INSTITUTO TECNOLOGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY

CAMPUS MONTERREY

DIVISION DE INGENIERIA Y ARQUITECTURA

UNA METODOLOGIA DE INVESTIGACION EN LA ESTIMACION DEL PARAMETRO DE FORMA DE LA DISTRIBUCION GAUSSIANA GENERALIZADA

TESIS

PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO ACADEMICO DE
MAESTRO EN CIENCIAS
CON ESPECIALIDAD EN SISTEMAS Y CALIDAD

POR

JUAN ANTONIO LOPEZ ESQUIVEL

ABRIL DEL 2000

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY
CAMPUS MONTERREY
DIVISIÓN DE INGENIERÍA Y ARQUITECTURA

UNA METODOLOGÍA DE INVESTIGACIÓN EN LA ESTIMACIÓN DEL PARÁMETRO DE FORMA DE LA DISTRIBUCIÓN GAUSSIANA GENERALIZADA

TESIS
PRESENTADA COMO REQUISITO PARCIAL PARA OBTENER EL GRADO ACADÉMICO DE MAESTRO EN CIENCIAS CON ESPECIALIDAD EN SISTEMAS Y CALIDAD POR

JUAN ANTONIO LÓPEZ ESQUIVEL

ABRIL DEL 2000

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY

CAMPUS MONTERREY

DIVISIÓN DE INGENIERÍA Y ARQUITECTURA

Los miembros del comité de tesis recomendamos que la presente tesis del Lic. Juan Antonio López Esquivel sea aceptada como requisito parcial para obtener el grado académico de maestro en ciencias con especialidad en

SISTEMAS Y CALIDAD

Comité de Tesis

Dra. Graciela González Farías
Asesor

Dr. Rogelio Ramos Quiroga
Sinodal

M.C. Sigfrido Iglesias González
Sinodal

APROBADO

Dr. Federico Viramontes Brown

Director del Programa de Graduados en Ingeniería

Abril del 2000

AGRADECIMIENTOS

Estoy en el momento propicio para retribuirles, sólo mínimamente, a las personas que han girado alrededor de mí (o alrededor de las cuales yo he girado) hasta el momento en que he llegado. Aunque únicamente en esta ocasión sea publicado mi sentir, les aseguro que en mi "Journal de páginas personales" es impreso diariamente.

Creo que lo que diga a continuación es sólo un mero acercamiento al mérito de aquellos en los que me he apoyado.

A veces he pensado que los problemas o situaciones complicadas con las que me enfrento, no son más que manifestaciones de los deseos de Dios, al enviarme siempre la mano indicada que me asistirá en tales casos. Esas manos que sin pensarlo dos veces, lo mismo se abren para exigir, que para sujetar, consolar, animar, y dar afecto. Aunque en ocasiones los impactos que se presentan en la vida son en verdad devastadores, siempre hay manos que intentan suavizar sus efectos. Vaya este pensamiento para todas aquellas personas que sin darse, o dándose cuenta, tuvieron un gesto de apoyo para con mi persona.

Quiero externar mi profundo y más sincero agradecimiento a Graciela María González Farías, conocida también como, La Dra. González (o simplemente Chela). Sus oportunos comentarios, su confianza, sus habilidades cognitivas y el afecto que me brindó, están guardados en mi mente y corazón. Reconozco que la Dra. González me mostró muchas de las maravillas del mundo de la Estadística; gracias por las provechosas discusiones que me otorgó. Gracias por sus infalibles recomendaciones. Gracias por ofrecerme su amistad, la cual espero seguir siendo honrado con ella.

Asimismo me congratulo de haber conocido a quien no menos gratitud me merece, el Dr. Rogelio Ramos Quiroga, un hombre con sensibilidad, del cual me he formado una imagen de persona, intachable, bondadosa, perseverante, luchador, y defensor incansable, con inteligencia, de sus convicciones. Tengo muchas personalidades a quien agradecer, pero en particular gracias a mis amigos Graciela y Rogelio (el Tec perdió a dos personalidades de la investigación ya que han dejado de laborar en este Instituto, ruego porque sigan viniendo personas de su misma calidad humana y científica).

Quiero agradecer también al Lic. Sigfrido Iglesias G., dudo mucho que este trabajo hubiese llegado a feliz termino sin su ayuda. Gracias Sig.

También quiero agradecer al Dr. Jorge H. Sierra C., quién me tomó en cuenta para entrar en contacto por primera vez con el mundo de la estadística en el departamento de matemáticas del ITESM.

Gracias a la Dra. Olivia Carrillo Gamboa, su alma bondadosa y serena me resulta un ejemplo continuo.

Gracias a los profesores de estadística del departamento de matemáticas: Dr. Christian Garrigoux, y Dra. Rebeca Romero, profesores que de una u otra forma han influenciado mi desarrollo profesional.

Gracias al Ing. Tomás Sánchez C. por el apoyo brindado a lo largo de la maestría, y la confianza que me manifiesta constantemente, por él he podido desarrollarme más en mi profesión. Gracias al Lic. Juan Manuel Silva, quien me dio la oportunidad de trabajar en el ITESM.

Gracias a mi esposa, Elvira, mi siempre bella mujer, mi infalible compañera, mi amada esposa, quien ha soportado buena parte de los sacrificios inherentes a mis actividades. Espero poder prodigarle en lo que me resta de vida el amor que siento por ella y a la vez recompensarle todo lo que me ha obsequiado. Dios nos ayude para tener la felicidad de seguir juntos en esta vida y gracias por permitirme tener la dicha de compartir los instantes con toda una mujer. Te amo Elvira.

Gracias al ITESM. Aquí me he realizado como profesional.

Gracias a maestros y personal administrativo del ITESM, quienes me han enseñado el significado de profesionalismo y dedicación. Gracias a muchos de mis alumnos por las lecciones que me han dado.

A mis padres. Mi cariño eterno para ellos. Gracias Papá y Mamá. Dios me permita mostrarles siempre mi cariño y gratitud. A mis hermanos. Espero que algún día nos juntemos los 9 en nuestra antigua, pero añorada casa.

Gracias a mis suegros, mis segundos padres, quienes me han abierto las puertas de su casa. Gracias a mis cuñados y cuñadas, en particular a Rosa, gracias comadre. A toda la familia de mi esposa que ahora es mi familia. Gracias a mi primo por convicción, Daniel, por su sensibilidad, don de gente y por brindarme una de las manifestaciones más puras del sentimiento humano, su amistad. Gracias a la comadre Amada quién no lo piensa dos veces para tendernos su mano amiga.

A mis hijos Juan Antonio y Josué Elías, mi amor. Sepan que los quiero y que pueden confiar en mí. Juanito, espero regresarte el tiempo que te he robado. Josué, lucha y no te dejes vencer ante las vicisitudes de la vida. Yo los cuidaré siempre que esté en mi mano.

Gracias a la Dra. Adriana, quién es de los pocos médicos que tiene corazón. Gracias por atender a mis hijos y gracias por su ayuda para con Josué.

Gracias a Todos.
Gracias a Dios.

Juan Antonio López Esquivel
ITESM Matemáticas

Índice General

1 INTRODUCCIÓN **3**

1.1 Propósitos del trabajo 3

1.2 Definición del problema y objetivos 5

1.3 Justificación y limitaciones 6

2 Antecedentes **8**

2.1 El método de momentos 8

2.2 El método de máxima verosimilitud 9

2.3 El MNS 10

2.4 Utilización de la desigualdad de Gurland 10

3 Expansión de Edgeworth **13**

3.1 Suma de variables aleatorias. 13

3.1.1 Convergencia y Cumulantes 14

3.1.2 Sumas de Variables Aleatorias 19

3.1.3 Expansión de Edgeworth 20

4 Estimador de Momentos de $Y = |X|$ e Intervalos de confianza para p **28**

4.1 Estimador de Momentos de $Y = |X|$ 28

4.2 Expansión de Edgeworth partiendo de la variable $|Y|$. 36

4.2.1 Intervalos de Confianza para M, y p 45

5 Estimación de p **61**

5.1 Curvas ajuste para $\gamma(p) = \frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}$ 61

5.1.1 Dos curvas de ajuste para $\gamma(p)$ 62

5.2 Simulación 69

6 Conclusiones 72

A Algoritmo para los intervalos de confianza para M y para p. Porcentaje de intervalos que cubren a los parámetros. 74

Capítulo 1

INTRODUCCIÓN

1.1 Propósitos del trabajo

En la rama de la ingeniería ocupada de los sistemas procesadores de señales, existe el problema del *ruido en las señales*, el cual es un factor común en el trabajo diario. Usualmente se asume se asume que el ruido es *ruido gaussiano*. Sin embargo, la experiencia muestra que es necesario modelarlo con una familia de distribuciones —modelos de probabilidad— más rica en cuanto a las formas características fundamentales asociadas. Tal familia puede ser, por ejemplo, la *distribución gaussiana generalizada*.

En lo que sigue, haremos una descripción general de esta familia de distribuciones y pondremos especial enfásis en la estimación de los parámetros de tal distribución, con la finalidad práctica de disponer de información adecuada para preparar los dispositivos adecuados que consideren tales fenómenos.

Así, uno de los propósitos de este trabajo consiste en efectuar una estimación del parámetro de forma de la distribución gaussiana generalizada. La estimación puntual se considera incompleta si no se proporciona información sobre la variabilidad de esa estimación, por lo cual, es de gran importancia la construcción de un *intervalo de confianza referente al parámetro p de la distribución gaussiana generalizada*. Este parámetro determina, en particular, la forma de la distribución. Para ciertos valores de p. se recuperan, por ejemplo, la distribución normal, y la distribución de Laplace.

Se propone una metodología para la construcción de dicho intervalo de

confianza, que produce coberturas aceptables y especifica el tamaño de muestra necesario para ello. La importancia que tiene el párametro de forma se hará evidente en el siguiente apartado.

Como parte de nuestra investigación, identificaremos los estimadores de momentos de la distribución gaussiana generalizada, su alcance y utilidad en la estimación del parámetro de forma. Se especifican las limitaciones en la estimación del parámetro de forma, al utilizar los estimadores de momentos, dando los órdenes respectivos de aproximación y encontrando los valores muestrales que dan el funcionamiento óptimo de los estimadores.

A su vez, mostraremos la utilidad de la expansión directa de Edgeworth, referente a la obtención de las distribuciones de probabilidad de los estimadores de interés.

Los tamaños de muestra son de primordial importancia desde el punto de vista ingenieril, dado que muchas decisiones de la planeación de captura y análisis de información dependen de esta selección. Así, encontramos los tamaños mínimos de muestra necesarios para la estimación de una función del parámetro p, y por lo tanto, como veremos más adelante, para el mismo parámetro, de acuerdo a las longitudes requeridas y válidos en el orden especificado.

Por otra parte, actualmente existe la tendencia de evitar los cálculos con ecuaciones, las cuales intervienen directamente en la estimación del parámetro de forma. En la sección 5.1.1, proponemos un algoritmo simple que prescinde de las ecuaciones trascendentales en la estimación del parámetro de forma. Este algoritmo se basa en el trabajo de Rodríguez y León [10].

La expansión directa de Edgeworth juega un papel importante en los análisis de los estimadores en los que intervienen sumas de variables aleatorias e incluso funciones de estas sumas, tales como las que aparecen en los procesos de estimación que se exponen en el presente trabajo. Por tal motivo, se presenta adicionalmente una solución haciendo uso de estas expansiones y se discuten los resultados obtenidos.

1.2 Definición del problema y objetivos

La función de densidad gaussiana generalizada (fdgg) es:

$$f_X(x) = \frac{1}{2\Gamma(1+1/p)A(p)} e^{-\left[\frac{|x-\mu|}{A(p)}\right]^p}, \quad x \in \Re,$$

donde, $A(p) = \left[\frac{\sigma^2\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}$; μ,σ, y p, son la media, la desviación estándar y la razón de decaimiento de la f.d.p.. En particular p es el parámetro de forma. Para evitar confusiones de aquí en adelante llamaremos σ_X a σ.

En ingeniería de señales, el ruido presente en la transmisión, suele ser modelado mediante una distribución gaussiana, de tal manera que los dispositivos se diseñan considerando tal modelo; para una introducción a estos procesos ver [11]. Sin embargo, en sistemas complejos en los que hay ambientes de ruido variable, es decir, donde la función de densidad no es únicamente la densidad normal, sino que puede tener otra forma, se hace imperativo modelar adecuadamente tal fenómeno. De hecho la familia de densidades de ruido que aparecen en los ambientes de señales queda modelada por la fdgg, con los valores respectivos de sus parámetros. Por lo anterior, las distribuciones a considerar pueden ser obtenidas de la familia de la fdgg. El problema principal radica en realizar una estimación adecuada de los parámetros de la fdgg, y con ello poder hacer frente a los distintos ambientes con los que se trabaja. El parámetro que nos atañe en este caso es el llamado parámetro de forma p, llamado así ya que indica la forma de la curva de que se trata. Para $p = 2$, por ejemplo, se tiene la distribución gaussiana (la normal), y para $p = 1$, la distribución de Laplace (doble exponencial). En otras palabras p es determinante en cuanto al peso de las colas: entre mayor es p (pensando sólo en la región de interés) menos pesadas las colas.

La estimación del parámetro de forma conlleva diversas dificultades [13], ya que por ejemplo, condiciones de eficiencias se dan únicamente para ciertos valores de p, también se han encontrado estimadores de máxima verosimilitud (con mejores propiedades), conociendo o no de antemano información

sobre uno o dos de los otros parámetros inmiscuidos, pero las ecuaciones resultantes suelen ser del tipo trascendentales, dificultando su manipulación numérica. En nuestro caso consideraremos que μ es cero (práctica común en la modelación del ruido aleatorio), por lo que la f.d.p., queda como

$$f_X(x) = \frac{1}{2\Gamma(1+1/p)A(p)} e^{-\left[\frac{|x|}{A(p)}\right]^p}, \quad x \in \Re,$$

esto produce una familia de curvas centradas en el origen (media cero). En las siguientes gráficas mostramos las curvas generadas por distintos valores del parámetro p de la fdgg, para $\sigma_X = 1$

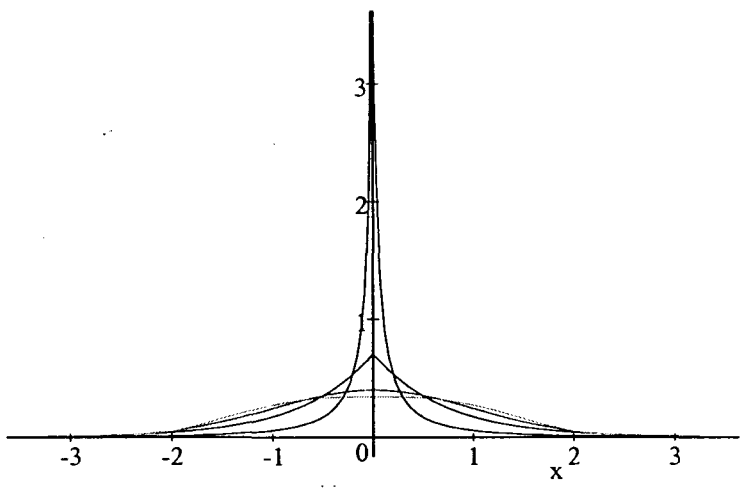

Figura 1. La fdgg, con $\sigma_x = 1$, para varios valores de p.

La curva más aguda corresponde al valor de $p = 0.3$, le sigue la que tiene el valor de $p = 1$, luego la que tiene el valor de $p = 2$, y finalmente la de colas más ligeras corresponde al valor de $p = 3$.

1.3 Justificación y limitaciones

Los resultados obtenidos se han validado primeramente mediante simulaciones computacionales, posteriormente se hará con datos reales provenientes

de la experimentación. También, en el desarrollo del trabajo especificamos las limitaciones, alcance y utilidad de los resultados. Además, pensamos que hemos aumentado el conocimiento que hasta el momento se tiene sobre el problema.

Capítulo 2

Antecedentes

Existen varios estudios acerca de la estimación de los parámetros de la fdgg, entre los más notables están los trabajos de Varana y Aazhang [13], al igual que los trabajos de Rodríguez y León [10] dirigidos estos últimos principalmente a la estimación del parámetro de forma de la distribución.

Una de las principales motivaciones para el estudio de la fdgg radica en que esta densidad modela, con los valores adecuados de los parámetros, el ruido aleatorio presente en los sistemas modernos de señales.

