



COLEGIO DE POSTGRADUADOS

---

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN  
EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA  
ESTADÍSTICA

**PRUEBAS DE RAZÓN DE VARIANZAS  
PARA LA DISTRIBUCIÓN INVERSA  
GAUSSIANA MEDIANTE EL USO DE  
BOOTSTRAP PARAMÉTRICO**

JUAN ISIDRO ESPINOZA CASIMIRO

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA  
OBTENER EL GRADO DE:

MAESTRO EN CIENCIAS

MONTECILLO, TEXCOCO, ESTADO DE MÉXICO  
ENERO DE 2019

---

**CARTA DE CONSENTIMIENTO DE USO DE LOS DERECHOS DE AUTOR Y DE LAS REGALIAS COMERCIALES DE PRODUCTOS DE INVESTIGACION**

En adición al beneficio ético, moral y académico que he obtenido durante mis estudios en el Colegio de Postgraduados, el que suscribe JUAN ISIDRO ESPINOZA CASIMIRO, Alumno (a) de esta Institución, estoy de acuerdo en ser partcipe de las regalías económicas y/o académicas, de procedencia nacional e internacional, que se deriven del trabajo de investigación que realicé en esta institución, bajo la dirección del Profesor DR. JOSÉ A. VILLASEÑOR ALVA, por lo que otorgo los derechos de autor de mi tesis PRUEBAS DE RAZÓN DE VARIANZAS PARA LA DISTRIBUCIÓN INVERSA GAUSSIANA MEDIANTE EL USO DE BOOTSTRAP PARAMÉTRICO

y de los productos de dicha investigación al Colegio de Postgraduados. Las patentes y secretos industriales que se puedan derivar serán registrados a nombre del colegio de Postgraduados y las regalías económicas que se deriven serán distribuidas entre la Institución, El Consejero o Director de Tesis y el que suscribe, de acuerdo a las negociaciones entre las tres partes, por ello me comprometo a no realizar ninguna acción que dañe el proceso de explotación comercial de dichos productos a favor de esta Institución.

Montecillo, Mpio. de Texcoco, Edo. de México, a 29 de ENERO de 2019



Firma del  
Alumno (a)



Vo. Bo. del Consejero o Director de Tesis

La presente tesis titulada: **Pruebas de razón de varianzas para la distribución Inversa Gaussiana mediante el uso de bootstrap paramétrico**, realizada por el alumno: **Juan Isidro Espinoza Casimiro**, bajo la dirección del Consejo Particular indicado ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

**MAESTRO EN CIENCIAS**

**SOCIOECONOMÍA, ESTADÍSTICA E INFORMATICA  
ESTADÍSTICA**

**CONSEJO PARTICULAR**

CONSEJERO



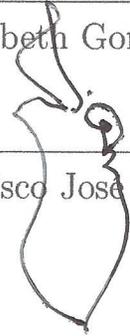
Dr. José Aurelio Villaseñor Alva

ASESOR



Dra. Elizabeth González Estrada

ASESOR



Dr. Francisco José Zamudio Sánchez

Montecillo, Texcoco, Estado de México, Enero de 2019.

# PRUEBAS DE RAZÓN DE VARIANZAS PARA LA DISTRIBUCIÓN INVERSA GAUSSIANA MEDIANTE EL USO DE BOOTSTRAP PARAMÉTRICO

JUAN ISIDRO ESPINOZA CASIMIRO, M.C.

Colegio de Postgraduados, 2019

## RESUMEN

Se hizo la implementación del método de bootstrap paramétrico para una estadística de prueba de bondad de ajuste para la distribución Inversa Gaussiana. Esta estadística es una razón de varianzas y tiene normalidad asintótica, por lo cual solo se emplea para muestras grandes. Mediante simulación se encontró que con el bootstrap paramétrico, la potencia de la prueba mejora con respecto a usar su aproximación asintótica, sin embargo, es baja en determinadas situaciones. Por esta razón se analizó implementar el bootstrap paramétrico a una versión más sencilla de la estadística de prueba, la cual mostro una mejor potencia. Después se comparó contra otras pruebas de bondad de ajuste la implementación de la estadística sencilla concluyendo que tiene una muy buena potencia de prueba cuando la distribución alternativa es una gamma o una weibull.

**Palabras claves:** Bootstrap paramétrico, potencia de prueba, Inversa Gaussiana, bondad de ajuste.

# VARIANCE RATIO TESTS FOR THE INVERSE GAUSSIAN DISTRIBUTION BY USING PARAMETRIC BOOTSTRAP

JUAN ISIDRO ESPINOZA CASIMIRO, M.Sc.

Colegio de Postgraduados, 2019

## ABSTRACT

The implementation of the parametric bootstrap method was made for a goodness-of-fit test statistic for the inverse Gaussian distribution. This statistic is a variance ratio and has asymptotic normality, which is why it is only used for large samples. Through simulation it was found that with parametric bootstrap, the power of the test improves with respect to using its asymptotic approach, however, it is low in certain situations. For this reason, it was analyzed to implement the parametric bootstrap to a simpler version of the test statistic, which showed better power. The implementation of the simple statistic was then compared against other goodness-of-fit tests, concluding that it has very good test power when the alternative distribution is a gamma or a weibull.

**Keywords:** Parametric bootstrap, power of the test, inverse Gaussian, goodness-of-fit.

---

## AGRADECIMIENTOS

Al CONACYT por el respaldo económico.

Al Colegio de Postgraduados por la oportunidad brindada.

Al Consejo particular por su apoyo. Además le agradezco al Dr. Villaseñor su valiosa dirección en la tesis, a la Dra. Elizabeth por la revisión esmerada de esta, al Dr. Zamudio por sus consejos.

A la planta docente del Departamento de Estadística por su excelente calidad.

A mis compañeros por haber tenido una buena convivencia.

Al área administrativa por su amable apoyo.

A mi Veronica por haberme apoyado durante este proceso.

A mi hermano y mis padres por su apoyo.

---

# DEDICATORIAS

A mis hijas Ariadna Sofía y Elisa Pilar.

# CONTENIDO

<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos	2
<b>2. Antecedentes</b>	<b>3</b>
2.1. Pruebas de Bondad de Ajuste	3
2.2. Distribución Inversa Gaussiana	4
<b>3. Marco teórico</b>	<b>5</b>
3.1. Distribución Inversa Gaussiana	5
3.1.1. Distribuciones relacionadas con la distribución IG	7
3.2. Pruebas de bondad de ajuste	7
3.3. Bootstrap paramétrico	9
<b>4. Pruebas de bondad de ajuste para la distribución Inversa Gaussiana</b>	<b>12</b>
4.1. Prueba de bondad de ajuste de razón de varianzas	12
4.2. Pruebas basadas en datos transformados	14
4.2.1. Pruebas t	14
4.2.2. Prueba de Anderson-Darling	15
4.3. Prueba de bondad de ajuste de razón de verosimilitud	16
<b>5. Comparación de tamaño de prueba y potencias de las pruebas</b>	<b>17</b>
5.1. Reproducción del artículo Villaseñor y González-Estrada (2015a)	18
5.2. Desempeño del bootstrap paramétrico vs versión asintótica	21
5.3. Desempeño bootstrap paramétrico vs otras pruebas	24
5.4. Discusión	26
<b>6. Ejemplo</b>	<b>27</b>
<b>7. Conclusiones</b>	<b>31</b>

# CONTENIDO

---

<b>Bibliografía</b>	<b>34</b>
<b>Anexos</b>	<b>35</b>
Anexo A: Aspecto computacionales . . . . .	35

# LISTA DE FIGURAS

3.1. Curvas de densidades de la distribución $IG(1, \lambda)$ para distintos valores de $\lambda$ y $\mu = 1$ . . . . .	5
3.2. Densidades IG vs otras densidades . . . . .	7
5.1. Tamaño estimado de las pruebas, $n=25,50,100,300,500$ . . . . .	18
5.2. Potencia estimada de las pruebas vs la distribución $\Gamma(\alpha, 1)$ , $n=25,50,100,300,500$	19
5.3. Potencia estimada de las pruebas vs la distribución $Weibull(\zeta, 1)$ , $n=25,50,100,300,500$	19
5.4. Potencia estimada de las pruebas vs la distribución $LogNormal(0, \sigma)$ , $n=25,50,100,300,500$	20
5.5. Tamaño estimado de las pruebas . . . . .	21
5.6. Potencia estimada de las pruebas vs la distribución $\Gamma(\alpha, 1)$ , $n=25,50,100,300,500$	22
5.7. Potencia estimada de las pruebas vs la distribución $Weibull(\zeta, 1)$ , $n=25,50,100,300,500$	23
5.8. Potencia estimada de las pruebas vs la distribución $LogNormal(0, \sigma)$ , $n=25,50,100,300,500$	23
5.9. Tamaño estimado de las pruebas . . . . .	24
5.10. Potencia estimada de las pruebas vs la distribución $\Gamma(\alpha, 1)$ , $n=25,50,100,300,500$	25
5.11. Potencia estimada de las pruebas vs la distribución $Weibull(\zeta, 1)$ , $n=25,50,100,300,500$	25
5.12. Potencia estimada de las pruebas vs la distribución $LogNormal(0, \sigma)$ , $n=25,50,100,300,500$	26
6.1. Distribución IG ajustada y Función de distribución empírica de los datos de precipitación . . . . .	29

# LISTA DE TABLAS

6.1. Precipitaciones mensuales de Huautla de Jimenez . . . . .	28
6.2. Aplicación de las pruebas del capítulo 4 a las precipitaciones mensuales de Huautla de Jiménez . . . . .	29

# Capítulo 1

## Introducción

De acuerdo con [RA Fisher \(1922\)](#), los problemas estadísticos se dividen en tres clases:

1. Problemas de especificación: La elección del modelo matemático que caracteriza a la población, principalmente especificar correctamente el comportamiento probabilístico de una o más variables aleatorias del modelo matemático con que se representa algún comportamiento de la población. Por esta razón es recomendable aplicar las pruebas de bondad de ajuste a los datos antes de sacar conclusiones a partir de estos.
2. Problemas de estimación: Los procedimientos para obtener los estadísticos que sirven para estimar los parámetros del modelo teórico de la población.<sup>1</sup>
3. Problemas de distribución: Qué distribución tienen los estadísticos obtenidos.

Las pruebas de bondad de ajuste ayudan en los problemas de especificación minimizando el riesgo de asignarle una distribución incorrecta a un conjunto de datos. Sean  $X_1, \dots, X_n$  una muestra aleatoria de una población con función de distribución continua con soporte en el conjunto de los reales positivos. En este trabajo se estudiará el problema de probar  $H_0 : X_1, \dots, X_n \sim IG(\mu, \lambda)$ ,  $\mu > 0$ ,  $\lambda > 0$  desconocidos vs  $H_1 : X_1, \dots, X_n \not\sim IG(\mu, \lambda)$ . En esta tesis se trabajará con 1 prueba de bondad de ajuste para la distribución Inversa Gaussiana (IG) presentada por [Villaseñor y González-Estrada \(2015a\)](#), el estadístico de esta prueba es  $T_{1,n} = \sqrt{\frac{n}{6\bar{X}_n V_n}} (R_n - 1)$ , donde  $R_n = \frac{S_n^2}{\sigma^2}$  que a su vez también puede servir como estadística de prueba,  $\bar{X}_n = \sum_{i=1}^n \frac{X_i}{n}$ ,  $V_n = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{X_i} - \frac{1}{\bar{X}_n} \right)$ . Cuando  $H_0$  es cierta y  $n \rightarrow \infty$  entonces  $T_{1,n} \xrightarrow{d} N(0, 1)$ , esta propiedad asintótica nos permite probar  $H_0$  cuando la muestra es grande. En el caso de muestras pequeñas no es válida su

---

<sup>1</sup>[RA Fisher \(1922\)](#) es el primero que especifica los terminos parámetros y estimadores.

## 1.1. Objetivos

---

aplicación, por esta razón surgió el interés de estudiar la factibilidad de la implementación del bootstrap paramétrico para aproximar la distribución del estadístico de prueba, de la cual se obtienen las constantes críticas requeridas para hacer la prueba.

En el capítulo 2 se tratará sobre los aspectos históricos de las pruebas de bondad de ajuste y la distribución Inversa Gaussiana, mientras tanto en el capítulo 3 abordará la parte estadística. A lo largo del capítulo 4 mostraremos como se derivan los estadísticos  $T_{1,n}$  y  $R_n$  y se describirán otras pruebas de bondad de ajuste mencionadas en [Villaseñor y González-Estrada \(2015a\)](#), dado que en el siguiente capítulo se hará un estudio de la potencia y tamaño de prueba mediante simulación para comparar el desempeño de la implementación del bootstrap paramétrico de los estadísticos  $T_{1,n}$  y  $R_n$  contra otras pruebas de bondad de ajustes disponibles.

## 1.1. Objetivos

- Estudiar la factibilidad del método de Bootstrap paramétrico para aproximar las distribuciones nulas de los estadísticos  $R_n$  y  $T_{1,n}$ .
- Comparar el tamaño de prueba y potencia de las pruebas basadas en los estadísticos de prueba  $R_n$  y  $T_{1,n}$  con la implementación de bootstrap paramétrico con otras pruebas de bondad de ajuste para la distribución IG.

# Capítulo 2

## Antecedentes

### 2.1. Pruebas de Bondad de Ajuste

Las pruebas de bondad de ajuste consisten en rechazar o no rechazar la hipótesis nula de que una muestra aleatoria proviene de una distribución dada, lo cual se puede realizar con métodos subjetivos o formales. Los métodos subjetivos son técnicas gráficas en las que el investigador mediante una inspección visual de los datos aplica la prueba de bondad de ajuste, la prueba más popular de este tipo es la gráfica cuantil-cuantil, que en sus orígenes fue esbozada por Francis Galton en su artículo [Galton \(1899\)](#).

Los métodos formales empiezan con la prueba de Chi-Cuadrada propuesta por Karl Pearson en [Pearson \(1900\)](#)<sup>1</sup>. Pearson introdujo este método cuando tuvo que abandonar la hipótesis de que todas las poblaciones biológicas provienen de distribuciones normales, como se explican en la siguiente cita:

*«In the course of his Mathematical Contributions to the Theory of Evolution, Karl Pearson abandoned the assumption that biological populations are normally distributed, introducing the Pearson system of distributions to provide other models. The need to test fit arose naturally in this context, and in 1900 Pearson invented his chi-squared test. This statistic and others related to it remain among the most used statistical procedures.»* [D'Agostino y Stephens \(1986, p. 63\)](#).

