \int_0^1 Y_n(t)\ 2\sqrt{\frac{2j+1}{j(j+1)}}\ \sqrt{t(1-t)}\ p_j'(2t-1)dt$$

$$= -\sqrt{2j+1} \int_0^1 Z_n(t) p_j'(2t-1)$$

$$= \sqrt{\frac{2j+1}{n}} \sum_{i=1}^{n} p_j(2\xi_i-1) \qquad j=1,2,\ldots\ ,$$

estas componentes no son v.a.i.i.d.. No obstante, $Z_{n1}^{*},\ldots,Z_{nK}^{*}$ son asintóticamente independientes N(0,1) y

$$T_{nK} = \sum_{j=1}^{k} Z_{nj}^{*2} \to_d \chi^2_{(K)} \quad \text{cuando } n \to \infty,$$

entonces podemos concluir que

$$A_n^2 \longrightarrow_d A^2 \sim \sum_{j=1}^{\infty} \frac{Z_j^{*2}}{j(j+1)} .$$

Una pregunta que surge en este momento de manera natural, es: ¿Cómo usar esta descomposición para determinar la distribución de estas estadísticas ?. Para responder la pregunta, consideremos la función característica para cada una de ellas.

Para W^2

$$\Phi(t) = E(\exp[itW^2]) = \prod_{j=1}^{\infty} (1-2it\lambda_j) \quad \text{con } \lambda_j = 1/(j\pi)^2.$$

Esta función fue invertida por Sminorv (1936), obteniéndose

$$P(W^2>x) = 1/\pi \sum_{j=1}^{\infty} (-1)^{j+1} \int_{(2j-1)^2\pi^2}^{(2j)^2\pi^2} 1/y \sqrt{\frac{-\sqrt{y}}{\operatorname{sen}\sqrt{y}}} \exp(-\frac{xy}{2})dy \quad x>0.$$

Para U^2

$$P(U^2>x) = 2\sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2k^2\pi^2 x) \quad x>0.$$

Finalmente, para A^2

$$P(A^2>x) = 1 - \left\{ \frac{\sqrt{2}}{x} \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(j+1/2)(4j+1)}{j!} \exp\left\{-\frac{(4j+1)^2\pi^2}{8x}\right\} \right.$$

$$\left. \int_0^{\infty} \exp\left\{\frac{x}{8(w^2+1)} - \frac{(4j+1)^2\pi^2 w^2}{8x}\right\} dw \right\} \quad x>0.$$

Otra manera de hacer uso de esta descomposición, consiste en truncar las series infinitas de sumas de v.a. χ^2 y utilizar los métodos numéricos de Imhof y Slepian para invertir las distribuciones. Estas son solamente algunas formas en que se puede usar la descomposición en componentes principales de estas estadísticas.

1.4 PRUEBAS DE POTENCIA.

Una manera de comparar estas estadísticas es a través de su potencia. A este respecto Durbin-Knott(1972) presentan los resultados sobre las potencias asintóticas de estas estadísticas para algunas alternativas consideradas. Estos resultados se muestran en la tablas 1.1 y 1.2.

Como puede observarse en estas tablas, la estadística más potente de las tres consideradas es la Anderson-Darling. En general, en Stephens(1974b) se muestra que en muchas ocaciones la Anderson-Darling es la más potente de estas estadísticas.

TABLA 1.1. Potencias asintóticas de W^2,A^2 y U^2 contra recorridos en la media y la varianza de una normal en una prueba de una cola

Prueba de una cola basada en	media recorrida potencia de la mejor prueba		varianza recorrida potencia de la mejor prueba	
	0.50	0.95	0.50	0.95
W^2	0.342	0.877	0.072	0.205
A^2	0.354	0.890	0.298	0.826
U^2	0.155	0.521	0.185	0.622

TABLA 1.2. Potencias asintóticas de W^2,A^2 y U^2 contra recorridos en la media y la varianza de una normal en una prueba de dos colas

Prueba de dos colas basada en	media recorrida potencia de la mejor prueba		varianza recorrida potencia de la mejor prueba	
	0.50	0.95	0.50	0.95
W^2	0.457	0.928	0.085	0.258
A^2	0.472	0.938	0.401	0.889
U^2	0.205	0.607	0.249	0.711

CAPITULO II

ESTADISTICAS CUADRATICAS CON MUESTRAS CENSURADAS

2.1 INTRODUCCION.

El desarrollo que se ha realizado hasta ahora de las estadísticas del tipo Cramér-von Mises, supone que se tiene una muestra completa, en el sentido de que un individuo en la muestra proporciona información completa sobre la característica de interés. Cuando un individuo sólo aporta información parcial sobre dicha característica se considera como una observación censurada. En este capítulo mostraremos cómo se modifican estas estadísticas cuando la muestra es censurada.

Los tipos de censura que se considerarán en este trabajo, son aquellos que aparecen, con mayor frecuencia, en el área estadística conocida como Análisis de Supervivencia, por lo que es conveniente hacer una introducción del tipo de variables que aparecen en ella.

En el Análisis de Supervivencia se presentan en general variables de tipo longitudinal y se está interesado en estudiar el tiempo que transcurre entre dos eventos perfectamente determinados, el tiempo de inicio de un individuo en el estudio (primer evento) y el tiempo en que ocurre el suceso de interés en el estudio (segundo evento), que recibe el nombre genérico de falla. Los tipos de censura más comunes en esta área son

- Censura tipo I. Ocurre cuando el investigador fija un tiempo

máximo de observación. Por lo tanto, todas las observaciones que no presentaron la falla al final del período de observación resultan censuradas.

- Censura tipo II. En este caso, el investigador decide prolongar el estudio hasta que ocurran r fallas de n posibles, el resto de las observaciones son censuras.

- Censura Aleatoria. Esta censura ocurre independientemente de la voluntad del investigador y corresponde a los casos en los cuales el individuo abandona el estudio, falla por una causa que no es de interés, permanece aún vivo al finalizar el estudio, etc..

En la mayoría de los estudios que se realizan en Análisis de Supervivencia, las censuras que se presentan se catalogan como censuras por la derecha, es decir, solamente se sabe que el tiempo en que ocurriría la falla es mayor que el tiempo de censura observado. Sin embargo, también se presenta, aunque no de manera frecuente, censura por la izquierda, en este caso lo que se sabe es que el tiempo de ocurrencia de la falla es menor que el tiempo de censura observado.

Para los casos de censura tipo I y II, Pettitt y Stephens (1976) proponen las siguientes modificaciones a las estadísticas del tipo Cramér-von Mises.

$$ {}_{q,p}W_n^2 = n \int_q^p \left\{G_n(t)-t\right\}^2 dt, $$

$$ {}_pU_n^2 = n \int_0^p \left\{G_n(t)-t - 1/p\int_0^p [G_n(s)-s]ds \right\}^2 dt, $$

$$ {}_{q,p}A_n^2 = n\int_q^p \frac{[G_n(t)-t]^2}{t(1-t)}dt $$

con $0 \le q < p \le 1$.

Es conveniente remarcar que en estos casos no se tiene censura en el interior del intervalo de observación, solamente en los extremos(censura por la izquierda y por la derecha). En el caso de la estadística ${}_pU_n^2$, únicamente tiene sentido la censura por la derecha ya que la estadística es invariante ante elecciones del origen.

Cuando se presenta censura tipo II en los datos, la muestra cumple con $Z_{(1)}<Z_{(2)}<\ldots<Z_{(r)}\le p<Z_{(r+1)}<\ldots<Z_{(n)}$ para algún valor de p, donde las primeras r observaciones son fallas y el resto censuras.

Las expresiones para calcular estas estadísticas son

$$ {}_pW_n^2 = n\int_0^{z(r)} \left\{G_n(t)-t\right\}^2 dt = \sum_{i=1}^{r}\left(z_{(i)} - \frac{2i-1}{2n}\right)^2 + \frac{r}{12n^2}(z_{(r)} - r/n)^3, $$

$$ {}_pU_n^2 = {}_pW_n^2 - nz_{(r)}^2\left\{\frac{r}{n} - \frac{z_{(r)}}{2} \quad \frac{r\bar{z}}{nz_{(r)}}\right\}^2 $$

y

$$ {}_pA_n^2 = -1/n\Big\{ \sum_{i=1}^{r}(2i-1)\log z_{(i)} + (2n-2i+1)\log(1-z_{(i)}) + $$

$$ r^2[\log(1-z_{(r)})-\log z_{(r)}]+2nr\log(1-z_{(r)})+n^2\log(1-z_{(1)})+z_{(1)} \Big\}. $$

Cuando la muestra presenta Censura tipo I se utilizan las mismas

expresiones anteriores, reemplazando r por r+1 y tomando $p = z_{(r+1)}$. Para este tipo de censura, ocurren r fallas antes del tiempo de censura fijo p, y al resto de las observaciones(n-p) se les asigna el valor p. En ambos casos se tiene

$$ {}_{q,p}W_n^2 = {}_pW_n^2 - {}_qW_n^2 $$

$$ {}_{q,p}A_n^2 = {}_pA_n^2 - {}_qA_n^2. $$

2.2 DISTRIBUCION ASINTOTICA DE LAS ESTADISTICAS.

Para encontrar la distribución asintótica de estas estadísticas, consideraremos que el proceso empírico

$$ Z_n(t) = \sqrt{n}\left\{G_n(t) - t\right\} \qquad q \le t \le p $$

converge al proceso Gaussiano Z(t) definido en [q,p] en lugar de [0,1] como se utiliza en el desarrollo del capítulo anterior. Ahora tenemos

$$ {}_{q,p}W^2 = \int_q^p Z^2(t)\,dt, $$

$$ {}_pU^2 = \int_0^p \left\{Z(t) - 1/p\int_0^p Z(s)\,ds\right\}^2 dt $$

y

$$ {}_{q,p}A^2 = \int_q^p \frac{Z^2(t)}{t(1-t)}\,dt. $$

Para estas estadísticas se puede realizar la descomposición en componentes principales, de la misma manera como se hizo

anteriormente.

Para ${}_{q,p}W^2$, necesitamos resolver la ecuación integral

$$\lambda f(t) = \int_q^p (\min(s,t)-st)f(s)ds \quad q\le t\le p, \tag{4}$$

que, como en el caso anterior, se reduce a resolver la ecuación diferencial

$$\lambda f''(t) + f(t) = 0 \quad q\le t\le p,$$

cuya solución general es

$$f(t) = a\cos(\lambda^{-1/2}t) + b\,\mathrm{sen}(\lambda^{-1/2}t).$$

Obsérvese que, de acuerdo a las expresiones de cálculo, basta tomar q=0, por lo que obtenemos la condición f(0)=0, con la cual la solución general queda

$$f(t) = b\,\mathrm{sen}(\lambda^{-1/2}t).$$

Si sustituimos esta expresión en (4), tenemos que, para encontrar los eigenvalores, hay que resolver la ecuación

$$\tan(\lambda^{-1/2}p) = -\lambda^{-1/2}(1-p). \tag{5}$$

La forma ortonormal de f(t) es

$$f(t) = \sqrt{2}\,\mathrm{sen}(\lambda^{-1/2}t)\left\{ p-\mathrm{sen}(\lambda^{-1/2}p)\cos(\lambda^{-1/2}p)\lambda^{1/2} \right\}^{-1/2},$$

en particular, si tomamos p=1 en (5) tenemos que $\lambda_j = 1/(j\pi)^2$, que coincide con los resultados obtenidos para W^2.

Ahora para ${}_pU^2$ la función de covarianza está dada por

$$K_1(s,t) = \min(s,t) - st - s + sp/2 + s^2/2p - t + tp/2 + t^2/2p + p/3 - p^2/4.$$

de donde la ecuación integral a resolver es

$$\lambda f(t) = \int_0^p K_1(s,t)f(s)ds \qquad 0 \le t \le p,$$

los eigenvalores se obtienen resolviendo

$$\operatorname{sen}\left(\frac{p}{2\sqrt{\lambda}}\right) = 0 , \tag{6}$$

o bien

$$\tan\left(\frac{p}{2\sqrt{\lambda}}\right) = -\frac{(1-p)}{2\sqrt{\lambda}}, \tag{7}$$

resolviendo (6) obtenemos

$$\lambda_j = \frac{p^2}{4(j\pi)^2} \qquad j=1,2,\ldots,$$

con correspondientes eigenfunciones ortonormales asociadas

$$f_j(t) = \sqrt{\frac{2}{p}}\cos\left(\frac{2j\pi t}{p}\right) \qquad j=1,2,\ldots,$$

las soluciones de la ecuación (7) se obtienen de la misma manera que en (5) y sus eigenfunciones ortonormales asociadas son

$$f_j(t) = C\left\{ \operatorname{sen}\left(\frac{2j\pi t}{p}\right) + \frac{j\pi(1-p)}{p}\cos\left(\frac{2j\pi t}{p}\right) \right\},$$

con C la constante de normalización dada por

$$C = \left\{ \frac{p}{2}\left(1+\frac{j^2\pi^2(1-p)^2}{p^2}\right) \right\}^{-1/2}.$$

Nuevamente, si tomamos p=1, recuperamos los eigenvalores para U^2.