En este capítulo mostraremos las propuestas y condiciones de aplicación de los métodos encontrados por los autores recién mencionados en las estimaciones de los parámetros relevantes de la fdgg, y mostraremos en que dirección nos moveremos en relación con las propuesta existentes.

2.1 El método de momentos

En articulo de Varana y Aazhang [13], desarrollan tres métodos de estimación para los tres parámetros de la fdgg, estos son: el método de momentos, el método de máxima verosimilitud, y el de moment/Newton-step (MNS).

En lo sucesivo denotaremos, como es usual, a la variable aleatoria por mayúsculas: X, Y, etc., y a sus realizaciones con minúsculas: x, y. El método de momentos, para la estimación de los parámetros de la fdgg lleva, para $r \geq 4$, a,

$$\hat{\mu} = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{\mu})^2, \quad \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{\mu})^r = \frac{[A(\hat{p})]^r \Gamma\left(\frac{r+1}{\hat{p}}\right)}{\Gamma\left(\frac{1}{\hat{p}}\right)},$$

notemos que la última ecuación (obtenida del r-ésimo momento centrado en la media para $r \geq 4$) no puede ser resuelta explícitamente para p, y se hacen necesarias técnicas iterativas. El paquete scientific work place (SWP). proporciona soluciones a un bajo costo temporal, sin embargo no lo hace para todos los valores de p, y no es programable.

2.2 El método de máxima verosimilitud

En [13] desarrollan varios estimadores de máxima verosimilitud: (EMV): EMV de la media dados σ^2, y p; EMV de la varianza dados μ, y p; EMV del parámetro de forma dados μ, y σ^2, y EMV de cada uno de los tres parámetros.

Existen desventajas (similares al método de momentos) de cálculo en todos estos estimadores, ciertas restricciones en los valores de p en varios de ellos, además de intervenir en las ecuaciones tanto funciones gamma como funciones digamma. Claro que las propiedades de tales estimadores son mejores, en muchos aspectos, que aquellas de los estimadores de momentos. Omitiremos las expresiones de los EMV, sin embargo estas pueden ser consultadas en la referencia correspondiente [13].

La comparación del método de momentos y el EMV, muestra que para $p \leq 1.5$, y $p \geq 2.5$, el estimador de momentos tiene baja eficiencia y el EMV tiene mayores dificultades computacionales[13].

2.3 El MNS

En este método se plantea la solución de una serie de ecuaciones lineales y una no lineal en el caso de estimar los tres parámetros (las cuales omitiremos), en las que interviene la matriz de información y la función de verosimilitud para la obtención de los estimadores. Mencionaremos que los estimadores proporcionados por este método son asintóticamente eficientes. Aunque tiene ventajas computacionales, nuevamente los estimadores tienen formas complicadas en las que interviene funciones gamma a evaluar.

Ahora mostraremos la propuesta hallada en el trabajo de Rodríguez y León [10].

2.4 Utilización de la desigualdad de Gurland

En [10], es relevante la propuesta de evitar las funciones gamma en la estimación del parámetro de forma p (considerando de antemano que $\mu = 0$), haciendo uso de la función de razón gaussiana generalizada y de la desigualdad de Gurland:

Teorema. Función de razón generalizada gaussiana

$$\frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)} = \frac{E^2[|x|]}{\sigma_X^2}.$$

Teorema. Desigualdad de Gurland

$$\frac{\Gamma^2(\alpha+\delta)}{\Gamma(\alpha)\Gamma(\alpha+2\delta)} \leq \frac{\alpha}{\alpha+\delta^2},$$

de tal manera que cuando hacen $\alpha = 1/p$, y $\delta = 1/p$, se obtiene una desigualdad para la función de razón gaussiana generalizada:

$$\frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)} \leq \frac{p}{1+p}.$$

la cual junto con la función de razón gaussiana generalizada, produce el resultado,

$$\frac{E^2[\|x\|]}{\sigma_X^2} \leq \frac{p}{1+p},$$

haciendo natural la proposición,

$$\frac{\left(\frac{1}{N}\sum_i |x_i|\right)^2}{\frac{1}{N}\sum_i x_i^2} = \frac{p}{1+p},$$

para obtener un estimador puntual para p; notemos, sin embargo, que el estimador de la izquierda no es el estimador de momentos para p, (ver al comienzo de este capítulo), y por lo tanto no sabemos las propiedades de que goza. Otro detalle está en la aproximación de considerar la desigualdad como una igualdad, aunque la aproximación es notable en el intervalo típico de interés para p, es decir, $0.3 \leq p \leq 3$. Una gran ventaja es la facilidad del cálculo del estimador que resulta.

En el presente trabajo primeramente mostraremos que el estimador anterior es en realidad el estimador de momentos de la variable aleatoria $Y = |X|$ (capítulo 3), por lo que podemos establecer sin ningún problema sus propiedades básicas. Posteriormente haremos uso de la aproximación de Edgeworth (capítulo 5), como punto de partida para enfrentar el problema de estimar el parámetro p de la fdgg, pero es en el siguiente capítulo que introduciremos la expansión.

Capítulo 3

Expansión de Edgeworth

3.1 Suma de variables aleatorias.

La suma de variables aleatorias, a menudo independientes e idénticamente distribuidas (iid), suelen aparecer comúnmente en la práctica, por ejemplo, bajo ciertas condiciones [4] y[7]: la suma de variable aleatorias geométricas produce una variable aleatoria binomial negativa, la suma de variables aleatorias exponenciales produce una gamma, la suma de variables aleatorias de una muestra aleatoria produce la media muestral y es de interés primordial conocer su comportamiento distribucional, etc., notemos que en los dos primeros casos se puede obtener en forma exacta la distribución de probabilidad de la suma resultante, pero en el tercer caso, a menos que se sepa algo sobre la población (en particular que sea normal), es practicamente imposible conocer la distribución de la suma resultante. De hecho, en el último caso es cuando se empieza a hacer uso de teoremas como el teorema del límite central [5], y [7]. Por lo anterior, resulta de suma utilidad, el poder encontrar al menos aproximadamente la función de densidad de la suma de las variables aleatorias de interés. En este capítulo mostraremos la teoría básica de la expansión de Edgeworth, para la obtención de funciones de densidad de sumas de variables aleatorias.

3.1.1 Convergencia y Cumulantes

Antes de mostrar la expansión de Edgeworth, introduciremos brevemente algunos conceptos sobre convergencia y cumulantes estandarizados [6].

Ordenes

Consideremos primero una sucesión de números reales $\{a_n\}$, diremos que:

1.- a_n es de orden uno: $a_n = O(1)$, si existe un número N tal que $|a_n| < N$ (sucesión limitada).

2.- a_n converge a cero: $a_n = o(1)$, si $a_n \longrightarrow 0$, cuando $n \longrightarrow \infty$.
Observación: si $a_n = o(1)$, entonces $a_n = O(1)$.

3.- Sea f_n una sucesión de números positivos, entonces $a_n = O(f_n)$ si $\frac{a_n}{f_n} = O(1)$, *i.e.* $|a_n| < Nf_n$; la sucesión a_n está limitada por la sucesión f_n.Observación: si $a_n = O(\frac{1}{n})$, entonces $a_n = o(1)$.

4.- $a_n = o(f_n)$ si $\frac{a_n}{f_n} = o(1)$, *i.e.* $\frac{a_n}{f_n} \longrightarrow 0$,cuando $n \longrightarrow \infty$.

5.- Si $a_n \longrightarrow a < \infty$, entonces $a_n = O(1)$.

6.- Si $a_n = O(f_n)$, entonces $a_n^m = O(f_n^m)$, para $m > 0$.

Ejemplos.

1.- $a_n = \frac{10n}{n+2} = 10\frac{n+2-2}{n+2} = 10 - \frac{20}{n+2} = O(1)$, en este caso N podría ser 11.

2.- $a_n = \frac{10n}{n^2+2} = \frac{10}{n+\frac{2}{n}} \longrightarrow 0$, cuando $n \longrightarrow \infty$, por lo tanto $a_n = o(1)$. En este mismo ejemplo notemos que $\frac{a_n}{\frac{1}{n}} = na_n = \frac{10n^2}{n^2+2} = 10 - \frac{20}{n^2+2} = O(1)$ esto implica que $a_n = O(\frac{1}{n})$, y por lo tanto $o(1)$.

Sea ahora $\{X_n\}$ una sucesión de variables aleatorias definidas en un espacio probabilístico, diremos que:

1.- X_n está límitada en probabilidad: $X_n = O_p(1)$, si dada $\epsilon > 0$, existe un $N_\epsilon < \infty$, tal que $P(-N_\epsilon \leq X_n \leq N_\epsilon) > 1 - \epsilon$, para todo valor de n, *i.e.*, X_n está limitada por N_ϵ, para todo valor de n, con una muy alta pobabilidad.

2.- X_n converge en probabilidad: $X_n = o_p(1)$, si dadas $\epsilon > 0$, $\delta > 0$, existe $N_{\epsilon,\delta}$, tal que $P(-\delta \leq X_n \leq \delta) > 1 - \epsilon$, para todo valor de $n > N_{\epsilon,\delta}$, se denota como $X_n \xrightarrow{P} 0$.

3.- En forma similar a las sucesiones de variables aleatorias, tenemos, con f_n una sucesión de números positivos, que,

i) $X_n = O_p(f_n)$ si $\frac{X_n}{f_n} = O_p(1)$. ii) $X_n = o_p(f_n)$ si $\frac{X_n}{f_n} = o_p(1)$.

4.- Con $m > 0$, i) si $X_n = O_p(f_n)$, entonces $|X_n^m| = O_p(f_n^m)$. ii) si $X_n = o_p(f_n)$, entonces $|X_n^m| = o_p(f_n^m)$.

5.- Para $X_n > 0$, si $E(X_n) = O(f_n)$, donde E, indica el valor esperado, entonces $X_n = O_p(f_n)$.

6.- $E(|X_n|) = O(f_n) \Longrightarrow X_n = O_p(f_n)$.

7.- Si $E(X_n^2) = O(f_n)$, entonces $X_n^2 = O_p(f_n^2)$, y $X_n = O_p(f_n)$.

8.- Si $E(X_n) = O(a_n), V(X_n) = O(b_n)$, donde E, V,son el valor esperado y la varianza, entonces,$E(X_n^2) = O(b_n) + O(a_n^2) = O(\text{máximo } (b_n, a_n^2)$, de donde, $X_n^2 = O_p(\sqrt{máximo(b_n, a_n^2)}$.

9.- Convergencia en distribución. La sucesión $\{X_n\}$ converge en distribución a la función de distribución acumulada $F(x)$ si para todo punto en que F es continua, $\lim_{n \longrightarrow \infty} P(X_n \leq x) = F(x)$.

Cumulantes

De aquí en adelante consideraremos variables aleatorias iid. De teoría básica de estadística sabemos que los momentos de una variable aleatoria pueden ser obtenidos a partir de su función generatriz de momentos en los casos que ésta exista[4] y[7]. Más específicamente, los momentos r-ésimos, respecto al origen y respecto a la media, respectivamente, de una variable aleatoria, definidos mediante los siguientes valores esperados,

$$\acute{\mu_r} = E(Y^r), \qquad \mu_r = E(Y - \mu)^r, \qquad \text{para } r = 1, 2, 3, \ldots,$$

están conectados con la función generatriz de momentos, definida por,

$$M_Y(t) = E(e^{tY}),$$

de la siguiente manera:

$$\mu_r^{'} = \frac{d^r}{dt^r}(M_Y(t))|_{t=0};$$

notemos que $M_{Y-\mu}(t) = e^{-\mu t}M_Y(t)$. También mostramos el desarrollo en serie de Maclaurin, para $M_Y(t)$, y $M_{Y-\mu}(t)$, válida en un cierto intervalo de valores reales de t. Las series correspondientes son:

$$\begin{aligned} M_Y(t) &= 1 + \sum \frac{\mu_r^{'}}{r!}t^r \\ M_{Y-\mu}(t) &= 1 + \sum \frac{\mu_r}{r!}t^r, \end{aligned}$$

de hecho, mediante estas series se suele mostrar la conexión entre los momento y las derivadas de la función generatriz de momentos. Con el fin de llegar a la definición de los cumulantes introduciremos la función generatriz cumulante [2] $K_Y(t)$:

$$K_Y(t) = \ln(M_Y(t)),$$

de la definición vemos que $K_{Y-\mu}(t) = \ln(M_{Y-\mu}(t)) = \ln(e^{-\mu t}M_Y(t)) = -\mu t + \ln M_Y(t), = -\mu t + K_Y(t)$, o equivalentemente

$$K_Y(t) = \mu t + \ln(M_{Y-\mu}(t)).$$

A continuación definimos los cumulantes de la variable aleatoria Y.

El r-ésimo cumulante de Y, κ_r, se define mediante la serie de Maclaurin de $K_Y(t)$:

$$K_Y(t) = \sum_{r=1}^{\infty} \kappa_r \frac{t^r}{r!}.$$

Se puede encontrar una relación entre los momentos y los cumulantes de la variabe aleatoria Y,utilizando el resultado $K_Y(t) = \ln(M_Y(t))$, y el desarrollo $M_{Y-\mu}(t) = 1 + \sum \frac{\mu_r}{r!} t^r$,

$$\begin{aligned} \sum_{r=1}^{\infty} \kappa_r \frac{t^r}{r!} &= \mu t + \ln(M_{Y-\mu}(t)) \\ &= \mu t + \ln(1 + \sum \frac{\mu_r}{r!} t^r), \end{aligned}$$

si en esta última relación, desarrollamos el logaritmo, tenemos que,

$$\kappa_1 t + \kappa_2 \frac{t^2}{2!} + \kappa_3 \frac{t^3}{3!} + \kappa_4 \frac{t^4}{4!} + \cdots = \mu t + \sum \frac{\mu_r}{r!} t^r - \frac{1}{2} \left(\sum \frac{\mu_r}{r!} t^r \right)^2 + \frac{1}{3} \left(\sum \frac{\mu_r}{r!} t^r \right)^3 + \cdots,$$

de aquí, recordando que $\mu_1 = 0$,

$$\kappa_1 = \mu, \qquad \kappa_2 = \mu_2 = \sigma^2, \qquad \kappa_3 = \mu_3, \qquad \kappa_4 = \mu_4 - 3\mu_2^2,$$

etc., una propiedad importante de los cumulantes es, con $a =$ cte.:

$$\kappa_r(aY) = a^r \kappa_r(Y),$$

es decir, el r-ésimo cumulante de la variable aleatoria transformada aY es a^r por el cumulante de la variable original.

Aún más importante por su carácter adimensional son los cumulantes estandarizados:

Definición. Los cumulantes estandarizados de la variable aleatoria Y, son

$$\rho_r = \frac{\kappa_r}{\sigma^r}, \text{ para } r = 3, 4, \ldots.$$

En particular, ρ_3 es el coeficiente de asimetría, y ρ_4 es la curtosis.

3.1.2 Sumas de Variables Aleatorias

Consideremos una suma de variables aleatorias (iid) con media μ y varianza σ^2,

$$S_n = Y_1 + Y_2 + \cdots + Y_n.$$

Las expansión de Edgeworth [2], no se aplica directamente a la suma S_n, sino a su forma estandarizada, la forma típica de estandarizar es,

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$

entre las ventajas de esta transformación, tenemos que el valor esperado de S^*, es cero, y su varianza uno, además se evita trabajar con distribuciones límite no degeneradas.

Resultados importantes [2] en los que intervienen los cumulantes de S, S^*, y Y, son:

$$\begin{aligned}
\kappa_r(S_n) &= n\kappa_r(Y), \quad \text{para } r = 1, 2, \ldots, \\
\rho_r(S_n^*) &= \rho_r(S_n), \quad \text{para } r = 3, 4, \ldots, \\
\rho_r(S_n) &= \frac{\rho_r(Y)}{n^{\frac{r}{2}-1}}, \quad \text{para } r = 2, 4, \ldots.
\end{aligned}$$

Ahora sí estamos en el momento justo para desarrollar la expansión de Edgeworth.