A partir de esta prueba han surgido variantes, conocidas como pruebas tipo chi-cuadrada. Las pruebas tipo chi-cuadrada son más universales en cuanto al tipo de hipótesis que se puede probar, pero menos potentes que otros tipos de pruebas. Aparte de las pruebas de tipo de Chi-Cuadrada existen también las pruebas basadas en la función de distribución empírica, o pruebas basadas en caracterizaciones de la

---

<sup>1</sup>El nacimiento de la estadística matemática se considera que inicia a partir de la publicación de este artículo.

## 2.2. Distribución Inversa Gaussiana

---

distribución. De acuerdo a [Ochoa \(2015\)](#) para la distribución IG existen pruebas disponibles basadas en la función de distribución empírica, por ejemplo Anderson-Darling, Kolmogorov-Smirnov, [O'Reilly y Rueda \(1992\)](#), pruebas basadas en caracterizaciones, por ejemplo [Natarajan y Mudholkar \(2004\)](#). [D'Agostino y Stephens \(1986\)](#) es un texto muy completo sobre las pruebas de bondad de ajuste, en donde se ilustran varios aspectos sobre ellas, por lo cual se recomienda consultarlo en caso de querer profundizar su estudio.

## 2.2. Distribución Inversa Gaussiana

El experimento del gato de Schrödinger es una paradoja muy popular en la física cuántica y fue propuesta por el físico austriaco Erwin Schrödinger. Este investigador también contribuyó a la probabilidad y a la estadística, dentro de sus aportes en estos campos, fue el primero en derivar la distribución Inversa Gaussiana en su artículo [Schrödinger \(1915\)](#) en el cual resuelve un problema de cuando las partículas brownianas son afectadas por fuerzas externas. De acuerdo a [Banerjee \(1986\)](#), Abraham Wald obtuvo de manera independiente la distribución IG como una distribución límite del tamaño muestral en una prueba secuencial de hipótesis, por este hecho a la distribución IG se le conoce también como distribución de Wald, sobre todo en la literatura rusa.

El nombre de distribución Inversa Gaussiana se debe a que [Tweedie \(1945\)](#) encontró que su función generadora de cumulantes tiene una relación inversa con la de la distribución normal, el mismo autor hace un desarrollo de las propiedades estadísticas de la distribución IG en [Tweedie \(1957\)](#) y también engloba en [Tweedie \(1984\)](#) a la distribución IG en una familia de distribuciones más general que será llamada distribución de Tweedie en su honor. Dentro de las bibliografías más completas sobre la distribución IG se puede citar a [Chhikara \(1988\)](#) y a [Seshadri \(2012\)](#), por lo cual si se desea profundizar en el estudio de esta distribución es recomendable acudir a estas fuentes.

Hay varias alternativas de pruebas de bondad de ajuste para la distribución IG, cada una con sus ventajas y desventajas. Por ejemplo es posible aplicar una prueba chi-cuadrada, pero tiene poca potencia, lo cual implica que podemos concluir que una muestra proviene de la distribución IG aunque esto sea falso, por eso es conveniente explorar otras alternativas más prometedoras como la que se estudia en este trabajo.

# Capítulo 3

## Marco teórico

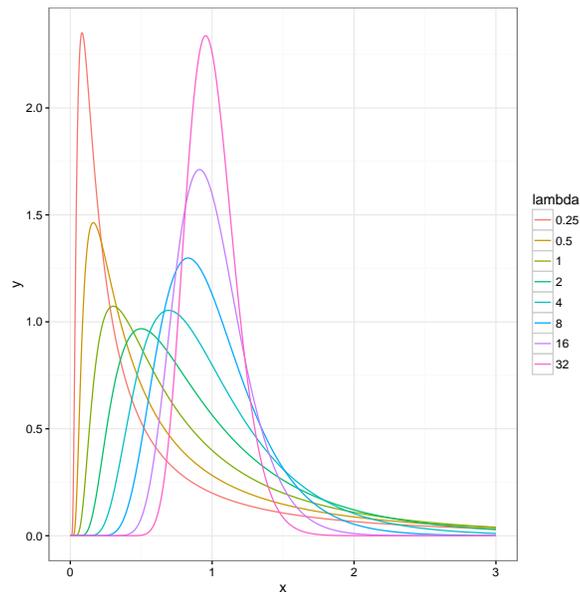
### 3.1. Distribución Inversa Gaussiana

Tomaremos como referencia en esta subsección a [Tweedie \(1957\)](#). De este artículo presentaremos algunos resultados. La variable aleatoria  $X$  tiene distribución Inversa Gaussiana si tiene la siguiente función de densidad

$$f_X(x; \mu, \lambda) = \exp\left\{-\frac{\lambda}{2\mu^2 x}(x - \mu)^2\right\} \sqrt{\frac{\lambda}{2\pi x^3}}; x > 0, \mu > 0, \lambda > 0.$$

A esta distribución se le denotará como  $IG(\mu, \lambda)$ .

Figura 3.1: Curvas de densidades de la distribución  $IG(1, \lambda)$  para distintos valores de  $\lambda$  y  $\mu = 1$



### 3.1. Distribución Inversa Gaussiana

---

La función de distribución acumulada de  $X$  es:

$$F_X(x; \mu, \lambda) = \Phi \left[ \sqrt{\frac{\lambda}{x}} \left( \frac{x}{\mu} - 1 \right) \right] + e^{2\frac{\lambda}{\mu}} \Phi \left[ -\sqrt{\frac{\lambda}{x}} \left( \frac{x}{\mu} + 1 \right) \right]$$

donde  $\Phi(\cdot)$  denota la distribución normal estándar.

Su función generadora de momentos es:

$$M_X(t) = \exp \left\{ \frac{\lambda}{\mu} \left[ 1 - \left( 1 - \frac{2\mu^2 t}{\lambda} \right)^{1/2} \right] \right\}, t < \frac{\lambda}{2\mu^2}$$

Los primeros momentos de  $X$  son:

$$E(X) = \mu$$

$$E(X^2) = \mu^2 + \frac{\mu^3}{\lambda}$$

En ocasiones es conveniente hacer reparametrizaciones como la que se muestra a continuación:

$$f_X(x; \mu, \phi) = \exp \left\{ -\frac{\phi x}{2\mu} + \phi - \frac{\mu\phi}{2x} \right\} [\mu\phi/2\pi x^3]^{1/2}$$

o

$$f_X(x; \phi, \lambda) = \exp \left\{ -\frac{\phi^2 x}{2\lambda} + \lambda - \frac{\lambda}{2x} \right\} [\lambda/2\pi x^3]^{1/2}$$

donde  $\phi = \lambda/\mu$ .

Sea  $X_1, \dots, X_n$  una muestra aleatoria de la distribución IG. Se tiene que  $\hat{\mu} = \bar{X}$  es el estimador de momentos, de máxima verosimilitud (EMV) y el estimador insesgado de varianza uniformemente mínima (EIVUM) para  $\mu$ . Mientras que para  $\lambda$  se tiene que el estimador de momentos es:

$$\hat{\lambda}_{MM} = \frac{\bar{X}^3}{S_X^2}$$

El EMV de  $\frac{1}{\lambda}$  es:

$$\frac{1}{\hat{\lambda}_{ML}} = \frac{\sum_{i=1}^n (X_i^{-1} - \bar{X}^{-1})}{n}$$

y el EIVUM es:

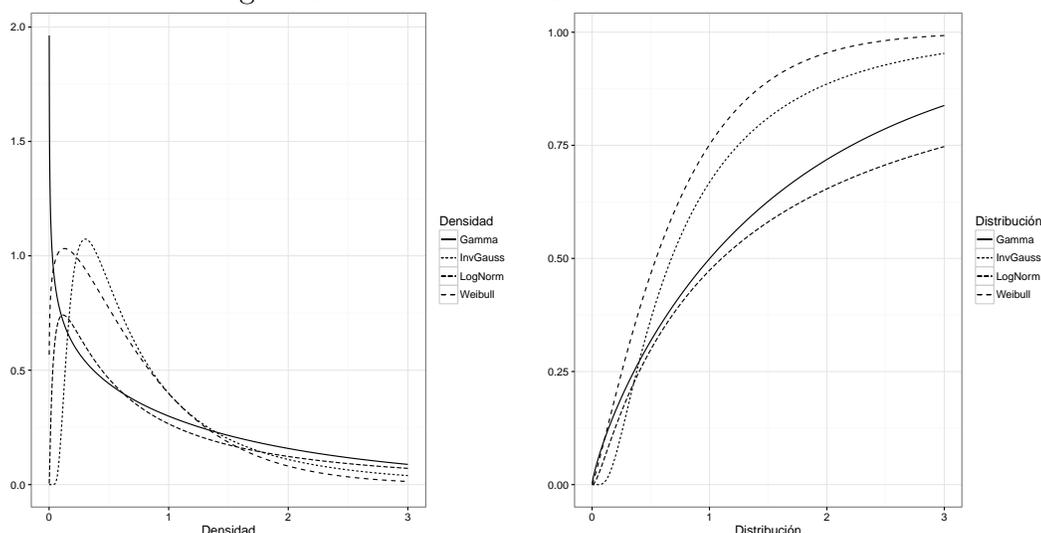
## 3.2. Pruebas de bondad de ajuste

$$\frac{1}{\hat{\lambda}_{MVUE}} = \frac{\sum_{i=1}^n (X_i^{-1} - \bar{X}^{-1})}{n-1}$$

### 3.1.1. Distribuciones relacionadas con la distribución IG

La distribución IG es una distribución asimétrica, comparte con las distribuciones Gamma, Log-Normal la convergencia asintótica a la distribución normal, cuando  $\lambda/\mu \rightarrow \infty$ , la distribución de  $X$  es asintóticamente normal con media  $\mu$  y varianza  $\mu^3/\lambda$  (Folks y Chhikara, 1978). Además la distribución IG tiene densidades similares a las densidades de las distribuciones Gamma, Log-normal y Weibull. En muchas ocasiones conjuntos de datos han sido tomados como una u otra de estas distribuciones sin considerar otras opciones, por lo cual sería conveniente aplicar pruebas de bondad de ajuste. Las distribuciones IG, Weibull, Log-normal, Gamma tienen como campo de aplicación principal la confiabilidad y el análisis de sobrevivencia.

Figura 3.2: Densidades IG vs otras densidades



El siguiente resultado relaciona a la distribución IG con la distribución Gamma.

**Teorema 3.1.** Si  $X \sim IG(\mu, \lambda)$  entonces  $\frac{\lambda(X-\mu)^2}{\mu^2 X} \sim \chi_1^2$ . Sea  $Z^{(\mu)} = \frac{(X-\mu)^2}{X}$ , por lo tanto por propiedades de la distribución Gamma  $Z^{(u)} \sim \Gamma\left(\frac{1}{2}, \beta\right)$ , donde  $\beta = 2\frac{\mu^2}{\lambda}$ .

## 3.2. Pruebas de bondad de ajuste

Sea  $X_1, \dots, X_n$  una muestra aleatoria con densidad  $f_X(x; \underline{\theta})$ , donde  $\underline{\theta}$  es un vector de parámetros de tamaño  $k$ , tal que  $\underline{\theta} \subset \Omega \subset \mathbb{R}^k$ , donde  $\Omega$  es el espacio de valores de  $\underline{\theta}$ .

## 3.2. Pruebas de bondad de ajuste

---

**Definición 3.1.** Un parámetro es una característica numérica fija de la distribución de una variable aleatoria obtenida de una población. En general este parámetro es de nuestro interés y debido a que se requiere para su cálculo conocer todos los valores de la población es desconocido en la mayoría de los casos. Típicamente los parámetros se representan con letras griegas.

**Definición 3.2.** Una hipótesis estadística es una conjetura acerca de la distribución de una variable aleatoria. Una hipótesis estadística es una suposición acerca de los valores de un parámetro  $\theta$ , donde  $\theta \in \Omega$ .

Las pruebas de bondad de ajuste son reglas de decisión en base a un estadístico que se construye bajo la hipótesis de que los datos provienen de una distribución  $F_X(x; \theta)$ , se rechaza la hipótesis si el valor del estadístico calculado a partir de la muestra cae en la zona de rechazo o no se rechaza si sucede lo contrario, para definir la zona de rechazo y no rechazo se usan los valores críticos. Dependiendo de la ubicación de la zona de rechazo, una prueba de hipótesis puede ser de una cola, si se ubica en un solo extremo del soporte de la distribución del estadístico, o dos colas si se ubica en ambos extremos.

**Definición 3.3.** Una hipótesis nula  $H_0$  es una afirmación que  $\theta \in w$ , donde  $w \in \Omega$ .

**Definición 3.4.** Una hipótesis alternativa  $H_1$  consiste en suponer que  $\theta \in (\Omega - w)$ , es decir suponer que  $H_0$  es falsa

**Definición 3.5.** Una prueba de hipótesis  $\phi$  es una regla de decisión basada en una muestra aleatoria ( $\underline{X}$ ). Si  $\chi$  denota al conjunto de los valores posibles de  $\underline{X}$ ,  $\underline{X}$  se separa en dos subconjuntos ajenos  $\chi_A$  y  $\chi_R$ . Si  $\underline{X} \in \chi_A$  entonces no se rechaza  $H_0$  y si  $\underline{X} \in \chi_R$  se rechaza  $H_0$ .

**Definición 3.6.** Una estadística de prueba, es una función de la m.a  $\underline{X}$ , digamos  $T = t(\underline{X})$ . En la definición 3.5 se menciona que se toma la decisión con base en la m.a., pero en la práctica se toma en base a  $T$ , ya que la función  $t$  realiza una reducción de dimensionalidad de  $\mathbb{R}^n \rightarrow \mathbb{R}$  haciendo la transformación  $\chi \rightarrow \tau$ , y se particiona  $\tau$  en  $\tau_R$  y  $\tau_A$ .

**Definición 3.7.** Error tipo I, E.I, es cuando se rechaza  $H_0$  siendo que es verdadera.

**Definición 3.8.** Error tipo II, E.II, es cuando no se rechaza  $H_0$ , siendo que es falsa.

**Definición 3.9.** El tamaño de prueba  $\alpha$ , es equivalente a que se cumpla  $\max_{\theta \in W} P(E.I) = \alpha$ .

