



COLEGIO DE POSTGRADUADOS

INSTITUCION DE ENSEÑANZA E INVESTIGACION EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

POSTGRADO DE SOCIOECONOMÍA, ESTADÍSTICA E INFORMATICA

ESTADISTICA

Una prueba de razón de varianzas para la distribución Gumbel

BERESTRAYN CRUZ SIERRA

T E S I S

**PRESENTADA COMO REQUISITO PARCIAL
PARA OBTENER EL GRADO DE:**

MAESTRO EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MEXICO

2016

La presente tesis titulada: **Una prueba de razón de varianzas para la distribución Gumbel**, realizada por el alumno: **Berestravn Cruz Sierra**, 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 INFORMÁTICA
ESTADÍSTICA

CONSEJO PARTICULAR

CONSEJERO



Dr. José Aurelio Villaseñor Alva

ASESOR



Dra. Elizabeth González Estrada

ASESOR



M.C. Margarito Soriano Montero

Montecillo, Texcoco, Estado de México, Noviembre de 2016

Una prueba de razón de varianzas para la distribución Gumbel

Berestravn Cruz Sierra, M.C.

Colegio de Postgraduados, 2016

RESUMEN

Las pruebas de bondad de ajuste son importantes en cualquier análisis de datos. Estas pruebas son útiles para saber si los datos que se están manejando satisfacen los supuestos distribucionales requeridos en los métodos estadísticos que se utilizarán para realizar la inferencia con los mismos. En el caso de la distribución Gumbel, que es ampliamente utilizada en análisis de datos extremos en áreas como la hidrología, finanzas y ciencias ambientales, entre otras; existen diversas pruebas de bondad de ajuste, pero como es bien conocido, ninguna de ellas es la más potente contra cualquier distribución alternativa. En este trabajo se propone una prueba de bondad de ajuste para la distribución Gumbel, basada en la razón de dos estimadores de la varianza. Para esto se propone utilizar un estimador de la varianza que sea una función lineal de las estadísticas de orden y dividirlo por el estimador de la varianza muestral, de manera similar a la estadística de Shapiro-Wilk para probar la hipótesis de normalidad. Los valores críticos para la prueba propuesta son obtenidos utilizando simulación Monte de Carlo. Posteriormente para evaluar sus propiedades, se discuten los resultados de un estudio de simulación en donde se analiza el tamaño y la potencia de la prueba contra algunas distribuciones alternativas como son la Cauchy, Normal, logística y valores extremos generalizada.

Palabras clave: Prueba de Shapiro-Wilk, Bondad de Ajuste, Distribución de valores extremos.

A ratio goodness-of-fit test for extreme value distributions

Berestravn Cruz Sierra, M.Sc.

Colegio de Postgraduados, 2016

ABSTRACT

The goodness of fit tests are important in any analysis. These tests are useful to know whether the data being handled meet the distributional assumptions required in the statistical methods to be used for inference with them. In the case of the Gumbel distribution, which is widely used in analysis of extreme data in areas such as hydrology, finance and environmental sciences, among others; There are several tests for goodness of fit, but as is well known, none of them is more powerful against any alternative distribution. In this work a goodness of fit test for the Gumbel distribution, based on the ratio of two variance estimators is proposed. For this it is proposed to use a variance estimator that is a linear function of the order statistics and divide it by the estimator of the sample variance, similar to the Shapiro-Wilk statistic to test the hypothesis of normal way. The critical values for testing proposal are obtained using Monte Carlo simulation. Subsequently to evaluate their properties, the results of a simulation study where the size and power of the test against some alternatives such as the Cauchy distributions, Normal, logistics and generalized extreme values discussed are discussed.

Keywords: Shapiro-Wilk test, Goodness of Fit, extreme value distribution.

Contenido

| | |
|--|------------|
| Lista de figuras | VII |
| Lista de cuadros | IX |
| 1. Introducción | 1 |
| 2. Objetivos | 3 |
| 2.1. Objetivos Generales | 3 |
| 2.2. Objetivos Particulares | 3 |
| 3. Marco Teórico | 4 |
| 3.1. Desarrollo histórico de la teoría de valores extremos | 4 |
| 3.2. Teoría de valores extremos y modelos | 5 |
| 3.3. Distribución de Valores Extremos Generalizada | 7 |
| 3.4. Distribución Gumbel | 8 |
| 3.4.1. Transformaciones | 9 |

| | | |
|-----------|--|-----------|
| 3.4.2. | Función Generatriz de momentos | 9 |
| 3.4.3. | Media y Varianza | 10 |
| 3.4.4. | Generación de variables aleatorias Gumbel | 10 |
| 3.5. | Prueba de Hipótesis | 10 |
| 3.6. | Pruebas de Bondad de Ajuste | 12 |
| 3.6.1. | Prueba de Anderson-Darling | 14 |
| 3.7. | Estimador de Kimball | 15 |
| 4. | Prueba de razón de varianzas para la distribución Gumbel | 17 |
| 4.1. | Planteamiento del problema | 17 |
| 4.2. | Construcción de la estadística y prueba de hipótesis | 17 |
| 4.3. | Distribución del estadístico de prueba | 18 |
| 4.4. | Constantes Críticas | 22 |
| 5. | Estudio de Simulación Monte Carlo para estimar y comparar la potencia y tamaño de la prueba, contra algunas distribuciones alternativas | 24 |
| 5.1. | Tamaño de la Prueba | 24 |
| 5.1.1. | Discusión de los resultados de simulación | 25 |
| 5.2. | Potencia de la Prueba | 25 |
| 5.2.1. | Discusión de los resultados de simulación | 26 |
| 5.3. | Estudio comparativo de la prueba V_n con la Prueba de Anderson-Darling | 29 |
| 6. | Ejemplos de aplicación de la prueba propuesta | 33 |
| 6.1. | Ejemplo 1 | 33 |
| 6.1.1. | Datos de Ozono de la Ciudad de México | 33 |
| 6.2. | Ejemplo 2 | 37 |
| 6.2.1. | Lluvias máximas consecutivas para 1 día en Álamo, Veracruz | 37 |

| | |
|--|-----------|
| 7. Conclusiones | 41 |
| Bibliografía | 42 |
| Anexos A | 45 |
| .1. Propiedades de la distribución de valores extremos Tipo Gumbel | 45 |
| .1.1. Función de densidad | 45 |
| .1.2. Función de distribución | 45 |
| .1.3. Función generatriz de momentos | 45 |
| .1.4. Media y Varianza | 46 |
| .1.5. Generación de variables aleatorias Gumbel | 46 |
| .2. Estimadores de Máxima Verosimilitud de la Distribución Gumbel | 46 |
| Anexos B | 47 |
| .3. Cálculo de las constantes críticas | 47 |
| .4. Distribución de la estadística V_n para varios tamaños de muestra | 48 |
| .5. Distribución de la estadística V_n variando parámetros de localidad y escala | 50 |
| .6. Tamaño de la prueba de la estadística V_n para distintos tamaños de la muestra | 52 |
| .7. Potencias de la prueba estadística V_n para distintos tamaños de la muestra | 55 |
| .7.1. Distribuciones Normal, Logística y Cauchy variando el parámetro de localid- dad ξ | 55 |
| .7.2. Distribuciones Normal, Logística y Cauchy variando el parámetro de escala θ | 61 |
| .7.3. Distribución t-student variando los grados de libertad | 66 |
| .7.4. Distribución de Valores Extremos Generalizada $GEV(\xi, \theta, \kappa)$ | 69 |
| .8. Comparación de las potencias de la estadística de prueba v vs Anderson-Darling variando los tamaños de la muestra | 71 |
| .8.1. Distribución Normal variando el parámetro de escala θ | 71 |
| .8.2. Distribución Logística variando el parámetro de escala θ | 75 |

| | | |
|-------|--|----|
| .8.3. | Distribución t-student variando los grados de libertad | 78 |
| .8.4. | Distribución de Valores Extremos Generalizada variando su parámetro de forma κ | 82 |
| .9. | Ejemplos de aplicación | 85 |

Lista de figuras

| | | |
|------|---|----|
| 4.1. | Distribución de la estadística V_n para varios tamaños de muestra. | 19 |
| 4.2. | Distribución de la estadística V_n cuando $n = 10$ variando los valores del parámetro de localidad ξ (izquierda) y escala θ (derecha), respectivamente. | 20 |
| 4.3. | Distribución de la estadística V_n cuando $n = 30$ variando los valores del parámetro de localidad ξ (izquierda) y escala θ (derecha), respectivamente. | 20 |
| 4.4. | Distribución de la estadística V_n cuando $n = 50$ variando los valores del parámetro de localidad ξ (izquierda) y escala θ (derecha), respectivamente. | 21 |
| 4.5. | Distribución de la estadística V_n cuando $n = 100$ variando los valores del parámetro de localidad ξ (izquierda) y escala θ (derecha), respectivamente. | 21 |
| 4.6. | Distribución de la estadística V_n cuando $n = 200$ variando los valores del parámetro de localidad ξ (izquierda) y escala θ (derecha), respectivamente. | 22 |
| 5.1. | Tamaño estimado de la prueba variando los tamaños de muestra iguales a $n = 10, 30, 50, 100$ y 200 , con $\alpha = 0.05$ | 25 |
| 5.2. | Potencia estimada de la prueba contra las distribuciones Normal, Logística y Cauchy variando los tamaños de muestra $n = 10, 30, 50, 100$ y 200 , con $\alpha = 0.05$ | 27 |

| | | |
|------|--|----|
| 5.3. | Potencia estimada de la prueba contra la distribución t-student variando los grados de libertad $\nu \in (1, 20)$, para distintos tamaños de muestra y $\alpha = 0.05$ | 28 |
| 5.4. | Potencia estimada de la prueba contra la Distribución de Valores Extremos Generalizada ($\xi = 0, \theta = 1$) variando el parámetro de forma, para distintos tamaños de muestra y $\alpha = 0.05$ | 29 |
| 5.5. | Comparación de la potencia de la prueba V_n y Anderson-Darling para las distribuciones Normal y Logística variando los tamaños de muestra $n = 10, 30, 50, 100$ y 200 , con $\alpha = 0.05$ | 30 |
| 5.6. | Comparación de la potencia de la prueba V_n y Anderson-Darling para la distribución T-Student variando los grados de libertad $\nu = 2, 3, 4, 5, 6, 7, 8, 9, \dots, 20$, para tamaños de muestra $n = 10, 30, 50, 100$ y 200 | 31 |
| 5.7. | Comparación de la potencia de la prueba V_n y Anderson-Darling para la distribución de Valores Extremos variando el parámetro forma, para tamaños de muestra $n = 10, 30, 50, 100$ y 200 | 32 |
| 6.1. | Observaciones de los cluster de ozono que excedieron el umbral $u = 0.11$, resultantes de los niveles máximos diarios de ozono correspondientes a las estaciones en verano de los años 2003 a 2007. | 35 |
| 6.2. | Auto correlación parcial para las observaciones de los cluster de ozono que excedieron el umbral $u = 0.11$, resultantes de los niveles máximos diarios de ozono correspondientes a las estaciones en verano de los años 2003 a 2007. | 36 |
| 6.3. | Gráfica Q-Q para los cluster de ozono que excedieron el umbral $u = 0.11$, resultantes de los niveles máximos diarios de ozono correspondientes a las estaciones en verano de los años 2003 a 2007. | 36 |
| 6.4. | Lluvias máximas consecutivas para 1 día en Álamo, Veracruz. | 38 |
| 6.5. | Auto correlación parcial para los datos de lluvias máximas consecutivas para 1 día en Álamo, Veracruz. | 39 |
| 6.6. | Gráfica Q-Q para las lluvias máximas consecutivas para 1 día en Álamo, Veracruz. | 39 |

Lista de cuadros

| | |
|---|----|
| 4.1. Constantes críticas para distintos tamaños de muestra $n = 10, 30, 50, 100$ y 200 con niveles de significancia $\alpha = 0.01, 0.05, 0.1$ y 0.25 | 23 |
| 6.1. Cluster de ozono y tamaños de cada cluster que excedieron el umbral $u = 0.11$. . . | 34 |
| 6.2. Lluvias máximas consecutivas para 1 día en Álamo, Veracruz. | 38 |

Introducción

Las pruebas de bondad de ajuste son importantes en cualquier análisis de datos. Estas pruebas permiten saber si los datos que se están manejando satisfacen los supuestos distribucionales requeridos en los métodos estadísticos, que se utilizarán para hacer el proceso de inferencia con los mismos.

En el caso de la distribución Gumbel, que es ampliamente utilizada en análisis de datos extremos en áreas como la hidrología, finanzas, ciencias ambientales, entre otras; existen ya diversas pruebas de bondad de ajuste, pero ninguna de ellas es la mejor en forma absoluta.

Mann et al. (1974) crearon un método basado en las diferencias de las estadísticas de orden adyacentes divididas por sus esperanzas. Tsujitani et al. (1980) presentaron un método basado en la entropía muestral. Chandra et al. (1981) desarrollaron una estadística parecida a la de Kolmogorov, permitiendo además la estimación de los parámetros desconocidos, por el método de máxima verosimilitud. Kinnison (1989) obtuvo una tabla por medio de simulación Monte Carlo para hacer la prueba de bondad de ajuste para esta distribución, utilizando como estadística de prueba el coeficiente de correlación.

Las pruebas de bondad de ajuste basadas en la razón de varianzas son invariantes a parámetros de localidad y escala y tienen propiedades deseables de potencia, lo cual las hace muy atractivas para los investigadores en el campo de la Estadística.

En el presente trabajo se desarrolla una prueba de bondad de ajuste para la distribución Gumbel, utilizando una razón de varianzas, para ello, se utilizan dos estimadores distintos de la varianzas, uno en términos de las estadísticas de orden, algo similar a la forma de como esta construida la prueba de Shapiro Wilk. La estructura del trabajo esta conformada por seis capítulos.

En el Capítulo 2, se enuncian los objetivos generales y particulares a cubrir. En el Capítulo 3, se presenta un marco teórico necesario, que incluye revisión de conceptos básicos sobre pruebas de hipótesis y algunos criterios para clasificar las pruebas de bondad de ajuste. También se muestran algunas propiedades de la distribución Gumbel y por último, el estimador de Kimball, el cual será importante para la creación de la prueba propuesta.

En el Capítulo 4, se muestra el planteamiento y construcción de la prueba para la distribución Gumbel, así como la distribución del estadístico de prueba. Por su parte, en el Capítulo 5 se realiza la simulación Monte Carlo para estimar la potencia y tamaño de la prueba contra algunas distribuciones alternativas y comparar la misma contra la prueba de Anderson-Darling. Se generaron tablas de valores críticos, para diferentes tamaños de muestra, niveles de significancia.

En el Capítulo 6 se implementa la prueba para la distribución Gumbel con algunos ejemplos prácticos; finalmente, en el Capítulo 7 se discuten las conclusiones obtenidas mediante los resultados de simulación Monte Carlo.

Objetivos

En este trabajo se tienen como metas la realización de los siguientes puntos:

2.1. Objetivos Generales

- Desarrollar una prueba de bondad de ajuste para la Distribución de Valores Extremos Tipo Gumbel que posea buena potencia, utilizando razón de varianzas.

2.2. Objetivos Particulares

- Revisar algunas propiedades de la Distribución de Valores Extremos Tipo Gumbel.
- Obtener la tabla de valores críticos para la prueba de bondad de ajuste de la distribución antes mencionada para diferentes niveles de significancia.
- Estudiar el comportamiento de la prueba con respecto al tamaño de la muestra.
- Comparar la potencia de la propuesta con la potencia de la prueba de Anderson Darling bajo las distribuciones alternativas siguientes: T-Student, Cauchy, Normal, Logística y Distribución de Valores Extremos Generalizada.

3.1. Desarrollo histórico de la teoría de valores extremos

El desarrollo de distribuciones de valores extremos procedió hasta cierto punto fuera de la corriente principal de la teoría de la distribución estadística, con su primera etapa dominada por el trabajo de ajuste de curvas y posteriormente por problemas encontrados en la inferencia estadística.

La teoría del valor extremo parece haberse originado principalmente por las necesidades de los astrónomos de la utilización o el rechazo de las observaciones periféricas.

Los primeros trabajos por Fuller (1914) y Griffith (1921) sobre el tema fueron altamente especializados tanto en el campo de las aplicaciones como en los métodos de análisis matemático.

Un desarrollo sistemático de la teoría general puede haber comenzado con un artículo de Borthiewicz (1922) que se ocupa de la distribución del rango de muestras aleatorias de una distribución normal. Desde el punto de vista actual, es suficiente decir que la importancia del artículo de Borthiewicz reside en el hecho de que el concepto de distribución del máximo se introdujo claramente en ella. Von Misses (1936) evaluó el valor esperado de esta distribución y Dodd (1923) calculó la mediana y también discutió sobre algunas distribuciones no normales. De igual relevancia es un trabajo de Fréchet (1927) en la que se consideran las distribuciones asintóticas de valores

máximos.

En el año siguiente Fisher y Tippett publicaron los resultados de una investigación independiente sobre el mismo problema. Mientras Fréchet (1927) había identificado una posible distribución límite de la estadística de orden para el máximo, Fisher y Tippett (1928) mostraron que las distribuciones extremas solo pueden ser de tres tipos.

Fue Gnedenko (1943), quien presentó un fundamento riguroso de la teoría del valor extremo y probó las condiciones necesarias y suficientes para la convergencia débil de las estadísticas de orden extremas. de Haan (1970) perfeccionó el trabajo de Gnedenko.

Los desarrollos teóricos de la década de 1920 y mediados de 1930 se siguieron en las décadas siguientes por una serie de documentos relacionados con las aplicaciones prácticas de las estadísticas de valores extremos en la distribución de vidas humanas, las emisiones radiactivas, resistencia de materiales, análisis de inundaciones, análisis sísmico y análisis de las precipitaciones.

Desde el punto de vista de aplicación, Gumbel hizo varias contribuciones importantes en el análisis de valores extremos.

3.2. Teoría de valores extremos y modelos

Sea X_1, \dots, X_n una sucesión de variables aleatorias independientes que tienen una función de distribución común F .

El modelo se centra en el comportamiento estadístico de:

$$M_n = \text{Max}\{X_1, \dots, X_n\} \tag{3.2.1}$$

En aplicaciones, X_i representa los valores de un fenómeno medido en tiempos regulares; por ejemplo, mediciones tomadas cada cierto tiempo del nivel del mar, por lo que M_n representa el máximo de n observaciones unitarias de un fenómeno. Si n es el número de observaciones de un año, entonces M_n corresponde al máximo anual.

En teoría, la distribución de M_n puede derivarse exactamente de todos los valores de n :

$$\begin{aligned}
P(M_n \leq z) &= P(X_1 \leq z, \dots, X_n \leq z) \\
&= P(X_1 \leq z) \times \dots \times P(X_n \leq z) \\
&= \prod_{i=1}^n P(X_i \leq z) \\
&= (F(z))^n
\end{aligned} \tag{3.2.2}$$

Note que $P(M_n \leq z) \rightarrow 0$ cuando $n \rightarrow \infty$.

Sin embargo, esto no resulta inmediatamente útil en la práctica ya que generalmente la función de distribución F es desconocida. Una posible solución es utilizar técnicas estadísticas para estimar F a partir de las observaciones y sustituir esta estimación en (3.2.2). Por desgracia, muy pequeñas discrepancias en la estimación de F pueden llevar a discrepancias sustanciales para F^n .

Un enfoque alternativo consiste en aceptar que F sea desconocida y buscar familias aproximadas de modelos para F , los cuales pueden ser estimados tomando en cuenta sólo los valores extremos. Esto es similar a la práctica usual de aproximar la distribución de las medias muestrales con la distribución normal que es justificada por el Teorema Central del Límite.

Análogo al Teorema Central del Límite se toma este argumento para encontrar las distribuciones límite de valores extremos.

La distribución de valor extremo es la distribución que se ajusta a los máximos de bloques independientes de observaciones adecuadamente normalizadas y surge a partir del desarrollo del teorema de Fisher-Tippett, Gnedenko.

Teorema 3.1 (Fisher-Tippett, Gnedenko). *Sea $\{X_n\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con una función de distribución común $F(x)$. Si existen sucesiones de constantes de normalización $\{a_n > 0\}$ y $\{b_n\} \in \mathfrak{R}$ y una función de distribución G no degenerada, tal que:*

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq z\right) = G(z) \tag{3.2.3}$$

entonces $G(z)$ pertenece a alguna de las siguientes tres familias:

1. Tipo I (Distribución tipo Gumbel)

$$P[Z \leq z] = \exp \left\{ -e^{-\frac{(z-\mu)}{\theta}} \right\} \quad (3.2.4)$$

2. Tipo II (Distribución tipo Fréchet)

$$P[Z \leq z] = \begin{cases} 0, & z < \mu \\ \exp \left\{ -\left(\frac{z-\mu}{\theta}\right)^{-\xi} \right\}, & z \geq \mu \end{cases} \quad (3.2.5)$$

3. Tipo III (Distribución tipo Weibull)

$$P[Z \leq z] = \begin{cases} \exp \left\{ -\left(\frac{z-\mu}{\theta}\right)^{-\xi} \right\}, & z \leq \mu \\ 1, & z > \mu \end{cases} \quad (3.2.6)$$

Donde $\mu \in \mathfrak{R}$ y $\xi > 0$ son los parámetros.

Las distribuciones correspondientes de $(-Z)$ también se llaman distribuciones de valores extremos (Observe que las distribuciones Fréchet y Weibull están relacionadas por un simple cambio de signo).