3.1.3 Expansión de Edgeworth

Sean $Y_1, Y_2, \ldots, Y_n$, iid, con $E(Y_i) = \mu$, y $V(Y_i) = \sigma^2$, y consideremos la suma estandarizada,

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$

su función generatriz cumulante es,

$$\begin{aligned}
K_{S^*}(t) &= K_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) \\
&= K_{\frac{S_n}{\sigma\sqrt{n}} - \frac{n\mu}{\sigma\sqrt{n}}}(t),
\end{aligned}$$

y recordando que $K_{Y-\mu}(t) = -\mu t + K_Y(t)$, obtenemos,

$$\begin{aligned}
K_{S^*}(t) &= -\frac{n\mu}{\sigma\sqrt{n}} + K_{\frac{S_n}{\sigma\sqrt{n}}}(t) \\
&= -\frac{\sqrt{n}\mu}{\sigma} + K_{\frac{1}{\sigma\sqrt{n}}S_n}(t) \\
&= -\frac{\sqrt{n}\mu}{\sigma} + \ln(M_{\frac{1}{\sigma\sqrt{n}}S_n}(t)) \\
&= -\frac{\sqrt{n}\mu}{\sigma} + \ln(M_{S_n}(\frac{t}{\sigma\sqrt{n}})) \\
&= -\frac{\sqrt{n}\mu}{\sigma} + K_{S_n}(\frac{t}{\sigma\sqrt{n}}) \\
&= -\frac{\sqrt{n}\mu}{\sigma} + nK_Y(\frac{t}{\sigma\sqrt{n}}),
\end{aligned}$$

desarrollemos $K_Y(t)$:

$$\begin{aligned}
K_Y(t) &= \kappa_1 t + \kappa_2\frac{t^2}{2!} + \kappa_3\frac{t^3}{3!} + \kappa_4\frac{t^4}{4!} + \cdots \\
&= \mu t + \sigma^2\frac{t^2}{2} + \rho_3\sigma^3\frac{t^3}{6} + \rho_4\sigma^4\frac{t^4}{24} + \cdots,
\end{aligned}$$

de aquí que la expresión que resulta para $K_{S^*}(t)$, es,

$$K_{S^*}(t) = -\frac{\sqrt{n}\mu}{\sigma} + nK_Y(\frac{t}{\sigma\sqrt{n}})$$

$$
\begin{aligned}
&= -\frac{\sqrt{n}\mu}{\sigma} + n(\mu\frac{t}{\sigma\sqrt{n}} + \sigma^2\frac{\left(\frac{t}{\sigma\sqrt{n}}\right)^2}{2} + \rho_3\sigma^3\frac{\left(\frac{t}{\sigma\sqrt{n}}\right)^3}{6} + \\
&\quad \rho_4\sigma^4\frac{\left(\frac{t}{\sigma\sqrt{n}}\right)^4}{24} + \cdots) \\
&= \frac{t^2}{2} + \rho_3\frac{t^3}{6n^{\frac{1}{2}}} + \rho_4\frac{t^4}{24n} + O(n^{-\frac{3}{2}}).
\end{aligned}
$$

Ya obtenida la función generatriz cumulante, podemos obtener la función generatriz de momento en términos de los cumulantes estandarizados ρ_r, tomando el exponencial en la última expresión,

$$
\begin{aligned}
M_{S^*}(t) &= \exp(\frac{t^2}{2})\exp(\rho_3\frac{t^3}{6n^{\frac{1}{2}}} + \rho_4\frac{t^4}{24n} + O(n^{-\frac{3}{2}})) \\
&= \exp(\frac{t^2}{2})\{1 + \rho_3\frac{t^3}{6n^{\frac{1}{2}}} + \rho_4\frac{t^4}{24n} + \rho_3^2\frac{t^6}{72n} + O(n^{-\frac{3}{2}})\},
\end{aligned}
$$

la última expresión se obtuvo desarrollando el exponencial en serie de Maclaurin. El primer término,$\exp(\frac{t^2}{2})$, es llamado el término guía, y en este caso corresponde a la función generatriz de momentos de la normal estándar.

A partir de esta expresión se puede realizar la inversión en el sentido de obtener la función de densidad de probabilidad de la cual $M_{S^*}(t)$ es su función generatriz de momentos. Para ello necesitamos introducir los llamados polinomios ortogonales de Hermite, de amplio uso en las ciencias, en cuanto a desarrollo en series de funciones ortogonales se refiere.

Polinomios ortogonales de Hermite.

$$H_0 = 1,\ H_1 = x,\ H_2 = x^2 - 1,\ H_3 = x^3 - 3x, \text{ etc.},$$

existe en la bibliogafía un amplio estudio de estos polinomios, para una fórmula de obtención de cualquiera de los polinomios, ver por ejemplo [1], y [3]. Consideramos que no es de vital importancia dar la demostración rigurosa de la inversión de $M_{S^*}(t)$, en nuestro caso bastará con observar, haciendo uso del resultado [2],

$$\int_{-\infty}^{\infty} e^{tx} e^{(-\frac{1}{2}x^2)} H_r(x) dx = t^r e^{\frac{1}{2}t^2}.$$

Si calculamos la función generatriz de momentos del valor esperado de la expresión,

Proposición. Expansión de Edgeworth

$$f_{S^*}(x) = e^{(-\frac{1}{2}x^2)}\{1 + \rho_3 \frac{H_3(x)}{6n^{\frac{1}{2}}} + \rho_4 \frac{H_4(x)}{24n} + \rho_3^2 \frac{H_6(x)}{72n}\} + O(n^{-\frac{3}{2}}),$$

vemos que recuperamos la expresión de la generatriz de momentos. Hemos presentado esta forma de ver la expansión en vista de que una demostración rigurosa utilizaría la función característica [8], y esto escapa a las intenciones de este trabajo.

Comentarios [2]:

1.- También se puede obtener una expresión para la función de distribución acumulada de S^*.

2.- $e^{(-\frac{1}{2}x^2)}\{1 + \rho_3 \frac{H_3(x)}{6n^{\frac{1}{2}}}\}$ puede no representar una función de densidad para pequeños valores de n, ya que H_3 sería negativo para pequeños valores de x.

3.- Existe una inversión sobre la función de densidad que reproduce una expresión para la variable aleatoria S^*

Ejemplo. Sean $Y_1, Y_2, \ldots, Y_n$, exponenciales con parámetro $\beta = 1$, entonces la suma

$$S_n = Y_1 + Y_2 + \cdots + Y_n,$$

tiene la distribución de probabilidad

$$f_{S_n}(x) = \frac{1}{\Gamma(n)1^n} x^{n-1} e^{-\frac{x}{1}} = \frac{x^{n-1}e^{-x}}{(n-1)!}, \quad \text{para } x \geq 0,$$

es decir, la suma de n variables aleatorias exponenciales, con $\beta = 1$, iid, produce una Gamma $(n, 1)$. En este caso $E(S_n) = n$, y $V(S_n) = n$, también se pueden obtener los cumulantes estandarizados a partir de su relación con los momentos, los cuales se pueden calcular , ver [4], de $E(S_n)^r = 1^r \frac{\Gamma(n+r)}{\Gamma(n)}$, mediante algunos calculos se encuentra que $\rho_3 = 2, \rho_4 = 6$. Entonces la suma estandarizada, $S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}$, tiene la expansión de Edgeworth,

$$\begin{aligned} f_{S_n^*}(x) &= e^{(-\frac{1}{2}x^2)}\{1 + \rho_3 \frac{H_3(x)}{6n^{\frac{1}{2}}} + \rho_4 \frac{H_4(x)}{24n} + \rho_3^2 \frac{H_6(x)}{72n}\} + O(n^{-\frac{3}{2}}) \\ &= e^{(-\frac{1}{2}x^2)}\{1 + \frac{H_3(x)}{3n^{\frac{1}{2}}} + \frac{H_4(x)}{4n} + \frac{H_6(x)}{18n}\} + O(n^{-\frac{3}{2}}), \end{aligned}$$

mediante alguno de los métodos de transformación [7] se puede encontrar que la función de densidad exacta de la suma estandarizada es,

$$\sqrt[2]{n}(n + x\sqrt[2]{n})^{n-1} e^{\frac{-n-x\sqrt[2]{n}}{(n-1)!}}.$$

Entonces podemos comparar la expansión de Edgeworth con la densidad exacta, hagámoslo para términos de orden $O(n^{\frac{1}{2}})$, y términos de $O(n^{-1})$ entonces, sustituyendo H_3, H_4, y H_6, tenemos,

$$f_{S_n^*}(x) = \frac{1}{\sqrt[2]{2\pi}} e^{-\frac{x^2}{2}} (1 + \frac{(x^3-3x)}{3\sqrt[2]{n}}) + O(n^{-1}),$$

y

$$f_{S_n^*}(x) = \frac{1}{\sqrt[2]{2\pi}} e^{-\frac{x^2}{2}} (1 + \frac{(x^3-3x)}{3\sqrt[2]{n}} + \frac{(x^4-6x^2+3)}{4n} + \frac{(x^6-15x^4+45x^2-15)}{18n}) + O(n^{-\frac{3}{2}}),$$

podríamos hacer una comparación numérica, pero preferimos realizar las gráficas correspondientes con $n = 10$:

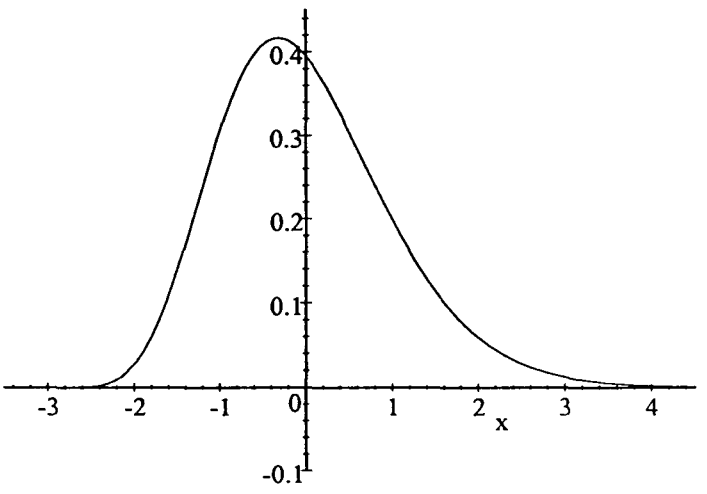

Figura 2. Distribución exacta de S^*.

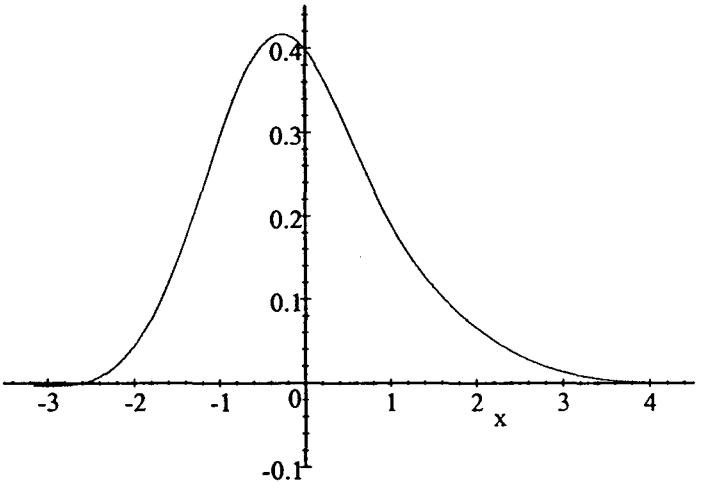

Figura 3 Aproximación de Edgeworth de orden $\frac{1}{\sqrt{n}}$,con $n = 10$.

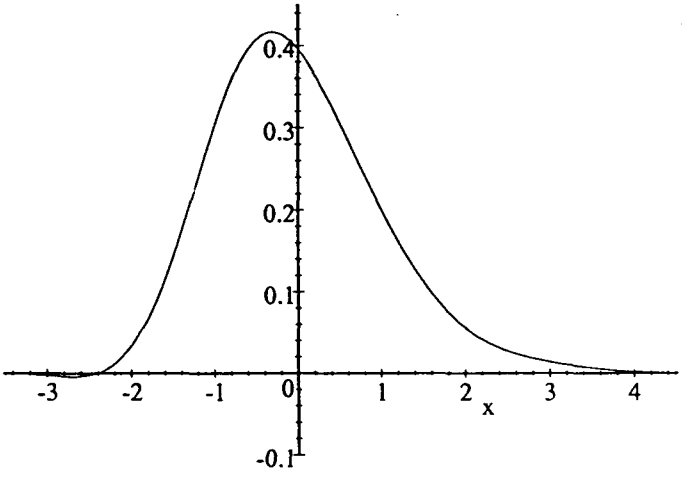

Figura 4. Aproximación de Edgeworth de orden $\frac{1}{n}$, con $n = 10$.

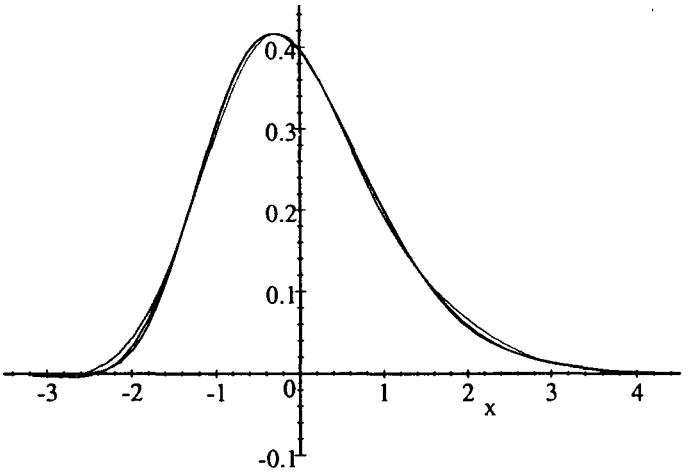

Figura 5. Distribución de S^*. Exacta, orden $\frac{1}{\sqrt{n}}$, y $\frac{1}{n}$.

Observemos la buena exactitud de esta técnica asimptótica. Con tales propiedades nos parece viable aplicar esta expansión a nuestro tema de investigación.

Capítulo 4

Estimador de Momentos de $Y = |X|$ e Intervalos de confianza para p

En este capítulo analizaremos los estimadores que intervienen en los procesos de estimación del parámetro p, mostraremos sus propiedades y ventajas. También aplicaremos la expansión de Edgeworth, especificando el orden de aproximación, y aprovecharemos los resultados obtenidos para fabricar intervalos de confianza para la función de razón generalizada, ver abajo. La idea final de los intervalos de confianza, encontrados en este capítulo, es hallar un intervalo de confianza para el parámetro de forma p de la fdgg.

4.1 Estimador de Momentos de $Y = |X|$

La función de razón gaussiana generalizada, obtenida asumiendo que μ es cero, y utilizada para estimar el parámetro de forma es

Teorema. $\frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)} = \frac{E^2[|x|]}{\sigma_X^2} = M.$

En el trabajo de Rodríguez y León [10] consideran el estimador del lado derecho del teorema, como

$$\hat{M} = \frac{\left(\frac{1}{N}\sum_i |x_i|\right)^2}{\frac{1}{N}\sum_i x_i^2}.$$

Aquí justificaremos el uso de estos estadísticos los cuales no son los estimadores de momentos de la distribución gaussiana generalizada ya que fueron propuestos sólo como estimadores naturales en el lado derecho del teorema anterior (posteriormente compararemos con el estimador del método de momentos).

Hagamos para ello $Y = |X|$, y hallemos la distribución de Y con sus momentos.

Usando el teorema de transformación con $Y = |X| = \begin{cases} x & x > 0 \\ -x & x < 0 \end{cases}$, tenemos que $x = -y$ para x negativa y $x = y$ para x positiva, es decir, $y > 0$, por lo que

$$\begin{aligned}
f_Y(y) &= f_X(-y)|\frac{dx}{dy}| + f_X(y)|\frac{dx}{dy}| \\
&= f_X(-y)|-1| + f_X(y)|1| \\
&= \frac{1}{2\Gamma(1+1/p)A(p)} e^{-\left[\frac{y}{A(p)}\right]^p} + \frac{1}{2\Gamma(1+1/p)A(p)} e^{-\left[\frac{y}{A(p)}\right]^p} \\
&= \frac{2}{2\Gamma(1+1/p)A(p)} e^{-\left[\frac{y}{A(p)}\right]^p} \\
&= \frac{1}{\Gamma(1+1/p)A(p)} e^{-\left[\frac{y}{A(p)}\right]^p},
\end{aligned}$$

como $A(p) = \left[\frac{\sigma_X^2 \Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}$, podemos sacar sigma cuadrada, y nos queda $A(p) = \sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}$, reemplazando en $f_Y(y)$, tenemos que

$$
\begin{aligned}
f_Y(y) &= \frac{1}{\Gamma(1+1/p)\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}} e^{-\left[\frac{y}{\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} \\
&= \frac{p}{\Gamma(1/p)\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}} e^{-\left[\frac{y}{\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p}, y > 0.
\end{aligned}
$$

Emplearemos el método de momentos para hallar un estimador [7] de $\frac{E^2[\|x\|]}{\sigma_X^2}$, para ello calcularemos los primeros dos momento de la distribución de Y. Para el cálculo de los momentos nos será útil la distribución gamma generalizada (GG) [13];

$$
\begin{aligned}
f_{GG}(x) &= \frac{p}{a\Gamma(d/p)} \left(\frac{x}{a}\right)^{d-1} e^{-\left(\frac{x}{a}\right)^p} \\
&= \frac{p}{a^d\Gamma(d/p)} (x)^{d-1} e^{-\left(\frac{x}{a}\right)^p}, x > 0,
\end{aligned}
$$

notemos que esta se reduce a la distribución gausiana generalizada con $\mu = 0$, si $d = 1$.