**Definición 3.10.** La potencia de la prueba es igual a  $1 - P(E.II)$ .

### 3.3. Bootstrap paramétrico

---

**Algoritmo 3.1** Prueba de bondad de ajuste

---

**Entrada:**  $x_1, \dots, x_n$ : Muestra de datos

$\alpha$ : Tamaño de prueba, máxima probabilidad de error tipo I

$k$ : Valores críticos

**VARIABLES DE PROCESO:**

**Salida:** Rechazar o no rechazar la hipótesis nula.

1. Calcular el estadístico de prueba  $T$
  2. Comparar valor calculado de  $T$  contra valores críticos y concluir.
- 

Cuando se comparan varias pruebas de hipótesis, en general se fija el tamaño de prueba, y se elige la prueba que tenga la mayor potencia.

**Definición 3.11.** Una prueba de bondad de ajuste es una prueba de hipótesis donde  $H_0$ : La m.a.  $X_1, X_2, \dots, X_n \sim f_X(x)$  vs  $H_1$ :  $X_1, \dots, X_n \approx f_X(x)$ , donde  $f_X(x)$  es una distribución conocida que puede depender de parámetros desconocidos.

En pruebas de bondad de ajuste se busca no rechazar  $H_0$ . En caso de que se rechace, se concluye que no tiene la distribución especificada en  $H_0$ , pero no aporta información adicional. Dentro de las pruebas de bondad de ajuste, dependiendo de si son conocidos o desconocidos los parámetros de la distribución, tenemos las siguientes situaciones.

**Definición 3.12.** Prueba de bondad de ajuste con hipótesis simple, es cuando todos los parámetros de la distribución  $f_X$  son conocidos.

**Definición 3.13.** Prueba de bondad con hipótesis compuesta, es cuando al menos un parámetro de la distribución en  $H_0$  es desconocido. Por ejemplo cuando  $H_0$ :  $X_1, \dots, X_n \sim IG(1, 1)$  es una hipótesis simple, en cambio  $H_0$ :  $X_1, \dots, X_n \sim IG(\mu, \lambda)$ ,  $\mu, \lambda$  son desconocidos, es una hipótesis compuesta.

### 3.3. Bootstrap paramétrico

Uno de los objetivos de la labor estadística es calcular las distribuciones de los estadísticos, preferentemente estos se obtienen de manera analítica, sin embargo, en ocasiones es imposible poder hacerlo de esta manera y se requiere emplear métodos numéricos como el remuestreo. En [Efron et al. \(1979\)](#) se presentó el método de

### 3.3. Bootstrap paramétrico

---

bootstrap que es una técnica de remuestreo para estimar las distribuciones empíricas de los estadísticos bajo  $H_0$ .

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de la cual obtuvimos un estadístico  $T$  pero no conocemos su distribución de este por diversos motivos, por ejemplo, la muestra es muy pequeña para poder emplear propiedades asintóticas. Hay dos formas de aplicar la técnica de bootstrap:

1. Sea  $X \sim F_X(x; \underline{\theta})$ , donde  $F_X(x; \theta)$  es una distribución que desconocemos. Cuando no se conoce la naturaleza de la distribución de  $X$ , se puede aplicar el bootstrap no paramétrico, que consiste en aplicar un remuestreo con reemplazo a partir de la muestra original para ir obteniendo nuevas muestras, con cada muestra se calcula el estadístico de nuestro interés, con el conjunto de valores calculados del estadístico se puede calcular la función de distribución empírica la cual estima a la función de distribución de dicho estadístico.
2. Sea  $X \sim F_X(x; \underline{\theta})$ , donde  $F_X(x; \theta)$  es una distribución que conocemos. Aprovechando que se conoce la naturaleza de la distribución de  $X$ , se puede aplicar el bootstrap paramétrico, que consiste en remuestrear  $F_X(x; \hat{\theta})$  aplicando el siguiente procedimiento. :

### 3.3. Bootstrap paramétrico

---

#### Algoritmo 3.2 Bootstrap paramétrico

---

**Entrada:**  $X_1, \dots, X_n$ : Muestra de datos

$B$ : Número de réplicas que se generarán a partir de la distribución  $F_X(x; \hat{\theta})$ , donde  $\hat{\theta}$  son los parámetros estimados de  $F_X$ .

$p_1, p_2$ : Probabilidades que definen la zona de rechazo permitida.

**VARIABLES DEL PROCESO:**  $i$ :  $i$ -ésima muestra generada

$T, \hat{\theta}$ : Valor del estadístico de prueba y los parámetros estimados a partir de la muestra.

$T^{(i)}$ :  $i$ -ésima realización del estadístico  $T$  a partir de la  $i$ -ésima muestra.

$\hat{\theta}$ : Vector de los parámetros estimados.

**Salida:** Valores críticos en forma de cuantiles.

1. A partir de la muestra original estimar los parámetros desconocidos de  $F_X(x; \theta)$ , esto se realiza generalmente mediante la técnica de máxima verosimilitud, con estos parámetros estimados se obtiene una distribución estimada  $F_X(x; \hat{\theta})$ .
  2. Fijar  $B$  como el número de réplicas de muestras de  $F_X(x; \hat{\theta})$  que se van a generar.
  3. Fijar  $i = 1$ , donde  $i$  representa la  $i$ -ésima réplica generada.
  4. Generar una muestra pseudoaleatoria del mismo tamaño que la muestra original, con distribución  $F_X(x; \hat{\theta})$ .
  5. Calcular el estadístico a partir de la muestra originada y guardarla
  6. Aumentar  $i = i + 1$
  7. Si  $i \leq B$  regresar al paso 4.
  8. Ordenar en orden ascendentes los valores  $T^{(i)}$ .
  9. Obtener los cuantiles correspondientes a los valores de probabilidades  $p_1, p_2$ , los cuales serán empleados como valores críticos.
-

# Capítulo 4

## Pruebas de bondad de ajuste para la distribución Inversa Gaussiana

El objetivo de esta sección es discutir las pruebas de bondad de ajuste que se comparan en el capítulo 5. La primera prueba que se muestra es una prueba de razón de varianzas desarrollada en [Villaseñor y González-Estrada \(2015a\)](#), de este mismo artículo se mostrarán otras dos pruebas más.

### 4.1. Prueba de bondad de ajuste de razón de varianzas

El desarrollo fue presentado en [Villaseñor y González-Estrada \(2015a\)](#), aquí lo mostramos con propósitos de ilustración. Sea  $X_1, \dots, X_n$  una m.a. de tamaño  $n$ , supongamos que se cumple la hipótesis nula compuesta  $H_0 : X_1, \dots, X_n \sim IG(\mu, \lambda)$ , donde  $\mu$  y  $\lambda$  son desconocidos. Sea  $Z^{(u)}$  la variable aleatoria descrita en el teorema 3.1, entonces para  $Y = Z^{(u)}$ :

$$\begin{aligned} \text{cov}(X, Y) &= E(XY) - E(X)E(Y) \\ &= E\left(X \frac{(X - \mu)^2}{X}\right) - \mu \frac{\mu^2}{\lambda} \\ &= E((X - \mu)^2) - \mu \frac{\mu^2}{\lambda} \\ &= \text{cov}(X) - \frac{\mu^3}{\lambda} \\ &= 0 \end{aligned}$$

#### 4.1. Prueba de bondad de ajuste de razón de varianzas

Si reemplazamos  $\mu$  por su estimador  $\bar{X}_n$  y se define

$$Z_i = \frac{(X_i - \bar{X}_n)}{X_i}, \quad (4.1)$$

entonces se puede proponer un estimador de momentos de  $\text{cov}(X, Z)$ , que se le denotará como  $\tilde{\text{cov}}(X, Z)$ .  $\tilde{\text{cov}}(X, Z)$  se puede emplear para una prueba de bondad de ajuste tal que se rechaza  $H_0 : X_1, \dots, X_n \sim IG(\mu, \lambda)$  si  $\tilde{\text{cov}}(X, Z)$  está lejos de 0, pero además  $\tilde{\text{cov}}(X, Z)$  tiene una propiedad más útil que se mostrará a continuación.

$$\begin{aligned} \tilde{\text{cov}}(X, Z) &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}) (Z_i - \bar{Z}) \\ &= \frac{1}{n} \sum_{i=1}^n (X_i Z_i - \bar{Z} X_i - \bar{X} Z_i + \bar{X} \bar{Z}) \\ &= \frac{1}{n} \sum_{i=1}^n \left( X_i \frac{(X_i - \bar{X})^2}{X_i} - \bar{X} \bar{Z} \right) \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 - \frac{1}{n} \sum_{i=1}^n \bar{X} \bar{Z} \\ &= S_n^2 - \bar{X} \bar{Z} \\ &= S_n^2 - \frac{\bar{x}}{n} \sum_{i=1}^n Z_i \\ &= S_n^2 - \frac{\bar{X}}{n} \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{X_i} \\ &= S_n^2 - \frac{\bar{X}}{n} \sum_{i=1}^n \frac{(X_i^2 - \bar{X} X_i + \bar{X}^2)}{X_i} \\ &= S_n^2 - \frac{\bar{X}}{n} \sum_{i=1}^n \left( X_i - 2\bar{X} + \frac{\bar{X}^2}{X_i} \right) \\ &= S_n^2 - \frac{\bar{X}}{n} \sum_{i=1}^n \left( \frac{\bar{X}^2}{X_i} - \bar{X} \right) \\ &= S_n^2 - \frac{\bar{X}^3}{n} \sum_{i=1}^n \left( \frac{1}{X_i} - \frac{1}{\bar{X}} \right) \\ &= S_n^2 - \bar{X}^3 V_n \\ &= S_n^2 - \hat{\sigma}_X^2 \end{aligned}$$

Por lo tanto si se cumple  $H_0$ , entonces  $S_n^2 = \hat{\sigma}_X^2$ . Sea  $R_n = \frac{S_n^2}{\hat{\sigma}_X^2}$ , de acuerdo al

## 4.2. Pruebas basadas en datos transformados

---

corolario 1 de Villaseñor y González-Estrada (2015a),  $T_{1,n} = \sqrt{n/6\bar{X}_n V_n}(R_n - 1) \xrightarrow{d} N(0, 1)$  cuando  $H_0$  es cierta, gracias a esta propiedad  $T_{1,n}$  se puede emplear para probar la bondad de ajuste de los datos a la distribución IG, los valores críticos de esta prueba son  $K_{\frac{\alpha}{2}} = K_{1-\frac{\alpha}{2}} = \Phi\left(1 - \frac{\alpha}{2}\right)$ , donde  $\Phi$  es la función de distribución acumulada de la normal estándar. En este trabajo se evalúa la potencia de esta prueba usando valores críticos obtenidos a partir del método de bootstrap paramétrico.

Note que para el problema de interés se puede considerar la siguiente prueba basada en la estadística  $R_n$ , la cual se espera que este cerca de 1 bajo  $H_0$ . Rechazar  $H_0$  con un tamaño de prueba  $\alpha \in (0, 1)$  si  $R_n$  se aleja del uno, es decir, rechazar  $H_0$  si  $R_n < C_{\alpha/2}$  o  $R_n > C_{1-\alpha/2}$ , donde los valores críticos  $C_{\alpha/2}$  y  $C_{1-\alpha/2}$  son tales que la prueba es de tamaño  $\alpha$ . Estos valores críticos se aproximarán por bootstrap paramétrico.

## 4.2. Pruebas basadas en datos transformados

Se presenta en esta sección dos pruebas basadas en la transformación de la ecuación 4.1, ambas pruebas fueron propuestas en Villaseñor y González-Estrada (2015a), donde se muestra que bajo  $H_0 : X_1, \dots, X_n \sim IG(\mu, \lambda)$ ,  $Z_i = \frac{(X_i - \bar{X}_n)}{X_i}$  se distribuye asintóticamente como  $\Gamma\left(\frac{1}{2}, \frac{2\mu^2}{\lambda}\right)$ . En ese artículo se muestra que  $H_0 : X_1, \dots, X_n \sim IG(\mu, \lambda)$  es equivalente a:

$$H'_0 : Z_1, \dots, Z_n \sim \Gamma\left(\frac{1}{2}, \beta\right)$$

Los autores del artículo proponen considerar el uso de una adaptación de la prueba  $t$  y también la prueba de Anderson-Darling a los datos transformados.

### 4.2.1. Pruebas $t$

La prueba  $t$  se usa en general para probar si la media de una población es igual a un valor dado, cuando la población tiene distribución normal. En Villaseñor y González-Estrada (2015b) se propone una prueba de bondad de ajuste para la distribución Gamma con parámetro de forma  $a_0$  conocido: sea  $Z_1, \dots, Z_n$  una muestra aleatoria de la distribución  $\Gamma(a, b)$  de tamaño  $n$ , donde  $a > 0$  es el parámetro de forma y  $b > 0$  es el parámetro de escala. Sea  $\bar{Z}_n = \sum_{i=1}^n Z_i/n$ ,  $U_i = \log Z_i$ . En el mismo artículo proponen como estimador de  $a$  al estadístico  $\check{a}_n = \bar{Z}_n/\check{b}_n$ , donde  $\check{b}_n = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)(Z_i - \bar{Z}_n)$ . En el artículo se muestra que

## 4.2. Pruebas basadas en datos transformados

---

$\sqrt{n}(\check{a}_n - a) \xrightarrow{d} N(0, k^2)$  cuando  $n \rightarrow \infty$ , donde  $k^2 = a[a(1 + a\psi'(a)) - 1]$ , cuando  $\psi'$  es la función trigamma. Entonces en Villaseñor y González-Estrada (2015a) proponen una prueba parecida a las pruebas t, para probar  $H_0 : a = a_0$  con un parámetro de forma conocido  $a_0$  con el siguiente estadístico de prueba:

$$T_{2,n} = \frac{\check{a}_n - a_0}{k/\sqrt{n}}$$

Los autores eligieron  $a_0 = 1/2$ , dado que bajo  $H'_0$  el parámetro de forma tiene este valor, entonces  $k^2 = 0.5[0.5(1 + 0.5 * \psi'(0.5)) - 1] = 0.5[0.5(1 + 0.5 * 4.934801) - 1] = 0.3669$ . Cuando la muestra es grande por el teorema del límite central  $T_{2,n} \xrightarrow{d} N(0, 1)$ , por lo que los valores críticos de esta prueba para un tamaño de prueba  $\alpha$  se obtienen de la distribución  $N(0, 1)$ .

Para muestras pequeñas recomiendan obtener los valores críticos mediante simulación Monte Carlo, fijando el valor del parámetro de escala  $\beta$ , por convención sugieren el valor de  $\beta = 1$ , aprovechando la propiedad de invarianza de  $T_{2,n}$ . Cuando  $H'_0$  se rechaza, entonces  $Z_i \approx \Gamma(1/2, \beta)$ , entonces por el teorema 3.1 se tiene que  $X_i \approx IG(\mu, \lambda)$ , por lo tanto  $H_0$  se rechaza.