Para la estadística ${}_{q,p}A^2$, debemos encontrar la solución de la ecuación integral

$$\lambda f(t) = \int_q^p \frac{\min(s,t)-st}{\sqrt{t(1-t)s(1-s)}} f(s)ds \qquad q\le t\le p, \tag{8}$$

en el caso usual en el que q=0 y p=1 los eigenvalores son $\lambda_j=(j(j+1))^{-1}$ y la solución de la ecuación diferencial es

$$f_j(t) = 2\sqrt{\frac{2j+1}{j(j+1)}}\sqrt{t(1-t)}\; p_j'(2t-1) \qquad j=1,2,\ldots,$$

en este caso, ya que q y p no son enteros, tenemos que $p_\nu(x)$ es un polinomio de Legendre infinito y definimos $p_\nu(-x)=q_\nu(x)$.

Utilizando estos polinomios, las eigenfunciones solución de (8) son

$$f_\nu(t) = \sqrt{t(1-t)}\left\{ a\, p_\nu'(2t-1) + b\, q_\nu'(2t-1) \right\},$$

donde

$$a/b = \left\{ q_\nu(u)(\nu-\nu u+1)+q_{\nu-1}(u) \right\} / \left\{ p_{\nu-1}(u)-p_\nu(u)(\nu-\nu u+1) \right\},$$

para encontrar los eigenvalores $\lambda=(\nu(\nu+1))^{-1}$. Tomamos $q=0$ y encontramos ν resolviendo

$$(1+\nu w+\nu)p_\nu(w) + p_{\nu-1}(w) = 0$$

con $w = 1-2p$.

La distribución de estas estadísticas se puede aproximar (Durbin-Knott(1972)) por

$$B = \sum_{j=1}^{k} \lambda_j C_j + aC_v,$$

donde las λ_j son los eigenvalores de la estadística correspondiente, C_j son v.a.i.i.d. $\chi^2_{(1)}$, a y v son constantes por determinar y C_v v.a. $\chi^2_{(v)}$. Las constantes a y v se eligen de tal manera que la estadística en cuestión y B tengan los dos primeros momentos iguales.

Una manera de usar esta descomposición consiste en considerar nuevamente la función característica de B

$$\Phi(t) = \prod_{j=1}^{k} (1-2i\lambda_j t)^{-1/2} (1-2iat)^{-1/2v},$$

que puede invertirse usando el método de Imhof(1961).

Con respecto a las correspondientes integrales de los procesos empíricos, cabe mencionar que la convergencia de éstas a las integrales de los procesos teóricos, sólo se ha demostrado por

medio de estudios de Monte Carlo (Pettitt y Stephens(1976)), es decir, se tiene solamente evidencia numérica de que

$$ {}_pW_n^2 \longrightarrow_d {}_pW^2, $$
$$ {}_pU_n^2 \longrightarrow_d {}_pU^2 $$

y

$$ {}_pA_n^2 \longrightarrow_d {}_pA^2. $$

2.3 PRUEBAS DE POTENCIA.

Para estas estadísticas se realizaron pruebas de potencia similares a las de Durbin-Knott(1972)(ver tablas 1.1 y 1.2 de esta tesis), y se encontraron los siguientes resultados.

Para recorridos en la media de una normal(μ,σ^2) se tiene que la potencia de ${}_pW^2$ para p=0.9 difiere de la de W^2 sólo en el tercer dígito, mientras que para p=0.5 es aproximadamente 10% menor que la de W^2, para las mismas alternativas y el mismo nivel de significancia considerados por Durbin-Knott. Resultados similares se obtuvieron al comparar ${}_pA^2$ con A^2 en las pruebas de recorrido de medias. En el caso de recorrido en la varianza, W^2 y A^2 tienen poca potencia mismo comportamiento que tuvieron ${}_pW^2$ y ${}_pA^2$.

2.4 MUESTRAS CON CENSURA ALEATORIA.

Ahora corresponde presentar los trabajos que se han realizado en estadísticas del tipo Cramér-Von Mises para censura aleatoria.

Primeramente, podemos caracterizar el modelo de censura aleatoria de la siguiente manera:
Sean $X_1^0,\ldots,X_n^0$ v.a.i.i.d. que representan tiempos de falla, con

función de distribución F^0 y $C_1,\ldots,C_n$ v.a.i.i.d. que representan tiempos de censura, con función de distribución H, los tiempos de falla y censura se suponen independientes. Las observaciones consisten en las parejas (X_i,δ_i), $i=1,\ldots,n$, donde

$$X_i = \min(X_i^0, C_i)$$

y

$$\delta_i = \begin{cases} 1 & \text{si } X_i = X_i^0 \\ 0 & \text{si } X_i = C_i \end{cases}$$

Las observaciones X_i constituyen una muestra aleatoria con función de distribución F que satisface

$$(1-F) = (1-F^0)(1-H).$$

Cuando se tiene censura en la muestra, la F.D.E. no es un buen estimador de la verdadera función de distribución. Sin embargo, podemos calcular un estimador análogo a la F.D.E., $\hat{F}_n^0$ el estimador Kaplan-Meier(K-M) de F^0,dado por

$$\hat{F}_n^0(t) = 1 - \prod_{j:t_j \le t} \left\{ \frac{n_j - d_j}{n_j} \right\},$$

con n_j el número de individuos expuestos al riesgo en t_j(esto es, el número de individuos que continuan en el estudio justo antes de t_j) y d_j el número de muertes al tiempo t_j, $j=1,\ldots,k$ y k=número de fallas observadas.

Cuando no hay empates en la muestra, el estimador (K-M) se puede escribir como

$$\hat{F}_n^0(t) = \begin{cases} 0 & t<X_{(1)} \\ 1-\prod_{j:X_j\le t}\left\{\dfrac{n-R_j}{n-R_j+1}\right\}^{\delta_j} & t<X_{(n)} \\ 1 & t\ge X_{(n)}, \end{cases}$$

con R_j el rango de X_j en la muestra ordenada y δ_j como antes. Si tenemos empate entre una falla y una censura se suele considerar que la falla se presenta antes que la censura.

Tomando en cuenta lo anterior, las expresiones de cálculo para las tres estadísticas son

$$W_n^2 = n\int_0^1 \left\{\hat{F}_n^0(t)-t\right\}^2 dt = n\sum_{j=1}^{n+1}\int_{Y_{(j-1)}}^{Y_{(j)}} \left\{\hat{F}_n^0(Y_{(j-1)})-t\right\}^2 dt$$

$$= n\sum_{j=1}^{n+1}\hat{F}_n^0(Y_{(j-1)})(Y_{(j)}-Y_{(j-1)})[\hat{F}_n^0(Y_{(j-1)})-(Y_{(j)}+Y_{(j-1)})] + n/3.$$

$$U_n^2 = n\sum_{j=1}^{n+1}\hat{F}_n^0(Y_{(j-1)})(Y_{(j)}-Y_{(j-1)})[\hat{F}_n^0(Y_{(j-1)})-(Y_{(j)}+Y_{(j-1)})] + n/3$$

$$-n\sum_{j=1}^{n+1}[\hat{F}_n^0(Y_{(j-1)})(Y_{(j)}-Y_{(j-1)})-1/2]^2$$

$$= W_n^2 - n\sum_{j=1}^{n+1}[\hat{F}_n^0(Y_{(j-1)})(Y_{(j)}-Y_{(j-1)})-1/2]^2.$$

$$A_n^2 = \int_0^1 \frac{[\hat{F}_n^0(t)-t]^2}{t(1-t)}dt = n\sum_{j=1}^{n}\int_{Y_{(j)}}^{Y_{(j+1)}} \frac{[\hat{F}_n^0(Y_{(j)})-t]^2}{t(1-t)}dt$$

$$= -n -n\Big\{ \sum_{j=1}^{n} [\hat{F}_n^{0^2}(Y_{(j)}) - \hat{F}_n^{0^2}(Y_{(j-1)})]\log Y_{(j)} + [(1-\hat{F}_n^0(Y_{(j-1)}))^2$$

$$-(1-\hat{F}_n^0(Y_{(j)}))^2]\log(1-Y_{(j)}) \Big\} \qquad (9)$$

con $Y_{(j)} = F^0(X_{(j)}^0)$ y $Y_{(0)}=0$.

En cada caso, si reemplazamos el estimador (K-M) por la función de distribución empírica obtenemos los mismos resultados que para muestras sin censura.

Koziol y Green(1976)(K-G) estudian la estadística Cramér-von Mises para hacer pruebas de bondad de ajuste cuando los datos estan sujetos a un particular mecanismo de censura aleatoria. El modelo que presentan (K-G) considera $(1-H)=(1-F^0)^{\beta}$, con β una constante no negativa, que puede interpretarse como "parámetro de censura" y $\beta=0$ corresponde a no censura. La estadística que estudian es

$$_cW_n^2 = n\int_0^1 \left\{\hat{G}_n^0(t)-t\right\}^2 dt,$$

con $\hat{G}_n^0(\cdot)$ el estimador (K-M) de una muestra uniforme (0,1) con censura aleatoria.

2.4.1 DISTRIBUCION ASINTOTICA DE $_cW_n^2$.

Para este tipo de censura (K-G) obtienen la distribución asintótica de la estadística Cramér-von Mises. En este caso,

tenemos resultados análogos a los que teníamos anteriormente. El proceso empírico (Breslow-Crowley (1974))

$$Y_n(t) = \sqrt{n}\left\{\hat{F}_n^0(t)-F(t)\right\} \qquad t>0$$

converge débilmente a un proceso Gaussiano. En particular, bajo el modelo de (K-G), tenemos que el proceso empírico converge a un proceso Gaussiano de media cero y función de covarianza

$$\mathrm{Cov}(Y(s),Y(t)) = [1-F^0(s)][1-F^0(t)]\int_0^s (1-F^0)^{-(2+\beta)}dF^0 \qquad s\le t.$$

De nuevo basta considerar F^0 uniforme, por lo que la estadística Cramér-von Mises esta dada por

$$cW_n^2 = \int_0^1 Y_n^2(t)dt = n\int_0^1 \left\{\hat{G}_n^0(t)-t\right\}^2 dt,$$

y además se tiene el siguiente resultado de convergencia

$$cW_n^2 = \int_0^1 Y_n^2(t)dt \longrightarrow_d cW^2 = \int_0^1 Y^2(t)dt.$$

La función de covarianza del proceso límite se puede escribir como

$$K(s,t) = (1+\beta)^{-1}(1-s)(1-t)[(1-\min(s,t))^{-(1+\beta)}-1] \qquad 0\le s,t<1$$

Se puede demostrar fácilmente que K no es continua en el punto(1,1) y que no es integrable para $\beta\ge 2$. No obstante, para $\beta<2$ podemos aplicar el teorema de Hammerstein(ver anexo II), lo que implica que en cualquier dominio del cuadrado unitario, la

función $K(s,t)$ puede expanderse en una serie uniformemente convergente dada por

$$K(s,t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \qquad 0 \le s,t < 1,$$

que nos permite nuevamente, siguiendo la teoría de Kac y Siegert (ver anexo II), descomponer la estadística como

$$_cW^2 = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2},$$

con Z_j^{*2} v.a.i.i.d. $\chi^2_{(1)}$.

De la misma manera, ahora siguiendo la teoría de Durbin y Knott(ver anexo II), podemos descomponer a $_cW_n^2$ como

$$_cW_n^2 = \sum_{j=1}^{\infty} \lambda_j Z_{nj}^{*2},$$

de donde concluimos que

$$_cW_n^2 \longrightarrow_d {}_cW^2 = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2}.$$

Para encontrar los eigenvalores y sus correspondientes eigenfunciones debemos resolver la ecuación integral

$$\lambda f(t) = \int_0^1 K(s,t) f(s) ds \qquad 0 \le t < 1$$

que, si tomamos

$$h(t) = f(t)(1-t)^{\beta} \qquad 0 \le t \le 1$$

la ecuación integral se convierte en la siguiente ecuación diferencial

$$(1-t)^{\beta}h''(t) + \beta(1-t)^{\beta-1}h'(t) + \lambda^{-1}h(t) = 0$$

o equivalentemente en términos de f(t)

$$(1-t)^{2}f''(t) - \beta(1-t)f'(t) + [\lambda^{-1}(1-t)^{2-\beta} - \beta]f(t) = 0$$

que puede transformarse a una ecuación de Bessel usando los siguientes cambios de variables

$$\Phi = (1-t)^{-p}f(t) \quad y \quad T = a(1-t)^{q},$$

con $p = \frac{1}{2} - \frac{\beta}{2} \quad q = 1 - \frac{\beta}{2} \quad a = 2(1+\beta)^{1/2}\lambda^{-1/2}(2-\beta)^{-1}$,

la solución general es

$$f(t) = (1-t)^{p}\left\{c_1 J_{\nu}(a(1-t)^{q}) + c_2 K_{\nu}(a(1-t)^{q})\right\},$$

con J una función de Bessel de primer tipo, K una función de Bessel modificada y

$$\nu = [(1-\beta)^{2}+4\beta]^{1/2}(2-\beta)^{-1} = (1+\beta)/(2-\beta).$$

Las eigenfunciones son

$$f_j(t) = c_j\,(1-t)^{1/2-1/2\beta}J_{\nu}(\,2(1+\beta)^{1/2}\lambda_j^{-1/2}(2-\beta)^{-1}(1-t)^{1-(1/2)\beta}),$$

en donde c_j es la constante de normalización. Los eigenvalores pueden determinarse resolviendo

$$J_\nu(\ 2(1+\beta)^{1/2}\lambda_j^{-1/2}(2-\beta)^{-1}) = 0.$$

Para obtener la distribución de ${}_cW^2$, podemos aproximarla (Durbin-Knott(1972)) nuevamente por

$$B = \sum_{j=1}^{k} \lambda_j C_j + aC_v,$$

donde las λ_j son los eigenvalores de ${}_cW^2$, C_j son v.a.i.i.d $\chi^2_{(1)}$, a y v son constantes por determinar y C_v v.a. $\chi^2_{(v)}$. Las constantes a y v se eligen de tal manera que la estadística ${}_cW^2$ y B tengan los dos primeros momentos iguales.