Entonces,

$$
E(Y) = \int_0^\infty \frac{yp}{\Gamma(1/p)\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}} e^{-\left[\frac{y}{\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy
$$

$$
\begin{aligned}
&= \frac{\Gamma(2/p)}{\Gamma(1/p)\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{-\frac{1}{2}}} \int_0^\infty \frac{yp}{\frac{\Gamma(1/p)}{\Gamma(3/p)}\sigma_X\,\Gamma(2/p)} e^{-\left[\frac{y}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy \\
&= \frac{\Gamma(2/p)}{\Gamma(1/p)\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{\frac{1}{2}}} \int_0^\infty \frac{yp}{\frac{\Gamma(1/p)}{\Gamma(3/p)}\sigma_X\,\Gamma(2/p)} e^{-\left[\frac{y}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy \\
&= \frac{\Gamma(2/p)\sigma}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}} \int_0^\infty \frac{yp}{\frac{\Gamma(1/p)}{\Gamma(3/p)}\sigma_X^2\Gamma(2/p)} e^{-\left[\frac{y}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy.
\end{aligned}
$$

Hemos hecho algunos acomodos para poder identificar el integrando con una gamma generalizada con $a = \sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}$, $d = 2$, $p = p$:

$$
\begin{aligned}
f_{GG}(x) &= \frac{p}{a^d\Gamma(d/p)}(x)^{d-1} e^{-\left(\frac{x}{a}\right)^p} \\
&= \frac{p}{\sigma_X^2\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{2}{2}}\Gamma(2/p)}(x)^{2-1} e^{-\left(\frac{x}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right)^p} \\
&= \frac{px}{\frac{\Gamma(1/p)}{\Gamma(3/p)}\sigma_X^2\Gamma(2/p)} e^{-\left(\frac{x}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right)^p}.
\end{aligned}
$$

Esto significa que la integral anterior es $\frac{\Gamma(2/p)\sigma_X}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}} \cdot 1$, así, el valor esperado de Y queda como

$$
E(Y) = \frac{\Gamma(2/p)\sigma_X}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}}.
$$

Podemos apreciar que

Resultado. $E^2(Y) = \frac{\Gamma^2(2/p)\sigma_X^2}{\Gamma(1/p)\Gamma(3/p)}$. Ahora calculemos el valor esperado de Y^2:

$$
\begin{aligned}
E(Y^2) &= \int_0^\infty \frac{y^2 p}{\Gamma(1/p)\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}} e^{-\left[\frac{y}{\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy \\
&= \frac{\Gamma(3/p)}{\Gamma(1/p)\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{-1}} \int_0^\infty \frac{y^2 p}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{3}{2}} \sigma_X\, \Gamma(3/p)} e^{-\left[\frac{y}{\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy \\
&= \frac{\Gamma(3/p)}{\Gamma(1/p)\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]} \int_0^\infty \frac{y^2 p}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{3}{2}} \sigma_X\, \Gamma(3/p)} e^{-\left[\frac{y}{\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy \\
&= \sigma^2 \int_0^\infty \frac{y^2 p}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{3}{2}} \sigma_X^3 \Gamma(3/p)} e^{-\left[\frac{y}{\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy.
\end{aligned}
$$

De nuevo hicimos algunos acomodos para poder identificar el integrando con una gamma generalizada con $a = \sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}$, $d = 3$, $p = p$:

$$
\begin{aligned}
f_{GG}(x) &= \frac{p}{a^d \Gamma(d/p)} (x)^{d-1} e^{-\left(\frac{x}{a}\right)^p} \\
&= \frac{p}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{3}{2}} \sigma_X^3 \Gamma(3/p)} (x)^{3-1} e^{-\left(\frac{x}{\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right)^p}
\end{aligned}
$$

$$= \frac{px^2}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{3}{2}} \sigma_X^3 \Gamma(3/p)} e^{-\left(\frac{x}{\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right)^p}$$

$$= \frac{px^2}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{3}{2}} \sigma_X^3 \Gamma(3/p)} e^{-\left(\frac{x}{\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right)^p}, \qquad x > 0.$$

Esto significa que la integral anterior es σ_X^2, así,

Resultado. $E(Y^2) = \sigma_X^2$ Las ecuaciones que resultan del método de momentos aplicado a la variable aleatoria Y son:

$$\mu_{1Y}' = m_1'(= \frac{1}{N}\sum_{i=1}^{N} Y_i), \qquad \mu_{2Y}' = m_2'(= \frac{1}{N}\sum_{i=1}^{N} Y_i^2).$$

donde μ_{1Y}', μ_{2Y}', son respectivamente el primer y segundo momentos poblacionales de la variable aleatoria Y; m_1', m_2', son el primer y segundo momentos muestrales de la variable aleatoria Y (centrados en el origen).

Entonces las anteriores ecuaciones quedan, reemplazando $\mu_1' = E(Y)$, $\mu_2' = E(Y^2)$, ya encontrados anteriormente,

$$\frac{\Gamma(2/p)\sigma_X}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}} = \frac{1}{N}\sum_{i=1}^{N} Y_i = \bar{Y}, \qquad E(Y^2)(= \sigma_X^2) = \frac{1}{N}\sum_{i=1}^{N} Y_i^2 = C_Y^2,$$

de aquí que el estimador de momentos para p, resultaría de despejar de

$$\frac{\Gamma(2/\hat{p})}{[\Gamma(1/\hat{p})\Gamma(3/\hat{p})]^{\frac{1}{2}}} = \frac{\bar{Y}}{\hat{\sigma}_X} = \frac{\bar{Y}}{C_Y},$$

a $\hat{p}$, claro que esta es una ecuación trascendental y no es posible escribir a p en términos de funciones elementales, sin embargo, existen paquetes en la actualidad que pueden realizar directamente el calculo, aunque con ciertas restricciones en los valores de p.

Hemos encontrado que, el estimador planteado en el trabajo de Rodríguez y León sería el obtenido a partir del método de momentos usando la variable aleatoria Y:

$$\begin{aligned}\hat{M} &= \frac{\bar{Y}^2}{C_Y^2}\left(= \frac{\Gamma^2(2/\hat{p})}{\Gamma(1/\hat{p})\Gamma(3/\hat{p})}\right). \\ &= \frac{\bar{Y}^2}{S_X^2}\left(= \frac{\left(\frac{1}{N}\sum_i |x_i|\right)^2}{\frac{1}{N}\sum_i x_i^2}\right).\end{aligned}$$

Se ha hecho uso de la igualdad $C_Y^2 = S_X^2$, ya que

$$C_Y^2 = \frac{1}{N}\sum_{i=1}^{N} Y_i^2 = \frac{1}{N}\sum_{i=1}^{N} |X_i|^2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2 = S_X^2$$

Conviene notar la siguiente relación entre S_X^2 y S_Y^2, la cual será utilizada posteriormente en el intervalo de confianza, el cual utiliza σ_Y:

$$S_Y^2 = \frac{1}{N}\sum_i \left(Y_i - \bar{Y}\right)^2 = \frac{1}{N}\sum_i Y_i^2 - \bar{Y}^2 = S_X^2 - \bar{Y}^2,$$

es decir,

$$S_Y^2 = S_X^2 - \bar{Y}^2.$$

Las propiedades asintóticas de los estimadores de momentos son la de consistencia y normalidad pero no son asintóticamente eficientes.

Debemos recalcar que esto no es la aplicación del método de momentos usando la distribución gaussiana generalizada, sino la función de densidad de la variable aleatoria $Y = |X|$. Es decir, el estimador natural utilizado en el trabajo de Rodríguez y León resulta ser el estimador de momentos para p empleando la variable transformada.

Mediante el método de momentos empleado directamente sobre la variable aleatoria X (con fdgg), se determina, con $\mu = 0$, y $r \geq 4$, que [13],

$$\frac{1}{N}\sum_{i=1}^{N} x_i^r = \frac{[A(\hat{p})]^r \Gamma\left(\frac{r+1}{\hat{p}}\right)}{\Gamma\left(\frac{1}{\hat{p}}\right)},$$

para $r = 4$,

$$\frac{1}{N}\sum_{i=1}^{N} x_i^4 = \frac{[A(\hat{p})]^4 \Gamma\left(\frac{5}{\hat{p}}\right)}{\Gamma\left(\frac{1}{\hat{p}}\right)},$$

donde $A(\hat{p}) = \left[\frac{\hat{\sigma}_X^2 \Gamma(1/\hat{p})}{\Gamma(3/\hat{p})}\right]^{\frac{1}{2}}$; podemos darle otra forma, con $r = 4$, para comparar, al menos algebraícamente con el anterior estimador (las propiedades estadísticas son equivalentes al ser ambos estimadores de momentos). Dejémos a las expresiones, que envuelven a $\hat{p}$, del lado derecho;

$$\frac{\frac{1}{N}\sum_{i=1}^{N} x_i^4}{\hat{\sigma}_X^4} = \frac{\left[\frac{\Gamma(1/\hat{p})}{\Gamma(3/\hat{p})}\right]^2 \Gamma\left(\frac{5}{\hat{p}}\right)}{\Gamma\left(\frac{1}{\hat{p}}\right)},$$

es decir,

$$\hat{G} = \frac{\frac{1}{N}\sum_{i=1}^{N} x_i^4}{C_Y^4} = \frac{\Gamma(\frac{1}{p})\Gamma(\frac{5}{p})}{\Gamma^2(\frac{3}{p})},$$

donde hemos llamado $\hat{G}$ al estadístico.

Comparaciones:

Sea $G(p) = \frac{\Gamma(\frac{1}{p})\Gamma(\frac{5}{p})}{\Gamma^2(\frac{3}{p})}$, y $M(p) = \frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}$. Para valores pequeños de p, G crece infinitamente y M tiende a cero; cuando p se acerca a 3, G se acerca 2.4181, y M a 0.68446. M tiene un comportamiento más estable. En el intervalo de cero a tres, M es creciente, y G es decreciente. En cuanto a los estimadores: el número de operaciones en $\hat{M}$ es igual al número de operaciones para $\hat{G}$. El estimador $\hat{G}$ conlleva más errores de acumulación en el proceso de elevar cada valor a la cuarta antes de promediar, al igual que al elevar al cuadrado C_Y^2.

Resumiendo:

$M(p)$ es una función creciente. Cuando $p \in (0,3], M \in (0, 0.68446]$. $\hat{M}$ acumula menos error

$G(p)$ es una función decreciente.Cuando $p \in (0,3], M \in [2.4184, \infty]$. $\hat{G}$ acumula más error

Calculos realizados con el software SWP.

4.2 Expansión de Edgeworth partiendo de la variable $|Y|$

Otra ventaja del estadístico $\hat{M}$, es que podemos encontrar una función de densidad de probabilidad en forma aproximada para $\left(\frac{1}{N}\sum_i |x_i|\right)^2$ mediante el uso de la expansión de Edgeworth. Aunque no se puede obtener en forma directa, proponemos aplicar una transformación sencilla [7] para llegar a su función de densidad de probabilidad.

Nos concentraremos inicialmente en el cálculo de la densidad de probabilidad de $\zeta = \sum_{i=1}^{N} Y_i$. Recordemos que $\hat{M}$ es función de $\sum_{i=1}^{N} Y_i$, más adelante veremos que basta con encontrar la densidad aproximada de esta cantidad.

Hagamos la expansión de Edgeworth para $N\bar{Y} = \sum_{i=1}^{N} Y_i$.Consideremos entonces $\zeta = \sum_{i=1}^{N} Y_i$, y observemos que $\zeta > 0$. La expansión de Edgeworth de orden N^{-1}, para la variable estandarizada

$$\zeta^* = \frac{\zeta - N\mu_Y}{\sigma_Y \sqrt[2]{N}} \left(= \frac{\frac{\zeta}{N} - \mu_Y}{\sigma_Y / \sqrt[2]{N}}\right),$$

es

$$f_{\zeta^*}(q) = \frac{e^{-\frac{1}{2}q^2}}{\sqrt[2]{2\pi}}\left[1 + \rho_3 \frac{H_3(q)}{6\sqrt[2]{N}}\right] + O(N^{-1}), \quad \text{con } q > -\frac{N\mu Y}{\sigma_Y \sqrt[2]{N}},$$

donde ρ_3 es el tercer cumulante estandarizado de la variable aleatoria Y, y $H_3(q) = q^3 - 3q$, es el polinomio de Hermite de tercer grado. Calculemos ρ_3;

$$\sigma_Y^3 \rho_3 = E(Y - \mu_Y)^3 = E(Y - \mu_Y)^3 = E[Y^3] - 3E[Y^2]\mu_Y + 3E[Y]\mu_Y^2 - \mu_Y^3,$$

donde $E(Y - \mu_Y)^3$ es el tercer cumulante. Necesitamos los momentos de Y:

$$E[Y] = \frac{\Gamma(2/p)\sigma_X}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}}.$$

Ahora calculemos el valor esperado de Y^3:

$$\begin{aligned}
E(Y^3) &= \int_0^\infty \frac{y^3 p}{\Gamma(1/p)\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}} e^{-\left[\frac{y}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy \\
&= \frac{\Gamma(4/p)}{\Gamma(1/p)\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{-\frac{3}{2}}} \int_0^\infty \frac{y^3 p}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^2 \sigma_X \Gamma(4/p)} e^{-\left[\frac{y}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy \\
&= \frac{\Gamma(4/p)\sigma^3}{\Gamma(1/p)\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{\frac{3}{2}}} \int_0^\infty \frac{y^3 p}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^2 \sigma_X^4 \Gamma(4/p)} e^{-\left[\frac{y}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy.
\end{aligned}$$

Una vez más hicimos algunos acomodos para poder identificar el integrando con una gamma generalizada con $a = \sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}$, $d = 4$, $p = p$:

$$\begin{aligned}
f_{GG}(x) &= \frac{p}{a^d \Gamma(d/p)} (x)^{d-1} e^{-\left(\frac{x}{a}\right)^p} \\
&= \frac{p}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^2 \sigma_X^4 \Gamma(4/p)} (x)^{4-1} e^{-\left(\frac{x}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right)^p} \\
&= \frac{px^3}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^2 \sigma_X^4 \Gamma(4/p)} e^{-\left(\frac{x}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right)^p}, \qquad x > 0.
\end{aligned}$$

Esto significa que la integral anterior es $\frac{\Gamma(4/p)\sigma_X^3}{\Gamma(1/p)\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{\frac{3}{2}}} \cdot 1$, así, el valor esperado de Y^3 queda como

Resultado. $E(Y^3) = \frac{\Gamma(4/p)\sigma_X^3}{\Gamma(1/p)\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{\frac{3}{2}}}$. Procediendo de manera similar,

Resultado. $E(Y^4) = \frac{\Gamma(5/p)\sigma_X^4}{\Gamma(1/p)\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{2}}$. En general, el valor esperado de Y^k se puede obtener de,

$$
\begin{aligned}
E(Y^k) &= \int_0^\infty \frac{y^k p}{\Gamma(1/p)\sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}} e^{-\left[\frac{y}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy \\
&= \frac{\Gamma(\frac{k+1}{p})}{\Gamma(1/p)\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{-\frac{k}{2}}} \int_0^\infty \frac{y^k p}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{k+1}{2}} \sigma_X \Gamma(\frac{k+1}{p})} e^{-\left[\frac{y}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy \\
&= \frac{\Gamma(\frac{k+1}{p})\sigma^k}{\Gamma(1/p)\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{\frac{k}{2}}} \int_0^\infty \frac{y^k p}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{k+1}{2}} \sigma_X^{k+1} \Gamma(\frac{k+1}{p})} e^{-\left[\frac{y}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right]^p} dy.
\end{aligned}
$$