### 4.2.2. Prueba de Anderson-Darling

La prueba de Anderson-Darling es una prueba de bondad de ajuste que se pueden emplear para cualquier tipo de distribución continua. El estadístico de prueba es:

$$A^2 = -n - S$$

donde  $N$  es el número de datos en la muestra,  $S = \sum_{i=1}^n \frac{2i-1}{n} [\ln(Z_i) + \ln(1 - Z_i)]$  y  $Z_i = \frac{(X_i - \bar{X}_n)}{X_i}$ .

La propuesta de Villaseñor y González-Estrada (2015a) es aplicar a los datos la transformación mediante la ecuación 4.1 y luego aplicarle la prueba de Anderson-Darling a los datos transformados. Los valores críticos puede obtenerse aplicando la técnica MonteCarlo generando muestras aleatorias de la distribución  $\Gamma(\frac{1}{2}, \beta)$ , donde  $\beta = 1$  dado que la prueba Anderson-Darling es invariante ante cambio en el parámetro de escala. El estadístico  $A^2$  se denominará como estadístico  $T_{3,n}$  para cuando la hipótesis nula es que los datos provienen de una distribución Inversa Gaussiana.

## 4.3. Prueba de bondad de ajuste de razón de verosimilitud

Esta prueba fue propuesta por [Vexler \*et al.\* \(2011\)](#) et al, como su nombre lo dice está basada en una razón de verosimilitud. Los autores de esta prueba indica que tiene buena potencia estadística. Sea  $X_1, \dots, X_n$  una m.a., sea  $Y_i = (\sqrt{X_i})^{-1}$ , entonces el estadística de la prueba de [Vexler \*et al.\* \(2011\)](#) es:

$$TK_n = \frac{\min_{1 \leq m < n^\delta} \left[ \prod_{j=1}^n \frac{2m}{n(Y_{(j+m)} - Y_{(j-m)})} \right]}{\left( \frac{2\hat{\lambda}}{\pi e} \right)^{n/2}}$$

donde  $\hat{\lambda}$  es el EMV del parámetro  $\lambda$  de la distribución IG. En la tabla 1 del artículo están los valores críticos para varios tamaños de muestra y de prueba, los cuales fueron calculados mediante bootstrap paramétrico.

# Capítulo 5

## Comparación de tamaño de prueba y potencias de las pruebas

En este capítulo se muestran los resultados de tamaño de prueba y potencias de las implementaciones de bootstrap paramétrico del estadístico  $T_{1,n}$  y  $R_n$ , además de otras pruebas de hipótesis para la distribución IG para contrastar el desempeño. Para estimar el tamaño y potencia en cada caso se generaron 10000 muestras aleatorias generadas en R en cada caso. Las pruebas de bondad de ajuste analizadas son las siguientes:

1. Prueba basada en la razón de varianzas, con el estadístico  $T_1$  usando su propiedad asintótica y el método bootstrap paramétrico, también se evaluó la prueba basada en el estadístico  $R_n$ , véase sección 4.1. En el método de bootstrap paramétrico se emplearon 500 réplicas.
2. Pruebas  $t$  a los datos que se aplica transformación gamma, véase subsección 4.2.2.
3. Pruebas Anderson-Darling a los datos que se aplica transformación Gamma, véase subsección 4.2.2
4. Prueba de razón de verosimilitud empírica, véase sección 4.3.

Por caso, nos referimos a la combinación de dos factores analizados. Primero cómo se comporta cada prueba fijando un tamaño de muestra y luego eligiendo un valor del parámetro que nos interesa estudiar su evolución. Para la estimación del tamaño de prueba se analizaron como tamaños de muestras a  $n = 25, 50, 100, 300, 500$ . En cada caso se fijó un tamaño de prueba nominal de 5%, y se estudió el tamaño de prueba y su potencia usando distribuciones alternativas. Para las distribuciones alternativas se consideraron a la distribución Weibull, Gamma, Log-Normal. El software empleado fue

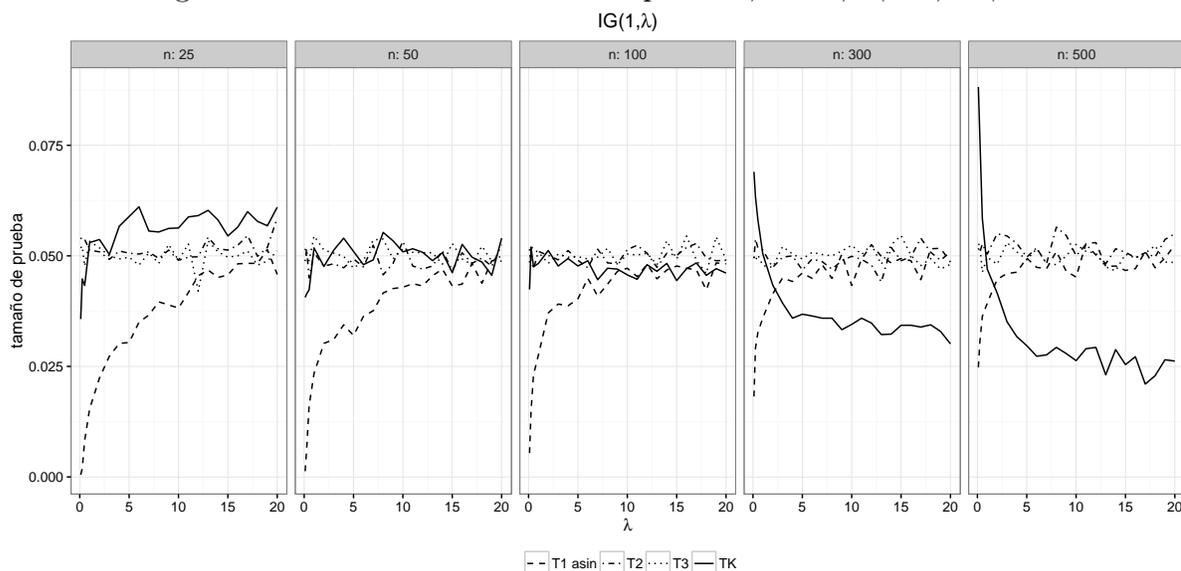
## 5.1. Reproducción del artículo Villaseñor y González-Estrada (2015a)

R Core Team (2018), se empleó el paquete `statmod` para generar números aleatorios de la distribución IG. En Giner y Smyth (2016) se muestra el funcionamiento de este paquete.

### 5.1. Reproducción del artículo Villaseñor y González-Estrada (2015a)

El siguiente gráfico muestra una reproducción de los resultados reportados por Villaseñor y González-Estrada (2015a), lo mostramos para mostrar las propiedades de la versión asintótica del estadístico  $T_{1,n}$  comparado contra los estadísticos  $T_{2,n}$ ,  $T_{3,n}$  y  $TK_n$ .

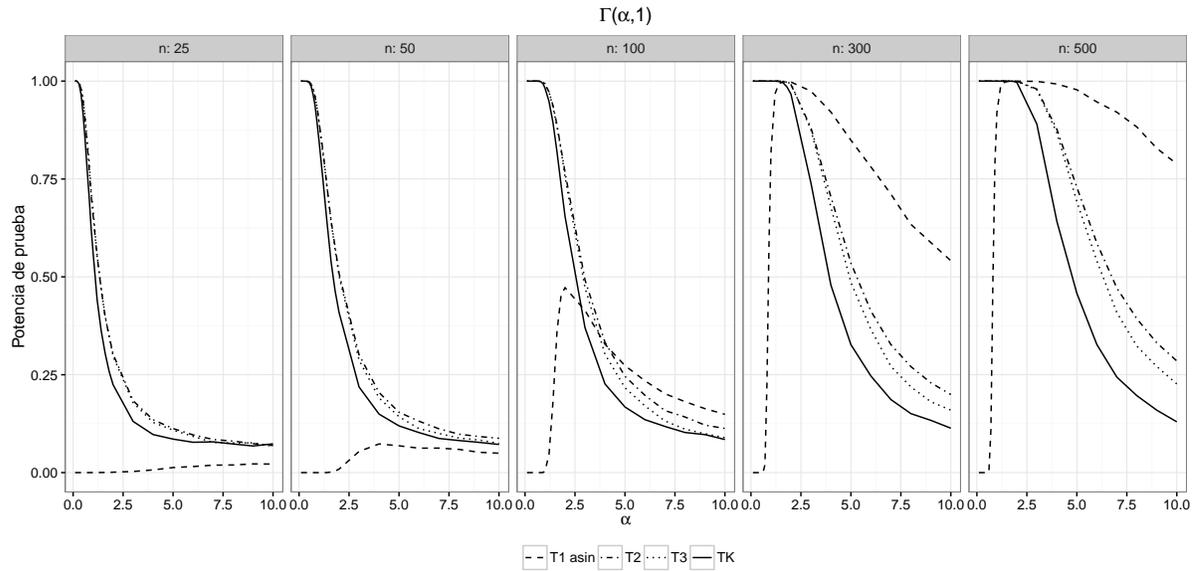
Figura 5.1: Tamaño estimado de las pruebas,  $n=25,50,100,300,500$



La prueba de razón de varianzas usando la aproximación asintótica de  $T_{1,n}$  tiene un menor tamaño de prueba a medida que  $\lambda$  disminuye, a diferencia de la prueba  $TK_n$  que tiene un comportamiento inverso, las pruebas  $T_{2,n}$  y  $T_{3,n}$  tienen un tamaño de prueba estimado aproximado al tamaño de prueba nominal.

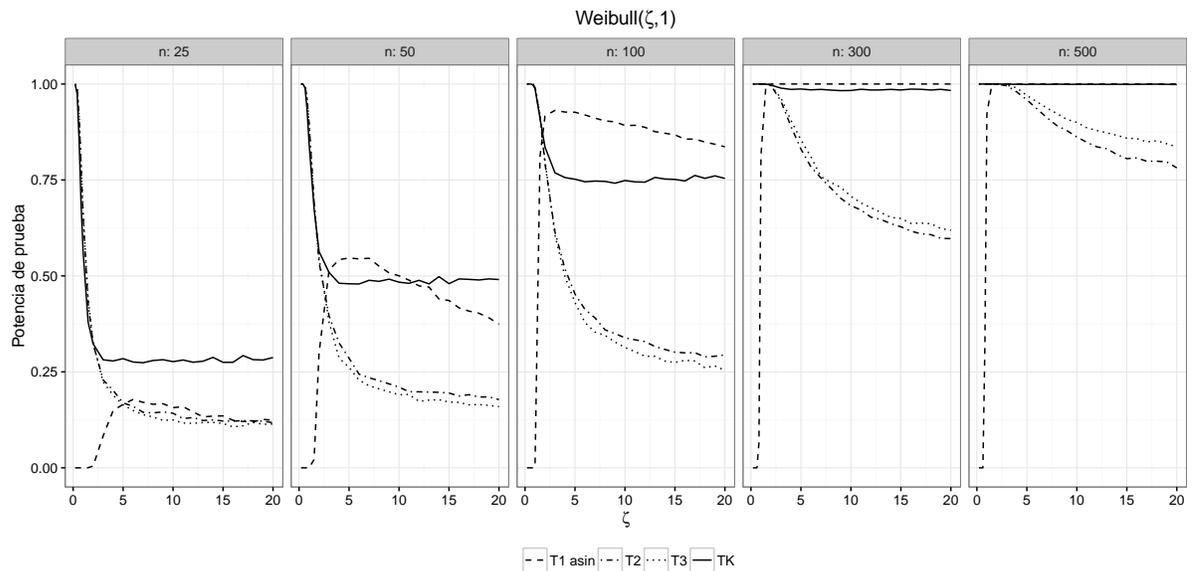
## 5.1. Reproducción del artículo Villaseñor y González-Estrada (2015a)

Figura 5.2: Potencia estimada de las pruebas vs la distribución  $\Gamma(\alpha, 1)$ ,  $n=25, 50, 100, 300, 500$



La prueba  $T_{1,n}$  tiene baja potencia en muestras pequeñas.

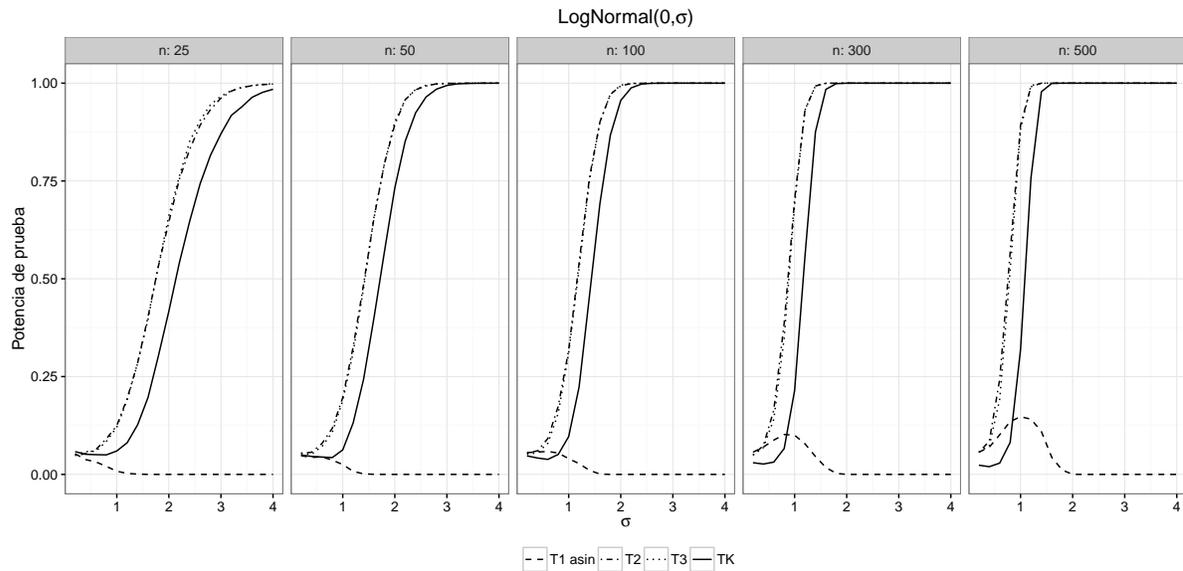
Figura 5.3: Potencia estimada de las pruebas vs la distribución  $Weibull(\zeta, 1)$ ,  $n=25, 50, 100, 300, 500$



La prueba  $T_{1,n}$  tiene baja potencia en muestra pequeñas.