De estas familias de distribuciones, la tipo I es la más comúnmente referida en las discusiones de los valores extremos. De hecho, algunos autores llaman a (3.2.4) la distribución de valores extremos. En vista de esto, y el hecho de que las distribuciones (3.2.5) y (3.2.6) pueden ser transformadas a la distribución tipo I por transformaciones sencillas:

$$Z = \log(X - \mu) \quad y \quad Z = -\log(\mu - X) \quad (3.2.7)$$

3.3. Distribución de Valores Extremos Generalizada

Los tres tipos de distribuciones en (3.2.4) - (3.2.6) pueden ser representados como miembros de una única familia llamada Distribución de Valores Extremos Generalizada (GEV) con función de distribución acumulativa:

$$P[Z \leq z] = \exp \left\{ -\left[1 + \xi \left(\frac{z-\mu}{\theta} \right) \right]^{-\frac{1}{\xi}} \right\}, \quad 1 + \xi \left(\frac{z-\mu}{\theta} \right) > 0 \quad (3.3.1)$$

$$\xi, \mu \in \mathfrak{R}, \theta > 0$$

Con las siguientes características,

1. Para $\xi > 0$, la distribución (3.3.1) es de la misma forma que (3.2.5).
2. Para $\xi < 0$, la distribución (3.3.1) es de la misma forma que (3.2.4).
3. Finalmente, cuando $\xi \rightarrow 0$, la distribución (3.3.1) toma la forma de la distribución de valor extremo tipo I, (3.2.4).
4. La función de distribución de (3.3.1) también es conocida como la Distribución de Valor Extremo tipo Von Mises o la Distribución tipo Mises-Jenkinson.

3.4. Distribución Gumbel

La distribución Gumbel es una distribución fundamental en la teoría de valores extremos. La distribución tipo I también se le llama distribución doblemente exponencial, a causa de la forma funcional de (3.2.4). Mediante la sustitución de Z por $-Z$, las distribuciones límites de los valores mínimos son obtenidas.

La distribución Gumbel es llamada de esta manera en honor a Emil Julius Gumbel (1891-1966). Es un caso particular de la distribución de valores extremos generalizada. Se emplea frecuentemente para predecir eventos en áreas tales como: hidrología, meteorología, finanzas, resistencia de materiales, emisiones radioactivas, entre otras.

Según Reiss y Thomas (2007), la distribución Gumbel tiene la misma importancia que la distribución normal en otras aplicaciones.

Si X es una variable aleatoria con distribución Gumbel con parámetros ξ y θ , su función de densidad está dada por:

$$f_X(x) = \begin{cases} \frac{1}{\theta} \exp \left\{ -e^{-\frac{x-\xi}{\theta}} - \frac{x-\xi}{\theta} \right\} I_{(-\infty, \infty)}(x), & \xi \in \mathfrak{R}, \theta > 0 \\ 0, & \text{d.o.m.} \end{cases} \quad (3.4.1)$$

Mientras que su función de distribución está dada por:

$$F_X(x) = \exp \left\{ -e^{-\frac{(x-\xi)}{\theta}} \right\} I_{(-\infty, \infty)}(x) \quad (3.4.2)$$

3.4.1. Transformaciones

Sea X una variable aleatoria que se distribuye Gumbel con parámetros ξ y θ , lo cual se denota como $X \sim Gumbel(\xi, \theta)$, entonces:

- Si $\xi = 0$ y $\theta = 1$ o equivalente si se hace la transformación $Z = \frac{X-\xi}{\theta}$ se obtiene la forma estándar de la distribución Gumbel.

$$f_Z(z) = \exp\{-e^{-z} - z\} I_{(-\infty, \infty)}(z) \quad (3.4.3)$$

- Ahora, sea la transformación $Y = e^{-\frac{X-\xi}{\theta}}$, entonces Y tiene distribución exponencial dada por la siguiente expresión:

$$f_Y(y) = \exp\{-y\} I_{(-\infty, \infty)}(y) \quad (3.4.4)$$

Conocida como la distribución exponencial con parámetro de escala igual a 1.

3.4.2. Función Generatriz de momentos

Sea $Z \sim Gumbel(0, 1)$, una variable aleatoria Gumbel estándar presentada en (3.4.3), entonces:

$$\begin{aligned} m_Z(t) &= E\{e^{tZ}\} = E\left\{e^{\frac{t(X-\xi)}{\theta}}\right\} = E\left\{\left(e^{-\frac{X-\xi}{\theta}}\right)^t\right\} = E\{Y^{-t}\} \\ &= \int_0^{\infty} y^{-t} e^{-y} dy = \int_0^{\infty} y^{(1-t)-1} e^{-y} dy = \Gamma(1-t), \quad \text{para } t < 1 \end{aligned} \quad (3.4.5)$$

Observe que $X = \xi + \theta Z$ y debido a la propiedad de la función generatriz de momentos que nos dice que si se transforma una variable aleatoria Z con una función lineal $X = a + bZ$, entonces la función generatriz de momentos para X esta dada por:

$$m_X(t) = e^{at} m_Z(bt) \quad (3.4.6)$$

Por lo tanto,

$$m_X(t) = e^{\xi t} \Gamma(1 - \theta t) \quad \text{para } \theta t < 1 \quad (3.4.7)$$

3.4.3. Media y Varianza

A partir de la función Generatriz de Momentos denotada en la expresión (3.4.7) se puede obtener la media y la varianza de la distribución, dadas por:

$$\mu = E(X) = \left. \frac{\partial M_X(t)}{\partial t} \right|_{t=0} = \xi + \theta\gamma_0 \quad (3.4.8)$$

Donde $\gamma_0 = 0.577216$ es llamada la constante de Euler. De la misma forma para la varianza.

$$\sigma^2 = E\{(X - \mu)^2\} = \frac{\pi^2\theta^2}{6} \quad (3.4.9)$$

3.4.4. Generación de variables aleatorias Gumbel

Sea U una variable aleatoria uniforme en el intervalo $(0, 1)$, entonces la variable aleatoria $X = \xi - \theta \ln(-\ln(U))$ tiene distribución $Gumbel(\xi, \theta)$, lo que se demuestra con el método de la transformada inversa.

3.5. Prueba de Hipótesis

Sea X_1, X_2, \dots, X_n una muestra aleatoria de una densidad $f_X(x; \theta)$ donde $\theta \in \Omega \subset \mathfrak{R}^k$, $k \geq 1$.

Definición 3.1. Una hipótesis estadística es una afirmación o conjetura acerca de la distribución de una o más variables aleatorias.

Sea $\omega \subset \Omega$. Considere la hipótesis:

$$H_0 : \theta \in \omega \quad y \quad H_1 : \theta \in \Omega - \omega$$

H_0 es llamada la hipótesis nula y H_1 la hipótesis alternativa.

Definición 3.2. Una prueba de hipótesis es una regla de decisión basada en $X = (X_1, X_2, \dots, X_n)$ que especifica para qué valores muestrales no se rechaza que H_0 es verdadera y para que valores se rechaza H_0 y se acepta que H_1 es verdadera.

Si χ denota al conjunto de todos los valores posibles del vector de observaciones X , entonces una prueba es una partición de χ en dos conjuntos, χ_R y χ_A , tales que $\chi_R \cup \chi_A = \chi$ y $\chi_R \cap \chi_A = \phi$. La regla de decisión es: rechazar H_0 si $x \in \chi_R$ o no rechazar H_0 si $x \in \chi_A$ donde x es una realización de X .

Típicamente, χ^R se describe en términos de una estadística $T = t(X)$ la cual es llamada la estadística de prueba. Al considerar $T = t(X)$, el espacio χ es transformado en el espacio τ el cual es el conjunto de todos los valores posibles de T .

Al tomar la decisión de rechazar o no rechazar H_0 puede ocurrir alguno de los siguientes tipos de errores:

Error de tipo I: se rechaza H_0 cuando H_0 es verdadera.

Error de tipo II: no se rechaza H_0 cuando H_0 es falsa.

Las pruebas de hipótesis se evalúan y comparan a través de las probabilidades de los errores de tipo I y II, las cuales son:

$$\begin{aligned} P\{\text{error de tipo I usando } \chi_R\} &= P\{x \in \chi_R | H_0 \text{ es verdadera}\} \\ &= P\{x \in \chi_R | \theta \in \omega\} \\ &= \int_{\chi_R} f_X(x; \theta) dx, \quad \theta \in \omega \\ &= \int_{\chi_R} f_T(t; \theta) dt, \quad \theta \in \omega \end{aligned}$$

$$\begin{aligned} P\{\text{error de tipo II usando } \chi_R\} &= P\{x \in \chi_A | H_0 \text{ es falsa}\} \\ &= P\{x \in \chi_A | \theta \in \Omega - \omega\} \\ &= \int_{\chi_A} f_X(x; \theta) dx, \quad \theta \in \Omega - \omega \\ &= \int_{\chi_A} f_T(t; \theta) dt, \quad \theta \in \Omega - \omega \end{aligned}$$

Definición 3.3. Se dice que una prueba χ_R es de tamaño α si

$$P\{\text{error de tipo I usando } X_R\} \leq \alpha, \quad \theta \in \omega, \quad \alpha \in (0, 1)$$

Es decir, χ_R es de tamaño α si

$$\sup_{\theta \in \omega} \{P\{\text{error de tipo I usando } X_R\}\} \leq \alpha, \quad \alpha \in (0, 1)$$

Definición 3.4. La función de potencia de una prueba χ_R es una función $\beta_{\chi_R}(\theta) : \Omega \rightarrow [0, 1]$ tal que

$$\begin{aligned}\beta_{\chi_R}(\theta) &= P \{ \text{rechazar } H_0 \text{ usando } \chi_R | \theta \} \\ &= P \{ x \in \chi_R | \theta \} \\ &= \int_{\chi_R} f_X(x; \theta) dx\end{aligned}$$

En general, no existe una prueba χ_R óptima. Por lo tanto, típicamente se fija un tamaño de prueba, digamos α_0 , y se trata de identificar la prueba uniformemente más potente de tamaño α_0 bajo la hipótesis alternativa.

3.6. Pruebas de Bondad de Ajuste

Definición 3.5. Una prueba de bondad de ajuste es un procedimiento estadístico para probar la hipótesis de que una función de distribución particular ajusta un conjunto de datos observados sin especificar una hipótesis alternativa.

Sea X_1, X_2, \dots, X_n una muestra aleatoria de una función de distribución $F_X(x)$. Cuando se enfrenta el problema de construir una prueba de bondad de ajuste para la hipótesis nula

$$H_0 : F_X(x) = F_0(x) \tag{3.6.1}$$

donde $F_0(x)$ denota una familia de distribuciones, no se cuenta con un criterio general para construirla. Kendall y Stuart (1973) mencionan que el problema de probar (3.6.1) es denominado el problema de la "prueba de bondad de ajuste". Es decir, el problema consiste en verificar si un conjunto de datos vienen de una distribución en particular, para lo cual se observa los mismos y posteriormente se decide si se rechaza o no se rechaza H_0 . Sin embargo, al revisar la literatura, D' Agostino y Stephens (1986) muestran muchas pruebas se han obtenido con base en alguno de los siguientes criterios.

1. Pruebas basadas en la distribución empírica
2. Pruebas basadas en la función característica empírica
3. Pruebas basadas en momentos

4. Pruebas de correlación y regresión

5. Pruebas basadas en caracterizaciones

Con respecto al punto (4) de la lista anterior, se tiene lo siguiente: Una prueba de correlación compara dos estimadores de σ^2 . Un ejemplo de este tipo es la prueba de Shapiro y Wilk (1965) para probar normalidad univariada cuya estadística de prueba es

$$W = \frac{\left[\sum_{i=1}^n a_i x_{(i)} \right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

para ciertas constantes a_1, a_2, \dots, a_n .

Reiss y Thomas (2001) dicen que para el caso de la distribución de valores extremos Tipo Gumbel, existen varios procedimientos para verificar si un conjunto de datos vienen de esta distribución, entre los más usados destacan los siguientes:

- Las pruebas basadas en espacios (Diferencias entre las estadísticas de orden)
- La prueba de χ^2
- La prueba de Kolmogorov-Smirnov

Además de los procedimientos mostrados arriba, en la literatura se pueden encontrar referencias a otras pruebas de bondad de ajuste. Stephens (1977) propuso una prueba de bondad de ajuste para la distribución de valores extremos, basada en la distribución empírica de las estadísticas W^2 , U^2 y A^2 :

$$W^2 = \sum_i \left(z_i - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n} \quad (3.6.2)$$

$$U^2 = W^2 - n \left(\bar{z} - \frac{1}{2} \right)^2 \quad (3.6.3)$$

$$A^2 = -n - \frac{1}{n} \sum_i (2i-1) \{ \log(z_i) + \log(1 - z_{n+1-i}) \} \quad (3.6.4)$$

Donde, n es el tamaño de la muestra, $\bar{z} = 1/n \sum_i z_i$, $z_i = F(x_i)$, $x_1 \leq x_2 \leq \dots \leq x_n$, $F(\cdot)$ está dada por (3.4.2), si los parámetros son desconocidos entonces son substituidos por sus correspondientes estimadores de máxima verosimilitud.

Stephens obtuvo los puntos críticos en los tres casos, cuando uno o ambos de los parámetros deben ser estimados a partir de la muestra.

Tsujitani et al. (1980) presentaron un método basado en la entropía muestral.

Chandra et al. (1981) obtuvieron los puntos porcentuales para la estadística de Kolmogorov-Smirnov $D+$, $D-$ y D para probar si un conjunto de datos se ajusta a la distribución de valores extremos.

Öztürk (1986) obtuvo una prueba de bondad de ajuste para la misma distribución basado en la estadística W de Shapiro-Wilk. Los valores críticos para la prueba fueron obtenidos utilizando simulación Monte Carlo.

Kinnison (1989), propuso una prueba de bondad de ajuste para la distribución Gumbel basada en el coeficiente de correlación, obteniendo los puntos críticos por medio de simulación Monte Carlo.

3.6.1. Prueba de Anderson-Darling

La prueba de Anderson-Darling es una prueba no paramétrica sobre si los datos de una muestra provienen de una distribución específica. La fórmula para el estadístico A determina si los datos $\{Y_1 < \dots < Y_N\}$ (observar que los datos se deben ordenar) vienen de una distribución con función acumulativa F .

$$A^2 = -N - S \quad (3.6.5)$$

Donde

$$S = \sum_{k=1}^N \frac{2k-1}{N} [\ln F(Y_k) + \ln (1 - F(Y_{N+1-k}))] \quad (3.6.6)$$

El estadístico de la prueba se puede entonces comparar contra las distribuciones del estadístico de prueba (dependiendo que F se utiliza) para determinar el P-valor.

La prueba de Anderson-Darling es una prueba estadística que permite determinar si una muestra de datos se extrae de una distribución de probabilidad.

En su forma básica, la prueba asume que no existen parámetros a estimar en la distribución que se está probando, en cuyo caso la prueba y su conjunto de valores críticos siguen una distribución libre. Sin embargo, la prueba se utiliza con mayor frecuencia en contextos en los que se está probando una familia de distribuciones, en cuyo caso deben ser estimados los parámetros de esa familia y debe tenerse estos en cuenta a la hora de ajustar la prueba estadística y sus valores críticos.

Cuando se aplica para probar si una distribución normal describe adecuadamente un conjunto de datos, es una de las herramientas estadísticas más potentes para la detección de la mayoría de las desviaciones de la normalidad.

3.7. Estimador de Kimball

Con base en una muestra aleatoria X_1, \dots, X_n es posible establecer las ecuaciones score para obtener los estimadores de máxima verosimilitud de los parámetros ξ y θ :

$$\hat{\theta} = \bar{x} - \frac{\sum_{i=1}^n x_i e^{-x_i/\hat{\theta}}}{\sum_{i=1}^n e^{-x_i/\hat{\theta}}} \quad (3.7.1)$$

y

$$\hat{\xi} = -\hat{\theta} \log \left[\frac{1}{n} \sum_{i=1}^n e^{-x_i/\hat{\theta}} \right]; \quad (3.7.2)$$

Sin embargo, estas ecuaciones no admiten una solución analíticamente cerrada para obtener los estimadores de máxima verosimilitud.

De la expresión (3.7.2), la expresión (3.7.1) se puede reescribir de la siguiente manera:

$$\begin{aligned} \hat{\theta} &= \bar{x} - \frac{1}{n} \sum_{i=1}^n x_i e^{-(x_i - \hat{\xi})/\hat{\theta}} \\ &= \bar{x} - \frac{1}{n} \sum_{i=1}^n x_i \log \hat{F}_X(x_i) \end{aligned} \quad (3.7.3)$$

donde $\hat{F}_X(x_i)$ es un estimador de la función de distribución Gumbel. Reemplazando $\log \hat{F}_X(x_i)$ en (3.7.3) por el valor esperado de $\log \hat{F}_X(x_i)$, Kimball (1956) deriva el siguiente estimador lineal para θ :

$$\hat{\theta} = \bar{x} - \frac{1}{n} \sum_{i=1}^n x_{(i)} \sum_{j=i}^n \frac{1}{j} \quad (3.7.4)$$

Note que el estimador en (3.7.4) es una función lineal de las estadísticas de orden, y por lo tanto su sesgo y el error cuadrado medio se pueden determinar fácilmente.

Prueba de razón de varianzas para la distribución Gumbel

4.1. Planteamiento del problema

Sea X_1, X_2, \dots, X_n una muestra aleatoria de una distribución F , con función de densidad $f(x; \cdot)$ con soporte en R y media finita.

Se tiene interés en probar el siguiente juego de hipótesis:

$$\begin{aligned} H_0 & : \text{ Las observaciones se ajustan a una distribución Gumbel.} \\ & \textit{vs} \\ H_1 & : \text{ Las observaciones no se ajustan a una distribución Gumbel.} \end{aligned} \tag{4.1.1}$$

4.2. Construcción de la estadística y prueba de hipótesis

La prueba de Shapiro y Wilk (1965) es la prueba clásica para probar normalidad univariada. La estadística de prueba es la razón del cuadrado de una combinación lineal de las estadísticas de orden entre la varianza muestral. Varios estudios de potencia dan evidencia de que esta prueba tiene buena potencia cuando se consideran tamaños de muestra pequeños (Gan y Koehler, 1990 y Seier, 2002).

La estadística de Shapiro-Wilk esta construida mediante una razón de la estimación del parámetro de escala, en la parte del denominador utiliza la estimación habitual S^2 , es decir, la varianza muestral. En la parte del numerador, utiliza un estimador de la varianza en términos de la estadísticas de orden, dada una muestra aleatoria.

Note que, la varianza de una distribución Gumbel esta definida sólo en términos del parámetro de escala, es decir, se encontró un estimador para el parámetro de escala θ de la distribución Gumbel en términos de las estadísticas de orden.

Utilizando las ideas de Shapiro-Wilk y el estimador de Kimball (1956), se construye la estadística de prueba V de la siguiente manera:

$$V_n = \frac{\frac{\pi^2}{6} \left(\bar{x} - \frac{1}{n} \sum_{i=1}^n x_{(i)} \sum_{j=i}^n \frac{1}{j} \right)^2}{S^2} \quad (4.2.1)$$

La prueba de tamaño $\alpha \in (0, 1)$ que se propone es dada por la siguiente regla de decisión:

Rechazar H_0 si:

$$V_n < C_{\alpha/2} \quad \text{o} \quad V_n > C_{1-\alpha/2}$$

donde V_n es la estadística de prueba dada en (4.2.1) y las constantes críticas $C_{\alpha/2}$ y $C_{1-\alpha/2}$ son obtenidas por simulación de Monte Carlo, ya que la estadística V_n no tiene una expresión cerrada conocida por una distribución además de ser invariante a cambios de escala y localidad.

4.3. Distribución del estadístico de prueba

La distribución de la estadística de prueba es difícil de obtener en forma analítica, razón por la cual se utiliza simulación Monte Carlo. La distribución de V_n puede obtenerse usando el procedimiento siguiente:

1. Se simula una muestra aleatoria m de la Distribución Gumbel con parámetros $\xi = 0$ y $\theta = 1$ (`library(evd)`).
2. Se calcula la estadística V_n , antes ordenar m de forma ascendente para obtener las estadísticas de orden y también calcular S^2 .
3. Se repite los pasos 1 y 2 B veces.

En la Figura 4.1 se muestra la distribución de la estadística de prueba V_n para distintos tamaños de muestra $n = 10, 30, 50, 100$ y 200 , $B = 20,000$ y para diferentes valores de los parámetros ξ y θ .

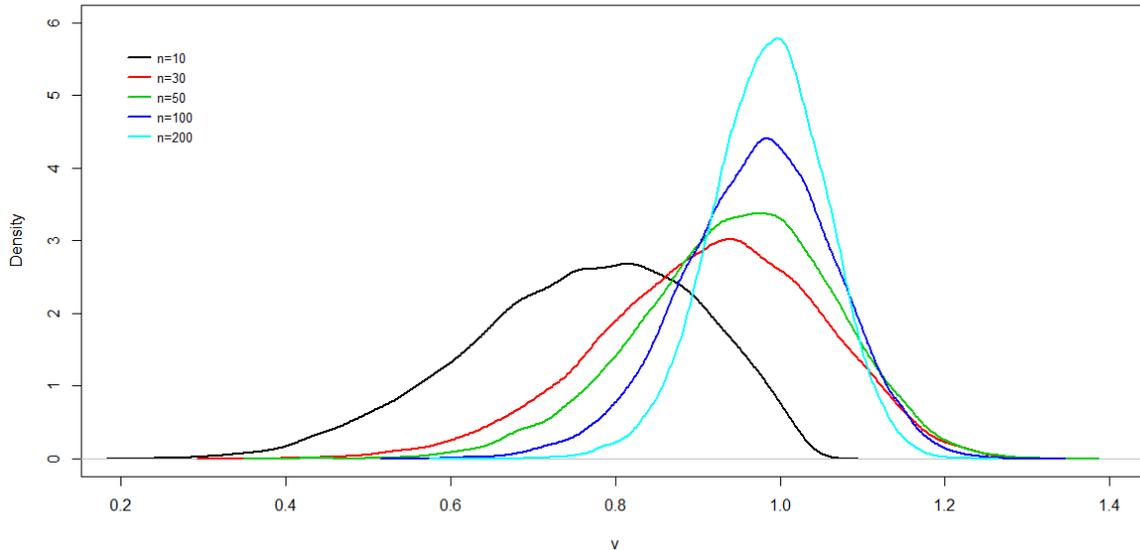


Figura 4.1: Distribución de la estadística V_n para varios tamaños de muestra.