Nuevamente hicimos algunos acomodos para poder identificar el integrando con una gamma generalizada con $a = \sigma_X \left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}$, $d = k+1$, $p = p$:

$$
\begin{aligned}
f_{GG}(x) &= \frac{p}{a^{k+1}\Gamma(\frac{k+1}{p})}(x)^{k+1-1}\, e^{-\left(\frac{x}{a}\right)^p} \\
&= \frac{p}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{k+1}{2}} \sigma_X^{k+1}\Gamma(\frac{k+1}{p})}(x)^k\, e^{-\left(\frac{x}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right)^p} \\
&= \frac{px^k}{\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{k+1}{2}} \sigma_X^{k+1}\Gamma(\frac{k+1}{p})} e^{-\left(\frac{x}{\sigma_X\left[\frac{\Gamma(1/p)}{\Gamma(3/p)}\right]^{\frac{1}{2}}}\right)^p}, \quad x > 0.
\end{aligned}
$$

Esto significa que la integral anterior es $\frac{\Gamma(\frac{k+1}{p})\sigma_X^k}{\Gamma(1/p)\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{\frac{k}{2}}} \cdot 1$, así, el valor esperado de Y^k queda como,

Resultado. $E(Y^k) = \frac{\Gamma(\frac{k+1}{p})\sigma_X^k}{\Gamma(1/p)\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{\frac{k}{2}}}$. El tercer cumulante es, sustituyendo los valores esperados,

$$
\begin{aligned}
k_3 &= E[Y^3] - 3E[Y^2]\mu_Y + 3E[Y]\mu_Y^2 - \mu_Y^3 \\
&= \frac{\Gamma(4/p)\sigma_X^3}{\Gamma(1/p)\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{\frac{3}{2}}} - 3(\sigma_X^2)\frac{\Gamma(2/p)\sigma_X}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}} + 3\left[\frac{\Gamma(2/p)\sigma_X}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}}\right]^3 - \\
&\quad \left[\frac{\Gamma(2/p)\sigma_X}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}}\right]^3 \\
&= \sigma_X^3\left\{\frac{\Gamma(4/p)}{\Gamma(1/p)\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{\frac{3}{2}}} - \frac{3\Gamma(2/p)}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}} + 2\left[\frac{\Gamma(2/p)}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}}\right]^3\right\}.
\end{aligned}
$$

En el caso particular de $p = 2$ (Normal), $p = 1$ (Laplace), tenemos, respectivamente, que

$$k_3 = \frac{\sigma_X^3 2(4-\pi)}{\pi \sqrt[2]{2\pi}}, \qquad k_3 = \frac{3\sigma_X^3}{\sqrt[2]{2}}.$$

El cumulante estandarizado ρ_3, es

$$\rho_3 = \frac{k_3}{\sigma_Y^3} = \frac{\sigma_X^3}{\sigma_Y^3}\left[\frac{\Gamma(4/p)}{\Gamma(1/p)\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{\frac{3}{2}}} - \frac{3\Gamma(2/p)}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}} + 2\left[\frac{\Gamma(2/p)}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}}\right]^3\right].$$

Calculando $\sigma_Y^2 = E[Y^2] - [\mu_Y]^2$, tenemos,

$$\begin{aligned}
\sigma_Y^2 &= \sigma_X^2 - \left[\frac{\Gamma(2/p)\sigma_X}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}}\right]^2 \\
&= \sigma_X^2\left[1 - \frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}\right],
\end{aligned}$$

entonces,

$$\begin{aligned}
\sigma_Y &= \sigma_X\left[1 - \frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}\right]^{\frac{1}{2}} \\
&= \sigma_X\left[1 - M(p)\right]^{\frac{1}{2}}
\end{aligned}$$

(notemos que $\sigma_Y < \sigma_X$, por ello el Intervalo de Confianza basado en σ_Y se espera que sea más angosto), así, el tercer cumulante estandarizado es,

$$\rho_3 = \frac{1}{\left(1-\frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}\right)^{\frac{3}{2}}}\left[\frac{\Gamma(4/p)}{\Gamma(1/p)\left[\frac{\Gamma(3/p)}{\Gamma(1/p)}\right]^{\frac{3}{2}}} - \frac{3\Gamma(2/p)}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}} + 2\left[\frac{\Gamma(2/p)}{[\Gamma(1/p)\Gamma(3/p)]^{\frac{1}{2}}}\right]^3\right]$$

(en el caso particular de $p = 2$, $p = 1$, tenemos, respectivamente, que $\rho_3 = \frac{\sqrt[2]{2}(4-\pi)}{(\pi-2)^{\frac{3}{2}}}$, $\rho_3 = 3\sqrt[4]{2}$). Este valor es el que se introduciría en la expansión de Edgeworth,

$$f_{\zeta^*}(q) = \frac{e^{-\frac{1}{2}q^2}}{\sqrt[2]{2\pi}}[1 + \rho_3\frac{H_3(q)}{6\sqrt[2]{N}}] + O(N^{-1}), \quad \text{con } q > -\frac{N\mu_Y}{\sigma_Y\sqrt[2]{N}} = -\sqrt[2]{\frac{NM}{1-M}}.$$

Realicemos una transformación sencilla para encontrar la densidad de probabilidad para la variable aleatoria $U^* = (\zeta^*)^2$. Tenemos que,

$$f_{U^*}(r) = f_{\zeta^*}(-\sqrt[2]{r})|-\tfrac{1}{2}r^{-\frac{1}{2}}| + f_{\zeta^*}(\sqrt[2]{r})|\tfrac{1}{2}r^{-\frac{1}{2}}|, \qquad 0 < r < \frac{N\mu_Y^2}{\sigma_Y^2},$$

sustituyendo:

$$\begin{aligned}
f_{U^*}(r) &= \frac{1}{2\sqrt[2]{r}}\frac{e^{-\frac{1}{2}r}}{\sqrt[2]{2\pi}}\left[1 + \rho_3\frac{H_3(-\sqrt[2]{r})}{6\sqrt[2]{N}}\right] + \\
&\quad \frac{1}{2\sqrt[2]{r}}\frac{e^{-\frac{1}{2}r}}{\sqrt[2]{2\pi}}\left[1 + \rho_3\frac{H_3(\sqrt[2]{r})}{6\sqrt[2]{N}}\right] + O(N^{-1}) \\
&= \frac{1}{2\sqrt[2]{r}}\frac{e^{-\frac{1}{2}r}}{\sqrt[2]{2\pi}}\left\{2 + \frac{\rho_3}{6\sqrt[2]{N}}\left[H_3(-\sqrt[2]{r}) + H_3(\sqrt[2]{r})\right]\right\} + O(N^{-1}) \\
&= \frac{1}{2\sqrt[2]{r}}\frac{e^{-\frac{1}{2}r}}{\sqrt[2]{2\pi}}\cdot 2 + O(N^{-1}) \\
&= \frac{e^{-\frac{1}{2}r}}{\sqrt[2]{2\pi r}} + O(N^{-1}), \qquad 0 < r < \frac{N\mu_Y^2}{\sigma_Y^2} = \frac{NM}{1-M},
\end{aligned}$$

donde se canceló $H_3(-\sqrt[2]{r})+H_3(\sqrt[2]{r})$, ya que H_3 es una función impar: $H_3(q) = q^3 - 3q$, es decir, $H_3(-\sqrt[2]{r}) = (-\sqrt[2]{r})^3 - 3(-\sqrt[2]{r}) = -r^{\frac{3}{2}} + r^{\frac{1}{2}}$, $H_3(\sqrt[2]{r}) = (\sqrt[2]{r})^3 - 3(-\sqrt[2]{r}) = r^{\frac{3}{2}} - r^{\frac{1}{2}}$, de donde $H_3(-\sqrt[2]{r}) + H_3(\sqrt[2]{r}) = 0$.

Vemos que la distribución de U^*, de orden N^{-1}, es una Gamma($\alpha = \frac{1}{2}$, $\beta = 2$), es decir, una Ji-cuadrada con un grado de libertad, para cualquier p y σ_X (la distribución asíntótica de $\left(\frac{\bar{Y}-\mu_{\bar{Y}}}{\sigma_{\bar{Y}}}\right)^2 = \left(\frac{\bar{Y}-\mu_Y}{\frac{\sigma_Y}{\sqrt[2]{N}}}\right)^2$ es una χ_1^2). De este resultado sabemos que el valor esperado de la variable aleatoria U^* es, $E[U^*] = 1$, y su varianza $V(U^*) = 2$, o equivalentemente, $\sigma_{U^*} = \sqrt[2]{2}$, entonces

$$E[U^*] = E\left[\left(\frac{\zeta - N\mu_Y}{\sigma_Y \sqrt[2]{N}}\right)^2\right] = \frac{1}{\sigma_Y^2 N}\left[E(\zeta^2) - N^2\mu_Y^2\right] = 1.$$

Despejando $E(\zeta^2)$,

$$E(\zeta^2) = \sigma_Y^2 N + N^2\mu_Y^2.$$

De aquí,

$$E\left(\left(\frac{\zeta}{N}\right)^2\right) = \frac{\sigma_Y^2}{N} + \mu_Y^2,$$

La varianza de Y ya fue obtenida, y es

$$\sigma_Y^2 = \sigma_X^2\left[1 - \frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}\right].$$

Entonces

$$E(\bar{Y}^2) = \frac{\sigma_X^2\left[1 - \frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}\right]}{N} + \frac{\sigma_X^2\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)} = \frac{\Gamma^2(2/p)\sigma_X^2}{\Gamma(1/p)\Gamma(3/p)}\left[1 - \frac{1}{N}\right] + \frac{\sigma_X^2}{N},$$

es decir,

$$\begin{aligned} E(\frac{\bar{Y}^2}{\sigma_X^2}) &= \frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}\left[1-\frac{1}{N}\right]+\frac{1}{N} \\ &= \frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}+O(N^{-1}) \\ &= M(p)+O(N^{-1}). \end{aligned}$$

Observemos que en el límite cuando N tiende a infinito, $\lim_{N\to\infty} E(\frac{\bar{Y}^2}{\sigma_X^2}) = \frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)} = M(p)$.

Veamos ahora un teorema de convergencia[2]; sean, U_N, V_N, sucesiones de variables aleatorias, y sea c, una constante distinta de cero, entonces si U_N converge en distribución a una función de distribución $F(u)$, y si V_N converge en probabilidad a c, entonces $V_N U_N$ converge en distribución a $F(cu)$, en símbolos,

Teorema. Si $U_N \xrightarrow{D} F(u)$, y $V_N \xrightarrow{P} c$, entonces $V_N U_N \xrightarrow{D} F(cu)$. Además sabemos que $\frac{\sigma_Y^2}{S_Y^2}$ converge en probabilidad a 1:

$$\frac{\sigma_Y^2}{S_Y^2} \xrightarrow{P} 1.$$

Apliquemos estos resultados a nuestro tratamiento: Hemos visto que la distribución asintótica de la variable $\left(\frac{\bar{Y}-\mu_Y}{\frac{\sigma_Y}{\sqrt[2]{N}}}\right)^2$, es una gamma con $\alpha = \frac{1}{2}, \beta = 2$, es decir, $\left(\frac{\bar{Y}-\mu_Y}{\frac{\sigma_Y}{\sqrt[2]{N}}}\right)^2 \xrightarrow{D} \chi_1^2$, y $\frac{\sigma_Y^2}{S_Y^2} \xrightarrow{P} 1$, entonces $\frac{\sigma_Y^2}{S_Y^2}\left(\frac{\bar{Y}-\mu_Y}{\frac{\sigma_Y}{\sqrt[2]{N}}}\right)^2 = \left(\frac{\bar{Y}-\mu_Y}{\frac{S_Y}{\sqrt[2]{N}}}\right)^2$, pero, $S_Y = \sqrt{S_X^2 - \bar{Y}^2}$, por lo tanto,

$$\left(\frac{\bar{Y}-\mu_Y}{\frac{\sqrt{S_X^2-\bar{Y}^2}}{\sqrt[2]{N}}}\right)^2 \xrightarrow{D} \chi_1^2.$$

Se podría usar este resultado para efectuar una prueba de hipótesis referente a la cantidad μ_Y, sin embargo nuestro propósito principal es el de estimar, por ello, lo utilizaremos para obtener un intervalo de confianza para μ_Y. Aunque el parámetro μ_Y no es el de interés directo, sabemos que está en función del valor del parámetro de forma p.

4.2.1 Intervalos de Confianza para M, y p

Utilizaremos la técnica del pivote [4], [5], y [7] para hallar un intervalo de confianza de $(1-\alpha)\%$ para la cantidad μ_Y. El pivote en cuestión es $W = \left(\frac{\bar{Y}-\mu_Y}{\frac{S_Y}{\sqrt[2]{N}}}\right)^2$. Consideraremos primero un intervalo típico de dos lados, para posteriormente ver la ventaja de utilizar un intervalo de confianza de un sólo lado.

Primer intervalo propuesto: se acostumbra utilizar el resultado, para un intervalo de dos lados, $P(\chi^2_{1-\frac{\alpha}{2},1} < W < \chi^2_{\frac{\alpha}{2},1}) = 1-\alpha$, de donde, sustituyendo W:

$$\begin{aligned}
\chi^2_{1-\frac{\alpha}{2},1} &< W < \chi^2_{\frac{\alpha}{2},1} \\
\chi^2_{1-\frac{\alpha}{2},1} &< \left(\frac{\bar{Y}-\mu_Y}{\frac{S_Y}{\sqrt[2]{N}}}\right)^2 < \chi^2_{\frac{\alpha}{2},1}.
\end{aligned}$$

Al sacar la raíz cuadrada, se tienen dos intervalos, siendo la unión de ellos el intervalo buscado,

$$\sqrt{\chi^2_{1-\frac{\alpha}{2},1}} < \frac{\bar{Y}-\mu_Y}{\frac{S_Y}{\sqrt[2]{N}}} < \sqrt{\chi^2_{\frac{\alpha}{2},1}}, \qquad -\sqrt{\chi^2_{\frac{\alpha}{2},1}} < \frac{\bar{Y}-\mu_Y}{\frac{S_Y}{\sqrt[2]{N}}} < -\sqrt{\chi^2_{1-\frac{\alpha}{2},1}}.$$

En el primer intervalo, el término central es positivo, y en el segundo caso es negativo, considerando ambos intervalos a la vez, tenemos entonces que el intervalo deseado es,

$$\begin{array}{rcl} \frac{S_Y}{\sqrt[2]{N}}\sqrt{\chi^2_{1-\frac{\alpha}{2},1}} & < & \bar{Y}-\mu_Y < \frac{S_Y}{\sqrt[2]{N}}\sqrt{\chi^2_{\frac{\alpha}{2},1}} \\ & & \cup \\ \frac{-S_Y}{\sqrt[2]{N}}\sqrt{\chi^2_{\frac{\alpha}{2},1}} & < & \bar{Y}-\mu_Y < \frac{-S_Y}{\sqrt[2]{N}}\sqrt{\chi^2_{1-\frac{\alpha}{2},1}}, \end{array}$$

manipulando un poco;

$$\begin{array}{rcl} -\bar{Y}+\frac{S_Y}{\sqrt[2]{N}}\sqrt{\chi^2_{1-\frac{\alpha}{2},1}} & < & -\mu_Y < -\bar{Y}+\frac{S_Y}{\sqrt[2]{N}}\sqrt{\chi^2_{\frac{\alpha}{2},1}} \\ & & \cup \\ -\bar{Y}+\frac{-S_Y}{\sqrt[2]{N}}\sqrt{\chi^2_{\frac{\alpha}{2},1}} & < & -\mu_Y < -\bar{Y}+\frac{-S_Y}{\sqrt[2]{N}}\sqrt{\chi^2_{1-\frac{\alpha}{2},1}}, \end{array}$$

finalmente,

$$
\begin{array}{rcl}
\bar{Y} - \frac{S_Y}{\sqrt[2]{N}}\sqrt{\chi^2_{\frac{\alpha}{2},1}} & < & \mu_Y < \bar{Y} - \frac{S_Y}{\sqrt[2]{N}}\sqrt{\chi^2_{1-\frac{\alpha}{2},1}} \\
 & & \cup \\
\bar{Y} + \frac{S_Y}{\sqrt[2]{N}}\sqrt{\chi^2_{1-\frac{\alpha}{2},1}} & < & \mu_Y < \bar{Y} + \frac{S_Y}{\sqrt[2]{N}}\sqrt{\chi^2_{\frac{\alpha}{2},1}}.
\end{array}
$$