## 5.1. Reproducción del artículo Villaseñor y González-Estrada (2015a)

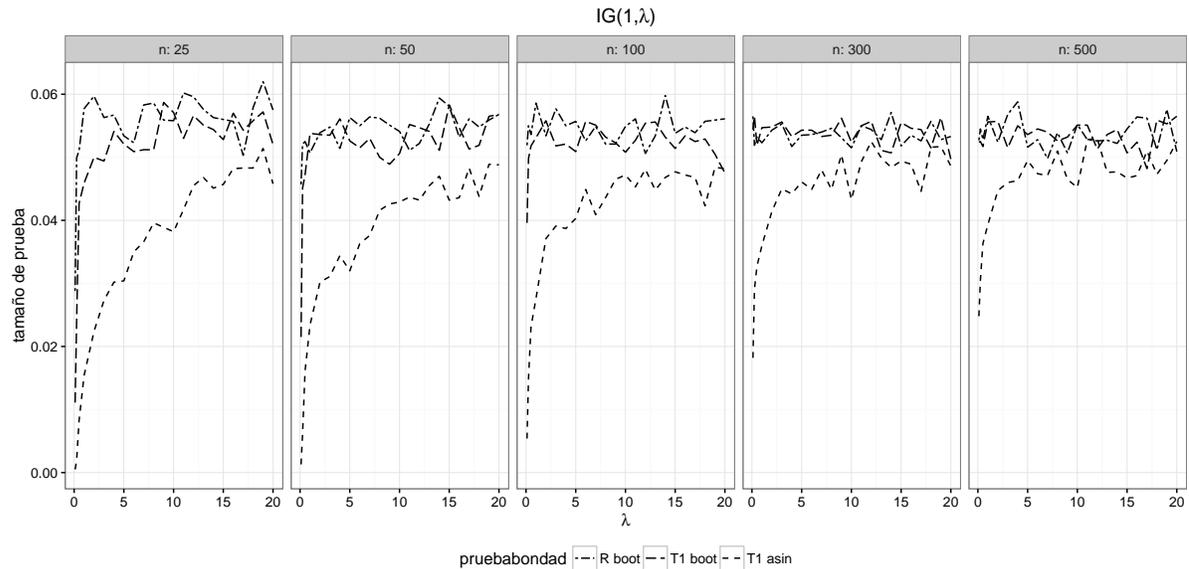
Figura 5.4: Potencia estimada de las pruebas vs la distribución  $LogNormal(0, \sigma)$ ,  $n=25, 50, 100, 300, 500$



Cuando la distribución alternativa es LogNormal, la versión asintótica de la prueba de razón de varianzas tiene baja potencia, lo cual implica que no hay diferencia entre una distribución LogNormal y una distribución Inversa Gaussiana.

## 5.2. Desempeño del bootstrap paramétrico vs versión asintótica

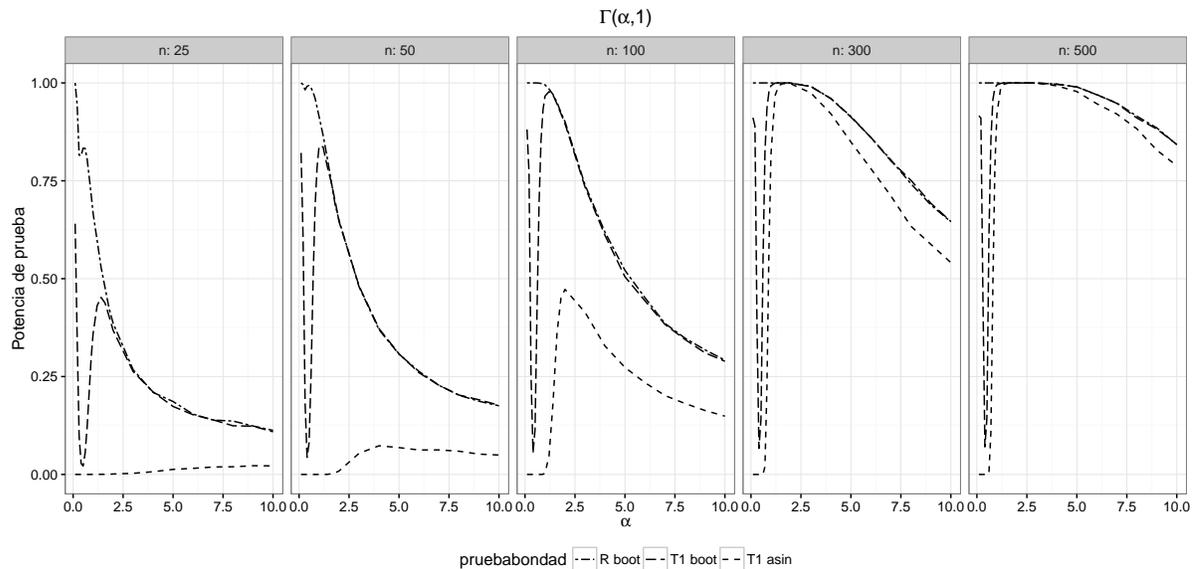
Figura 5.5: Tamaño estimado de las pruebas



La prueba basada en el estadístico  $R_n$  es la que tiene un tamaño de prueba más aproximado al tamaño de prueba nominal de 5%.

## 5.2. Desempeño del bootstrap paramétrico vs versión asintótica

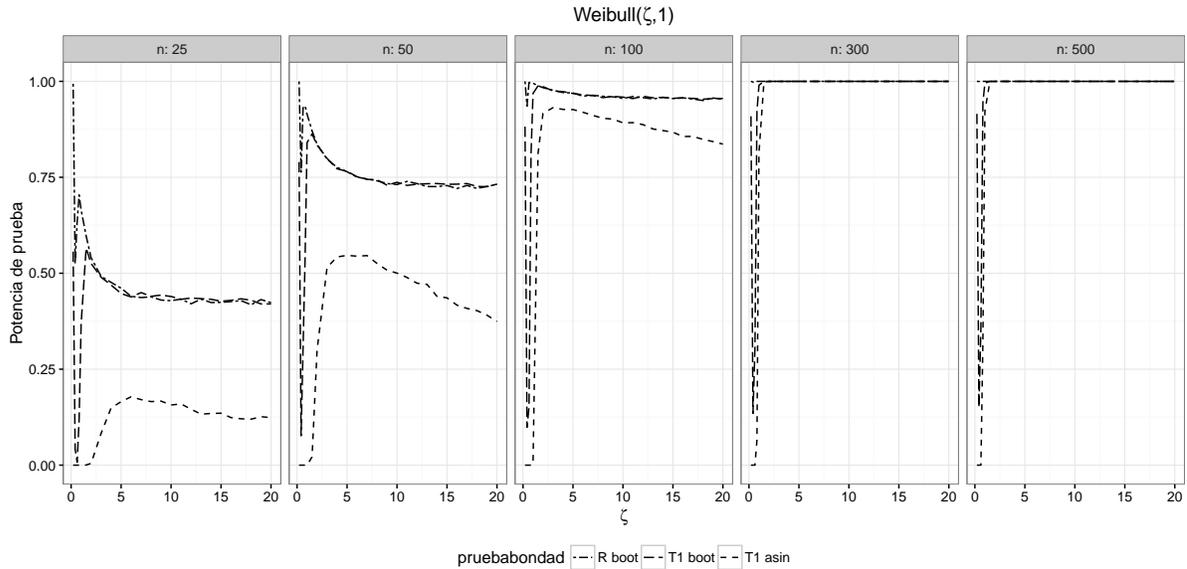
Figura 5.6: Potencia estimada de las pruebas vs la distribución  $\Gamma(\alpha, 1)$ ,  $n=25, 50, 100, 300, 500$



La prueba  $R_n$  es la que tiene mejor potencia cuando la distribución alternativa es una Gamma. El comportamiento de la prueba  $T_{1,n}$  con la implementación del bootstrap paramétrico tiene un comportamiento inconsistente ya que baja su potencia de prueba conforme el parámetro de forma es cercano o igual 1.

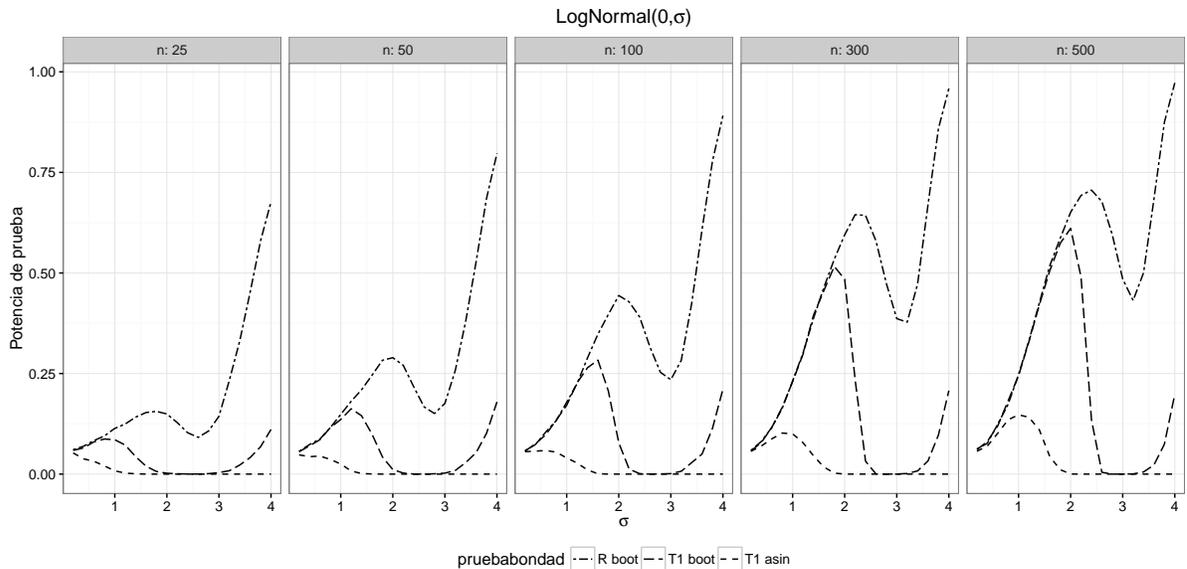
## 5.2. Desempeño del bootstrap paramétrico vs versión asintótica

Figura 5.7: Potencia estimada de las pruebas vs la distribución  $Weibull(\zeta, 1)$ ,  $n=25, 50, 100, 300, 500$



La prueba  $R_n$  es la que tiene mejor potencia cuando la distribución alternativa es Weibull. El comportamiento del estadístico  $T_{1,n}$  con la implementación del bootstrap paramétrico tiene un comportamiento inconsistente ya que baja su potencia de prueba conforme el parámetro de forma es cercano o igual 1.

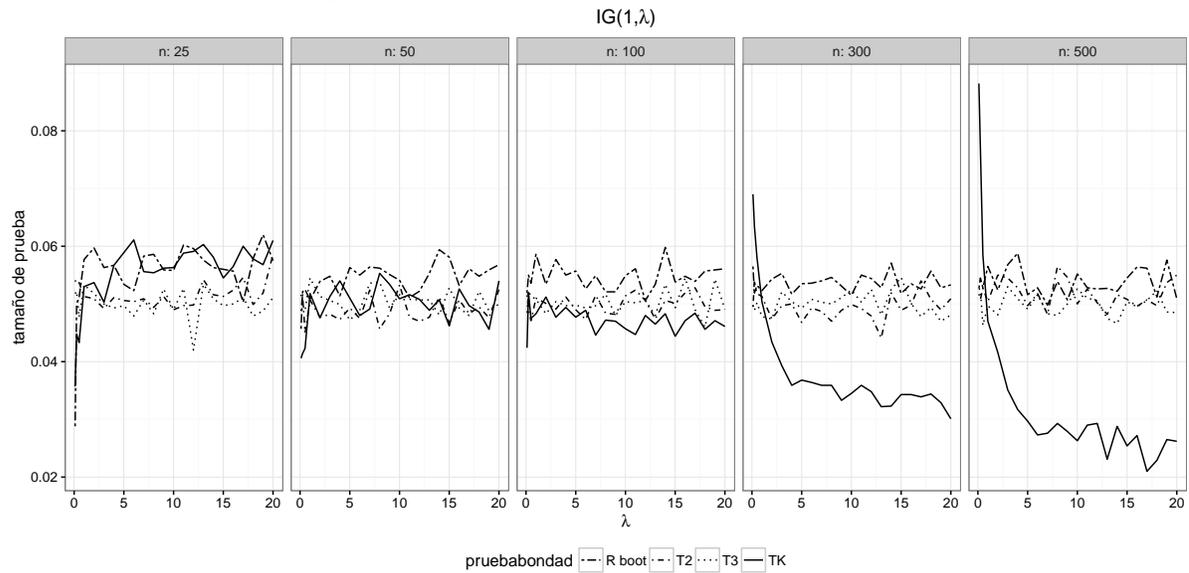
Figura 5.8: Potencia estimada de las pruebas vs la distribución  $LogNormal(0, \sigma)$ ,  $n=25, 50, 100, 300, 500$



La prueba  $R_n$  es la que tiene mejor potencia cuando la distribución alternativa es una Log-Normal. Curiosamente la potencia de cada prueba tiene un máximo local.