En las Figuras 4.2–4.6 se muestra la distribución de la estadística de prueba V_n donde se varía el parámetro de localidad $\xi = \{-10, 0, 10\}$ (izquierda) y de escala $\theta = \{1, 10, 100\}$ (derecha), en ambos casos se puede ver que la distribución de la estadística de prueba no se ve afectada tanto en localidad como en escala sin importar la variación de los parámetros de localidad y escala y se induce que la estadística es invariante a localidad y escala, todo esto para distintos tamaños de muestra $n = 10, 30, 50, 100$ y 200 .

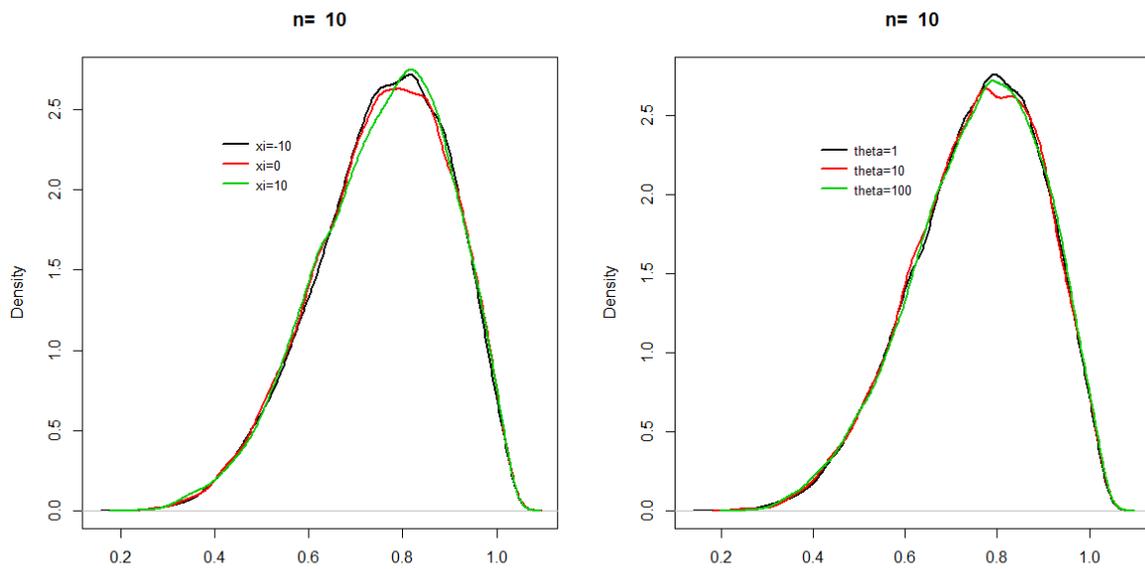


Figura 4.2: Distribución de la estadística V_n cuando $n = 10$ variando los valores del parámetro de localización ξ (izquierda) y escala θ (derecha), respectivamente.

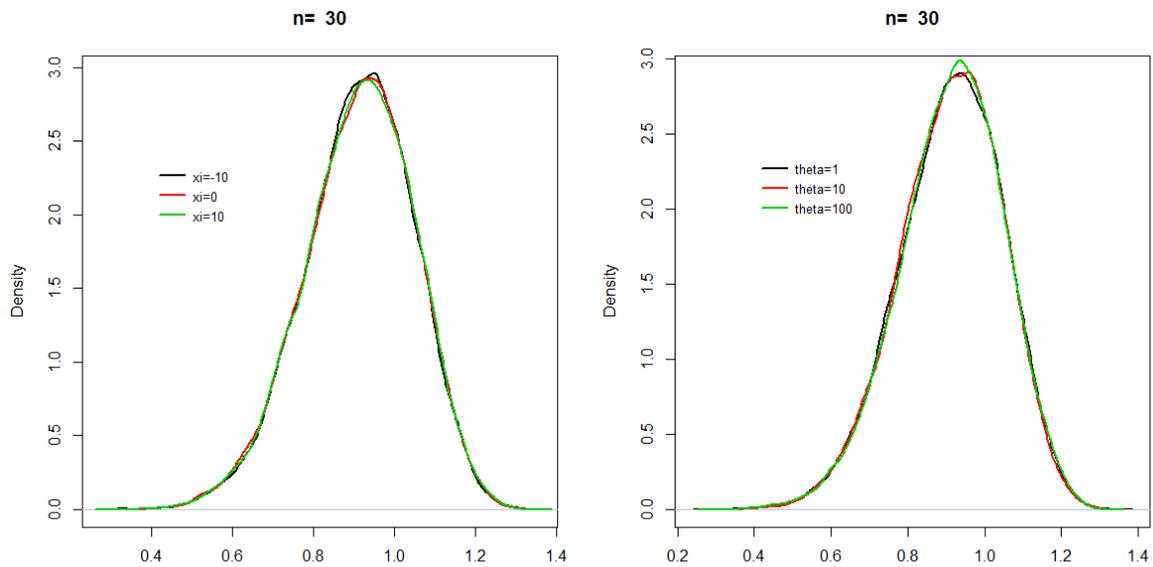


Figura 4.3: Distribución de la estadística V_n cuando $n = 30$ variando los valores del parámetro de localización ξ (izquierda) y escala θ (derecha), respectivamente.

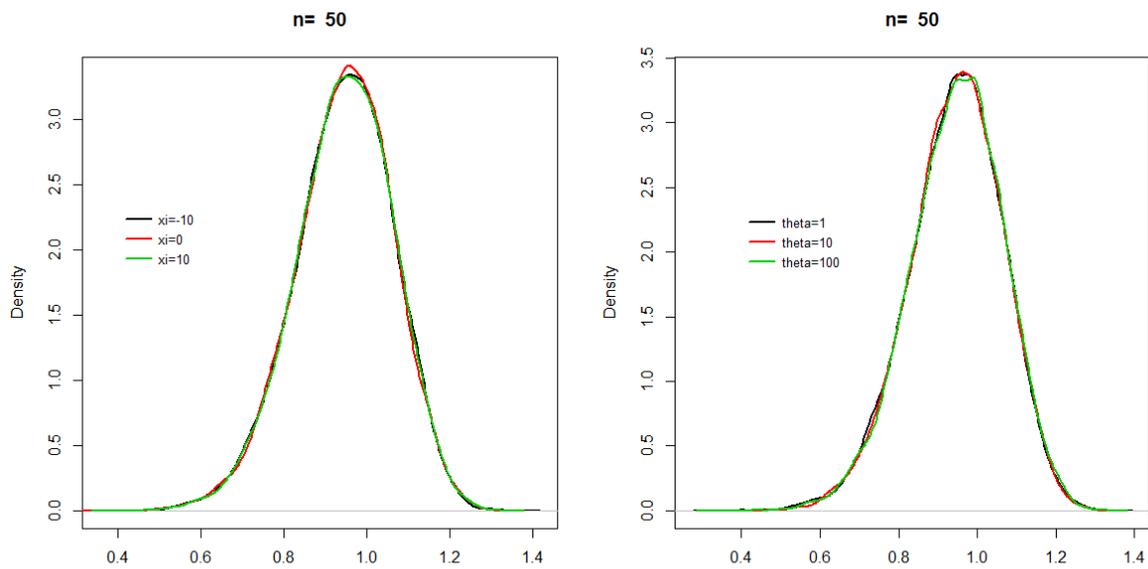


Figura 4.4: Distribución de la estadística V_n^v cuando $n = 50$ variando los valores del parámetro de localización ξ (izquierda) y escala θ (derecha), respectivamente.

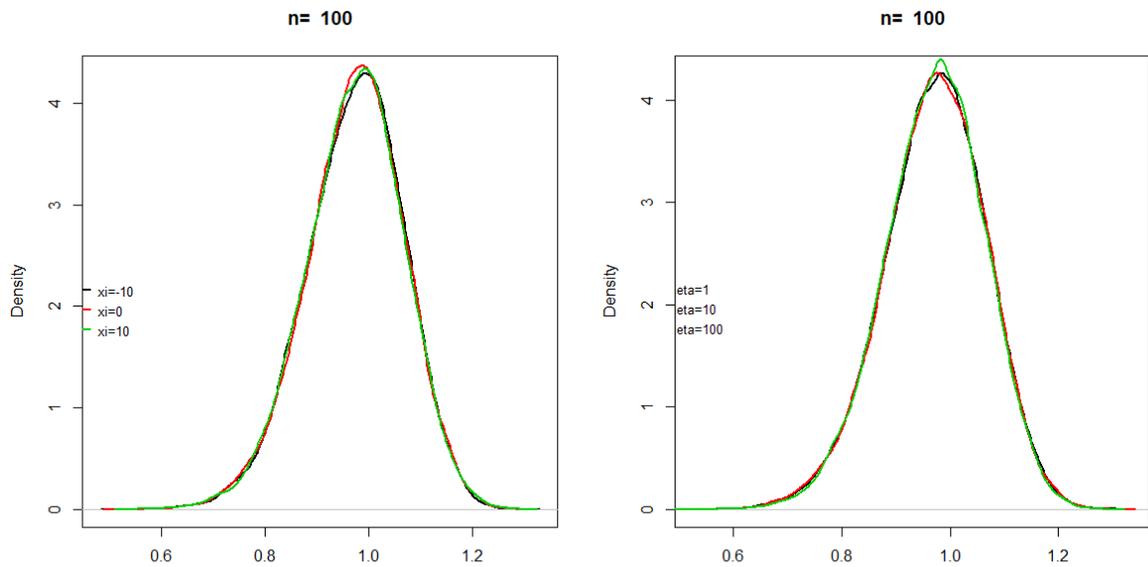


Figura 4.5: Distribución de la estadística V_n^v cuando $n = 100$ variando los valores del parámetro de localización ξ (izquierda) y escala θ (derecha), respectivamente.

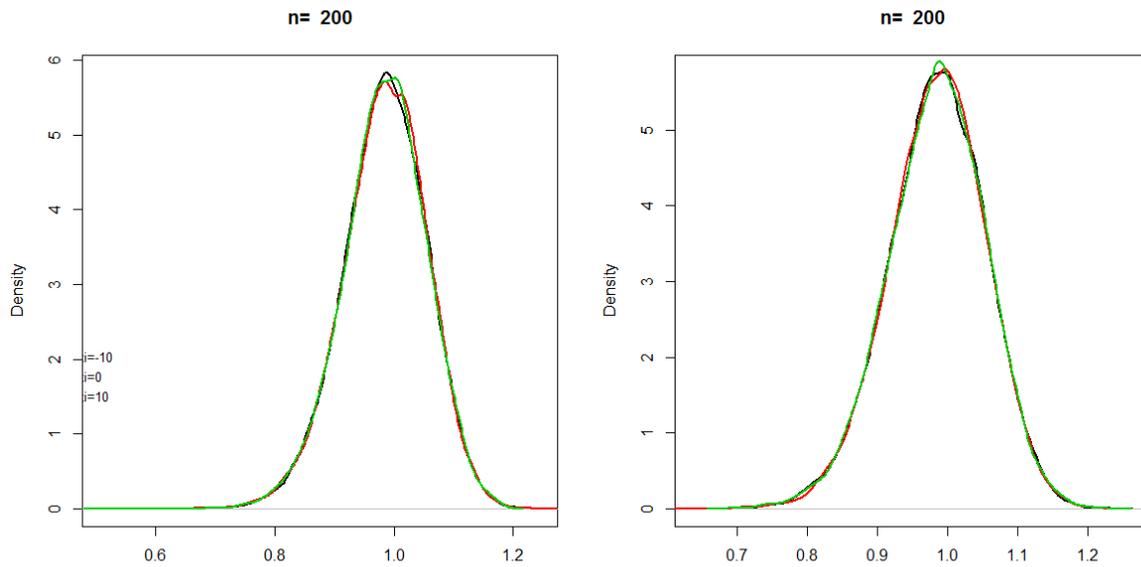


Figura 4.6: Distribución de la estadística V_n cuando $n = 200$ variando los valores del parámetro de localidad ξ (izquierda) y escala θ (derecha), respectivamente.

4.4. Constantes Críticas

Debido a que la distribución de V_n no tiene forma analítica y es invariante a transformaciones de localidad y escala, los valores críticos V_n fueron obtenidos mediante simulación Monte Carlo. Los niveles de significancia utilizados fueron $\alpha = 0.01, 0.05, 0.10$ y 0.25 .

Se generaron muestras aleatorias de la distribución Gumbel estándar para $n = 10, 30, 50, 100$ y 200 con $B = 20,000$. El valor de V_n fue calculado para cada tamaño de muestra y nivel de significancia. Los valores críticos fueron calculados para cada n con los cuantiles $\alpha/2 \times 100\%$ y $(1 - \alpha/2) \times 100\%$ de la distribución empírica V_n . En la Tabla 4.1 se presentan los valores críticos obtenidos mediante el proceso de simulación ya descrito para diferentes tamaños de muestra y diferentes tamaños de prueba. El programa que implementa el algoritmo está en el Anexo II del trabajo.

Para el cálculo de la constante crítica de la estadística V_n , el algoritmo se describe a continuación:

- 1) Se simula una muestra aleatoria m de la Distribución Gumbel con parámetros $\xi = 0$ y $\theta = 1$ (`library(evd)`).
- 2) Se calcula la estadística V_n , antes ordenar m de forma ascendente para obtener las estadísticas de orden y también calcular S^2 .

- 3) Se repiten los pasos 1) y 2) B veces.
- 4) Se obtienen los cuantiles de V_n , $C_{\alpha/2}$ y $C_{1-\alpha/2}$, respectivamente.

| n | 0.01 | | 0.05 | | 0.1 | | 0.25 | |
|------------|-------------|-------------|-------------|-------------|------------|------------|-------------|-------------|
| | $C_{0.005}$ | $C_{0.995}$ | $C_{0.025}$ | $C_{0.975}$ | $C_{0.05}$ | $C_{0.95}$ | $C_{0.125}$ | $C_{0.875}$ |
| 10 | 0.367 | 1.018 | 0.457 | 0.989 | 0.507 | 0.967 | 0.585 | 0.921 |
| 30 | 0.537 | 1.213 | 0.629 | 1.159 | 0.676 | 1.122 | 0.754 | 1.064 |
| 50 | 0.608 | 1.219 | 0.702 | 1.164 | 0.742 | 1.133 | 0.810 | 1.078 |
| 100 | 0.701 | 1.191 | 0.783 | 1.142 | 0.811 | 1.118 | 0.865 | 1.079 |
| 200 | 0.788 | 1.151 | 0.841 | 1.117 | 0.866 | 1.096 | 0.906 | 1.064 |

Tabla 4.1: Constantes críticas para distintos tamaños de muestra $n = 10, 30, 50, 100$ y 200 con niveles de significancia $\alpha = 0.01, 0.05, 0.1$ y 0.25 .

Estudio de Simulación Monte Carlo para estimar y comparar la potencia y tamaño de la prueba, contra algunas distribuciones alternativas

5.1. Tamaño de la Prueba

Para la estimación del tamaño de la prueba V_n , se usaron 10,000 muestras Monte Carlo. Se consideró un tamaño de prueba $\alpha = 0.05$ y los tamaños de muestra $n = 10, 30, 50, 100$ y 200 .

Las computaciones se hicieron en el lenguaje R Development (2015). El algoritmo se describe a continuación:

- 1) Se simula una muestra aleatoria m de la Distribución Gumbel con parámetros ξ y θ .
- 2) Se calcula la estadística V_n , antes ordenar m de forma ascendente para obtener las estadísticas de orden y también calcular S^2 .
- 3) Se compara $V_n < C_{\alpha/2}$ y $V_n > C_{1-\alpha/2}$, si se cumple al menos una de las condiciones se rechaza la hipótesis nula, se guarda la decisión en un vector binario con el valor de 1, 0 de otro modo.
- 4) Se repiten los pasos 1) al 3) B veces.

- 5) Para obtener el tamaño de la prueba, se suma el vector binario del paso 3) y se divide entre B .

5.1.1. Discusión de los resultados de simulación

En la Figura 5.1 se muestran los resultados de la simulación para el tamaño de la prueba con el algoritmo anteriormente descrito. En ella se puede apreciar que la prueba mantiene su tamaño de prueba variando el tamaño de muestra igual a $n = 10, 30, 50, 100$ y 200 .

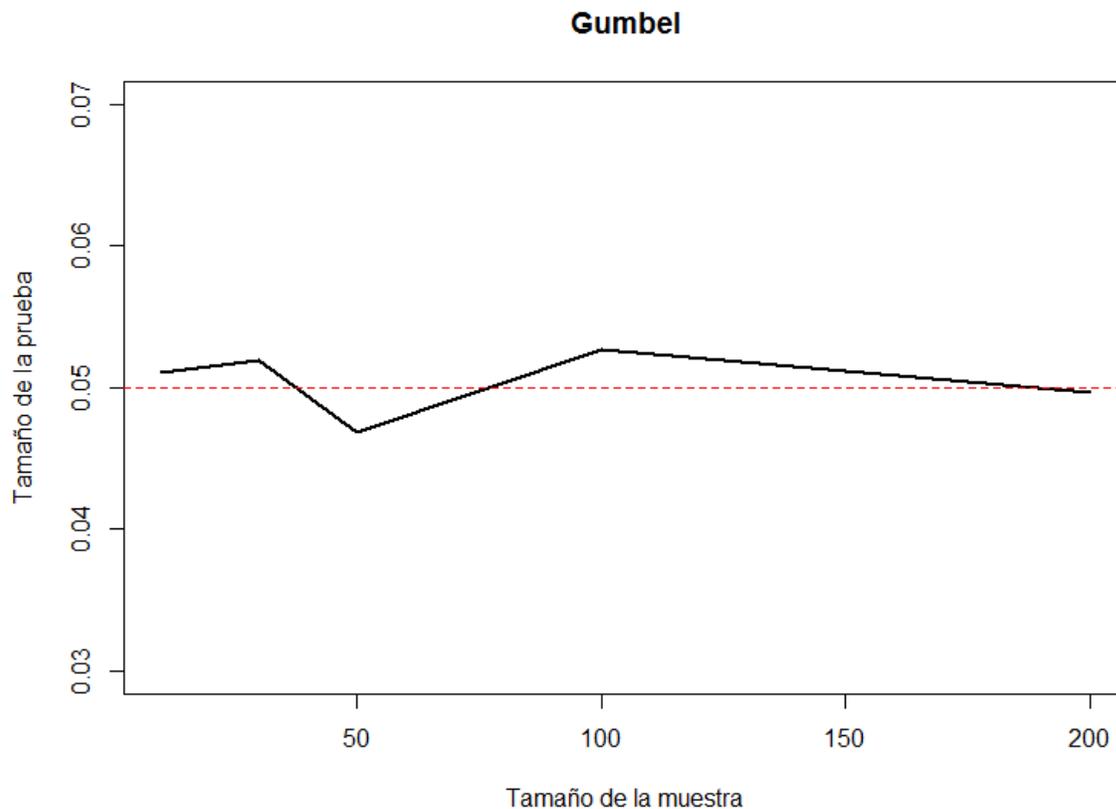


Figura 5.1: *Tamaño estimado de la prueba variando los tamaños de muestra iguales a $n = 10, 30, 50, 100$ y 200 , con $\alpha = 0.05$.*

5.2. Potencia de la Prueba

La potencia de la prueba se calculó bajo las distribuciones alternativas: Normal, Logística, Cauchy, T-Student y Valores Extremos Generalizada con función de distribución denotada en **3.3.1**, donde $1 + \kappa(x - \xi)/\theta > 0$ y $\kappa \in \mathfrak{R}$.

Para la estimación de la potencia de la prueba V_n , se usaron 20,000 muestras Monte Carlo.

Se consideró un nivel de significancia $\alpha = 0.05$ y los tamaños de muestra $n = 10, 30, 50, 100$ y 200 .

Las computaciones se hicieron en el lenguaje R Development (2015). El algoritmo se describe a continuación:

- 1) Se simula una muestra aleatoria m de la Distribución Alternativa.
- 2) Se calcula la estadística V_n , antes ordenar m de forma ascendente para obtener las estadísticas de orden y también calcular S^2 .
- 3) Se compara $V_n < C_{\alpha/2}$ y $V_n > C_{1-\alpha/2}$, si se cumple al menos una de las condiciones se rechaza la hipótesis nula, se guarda la decisión en un vector binario con el valor de 1, 0 de otro modo.
- 4) Se repiten los pasos 1) al 3) B veces.
- 5) Para obtener la potencia de la prueba, se suma el vector binario del paso 3) y se divide entre B .

5.2.1. Discusión de los resultados de simulación

En las siguientes figuras se muestran los resultados de la simulación para la potencia de la prueba con el algoritmo anteriormente descrito.

En la Figura 5.2 se muestran las potencias de la prueba para las distribuciones alternativas: Normal, Logística y Cauchy variando el tamaño de muestra $n = 10, 30, 50, 100$ y 200 , en el cual podemos observar lo siguiente:

- Cuando el tamaño de muestra aumenta la potencia de la prueba para las distribuciones alternativas Normal, Logística y Cauchy aumentan.
- Para tamaños de muestra menores a 30 la potencia de la prueba contra la Distribución Cauchy fue superior a la mostrada por las distribuciones Normal y Logística.
- Cuando $n = 50$ las potencias de las tres distribuciones alternativas fueron similares, no hubo tanta variación, dichas probabilidades oscilaron entre 0.6 al 0.7.
- Cuando el tamaño de muestra fue $n = 100$ las potencias de la distribución Normal fueron las de mayor probabilidad, superiores a 0.95.

- Cuando $n = 200$ las potencias de las distribuciones Normal y Logística son prácticamente iguales, sin embargo, la distribución alternativa Normal tuvo mayor potencia.