Ahora podemos sustituir $\mu_Y = \sigma_X \sqrt[2]{M(p)}$, $S_Y = \left(S_X^2 - \bar{Y}^2\right)^{\frac{1}{2}}$ y dividir por σ_X; haremos los pasos para el intervalo de la izquierda, y escribiremos el resultado final para los dos intervalos,

$$
\begin{array}{rcl}
\frac{\bar{Y}}{\sigma_X} - \frac{\frac{\left(S_X^2-\bar{Y}^2\right)^{\frac{1}{2}}}{\sigma_X}}{\sqrt{N}}\sqrt{\chi^2_{\frac{\alpha}{2},1}} & < & \frac{\sigma_X\sqrt{M(p)}}{\sigma_X} < \frac{\bar{Y}}{\sigma_X} - \frac{\frac{\left(S_X^2-\bar{Y}^2\right)^{\frac{1}{2}}}{\sigma_X}}{\sqrt[2]{N}}\sqrt{\chi^2_{1-\frac{\alpha}{2},1}} \\
\frac{\bar{Y}}{\sigma_X} - \frac{S_X}{\sigma_X}\frac{\left[1-\frac{\bar{Y}^2}{S_X^2}\right]^{\frac{1}{2}}}{\sqrt{N}}\sqrt{\chi^2_{\frac{\alpha}{2},1}} & < & \sqrt{M(p)} < \frac{\bar{Y}}{\sigma_X} - \frac{S_X}{\sigma_X}\frac{\left[1-\frac{\bar{Y}^2}{S_X^2}\right]^{\frac{1}{2}}}{\sqrt{N}}\sqrt{\chi^2_{1-\frac{\alpha}{2},1}}.
\end{array}
$$

Ahora, en el caso de que no conozcamos σ_X podemos usar su estimador de momentos (además, $\frac{\sigma_X^2}{S_X^2} \xrightarrow{P} 1$). Entonces, elevando al cuadrado,

$$
\left(\frac{\bar{Y}}{S_X} - \frac{\left[1-\frac{\bar{Y}^2}{S_X^2}\right]^{\frac{1}{2}}}{\sqrt{N}}\sqrt{\chi^2_{\frac{\alpha}{2},1}}\right)^2 < M(p) < \left(\frac{\bar{Y}}{S_X} - \frac{\left[1-\frac{\bar{Y}^2}{S_X^2}\right]^{\frac{1}{2}}}{\sqrt{N}}\sqrt{\chi^2_{1-\frac{\alpha}{2},1}}\right)^2,
$$

que junto con,

$$\left(\frac{\bar{Y}}{S_X}+\frac{\left[1-\frac{\bar{Y}^2}{S_X^2}\right]^{\frac{1}{2}}}{\sqrt{N}}\sqrt{\chi^2_{1-\frac{\alpha}{2},1}}\right)^2 < M(p) < \left(\frac{\bar{Y}}{S_X}+\frac{\left[1-\frac{\bar{Y}^2}{S_X^2}\right]^{\frac{1}{2}}}{\sqrt{N}}\sqrt{\chi^2_{\frac{\alpha}{2},1}}\right)^2,$$

produce el intervalo buscado. Vemos que nuestro intervalo es la unión de dos intervalos.

Aunque $M(p)$ no es el fin último en la estimación, se sabe que es una función creciente en el intervalo de interés ingenieril para p, que resulta ser, $0.3 < p < 3$. De esta manera, de acuerdo al valor del intervalo para $M(p)$, tendríamos uno para p. Por ejemplo, usando una simulación con datos, generados por el paquete excel, provenientes de una normal ($p = 2$), y utilizando $\sigma_X = 1$, contemplando $N = 324$, al 98% de confianza, y usando las tablas de la Ji-cuadrada, obteniendo de ellas que $\chi^2_{\frac{\alpha}{2},1} = \chi^2_{.01,1} = 6.6349$, $\chi^2_{1-\frac{\alpha}{2},1} = \chi^2_{.99,1} = .00016$ (podríamos despreciar este valor ya que es cercano a cero), nos produce un cociente de $\frac{\bar{Y}}{S_X} = 0.82590$. Así, el IC para $M(p)$ es,

$$\begin{aligned}
\left[\frac{\bar{Y}}{S_X}-\frac{\left[1-\frac{\bar{Y}^2}{S_X^2}\right]^{\frac{1}{2}}\sqrt{\chi^2_{\frac{\alpha}{2},1}}}{\sqrt{N}}\right]^2 &< M < \left[\frac{\bar{Y}}{S_X}-\frac{\left[1-\frac{\bar{Y}^2}{S_X^2}\right]^{\frac{1}{2}}\sqrt{\chi^2_{1-\frac{\alpha}{2},1}}}{\sqrt{N}}\right]^2 \\
[0.82590-0.080683]^2 &< M < [0.82590-0.000396]^2 \\
0.55535 &< M < 0.68146,
\end{aligned}$$

el cual hay que unir con,

$$\left[\frac{\bar{Y}}{S_X}+\frac{\left[1-\frac{\bar{Y}^2}{S_X^2}\right]^{\frac{1}{2}}\sqrt{\chi^2_{1-\frac{\alpha}{2},1}}}{\sqrt{N}}\right]^2 < M < \left[\frac{\bar{Y}}{S_X}+\frac{\left[1-\frac{\bar{Y}^2}{S_X^2}\right]^{\frac{1}{2}}\sqrt{\chi^2_{\frac{\alpha}{2},1}}}{\sqrt{N}}\right]^2$$

$$\begin{aligned}(0.82590+0.000396)^2 &< M < (0.82590+0.080683)^2\\ 0.68277 &< M < 0.82189.\end{aligned}$$

Entonces, el IC, del 98% es,

$$0.55535 < M(p) < 0.68146 \ \cup \ 0.68277 < M(p) < 0.82189.$$

El valor correcto de $M(p)$, es 0.63662. Nótese la longitud del intervalo: $0.12611 + .13912 = 0.2652$.

La anterior no es una forma típica de producir un intervalo para una media poblacional, lo usual es considerar la distribución normal estándar, sin embargo hemos sido capaces de construir un intervalo de confianza aprovechando el resultado obtenido, para nuestro problema de estimación.

Una dificultad que se puede observar es que existe un pequeño segmento que no es cubierto por la unión de los dos intervalos. Esto es debido a que se usó un intervalo de dos lados en el proceso de construcción, así que existe la posibilidad de que nuestro intervalo no cubra con la confianza establecida el valor del parámetro deseado. Podemos evitar la deficiencia del intervalo si consideramos un intervalo de un sólo lado (de una cola) [4],[5].

Intervalo de un sólo lado. Usando el hecho de que, $P(W < \chi^2_{\alpha,1}) = 1-\alpha$, construiremos el intervalo a partir de,

$$W < \chi^2_{\alpha,1}.$$

Evidentemente $W > 0$. No es forzosamente el mejor intervalo aquel que tiene probabilidades de colas iguales, de hecho eso sería lo adecuado para una distribución simétrica, pero en nuestro caso la distribución Ji-cuadrada con un grado de libertad es altamente asimétrica; tiende a infinito cerca de $\chi^2_1 = 0$, y decae a cero para valores alejados del cero. De hecho, la mediana de esta distribución es $\tilde{\chi}^2_1 = 0.455$, con su valor esperado igual a 1. Con el anterior intervalo podemos cubrir el segmento dejado fuera por el primer intervalo, incluso la longitud del intervalo disminuye, por ejemplo, para el primer intervalo propuesto, con $\alpha = .1$,

$$\chi^2_{\frac{\alpha}{2},1} - \chi^2_{1-\frac{\alpha}{2},1} = 3.84146 - .00393 = 3.83753,$$

y para el segundo intervalo propuesto,

$$\chi^2_{\alpha,1} - 0 = 2.70554.$$

Construyamos el segundo intervalo:

$$\begin{aligned} W &< \chi^2_{\alpha,1} \\ \left(\frac{\bar{Y} - \mu_Y}{\frac{S_Y}{\sqrt[2]{N}}}\right)^2 &< \chi^2_{\alpha,1}, \end{aligned}$$

sacando raíz cuadrada,

$$\left|\frac{\bar{Y}-\mu_Y}{\frac{S_Y}{\sqrt[2]{N}}}\right| < \sqrt{\chi^2_{\alpha,1}},$$

o equivalentemente,

$$-\sqrt{\chi^2_{\alpha,1}} < \frac{\bar{Y}-\mu_Y}{\frac{S_Y}{\sqrt[2]{N}}} < \sqrt{\chi^2_{\alpha,1}},$$

el intervalo resultante para M, despúes de las manipulaciones, sería,

$$\left(\frac{\bar{Y}}{S_X} - \sqrt[2]{\frac{\chi^2_{\alpha,1}\left(1-\left(\frac{\bar{Y}}{S_X}\right)^2\right)}{N}}\right)^2 < M < \left(\frac{\bar{Y}}{S_X} + \sqrt[2]{\frac{\chi^2_{\alpha,1}\left(1-\left(\frac{\bar{Y}}{S_X}\right)^2\right)}{N}}\right)^2.$$

Podemos también obtener un tamaño de muestra para una longitud específica del intervalo, consideremos una longitud de $L = 0.2$, la longitud del intervalo, para el intervalo de dos lados, es

$$L = \left[\frac{\bar{Y}}{S_X} - \sqrt[2]{\frac{\chi^2_{1-\frac{\alpha}{2},1}\left(1-\frac{\bar{Y}}{S_X}\right)^2}{N}}\right]^2 - \left[\frac{\bar{Y}}{S_X} - \sqrt[2]{\frac{\chi^2_{\frac{\alpha}{2},1}\left(1-\frac{\bar{Y}}{S_X}\right)^2}{N}}\right]^2 +$$

$$\left[\frac{\bar{Y}}{S_X}+\sqrt[2]{\frac{\chi^2_{\frac{\alpha}{2},1}\left(1-\left(\frac{\bar{Y}}{S_X}\right)^2\right)}{N}}\right]^2-\left[\frac{\bar{Y}}{S_X}+\sqrt[2]{\frac{\chi^2_{1-\frac{\alpha}{2},1}\left(1-\left(\frac{\bar{Y}}{S_X}\right)^2\right)}{N}}\right]^2$$

$$= 4\frac{\bar{Y}}{S_X}\sqrt{\frac{\chi^2_{\frac{\alpha}{2},1}}{N}\left(1-\left(\frac{\bar{Y}}{S_X}\right)^2\right)}.$$

Hemos despreciado el término $\sqrt[2]{\frac{\chi^2_{1-\frac{\alpha}{2},1}\left(1-\left(\frac{\bar{Y}}{S_X}\right)^2\right)}{N}}$, ya que $\chi^2_{1-\frac{\alpha}{2},1}$ es aproximadamente cero: $0 < 1-\left(\frac{\bar{Y}}{S_X}\right)^2 < 1$, y $N > 1$. Esto ayuda a tener una fórmula más simple para el tamaño de muestras. Entonces,despejando, nos queda que N es igual a,

$$16\left(1-b^2\right)\frac{b^2}{L^2}r^2.$$

donde hemos llamado r^2 a $\chi^2_{\frac{\alpha}{2},1}$, y b es $\frac{\bar{Y}}{S_X}$.

Supongamos que queremos saber el tamaño de muestra necesario para una longitud del intervalo de 0.2, con $\alpha = 0.02$; en el caso de que tengamos una normal ($p = 2$), el valor que se estaría buscando sería $M(2) = 0.63662$, utilicemos este valor en lugar de b^2, obtenemos

$$N = 614.$$

En el caso de la distribución de Laplace ($p = 1$), $M(1) = 0.5$, utilizando este valor en lugar de b^2, tenemos, con la misma confianza y longitud del intervalo que,

$$N = 664.$$

Los resultados análogos para el intervalo de un lado son:

$$L = 4b\sqrt{\frac{r^2}{N}(1-b^2)},$$

sin ninguna aproximación en el desarrollo algebraíco, y

$$N = 16(1-b^2)\frac{b^2}{L^2}r^2,$$

donde r^2 es $\chi^2_{\alpha,1}$, y b es $\frac{\bar{Y}}{S_X}$.

Repitiendo los calculos para saber el tamaño de muestra necesario para una longitud del intervalo de 0.2, con $\alpha = 0.02$; en el caso de que tengamos una normal, con $M(2) = 0.63662$, obtenemos de la última fórmula,

$$N = 501.$$

En el caso de la distribución de Laplace ($p = 1$), $M(1) = 0.5$, tenemos, con la misma confianza y longitud del intervalo que,

$$N = 542.$$

Los tamaños de muestra disminuyen en aproximadamente 100 unidades.

Debido a su característica de cubrimiento, en lo que sigue utilizaremos el intervalo de un sólo lado.

Elaboremos una tabla para el intervalo, con diferentes confianzas y con diferentes tamaños usando valores simulados de una normal, obtenidos del paquete excel, en la que $M(p = 2) = 0.63662$, $\sigma_X = 1$:

N	IC(90%)	IC(95%)	IC(98%)	$\frac{\bar{Y}}{S_X}$
10	$(0.3686, 1.2682)$	$(0.3107, 1.3827)$	$(0.2496, 1.522)$	0.86664
50	$(0.46413, 0.90313)$	$(0.42968, 0.95278)$	$(0.39129, 1.0122)$	0.8158
100	$(0.52916, 0.83745)$	$(0.50333, 0.87068)$	$(0.47411, 0.91012)$	0.82128
324	$(0.59966, 0.76987)$	$(0.58447, 0.78724)$	$(0.56706, 0.80778)$	0.82590
565	$(0.59723, 0.72825)$	$(0.58536, 0.74148)$	$(0.57171, 0.75701)$	0.81309
596	$(0.59938, 0.7269)$	$(0.58781, 0.73976)$	$(0.5745, 0.75485)$	0.81339
625	$(0.60031, 0.72488)$	$(0.589, 0.73743)$	$(0.57598, 0.75216)$	0.8131
2500	$(0.61333, 0.67632)$	$(0.60746, 0.68251)$	$(0.60067, 0.68975)$	0.80277

Todos los intevalos mostrados contienen el valor verdadero de $M(p = 2)$, notemos que en todos los casos $\left(\frac{\bar{Y}}{S_X}\right)^2$ es mayor que $M(2)$. Las longitudes L de los intervalos son presentadas en la siguiente tabla:

N	$L(90\%)$	$L(95\%)$	$L(98\%)$
10	0.8996	1.072	1.2724
50	0.439	0.52310	0.62091
100	0.30829	0.36735	0.43601
324	0.17021	0.20277	0.24072
565	0.13102	0.15612	0.1853
596	0.12752	0.15195	0.18035
625	0.12457	0.14843	0.17618
2500	0.06299	0.07505	0.08908

Ahora presentamos tablas análogas para una distribución de Laplace. En la generación de los números aleatorios usamos también el paquete excell, pero en vista de que el paquete no proporciona directamente los valores requeridos hicimo uso de una transformación, la cual relaciona el producto de una variable aleatoria Bernoulli con una exponencial, produciéndo con ello una variable aleatoria con distribución de Laplace, ver [9]. En este caso $M(p=1)=\frac{1}{2}$:

N	IC(90%)	IC(95%)	IC(98%)	$\frac{Y}{S_X}$
10	$(0.28933, 1.2541)$	$(0.23247, 1.3821)$	$(0.17417, 1.5387)$	0.82888
50	$(0.37817, 0.83664)$	$(0.34369, 0.88999)$	$(0.30566, 0.95409)$	0.76482
100	$(0.37063, 0.69928)$	$(0.34458, 0.73619)$	$(0.31548, 0.7803)$	0.72251
324	$(0.39233\ , 0.57493)$	$(0.37666, 0.59425)$	$(0.35885, 0.61711)$	0.69230
565	$(0.43373, 0.57213)$	$(0.42148, 0.58639)$	$(0.40746, 0.6032)$	0.70749
596	$(0.43118, 0.56592)$	$(0.41923, 0.57979)$	$(0.40555, 0.59613)$	0.70446
625	$(0.44629, 0.57786)$	$(0.43458, 0.59135)$	$(0.42116, 0.60723)$	0.71411
2500	$(0.47464, 0.54043)$	$(0.46856, 0.54695)$	$(0.46155, 0.55459)$	0.71204

Todos los intervalos encontrados contienen el valor de $M(p=1)$.