## 5.3. Desempeño bootstrap paramétrico vs otras pruebas

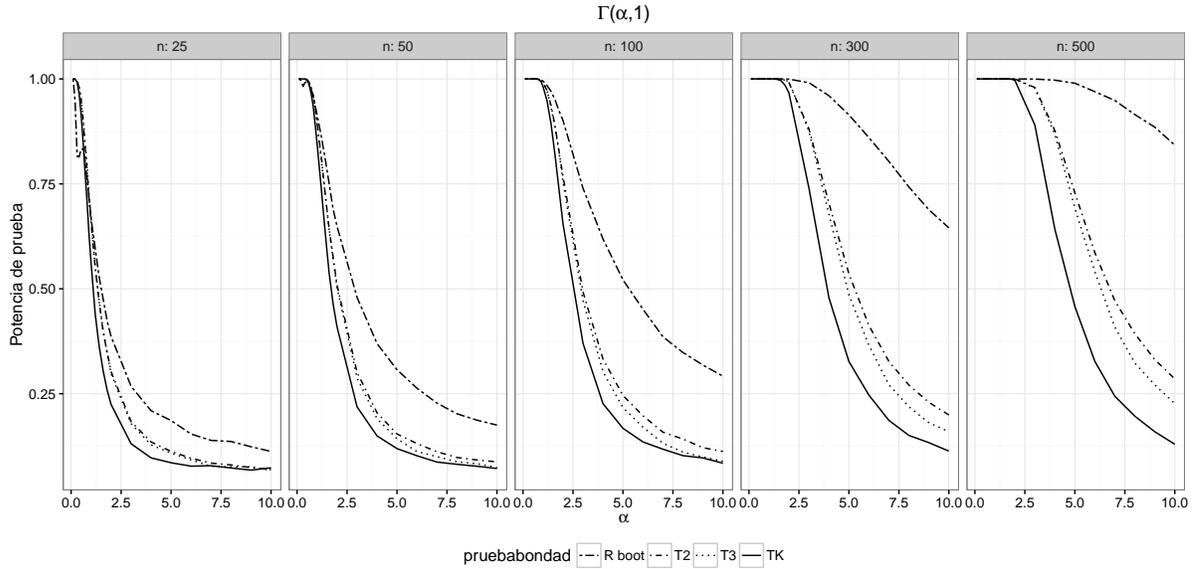
Figura 5.9: Tamaño estimado de las pruebas



La prueba  $R_n$  es la que tiene un tamaño de prueba más aproximado al tamaño de prueba nominal de 5%.

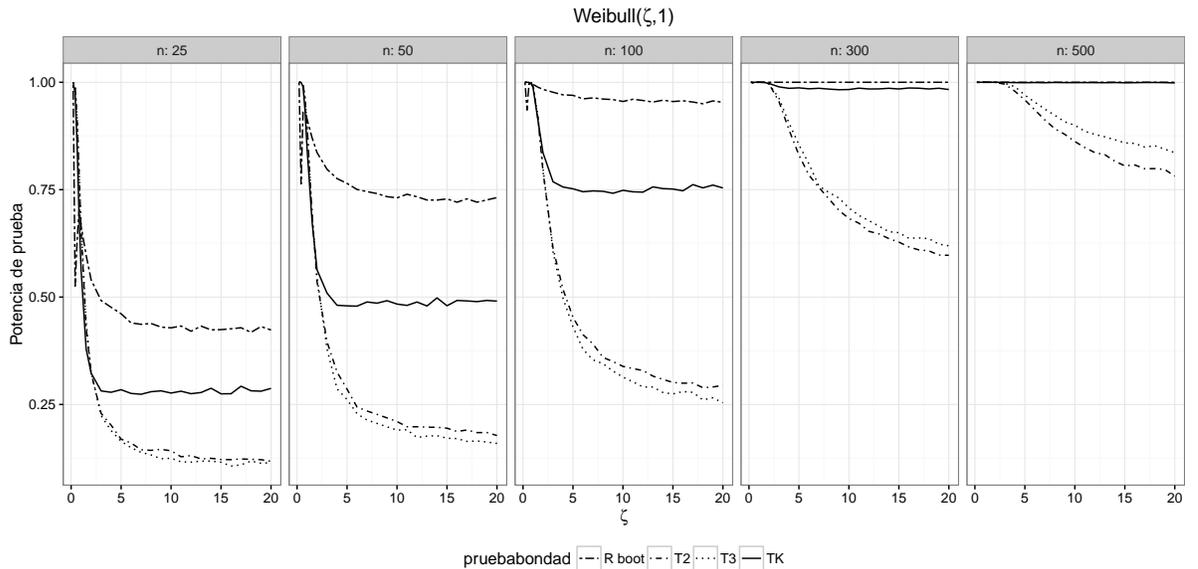
### 5.3. Desempeño bootstrap paramétrico vs otras pruebas

Figura 5.10: Potencia estimada de las pruebas vs la distribución  $\Gamma(\alpha, 1)$ ,  $n=25, 50, 100, 300, 500$



La prueba  $R_n$  es la que tiene mejor potencia cuando la distribución alternativa es una Gamma.

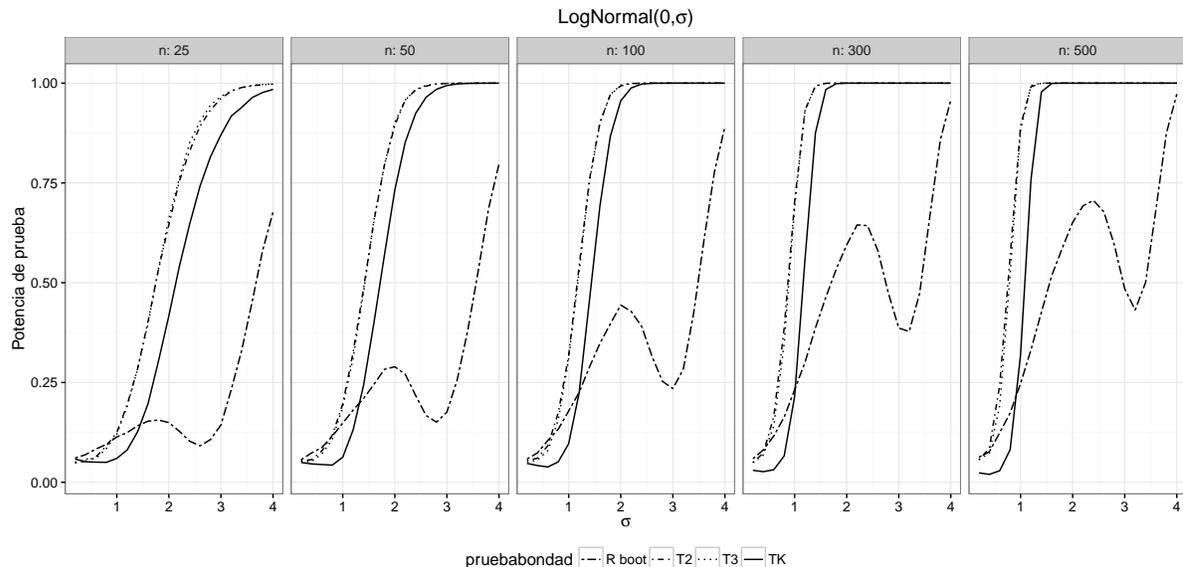
Figura 5.11: Potencia estimada de las pruebas vs la distribución Weibull  $(\zeta, 1)$ ,  $n=25, 50, 100, 300, 500$



La prueba  $R_n$  es la que tiene mejor potencia cuando la distribución alternativa es una Weibull.

## 5.4. Discusión

Figura 5.12: Potencia estimada de las pruebas vs la distribución  $LogNormal(0, \sigma)$ ,  $n=25, 50, 100, 300, 500$



Cuando la distribución alternativa es una Log-Normal, la prueba basada en el estadístico  $R_n$  es la que tiene una menor potencia.

## 5.4. Discusión

En la sección 5.1 se observa que la prueba  $T_{1,n}$  utilizando su distribución asintótica tiene poca potencia en muestras pequeñas. Por esta razón se analizó cuánto mejora su desempeño usando el método de bootstrap paramétrico. En la sección 5.2 se observa que la potencia del estadístico  $T_{1,n}$  mejora, pero tiene un comportamiento poco consistente, por esta razón se contempló la posibilidad de ver qué sucedía si empleamos el estadístico  $R_n$ , el cual tenía un tamaño de prueba cercano al tamaño de prueba nominal y una mejor potencia. Finalmente se comparó el desempeño de la prueba  $R_n$  con su implementación del bootstrap paramétrico contra las demás pruebas del artículo Villaseñor y González-Estrada (2015a), con el objetivo de ver si esta prueba tiene mejores propiedades de potencia. Se observó que tiene una mejor potencia cuando las muestras son provenientes de distribuciones Weibull o Gamma; sin embargo, tiene baja potencia cuando la muestra es proveniente de la distribución Log-Normal.

# Capítulo 6

## Ejemplo

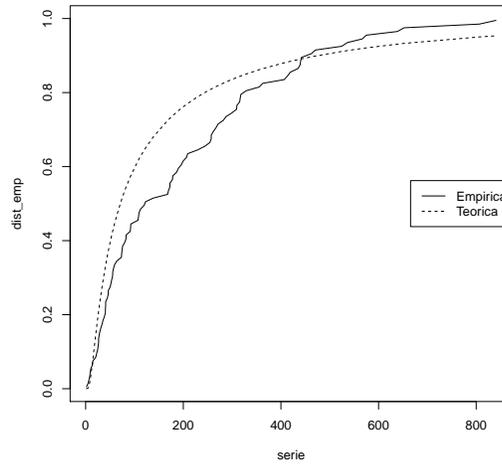
El presente ejemplo se escogió gracias a la lectura de [Sukrutha y Desai \(2018\)](#), en este artículo se intenta encontrar la distribuciones que mejor caracterizan las lluvias mensuales en 20 ciudades de India, resultando ser la distribución Inversa Gaussiana una de las distribuciones que más se seleccionan para modelar la precipitación. El interés surgió ya que México es un país que 75 % de su superficie de cultivo es de temporal, y decidimos estudiar este fenómeno en escala local.

Los datos que se seleccionaron fueron obtenidos de la base de datos Clicom perteneciente al Sistema Meteorológico Nacional, y corresponden a las precipitaciones en milímetros de la base meteorológica de la localidad de Huautla de Jimenez, Oaxaca desde marzo de 1993 hasta diciembre de 2005. En ese periodo hubo datos faltantes y a su vez se tuvieron que descartar los meses con nula precipitación debido a que la distribución Inversa Gaussiana no incluye en su soporte al valor 0. En total se tuvieron 100 datos para aplicar la prueba de bondad de ajuste, los cuales se presentan en la [Tabla 6.1](#)

Tabla 6.1: Precipitaciones mensuales de Huautla de Jimenez

Periodo	PP	Periodo	PP	Periodo	PP	Periodo	PP
1993-3	27.7	1997-4	63.9	2001-8	435.2	2003-12	50.3
1993-4	35.7	1997-5	82.8	2001-9	406.7	2004-1	53.5
1993-5	55.5	1997-7	254.4	2001-10	566.8	2004-2	14.1
1993-6	806.8	1997-8	229.3	2001-11	59.2	2004-3	82.7
1993-7	470.8	1997-9	186.5	2002-1	14.8	2004-4	40.4
1993-8	574.9	1997-10	271.2	2002-2	44.5	2004-5	92.1
1993-9	282	1997-11	178.5	2002-3	24.6	2004-6	206.5
1993-10	172.5	2000-1	74.6	2002-4	7	2004-7	177.8
1993-11	169.6	2000-2	73.2	2002-5	119.7	2004-8	257.2
1993-12	29.3	2000-3	55.3	2002-6	309.3	2004-9	138.6
1995-1	33.7	2000-4	51.9	2002-7	363	2004-10	172.6
1995-2	46.4	2000-5	299.4	2002-8	442.1	2004-11	112.3
1995-3	26.6	2000-6	419.1	2002-10	287.6	2004-12	41
1995-4	10.8	2000-7	316.6	2002-11	196	2005-1	57.1
1995-5	75.7	2000-8	840.5	2003-1	91.3	2005-2	31.4
1995-6	318	2000-9	329	2003-2	5.1	2005-3	1.8
1995-7	638.7	2000-10	309.7	2003-3	8.6	2005-4	25.5
1995-8	652.3	2000-11	189.7	2003-4	20.7	2005-5	41
1995-9	261.6	2001-1	46.2	2003-5	107.2	2005-6	413.6
1995-10	199.7	2001-2	73.8	2003-6	315.5	2005-7	524.5
1995-11	257.2	2001-3	26.6	2003-7	267	2005-8	440.4
1995-12	108.3	2001-4	40.5	2003-8	244.1	2005-9	536
1996-12	38.5	2001-5	80	2003-9	209.1	2005-10	439.4
1997-1	9.4	2001-6	462.3	2003-10	167.9	2005-11	122.6
1997-3	92.5	2001-7	355.3	2003-11	109.3	2005-12	22.5

Figura 6.1: Distribución IG ajustada y Función de distribución empírica de los datos de precipitación



La distribución teórica se obtuvo asumiendo que los datos provienen de una distribución IG y los parámetros de esta se estimaron mediante máxima verosimilitud. La figura 6.1 muestra la distribución empírica de los datos de precipitación y la distribución IG ajustada a los mismos. Los parámetros se estimaron por máxima verosimilitud y se obtuvo  $\hat{\mu} = 196.21700$  y  $\hat{\lambda} = 51.85884$

En la Figura 6.1 se observa que los datos difícilmente parece que provienen de una distribución IG. Para comprobarlo se aplicaron las pruebas de bondad de ajuste del capítulo 4 con un nivel de prueba de 5%. Los resultados son los siguientes:

Tabla 6.2: Aplicación de las pruebas del capítulo 4 a las precipitaciones mensuales de Huautla de Jiménez

Prueba	Constantes críticas			Valor estadística	Veredicto
	$K_{\alpha/2}$	$K_{1-\alpha/2}$	$K_{\alpha}$		
$T_{1,n}$ asintótica	-1.96	1.96		-1.584	No se rechaza $H_0$
$T_{2,n}$	-1.65	2.38		-3.53	Se rechaza $H_0$
$T_{3,n}$			1.49	6.11	Se rechaza $H_0$
$TK_n$			12.38	24.48	Se rechaza $H_0$
$T_{1,n}$ Bootstrap paramétrico	-1.02	1.48		-1.584	Se rechaza $H_0$
$R_n$ Bootstrap paramétrico	0.51	1.76		0.24	Se rechaza $H_0$

La prueba  $T_{1,n}$  usando su versión asintótica nos indica que los datos de precipitación si provienen de una distribución Inversa Gaussiana, en cambio las demás pruebas nos indican que esta hipótesis debe ser rechazada. Por otra parte, los resultados de la implementación de bootstrap paramétrico de  $T_{1,n}$  y  $R_n$  dan resultados similares a las

---

demás pruebas. En este caso caso por lo tanto la implementación del bootstrap paramétrico ha sido satisfactoria.