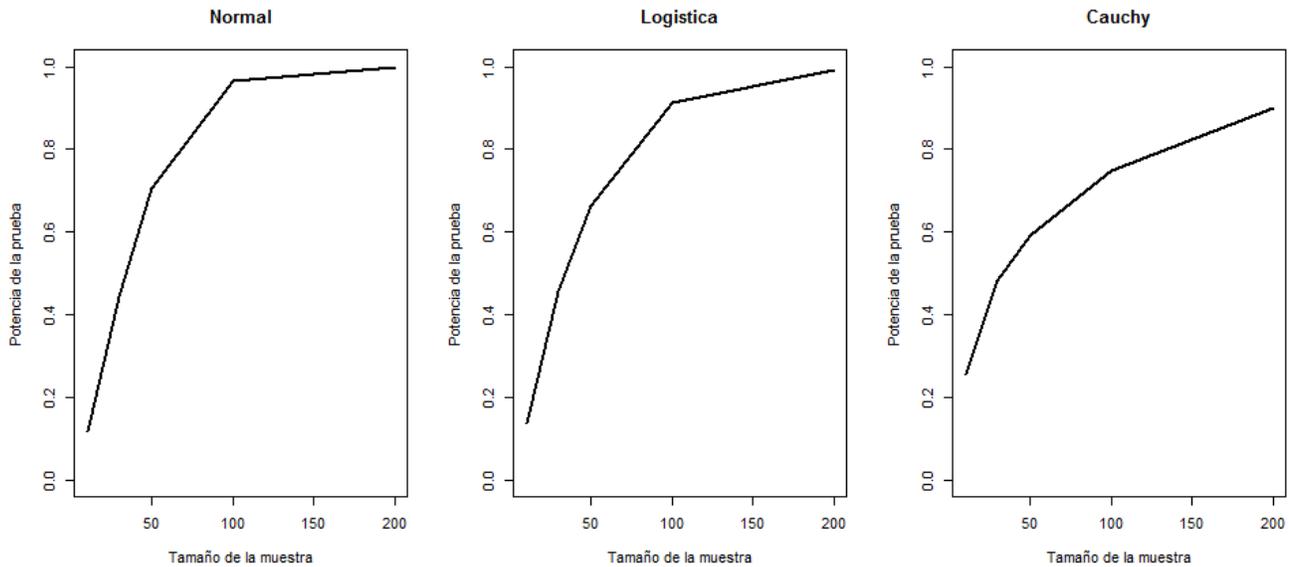


Figura 5.2: Potencia estimada de la prueba contra las distribuciones Normal, Logística y Cauchy variando los tamaños de muestra $n = 10, 30, 50, 100$ y 200 , con $\alpha = 0.05$.

En la Figura 5.3 se muestran las potencias de la prueba para la distribución alternativa T-Student con distintos tamaños de muestra $n = 10, 30, 50, 100$ y 200 variando los grados de libertad en el cual podemos observar lo siguiente:

- Cuando el tamaño de muestra aumenta la potencia de la prueba también aumenta.
- Cuando $n = 10$ la potencia de la prueba disminuye ligeramente cuando aumentan los grados de libertad.
- Cuando $n = 200$ la potencia de la prueba aumenta notoriamente cuando los grados de libertad son más de 5.

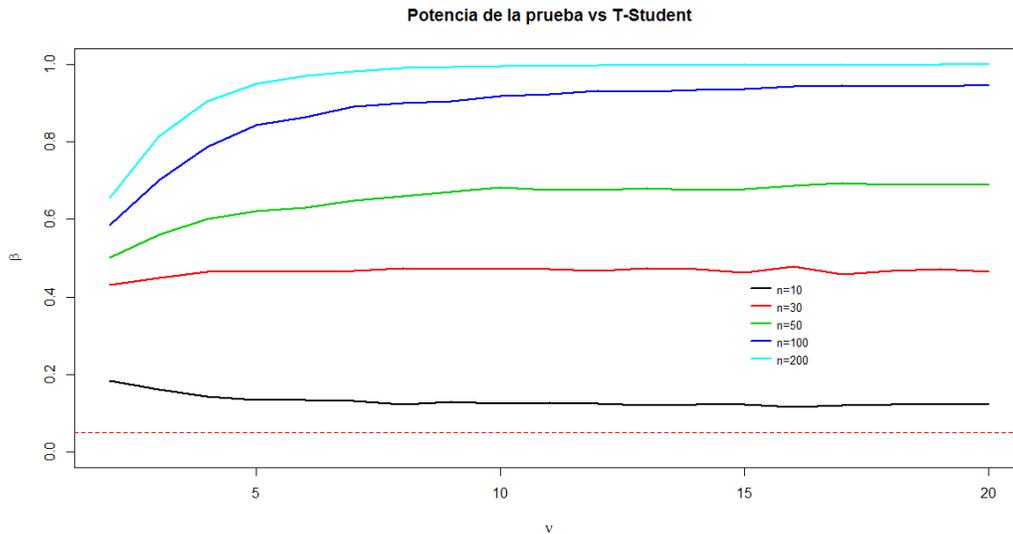


Figura 5.3: Potencia estimada de la prueba contra la distribución t -student variando los grados de libertad $\nu \in (1, 20)$, para distintos tamaños de muestra y $\alpha = 0.05$.

En la Figura 5.4 se muestran las potencias de la prueba para la distribución alternativa de Valores Extremos Generalizada (GEV) con distintos tamaños de muestra $n = 10, 30, 50, 100$ y 200 variando el parámetro de forma κ en el cual podemos observar lo siguiente:

- Cuando la distribución de Valores Extremos Generalizada toma valores en los intervalos $(-\infty, -1)$ y $(1, \infty)$, la potencia de la prueba es prácticamente igual a 1 cuando el tamaño de la muestra es mayor a 30.
- Cuando el parámetro de forma κ es igual a 0, la potencia de la prueba es aproximadamente igual a 0.05, es decir, en ese punto se convierte en el tamaño de la prueba ya que cuando $\kappa = 0$ la distribución GEV se convierte en una distribución Gumbel (ξ, θ) . La distribución de la hipótesis alternativa es la misma que la distribución de la hipótesis nula. Lo anterior funciona para cualquier tamaño de muestra.
- cuando el tamaño de la muestra es igual a 10, la potencia de la prueba aumenta poco comparado con otros tamaños de muestra más grandes, en el gráfico muestra que la potencia más alta la consigue cuando $\kappa = 1$ donde la potencia es superior a 0.6.

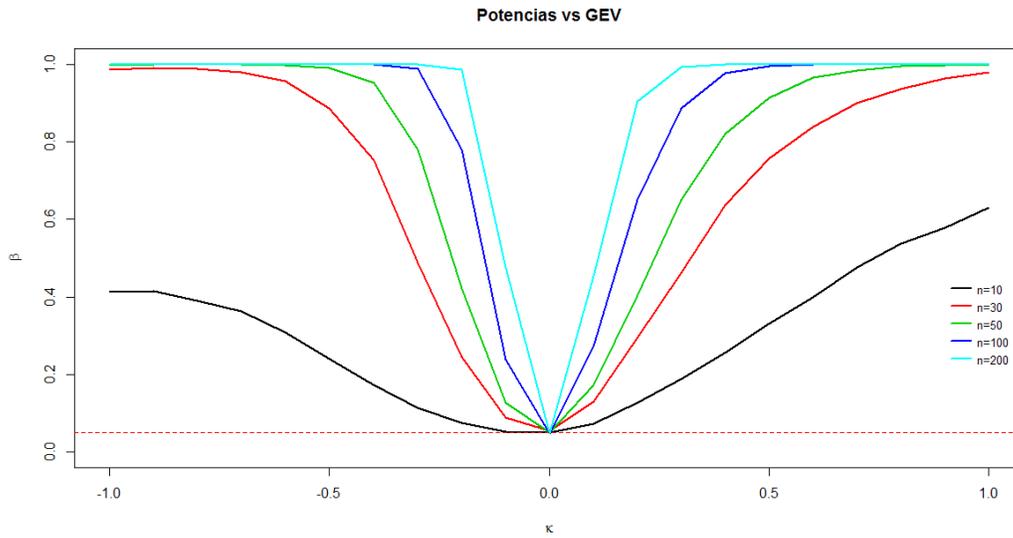


Figura 5.4: Potencia estimada de la prueba contra la Distribución de Valores Extremos Generalizada ($\xi = 0, \theta = 1$) variando el parámetro de forma, para distintos tamaños de muestra y $\alpha = 0.05$.

5.3. Estudio comparativo de la prueba V_n con la Prueba de Anderson-Darling

Se comparó la prueba propuesta V_n contra la prueba de Anderson-Darling, la que a su vez, mediante literatura revisada es muy potente para probar muestras aleatorias que provengan de una distribución Gumbel. Las distribuciones alternativas a considerar fueron: Normal, Logística, Cauchy, T-Student y Valores Extremos Generalizada con función de distribución denotada en **3.3.1**, donde $1 + \kappa(x - \xi)/\theta > 0$ y $\kappa \in \mathfrak{R}$.

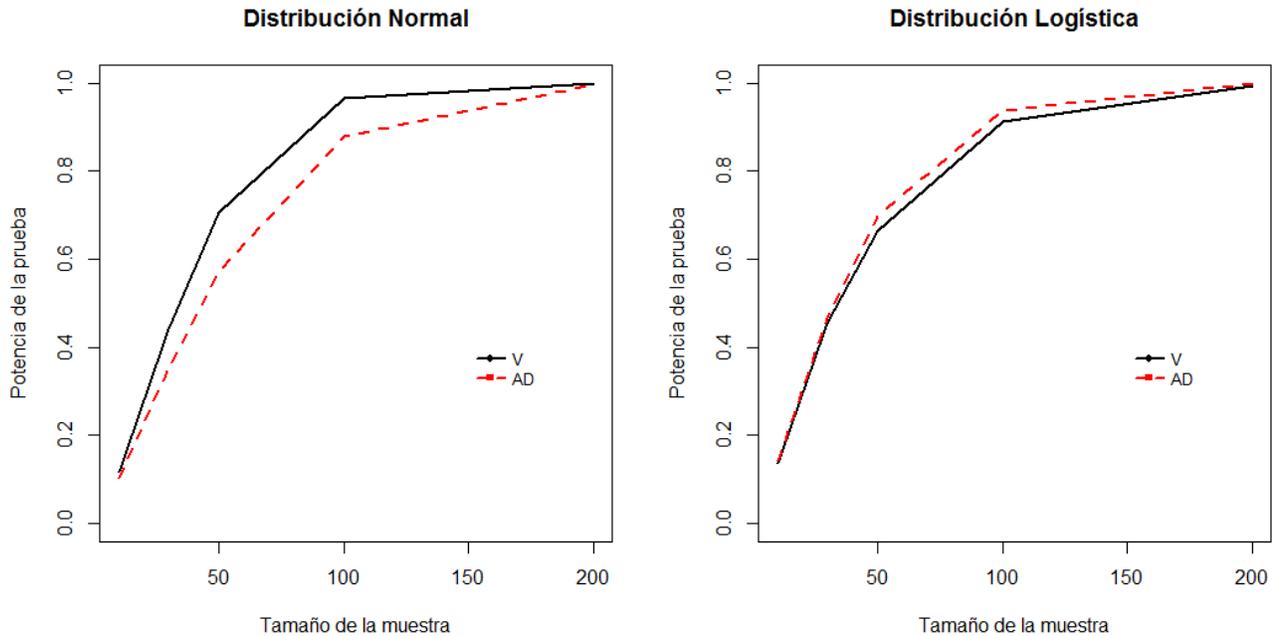


Figura 5.5: Comparación de la potencia de la prueba V_n y Anderson-Darling para las distribuciones Normal y Logística variando los tamaños de muestra $n = 10, 30, 50, 100$ y 200 , con $\alpha = 0.05$.

En la Figura 5.5 se muestran las potencias de las pruebas V_n y Anderson-Darling para las distribuciones alternativas Normal y Logística con tamaños de muestra $n = 10, 30, 50, 100$ y 200 en el cual podemos observar lo siguiente:

- En el caso de la distribución Normal, se puede ver que las potencias de la prueba V_n fueron más altas en los distintos tamaños de muestra a los obtenidos por la prueba de Anderson Darling.
- En el caso de la distribución Logística, se observa que las potencias de ambas pruebas fueron muy similares para cada uno de los distintos tamaños de muestra, ligeramente la prueba de Anderson Darling obtuvo mayor Potencia.

En la Figura 5.6 se muestran las potencias de las pruebas V_n y Anderson-Darling para la distribución alternativa T-Student variando sus grados de libertad con distintos tamaños de muestra $n = 10, 30, 50, 100$ y 200 en el cual podemos observar lo siguiente:

- La potencia de la prueba V_n aumenta conforme aumentan los grados de libertad de la distribución T-Student.

- La potencia de la prueba Anderson - Darling disminuye conforme aumentan los grados de libertad de la distribución T-Student.
- Cuando los grados de libertad son menores a 10, la distribución de Anderson - Darling tiene mayor potencia comparada a la prueba V_n .
- Cuando los grados de libertad son mayores a 10, la distribución de Anderson - Darling tiene menor potencia comparada a la prueba V_n .

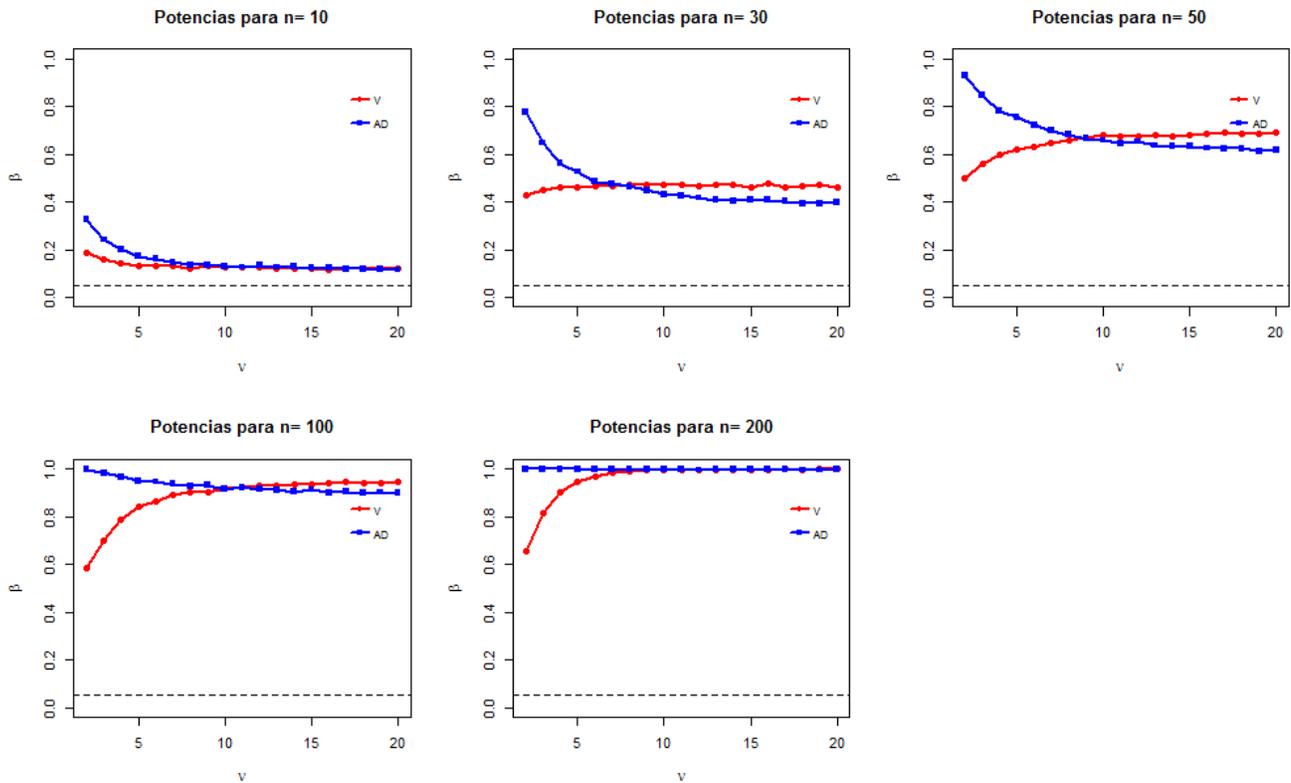


Figura 5.6: Comparación de la potencia de la prueba V_n y Anderson-Darling para la distribución T-Student variando los grados de libertad $\nu = 2, 3, 4, 5, 6, 7, 8, 9, \dots, 20$, para tamaños de muestra $n = 10, 30, 50, 100$ y 200 .

En la Figura 5.7 se muestran las potencias de las pruebas V_n y Anderson-Darling para la distribución alternativa de Valores Extremos Generalizada variando los valores del parámetro de forma κ con distintos tamaños de muestra $n = 10, 30, 50, 100$ y 200 en el cual podemos observar lo siguiente:

- Se consideraron valores de κ entre -1 y 0.5 , posterior a este último valor el algoritmo para la distribución Anderson - Darling no logra obtener un valor de potencia.
- Cuando el tamaño de la muestra es igual a 10 las potencias de ambas pruebas es variante, es decir, en ciertos valores de κ una prueba es más potente que la otra y viceversa.
- Para tamaños de muestra $n = 30, 50, 100$ y 200 la prueba V_n resultó ser más potente comparada a la prueba de Anderson - Darling, estas potencias se hace más distantes cuando el parámetro de forma es cercano a 0.

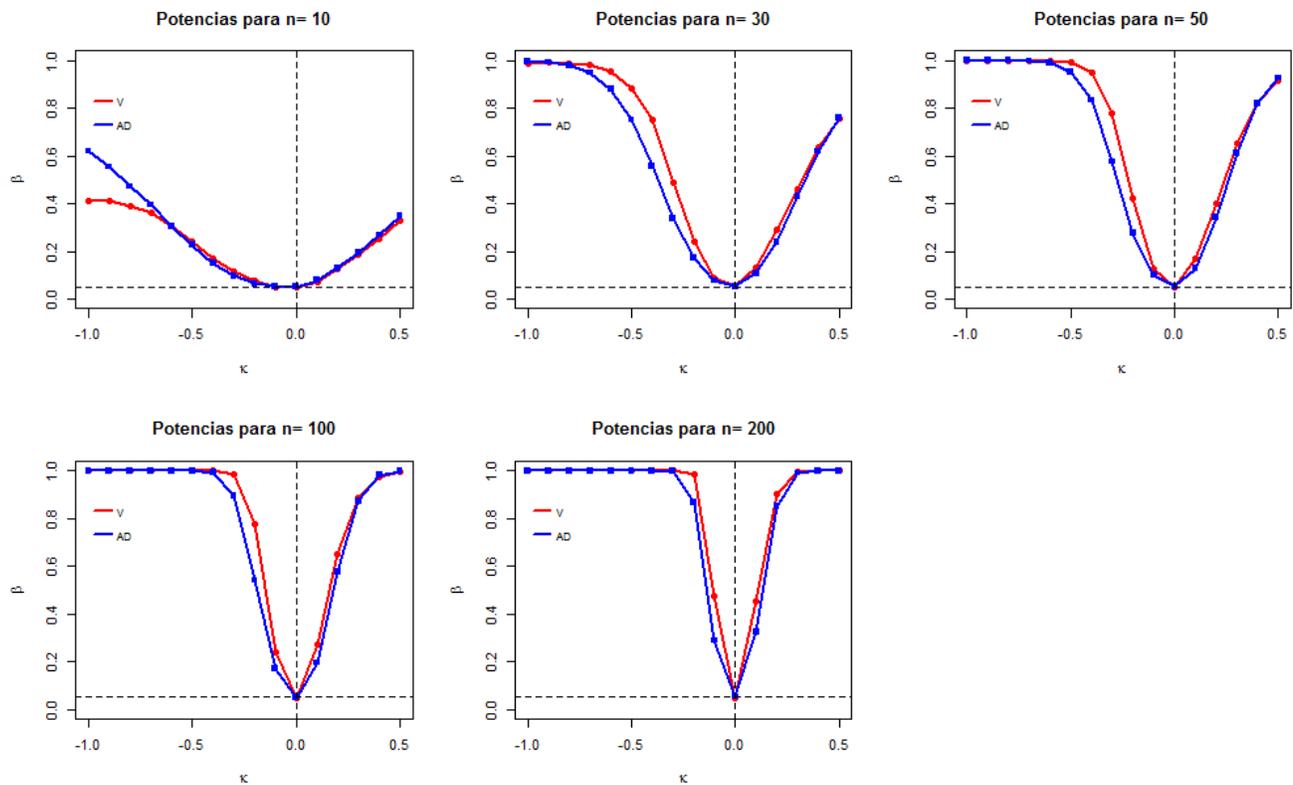


Figura 5.7: Comparación de la potencia de la prueba V_n y Anderson-Darling para la distribución de Valores Extremos variando el parámetro forma, para tamaños de muestra $n = 10, 30, 50, 100$ y 200 .

Ejemplos de aplicación de la prueba propuesta

En este capítulo se presentan dos ejemplos en los que se ilustran los pasos que hay que seguir para aplicar la prueba de bondad de ajuste descrita en el Capítulo 4.

6.1. Ejemplo 1

6.1.1. Datos de Ozono de la Ciudad de México

La razón para el estudio de la capa de ozono es principalmente porque se ha observado que altos niveles de ozono son en general, seguido por altos niveles de otros contaminantes. Es particularmente importante estudiar los niveles de ozono por encima de 0.11 partes por millón (ppm) por unidad de volumen porque son considerados peligrosos para la salud humana según la Agencia de Protección Ambiental de los Estados Unidos.

Los datos que aquí se consideran son los cluster con excedencia del umbral propuesto igual a $u = 0.11$ ppm, resultantes de los niveles diarios de ozono correspondientes a las estaciones en verano de los años 2003 a 2007 y proceden de la estación de monitoreo de la calidad del aire Pedregal en la Ciudad de México. El número total de cluster con dichas excedencias es 78.