Mostramos ahora las longitudes de los intervalos:

N	$L(90\%)$	$L(95\%)$	$L(98\%)$
10	0.96477	1.14963	1.36453
50	0.45847	0.5463	0.64843
100	0.32865	0.39161	0.46482
324	0.1826	0.21759	0.25826
565	0.1384	0.16491	0.19574
596	0.13474	0.16056	0.19058
625	0.13157	0.15677	0.18607
2500	0.06579	0.07839	0.09304

Ahora calcularemos veinte intervalos de confianza, con diversas muestras, para la distribución normal al 90% de confianza, con un tamaño de muestra de 324.Valor de M; 0.63662,

$(0.57878, 0.75163)$
$(0.53664, 0.71387)$
$(0.58636, 0.75829)$
$(0.53182, 0.70947)$
$(0.55151, 0.72734)$
$(0.54641, 0.72274)$
$(0.58939, 0.76094)$
$(0.56476, 0.73921)$
$(0.52592, 0.70405)$
$(0.57204, 0.74568)$
$(0.5638, 0.73835)$
$(0.55722, 0.73247)$
$(0.54039, 0.71728)$
$(0.56632, 0.7406)$
$(0.55344, 0.72908)$
$(0.56974, 0.74364)$
$(0.55635, 0.73169)$
$(0.57277, 0.74633)$
$(0.57374, 0.74719)$
$(0.6122, 0.78066)$

La gráfica de estos intervalos se muestra a continuación:

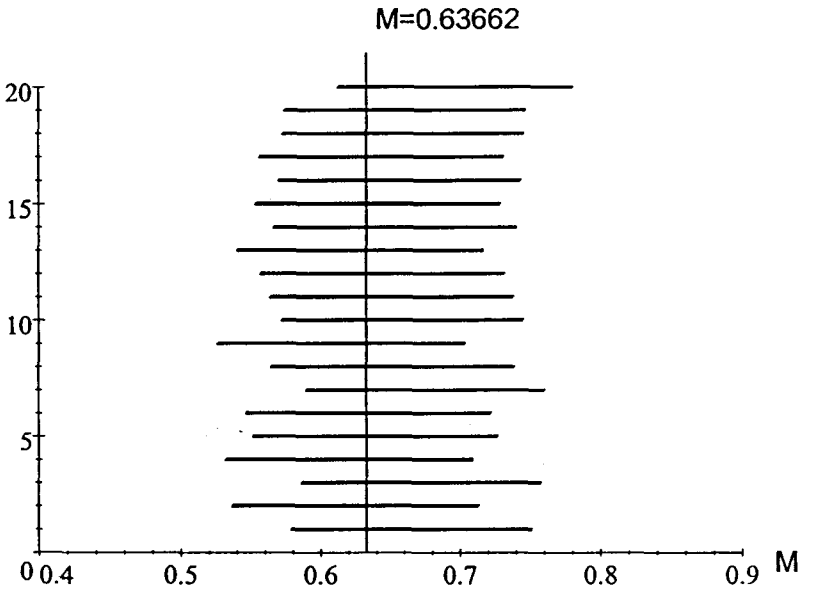

Figura 6. 20 intervalos de confianza del 98%.

Todos los intervalos contienen el verdadero valor de M.

Para terminar este capítulo mostramos los tamaños de muestra mínimos para obtener intervalos de confianza del 90%, 95%, y 98% para M, con una longitud de 0.2, considerando los valores de $p \in (0.3, 3)$.

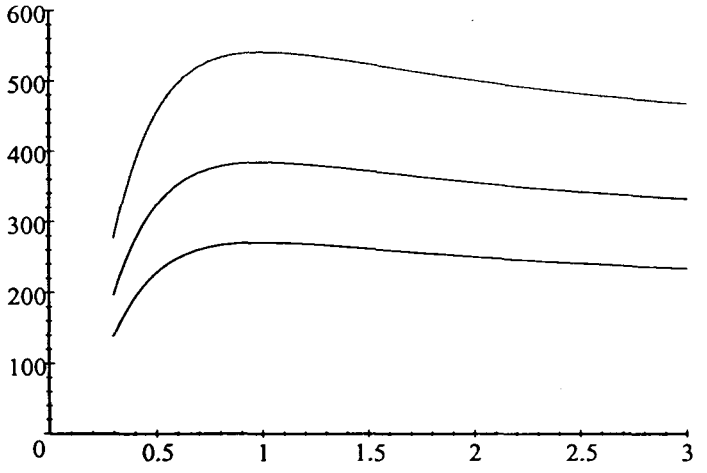

Figura 7. Tamaños de muestra mínimos.

La curva superior corresponde a la confianza del 98%, la central a la confianza del 95%, y la inferior a la del 90%. Los máximos se pueden calcular fácilmente y son, para las confianzas 90%, 95%, y 98%, respectivamente: 271, 385, 543.

Agregamos una gráfica del tamaño de muestra dependiendo de la longitud L del intervalo, con una confianza del 90%.

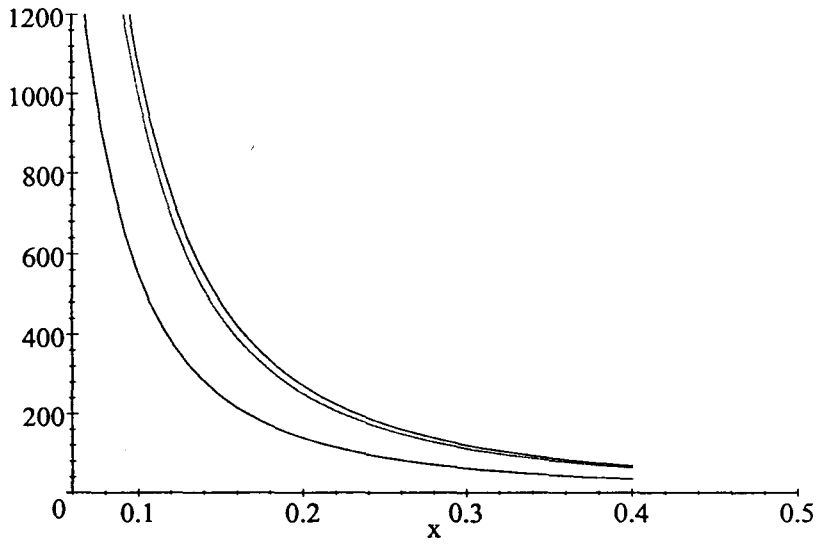

Figura 8. Tamaños de muestra contra longitud de intervalo.

Empezando desde la curva inferior a la superior, los valores de p que se consideraron son: $0.3, 3.0, 2.0$, y 1.0. Observamos que el máximo tamaño de muestra siempre está determinado por el valor de $p = 1$, es decir la distribución de Laplace.

Ya tenemos un intervalo de confianza para M, sin embargo los tamaños requeridos pueden no ser muy satisfactorios en la práctica, más aún, de la gráfica de M Vs p, un intervalo pequeño para M, no implica uno pequeño para p. Propondremos otro intervalo en el cual la longitud del intervalo se disminuya de acuerdo al tamaño de muestra escogido, esto es, mediante un factor conveniente aumentaremos el límite inferior y disminuiremos el límite superior, esto basándonos en que la cantidad M aparece aceptablemente centrada en los intervalos hallados.

Pensando un poco en la expresión para la longitud L del intervalo de confianza, $4b\sqrt{\frac{r^2}{N}(1-b^2)}$, donde r^2 es $\chi^2_{\alpha,1}$, y b es $\frac{\bar{Y}}{S_X}$, notamos que $L \propto \frac{1}{N^{\frac{1}{2}}}$, podemos hacer la conjetura de que el factor de corrección tiene esta forma, es decir, que el límite inferior sea el límite obtenido del intervalo más $\frac{1}{N^{\frac{1}{2}}}$, y el

límite superior sea el límite obtenido del intervalo menos $\frac{1}{N^{\frac{1}{2}}}$. Considerando unos pocos casos (usando la tabla de intervalos con distintos tamaños de N) notamos que es aún mejor restarle al límite superior $\frac{2}{N^{\frac{1}{2}}}$, en lugar de $\frac{1}{N^{\frac{1}{2}}}$. Al hacer estos ajustes a nuestros intervalos la confianza de cubrimiento puede cambiar, por ello, posteriormente, realizaremos unas simulaciones que nos den información sobre tales confianzas.

Hemos trabajado en la estimación de M, sin embargo el parámetro importante es, como ya sabemos, p. En el próximo capítulo consideraremos una forma sencilla para estimar puntualmente el valor de p, aprovechando los resultados obtenidos hasta el momento.

Capítulo 5

Estimación de p

Ahora desarrollaremos un intervalo de confianza para el parámetro de forma p, con la ayuda de los resultados del capítulo anterior, usando para ello un ajuste para la función $\frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}$. Esto es una ejemplificación de la ventaja de haber hallado la aproximación de Edgeworth de la variable $|Y|$.

5.1 Curvas ajuste para $\gamma(p) = \frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}$

La función $\gamma(p) = \frac{\Gamma^2(2/p)}{\Gamma(1/p)\Gamma(3/p)}$, es una función creciente de p, podemos observar su gráfica en la siguiente figura:

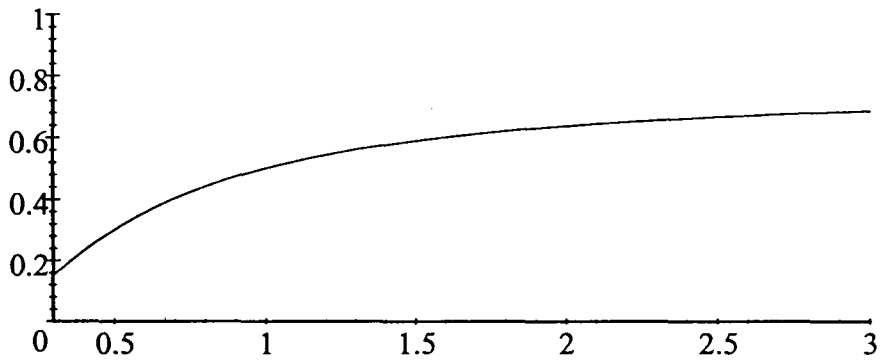

Figura 9. $\gamma(p)$.

Entonces, si encontramos un intervalo de confianza para M, esto equivaldría a encontrar un intervalo sobre el eje vertical, corespondiéndole por lo tanto un intervalo en el eje horizontal. Proponemos realizar un ajuste de curvas, de una manera particular, de tal manera que realizando un despeje algebraíco sencillo podamos realizar una estimación puntual de p. En el presente capítulo mostramos como pudiera hacerse esto.

5.1.1 Dos curvas de ajuste para $\gamma(p)$

En el trabajo de Rodríguez y León [10], mencionan que la función $\gamma(p)$ cumple con la siguiente desigualdad:

$$\gamma(p) \leq \frac{p}{1+p} = R_g(p).$$

Las gráficas de $\gamma(p)$, y $R_g(p)$, son mostradas acontinuación:

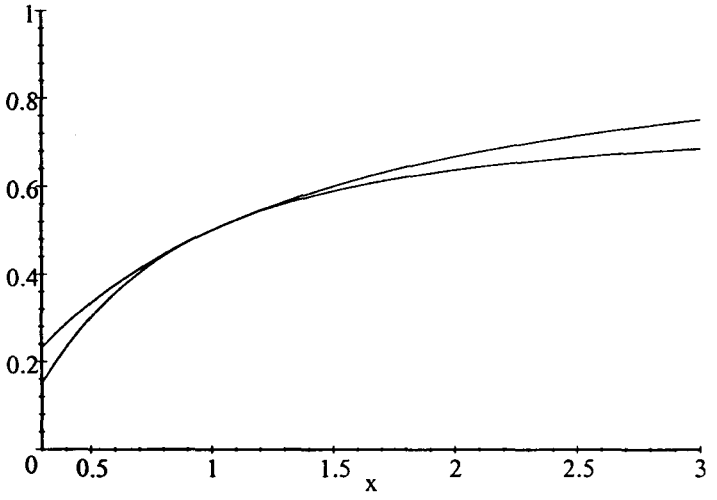

Figura 10. $\gamma(p)$ y $R_g(p)$.

Las gráficas nos sugieren que podríamos proponer ajustar curvas de la forma:

$$A(p) = \begin{cases} ax^2 + bx + c \\ \frac{p}{1+f(p)} \end{cases},$$

la primera posibilidad es la forma más sencilla posible con característica no lineal, y la segunda curva está inspirada en la desigualdad $\gamma(p) \leq \frac{p}{1+p}$, donde $f(p)$, sería una función polinomial.

En el dominio de interés para p, no hay buenos ajustes polinomiales sino hasta uno de grado cuarto, pero si dividimos el dominio en dos intervalos para p, podemos ajustar una polinomio de grado 2 para $p \leq 1 (M \leq 0.5)$, y otro para $p \geq 1 (M \geq 0.5)$, en ésta parte resulta un mejor ajuste utilizando la forma $\frac{p}{1+f(p)}$, además de que $f(p)$ resulta ser una recta.

Llamemos $A_1(p)$ a la curva ajustada para $p \leq 1$, y $A_2(p)$, a la curva ajustada para $p \geq 1$:

$$\gamma(p) = A(p) = \begin{cases} A_1(p) = ap^2 + bp + c, & p \leq 1 \\ A_2(p) = \frac{p}{1+f_2(p)}, & p \geq 1 \end{cases},$$

donde $f_2(p)$ es un polinomio de grado uno. Con estas proposiciones el despeje de p resulta sencillo.

Las curvas que resultan de los ajustes de mínimos cuadrados son:

$$A_1(p) = -0.1299 + 1.07954p - 0.45273p^2, \qquad A_2(p) = \frac{p}{0.7582+1.20007p},$$

estas se hicieron considerando ocho valores para p:

$$p = \{0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\},$$

en el caso de $p \leq 1$; y veintiún puntos:

$$p = \begin{array}{c} \{1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, \\ 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0\} \end{array}$$

en el caso de $p \geq 1$.

La siguientes gráficas muestran a $\gamma(p)$, $A_1(p)$, y $A_2(p)$:

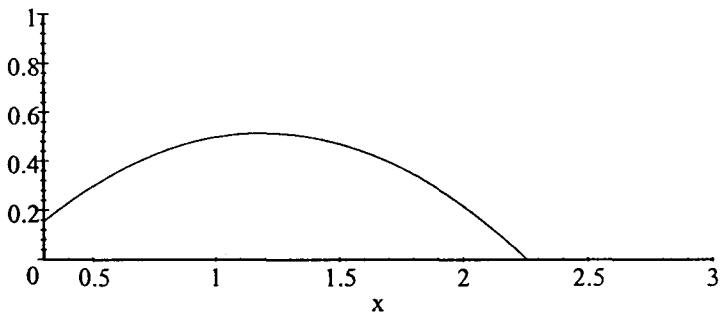

Figura.11. $A_1(p)$, para $p \in (0.3, 3)$.

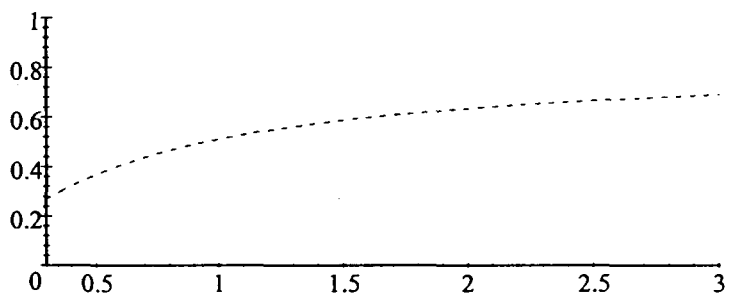

Figura 12. $A_2(p)$, para $p \in (0.3, 3)$.

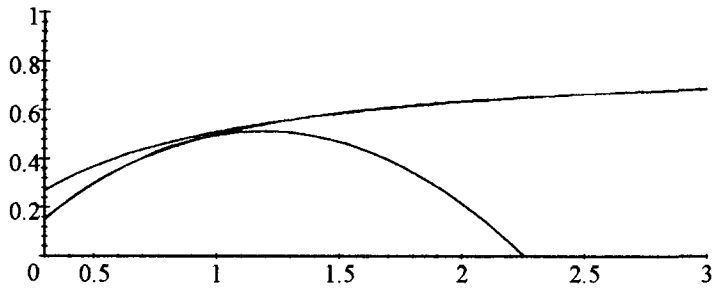

Figura 13. $A_1(p)$, $A_2(p)$, y $\gamma(p)$.