# Capítulo 7

## Conclusiones

En el ejemplo del capítulo 6 se observa que debido a la baja potencia de la prueba de bondad de ajuste del estadístico  $T_{1,n}$  usando su propiedad asintótica no rechaza  $H_0$ , y las demás pruebas si, sin embargo empleando el método de bootstrap paramétrico para aproximar los valores críticos de la prueba de razón rechaza  $H_0$ , tanto aplicando al estadístico  $T_{1,n}$  como al estadístico  $R_n$ . En la comparación de pruebas de bondad de ajuste del capítulo 5 se observa que hay una gran mejoría empleando el método de bootstrap paramétrico para el estadístico  $T_{1,n}$ , pero su forma más básica el estadístico  $R_n$  usando la misma implementación tiene un mejor desempeño, además tiene una potencia muy buena cuando las muestras provienen de las distribuciones Gamma y Weibull, pero una potencia mala cuando las muestras provienen de una distribución Log Normal. Por su buena potencia para cuando las distribuciones son provenientes de una Gamma o una Weibull, la prueba de bondad de ajuste con el estadístico  $R_n$  empleando el método de bootstrap paramétrico para aproximar sus valores críticos debe ser considerada como una opción cuando se desea probar que una muestra aleatoria proviene de una población con distribución Inversa Gaussiana.

# Bibliografía

- Banerjee, A. K. (1986). *A Bivariate Inverse Gaussian Distribution*. University of Wisconsin, Department of Statistics.
- Bitbol, M. y Darrigol, O. (1992). Erwin Schroedinger, Philosophy and the birth of quantum mechanics.
- Canal, S. Y. (2000). La estadística una ciencia del siglo XX. RA Fisher, el genio. *Revista Colombiana de Estadística*, 23, 2, 1–14.
- Chhikara, R. (1988). *The Inverse Gaussian Distribution: Theory: Methodology, and Applications*, tomo 95. CRC Press.
- Chhikara, R. y Folks, J. (1977). The inverse Gaussian distribution as a lifetime model. *Technometrics*, 19, 4, 461–468.
- Chhikara, R. S. (1972). *Statistical inference related to the inverse Gaussian distribution*. Tesis Doctoral, Oklahoma State University.
- D’Agostino, R. B. y Stephens, M. A. (1986). Goodness-of-fit techniques. *Statistics: Textbooks and Monographs, New York: Dekker, 1986, edited by D’Agostino, Ralph B.; Stephens, Michael A.*
- Deheuvels, P. y Steinebach, J. (1990). On the sample path behavior of the first passage time process of a brownian motion with drift. En *Annales de l’IHP Probabilités et statistiques*, tomo 26, 145–179. Gauthier-Villars.
- Efron, B. *et al.* (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7, 1, 1–26.
- Folks, J. L. y Chhikara, R. S. (1978). The inverse Gaussian distribution and its statistical application—a review. *Journal of the Royal Statistical Society. Series B (Methodological)*, 263–289.

## Bibliografía

---

- Galton, F. (1899). A geometric determination of the median value of a system of normal variants, from two of its centiles. *Nature*, 61, 1570, 102.
- Giner, G. y Smyth, G. K. (2016). statmod: Probability Calculations for the Inverse Gaussian Distribution. *The R Journal*, 8, 1, 339–351.
- Hald, A. (2008). *A history of parametric statistical inference from Bernoulli to Fisher, 1713-1935*. Springer Science & Business Media.
- Hand, D. J. (2015). From evidence to understanding: a commentary on Fisher (1922)'On the mathematical foundations of theoretical statistics'. *Phil. Trans. R. Soc. A*, 373, 2039, 20140252.
- Lenhard, J. (2006). Models and statistical inference: The controversy between Fisher and Neyman–Pearson. *The British journal for the philosophy of science*, 57, 1, 69–91.
- Natarajan, R. y Mudholkar, G. S. (2004). Moment-based goodness-of-fit tests for the inverse Gaussian distribution. *Technometrics*, 46, 3, 339–347.
- Ochoa, A. (2015). *Pruebas para la distribución Gaussiana Inversa basadas en transformaciones*. Proyecto Fin de Carrera, Colegio de Postgraduados.
- O'Reilly, F. J. y Rueda, R. (1992). Goodness of fit for the inverse Gaussian distribution. *Canadian Journal of Statistics*, 20, 4, 387–397.
- Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50, 302, 157–175.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- RA Fisher, M. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc. Lond. A*, 222, 594-604, 309–368.
- Schrödinger, E. (1915). Zur theorie der fall-und steigversuche an teilchen mit brownscher bewegung. *Physikalische Zeitschrift*, 16, 289–295.
- Seshadri, V. (2012). *The inverse Gaussian distribution: statistical theory and applications*, tomo 137. Springer Science & Business Media.

## Bibliografía

---

- Sukrutha, D. S. R., Anumandla y Desai, S. (2018). Multimodel response assessment for monthly rainfall distribution in some selected Indian cities using best-fit probability as a tool. *Applied Water Science*, 8, 5, 145. ISSN 2190-5495.
- Tweedie, M. (1984). An index which distinguishes between some important exponential families. En *Statistics: Applications and new directions: Proc. Indian statistical institute golden Jubilee International conference*, tomo 579, 604.
- Tweedie, M. C. (1945). Inverse statistical variates. *Nature*, 155, 3937, 453.
- Tweedie, M. C. (1957). Statistical Properties of Inverse Gaussian Distributions. I. *The Annals of Mathematical Statistics*, 362–377.
- Vexler, A., Shan, G., Kim, S., Tsai, W.-M., Tian, L. y Hutson, A. D. (2011). An empirical likelihood ratio based goodness-of-fit test for Inverse Gaussian distributions. *Journal of Statistical Planning and Inference*, 141, 6, 2128–2140.
- Villaseñor, J. A. y González-Estrada, E. (2015a). Tests of fit for Inverse Gaussian distributions. *Statistics & Probability Letters*, 105, 189–194.
- Villaseñor, J. A. y González-Estrada, E. (2015b). A variance ratio test of fit for Gamma distributions. *Statistics & Probability Letters*, 96, 281–286.
- Vladimirescu, I. y Tunaru, R. (2003). Estimation functions and uniformly most powerful tests for inverse Gaussian distribution. *Comment. Math. Univ. Carolin.*, 44, 1, 153–164.
- Wainer, H. y Thissen, D. (1981). Graphical data analysis. *ETS Research Report Series*, 1981, 1, 191–241.
- Wald, A. (1944). On Cumulative Sums of Random Variables. *Ann. Math. Statist.*, 15, 3, 283–296.

# Anexos

## Anexo A: Aspecto computacionales

Funciones para estimar parámetros de IG mediante máxima verosimilitud

Listing .1: estimacionparametros.R

---

```
1 # Funcion para estimar parametros de maxima verosimilitud
2 # de la Inversa Gaussiana
3 calculo_ML_invgauss <- function(x) {
4   n <- length(x)
5   mu <- mean(x)
6   vn <- sum(1/x-1/mu)/n
7   if(vn < 0) stop("Varianza negativa")
8   lambda <- 1/vn
9   return(c("mu"=mu,"lambda"=lambda))
10 }
```

---

Esta parte nos sirve para obtener los valores críticos de varias pruebas.

Listing .2: obtencionvalorescriticos.R

---

```
1 source("estadisticadeprueba.R")
2 source("pruebasbondadajuste.R")
3
4 ### Calcular MC fde AD
5 ### Anderson Darling
6 calculo_ad_value <- function(x, scale) {
7   x <- sort(x)
8   Ti <- pgamma(x,shape=1/2, scale=scale)
9   Ti_inv <- sort(Ti, decreasing=T)
```

```
10  n <- length(x)
11  i <- 1:n
12  S <- (2*i - 1) * (log(Ti) + log(1 - Ti_inv))
13  - n - sum(S)/n
14  }
15
16  obtener_fde_AD <- function(n,B=10000) {
17    resultados <- rep(NA, B)
18    for(i in 1:B) {
19      muestra <- rgamma(n, shape=1/2, scale=1)
20      resultados[i] <- calculo_ad_value(muestra, 1)
21    }
22    resultados <- sort(resultados)
23    return(resultados)
24  }
25
26  ## Estimar el valor critico del estadistico T3, ver la publicacion
27  ## Tests of fit for Inverse Gaussian distributions
28  ### Jose A. Villasenor, Elizabeth Gonzalez-Estrada
29  obtener_fde_T3 <- function(n, alpha=0.05) {
30    R <- 1000000
31    resultados <- rep(NA, R)
32    for(i in 1:R) {
33      x_star <- rgamma(n, 1/2, 1)
34      n <- length(x_star)
35      b. <- 2*mean(x_star)
36      s <- sort(x_star)/b.
37      theop <- pgamma(s,.5)
38      # Anderson-Darling statistic
39      ad <- -n-sum((2*(1:n)-1)*log(theop) + (2*n+1-2*(1:n))*log(1-theop
40      ))/n
41      resultados[i] <- ad
42    }
43    resultados <- sort(resultados)
44  }
```

```
45 ## Calculo T2n
46 Calculo de T2n
47 calculo_T2n <- function(x) {
48   x_mean <- mean(x)
49   y <- (x-x_mean)^2/x
50   y_mean <- mean(y)
51   u <- log(y)
52   u_mean <- mean(u)
53   n <- length(x)
54   bn <- sum((y-y_mean)*(u-u_mean))/(n-1)
55   an <- y_mean /bn
56   a0 <- 1/2
57   k <- sqrt(a0 * (a0 * (1 + a0 * trigamma(a0)) - 1))
58   (an - a0) / (k / sqrt(n))
59 }
60
61 set.seed(0)
62 fde_T2_300 <- obtener_fde_MC(c("mu"=1,"lambda"=1), R=1000000, M=300,
63   calculo_T2n)
64 cuantiles_T2_300 <- quantile(fde_T2_300, c(0.025,0.975))
65
66 ## Estimar el valor critico del estadistico TKN, ver la publicacion
67 ## An empirical likelihood ratio based goodness-of-fit test
68 ## for Inverse Gaussian distributions
69 ## Los valores para los tamanos de muestras 25,50,100, 200, 300 se
70 ## obtuvo de esta publicacion, esta funcion se diseno para estimar para
71 ## tamano de muestra 500
72 set.seed(0)
73 fde_TKN_500 <- obtener_fde_MC(c("mu"=1,"lambda"=1), R=1000000, M=500,
74   calculo_TKN)
75 cuantil_TKN_500_0.95 <- quantile(fde_TKN_500, 0.95)
76
77 set.seed(0); fde_T3_100 <- obtener_fde_T3(100)
78 cuantil_T3_100_0.95 <- quantile(fde_T3_100, 0.95)
79 set.seed(0); fde_T3_300 <- obtener_fde_T3(300)
```

```
79 cuantil_T3_300_0.95 <- quantile(fde_T3_300, 0.95)
80 set.seed(0); fde_T3_500 <- obtener_fde_T3(500)
81 cuantil_T3_500_0.95 <- quantile(fde_T3_500, 0.95)
```

---

La implementación de prueba de bondad con bootstrap paramétrico para aplicarlo en pruebas de dos colas, se devuelve un 0 si no se rechazó  $H_0$  y 1 si se rechaza.

Listing .3: pruebasbondadajuste.R

---

```
1 library("statmod")
2 source("estimacionparametros.R")
3 source("estadisticadeprueba.R")
4
5 # Bootstrap parametrico MC
6 obtener_fde_MC <- function(parametros, R=100, M=100, fn) {
7   resultados <- rep(NA, R)
8   for(i in 1:R) {
9     x_star <- rinvgauss(M, parametros["mu"], parametros["lambda"])
10    resultados[i] <- fn(x_star)
11  }
12  resultados <- sort(resultados)
13  return(resultados)
14 }
15
16 ### La prueba de bondad de ajuste para realizar bootstrap parametrico
17 pruebabondad_bootMC <- function(x, argumentos, metodoiter, metodo_est) {
18   par_est <- metodoiter(x)
19   valor_est <- metodo_est(x)
20   n <- length(x)
21   fde_estimado <- obtener_fde_MC(par_est, R=argumentos["R"], M=n,
22     metodo_est)
23   alpha <- argumentos["alpha"]
24   cuantil1 <- quantile(fde_estimado, alpha/2)
25   cuantil2 <- quantile(fde_estimado, 1-alpha/2)
26   ifelse((cuantil1 > valor_est) || (cuantil2 < valor_est), 1, 0)
27 }
```

---

Este archivo contiene un código que sirve para ir registrando en una base de datos sqlite el número de veces que se genero una muestra aleatoria con una determinada

distribución y parámetros, y el número de veces que se rechazaron esas muestras como proveniente de la distribución IG aplicando una prueba de bondad ajuste que se desea analizar.