En la Tabla 6.1 se muestran los datos de los cluster de ozono y tamaños de cada cluster que excedieron el umbral $u = 0.11$ ppm.

| Cluster | Fecha | M | N | Cluster | Fecha | M | N | Cluster | Fecha | M | N |
|-----------|------------|-------|----|-----------|------------|-------|---|-----------|------------|-------|---|
| 1 | 06/23/2003 | 0.025 | 1 | 27 | 09/03/2004 | 0.025 | 2 | 53 | 07/18/2006 | 0.009 | 2 |
| 2 | 07/03/2003 | 0.045 | 3 | 28 | 09/05/2004 | 0.005 | 1 | 54 | 07/28/2006 | 0.015 | 1 |
| 3 | 07/05/2003 | 0.014 | 3 | 29 | 09/19/2004 | 0.01 | 1 | 55 | 07/31/2006 | 0.04 | 3 |
| 4 | 07/09/2003 | 0.037 | 4 | 30 | 07/03/2005 | 0.017 | 2 | 56 | 08/13/2006 | 0.047 | 5 |
| 5 | 07/15/2003 | 0.039 | 2 | 31 | 07/06/2005 | 0.003 | 1 | 57 | 08/18/2006 | 0.024 | 3 |
| 6 | 07/19/2003 | 0.008 | 1 | 32 | 07/09/2005 | 0.05 | 3 | 58 | 08/22/2006 | 0.043 | 1 |
| 7 | 07/23/2003 | 0.071 | 5 | 33 | 07/12/2005 | 0.071 | 1 | 59 | 08/25/2006 | 0.029 | 3 |
| 8 | 07/31/2003 | 0.07 | 4 | 34 | 07/14/2005 | 0.038 | 1 | 60 | 09/02/2006 | 0.026 | 1 |
| 9 | 08/08/2003 | 0.055 | 11 | 35 | 07/20/2005 | 0.073 | 3 | 61 | 09/06/2006 | 0.02 | 1 |
| 10 | 08/18/2003 | 0.001 | 1 | 36 | 07/29/2005 | 0.047 | 3 | 62 | 09/08/2006 | 0.05 | 4 |
| 11 | 08/26/2003 | 0.016 | 1 | 37 | 08/03/2005 | 0.051 | 3 | 63 | 09/17/2006 | 0.047 | 2 |
| 12 | 08/28/2003 | 0.039 | 3 | 38 | 08/07/2005 | 0.037 | 1 | 64 | 06/23/2007 | 0.084 | 6 |
| 13 | 09/04/2003 | 0.037 | 2 | 39 | 08/12/2005 | 0.064 | 4 | 65 | 06/30/2007 | 0.002 | 1 |
| 14 | 09/07/2003 | 0.036 | 2 | 40 | 08/15/2005 | 0.009 | 1 | 66 | 07/02/2007 | 0.071 | 3 |
| 15 | 09/09/2003 | 0.021 | 1 | 41 | 08/27/2005 | 0.013 | 1 | 67 | 07/07/2007 | 0.05 | 5 |
| 16 | 09/12/2003 | 0.085 | 3 | 42 | 08/29/2005 | 0.047 | 1 | 68 | 07/14/2007 | 0.022 | 2 |
| 17 | 06/23/2004 | 0.028 | 3 | 43 | 08/31/2005 | 0.029 | 1 | 69 | 07/19/2007 | 0.006 | 1 |
| 18 | 06/25/2004 | 0.039 | 1 | 44 | 09/08/2005 | 0.001 | 1 | 70 | 07/24/2007 | 0.008 | 4 |
| 19 | 07/03/2004 | 0.016 | 1 | 45 | 09/10/2005 | 0.061 | 3 | 71 | 07/27/2007 | 0.073 | 2 |
| 20 | 07/09/2004 | 0.022 | 3 | 46 | 09/14/2005 | 0.062 | 1 | 72 | 07/31/2007 | 0.04 | 5 |
| 21 | 07/16/2004 | 0.006 | 1 | 47 | 09/16/2005 | 0.012 | 1 | 73 | 08/08/2007 | 0.023 | 3 |
| 22 | 07/18/2004 | 0.016 | 1 | 48 | 09/18/2005 | 0.011 | 1 | 74 | 08/12/2007 | 0.001 | 1 |
| 23 | 07/28/2004 | 0.03 | 1 | 49 | 06/21/2006 | 0.013 | 1 | 75 | 08/20/2007 | 0.027 | 3 |
| 24 | 08/19/2004 | 0.003 | 1 | 50 | 06/27/2006 | 0.002 | 1 | 76 | 08/25/2007 | 0.02 | 1 |
| 25 | 08/23/2004 | 0.029 | 2 | 51 | 07/01/2006 | 0.041 | 6 | 77 | 09/04/2007 | 0.047 | 2 |
| 26 | 08/26/2004 | 0.013 | 1 | 52 | 07/11/2006 | 0.048 | 5 | 78 | 09/17/2007 | 0.032 | 2 |

Tabla 6.1: Cluster de ozono y tamaños de cada cluster que excedieron el umbral $u = 0.11$.

En la Figura 6.1 se observan gráficamente los datos seleccionados para realizar la prueba.

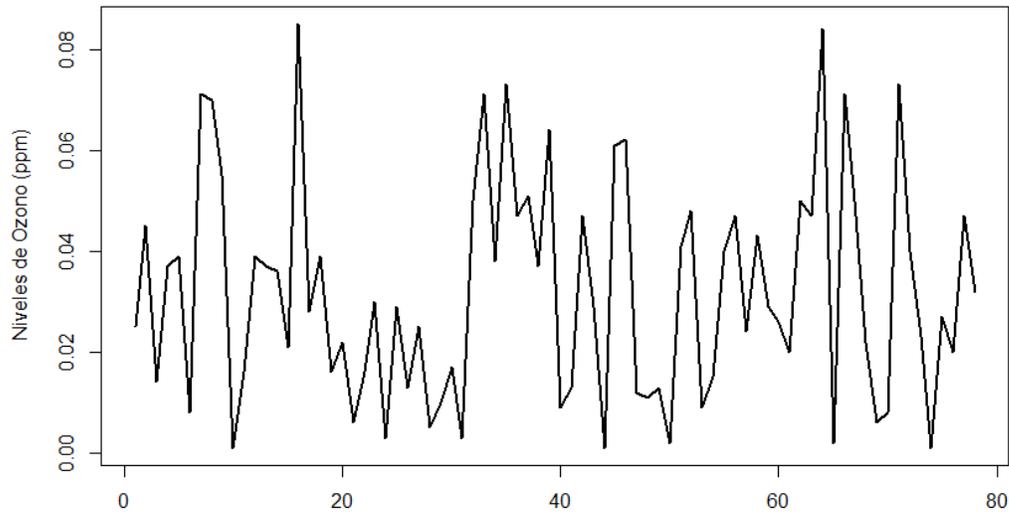


Figura 6.1: Observaciones de los cluster de ozono que excedieron el umbral $u = 0.11$, resultantes de los niveles máximos diarios de ozono correspondientes a las estaciones en verano de los años 2003 a 2007.

Ahora, hay que verificar que no exista ningún tipo de auto correlación en la serie de datos, lo cual se puede ver al estudiar la gráfica siguiente:

En la Figura 6.2 se observa que existe una baja auto correlación entre los datos (es menor en valor absoluto a 0.2), por lo que se continúa con el análisis de los mismos.

Al realizar el gráfico Q-Q de los cluster de ozono que excedieron el umbral $u = 0.11$ contra los cuantiles teóricos de una distribución Gumbel estándar ($\nu = 0$ y $\sigma = 1$), se obtiene lo siguiente:

La Figura 6.3 proporciona fuerte evidencia de que los datos pueden ser ajustados a una distribución Gumbel, ya que se observa una relación lineal entre los cluster de ozono que excedieron el umbral $u = 0.11$ contra los cuantiles teóricos,

Ahora se procede de manera rigurosa a verificar si los datos vienen de una distribución Gumbel, utilizando dos pruebas de prueba de bondad de ajuste, la primera es la Prueba de Anderson Darling y la segunda la propuesta en este escrito basada en el cociente de dos estimadores de la varianza uno de ellos en términos de las estadísticas de orden, donde el juego de hipótesis de interés es el planteado en 4.1.1.

El tamaño de muestra $n = 78$, para un nivel de significancia $\alpha = 0.05$, de la sección 4.4 se

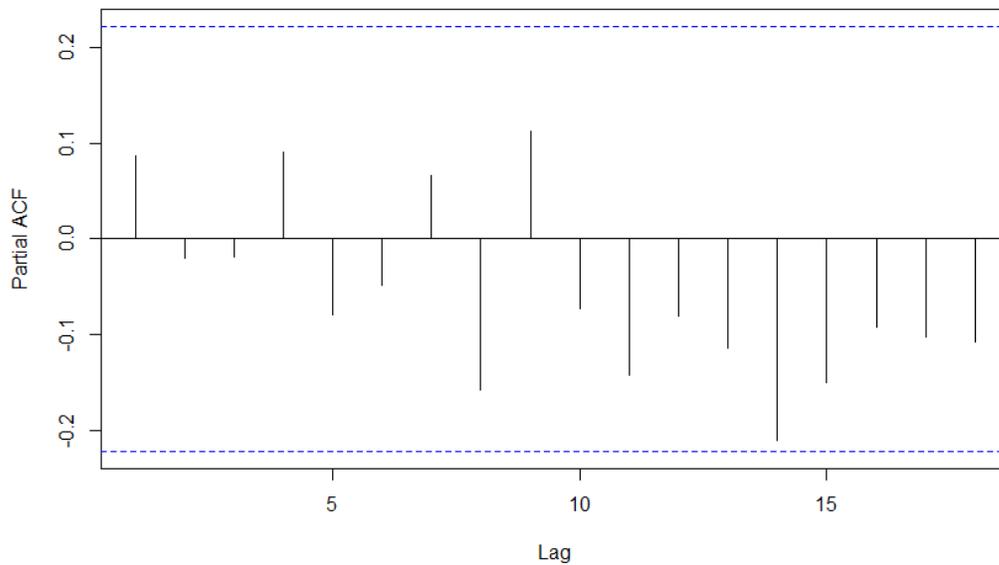


Figura 6.2: Auto correlación parcial para las observaciones de los cluster de ozono que excedieron el umbral $u = 0.11$, resultantes de los niveles máximos diarios de ozono correspondientes a las estaciones en verano de los años 2003 a 2007.

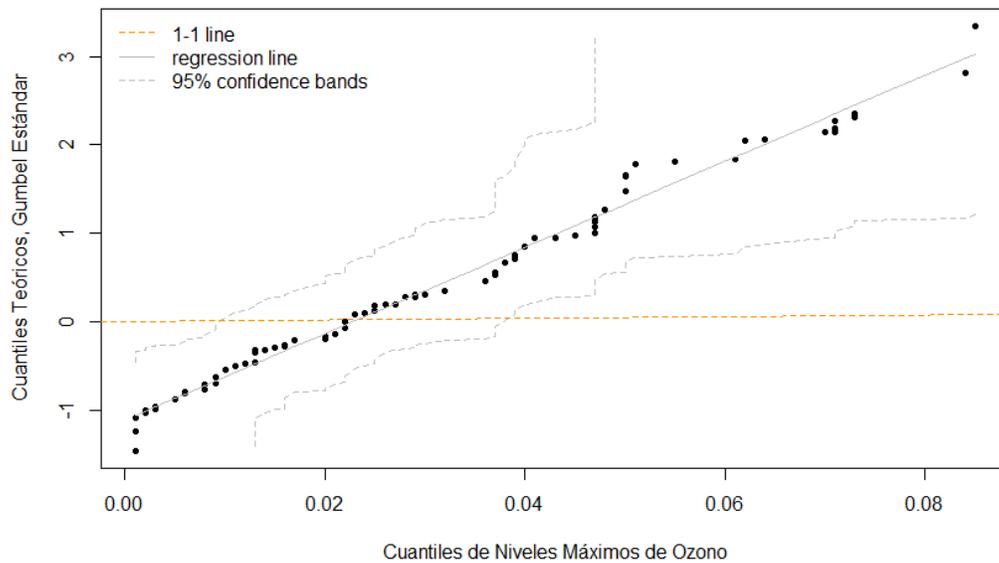


Figura 6.3: Gráfica Q-Q para los cluster de ozono que excedieron el umbral $u = 0.11$, resultantes de los niveles máximos diarios de ozono correspondientes a las estaciones en verano de los años 2003 a 2007.

calculan las constantes críticas $C_{\alpha/2}$ y $C_{1-\alpha/2}$ y el estimador propuesto en la Ecuación 4.2.1 con un nivel de $\alpha = 0.05$, obteniéndose los valores siguientes:

$$C_{\alpha/2} = 0.7496241 \quad , \quad C_{1-\alpha/2} = 1.1484416 \quad \text{y} \quad V_n = 1.023557$$

Finalmente, se utiliza la siguiente regla de decisión:

Rechazar H_0 si:

$$V_n < C_{\alpha/2} \quad \text{o} \quad V_n > C_{1-\alpha/2}$$

como $C_{\alpha/2} < V_n < C_{1-\alpha/2}$ no se rechaza H_0 y se concluye que los datos no contradicen la hipótesis nula.

Paralelamente, la prueba de Anderson Darling obtuvo la constante $AD = 0.6372882$ la cual es menor a la constante 0.757 dada por Stephens (1977), y se concluye que los datos tampoco contradicen la hipótesis nula. Por lo tanto, los cluster de ozono que excedieron el umbral $u = 0.11$, resultantes de los niveles máximos diarios de ozono correspondientes a las estaciones en verano de los años 2003 a 2007 provienen de una distribución Gumbel.

6.2. Ejemplo 2

6.2.1. Lluvias máximas consecutivas para 1 día en Álamo, Veracruz

El conjunto de datos de este ejemplo corresponden a lluvias máximas consecutivas para 1 día, en el periodo 1967-2001 en la estación meteorológica, obtenido de las bases de datos del Servicio Meteorológico Nacional con la ayuda del ERIC (Extractor Rápido de Información Climática) del Instituto Mexicano de Tecnología del Agua (IMTA). Este tipo de datos se utiliza en el diseño de drenaje agrícola superficial.

En la Tabla 6.2 y Figura 6.4 se observan los datos seleccionados para realizar la prueba.

Ahora, hay que verificar que no exista ningún tipo de auto correlación en la serie de datos, lo cual se puede ver al estudiar la gráfica 6.5.

En la figura 6.5 se observa que existe una baja auto correlación entre los datos (es menor en valor absoluto a 0.2), por lo que se continúa con el análisis de los mismos.

Al realizar el gráfico Q-Q de las lluvias máximas consecutivas para 1 día en Álamo contra los cuantiles teóricos de una distribución Gumbel estándar ($\nu = 0$ y $\sigma = 1$), se obtiene lo siguiente:

La figura 6.6 proporciona fuerte evidencia de que los datos pueden ser ajustados a una distribución Gumbel, ya que se observa una relación lineal entre las lluvias máximas consecutivas para 1 día en Álamo contra los cuantiles teóricos,

| Año | Precipitación (mm) | Año | Precipitación (mm) |
|------|--------------------|------|--------------------|
| 1967 | 86.8 | 1986 | 142.6 |
| 1968 | 78.5 | 1987 | 42.9 |
| 1969 | 93.1 | 1988 | 60.2 |
| 1970 | 95.5 | 1989 | 100 |
| 1971 | 78.1 | 1990 | 64.3 |
| 1973 | 89.9 | 1991 | 98 |
| 1974 | 109.5 | 1992 | 30.7 |
| 1975 | 161.6 | 1993 | 37.9 |
| 1976 | 187.6 | 1994 | 60.7 |
| 1977 | 89.9 | 1995 | 48.7 |
| 1978 | 73.4 | 1996 | 39.7 |
| 1979 | 78.1 | 1997 | 80.3 |
| 1980 | 73.3 | 1998 | 116.4 |
| 1981 | 130.1 | 1999 | 120 |
| 1982 | 188.3 | 2000 | 160 |
| 1983 | 113.9 | 2001 | 129 |
| 1984 | 42.5 | 2002 | 80 |
| 1985 | 80 | | |

Tabla 6.2: Lluvias máximas consecutivas para 1 día en Álamo, Veracruz.

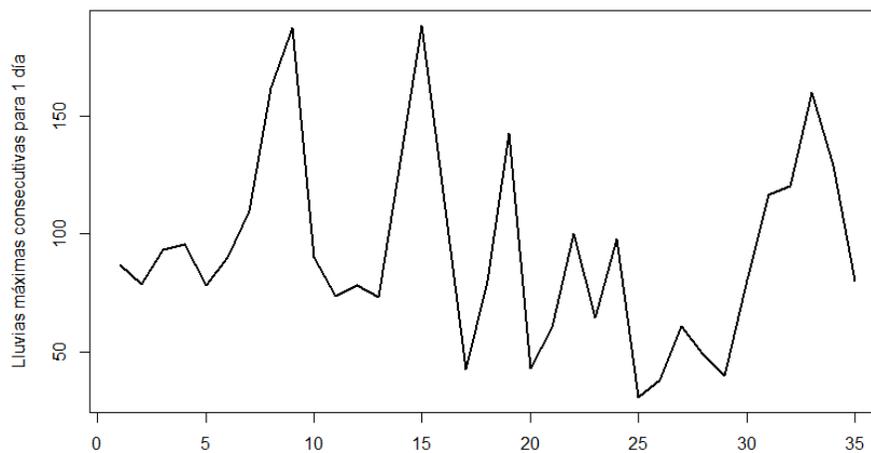


Figura 6.4: Lluvias máximas consecutivas para 1 día en Álamo, Veracruz.

Ahora se procede de manera rigurosa a verificar si los datos vienen de una distribución Gumbel, utilizando dos pruebas de prueba de bondad de ajuste, la primera es la Prueba de Anderson Darling y la segunda la propuesta en este escrito basada en el cociente de dos estimadores de la varianza uno de ellos en términos de las estadísticas de orden, donde el juego de hipótesis

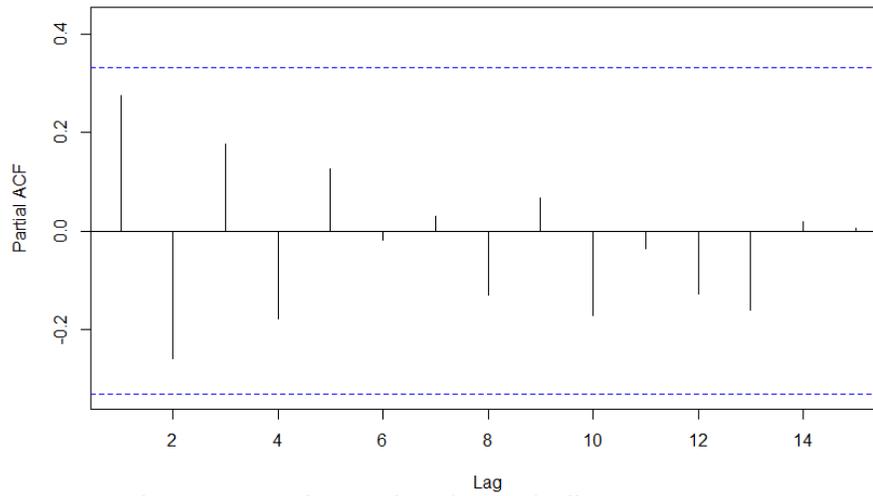


Figura 6.5: Auto correlación parcial para los datos de lluvias máximas consecutivas para 1 día en Álamo, Veracruz.

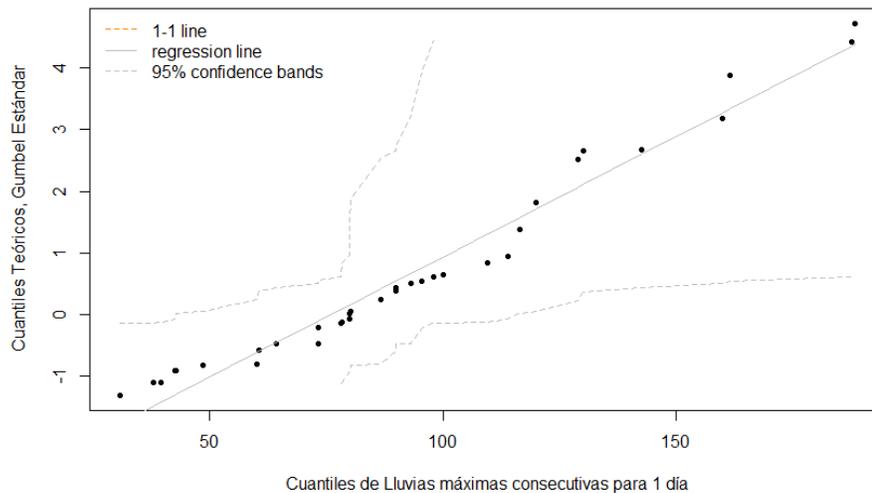


Figura 6.6: Gráfica Q-Q para las lluvias máximas consecutivas para 1 día en Álamo, Veracruz.

de interés es el planteado en 4.1.1.

El tamaño de muestra $n = 46$, para un nivel de significancia $\alpha = 0.05$, de la sección 4.4 se calculan las constantes críticas $C_{\alpha/2}$ y $C_{1-\alpha/2}$ y el estimador propuesto en la Ecuación 4.2.1 con un nivel de $\alpha = 0.05$, obteniéndose los valores siguientes:

$$C_{\alpha/2} = 0.6516656 \quad , \quad C_{1-\alpha/2} = 1.1629873 \quad \text{y} \quad V_n = 0.9622739$$

Finalmente, se utiliza la siguiente regla de decisión:

Rechazar H_0 si:

$$V_n < C_{\alpha/2} \quad \text{o} \quad V_n > C_{1-\alpha/2}$$

como $C_{\alpha/2} < V_n < C_{1-\alpha/2}$ no se rechaza H_0 y se concluye que los datos no contradicen la hipótesis nula.

Paralelamente, la prueba de Anderson Darling obtuvo la constante $AD = 0.269$ la cual es menor a la constante 0.757 dada por Stephens (1977), y se concluye que los datos tampoco contradicen la hipótesis nula. Por lo tanto, las lluvias máximas consecutivas para 1 día en Álamo, Veracruz provienen de una distribución Gumbel.

Conclusiones

De los resultados de los estudios de simulación realizados se puede concluir lo siguiente:

- 1).- La prueba $V_n(\xi, \theta)$ es uniformemente más potente que la prueba de Anderson-Darling en el caso de la distribución alternativa de Valores Extremos Generalizada, cuando se hace variar el parámetro de forma κ .
- 2).- Se presenta evidencia de que la prueba es consistente contra la distribución de valores extremos generalizada.
- 3).- La prueba $V_n(\xi, \theta)$ es uniformemente más potente que la prueba de Anderson-Darling en el caso de la distribución alternativa Normal, cuando se hace variar el parámetro de escala θ para tamaños de muestra $n \geq 30$.
- 4).- Se presenta evidencia de que la distribución nula asintótica de la estadística de prueba V es aproximadamente normal.