Para usar esta descomposición consideremos nuevamente la función característica de B

$$\Phi(t) = \prod_{j=1}^{k} (1-2i\lambda_j t)^{-1/2}(1-2iat)^{-1/2v},$$

que podemos invertir usando el método de Imhof(1961).

2.4.2 PRUEBAS DE POTENCIA.

Con respecto a pruebas de potencia para ${}_cW^2$, (K-G) realizan las mismas pruebas propuestas por Durbin-Knott, además de pruebas sobre recorrido del parámetro de escala en una exponencial(θ) y distribuciones Weibull como alternativas a esta distribución exponencial. Algunos resultados que se obtuvieron son los siguientes.

Para pruebas sobre recorrido de la media en distribuciones normales, la primer componente de ${}_cW^2$ es asintóticamente tan potente como la estadística completa para cualquier valor del parámetro de censura β. Sin embargo, cuando β crece, lo que implica que la proporción de censura esperada se incrementa, tanto la primer componente como ${}_cW^2$ reducen su potencia asintótica. En las pruebas sobre recorrido de la varianza en distribuciones normales, solamente la segunda componente de ${}_cW^2$ es asintóticamente más potente que la estadística completa. En las pruebas para recorrido del parámetro de escala en la exponencial, la primer componente es tan potente como ${}_cW^2$, pero ambas tienen poca potencia. Finalmente, para las alternativas Weibull a la exponencial, la segunda componente de la estadística es más potente que ${}_cW^2$ excepto para la cola superior en $\beta=3/2$.

CAPITULO III

LA ESTADISTICA ANDERSON-DARLING PARA MUESTRAS CON CENSURA ALEATORIA.

3.1 INTRODUCCION.

En los capítulos anteriores se presentó la metodología que se tiene para hacer pruebas de Bondad de Ajuste con las estadísticas del tipo Cramér-von Mises. Se puede notar que, con algunas modificaciones, la metodología es estándar para todos los casos. Ahora corresponde presentar el trabajo que es el objetivo central de esta tesis.

El objetivo de este trabajo es proponer una modificación a la estadística A^2 y analizar su comportamiento para realizar una prueba de Bondad de Ajuste en muestras con censura aleatoria. El desarrollo de esta propuesta se sujetará al modelo que proponen Koziol-Green para el mecanismo de censura, esto es,

$$(1-H) = (1-F^0)^{\beta},$$

con $\beta \geq 0$ parámetro de censura.

Se propone utilizar la estadística

$${}_cA_n^2 = n\int_0^1 \frac{(\hat{G}_n^0(t)-t)^2}{t(1-t)}\,dt,$$

con $\hat{G}_n^0(\cdot)$ como se definió anteriormente, para realizar la prueba

$$H_0: F^0 = F^*,$$

con F^* una función de distribución totalmente especificada y suponiendo que se tiene una muestra censurada.

Obsérvese que el peso que se da a cada observación con esta estadística no es en realidad el que le correspondería si este peso fuera el asignado bajo la verdadera varianza asintótica del estimador (K-M). Sin embargo, no es posible asignar el peso verdadero porque se desconoce la expresión exacta para la varianza del (K-M) y además las aproximaciones que hay a la varianza de él para muestras finitas dependen de la muestra particular, por lo que se decidió utilizar la misma función de peso que para la estadística de Anderson-Darling en el caso de no censura.

Para encontrar la distribución asintótica de esta estadística consideremos primeramente los siguientes resultados.

Haciendo una extensión del resultado utilizado por (K-G) para el caso de la Cramér-von Mises con censura aleatoria, el proceso ponderado

$$Y_n(t) = \frac{\sqrt{n}\left\{\hat{F}_n^0(t) - F^0(t)\right\}}{\sqrt{F^0(t)(1-F^0(t))}} \longrightarrow_d Y(t) \qquad t>0,$$

con $Y(t)$ un proceso Gaussiano que cumple con

$$E(Y(t)) = 0 \qquad t>0$$

y

$$\mathrm{cov}(Y(s),Y(t)) = \frac{(1-F^0(s))(1-F^0(t))}{\sqrt{F^0(s)(1-F^0(s))F^0(t)(1-F^0(t))}} \int_0^s (1-F^0)^{-(2+\beta)} dF^0 \qquad s\le t.$$

Utilizando la T.I.P., H_0 es equivalente a considerar F^0 unifome en (0,1). La función de covarianza de este proceso estocástico se puede escribir como

$$K(s,t) = (1+\beta)^{-1}\frac{(1-s)(1-t)}{\sqrt{s(1-s)t(1-t)}}\left[(1-\min(s,t))^{-(1+\beta)}-1\right] \qquad 0<s,t<1.$$

Entre las características de esta función podemos mencionar las siguientes:

- No es continua en ninguna de las esquinas del cuadrado unitario
- No es continua en los puntos de la forma (0,t) $0<t<1$ ni en los de la forma (s,0) $0<s<1$.
- La carcterística anterior impide aplicar el Teorema de Hammerstein (ver anexo 2) para justificar una descomposición en una serie uniformemente convergente.
- Se puede mostrar que $\int_0^1\int_0^1 K(s,t)\,dsdt < \infty$ para $\beta<2$.
- La correspondiente ecuación integral, puede transformarse en una ecuación diferencial de la siguiente manera

$$\lambda f(t) = \int_0^1 (1+\beta)^{-1}\frac{(1-s)(1-t)}{\sqrt{s(1-s)t(1-t)}}\left[(1-\min(s,t))^{-(1+\beta)}-1\right]f(s)ds \quad 0<t<1$$

si hacemos el cambio de variable

$$h(t) = \sqrt{t(1-t)}(1-t)^{\beta}f(t) \qquad 0\le t\le 1,$$

que cumple la condición $h(0)=h(1)=0$, obtenemos la ecuación diferencial

$$t(1-t)^{\beta+1}h''(t) + \beta t(1-t)^{\beta}h'(t) + \lambda^{-1}h(t) = 0$$

y en términos de f(t)

$$2t^2(1-t)^2f''(t) + \left\{ 2t(1-t)(1-(2+\beta))t \right\}f'(t) + \left\{ -2t^2 + (2-\beta)t + \lambda^{-1}2t(1-t)^{1-\beta} \right\}f(t) = 0.$$

Por lo general, resolver estas ecuaciones diferenciales resulta muy complicado, por lo que es usual hacer una aproximación a la descomposición de la integral del proceso límite.

La imposibilidad de aplicar el Teorema de Hammerstein nos lleva a considerar dos caminos para continuar el análisis de la estadística ${}_cA_n^2$.

El primer camino es considerar, para cada valor de β, la matriz $\mathbb{K}$ construida a partir de una versión discreta de la función de covarianza del proceso límite Y(t), y estudiar el comportamiento de la variable aleatoria

$$A_m^{*2} = \sum_{j=1}^{m} \lambda_j Z_j^2,$$

con λ_j los eigenvalores de la matriz $\mathbb{K}$ y Z_j^2 v.a.i.i.d. $\chi^2_{(1)}$.

El segundo camino es simular, para cada valor de β, la distribución de la estadística ${}_cA_n^2$ utilizando la expresión (9) del capítulo anterior, y analizar su comportamiento para distintos tamaños de muestra.

Hay que notar que si los resultados que se obtengan por ambos caminos son similares, tendriamos alguna evidencia para determinar la distribución asintótica de la estadística propuesta.

3.2 COMPORTAMIENTO DE A_m^{*2}.

Para analizar el comportamiento de esta variable aleatoria necesitamos una versión discreta de la función de covarianza, a saber

$$K(t_i,t_j) = \frac{1}{m}\left\{ (1+\beta)^{-1}\frac{(1-t_i)(1-t_j)}{\sqrt{t_i(1-t_i)t_j(1-t_j)}}\left[(1-\min(t_i,t_j))^{-(1+\beta)}-1\right] \right\}$$

con $t_i,t_j \in (0,1)$, $i,j = 1,\ldots,m-1$.
donde m es el número de puntos en que se dividen los lados del cuadrado unitario.

Los puntos (t_i,t_j) se pueden tomar de la forma $(i/m,j/m)$ y así evitar las discontinuidades que presenta la función en t= 0 y 1. Si m es suficientemente grande se espera que los valores $K(t_i,t_j)$ proporcionen un buen espectro de la verdadera función de covarianza. Estos valores se pueden arreglar en una matriz $\mathbb{K}=K(t_i,t_j)$ y a partir de ella trabajar como se hace comúnmente en estadística, encontrando sus eigenvalores y sus correspondientes eigenvectores, en el entendido de que se espera que para m grande los eigenvalores de la matriz converjan a los verdaderos eigenvalores y que los eigenvectores asociados proporcionen una versión discreta de las eigenfunciones correspondientes.

Para calcular los eigenvalores y los eigenvectores asociados, se usaron varias cuadrículas (que definen la dimensión de la matriz $\mathbb{K}$) para dividir el cuadrado unitario, de 50 x 50, 100 x 100, 500 x 500 y de 1000 x 1000. Ya que para β=0 se tiene la estadística de Anderson-Darling usual, de la que conocemos sus eigenvalores, ésto nos ayudó para decidir sobre el tamaño de la cuadrícula a usar. Pese a que al aumentar el tamaño de la cuadrícula la aproximación a los verdaderos eigenvalores es mejor, decidimos usar la cuadrícula de 100 x 100 ya que proporciona una buena

aproximación y simplifica mucho el trabajo computacional.

Tomando en cuenta que $\mathbb{K}$, la matriz construida, es simétrica sus eigenvalores son reales. Además tendremos

$$\sum_{j=1}^{m} \lambda_j = \text{traza } \mathbb{K},$$

y se pudo mostrar que para $\beta<1$

$$\lim_{m\to\infty} \sum_{j=1}^{m} \lambda_j < \infty.$$

Se calcularon algunos valores de convergencia. Estos valores se muestran en la tabla 3.1

TABLA 3.1. Valores de convergencia de la serie de eigenvalores de la matriz $\mathbb{K}$ para diferentes valores de β

β	$\sum_{j=1}^{\infty} \lambda_j$
0	1.0
0.1	1.07064
0.2	1.15648
0.3	1.26365
0.4	1.40222
0.5	1.59042
0.6	1.86255
0.7	2.29602
0.8	3.08779
0.9	5.03902

A partir de estos eigenvalores y eigenfunciones se puede construir la variable aleatoria

$$A_m^{*2} = \sum_{j=1}^{m} \lambda_j Z_j^2,$$

con λ_j y Z_j^2 como se definieron anteriormente y

$$Z_j = \lambda_j^{-1/2} \sum_{k=1}^{m} \gamma_{jk} Y(t_j) \qquad t_j \in (0,1),$$

con γ_{jk} el eigenvector asociado a λ_j. Para encontrar la distribución de A_m^{*2} usaremos la correspondiente aproximación a su función característica, es decir

$$\Phi(A_m^{*2}) \cong \prod_{j=1}^{k} (1-2it\lambda_j)^{-1/2},$$

la cual podemos invertir usando el método de Imhof(1961) para encontrar los cuantiles de la distribución.

Los cuantiles de la distribución para algunos valores del parámetro de censura β se muestran en la tabla 3.2.

Estos valores de la estadística se calcularon utilizando los primeros 30 eigenvalores de la descomposición. Por lo general, el número de eigenvalores que se considera para usar el método de Imhof es 20 (Durbin-Knott(1972)), en este caso decidimos utilizar un número mayor, porque al crecer el parámetro de censura β se espera que el porcentaje de varianza que explica este conjunto de eigenvalores vaya decreciendo. En seguida se muestran los porcentajes de varianza explicada por estos 30 eigenvalores para los valores de β utilizados anteriormente. Estos porcentajes se calcularon considerando como el total de la varianza los valores de convergencia de la serie de eigenvalores que se muestran en la

tabla 3.1.