Al parecer el ajuste es bueno. $A_1(p)$ en azul, $A_2(p)$ en rojo, y $\gamma(p)$ en negro.

Las soluciones a las ecuaciones $A(p) = L_1$, $A(p) = L_2$, donde L_1 y L_2 son los límites de los intervalos de confianza encontrados para $M(p)$, nos produciría un intervalo de confianza para p. Considerando las ecuaciones $A_1(p) = L_1, A_1(p) = L_2$, encontramos, despejando, que

$$Lp_i = \frac{53\,977}{45\,273} - \frac{1}{45\,273}\sqrt{(2325\,420\,259 - 4527\,300\,000 L_i)}, \quad \text{para} \quad i = 1, 2;$$

esto en el caso $p \leq 1 (M \leq 0.5)$, en el caso de $p \geq 1 (M \geq 0.5)$, tendríamos,

$$Lp_i = -75820 \frac{L_i}{-1.0 \times 10^5 + 1.200\,1 \times 10^5 L_i}, \quad \text{para} \quad i = 1, 2.$$

Ejemplo del uso de los ajustes junto con los intervalos de confianza.

Paso 1. Elección del tamaño de muestra para la generación del intervalo.

En vista de que en principio no se tiene información del posible valor de p, usaremos un valor ligeramente más grande que el sugerido por la gráfica del tamaño de muestra mínimo para una longitud del intervalo de 0.2,en M, con una confianza del 90%, es decir, tomaremos $N = 288$. Entonces realizando una simulación con una normal ($p = 2$), tenemos con $\chi^2_{\alpha,1} = 2.70554$, y $\frac{\bar{Y}}{S_X} = 0.81161$,

$$\begin{aligned}
\left[\frac{\bar{Y}}{S_X} - \sqrt{\frac{\chi^2_{\alpha,1}\left(1-(\frac{\bar{Y}}{S_X})^2\right)}{N}}\right]^2 &< M < \left[\frac{\bar{Y}}{S_X} + \sqrt{\frac{\chi^2_{\alpha,1}\left(1-(\frac{\bar{Y}}{S_X})^2\right)}{N}}\right]^2 \\
\left[0.81161 - \sqrt[2]{\frac{0.92337}{288}}\right]^2 &< M < \left[0.81161 + \sqrt[2]{\frac{0.92337}{288}}\right]^2 \\
(0.81161 - 0.056623)^2 &< M < (0.81161 + 0.056623)^2 \\
0.57001 &< M < 0.75383.
\end{aligned}$$

En este caso, $L_1 = 0.57001$, y $L_2 = 0.75383$.

Paso 2. Los dos límites son mayores que 0.5, entonces, sustituímos los límites en,

$$\begin{aligned}
Lp_1 &= -75820\frac{L_1}{-1.0 \times 10^5 + 1.200\,1 \times 10^5 L_1} \\
&= -75820\frac{0.57001}{-1.0 \times 10^5 + 1.200\,1 \times 10^5 (0.57001)} \\
&= 1.368,
\end{aligned}$$

y lo mismo hacemos con el segundo límite,

$$\begin{aligned} Lp_2 &= -75820\frac{L_2}{-1.0\times 10^5 + 1.200\,1\times 10^5 L_2} \\ &= -75820\frac{0.75383}{-1.0\times 10^5 + 1.200\,1\times 10^5(0.75383)} \\ &= 5.9956. \end{aligned}$$

Así, un intervalo para p, sería, con una confianza del 90%,

$$1.368 < p < 5.9956.$$

El intervalo encontrado es relativamente grande, sin embargo, queda ejemplificada la ventaja de obtener un intervalo para M. Podemos disminuir el intervalo para M con el fin de obtener una disminución sustancial para el intervalo de p,investigando qué factor se le puede añadir al límite inferior del intervalo para M, y qué factor se le puede restar al límite superior del intervalo para M. En principio tal factor dependerá del tamaño muestral y posiblemente de la confianza establecida.

Con el factor de corrección $1/\sqrt{N}$, y $2/\sqrt{N}$, los límites para M serían: $L_1 = 0.57001 + 0.05893 = 0.62894$, $L_2 = 0.75383 - .11785 = 0.63598$, así, los límites correspondientes para p, son

$$\begin{aligned} Lp_1 &= -75820\frac{L_1}{-1.0\times 10^5 + 1.200\,1\times 10^5 L_1} \\ &= -75820\frac{0.62894}{-1.0\times 10^5 + 1.200\,1\times 10^5(0.62894)} \\ &= 1.9447, \end{aligned}$$

y,

$$
\begin{aligned}
Lp_2 &= -75820\frac{L_2}{-1.0\times 10^5 + 1.200\,1\times 10^5(L_2)} \\
&= -75820\frac{0.63598}{-1.0\times 10^5 + 1.200\,1\times 10^5(0.63598)} \\
&= 2.0367.
\end{aligned}
$$

Por lo tanto, el intervalo para p, con el factor de corrección utilizado, sería,

$$1.9447 < p < 2.0367.$$

En este ejemplo se cubrió el valor verdadero de p y obtuvimos una longitud del intervalo de 0.092, para p. Hemos encontrado un intervalo más angosto, sin embargo, es posible que nuestra confianza halla cambiado en el desarrollo; para analizar este comportamiento realizamos unas simulaciones para la distribución Normal, para distintas confianzas.

5.2 Simulación

En este apartado mostramos los resultados de la simulación que se efectuó, con el fin de hallar una mejor corrección en cuanto a la disminución de la longitud de los intervalos estimados sin afectar, claro está, la confianza especificada.

Usando el paquete S-Plus, con base en el algoritmo que se muestra en el apéndice, se simularon 5000 muestras de tamaño 400 de una variable aleatoria

normal con media cero y desviación estándar de uno. Los porcentajes de cubrimiento teóricos para M, usando

$$L_1 < M < L_2,$$

donde,

$$L_1 = \left(\frac{\bar{Y}}{S_X} - \sqrt[2]{\frac{\chi^2_{\alpha,1}\left(1-\left(\frac{\bar{Y}}{S_X}\right)^2\right)}{N}} \right)^2, \qquad L_2 = \left(\frac{\bar{Y}}{S_X} + \sqrt[2]{\frac{\chi^2_{\alpha,1}\left(1-\left(\frac{\bar{Y}}{S_X}\right)^2\right)}{N}} \right)^2,$$

resultaron ser muy conservadores; usando la corrección propuesta anteriormente:

$$L_1 + \frac{1}{\sqrt{N}} < M < L_2 - \frac{2}{\sqrt{N}},$$

pudimos observar que los porcentajes se alteraban sustancialmente al disminuir la confianza, lo cual afectaba a su vez los porcentajes de cubrimiento para el parámetro p. En vista de los resultados decidimos mejorar nuestra corrección y realizar la simulación probando distintas correcciones. Los resultados mejoraron notoriamente. Los intervalos finales para M y p, fueron, respectivamente:

$$L_1 + \frac{0.985 + \alpha}{\sqrt{N}} < M < L_2 - \frac{(1.85 - 10\alpha)}{\sqrt{N}},$$

$$Lp_1 < p < Lp_2,$$

con;

$$Lp_1 = -75820\frac{L_1+\frac{0.985+\alpha}{\sqrt{N}}}{-1.0\times10^5+1.2001\times10^5(L_1+\frac{0.985+\alpha}{\sqrt{N}})},$$

$$Lp_2 = -75820\frac{L_2-\frac{(1.85-10\alpha)}{\sqrt{N}}}{-1.0\times10^5+1.2001\times10^5(L_2-\frac{(1.85-10\alpha)}{\sqrt{N}})},$$

ya que $L_1 + \frac{0.985+\alpha}{\sqrt{N}} \geq 0.5$, y $L_2 - \frac{(1.85-10\alpha)}{\sqrt{N}} \geq 0.5$. En caso de que fueran menor que 0.5 se usan las otras expresiones definidas anteriormente para Lp_1 y Lp_2. Nótese que el intervalo depende del tamaño de muestra y de la confianza utilizada. A continuación, mostramos una tabla con los porcentajes de cubrimiento hallados en la simulación usando la última correción:

α	Cobertura para M %	Cobertura para p%
0.01	0.9982	0.9992
0.02	0.9852	0.9900
0.03	0.9752	0.9808
0.04	0.9636	0.9690
0.05	0.9608	0.9628
0.06	0.9485	0.9462
0.07	0.9350	0.9234
0.08	0.9256	0.9036
0.09	0.9110	0.8854
0.10	0.8894	0.8492

Se puede apreciar una excelente cobertura, tanto para M como para p, sobre todo para $0.01 \leq \alpha \leq 0.07$.

Capítulo 6

Conclusiones

Nos hemos esforzado en realizar un estudio apegado lo más posible a la teoría estadística y matemática, y pensamos que hemos hallado resultados satisfactorios.

Hemos encontrado las propiedades del estimador $\frac{\bar{Y}}{\hat{\sigma}_X}$ al identificarlo como el estimador de momentos que interviene, considerando la variable aleatoria $Y = |X|$, en la estimación de p usando la función de razón gaussiana generalizada.

Se aplicó la expansión de Edgeworth con éxito en el cálculo de la función de densidad relevante, obteniendo a partir de ella, y en el orden ya especificado, los tamaños de muestra mínimos necesarios para una estimación confiable. En particular hallamos que la distribución de Laplace se muestra como un punto de referencia a la hora de escoger los tamaños de muestra necesarios. Hemos logrado fabricar intervalos de confianza para el parámetro de interés con las consecuentes ventajas, y dejamos para otras posibles investigaciones la posibilidad de realizar ajustes para hallar resultados óptimos, estos podrían hacerse por ejemplo en el ajuste de las dos curvas a la función $M(p)$, o buscar un mejor factor de corrección para los límites del intervalo para $M(p)$, que produzca intervalos con la confianza y longitud requeridas en la estimación de p. También sería deseable buscar una posible disminución en los tamaños de muestra, lo cual creemos se puede lograr ampliando la presente investigación.

El hallar intervalos de confianza para un parámetro determinado nos indica que, en principio, se puede plantear, al menos teóricamente, la posibilidad de construir pruebas de hipótesis para tal parámetro, aunque quedaría pendiente su aplicación.

Queda demostrada la aplicabilidad de la expansión de Edgeworth y animamos a que sea contemplada como una buena técnica de aproximación a sumas de variables aleatorias.

Nos queda también, para posibles investigaciones futuras, el uso de los teoremas de inversión para obtener las distribuciones acumuladas y valores de la suma de variables aleatorias, con su posible aplicación; para ello recomendamos la referencia [2].

Apéndice A

Algoritmo para los intervalos de confianza para M y para p. Porcentaje de intervalos que cubren a los parámetros.

El siguiente algoritmo calcula la frecuencia de cubrimiento de los intervalos de confianza usando la generación de muestras aleatorias de tamaño N, de una cierta distribución de probabilidad. Se recomienda usar un paquete estadístico tal como el S-Plus para la ejecución

Contadores: $I = 0$, $J = 0$, $Q = 0$

Variables de inicio: $H_1 = 0.63662$, $H_2 = 2$, $\alpha = .01$

1.- Generar muestras de tamaño 400 para una distribución normal estándar: $X_i \sim N(0,1)$

2.- Calcular los valores absolutos Y_i de la muestra: $Y_i = |X_i|$

3.- Calcular $\frac{\bar{y}}{s_X} = \frac{\frac{1}{N}\sum_i^N |x_i|}{\sqrt{\frac{1}{N}\sum_i^N x_i^2}} = \frac{\frac{1}{N}\sum_i^N y_i}{\sqrt{\frac{1}{N}\sum_i^N y_i^2}}$

4.- Para α dada, calcular los límites inferior: $L_1 = \left(\frac{\bar{Y}}{S_X} - \sqrt{\frac{\chi^2_{\alpha,1}\left(1-\left(\frac{\bar{Y}}{S_X}\right)^2\right)}{N}} \right)^2$,

y

superior: $L_2 = \left(\frac{\bar{Y}}{S_X} + \sqrt{\frac{\chi^2_{\alpha,1}\left(1-\left(\frac{\bar{Y}}{S_X}\right)^2\right)}{N}} \right)^2$.

5.- Contador: si $L_1 \leq H_1 \leq L_2$, entonces $I = I + 1$

6.- Aplicar corrección: $LC_1 = L_1 + \frac{1}{\sqrt[2]{N}}$, $LC_2 = L_2 - \frac{2}{\sqrt[2]{N}}$

7.- Contador: si $LC_1 \leq H_1 \leq LC_2$, entonces $J = J + 1$

8.- Si $LC_1 < \frac{1}{2}$, entonces calcular límite inferior del intervalo para p:

$Lp_1 = \frac{53\,977}{45\,273} - \frac{1}{45\,273}\sqrt{(2325\,420\,259 - 4527\,300\,000 LC_1)}$

9.- Si $LC_1 \geq \frac{1}{2}$, entonces calcular límite inferior del intervalo para p:

$Lp_1 = -75820 \frac{LC_1}{-1.0\times10^5 + 1.200\,1\times10^5 LC_1}$

10.- Si $LC_2 < \frac{1}{2}$, entonces calcular límite superior del intervalo para p:

$Lp_2 = \frac{53\,977}{45\,273} - \frac{1}{45\,273}\sqrt{(2325\,420\,259 - 4527\,300\,000 LC_2)}$

11.- Si $LC_2 \geq \frac{1}{2}$, entonces calcular límite superior del intervalo para p:

$Lp_2 = -75820 \frac{LC_2}{-1.0\times10^5 + 1.200\,1\times10^5 LC_2}$

12.- Contador: si $Lp_1 \leq H_2 \leq Lp_2$, entonces $Q = Q + 1$

13.- Repetir pasos 1 al 11: 1000 veces

14.- **Reportar**: $ICM = \frac{I}{N}$, $ICMC = \frac{J}{N}$, $ICP = \frac{Q}{N}$

15.- Repetir pasos 1 al 14 para $\alpha = 0.05$, y $\alpha = 0.02$

16.- **Repetir pasos 1 al 15 pero, para una distribución doble exponencial (Laplace) con media cero y desviación estándar de 1, y con un tamaño de muestra de 550, en este caso:** $H_1 = 0.5$, $H_2 = 1$, **con los contadores en ceros:** $I = 0$, $J = 0$. $Q = 0$.

Bibliografía

[1] Arfken, G.B., (1970) *Mathematical Methods for Physicists*, Academic Press.

[2] Barndorff-Nielsen, O. E., Cox, D.R., (1989) *AsymptoticTechniques for Use in Statistics*, Chapman and Hall.

[3] Butkov, E.,(1973) *Mathematical Physiscs*, Addison Wesley.

[4] Casella, G.,Berger, R.L., (1990) *Statistical Inference*, Duxbury Press.

[5] Devore, J.L., (1995) *Probability and Statistics for Engineering and the Sciences*, 4th Edition,Duxbury Press.

[6] Fuller, W.A., (1976) *Introduction to Statistical Time Series*, New York:John Wiley & Sons.

[7] Gonzalez, G., Iglesias, S., López, J., (1998) Notas de Estadística Matemática, Universidad Virtual, ITESM.

[8] Kendall, M., Stuart, A., (1979) *The Advanced Theory of Statistics, Volume II: Inference and Relationship.*

[9] Rice, J.A., (1995) *Mathematical Statistics and Data Analysis*, Second Edition, Duxbury Press.

[10] Rodríguez-Dagnino, R. M., León-García, A., (1998) "An Explicit Estimator for the Shape Parameter of the Generalized Gaussian Distribution". Comunicación Interna. Presentada en el XII Foro Nacional de Estadística, Monterrey, N. L.

[11] Stremler, F.G., (1990) *Introducción a los sistemas de comunicación*, 3a Ed.,Addison-Wesley Iberoamericana.

[12] Takeuchi,Yu, (1976) *Sucesiones y Series*, Tomo II, Ed. Limusa.

[13] Varanasi, M.K., Aazhang, B., "Parametric generalized Gaussian density estimation", J. Acoust. Soc. Am.86 (4), October 1989, pp. 1404.

Centro de Información-Biblioteca
3000200583977 4