Bibliografía

- Borthiewicz, L. (1922). Variations breite und mittlerer fehler, sitzungsberichtet. *Berliner Math. Gas.*, 21, 2.
- Chandra, M., Singpurwalla, N. D., y Stephens, M. A. (1981). Kolmogorov statistics for tests of fit for the extreme value and weibull distributions. *Journal of the American Statistical Association*, 74, 729-735.
- D' Agostino, R. y Massaro, J. M. (1992). *Goodness of fit tests*. In: *Handbook of the Logistic Distribution*. Marcel Dekker, Inc., New York, 291-371.
- D' Agostino, R. y Stephens, M. A. (1986). *Goodness of fit techniques*. New York: Marcel Dekker.
- de Haan, L. (1970). On regular variation and its application to the weak convergence of sample extremes. *Mathematical Centre Tract, vol 32*. Amsterdam: Mathematisch Centrum.
- Dodd, E. L. (1923). The greatest and the least variate under general laws of error. *Trans. Amer. Math. Soc.* 25, 525-539.
- Fisher, R. A. y Tippett, H. A. (1928). Limiting forms of the frequency distribution of the largest or smallest sample. *Proc. Cambridge Phil. Soc.*, 24, 180-190.
- Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Societé Polonaise de Mathématique Cracow*, 6, 93-117.
- Fuller, W. E. (1914). Flood flows. *Trans. ASCE*, 77, 564-617.

- Gan, F. F. y Koehler, K. J. (1990). Goodness of fit tests based on p-p probability plots. *Technometrics*, 32, 289-303.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Math, Second series* 44: 423-453.
- Griffith, A. A. (1921). The phenomena of rupture and flow in solids. *Phil. Trans. R. Soc. Lond. A* 1921 221 163-198; DOI: 10.1098/rsta.1921.0006.
- Jhonson, N. L., Kotz, S., y Balakrishnan, N. (1995). *Continuous Univariate Distributions, Vol. 2*. Wiley-Interscience; 2 edition.
- Kendall, M. y Stuart, A. (1973). *The Advanced Theory of Statistics. Volume 2, Inference and Relationship*. New York, NY:MacMillan Publishing Co., Inc.
- Kimball, B. F. (1956). The bias in certain estimates of the parameters of the extreme-value distribution. *Ann. Math. Statist.* 27, No. 3, 758-767.
- Kinnison, R. (1989). Correlation coefficient goodness of fit test for the extreme value distribution. *Am. Statistn.* 43, 98-100.
- Kotz, S. y Nadarajah, S. (2000). *Extreme Value Distributions: Theory and Applications*. London: Imperial College Press.
- Mann, N. R., Schafer, R. E., y Singpurwalla, N. D. (1974). *Methods for Statistical Analysis of Reliability and Life Data*. New York: John Wiley.
- R Development, C. T. (2015). R: A language and environment for statistical computing. *R Foundation for Statistical Computing Vienna, Austria, 2004*. URL: <http://www.R-project.org>.
- Reiss, D. y Thomas, M. (2001). *Statistical analysis of extreme values*. 2nd Ed. Berlin: Birkhäuser.
- Reiss, R. y Thomas, M. (2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhäuser Basel.
- Seier, E. (2002). Comparison of tests for univariate normality. *Disponible en* <http://interstat.statjournals.net/index.php>.
- Shapiro, S. y Wilk, M. B. (1965). An analysis of variance test for normality: complete samples. *Biometrika*, 52, No. 3/4, 591-611.
- Stephens, M. (1977). Goodness of fit for the extreme value distribution. *Biometrika*, 64, 583-588.

- Tsujitani, M., Ohta, H., y Kase, S. (1980). Goodness of fit for the extreme value distribution. *IEEE Transactions on Reliability*, 29, 151-153.
- Villaseñor Alva, J. A. y González Estrada, E. (2009). On modeling cluster maxima with applications to ozone data from mexico city. *Environmetrics 2010*; 21: 528-540.
- Von Misses, R. (1936). La distribution de la plus grande de n valeurs. *Reprinted 1954 in Selected Papers II 271-294. Amer. Math. Soc. Providence, RI.*
- Öztürk, A. (1986). On the w test for the extreme value distribution. *Biometrika*, 73, 738-740.

.1. Propiedades de la distribución de valores extremos Tipo Gumbel

.1.1. Función de densidad

$$f_X(x) = \begin{cases} \frac{1}{\theta} \exp \left\{ -e^{-\frac{x-\xi}{\theta}} - \frac{x-\xi}{\theta} \right\} I_{(-\infty, \infty)}(x), & \xi \in \mathfrak{R}, \theta > 0 \\ 0, & \text{d.o.m.} \end{cases} \quad (.1.1)$$

.1.2. Función de distribución

$$F_X(x) = \exp \left\{ -e^{-\frac{(x-\xi)}{\theta}} \right\} I_{(-\infty, \infty)}(x) \quad (.1.2)$$

.1.3. Función generatriz de momentos

$$\begin{aligned} m_Z(t) &= E \{ e^{tZ} \} = E \left\{ e^{\frac{t(X-\xi)}{\theta}} \right\} = E \left\{ \left(e^{-\frac{X-\xi}{\theta}} \right)^t \right\} = E \{ Y^{-t} \} \\ &= \int_0^\infty y^{-t} e^{-y} dy = \int_0^\infty y^{(1-t)-1} e^{-y} dy = \Gamma(1-t), \quad \text{para } t < 1 \end{aligned} \quad (.1.3)$$

.1.4. Media y Varianza

$$\mu = E(X) = \left. \frac{\partial M_X(t)}{\partial t} \right|_{t=0} = \xi + \theta\gamma_0 \quad (.1.4)$$

$$\sigma^2 = E\{(X - \mu)^2\} = \frac{\pi^2\theta^2}{6} \quad (.1.5)$$

Donde $\gamma_0 = 0.577216$ es llamada la constante de Euler.

.1.5. Generación de variables aleatorias Gumbel

Sea $U \sim U(0, 1)$, entonces

$$X = \xi - \theta \ln(-\ln(U)) \sim Gumbel(\xi, \theta) \quad (.1.6)$$

.2. Estimadores de Máxima Verosimilitud de la Distribución Gumbel

$$\hat{\theta} = \bar{x} - \frac{\sum_{i=1}^n x_i e^{-x_i/\hat{\theta}}}{\sum_{i=1}^n e^{-x_i/\hat{\theta}}} \quad (.2.1)$$

$$\hat{\xi} = -\hat{\theta} \log \left[\frac{1}{n} \sum_{i=1}^n e^{-x_i/\hat{\theta}} \right]; \quad (.2.2)$$

Se anexan los siguientes códigos que se utilizaron en el presente trabajo programados en el software R. version 3.1.3(2015 – 03 – 09) para las gráficas de las distribuciones mencionadas y el cálculo de las constantes críticas,

.3. Cálculo de las constantes críticas

```
# -----  
rm(list = ls())  
CC.Gumbel <- function(n=10,alpha=0.05,B=20000,xi=0,theta=1){  
w <- numeric(length=B)  
cc <- numeric(length=2)  
l <- 1  
while(l <= B){  
library(evd)  
m <- rgumbel(n,loc=xi,scale=theta)  
m_ord <- sort(m)  
xj_r <- numeric(length=n)  
i <- 1  
while(i <= n){  
j <- i:n  
xj_r[i] <- m_ord[i]*sum(1/j)  
i <- i+1
```

```

}
w[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)
l <- l+1
}
cc[1] <- as.numeric(quantile(w,alpha/2))
cc[2] <- as.numeric(quantile(w,1-alpha/2))
answer <- list(n = n, alpha = alpha, B = B,
constantes_criticas = cc)
return(answer)
}

n <- c(10,30,50,100,200)
alpha <- c(0.01,0.05,0.1,0.25)
const_crit <- matrix(data=NA,ncol=2*length(alpha),nrow=length(n))

s <- 1
while(s<=length(n)){
a <- 1
while(a<=length(alpha)){
CC <- CC.Gumbel(n=n[s],alpha=alpha[a])
const_crit[s,2*a-1] <- CC$constantes_criticas[1]
const_crit[s,2*a] <- CC$constantes_criticas[2]
a <- a+1
}
s <- s+1
}
rm(a,s)
View(cbind(n,const_crit))
write.csv(const_crit,file="constantes_criticas.csv")
# -----

```

.4. Distribución de la estadística V_n para varios tamaños de muestra

```

# -----
rm(list = ls())
CC.Gumbel <- function(n=10,alpha=0.05,B=20000,xi=0,theta=1){

```

```

v <- numeric(length=B)
cc <- numeric(length=2)
l <- 1
while(l <= B){
library(evd)
m <- rgumbel(n,loc=xi,scale=theta)
m_ord <- sort(m)
xj_r <- numeric(length=n)
i <- 1
while(i <= n){
j <- i:n
xj_r[i] <- m_ord[i]*sum(1/j)
i <- i+1
}
v[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)
l <- l+1
}
cc[1] <- as.numeric(quantile(v,alpha/2))
cc[2] <- as.numeric(quantile(v,1-alpha/2))
parm <- c(xi,theta)
names(parm) <- c("xi","theta")
answer <- list(v = v, n = n, alpha = alpha, parametros = parm,
constantes_criticas = cc)
return(answer)
}
# -----
size <- c(10,30,50,100,200)
aux <- matrix(data=NA, ncol=length(size),nrow=20000)
n <- 1
while(n <= length(size)){
aux[,n] <- CC.Gumbel(n=size[n])$v
n <- n+1
}

par(mfrow=c(1,1))
plot(density(aux[,1]),main="",xlab="v",lwd=2,xlim=c(0.2,1.4),ylim=c(0,6))

```

```
l <- 2
while(l <= length(size)){
lines(density(aux[,l]),lwd=2,col=1)
l <- l+1
}
legend(0.2,6, c("n=10","n=30", "n=50", "n=100", "n=200"), col = c(1, 2, 3, 4, 5),
lty=c(1,1,1,1,1), lwd=c(2,2,2,2,2),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)

write.csv(aux,file="Distribucion de V para distintos tamaños de muestra.csv")
# -----
```

.5. Distribución de la estadística V_n variando parámetros de localidad y escala

```
# -----
rm(list = ls())
CC.Gumbel <- function(n=10,alpha=0.05,B=20000,xi=0,theta=1){
v <- numeric(length=B)
cc <- numeric(length=2)
l <- 1
while(l <= B){
library(evd)
m <- rgumbel(n,loc=xi,scale=theta)
m_ord <- sort(m)
xj_r <- numeric(length=n)
i <- 1
while(i <= n){
j <- i:n
xj_r[i] <- m_ord[i]*sum(1/j)
i <- i+1
}
v[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)
l <- l+1
}
cc[1] <- as.numeric(quantile(v,alpha/2))
```

```

cc[2] <- as.numeric(quantile(v,1-alpha/2))
parm      <- c(xi,theta)
names(parm) <- c("xi","theta")
answer <- list(v = v, n = n, alpha = alpha, parametros = parm,
constantes_criticas = cc)
return(answer)
}

n  <- c(10,30,50,100,200)

i <- 1
while(i <= length(n)){
# -----
xi <- c(-10,0,10)

aux1 <- matrix(data=NA, ncol=length(xi),nrow=20000)
x <- 1
while(x <= length(xi)){
aux1[,x] <- CC.Gumbel(n=n[i], xi=xi[x])$v
x <- x+1
}
par(mfrow=c(1,2))
plot(density(aux1[,1]),main=paste("n= ",n[i]),xlab="v",lwd=2)
l <- 2
while(l <= length(xi)){
lines(density(aux1[,l]),lwd=2,col=1)
l <- l+1
}
legend(0.4,2.5, c("xi=-10","xi=0", "xi=10"), col = c(1, 2, 3),
lty=c(1,1,1), lwd=c(2,2,2),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
theta <- c(1,10,100)

aux2 <- matrix(data=NA, ncol=length(theta), nrow=20000)
t <- 1

```

```
while(t <= length(theta)){
aux2[,t] <- CC.Gumbel(n=n[i], theta=theta[t])$v
t <- t+1
}

plot(density(aux2[,1]),main=paste("n= ",n[i]),xlab="v",lwd=2)

l <- 2
while(l <= length(theta)){
lines(density(aux2[,1]),lwd=2,col=l)
l <- l+1
}
legend(0.4,2.5, c("theta=1","theta=10", "theta=100"), col = c(1,2,3),
lty=c(1,1,1), lwd=c(2,2,2),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
i <- i+1
}
rm(i)
# -----
```

.6. Tamaño de la prueba de la estadística V_n para distintos tamaños de la muestra

```
# -----
rm(list = ls())

n <- 200#c(10,30,50,100,200)
alpha <- 0.05

i <- 1
while(i <= length(n)){

if(n==10){
if(alpha==0.01) {cc <- c(0.367,1.018)}
if(alpha==0.05) {cc <- c(0.457,0.989)}
if(alpha==0.1) {cc <- c(0.507,0.967)}

```

```
if(alpha==0.25) {cc <- c(0.585,0.921)}
}
if(n[i]==30){
if(alpha==0.01) {cc <- c(0.537,1.213)}
if(alpha==0.05) {cc <- c(0.629,1.159)}
if(alpha==0.1) {cc <- c(0.676,1.122)}
if(alpha==0.25) {cc <- c(0.754,1.064)}
}
if(n[i]==50){
if(alpha==0.01) {cc <- c(0.608,1.219)}
if(alpha==0.05) {cc <- c(0.702,1.164)}
if(alpha==0.1) {cc <- c(0.742,1.133)}
if(alpha==0.25) {cc <- c(0.810,1.078)}
}
if(n[i]==100){
if(alpha==0.01) {cc <- c(0.701,1.191)}
if(alpha==0.05) {cc <- c(0.783,1.142)}
if(alpha==0.1) {cc <- c(0.811,1.118)}
if(alpha==0.25) {cc <- c(0.865,1.079)}
}
if(n[i]==200){
if(alpha==0.01) {cc <- c(0.788,1.151)}
if(alpha==0.05) {cc <- c(0.841,1.117)}
if(alpha==0.1) {cc <- c(0.866,1.096)}
if(alpha==0.25) {cc <- c(0.906,1.064)}
}

Size.Gumbel <- function(n=10,cc=c(0.457,0.989),alpha=0.05,xi=0,theta=1,B=20000){
lower.cc <- cc[1]
upper.cc <- cc[2]

v <- numeric(length=B)
rechazos <- numeric(length=B)

l <- 1
while(l <= B){
```

```
library(evd)
m      <- rgumbel(n,loc=xi,scale=theta)
m_ord <- sort(m)
xj_r  <- numeric(length=n)
i <- 1
while(i <= n){
  j      <- i:n
  xj_r[i] <- m_ord[i]*sum(1/j)
  i <- i+1
}
v[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)
if(v[l]<lower.cc || v[l]>upper.cc) rechazos[l]=1 else rechazos[l]=0
l <- l+1
}
size <- sum(rechazos)/B

answer <- list(v = v, rechazos= rechazos, n = n, alpha = alpha,
constantes_criticas = cc, Size = size)
return(answer)
}
# -----
xi      <- seq(-1,1,0.1)
size1 <- numeric(length=length(xi))

x <- 1
while(x <=length(xi)){
  size1[x] <- Size.Gumbel(xi=xi[x],n=n[i],cc=cc)$Size
  x <- x+1
}
rm(x)
par(mfrow=c(1,2))
plot(xi,size1, main=paste("Tamaño de prueba, n= ",n[i]),
xlab=expression(xi),lwd=2,type="l",ylim=c(0,0.5),ylab="Size")
abline(h=0.05,lty=2,col=2)
# -----
theta <- seq(1,10,0.5)
```

```

size2 <- numeric(length=length(theta))

t <- 1
while(t <=length(theta)){
size2[t] <- Size.Gumbel(theta=theta[t],n=n[i],cc=cc)$Size
t <- t+1
}
rm(t)
plot(theta,size2, main=paste("Tamaño de prueba, n= ",n[i]),
xlab=expression(theta),lwd=2,type="l",ylim=c(0,0.5),ylab="Size")
abline(h=0.05,lty=2,col=2)
write.csv(size1,file=paste("Tamaños de prueba n=",n[i],".csv"))
i <- i+1
}
# -----

```

.7. Potencias de la prueba estadística V_n para distintos tamaños de la muestra

.7.1. Distribuciones Normal, Logística y Cauchy variando el parámetro de localidad ξ

```

# -----
rm(list = ls())

CC.Gumbel <- function(n=50,alpha=0.05,B=100,xi=0,theta=1){
v <- numeric(length=B)
cc <- numeric(length=2)
l <- 1
while(l <= B){
library(evd)
m <- rgumbel(n,loc=xi,scale=theta)
m_ord <- sort(m)
xj_r <- numeric(length=n)
i <- 1
while(i <= n){
j <- i:n

```

```

xj_r[i] <- m_ord[i]*sum(1/j)
i <- i+1
}
v[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)
l <- l+1
}
cc[1] <- as.numeric(quantile(v,alpha/2))
cc[2] <- as.numeric(quantile(v,1-alpha/2))
parm <- c(xi,theta)
names(parm) <- c("xi","theta")
answer <- list(v = v, n = n, alpha = alpha, parametros = parm,
constantes_criticas = cc)
return(answer)
}
# -----
Power.Loc <- function(n=50,cc,alpha=0.05,loc=0,B=10000,distribution="normal"){
lower.cc <- cc[1]
upper.cc <- cc[2]

v <- numeric(length=B)
rechazos <- numeric(length=B)

l <- 1
while(l <= B){
if(distribution=="normal") m <- rnorm(n,mean=loc,sd=1)
if(distribution=="logistica") m <- rlogis(n,location=loc,scale=1)
if(distribution=="cauchy") m <- rcauchy(n,location=loc,scale=1)
m_ord <- sort(m)
xj_r <- numeric(length=n)
i <- 1
while(i <= n){
j <- i:n
xj_r[i] <- m_ord[i]*sum(1/j)
i <- i+1
}
v[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)

```

```
if(v[l]<lower.cc || v[l]>upper.cc) rechazos[l]=1 else rechazos[l]=0
l <- l+1
}
potencia <- sum(rechazos)/B
answer <- list(v = v, rechazos= rechazos, n = n, alpha = alpha,
constantes_criticas = cc, distribution = distribution,
Potencia = potencia)
return(answer)
}

# -----
n          <- c(10,30,50,100,200)
distribution <- c("normal","logistica","cauchy")
loc         <- seq(-10,10,1)
powernormal <- matrix(data=NA,ncol=length(n),nrow=length(loc))
powerlogis  <- matrix(data=NA,ncol=length(n),nrow=length(loc))
powercauchy <- matrix(data=NA,ncol=length(n),nrow=length(loc))

s <- 1
while(s <= length(n)){
  const.crit <- CC.Gumbel(n=n[s])$constantes_criticas
  cc <- c(const.crit[1],const.crit[2])
  l <- 1
  while(l <=length(loc)){
    powernormal[l,s] <- Power.Loc(n=n[s],loc=loc[l],cc=cc,
distribution="normal")$Potencia
    l <- l+1
  }
  s <- s+1
}

s <- 1
while(s <= length(n)){
  const.crit <- CC.Gumbel(n=n[s])$constantes_criticas
  cc <- c(const.crit[1],const.crit[2])
  l <- 1
```

```
while(l <=length(loc)){
powerlogis[l,s] <- Power.Loc(n=n[s],loc=loc[l],cc=cc,
distribution="logistica")$Potencia
l <- l+1
}
s <- s+1
}

s <- 1
while(s <= length(n)){
const.crit <- CC.Gumbel(n=n[s])$constantes_criticas
cc <- c(const.crit[1],const.crit[2])
l <- 1
while(l <=length(loc)){
powercauchy[l,s] <- Power.Loc(n=n[s],loc=loc[l],cc=cc,
distribution="cauchy")$Potencia
l <- l+1
}
s <- s+1
}
rm(s,l)

powers.10 <- matrix(data=NA,ncol=length(distribution),nrow=length(loc))
powers.30 <- matrix(data=NA,ncol=length(distribution),nrow=length(loc))
powers.50 <- matrix(data=NA,ncol=length(distribution),nrow=length(loc))
powers.100 <- matrix(data=NA,ncol=length(distribution),nrow=length(loc))
powers.200 <- matrix(data=NA,ncol=length(distribution),nrow=length(loc))

powers.10[,1] <- powernormal[,1]
powers.10[,2] <- powerlogis[,1]
powers.10[,3] <- powercauchy[,1]

powers.30[,1] <- powernormal[,2]
powers.30[,2] <- powerlogis[,2]
powers.30[,3] <- powercauchy[,2]
```

```
powers.50[,1] <- powernormal[,3]
powers.50[,2] <- powerlogis[,3]
powers.50[,3] <- powercauchy[,3]

powers.100[,1] <- powernormal[,4]
powers.100[,2] <- powerlogis[,4]
powers.100[,3] <- powercauchy[,4]

powers.200[,1] <- powernormal[,5]
powers.200[,2] <- powerlogis[,5]
powers.200[,3] <- powercauchy[,5]

# -----
par(mfrow=c(1,1))
plot(loc,powers.10[,1], main="Potencias vs Parm Localidad con n=10",
xlab=expression(xi),lwd=2,type="o",ylim=c(0,1),ylab=expression(beta),
pch=1,col=1)
lines(loc,powers.10[,2],lwd=2,col=2,pch=7,type="o")
lines(loc,powers.10[,3],lwd=2,col=3,pch=8,type="o")
abline(h=0.05,lty=2,col=2)
legend(8.5,1, c("normal","logistica","cauchy"), col = c(1, 2, 3),
lty=c(1,1,1), lwd=c(2,2,2),pch=c(1,7,8),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
par(mfrow=c(1,1))
plot(loc,powers.30[,1], main="Potencias vs Parm Localidad con n=30",
xlab=expression(xi),lwd=2,type="o",ylim=c(0,1),ylab=expression(beta),
pch=1,col=1)
lines(loc,powers.30[,2],lwd=2,col=2,pch=7,type="o")
lines(loc,powers.30[,3],lwd=2,col=3,pch=8,type="o")
abline(h=0.05,lty=2,col=2)
legend(8.5,1, c("normal","logistica","cauchy"), col = c(1, 2, 3),
lty=c(1,1,1), lwd=c(2,2,2),pch=c(1,7,8),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
par(mfrow=c(1,1))
```

```
plot(loc,powers.50[,1], main="Potencias vs Parm Localidad con n=50",
xlab=expression(xi),lwd=2,type="o",ylim=c(0,1),ylab=expression(beta),
pch=1,col=1)
lines(loc,powers.50[,2],lwd=2,col=2,pch=7,type="o")
lines(loc,powers.50[,3],lwd=2,col=3,pch=8,type="o")
abline(h=0.05,lty=2,col=2)
legend(8.5,0.5, c("normal","logistica","cauchy"), col = c(1, 2, 3),
lty=c(1,1,1), lwd=c(2,2,2),pch=c(1,7,8),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
par(mfrow=c(1,1))
plot(loc,powers.100[,1], main="Potencias vs Parm Localidad con n=100",
xlab=expression(xi),lwd=2,type="o",ylim=c(0,1),ylab=expression(beta),
pch=1,col=1)
lines(loc,powers.100[,2],lwd=2,col=2,pch=7,type="o")
lines(loc,powers.100[,3],lwd=2,col=3,pch=8,type="o")
abline(h=0.05,lty=2,col=2)
legend(8.5,0.5, c("normal","logistica","cauchy"), col = c(1, 2, 3),
lty=c(1,1,1), lwd=c(2,2,2),pch=c(1,7,8),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
par(mfrow=c(1,1))
plot(loc,powers.200[,1], main="Potencias vs Parm Localidad con n=200",
xlab=expression(xi),lwd=2,type="o",ylim=c(0,1),ylab=expression(beta),
pch=1,col=1)
lines(loc,powers.200[,2],lwd=2,col=2,pch=7,type="o")
lines(loc,powers.200[,3],lwd=2,col=3,pch=8,type="o")
abline(h=0.05,lty=2,col=2)
legend(8.5,0.5, c("normal","logistica","cauchy"), col = c(1, 2, 3),
lty=c(1,1,1), lwd=c(2,2,2),pch=c(1,7,8),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
write.csv(powers.10,file="Potencias vs Parm Localidad con n=10.csv")
write.csv(powers.30,file="Potencias vs Parm Localidad con n=30.csv")
write.csv(powers.50,file="Potencias vs Parm Localidad con n=50.csv")
write.csv(powers.100,file="Potencias vs Parm Localidad con n=100.csv")
```

```
write.csv(powers.200,file="Potencias vs Parm Localidad con n=200.csv")
```

```
# -----
```