β	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
% v.e.	97.7	97.4	96.8	96.0	94.6	92.2	88.2	81.4	69.9	50.5

TABLA 3.2. Cuantiles de la distribución de A_m^{*2}

	β=0	β=0.1	β=0.2	β=0.3	β=0.4	β=0.5
α	cuantil	cuantil	cuantil	cuantil	cuantil	cuantil
0.001	0.112	0.125	0.140	0.159	0.181	0.208
0.005	0.149	0.165	0.185	0.208	0.237	0.273
0.01	0.171	0.189	0.211	0.238	0.271	0.311
0.025	0.210	0.233	0.259	0.292	0.331	0.380
0.05	0.253	0.279	0.311	0.349	0.395	0.452
0.1	0.316	0.348	0.385	0.431	0.487	0.555
0.2	0.420	0.458	0.505	0.562	0.632	0.718
0.3	0.519	0.564	0.619	0.685	0.767	0.868
0.4	0.626	0.678	0.740	0.816	0.909	1.024
0.5	0.749	0.808	0.878	0.963	1.068	1.198
0.6	0.900	0.966	1.046	1.141	1.259	1.405
0.7	1.098	1.174	1.264	1.372	1.504	1.669
0.8	1.389	1.477	1.580	1.705	1.857	2.045
0.9	1.919	2.026	2.152	2.302	2.484	2.709
0.95	2.484	2.611	2.759	2.933	3.144	3.403
0.975	3.075	3.222	3.392	3.592	3.831	4.124
0.99	3.884	4.057	4.481	4.509	4.772	5.109
0.995	4.525	4.707	4.931	5.193	5.501	5.874
0.999	5.997	6.244	6.722	6.852	7.234	7.691

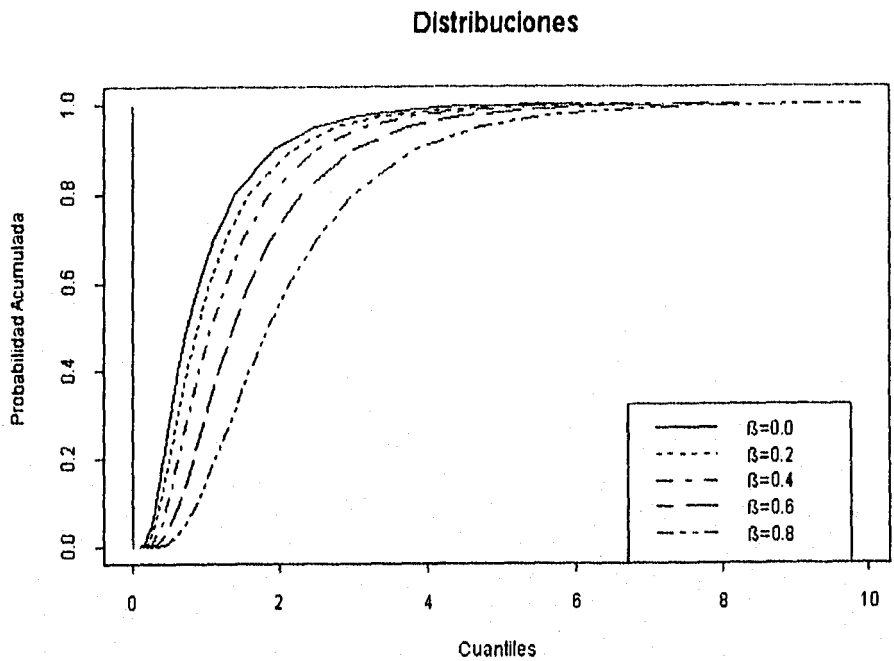
TABLA 3.2. Continuación

α	β=0.6 cuantil	β=0.7 cuantil	β=0.8 cuantil	β=0.9 cuantil
0.001	0.241	0.281	0.330	0.390
0.005	0.316	0.368	0.433	0.513
0.01	0.360	0.421	0.495	0.597
0.025	0.439	0.513	0.604	0.718
0.05	0.522	0.610	0.718	0.855
0.1	0.641	0.747	0.880	1.048
0.2	0.825	0.959	1.129	1.345
0.3	0.994	1.153	1.355	1.613
0.4	1.168	1.351	1.585	1.885
0.5	1.361	1.569	1.836	2.181
0.6	1.589	1.824	2.127	2.522
0.7	1.877	2.144	2.490	2.944
0.8	2.283	2.590	2.991	3.523
0.9	2.993	3.360	3.845	4.499
0.95	3.729	4.150	4.711	5.477
0.975	4.489	4.936	5.594	6.466
0.99	5.529	6.069	6.789	7.792
0.995	6.335	6.925	7.711	8.810
0.999	8.242	9.100	9.901	11.22

α es la probabilidad de la cola izquierda de la distribución.

En la gráfica 3.1 se muestran las distribuciones de esta estadística para algunos de los valores de β considerados. Las gráficas de estas funciones de distribución muestran que, a medida que el parámetro de censura crece, las distribuciones son más sesgadas hacia la derecha, lo que implica que conforme β crece se requiere de un mayor tiempo en el estudio para que ocurran un cierto número de fallas.

GRAFICA 3.1. Distribuciones de A_n^{*2} para diferentes valores de β.



3.3 SIMULACIONES DE LA ESTADISTICA $_cA_n^2$.

En el segundo camino, que consiste en hacer estudios de simulación para la estadística $_cA_n^2$, se simularon 5000 valores de la estadística de prueba utilizando la expresión(9) del capítulo II, con diferentes tamaños de muestra. Analizando estos valores, se puede observar que para los valores del parámetro de censura $0 \leq \beta \leq 0.6$ los cuantiles de esta distribución convergen rápidamente, ya que para los tamaños de muestra considerados, los valores de cada cuantil son prácticamente iguales. Por otra parte, para los

valores $0.7 \leq \beta \leq 0.9$ se observa que a medida que el tamaño de muestra crece los cuantiles de la distribución van creciendo también. Para buscar hacia donde convergen estos cuantiles se tomó un tamaño de muestra mayor(200) con el cual ya se puede observar un punto de convergencia. En la tabla 3.3 se muestran estas simulaciones. Se anexan a la tabla en negritas los valores de los cuantiles correspondientes para A_m^{*2}.

En este momento podemos hacer una comparación de los resultados obtenidos por las dos caminos. Podemos observar que, en general, los valores de los cuantiles de A_m^{*2} y los que se simularon de ${}_cA_n^2$ son casi iguales. Este hecho nos permite conjeturar que las distribuciones son iguales. Es decir, tenemos evidencia numérica de que

$$ {}_cA_n^2 = \int_0^1 Y_n^2(t)\,dt \longrightarrow_d A_m^{*2}. $$

Además esta igualdad proporciona elementos para suponer que la descomposición de la función de covarianza que se realizó debe ser factible, pese a que no se cuenta con los resultados teóricos que la sustenten.

TABLA 3.3. Cuantiles de la distribución de ${}_cA_n^2$.

		NIVEL DE SIGNIFICANCIA α							
${}_cA_n^2$	n	0.001	0.005	0.01	0.025	0.05	0.10	0.20	0.30
	10	0.148	0.178	0.202	0.243	0.288	0.347	0.446	0.542
	30	0.155	0.181	0.201	0.244	0.287	0.357	0.461	0.562
β=0.0	50	0.147	0.184	0.202	0.236	0.280	0.343	0.447	0.550
	100	0.143	0.178	0.196	0.235	0.281	0.354	0.453	0.537
		0.112	**0.149**	**0.171**	**0.210**	**0.253**	**0.316**	**0.420**	**0.519**
	10	0.158	0.195	0.214	0.254	0.300	0.370	0.473	0.581
	30	0.152	0.186	0.218	0.259	0.306	0.375	0.478	0.585
β=0.1	50	0.162	0.187	0.211	0.256	0.302	0.376	0.483	0.587
	100	0.171	0.200	0.228	0.272	0.318	0.381	0.492	0.596
		0.125	**0.165**	**0.189**	**0.233**	**0.279**	**0.348**	**0.458**	**0.564**
	10	0.158	0.207	0.234	0.285	0.339	0.406	0.519	0.638
	30	0.165	0.219	0.251	0.292	0.339	0.417	0.527	0.653
β=0.2	50	0.187	0.229	0.255	0.303	0.349	0.426	0.548	0.658
	100	0.168	0.210	0.250	0.301	0.360	0.429	0.535	0.638
		0.140	**0.185**	**0.211**	**0.259**	**0.311**	**0.385**	**0.505**	**0.619**
	10	0.174	0.229	0.259	0.307	0.360	0.429	0.561	0.679
	30	0.186	0.240	0.267	0.319	0.367	0.451	0.576	0.703
β=0.3	50	0.193	0.241	0.268	0.325	0.378	0.463	0.593	0.715
	100	0.195	0.250	0.276	0.338	0.389	0.471	0.611	0.737
		0.159	**0.208**	**0.238**	**0.292**	**0.349**	**0.431**	**0.562**	**0.685**
	10	0.193	0.250	0.279	0.337	0.389	0.471	0.604	0.735
	30	0.223	0.267	0.301	0.346	0.411	0.508	0.647	0.782
β=0.4	50	0.212	0.276	0.304	0.361	0.422	0.519	0.656	0.784
	100	0.218	0.274	0.310	0.366	0.438	0.530	0.680	0.818
		0.181	**0.237**	**0.271**	**0.331**	**0.395**	**0.487**	**0.632**	**0.767**

TABLA 3.3. Continuación

	10	0.216	0.268	0.315	0.359	0.418	0.513	0.654	0.793
	30	0.217	0.277	0.311	0.370	0.437	0.547	0.702	0.845
β=0.5	50	0.234	0.287	0.330	0.403	0.471	0.562	0.718	0.862
	100	0.246	0.315	0.359	0.426	0.495	0.590	0.750	0.900
		0.208	**0.273**	**0.311**	**0.380**	**0.452**	**0.555**	**0.718**	**0.868**
	10	0.221	0.271	0.321	0.385	0.451	0.554	0.704	0.864
	30	0.221	0.314	0.354	0.417	0.494	0.608	0.777	0.929
β=0.6	50	0.256	0.324	0.368	0.446	0.518	0.629	0.801	0.946
	100	0.256	0.357	0.394	0.464	0.553	0.658	0.837	0.997
		0.241	**0.316**	**0.360**	**0.439**	**0.522**	**0.641**	**0.825**	**0.994**
	10	0.242	0.298	0.334	0.398	0.477	0.599	0.782	0.952
	30	0.289	0.354	0.404	0.469	0.545	0.655	0.844	1.019
β=0.7	50	0.272	0.368	0.417	0.484	0.570	0.698	0.904	1.073
	100	0.333	0.426	0.463	0.563	0.620	0.742	0.952	1.139
	200	0.336	0.417	0.463	0.554	0.627	0.759	0.954	1.141
		0.281	**0.368**	**0.421**	**0.513**	**0.610**	**0.747**	**0.959**	**1.153**
	10	0.228	0.292	0.353	0.432	0.510	0.622	0.813	0.983
	30	0.297	0.370	0.415	0.499	0.589	0.727	0.930	1.123
β=0.8	50	0.291	0.389	0.446	0.523	0.620	0.761	0.981	1.189
	100	0.334	0.445	0.500	0.588	0.693	0.829	1.049	1.261
	200	0.398	0.457	0.515	0.625	0.738	0.882	1.105	1.309
		0.330	**0.433**	**0.495**	**0.604**	**0.718**	**0.880**	**1.129**	**1.355**
	10	0.270	0.324	0.366	0.457	0.532	0.660	0.879	1.072
	30	0.323	0.405	0.456	0.556	0.656	0.800	1.008	1.234
β=0.9	50	0.356	0.447	0.498	0.592	0.682	0.842	1.080	1.278
	100	0.419	0.517	0.562	0.662	0.785	0.943	1.176	1.405
	200	0.421	0.531	0.599	0.723	0.845	1.022	1.283	1.500
		0.390	**0.513**	**0.597**	**0.718**	**0.855**	**1.048**	**1.345**	**1.613**