Listing .4: medidortasarechazo.R

---

```
1 library("RSQLite")
2
3 # Se analiza para cada distribucion especifica el numero de rechazo
4 # C: Numero de veces que se realiza el metodo de bootsrap para obtener el
   numero de
5 # de rechazos para medir el tamano de prueba.
6 # N: El tamano de la muestra que se analiza la potencia
7 # R: El numero de replicas que se usa para el metodo de bootstrap
8 # M: El tamano de la muestra remuestrada por bootstrap
9 # lambda: el valor de lambda analizado
10 ejecutarprograma <- function(iter,distribucion,pruebabondad,argumentos,
   archivosdatos,nrechazos,i1) {
11 #   print(distribucion)
12 #   print(argumentos)
13 # Calculate the number of cores
14 no_cores <- detectCores()
15 # Initiate cluster
16 cl <- makeCluster(no_cores)
17
18 parametrosdistribucion <- distribucion$parametros
19 distribucion$parametros[["n"]] <- argumentos["n"]
20
21 llaves <- c(argumentos, parametrosdistribucion)
22 names(llaves) <- c(names(argumentos), names(parametrosdistribucion))
23
24 tipollaves <- c("id"="INTEGER PRIMARY KEY", "pruebabondad"="TEXT",
25   "distribucion"="TEXT", "n"="INTEGER", "alpha"="REAL", "mean"="REAL
   ", "shape"="REAL", "R" ="INTEGER", "M"="INTEGER",
26   "scale"= "REAL", "meanlog"="REAL", "sdlog"="REAL", "int"="INTEGER
   DEFAULT 0", "rechazos"="INTEGER DEFAULT 0")
27
28 ### Checar existencia tabla sqlite
```

```
29  instruccionesql <- paste(names(tipollaves), tipollaves, sep=" ",
30  collapse=", ")
31  instruccionesql <- paste("create table if not exists pruebas (",
32  instruccionesql, ")")
33  db <- dbConnect(SQLite(), dbname=archivosdatos)
34  dbExecute(db, instruccionesql)
35  dbDisconnect(db)
36
37  instruccionesql <- paste("SELECT * FROM pruebas WHERE ", sep="")
38  llavesdata <- llaves
39  llavesdata$distribucion <- distribucion$name
40  llavesdata$pruebabondad <- pruebabondad
41  # print(llavesdata)
42  llavesdata <- data.frame(llavesdata)
43  llaves <- paste(paste(names(llaves), llaves, sep="="), collapse=" AND
44  ")
45  instruccionesql <- paste(instruccionsql, llaves)
46  instruccionesql <- paste(instruccionsql, " AND pruebabondad=\"",
47  pruebabondad, "\",sep="")
48  instruccionesql <- paste(instruccionsql, " AND distribucion=\"",
49  distribucion$name, "\",sep="")
50
51  db <- dbConnect(SQLite(), dbname=archivosdatos)
52  pruebadb <- dbGetQuery(db, instruccionesql)
53  dbDisconnect(db)
54  if(nrow(pruebadb) == 0) {
55    print(llavesdata)
56    db <- dbConnect(SQLite(), dbname=archivosdatos)
57    dbWriteTable(db, "pruebas", llavesdata, append= TRUE)
58    dbDisconnect(db)
59    db <- dbConnect(SQLite(), dbname=archivosdatos)
60    pruebadb <- dbGetQuery(db, instruccionesql)
61    dbDisconnect(db)
62  }
63  intentos <- pruebadb[1,"int"]
64  rechazos <- pruebadb[1,"rechazos"]
```

```
60
61 while(intentos < iter) {
62     serechaza <- foreach(i2=1:iteracionesporciclo, .combine="c") %
63     dopar% {
64         mual <- do.call(distribucion$name, distribucion$parametros)
65         do.call(pruebabondad,list(x=mual, argumentos=argumentos))
66     }
67     intentos <- intentos + iteracionesporciclo
68     rechazos <- rechazos + sum(serechaza)
69     instruccionesql <- paste("UPDATE pruebas SET int=", intentos, ",
70     rechazos=", rechazos, " where id=", pruebadb[1,"id"])
71     db <- dbConnect(SQLite(), dbname=archivosdatos)
72     dbExecute(db, instruccionesql)
73     dbDisconnect(db)
74 }
75 print(rechazos/intentos)
76 stopCluster(cl)
77 return(rechazos)
78 }
79
80 ejecutarprograma2 <- function(iter, distribucion ,parfijos, parvars,
81     pruebabondad, argumentos,archivosdatos) {
82     print(distribucion)
83     resultados <- parvars
84     resultados$resultados <- NA
85     for(i in 1:nrow(parvars)) {
86         para <- parvars[i,]
87
88         distribucion[["parametros"]] <- list(parfijos, parvars[i,])
89         names(distribucion[["parametros"]]) <- c(names(parfijos), colnames(
90     parvars))
91     nrechazo <- ejecutarprograma(iter,distribucion,pruebabondad=
92     pruebabondad,argumentos=argumentos, archivosdatos=archivosdatos)
93     resultados[i,"resultados"] <- nrechazo
94 }
95 }
```

```
91   }
92   resultados$resultados <- resultados$resultados/iter
93 }
```

---

Programa para medir la tasa de rechazo para cada distribución y los parámetros que nos interesa analizar su evolución. El programa invoca el archivo `medidortasarechazo.R` cada vez para ir registrando los datos.

Listing .5: `tasarechazo.R`

---

```
1 options(warn=2)
2
3 library(parallel)
4 library(foreach)
5 options(warn=1)
6 library(doMC)
7 registerDoMC(cores=4)
8
9 # Libreria que incluye la funcion generadores de numeros aleatorios de la
10 # distribucion inversa gaussiana
11 library(statmod)
12 source("codigo-autor.R")
13 source("pruebasbondadajuste.R")
14 source("medidortasarechazo.R")
15
16 archivosdatos <- "archivosdatos3.sqlite"
17 tamanosmuestras <- c(25,50,100,300,500)
18 parshapeIG <- c(0.1,0.25,0.5,1:20)
19 parshapeGamma <- c
   (0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1,1.2,1.4,1.6,1.8,2:10)
20 parshapeWeibull <- c(0.2,0.4,0.6,0.8,1,1.5,2:20)
21 parsdlogLogNorm <- c
   (0.2,0.4,0.6,0.8,1,1.2,1.4,1.6,1.8,2,2.2,2.4,2.6,2.8,3,3.2,3.4,3.6,3.8,4)
22
23
24 for(n in tamanosmuestras) {
25   argumentos["n"] <- n
```

```
26 ## Se analiza en esta parte el tamaño de prueba
27 distribucion <- list(name="rinvgauss")
28 parfijos <- c("mean"=1)
29 parvars <- data.frame(shape=parshapeIG)
30 ejecutarprograma2(iteraciones, distribucion=distribucion, parfijos=
    parfijos,
31 parvars=parvars,pruebabondad, argumentos=argumentos, archivosdatos=
    archivosdatos)
32
33 ## En esta parte se analiza la potencia de la prueba
34 ## Primero se analiza la potencia cuando la muestra proviene de una
    distribucion
35 ## Gamma
36 distribucion <- list(name="rgamma")
37 parfijos <- c("scale"=1)
38 parvars <- data.frame(shape=parshapeGamma)
39 ejecutarprograma2(iteraciones, distribucion=distribucion, parfijos=
    parfijos,
40 parvars=parvars,pruebabondad, argumentos=argumentos, archivosdatos=
    archivosdatos)
41
42 ## Luego cuando proviene de una distribucion Weibull
43 distribucion <- list(name="rweibull")
44 parfijos <- c("scale"=1)
45 parvars <- data.frame(shape=parshapeWeibull)
46 ejecutarprograma2(iteraciones, distribucion=distribucion, parfijos=
    parfijos,
47 parvars=parvars,pruebabondad, argumentos=argumentos, archivosdatos=
    archivosdatos)
48
49 ## Finalmente se obtiene la potencia cuando los datos provienen de una
    distribucion
50 ## Log Normal
51 distribucion <- list(name="rlnorm")
52 parfijos <- c("meanlog"=0)
53 parvars <- data.frame(sdlog=parsdlogLogNorm)
```

```
54 ejecutarprograma2(iteraciones, distribucion=distribucion, parfijos=
    parfijos,
55 parvars=parvars,pruebabondad, argumentos=argumentos, archivosdatos=
    archivosdatos)
56
57 }
```

---

Se analiza la prueba de bondad de ajuste con el estadístico de prueba  $T_{1,n}$  en su versión asintótica. El código de la prueba de bondad de ajuste fue proporcionado por los autores de [Villaseñor y González-Estrada \(2015a\)](#).

Listing .6: analizarT1asin.R

---

```
1 # Las funciones t1, t2, t2. y t3 son las que se usaron para
2 # obtener los resultados del artículo publicado en SPL en 2015.
3
4 # Prueba T_{1,n} version asintotica
5 t1 <- function(x, argumentos)
6 {
7   n <- length(x)
8   m <- mean(x)
9   v <- mean(1/x-1/m)
10  s2 <- var(x)
11  r <- s2/(m^3*v)
12  t <- sqrt(n/(6*m*v))*(r-1)
13  q1 <- qnorm(1-argumentos["alpha"]/2)
14  return(ifelse(abs(t)>q1,1,0))
15
16 }
17
18 pruebabondad <- "t1"
19 iteraciones <- 10000
20 iteracionesporciclo <- 100
21 argumentos <- c(alpha=0.05)
22
23 source("tasarechazo.R")
```

---

Se analiza la prueba de bondad de ajuste con el estadístico de prueba  $T_{2,n}$ . El código de la prueba de bondad de ajuste fue proporcionado por los autores de [Villaseñor y](#)

[González-Estrada \(2015a\)](#).

Listing .7: analizarT1asin.R

---

```
1 ## Prueba T2: Version exacta para muestra pequenas
2 t2 <- function(x, argumentos)
3 {
4   n <- length(x)
5   z <- ((x - mean(x) )**2)/x
6   a.check <- mean(z)/cov(z,log(z))
7   t2      <- (a.check-0.5)/sqrt(.3669/n)
8
9   if(n==25)   return(ifelse( t2 > -1.420912 && t2 < 3.037665,0,1))
10  if(n==50)   return(ifelse( t2 > -1.552335 && t2 < 2.640791,0,1))
11  if(n==100)  return(ifelse( t2 > -1.651097 && t2 < 2.388018,0,1))
12  if(n==300)  return(ifelse( t2 > -1.767500 && t2 < 2.188671,0,1))
13  if(n==500)  return(ifelse( t2 > -1.790535 && t2 < 2.131355,0,1))
14 }
15
16 pruebabondad <- "t2"
17 iteraciones <- 10000
18 iteracionesporciclo <- 100
19 argumentos <- c(alpha=0.05)
20
21 source("tasarechazo.R")
```

---

Se analiza la prueba de bondad de ajuste con el estadístico de prueba  $T_{3,n}$ . El código de la prueba de bondad de ajuste fue proporcionado por los autores de [Villaseñor y González-Estrada \(2015a\)](#).

Listing .8: analizarT3.R

---

```
1
2 # Prueba de Anderson-Darling basada en datos transformados
3 # los valores criticos se obtuvieron simulando numeros de la distribucion
   Gamma(1/2, 1)
4 t3 <- function(x, argumentos)
5 {
6   z <- ((x - mean(x))**2)/x
7   n <- length(z)
```

```
8 b. <- 2*mean(z)
9 s <- sort(z)/b.
10 theop <- pgamma(s,.5)
11 # Anderson-Darling statistic
12 ad <- -n-sum((2*(1:n)-1)*log(theop) + (2*n+1-2*(1:n))*log(1-theop))/n
13
14 if(n==25) return(iffelse(ad > 1.469049,1,0))
15 if(n==50) return(iffelse(ad > 1.484123,1,0))
16 if(n==100) return(iffelse(ad > 1.489746,1,0))
17 if(n==300) return(iffelse(ad > 1.493149,1,0))
18 if(n==500) return(iffelse(ad > 1.498679,1,0))
19 }
20
21 pruebabondad <- "t3"
22 iteraciones <- 10000
23 iteracionesporciclo <- 100
24 argumentos <- c(alpha=0.05)
25
26 source("tasarechazo.R")
```

---

Se analiza la prueba de bondad de ajuste propuesta por [Vexler \*et al.\* \(2011\)](#), y se registra su tamaño de prueba y potencia.

Listing .9: analizarTKN.R

---

```
1 calculo_TKNM <- function(y, m) {
2   n <- length(y)
3   indice1 <- 1:n + m
4   indice2 <- 1:n -m
5   indice2[which(indice2 < 1)] <- 1
6   indice1[which(indice1 > n)] <- n
7   sum(log(2*m/(n * (y[indice1] - y[indice2])))))
8 }
9
10 ### TKNM
11 calculo_TKN <- function(x) {
12   x <- sort(x)
13   y <- sort(x, decreasing=T) ^(-1/2)
```

```
14   n <- length(x)
15   m_max <- floor(n^0.5)
16   TKNMs <- rep(NA, m_max)
17   for(i in 1:m_max) TKNMs[i] <- calculo_TKNM(y, i)
18   lambda <- calculo_ML_invgauss(x)[2]
19   TKN <- n/2*log(exp(1)*pi/(2*lambda)) + min(TKNMs)
20   return(TKN)
21 }
22
23 # Prueba TKN
24 # version exacta (muestras pequenas)
25 tkn <- function(x, argumentos)
26 {
27   n <- length(x)
28   tkn <- calculo_TKN(x)
29
30   if(n==25)   return(ifelse(tkn > 7.9547,1,0))
31   if(n==50)   return(ifelse(tkn > 9.9838,1,0))
32   if(n==100)  return(ifelse(tkn > 12.3785,1,0))
33   if(n==300)  return(ifelse(tkn > 17.3100,1,0))
34   if(n==500)  return(ifelse(tkn > 20.06328,1,0))
35 }
36
37 pruebabondad <- "tkn"
38 iteraciones <- 10000
39 iteracionesporciclo <- 100
40 argumentos <- c(alpha=0.05, R=500)
41
42 source("tasarechazo.R")
```

---

Se analiza la implementación del bootstrap paramétrico a la prueba de bondad de ajuste de razón de varianzas  $T_{1,n}$

Listing .10: analizarT1boot.R

---

```
1 #Calculo de T1n
2 calculo_T1n <- function(x) {
3   n <- length(x)
```

## Anexos

---

```
4   x_mean <- mean(x)
5   vn <- sum(1/x-1/x_mean)/n
6   if(vn < 0) stop("Varianza negativa")
7   Sn <- var(x)
8   sigma_emv <- x_mean^3*vn
9   Rn <- Sn/sigma_emv
10  sqrt(n/(6*x_mean*vn))*(Rn-1)
11 }
12
13 #Calculo de R
14 calculo_R <- function(x) {
15   n <- length(x)
16   x_mean <- mean(x)
17   vn <- sum(1/x-1/x_mean)/n
18   #   if(vn < 0) return(NA)
19   if(vn < 0) stop("varianza negativa")
20   Sn <- var(x)
21   sigma_emv <- x_mean^3*vn
22   Sn/sigma_emv
23 }
24
25
26
27 ##### Metodo de bootstrap parametrico MC simulacion ###
28 ### Estadistico T1 ML #####
29 pruebabondad_bootMC_T1n_ML <- function(x,argumentos) {
30   pruebabondad_bootMC(x, argumentos, calculo_ML_invgauss, calculo_T1n)
31 }
32
33 pruebabondad <- "pruebabondad_bootMC_T1n_ML"
34 iteraciones <- 10000
35 iteracionesporciclo <- 100
36 argumentos <- c(alpha=0.05, R=200)
37
38 source("tasarechazo.R")
```

---

Se analiza la implementación del bootstrap paramétrico a la prueba de bondad de ajuste de razón de varianza  $R_n$

Listing .11: analizarRboot.R

---

```
1 #Calculo de R
2 calculo_R <- function(x) {
3   n <- length(x)
4   x_mean <- mean(x)
5   vn <- sum(1/x-1/x_mean)/n
6   #   if(vn < 0) return(NA)
7   if(vn < 0) stop("varianza negativa")
8   Sn <- var(x)
9   sigma_emv <- x_mean^3*vn
10  Sn/sigma_emv
11 }
12
13 ### Estadistico R #####
14 pruebabondad_bootMC_R_ML <- function(x,argumentos) {
15   pruebabondad_bootMC(x, argumentos, calculo_ML_invgauss, calculo_R)
16 }
17
18 pruebabondad <- "pruebabondad_bootMC_R_ML"
19 iteraciones <- 10000
20 iteracionesporciclo <- 100
21 argumentos <- c(alpha=0.05, R=200)
22
23 source("tasarechazo.R")
```

---