.7.2. Distribuciones Normal, Logística y Cauchy variando el parámetro de escala θ

```
# -----
```

```
rm(list = ls())
```

```
CC.Gumbel <- function(n=50,alpha=0.05,B=10000,xi=0,theta=1){
```

```
  v  <- numeric(length=B)
```

```
  cc <- numeric(length=2)
```

```
  l <- 1
```

```
  while(l <= B){
```

```
    library(evd)
```

```
    m  <- rgumbel(n,loc=xi,scale=theta)
```

```
    m_ord <- sort(m)
```

```
    xj_r <- numeric(length=n)
```

```
    i <- 1
```

```
    while(i <= n){
```

```
      j  <- i:n
```

```
      xj_r[i] <- m_ord[i]*sum(1/j)
```

```
      i <- i+1
```

```
    }
```

```
    v[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)
```

```
    l  <- l+1
```

```
  }
```

```
  cc[1] <- as.numeric(quantile(v,alpha/2))
```

```
  cc[2] <- as.numeric(quantile(v,1-alpha/2))
```

```
  parm  <- c(xi,theta)
```

```
  names(parm) <- c("xi","theta")
```

```
  answer <- list(v = v, n = n, alpha = alpha, parametros = parm,
```

```
  constantes_criticas = cc)
```

```
  return(answer)
```

```
}
```

```
# -----
```

```
Power.Esc <- function(n=50,cc,alpha=0.05,scale=1,B=10000,distribution="normal"){
```

```
  lower.cc <- cc[1]
```

```
upper.cc <- cc[2]

v          <- numeric(length=B)
rechazos  <- numeric(length=B)

l <- 1
while(l <= B){
  if(distribution=="normal")    m <- rnorm(n,mean=0,sd=scale)
  if(distribution=="logistica") m <- rlogis(n,location=0,scale=scale)
  if(distribution=="cauchy")    m <- rcauchy(n,location=0,scale=scale)
  m_ord <- sort(m)
  xj_r  <- numeric(length=n)
  i <- 1
  while(i <= n){
    j          <- i:n
    xj_r[i] <- m_ord[i]*sum(1/j)
    i <- i+1
  }
  v[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)
  if(v[l]<lower.cc || v[l]>upper.cc) rechazos[l]=1 else rechazos[l]=0
  l <- l+1
}

potencia <- sum(rechazos)/B
answer <- list(v = v, rechazos= rechazos, n = n, alpha = alpha,
  constantes_criticas = cc, distribution = distribution,
  Potencia = potencia)
return(answer)
}

# -----
n          <- c(10,30,50,100,200)
distribution <- c("normal","logistica","cauchy")
scale      <- seq(0.5,10,0.5)
powernormal <- matrix(data=NA,ncol=length(n),nrow=length(scale))
powerlogis  <- matrix(data=NA,ncol=length(n),nrow=length(scale))
powercauchy <- matrix(data=NA,ncol=length(n),nrow=length(scale))
```

```
s <- 1
while(s <= length(n)){
  const.crit <- CC.Gumbel(n=n[s])$constantes_criticas
  cc <- c(const.crit[1],const.crit[2])
  l <- 1
  while(l <=length(scale)){
    powernormal[l,s] <- Power.Esc(n=n[s],scale=scale[l],cc=cc,
    distribution="normal")$Potencia
    l <- l+1
  }
  s <- s+1
}
```

```
s <- 1
while(s <= length(n)){
  const.crit <- CC.Gumbel(n=n[s])$constantes_criticas
  cc <- c(const.crit[1],const.crit[2])
  l <- 1
  while(l <=length(scale)){
    powerlogis[l,s] <- Power.Esc(n=n[s],scale=scale[l],cc=cc,
    distribution="logistica")$Potencia
    l <- l+1
  }
  s <- s+1
}
```

```
s <- 1
while(s <= length(n)){
  const.crit <- CC.Gumbel(n=n[s])$constantes_criticas
  cc <- c(const.crit[1],const.crit[2])
  l <- 1
  while(l <=length(scale)){
    powercauchy[l,s] <- Power.Esc(n=n[s],scale=scale[l],cc=cc,
    distribution="cauchy")$Potencia
    l <- l+1
  }
}
```

```
s <- s+1
}
rm(s,1)

powers.10 <- matrix(data=NA,ncol=length(distribution),nrow=length(scale))
powers.30 <- matrix(data=NA,ncol=length(distribution),nrow=length(scale))
powers.50 <- matrix(data=NA,ncol=length(distribution),nrow=length(scale))
powers.100 <- matrix(data=NA,ncol=length(distribution),nrow=length(scale))
powers.200 <- matrix(data=NA,ncol=length(distribution),nrow=length(scale))

powers.10[,1] <- powernormal[,1]
powers.10[,2] <- powerlogis[,1]
powers.10[,3] <- powercauchy[,1]

powers.30[,1] <- powernormal[,2]
powers.30[,2] <- powerlogis[,2]
powers.30[,3] <- powercauchy[,2]

powers.50[,1] <- powernormal[,3]
powers.50[,2] <- powerlogis[,3]
powers.50[,3] <- powercauchy[,3]

powers.100[,1] <- powernormal[,4]
powers.100[,2] <- powerlogis[,4]
powers.100[,3] <- powercauchy[,4]

powers.200[,1] <- powernormal[,5]
powers.200[,2] <- powerlogis[,5]
powers.200[,3] <- powercauchy[,5]
# -----
par(mfrow=c(1,1))
plot(scale,powers.10[,1], main="Potencias vs Parm Escala con n=10",
xlab=expression(theta),lwd=2,type="o",ylim=c(0,1),ylab=expression(beta),
pch=1,col=1)
lines(scale,powers.10[,2],lwd=2,col=2,pch=7,type="o")
lines(scale,powers.10[,3],lwd=2,col=3,pch=8,type="o")
```

```

abline(h=0.05,lty=2,col=2)
legend(8.5,1, c("normal","logistica","cauchy"), col = c(1, 2, 3),
lty=c(1,1,1), lwd=c(2,2,2),pch=c(1,7,8),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
par(mfrow=c(1,1))
plot(scale,powers.30[,1], main="Potencias vs Parm Escala con n=30",
xlab=expression(theta),lwd=2,type="o",ylim=c(0,1),ylab=expression(beta),
pch=1,col=1)
lines(scale,powers.30[,2],lwd=2,col=2,pch=7,type="o")
lines(scale,powers.30[,3],lwd=2,col=3,pch=8,type="o")
abline(h=0.05,lty=2,col=2)
legend(8.5,1, c("normal","logistica","cauchy"), col = c(1, 2, 3),
lty=c(1,1,1), lwd=c(2,2,2),pch=c(1,7,8),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
par(mfrow=c(1,1))
plot(scale,powers.50[,1], main="Potencias vs Parm Escala con n=50",
xlab=expression(theta),lwd=2,type="o",ylim=c(0,1),ylab=expression(beta),
pch=1,col=1)
lines(scale,powers.50[,2],lwd=2,col=2,pch=7,type="o")
lines(scale,powers.50[,3],lwd=2,col=3,pch=8,type="o")
abline(h=0.05,lty=2,col=2)
legend(8.5,0.5, c("normal","logistica","cauchy"), col = c(1, 2, 3),
lty=c(1,1,1), lwd=c(2,2,2),pch=c(1,7,8),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
par(mfrow=c(1,1))
plot(scale,powers.100[,1], main="Potencias vs Parm Escala con n=100",
xlab=expression(theta),lwd=2,type="o",ylim=c(0,1),ylab=expression(beta),
pch=1,col=1)
lines(scale,powers.100[,2],lwd=2,col=2,pch=7,type="o")
lines(scale,powers.100[,3],lwd=2,col=3,pch=8,type="o")
abline(h=0.05,lty=2,col=2)
legend(8.5,0.5, c("normal","logistica","cauchy"), col = c(1, 2, 3),
lty=c(1,1,1), lwd=c(2,2,2),pch=c(1,7,8),merge=T,cex=0.8,y.intersp=0.4,

```

```
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
par(mfrow=c(1,1))
plot(scale,powers.200[,1], main="Potencias vs Parm Escala con n=200",
xlab=expression(theta),lwd=2,type="o",ylim=c(0,1),ylab=expression(beta),
pch=1,col=1)
lines(scale,powers.200[,2],lwd=2,col=2,pch=7,type="o")
lines(scale,powers.200[,3],lwd=2,col=3,pch=8,type="o")
abline(h=0.05,lty=2,col=2)
legend(8.5,0.5, c("normal","logistica","cauchy"), col = c(1, 2, 3),
lty=c(1,1,1), lwd=c(2,2,2),pch=c(1,7,8),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
write.csv(powers.10,file="Potencias vs Parm Escala con n=10.csv")
write.csv(powers.30,file="Potencias vs Parm Escala con n=30.csv")
write.csv(powers.50,file="Potencias vs Parm Escala con n=50.csv")
write.csv(powers.100,file="Potencias vs Parm Escala con n=100.csv")
write.csv(powers.200,file="Potencias vs Parm Escala con n=200.csv")
# -----
```

.7.3. Distribución t-student variando los grados de libertad

```
# -----
rm(list = ls())

CC.Gumbel <- function(n=50,alpha=0.05,B=10000,xi=0,theta=1){
v <- numeric(length=B)
cc <- numeric(length=2)
l <- 1
while(l <= B){
library(evd)
m <- rgumbel(n,loc=xi,scale=theta)
m_ord <- sort(m)
xj_r <- numeric(length=n)
i <- 1
while(i <= n){
j <- i:n
```

```

xj_r[i] <- m_ord[i]*sum(1/j)
i <- i+1
}
v[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)
l <- l+1
}
cc[1] <- as.numeric(quantile(v,alpha/2))
cc[2] <- as.numeric(quantile(v,1-alpha/2))
parm <- c(xi,theta)
names(parm) <- c("xi","theta")
answer <- list(v = v, n = n, alpha = alpha, parametros = parm,
constantes_criticas = cc)
return(answer)
}
# -----
Power.T <- function(n=50,cc,alpha=0.05,gl=1,B=10000){
lower.cc <- cc[1]
upper.cc <- cc[2]

v <- numeric(length=B)
rechazos <- numeric(length=B)

l <- 1
while(l <= B){
m <- rt(n,df=gl)
m_ord <- sort(m)
xj_r <- numeric(length=n)
i <- 1
while(i <= n){
j <- i:n
xj_r[i] <- m_ord[i]*sum(1/j)
i <- i+1
}
v[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)
if(v[l]<lower.cc || v[l]>upper.cc) rechazos[l]=1 else rechazos[l]=0
l <- l+1
}
}

```

```

}
potencia <- sum(rechazos)/B
answer <- list(v = v, rechazos= rechazos, n = n, alpha = alpha,
constantes_criticas = cc, Potencia = potencia)
return(answer)
}
# -----
n      <- c(10,30,50,100,200)
gl     <- seq(2,20,1)
powers <- matrix(data=NA,ncol=length(n),nrow=length(gl))

s <- 1
while(s <= length(n)){
const.crit <- CC.Gumbel(n=n[s])$constantes_criticas
cc <- c(const.crit[1],const.crit[2])
g <- 1
while(g <=length(gl)){
powers[g,s] <- Power.T(n=n[s],gl=gl[g],cc=cc)$Potencia
g <- g+1
}
s <- s+1
}
# -----
par(mfrow=c(1,1))
plot(gl,powers[,1], main="Potencia de la prueba vs T-Student",
xlab=expression(nu),lwd=2,type="l",ylim=c(0,1),ylab=expression(beta))
l <- 2
while(l<=length(n)){
lines(gl,powers[,l],lwd=2,col=1)
l <- l+1
}
abline(h=0.05,lty=2,col=2)
legend(15,0.5, c("n=10","n=30", "n=50", "n=100", "n=200"), col = c(1, 2, 3, 4, 5),
lty=c(1,1,1,1,1), lwd=c(2,2,2,2,2),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
rm(g,s,l)

```

```

write.csv(powers,file="Potencias vs T-Student.csv")
# -----

    .7.4.  Distribución de Valores Extremos Generalizada  $GEV(\xi, \theta, \kappa)$ 

# -----
rm(list = ls())

CC.Gumbel <- function(n=50,alpha=0.05,B=10000,xi=0,theta=1){
v  <- numeric(length=B)
cc  <- numeric(length=2)
l <- 1
while(l <= B){
library(evd)
m    <- rgumbel(n,loc=xi,scale=theta)
m_ord <- sort(m)
xj_r <- numeric(length=n)
i <- 1
while(i <= n){
j      <- i:n
xj_r[i] <- m_ord[i]*sum(1/j)
i <- i+1
}
v[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)
l    <- l+1
}
cc[1] <- as.numeric(quantile(v,alpha/2))
cc[2] <- as.numeric(quantile(v,1-alpha/2))
parm      <- c(xi,theta)
names(parm) <- c("xi","theta")
answer <- list(v = v, n = n, alpha = alpha, parametros = parm,
constantes_criticas = cc)
return(answer)
}
# -----
Power.GEV <- function(n=50,cc,alpha=0.05,shape=0,B=10000){
lower.cc <- cc[1]

```

```
upper.cc <- cc[2]

v      <- numeric(length=B)
rechazos <- numeric(length=B)

l <- 1
while(l <= B){
  m      <- rgev(n,loc=0,scale=1,shape=shape)
  m_ord <- sort(m)
  xj_r   <- numeric(length=n)
  i <- 1
  while(i <= n){
    j      <- i:n
    xj_r[i] <- m_ord[i]*sum(1/j)
    i <- i+1
  }
  v[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)
  if(v[l]<lower.cc || v[l]>upper.cc) rechazos[l]=1 else rechazos[l]=0
  l <- l+1
}
potencia <- sum(rechazos)/B
answer <- list(v = v, rechazos= rechazos, n = n, alpha = alpha,
  constantes_criticas = cc, Potencia = potencia)
return(answer)
}
# -----
n      <- c(10,30,50,100,200)
shape <- seq(-1,1,0.1)
powers <- matrix(data=NA,ncol=length(n),nrow=length(shape))

s <- 1
while(s <= length(n)){
  const.crit <- CC.Gumbel(n=n[s])$constantes_criticas
  cc <- c(const.crit[1],const.crit[2])
  l <- 1
  while(l <=length(shape)){
```

```
powers[l,s] <- Power.GEV(n=n[s],shape=shape[l],cc=cc)$Potencia
l <- l+1
}
s <- s+1
}
# -----
par(mfrow=c(1,1))
plot(shape,powers[,1], main="Potencias vs GEV",
xlab=expression(kappa),lwd=2,type="l",ylim=c(0,1),ylab=expression(beta),
pch=1,col=1)
lines(shape,powers[,2],lwd=2,col=2,pch=2,type="l")
lines(shape,powers[,3],lwd=2,col=3,pch=3,type="l")
lines(shape,powers[,4],lwd=2,col=4,pch=4,type="l")
lines(shape,powers[,5],lwd=2,col=5,pch=5,type="l")
abline(h=0.05,lty=2,col=2)
legend(0.9,0.5, c("n=10","n=30","n=50","n=100","n=200"), col = c(1, 2, 3, 4, 5),
lty=c(1,1,1,1,1), lwd=c(2,2,2),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
write.csv(powers,file="Potencias vs GEV para varios tamaños de muestra.csv")
# -----
```

.8. Comparación de las potencias de la estadística de prueba v vs Anderson-Darling variando los tamaños de la muestra

.8.1. Distribución Normal variando el parámetro de escala θ

```
# -----
rm(list=ls())
# -----
AD.Test <- function(x, alpha=0.05){
library(evd)
n <- length(x)
x <- sort(x, decreasing = FALSE)
f <- function(p){
1/n*sum(x) - (sum(x*exp(-x/p)))/(sum(exp(-x/p))) - p
}
}
```

```
Pgumbel <- function(q, loc=0, scale=1){
answer = exp(-exp(-(q-loc)/scale))
answer[scale <= 0] = NaN
answer
}
itSol <- uniroot(f, c(-1000,1000), tol=1e-07, maxiter=1e+07)
theta <- as.numeric(itSol$root)

if(theta>0){
xi          <- -theta*log(sum(exp(-x/theta)/n))
parafit     <- numeric(2)
parafit[1]  <- xi
parafit[2]  <- theta
names(parafit) <- c("localidad", "escala")
pit         <- numeric(n)
pit[1:n]    <- Pgumbel(x[1:n], parafit[1], parafit[2])
aux <- numeric(n)
i <- 1
while(i <=n){
aux[i] <- (2*i-1)*(log(pit[i]) + log(1-pit[n+1-i]))
i <- i+1
}
A2 <- -sum(aux)/n-n
AD <- A2*(1+0.2/sqrt(n))
return(list(parameter_estimation = parafit, alpha=alpha, A_Cuadrada = A2,
Anderson_Darling = AD))
}
else{
names(theta) <- c("escala")
AD <- NaN
return(list(parameter_estimation = theta,
alpha=alpha, Anderson_Darling = AD))
}
}
# -----
Power.Norm.AD <- function(n=50,mean=0,sd=1,B=10000,alpha=0.05){
```

```
if (alpha==0.25) cc <- 0.474
if (alpha==0.10) cc <- 0.637
if (alpha==0.05) cc <- 0.757
if (alpha==0.025) cc <- 0.877
if (alpha==0.01) cc <- 1.038
rechazos <- numeric(B)
l <- 1
while(l <= B){
x <- rnorm(n,mean=0,sd=sd)
AD <- AD.Test(x)$Anderson_Darling
if(AD > cc || AD == "NaN") rechazos[l]=1 else rechazos[l]=0
l <- l+1
}
potencia <- sum(rechazos)/B
answer <- list(B = B, n = n, Anderson_Darling = AD, alpha = alpha,
Potencia = potencia)
return(answer)
}
# -----
n <- c(10,30,50,100,200)
scale <- seq(0.5,10,0.5)
powers.AD <- matrix(data=NA,ncol=length(n),nrow=length(scale))

s <- 1
while(s <= length(n)){
l <- 1
while(l <=length(scale)){
powers.AD[l,s] <- Power.Norm.AD(n=n[s],sd=scale[l])$Potencia
l <- l+1
}
s <- s+1
}
# -----
par(mfrow=c(1,1))
plot(scale,powers.AD[,1], main="Potencias vs Normal con AD",
xlab=expression(sigma),lwd=2,type="o",ylim=c(0,1),ylab=expression(beta),
```

```
pch=16,col=1)
lines(scale,powers.AD[,2],lwd=2,col=2,pch=16,type="o")
lines(scale,powers.AD[,3],lwd=2,col=3,pch=16,type="o")
lines(scale,powers.AD[,4],lwd=2,col=4,pch=16,type="o")
lines(scale,powers.AD[,5],lwd=2,col=5,pch=16,type="o")
abline(h=0.05,lty=2,col=2)
legend(9,0.9, c("n=10","n=30","n=50","n=100","n=200"), col = c(1, 2, 3, 4, 5),
lty=c(1,1,1,1,1), lwd=c(2,2,2,2,2),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
write.csv(powers.AD,file="Potencias vs Normal con Anderson Darling.csv")
# -----
V.10 <- read.csv(file="Potencias vs Parm Escala con n=10.csv", header=T)
V.30 <- read.csv(file="Potencias vs Parm Escala con n=30.csv", header=T)
V.50 <- read.csv(file="Potencias vs Parm Escala con n=50.csv", header=T)
V.100 <- read.csv(file="Potencias vs Parm Escala con n=100.csv", header=T)
V.200 <- read.csv(file="Potencias vs Parm Escala con n=200.csv", header=T)
AD <- read.csv(file="Potencias vs Normal con Anderson Darling.csv", header=T)

V.10 <- V.10[,2]; V.30 <- V.30[,2]; V.50 <- V.50[,2]
V.100 <- V.100[,2]; V.200 <- V.200[,2]

V <- cbind(1:length(V.10),V.10,V.30,V.50,V.100,V.200)
V <- as.data.frame(V)

scale <- seq(0.5,10,0.5)
n <- c(10,30,50,100,200)

par(mfrow=c(2,3))

i <- 2
while(i <= 6){
plot(scale, V[1:length(scale),i], type = "o" , lwd = 2,
main = paste("Potencias para n=",n[i-1]), xlab = expression(sigma),
ylab = expression(beta), col = 'red', ylim=c(0,1),pch=16)
lines(scale, AD[,i], type="o", lwd=2, col = 'blue',pch=15)
```

```
abline(h=0.05, lty=2, col='black')
if(i <= 4){
  legend(8,1, c("V","AD"), col = c('red', 'blue'), lty=c(1,1), lwd=c(2,2),
  pch=c(16,15), merge=T,cex=0.8,y.intersp=0.4,x.intersp=0.3,
  box.lty=0, bty="n", seg.len=1)
}
else{
  legend(8,0.45, c("V","AD"), col = c('red', 'blue'), lty=c(1,1), lwd=c(2,2),
  pch=c(16,15), merge=T,cex=0.8,y.intersp=0.4,x.intersp=0.3,
  box.lty=0, bty="n", seg.len=1)
}
i <- i+1
}
# -----
```