TABLA 3.3. Continuación

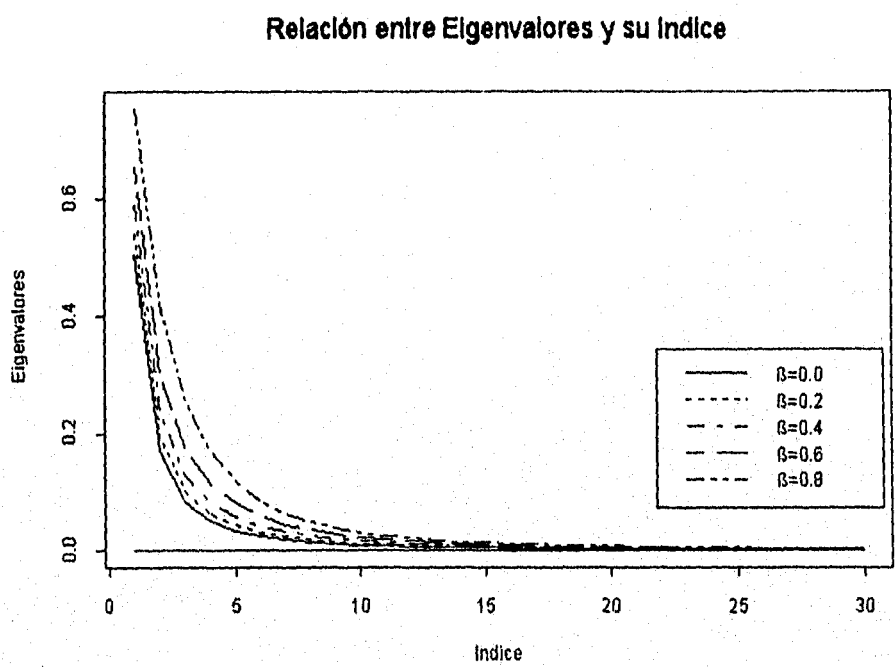
		NIVEL DE SIGNIFICANCIA α							
cA_n^2	n	0.70	0.80	0.90	0.95	0.975	0.990	0.995	0.999
	10	1.161	1.412	1.945	2.516	3.038	3.841	4.367	5.986
	30	1.116	1.404	1.930	2.506	3.065	4.055	4.597	6.081
β=0.0	50	1.134	1.454	1.943	2.463	3.014	3.771	4.256	5.794
	100	1.112	1.388	1.862	2.461	3.002	3.706	4.357	6.015
		1.098	**1.389**	**1.919**	**2.484**	**3.075**	**3.884**	**4.525**	**5.997**
	10	1.182	1.497	2.035	2.674	3.398	4.242	5.125	6.523
	30	1.195	1.514	2.073	2.616	3.295	3.976	4.413	6.293
β=0.1	50	1.168	1.471	1.977	2.591	3.155	3.946	4.516	6.243
	100	1.216	1.541	2.113	2.697	3.232	4.052	4.632	6.027
		1.174	**1.477**	**2.026**	**2.611**	**3.222**	**4.057**	**4.707**	**6.244**
	10	1.288	1.627	2.166	2.781	3.564	4.412	5.129	7.302
	30	1.320	1.627	2.174	2.824	3.492	4.294	5.104	6.746
β=0.2	50	1.326	1.653	2.226	2.814	3.464	4.308	4.856	6.188
	100	1.280	1.612	2.216	2.767	3.402	4.240	4.831	6.291
		1.126	**1.580**	**2.152**	**2.759**	**3.392**	**4.481**	**4.931**	**6.722**
	10	1.373	1.700	2.291	2.997	3.740	4.831	5.589	7.532
	30	1.377	1.720	2.333	3.065	3.847	4.794	5.808	8.050
β=0.3	50	1.400	1.730	2.336	2.968	3.602	4.536	5.162	6.518
	100	1.441	1.787	2.401	3.035	3.784	4.685	5.515	7.497
		1.372	**1.705**	**2.302**	**2.933**	**3.592**	**4.509**	**5.193**	**6.852**
	10	1.455	1.847	2.545	3.337	4.058	5.306	6.271	8.820
	30	1.544	1.917	2.615	3.287	3.903	5.182	6.240	6.913
β=0.4	50	1.538	1.903	2.611	3.359	4.110	4.916	5.720	6.913
	100	1.571	1.902	2.517	3.205	3.811	4.824	5.557	7.176
		1.504	**1.857**	**2.484**	**3.144**	**3.831**	**4.772**	**5.501**	**7.234**

TABLA 3.3. Continuación

	10	1.598	2.015	2.719	3.621	4.550	6.046	6.838	10.57
	30	1.629	2.052	2.749	3.472	4.249	5.405	6.346	8.541
β=0.5	50	1.660	2.089	2.762	3.530	4.342	5.653	6.734	9.949
	100	1.734	2.150	2.885	3.647	4.408	5.431	6.171	7.955
		1.669	**2.045**	**2.709**	**3.403**	**4.124**	**5.109**	**5.874**	**7.691**
	10	1.692	2.117	2.923	3.671	4.568	6.050	7.860	11.03
	30	1.803	2.215	2.982	3.920	4.835	6.313	7.248	11.43
β=0.6	50	1.828	2.213	2.939	3.811	4.715	6.080	7.152	10.23
	100	1.905	2.288	2.998	3.849	4.723	6.079	7.131	9.643
		1.877	**2.283**	**2.993**	**3.729**	**4.489**	**5.529**	**6.335**	**8.999**
	10	1.826	2.300	3.107	4.212	5.396	7.440	8.969	12.46
	30	1.966	2.402	3.243	4.122	5.112	6.832	7.744	10.36
β=0.7	50	2.009	2.463	3.271	4.240	5.242	6.623	7.975	11.84
	100	2.105	2.563	3.350	4.139	5.018	6.115	7.058	10.16
	200	2.116	2.572	3.347	4.148	5.087	6.289	7.536	10.00
		2.144	**2.590**	**3.360**	**4.150**	**4.963**	**6.069**	**6.925**	**9.100**
	10	1.999	2.448	3.329	4.398	5.745	7.868	9.287	17.53
	30	2.119	2.543	3.258	4.218	5.325	6.550	7.772	11.40
β=0.8	50	2.238	2.753	3.590	4.636	5.553	7.142	8.375	11.57
	100	2.320	2.808	3.749	4.627	5.581	7.100	8.375	13.50
	200	2.350	2.831	3.700	4.611	5.575	6.999	8.426	10.49
		2.490	**2.991**	**3.845**	**4.711**	**5.594**	**6.789**	**7.711**	**9.901**
	10	2.122	2.650	3.648	4.820	5.982	8.599	10.40	19.55
	30	2.308	2.842	3.750	4.800	5.854	7.648	8.652	13.39
β=0.9	50	2.242	2.941	3.903	4.957	5.997	7.527	8.819	12.46
	100	2.593	3.126	4.153	5.101	6.171	7.748	9.083	13.52
	200	2.712	3.226	4.189	5.190	6.170	7.694	9.085	12.07
		2.944	**3.523**	**4.499**	**5.477**	**6.466**	**7.429**	**8.810**	**11.22**

Como puede observarse, tanto los resultados de convergencia de la serie de eigenvalores, las distribuciones asintóticas y las simulaciones para muestras pequeñas, se realizaron para algunos valores del parámetro de censura β. Una pregunta importante que debemos hacernos es ¿ existe una manera general de encontrar los eigenvalores de la matriz $\mathbb{K}$ conociendo solamente un estimador del parámetro β ?. Esta pregunta se planteó al observar el patrón de comportamiento de los eigenvalores para cada valor de β y el índice(j=1,...,30) del correspondiente eigenvalor, patrón que se muestra en la gráfica 3.2.

GRAFICA 3.2.



Para intentar modelar este comportamiento se ajustaron varios modelos, de donde se obtuvo como primer resultado lo siguiente

Dado β el parámetro de censura y el índice del correspondiente eigenvalor se tiene que

$$\log(1+1.85\log(j)+\log(\text{eigen}[\beta,j])) = C_{1j}+C_{2j}\beta \quad j=1,\ldots,30,$$

con $\text{eigen}[\beta,j]$ el j-ésimo eigenvalor de la matriz con un parámetro β de censura en la muestra. Además

$$C_{1j} = 1.105\exp(-0.008j)-1.768\exp(-0.869j)-1.5$$

$$C_{2j} = 1.001\exp(-0.006j)-1.604\exp(-1.452j)+0.4 \ ,$$

entonces, se tiene finalmente que

$$\text{eigen}[\beta,j] = \exp(\exp[C_{1j}+C_{2j}\beta]-(1+\log(j))).$$

La utilidad de estos resultados es que, dado β el parámetro de censura de la muestra, se pueden calcular a partir de este valor los eigenvalores de la matriz y con ellos encontrar la distribución de la estadística.

Lo adecuado del ajuste se puede observar comparando los valores de los eigenvalores ya calculados con los ajustados con este modelo. Algunos de ellos se muestran en la tabla 3.4

TABLA 3.4. Tabla de eigenvalores ajustados y calculados

β=0		β=0.2		β=0.5		β=0.8	
ajust.	calc.	ajust.	calc.	ajust.	calc.	ajust.	calc.
0.5057	0.5050	0.5434	0.5424	0.6250	0.6192	0.7557	0.7508
0.1657	0.1683	0.1914	0.1961	0.2586	0.2641	0.4032	0.4185
0.0856	0.0842	0.1026	0.1021	0.1503	0.1501	0.2672	0.2644
0.0523	0.0505	0.0636	0.0627	0.0960	0.0969	0.1791	0.1731
0.0351	0.0337	0.0428	0.0425	0.0653	0.0672	0.1233	0.1182
0.0251	0.0241	0.0306	0.0307	0.0467	0.0489	0.0880	0.0843
0.0188	0.0180	0.0229	0.0232	0.0348	0.0369	0.0652	0.0624
0.0146	0.0140	0.0178	0.0181	0.0269	0.0287	0.0499	0.0478
0.0117	0.0112	0.0142	0.0145	0.0213	0.0229	0.0392	0.0375
0.0096	0.0091	0.0116	0.0119	0.0173	0.0186	0.0315	0.0301
0.0080	0.0076	0.0096	0.0099	0.0143	0.0154	0.0258	0.0246
0.0067	0.0064	0.0081	0.0083	0.0120	0.0129	0.0215	0.0205
0.0058	0.0055	0.0070	0.0071	0.0102	0.0100	0.0181	0.0172
0.0050	0.0048	0.0060	0.0062	0.0088	0.0094	0.0154	0.0147
0.0044	0.0042	0.0052	0.0054	0.0076	0.0082	0.0133	0.0126
0.0039	0.0037	0.0046	0.0047	0.0067	0.0071	0.0115	0.0110
0.0034	0.0033	0.0041	0.0042	0.0059	0.0063	0.0101	0.0096
0.0031	0.0029	0.0037	0.0037	0.0052	0.0056	0.0089	0.0084
0.0028	0.0026	0.0033	0.0034	0.0047	0.0050	0.0079	0.0075
0.0025	0.0024	0.0030	0.0030	0.0042	0.0045	0.0070	0.0067
0.0023	0.0021	0.0027	0.0027	0.0038	0.0040	0.0063	0.0060
0.0021	0.0020	0.0024	0.0025	0.0034	0.0037	0.0057	0.0054
0.0019	0.0018	0.0022	0.0023	0.0031	0.0033	0.0051	0.0049
0.0017	0.0016	0.0020	0.0021	0.0029	0.0030	0.0046	0.0044
0.0016	0.0015	0.0019	0.0019	0.0026	0.0028	0.0042	0.0040
0.0015	0.0014	0.0017	0.0018	0.0024	0.0026	0.0039	0.0037
0.0014	0.0013	0.0016	0.0016	0.0022	0.0024	0.0035	0.0034
0.0013	0.0012	0.0015	0.0015	0.0020	0.0022	0.0032	0.0031
0.0012	0.0011	0.0014	0.0014	0.0019	0.0020	0.0030	0.0029
0.0011	0.0010	0.0013	0.0013	0.0018	0.0019	0.0028	0.0026

3.4 PRUEBAS DE POTENCIA.

Para realizar las pruebas de potencia para esta estadística se consideraron algunas hipótesis alternativas que se alejaran poco a poco de la hipótesis nula. Ya que probar H_0 es equivalente a probar que la distribución especificada por esta hipótesis es Uniforme(0,1), o Beta(1,1). Las alternativas que se propusieron son betas con ligeros cambios en los parámetros (Beta(a1,a2)). Dentro de estas alternativas, consideramos aquéllas en las que cambia solamente el primer parámetro de forma, las que cambia el segundo parámetro de forma y en las que cambian los dos parámetros, dentro de estas últimas se propuso un grupo que mantuviera la simetría de la hipótesis nula. Para cada caso se simularon 1000 valores de la estadística y se contaron las veces que la prueba rechaza la hipótesis nula a niveles de significancia α=0.1 y α=0.05. Para obtener una aproximación a la potencia asintótica se tomó un tamaño de muestra de 500 en cada caso.

En primer lugar se consideraron alternativas para las cuales sólo varía el primer parámetro de forma, es decir, a1. En las gráficas 3.3(a) y 3.3(b) se observa el comportamiento de las funciones de densidad de las alternativas en este caso. Se analizaron diferentes valores de a1 y se encontró que para valores mayores a 1.3 y menores que 0.6 la potencia es prácticamente 1.00 para todos los valores de β elegidos. Los resultados se muestran en la tabla 3.5.

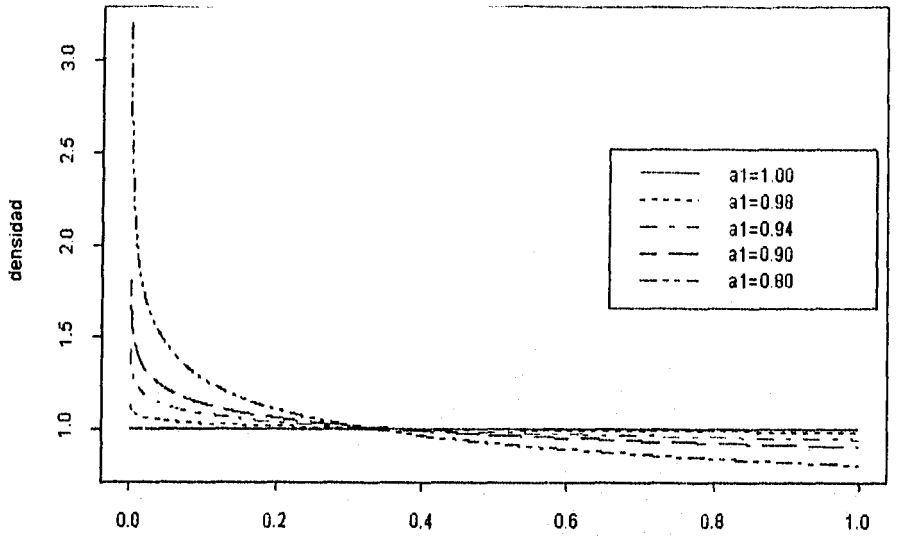
En esta tabla puede observarse que a medida que el valor del parámetro de forma se aleja del correspondiente de la hipótesis nula la potencia de la prueba aumenta. Además, a medida que el parámetro de censura β crece, que, como ya dijimos, implica un crecimiento en el porcentaje de censura esperado, la potencia de la prueba decrece en general. Este comportamiento es más evidente

conforme la hipótesis alternativa se aleja de la nula. Pueden observarse algunos casos en los que cuando el parámetro de censura crece no necesariamente decrece la potencia para alguno de los cuantiles considerados.

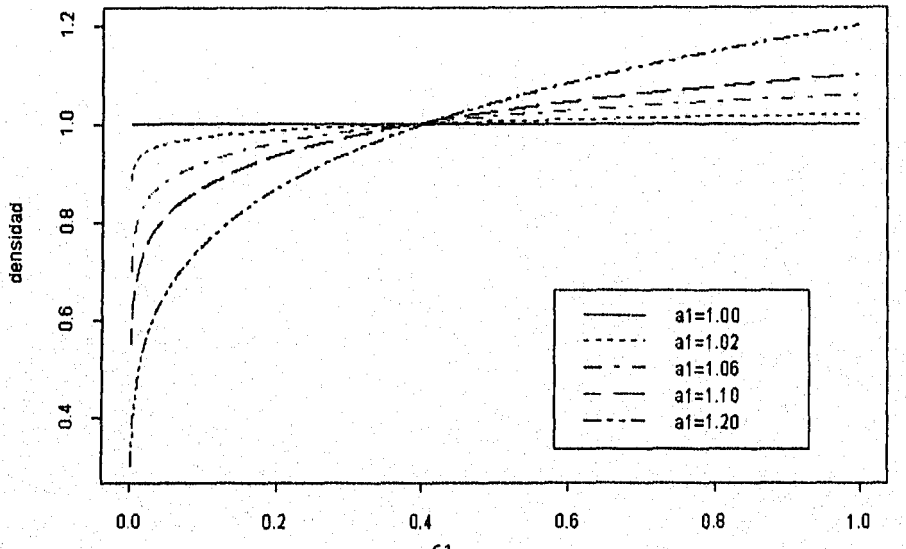
TABLA 3.5. Alternativas con recorrido en el primer parámetro de forma

Alternativa		β=0.0		β=0.2		β=0.4		β=0.6		β=0.8	
a1	a2	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
0.60	1.0	1000	1000	976	976	1000	1000	1000	1000	1000	1000
0.80	1.0	997	993	997	993	994	985	989	981	984	967
0.90	1.0	683	575	678	563	612	484	612	481	571	447
0.92	1.0	512	385	495	373	467	340	439	324	391	280
0.94	1.0	327	235	329	209	304	202	318	197	291	192
0.96	1.0	229	134	191	127	220	119	201	115	185	105
0.98	1.0	145	85	134	76	121	72	144	79	127	70
1.0	1.0	103	40	112	64	99	50	101	52	109	65
1.02	1.0	119	61	134	67	125	71	129	69	111	54
1.04	1.0	182	112	178	97	178	98	163	93	137	82
1.06	1.0	337	233	280	188	273	172	244	145	172	102
1.08	1.0	442	332	424	300	376	249	344	216	251	156
1.10	1.0	605	460	543	418	529	396	437	310	354	215
1.20	1.0	984	954	966	940	968	934	942	879	867	781
1.30	1.0	1000	1000	1000	999	999	997	1000	999	998	983

GRAFICA 3.3(a). ALTERNATIVAS CON RECORRIDO EN a1.



GRAFICA 3.3(b). ALTERNATIVAS CON RECORRIDO EN a1



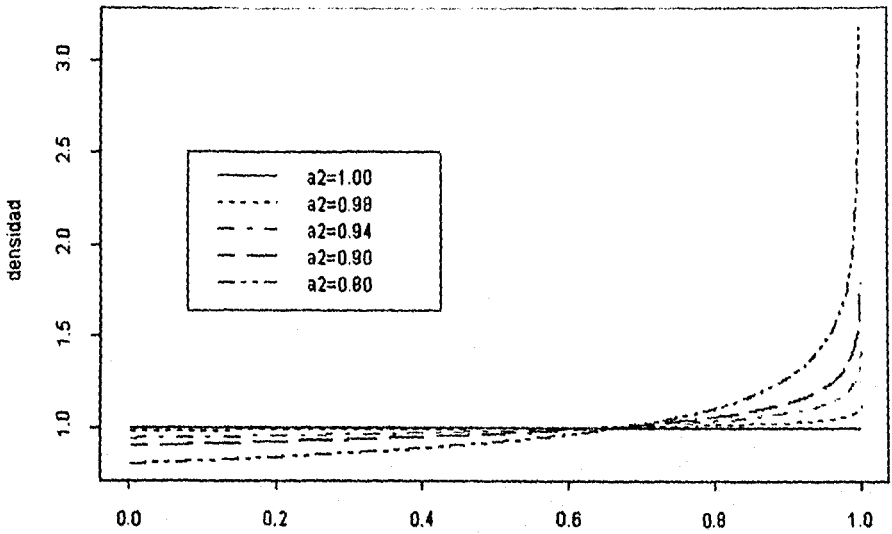
En segundo término, se tomaron alternativas en las que solamente variara el segundo parámetro de forma, a2. Las gráficas 3.4(a) y 3.4(b) muestran las funciones de densidad de estas alternativas. Se tomaron distintos valores de a2, y se observó que para valores menores que 0.8 y mayores que 1.3 se tiene una potencia aproximada de 1.00 para cualquiera de los valores de β elegidos. Para este caso, los resultados se muestran en la tabla 3.6.

TABLA 3.6. Alternativas con recorrido en el segundo parámetro de forma

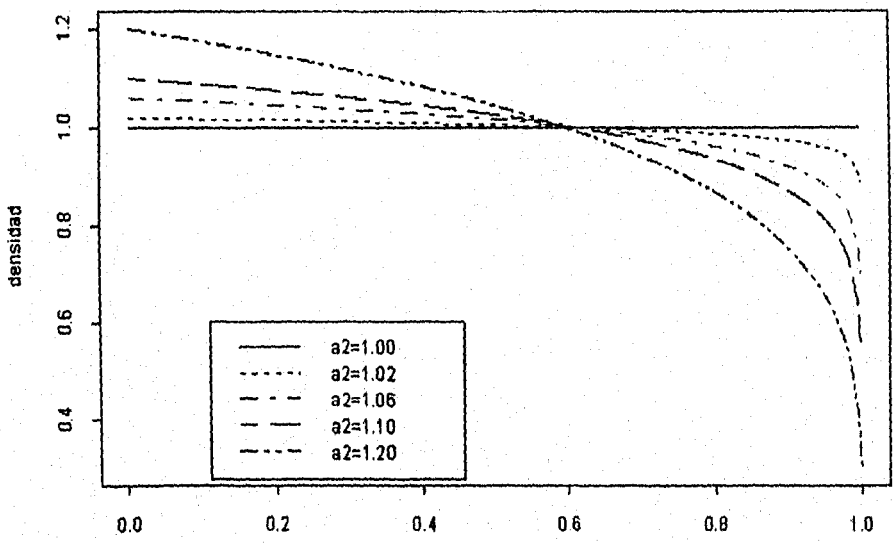
Alternativa		β=0.0		β=0.2		β=0.4		β=0.6		β=0.8	
a1	a2	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
1.0	0.80	999	994	996	989	989	976	977	965	951	907
1.0	0.90	684	567	654	546	617	490	528	413	471	328
1.0	0.92	499	384	469	343	449	332	390	280	300	216
1.0	0.94	326	209	329	225	286	192	276	190	224	133
1.0	0.96	191	118	216	131	202	126	153	87	136	80
1.0	0.98	134	75	124	68	142	73	124	72	104	53
1.0	1.0	103	40	112	64	99	50	101	52	109	65
1.0	1.02	132	71	115	59	145	84	129	72	130	72
1.0	1.04	205	122	202	120	176	94	164	93	147	91
1.0	1.06	293	199	321	221	278	183	270	174	257	159
1.0	1.08	433	305	407	301	398	283	381	269	357	256
1.0	1.10	563	445	572	450	518	393	502	375	473	363
1.0	1.20	972	950	969	931	955	913	937	898	914	848
1.0	1.30	1000	1000	997	997	1000	999	1000	998	998	991

Para este tipo de alternativas, se presentan comportamientos semejantes a los de la tabla 3.5, es decir, a medida que crece el parámetro de censura la potencia de la prueba decrece en general. Además, nuevamente se observa que hay algunos casos en los cuales no se sigue el comportamiento anterior.

GRAFICA 3.4(a). ALTERNATIVAS CON RECORRIDO EN a2.



GRAFICA 3.4(b). ALTERNATIVAS CON RECORRIDO EN a2.



El tercer grupo de alternativas que se consideró fueron aquéllas en las que ambos parámetros de forma variaran. Se hizo una leve variación en el primer parámetro a1, y se tomaron combinaciones con variaciones del segundo parámetro a2. En las gráficas 3.5(a), 3.5(b), 3.5(c) y 3.5(d) se muestran las densidades de estas alternativas. Se observó que para valores de a2 menores a 0.6 y mayores a 1.20 la prueba rechaza casi un 100% de las ocasiones para cualquier valor de β. En las tablas 3.7 y 3.8 se muestran estos resultados.

Cuando los dos parámetros cambian simultáneamente, la potencia de la prueba presenta un comportamiento muy parecido a los dos anteriores.

TABLA 3.7. Alternativas con recorrido en los dos parámetros de forma.

Alternativa		β=0.0		β=0.2		β=0.4		β=0.6		β=0.8	
a1	a2	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
0.98	0.60	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.98	0.80	996	990	989	979	984	971	953	920	897	851
0.98	0.90	549	424	550	426	486	373	418	294	312	214
0.98	0.92	387	257	387	265	331	230	282	180	244	155
0.98	0.94	257	164	237	149	222	152	183	111	146	88
0.98	0.96	156	90	139	70	165	88	148	79	121	73
1.0	1.0	103	40	112	64	99	50	101	52	109	65
0.98	1.02	198	131	187	101	199	136	209	122	168	101
0.98	1.04	306	211	283	182	265	166	238	154	273	177
0.98	1.06	462	336	423	325	382	260	378	270	369	250
0.98	1.08	597	450	577	460	523	402	511	372	459	359
0.98	1.10	730	611	696	561	649	536	617	500	625	501
0.98	1.20	994	987	987	974	978	955	971	944	947	904

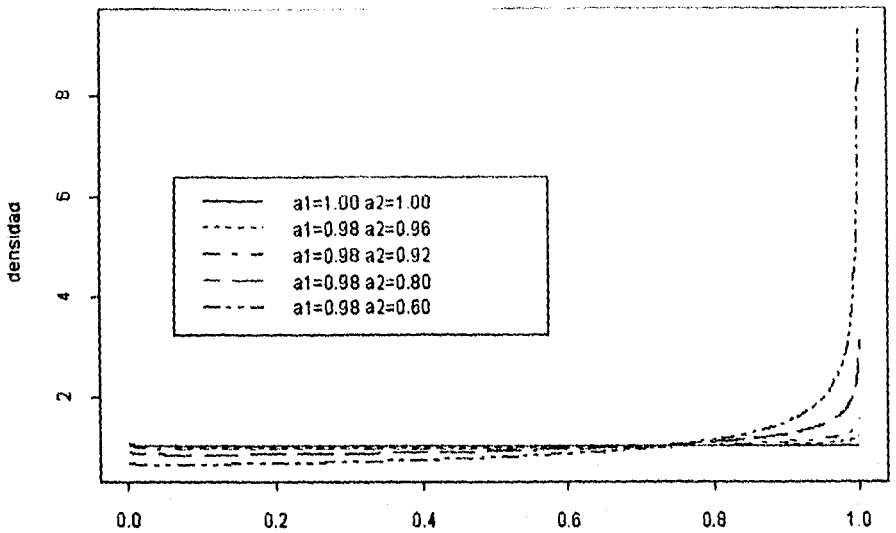
TABLA 3.8. Alternativas con recorrido de los dos parámetros de forma

Alternativa		β=0.0		β=0.2		β=0.4		β=0.6		β=0.8	
a1	a2	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
0.96	0.60	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.96	0.80	987	966	972	948	971	943	934	885	861	798
0.96	0.90	427	293	432	309	368	252	339	245	283	191
0.96	0.92	268	157	281	180	298	181	241	159	173	100
0.96	0.94	178	98	169	82	174	91	150	84	131	81
0.96	0.98	140	74	133	59	150	85	144	84	146	80
1.0	1.0	103	40	112	64	99	50	101	52	109	65
0.96	1.02	330	209	313	210	305	213	279	187	254	169
0.96	1.04	460	335	423	321	421	298	377	276	356	254
0.96	1.06	613	480	575	456	536	416	520	411	488	381
0.96	1.08	735	628	710	584	695	573	656	548	610	492
0.96	1.10	817	733	795	697	784	673	767	663	734	614
0.96	1.20	998	994	993	988	990	982	993	979	979	943

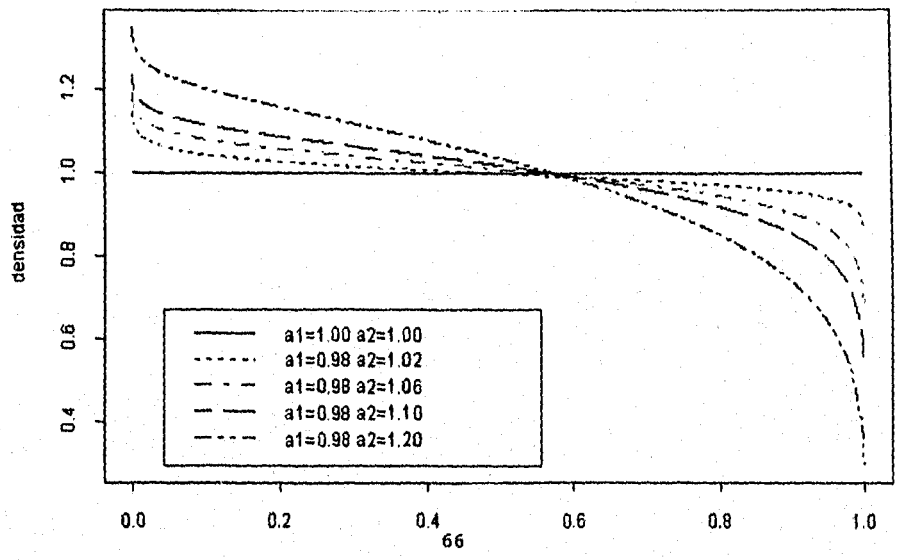
Finalmente, se consideraron alternativas que mantuvieran el comportamiento simétrico de la hipótesis nula. En las gráficas 3.6(a) y 3.6(b) se muestran las densidades correspondientes. En este caso, se observó que para valores de los dos parámetros menores que 0.6 y mayores de 1.5 se tiene una potencia casi de 1.00 para cualquiera de los valores de β presentados. En la tabla 3.9 se muestran los resultados correspondientes.

En este caso, es mucho más claro que cuando el parámetro de censura crece la potencia de la prueba decrece, solamente se presentan algunos comportamientos diferentes en las alternativas más cercanas a la hipótesis nula ((0.90,0.90) y (1.10,1.10)).

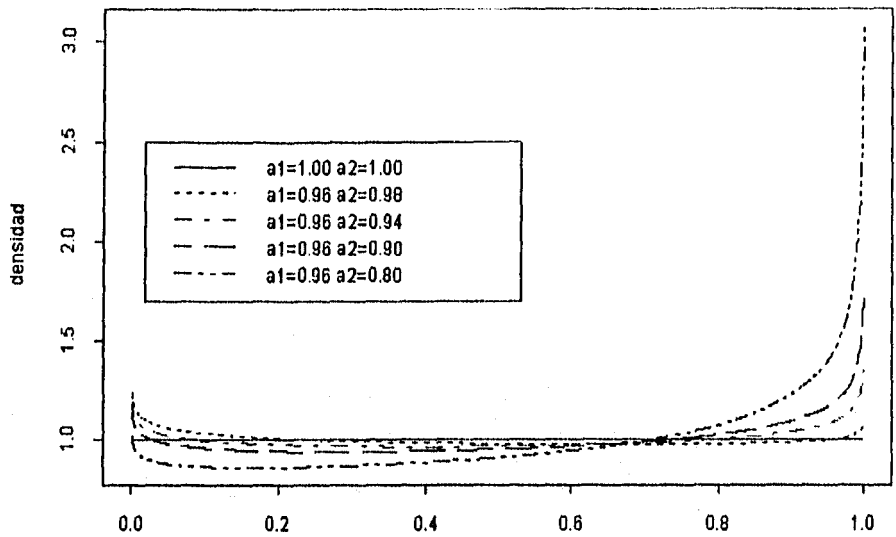
GRAFICA 3.5(a). ALTERNATIVAS CON RECORRIDO EN a1 Y a2.



GRAFICA 3.5(b). ALTERNATIVAS CON RECORRIDO EN a1 Y a2.



GRAFICA 3.5(c). ALTERNATIVAS CON RECORRIDO EN a1 Y a2.



GRAFICA 3.5(d). ALTERNATIVAS CON RECORRIDO EN a1 Y a2.

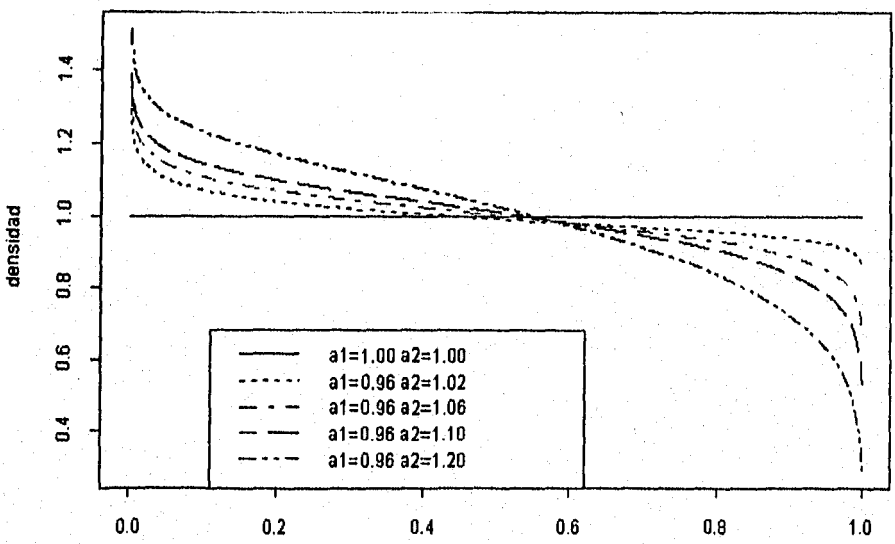
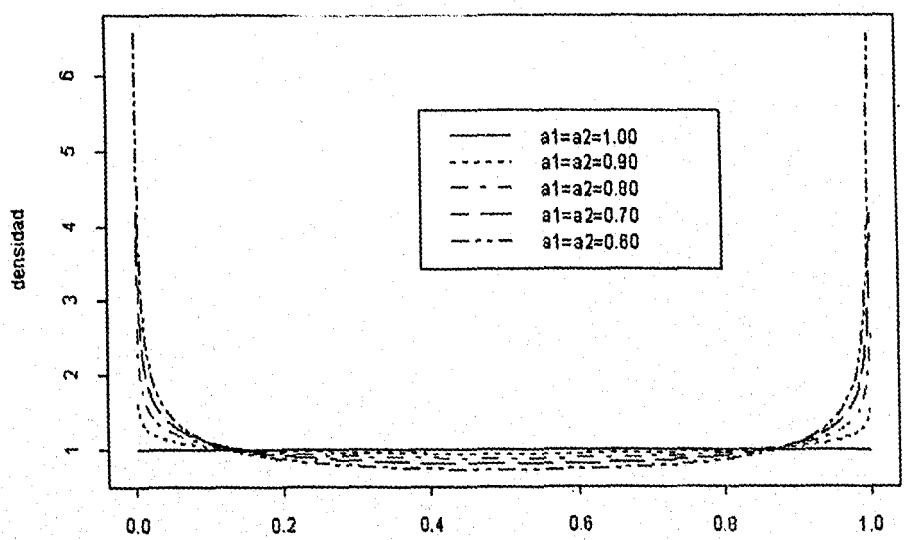


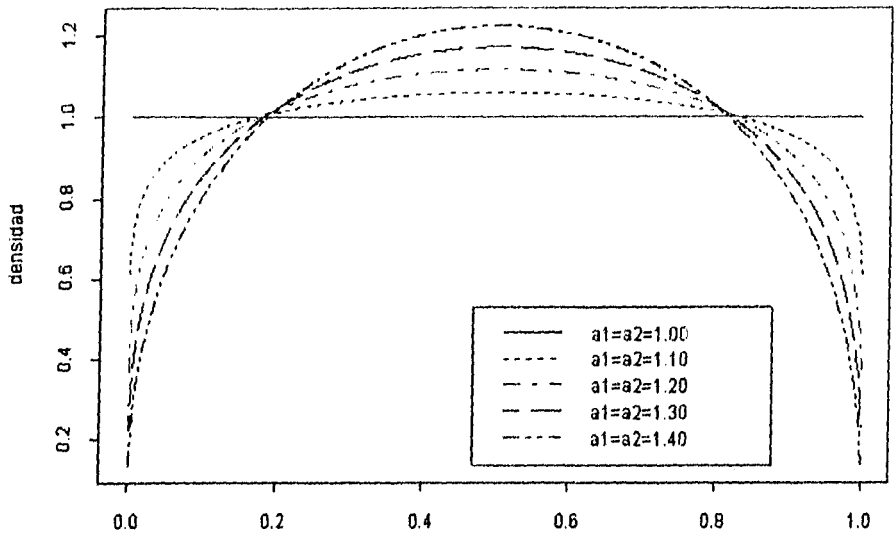
TABLA 3.9. Alternativas simétricas

Alternativa		β=0.0		β=0.2		β=0.4		β=0.6		β=0.8	
a1	a2	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
0.60	0.60	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
0.70	0.70	1000	1000	1000	998	998	987	988	973	975	925
0.80	0.80	901	792	850	724	831	685	749	606	620	471
0.90	0.90	317	178	268	145	307	181	275	164	214	121
1.0	1.0	103	40	112	64	99	50	101	52	109	65
1.10	1.10	206	96	205	99	200	97	185	87	147	70
1.20	1.20	618	381	567	348	511	317	432	277	348	190
1.30	1.30	933	815	896	763	842	688	779	598	654	449
1.40	1.40	998	987	990	969	980	947	953	885	874	762
1.50	1.50	1000	1000	1000	1000	999	993	998	989	968	922

GRAFICA 3.6(a). ALTERNATIVAS SIMETRICAS.

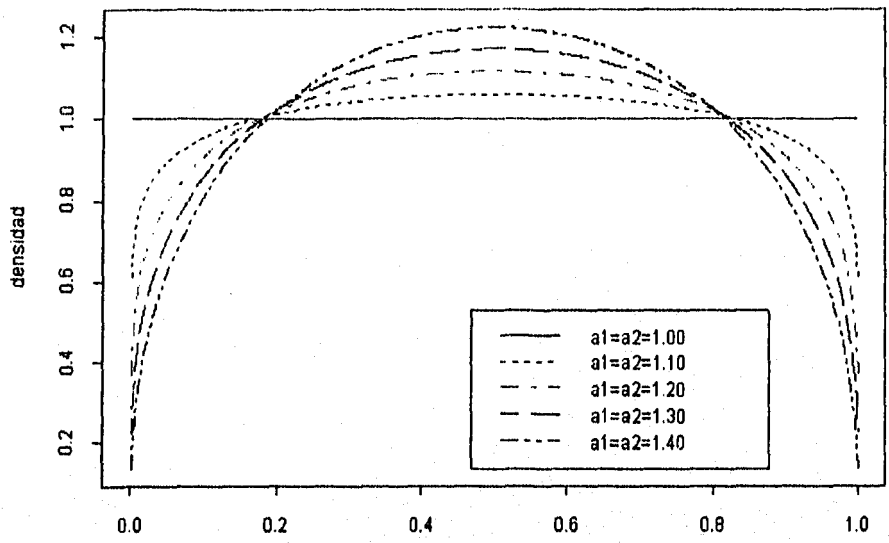


GRAFICA 3.6(b). ALTERNATIVAS SIMETRICAS.



De manera general, podemos observar que la potencia de la prueba decrece con el crecimiento del parámetro de censura para cualquiera de las alternativas consideradas. Este comportamiento es más evidente conforme la hipótesis alternativa se aleja de la nula. Los casos en los que no se sigue el comportamiento anterior usualmente se presentan para los valores del parámetro β=0.2 y β=0.4. Este tipo de comportamientos anómalos ya se habían presentado en el trabajo que desarrollaron Koziol y Green(1976) para la Cramér-von Mises con censura aleatoria.

GRAFICA 3.6(b). ALTERNATIVAS SIMETRICAS.



De manera general, podemos observar que la potencia de la prueba decrece con el crecimiento del parámetro de censura para cualquiera de las alternativas consideradas. Este comportamiento es más evidente conforme la hipótesis alternativa se aleja de la nula. Los casos en los que no se sigue el comportamiento anterior usualmente se presentan para los valores del parámetro β=0.2 y β=0.4. Este tipo de comportamientos anómalos ya se habían presentado en el trabajo que desarrollaron Koziol y Green(1976) para la Cramér-von Mises con censura aleatoria.

CAPITULO IV

CONCLUSIONES.

Como se mencionó en la introducción a este trabajo, el objetivo del mismo era, por una parte, dar a conocer las propuestas que se tienen para realizar pruebas de Bondad de Ajuste con datos que presentan algún tipo de censura. A este respecto, en el capítulo II se muestran de manera extensa varios procedimientos que se tienen para realizar dichas pruebas. En segundo término, y posiblemente el más importante, era proponer una modificación de la estadística de Anderson-Darling para hacer pruebas de Bondad de Ajuste para muestras con censura aleatoria. Sobre este último objetivo haremos los siguientes comentarios.

La propuesta de la estadística de prueba se realiza en el capítulo III y podemos observar que dicha estadística tiene un comportamiento estable, en el sentido de que, para los diferentes valores del parámetro de censura, la distribución de la estadística presenta un comportamiento muy regular (parecido al comportamiento observado por la Cramér-von Mises propuesta por Koziol-Green(1976)). Este comportamiento, además de los estudios de potencia que se realizaron, permiten afirmar que el uso de esta estadística para realizar las pruebas antes mencionadas, puede resultar adecuado, bajo las condiciones de que el parámetro de censura sea menor que uno y que se cumplan los supuestos sobre el modelo para el mecanismo de censura.

Este último punto, el modelo para el mecanismo de censura, hace que la propuesta no sea general. Sin embargo, lo que podemos

observar es que la distribución de la estadística depende del modelo del mecanismo de censura a través de la función de covarianza del proceso límite, por lo que esperaríamos que de tenerse un mecanismo de censura diferente al considerado aquí, la función de covarianza se modifique. No obtante esto, creemos que los caminos considerados para obtener la distribución de la estadística deben de seguir siendo válidos, y únicamente se deben hacer las modificaciones pertinentes para la nueva función de covarianza.

Como puede observarse, en este trabajo solamente se consideró el desarrollo de la estadística para el caso simple de Bondad de Ajuste (familia y parámetro(s) conocidos), resta abordar el caso más general en el cual los parámetros de la distribución a probar son total o parcialmente desconocidos. A este respecto, cabe mencionar que tampoco se conoce ningún trabajo en esta dirección para la Cramér-von Mises propuesta por Koziol-Green(1976). La importancia de desarrollar un trabajo de esta naturaleza es que en la práctica usualmente se desconocen estos parámetros.

Otra alternativa para seguir investigando en este tema, lo constituye la posibilidad de proponer modelos para distintos mecanismos de censura que puedan presentarse en algunas áreas del conocimiento en las que se trabaje con muestras censuradas. Para, posteriormente, desarrollar estas estadísticas bajo el modelo particular que se considere.

La estadística presentada en este trabajo es para hacer pruebas de Bondad de Ajuste para una muestra, por lo tanto, surge de manera natural la posibilidad de desarrollarla para hacer pruebas con más de una muestra.

Un problema relacionado con este tema, es encontrar una estadística de prueba para verificar si datos que presentan

censura, se ajustan a algún modelo, por ejemplo el modelo de Riesgos Proporcionales de Cox o el Modelo de Vida Acelerada.

Estos son solamente algunos temas de investigación en Bondad de Ajuste con muestras censuradas, pero como se comentó en la introducción falta mucho por realizarse aún.

ANEXO I

DESCOMPOSICION DE UN VECTOR ALEATORIO EN COMPONENTES PRINCIPALES

Dado un vector aleatorio Y(nx1), de media cero y matriz de varianza-covarianza Σ. Las componentes principales de Y son variables aleatorias dadas por

$$Z_j = \gamma_j' Y \quad j = 1,\ldots,n,$$

con γ_j el correspondiente eigenvector (ortonormal) asociado al eigenvalor λ_j, ambos son solución de la ecuación

$$\gamma'\Sigma = \lambda\gamma',$$

produciendo la descomposición de Σ dada por

$$\Sigma = \Gamma'\Lambda\Gamma = \sum_{j=1}^{n} \lambda_j \gamma_j \gamma_j'$$

donde Λ es una matriz diagonal con elementos $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$ y Γ es una matriz ortogonal con renglones $\gamma_1', \ldots, \gamma_n'$ que forman una base para el espacio de dimensión n.

La descomposición de Y en sus componentes principales es

$$Y = \sum_{j=1}^{n} Z_j \gamma_j,$$

o si denotamos por $Z_j^* = Z_j/\sqrt{\lambda_j}$ las componentes normalizadas, podemos escribir a Y como

$$Y = \sum_{j=1}^{n} \sqrt{\lambda}\, Z_j^{*} \gamma_j,$$

entonces, tenemos que

$$Y'Y = \sum_{j=1}^{n} \lambda_j Z_j^{*2}.$$

Finalmente, nótese que si $Y \sim N(0,\Sigma)$, entonces $Z_1^{*},\ldots,Z_n^{*}$ se distribuyen $N(0,1)$ independientes y, $Y'Y$ se distribuye como una suma ponderada de variables aleatorias independientes χ^2, donde las ponderaciones λ_j son los eigenvalores de la matriz de varianza-covarianza Σ y las Z_j's son las proyecciones de Y hacia los eigenvectores asociados

ANEXO II

En este anexo se encuentran los teoremas a los que se hace referencia a lo largo de la tesis.

TEOREMA DE MERCER. Si $K(s,t)$ es un kernel continuo para $0 \le s,t \le 1$ y es semidefinido positivo, entonces

$$K(s,t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \qquad 0 \le s,t \le 1,$$

donde la serie converge de manera uniforme y absoluta y además se tiene que $\sum_{j=1}^{\infty} \lambda_j < \infty$.

TEOREMA DE KAC Y SIEGERT. Supongamos que $K(s,t)$ es una función de covarianza de un proceso no trivial de media cero $\{X(t);\ 0 \le t \le 1\}$ cuyas trayectorias están c.s. en un subespacio L de funciones L_2, y supongamos que K satisface

$$\int_0^1 K(t,t)\,dt < \infty .$$

Supongamos también que $\lambda_1 \ge \lambda_2 \ge \ldots > 0$ comprende el espectro de eigenvalores de K, que las eigenfunciones asociadas $f_1, f_2, \ldots$ forman un conjunto completo para el subespacio L, y que tenemos la descomposición de Mercer

$$K(s,t) = \sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t) \qquad \text{válida para} \qquad 0 \le s,t \le 1 ,$$

entonces

i) $\sum_{j=1}^{\infty} \lambda_j < \infty$

ii) $Z_j = \langle \mathbf{X}, f_j \rangle$ son v.a $(0,\lambda_j)$ no correlacionadas

iii) $\sum_{j=1}^{m} Z_j f_j \overset{c.s}{\to} \mathbf{X}$ cuando $m \to \infty$

iv) $\int_0^1 \mathbf{X}^2(t)\,dt \underset{c.s.}{=} \sum_{j=1}^{\infty} Z_j^2 = \sum_{j=1}^{\infty} \lambda_j Z_j^{*2}$,

donde $Z_j^* = Z_j/\sqrt{\lambda_j}$ son v.a.(0,1) no correlacionadas

v) $E\left\{\mathbf{X}(t) - \sum_{j=1}^{m} Z_j f_j\right\}^2 \to 0$ para cada t cuando $m \to \infty$

vi) $\mathbf{X} \sim \sum_{j=1}^{\infty} \sqrt{\lambda_j}\, Z_j^* f_j$.

TEOREMA DE DURBIN-KNOTT. Las componentes principales normalizadas de $Z_n(t)$ son

i) $Z_{nj}^* = \lambda^{-1/2} \int_0^1 f_j(t) Z_n(t)\,dt = n^{-1/2} \sum_{k=1}^{n} \sqrt{2} \cos(j\pi\xi_k)$

con λ_j y f_j iguales a las obtenidas para W^2.

ii) $\left\{ Z_{nj}^{*} : j \geq 1 \right\}$ son v.a.i.i.d.(0,1) no correlacionadas

iii) $\sqrt{2}\cos(j\pi\xi_k) \sim \sqrt{2}\cos(\pi\xi_k) \sim (0,1)$ para toda $j \geq 1$

iv) $\sum_{j=1}^{m} \sqrt{\lambda_j} Z_{nj}^{*} f_j \to Z_n(t)$ cuando $m \to \infty$ para cada $0 \leq t \leq 1$

v) $Z_n(t) \sim \sum_{j=1}^{\infty} \sqrt{\lambda_j} Z_{nj}^{*} f_j = \sum_{j=1}^{\infty} \frac{Z_{nj}^{*} f_j}{j\pi}$

vi) $W_n^2 = \int_0^1 Z_n^2(t)dt = \sum_{j=1}^{\infty} \frac{Z_{nj}^{*2}}{(J\pi)^2}$.

TEOREMA DE HAMMERSTEIN. Sea $K(s,t)$ continua en el cuadrado unitario excepto posiblemente en las esquinas del cuadrado; sea $\partial K(s,t)/\partial s$ continua en el interior de los triángulos en que divide al cuadrado unitario la linea que une los puntos (0,0) y (1,1), y sea $\partial K(s,t)/\partial s$ acotada en el dominio $\varepsilon \leq s \leq 1-\varepsilon$ y $0 \leq t \leq 1$ para cada $\varepsilon > 0$. Entonces la serie

$$\sum_{j=1}^{\infty} \lambda_j f_j(s) f_j(t),$$

converge uniformemente a $K(s,t)$ en cualquier dominio en el interior del cuadrado unitario.

BIBLIOGRAFIA

ANDERSON,T.W. and DARLING,D.A.(1952). *Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes.* Ann. Math. Statist.,23,193-212.

BRESLOW,N. and CROWLEY,J.(1974). *A large sample study of the life table and product limit estimates under random censorship.* Ann. of Statistics, 2,num.3,437-453.

D'AGOSTINO,R.B. y STEPHENS,M.A.(1986).*Goodness-of-fit techniques.* Marcel Dekker,Inc.

DURBIN,J. and KNOTT,M.(1972). *Components of Cramér-von Mises Statistics I.* J. Roy. Statist. Soc.,B,34,216-237.

IMHOF,J.P.(1961). *Computing the distribution of quadratic forms in normal variables.* Biometrika,48,3 and 4,419-426.

KAC,M. and SIEGERT,A.J.F.(1947). *An explicit representation of a stationary Gaussian Process.* Ann.Math. Statist.,18,438-442.

KAPLAN,E.L. and MEIER,P.(1958). *Nonparametric estimation from incomplete observations.* J. Amer. Statist. Assoc.,53,457-481.

KOZIOL,J.A. and GREEN,S.B.(1976). *A Cramér-von Mises statistic for randomly censored data.* Biometrika,63,3,465-474.

PETTIT,A.N. and STEPHENS,M.A.(1976). *Modified Cramér-von Mises statistics for censored data.* Biometrika,63,2,291-298.

RUEDA DIAZ DEL CAMPO, R.(1995). Uso de la Distribución Predictiva en Bondad de Ajuste: Gaussiana Inversa. Tesis de Doctorado en Matemáticas. Universidad Nacional Autónoma de México.

SHORACK,G.R. and WELLNER,J.A.(1986). *Empirical processes with applications to statistics.* John Wiley & Sons.

STEPHENS,M.A.(1974a). *Components of goodness-of-fit statistics.* Ann. Inst. Henri Poincaré. Section B-vol.X, 1,37-54.

STEPHENS,M.A.(1974b). *EDF statistics for goodness-of-fit and some comparisons.* J. Amer. Statist. Assoc.,69,730-737.

WATSON,G.S.(1961). *Goodness-of-fit tests on a circle.* Biometrika, 8, 1 and 2,109-114.