.8.2. Distribución Logística variando el parámetro de escala θ

```
# -----
rm(list=ls())
# -----
AD.Test <- function(x, alpha=0.05){
  library(evd)
  n <- length(x)
  x <- sort(x, decreasing = FALSE)
  f <- function(p){
    1/n*sum(x) - (sum(x*exp(-x/p)))/(sum(exp(-x/p))) - p
  }
  Pgumbel <- function(q, loc=0, scale=1){
    answer = exp(-exp(-(q-loc)/scale))
    answer[scale <= 0] = NaN
    answer
  }
  itSol <- uniroot(f, c(-1000,1000), tol=1e-07, maxiter=1e+07)
  theta <- as.numeric(itSol$root)

  if(theta>0){
    xi <- -theta*log(sum(exp(-x/theta)/n))
```

```
parafit      <- numeric(2)
parafit[1]   <- xi
parafit[2]   <- theta
names(parafit) <- c("localidad", "escala")
pit          <- numeric(n)
pit[1:n]     <- Pgumbel(x[1:n], parafit[1], parafit[2])
aux <- numeric(n)
i <- 1
while(i <=n){
aux[i] <- (2*i-1)*(log(pit[i]) + log(1-pit[n+1-i]))
i <- i+1
}
A2 <- -sum(aux)/n-n
AD <- A2*(1+0.2/sqrt(n))
return(list(parameter_estimation = parafit, alpha=alpha, A_Cuadrada = A2,
Anderson_Darling = AD))
}
else{
names(theta) <- c("escala")
AD <- NaN
return(list(parameter_estimation = theta,
alpha=alpha, Anderson_Darling = AD))
}
}
# -----
Power.Logis.AD <- function(n=50,location=0,scale=1,B=10000,alpha=0.05){
if (alpha==0.25) cc <- 0.474
if (alpha==0.10) cc <- 0.637
if (alpha==0.05) cc <- 0.757
if (alpha==0.025) cc <- 0.877
if (alpha==0.01) cc <- 1.038
rechazos <- numeric(B)
l <- 1
while(l <= B){
x <- rlogis(n,location=0,scale=scale)
AD <- AD.Test(x)$Anderson_Darling
```

```
if(AD > cc || AD == "NaN") rechazos[l]=1 else rechazos[l]=0
l <- l+1
}
potencia <- sum(rechazos)/B
answer <- list(B = B, n = n, Anderson_Darling = AD, alpha = alpha,
Potencia = potencia)
return(answer)
}
# -----
n <- c(10,30,50,100,200)
scale <- seq(0.5,10,0.5)
powers.AD <- matrix(data=NA,ncol=length(n),nrow=length(scale))

s <- 1
while(s <= length(n)){
l <- 1
while(l <=length(scale)){
powers.AD[l,s] <- Power.Logis.AD(n=n[s],scale=scale[l])$Potencia
l <- l+1
}
s <- s+1
}
# -----
par(mfrow=c(1,1))
plot(scale,powers.AD[,1], main="Potencias vs Logistica con AD",
xlab=expression(sigma),lwd=2,type="o",ylim=c(0,1),ylab=expression(beta),
pch=16,col=1)
lines(scale,powers.AD[,2],lwd=2,col=2,pch=16,type="o")
lines(scale,powers.AD[,3],lwd=2,col=3,pch=16,type="o")
lines(scale,powers.AD[,4],lwd=2,col=4,pch=16,type="o")
lines(scale,powers.AD[,5],lwd=2,col=5,pch=16,type="o")
abline(h=0.05,lty=2,col=2)
legend(9,0.5, c("n=10","n=30","n=50","n=100","n=200"), col = c(1, 2, 3, 4, 5),
lty=c(1,1,1,1,1), lwd=c(2,2,2,2,2),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
```

```
write.csv(powers.AD,file="Potencias vs Logistica con Anderson Darling.csv")
# -----
V.10 <- read.csv(file="Potencias vs Parm Escala con n=10.csv", header=T)
V.30 <- read.csv(file="Potencias vs Parm Escala con n=30.csv", header=T)
V.50 <- read.csv(file="Potencias vs Parm Escala con n=50.csv", header=T)
V.100 <- read.csv(file="Potencias vs Parm Escala con n=100.csv", header=T)
V.200 <- read.csv(file="Potencias vs Parm Escala con n=200.csv", header=T)
AD <- read.csv(file="Potencias vs Logistica con Anderson Darling.csv", header=T)

V.10 <- V.10[,2]; V.30 <- V.30[,2]; V.50 <- V.50[,2]
V.100 <- V.100[,2]; V.200 <- V.200[,2]

V <- cbind(1:length(V.10),V.10,V.30,V.50,V.100,V.200)
V <- as.data.frame(V)

scale <- seq(0.5,10,0.5)
n <- c(10,30,50,100,200)

par(mfrow=c(2,3))

i <- 2
while(i <= 6){
plot(scale, V[1:length(scale),i], type = "o" , lwd = 2,
main = paste("Potencias para n=",n[i-1]), xlab = expression(sigma),
ylab = expression(beta), col = 'red', ylim=c(0,1),pch=16)
lines(scale, AD[,i], type="o", lwd=2, col = 'blue',pch=15)
abline(h=0.05, lty=2, col='black')
legend(8,0.5, c("V","AD"), col = c('red', 'blue'), lty=c(1,1), lwd=c(2,2),
pch=c(16,15), merge=T,cex=0.8,y.intersp=0.4,x.intersp=0.3,
box.lty=0, bty="n", seg.len=1)
i <- i+1
}
# -----
```

.8.3. Distribución t-student variando los grados de libertad

```
# -----
```

```
rm(list=ls())
# -----
AD.Test <- function(x, alpha=0.05){
  library(evd)
  n <- length(x)
  x <- sort(x, decreasing = FALSE)
  f <- function(p){
    1/n*sum(x)-(sum(x*exp(-x/p)))/(sum(exp(-x/p)))-p
  }
  Pgumbel <- function(q, loc=0, scale=1){
    answer = exp(-exp(-(q-loc)/scale))
    answer[scale <= 0] = NaN
    answer
  }
  itSol <- uniroot(f, c(-1000,1000), tol=1e-07, maxiter=1e+07)
  theta <- as.numeric(itSol$root)

  if(theta>0){
    xi <- -theta*log(sum(exp(-x/theta)/n))
    parafit <- numeric(2)
    parafit[1] <- xi
    parafit[2] <- theta
    names(parafit) <- c("localidad", "escala")
    pit <- numeric(n)
    pit[1:n] <- Pgumbel(x[1:n], parafit[1], parafit[2])
    aux <- numeric(n)
    i <- 1
    while(i <=n){
      aux[i] <- (2*i-1)*(log(pit[i]) + log(1-pit[n+1-i]))
      i <- i+1
    }
    A2 <- -sum(aux)/n-n
    AD <- A2*(1+0.2/sqrt(n))
    return(list(parameter_estimation = parafit, alpha=alpha, A_Cuadrada = A2,
      Anderson_Darling = AD))
  }
}
```

```
else{
names(theta) <- c("escala")
AD <- NaN
return(list(parameter_estimation = theta,
alpha=alpha, Anderson_Darling = AD))
}
}
# -----
Power.T.AD <- function(n=50,gl=10,alpha=0.05,B=10000){
if (alpha==0.25) cc <- 0.474
if (alpha==0.10) cc <- 0.637
if (alpha==0.05) cc <- 0.757
if (alpha==0.025) cc <- 0.877
if (alpha==0.01) cc <- 1.038
rechazos <- numeric(B)
l <- 1
while(l <= B){
x <- rt(n,df=gl)
AD <- AD.Test(x)$Anderson_Darling
if(AD > cc || AD == "NaN") rechazos[l]=1 else rechazos[l]=0
l <- l+1
}
potencia <- sum(rechazos)/B
answer <- list(B = B, n = n, Anderson_Darling = AD, alpha = alpha,
Potencia = potencia)
return(answer)
}
# -----
n <- c(10,30,50,100,200)
gl <- 2:20
powers.AD <- matrix(data=NA,ncol=length(n),nrow=length(gl))

s <- 1
while(s <= length(n)){
l <- 1
while(l <=length(gl)){
```

```
powers.AD[l,s] <- Power.T.AD(n=n[s],gl=gl[l])$Potencia
l <- l+1
}
s <- s+1
}
rm(l,s)
# -----
par(mfrow=c(1,1))
plot(gl,powers.AD[,1], main="Potencias vs T con AD",
xlab=expression(nu),lwd=2,type="o",ylim=c(0,1),ylab=expression(beta),
pch=16,col=1)
lines(gl,powers.AD[,2],lwd=2,col=2,pch=16,type="o")
lines(gl,powers.AD[,3],lwd=2,col=3,pch=16,type="o")
lines(gl,powers.AD[,4],lwd=2,col=4,pch=16,type="o")
lines(gl,powers.AD[,5],lwd=2,col=5,pch=16,type="o")
abline(h=0.05,lty=2,col=2)
legend(5,0.5, c("n=10","n=30","n=50","n=100","n=200"), col = c(1, 2, 3, 4, 5),
lty=c(1,1,1,1,1), lwd=c(2,2,2,2,2),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
write.csv(powers.AD,file="Potencias vs T con Anderson Darling.csv")
# -----
V <- read.csv(file="Potencias vs T-Student.csv", header=T)
AD <- read.csv(file="Potencias vs T con Anderson Darling.csv", header=T)

gl <- seq(2,20,1)
n <- c(10,30,50,100,200)

par(mfrow=c(2,3))
i <- 2
while(i <= 6){
plot(gl, V[,i], type = "o" , lwd = 2,
main = paste("Potencias para n=",n[i-1]), xlab = expression(nu),
ylab = expression(beta), col = 'red', ylim=c(0,1), pch=16)
lines(gl, AD[,i], type="o", lwd=2, col = 'blue',pch=15)
abline(h=0.05, lty=2, col='black')
```

```
legend(17,1, c("V","AD"), col = c('red', 'blue'), lty=c(1,1), lwd=c(2,2),
pch=c(16,15), merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0, bty="n", seg.len=1)
i <- i+1
}
# -----
```

.8.4. Distribución de Valores Extremos Generalizada variando su parámetro de forma

κ

```
# -----
rm(list=ls())
# -----
AD.Test <- function(x, alpha=0.05){
library(evd)
n <- length(x)
x <- sort(x, decreasing = FALSE)
f <- function(p){
1/n*sum(x)-(sum(x*exp(-x/p)))/(sum(exp(-x/p)))-p
}
Pgumbel <- function(q, loc=0, scale=1){
answer = exp(-exp(-(q-loc)/scale))
answer[scale <= 0] = NaN
answer
}
itSol <- uniroot(f, c(-1000,1000), tol=1e-07, maxiter=1e+07)
theta <- as.numeric(itSol$root)

if(theta>0){
xi <- -theta*log(sum(exp(-x/theta)/n))
parafit <- numeric(2)
parafit[1] <- xi
parafit[2] <- theta
names(parafit) <- c("localidad", "escala")
pit <- numeric(n)
pit[1:n] <- Pgumbel(x[1:n], parafit[1], parafit[2])
aux <- numeric(n)
```

```
i <- 1
while(i <=n){
aux[i] <- (2*i-1)*(log(pit[i]) + log(1-pit[n+1-i]))
i <- i+1
}
A2 <- -sum(aux)/n-n
AD <- A2*(1+0.2/sqrt(n))
return(list(parameter_estimation = parafit, alpha=alpha, A_Cuadrada = A2,
Anderson_Darling = AD))
}
else{
names(theta) <- c("escala")
AD <- NaN
return(list(parameter_estimation = theta,
alpha=alpha, Anderson_Darling = AD))
}
}
# -----
Power.GEV.AD <- function(n=50,alpha=0.05,shape=0.5,B=10000){
if (alpha==0.25) cc <- 0.474
if (alpha==0.10) cc <- 0.637
if (alpha==0.05) cc <- 0.757
if (alpha==0.025) cc <- 0.877
if (alpha==0.01) cc <- 1.038
rechazos <- numeric(B)
l <- 1
while(l <= B){
library(evd)
x <- rgev(n,loc=0,scale=1,shape=shape)
AD <- AD.Test(x)$Anderson_Darling
if(AD > cc || AD == "NaN") rechazos[l]=1 else rechazos[l]=0
l <- l+1
}
potencia <- sum(rechazos)/B
answer <- list(B = B, n = n, Anderson_Darling = AD, alpha = alpha,
Potencia = potencia)
```

```
return(answer)
}
# -----
n      <- c(10,30,50,100,200)
shape  <- seq(-1,0.5,0.1)
powers.AD <- matrix(data=NA,ncol=length(n),nrow=length(shape))

s <- 1
while(s <= length(n)){
  l <- 1
  while(l <=length(shape)){
    powers.AD[l,s] <- Power.GEV.AD(n=n[s],shape=shape[l])$Potencia
    l <- l+1
  }
  s <- s+1
}
# -----
par(mfrow=c(1,1))
plot(shape,powers.AD[,1], main="Potencias vs GEV con AD",
xlab=expression(kappa),lwd=2,type="l",ylim=c(0,1),ylab=expression(beta),
pch=1,col=1)
lines(shape,powers.AD[,2],lwd=2,col=2,pch=2,type="l")
lines(shape,powers.AD[,3],lwd=2,col=3,pch=3,type="l")
lines(shape,powers.AD[,4],lwd=2,col=4,pch=4,type="l")
lines(shape,powers.AD[,5],lwd=2,col=5,pch=5,type="l")
abline(h=0.05,lty=2,col=2)
legend(-1,0.5, c("n=10","n=30","n=50","n=100","n=200"), col = c(1, 2, 3, 4, 5),
lty=c(1,1,1,1,1), lwd=c(2,2,2,2,2),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
# -----
write.csv(powers.AD,file="Potencias vs GEV con Anderson Darling.csv")
# -----
V <- read.csv(file="Potencias vs GEV para varios tamaños de muestra.csv", header=T)
AD <- read.csv(file="Potencias vs GEV con Anderson Darling.csv", header=T)

shape <- seq(-1,0.5,0.1)
```

```

n      <- c(10,30,50,100,200)

par(mfrow=c(2,3))
i <- 2
while(i <= 6){
plot(shape, V[1:length(shape),i], type = "o" , lwd = 2,
main = paste("Potencias para n=",n[i-1]), ylab = expression(beta),
col = 2, ylim=c(0,1),xlab=expression(kappa),pch=16)
lines(shape, AD[,i], type="o" , lwd=2, col = 4,pch=15)
abline(h=0.05, lty=2)
abline(v=0, lty=2)
legend(-1,1, c("V","AD"), col = c(2, 4),
lty=c(1,1), lwd=c(2,2),merge=T,cex=0.8,y.intersp=0.4,
x.intersp=0.3,box.lty=0,bty="n",seg.len=1)
i <- i+1
}
# -----

```

.9. Ejemplos de aplicación

```

rm(list = ls())

data <- c(0.025,0.045,0.014,0.037,0.039,0.008,0.071,0.07,0.055,0.001,0.016,0.039,
0.037,0.036,0.021,0.085,0.028,0.039,0.016,0.022,0.006,0.016,0.03,0.003,0.029,0.013,
0.025,0.005,0.01,0.017,0.003,0.05,0.071,0.038,0.073,0.047,0.051,0.037,0.064,0.009,
0.013,0.047,0.029,0.001,0.061,0.062,0.012,0.011,0.013,0.002,0.041,0.048,0.009,0.015,
0.04,0.047,0.024,0.043,0.029,0.026,0.02,0.05,0.047,0.084,0.002,0.071,0.05,0.022,
0.006,0.008,0.073,0.04,0.023,0.001,0.027,0.02,0.047,0.032)

ts(data)

pacf(data,main = "")

plot(data,type = "l", lwd = 2, ylab = "Niveles de Ozono (ppm)",
xlab = "")

```

```

library(evd)
q <- rgumbel(length(data))
qqplot(data,q,xlab = "Cuantiles de Niveles Maximos de Ozono",
ylab = "Cuantiles Teoricos, Gumbel Estandar")

CC.Gumbel <- function(n=10,alpha=0.05,B=20000,xi=0,theta=1){
w  <- numeric(length=B)
cc  <- numeric(length=2)
l <- 1
while(l <= B){
library(evd)
m    <- rgumbel(n,loc=xi,scale=theta)
m_ord <- sort(m)
xj_r <- numeric(length=n)
i <- 1
while(i <= n){
j      <- i:n
xj_r[i] <- m_ord[i]*sum(1/j)
i <- i+1
}
w[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)
l    <- l+1
}
cc[1] <- as.numeric(quantile(w,alpha/2))
cc[2] <- as.numeric(quantile(w,1-alpha/2))
answer <- list(n = n, alpha = alpha, B = B,
constantes_criticas = cc)
return(answer)
}

# Calculo de las constantes criticas
CC.Gumbel(n = length(data),alpha=0.05)

# Calculo del estimador V
n <- length(data)
xj_r <- numeric(length=n)

```

```

i <- 1
while(i <= n){
  j      <- i:n
  m_ord  <- sort(data)
  xj_r[i] <- m_ord[i]*sum(1/j)
  i <- i+1
}
m_ord
xj_r
pi^2/6*(mean(data) - sum(xj_r)/n)^2/var(data)

# Calculo del estimador de Anderson Darling
AD.Test <- function(x, alpha){
  library(evd)
  n <- length(x)
  x <- sort(x, decreasing = FALSE)
  f <- function(p){
    1/n*sum(x) - (sum(x*exp(-x/p)))/(sum(exp(-x/p))) - p
  }
  Pgumbel <- function(q, loc = 0, scale = 1){
    answer = exp(-exp(-(q-loc)/scale))
    answer[scale <= 0] = NaN
    answer
  }
  itSol <- uniroot(f, c(-1000,1000), tol=1e-07, maxiter=1e+07)
  theta <- as.numeric(itSol$root)

  if(theta>0){
    xi      <- -theta*log(sum(exp(-x/theta)/n))
    parafit <- numeric(2)
    parafit[1] <- xi
    parafit[2] <- theta
    names(parafit) <- c("localidad", "escala")
    pit     <- numeric(n)
    pit[1:n] <- Pgumbel(x[1:n], parafit[1], parafit[2])
  }
}

```

```
aux <- numeric(n)
i <- 1
while(i <=n){
aux[i] <- (2*i-1)*(log(pit[i]) + log(1-pit[n+1-i]))
i <- i+1
}
A2 <- -sum(aux)/n-n
AD <- A2*(1+0.2/sqrt(n))
return(list(parameter_estimation = parafit, alpha=alpha, A_Cuadrada = A2,
Anderson_Darling = AD))
}
else{
names(theta) <- c("escala")
AD <- NaN
return(list(parameter_estimation = theta,
alpha=alpha, Anderson_Darling = AD))
}
}

AD.Test(data,alpha=0.05)

# -----
rm(list = ls())

data <- c(86.8,78.5,93.1,95.5,78.1,89.9,109.5,161.6,187.6,89.9,73.4,78.1,73.3,130.1,
188.3,113.9,42.5,80,142.6,42.9,60.2,100,64.3,98,30.7,37.9,60.7,48.7,39.7,80.3,116.4,
120,160,129,80)

ts(data)

pacf(data,main = "")

plot(data,type = "l", lwd = 2, ylab = "Lluvias maximas consecutivas para 1 dia",
xlab = "")

library(evd)
```

```

require(graphics)
q <- rgumbel(length(data))
qqplot(data,q,xlab = "Cuantiles de Lluvias maximas consecutivas para 1 dia",
ylab = "Cuantiles Teoricos, Gumbel Estandar")

CC.Gumbel <- function(n=10,alpha=0.05,B=20000,xi=0,theta=1){
w  <- numeric(length=B)
cc  <- numeric(length=2)
l <- 1
while(l <= B){
library(evd)
m    <- rgumbel(n,loc=xi,scale=theta)
m_ord <- sort(m)
xj_r <- numeric(length=n)
i <- 1
while(i <= n){
j      <- i:n
xj_r[i] <- m_ord[i]*sum(1/j)
i <- i+1
}
w[l] <- pi^2/6*(mean(m) - sum(xj_r)/n)^2/var(m)
l    <- l+1
}
cc[1] <- as.numeric(quantile(w,alpha/2))
cc[2] <- as.numeric(quantile(w,1-alpha/2))
answer <- list(n = n, alpha = alpha, B = B,
constantes_criticas = cc)
return(answer)
}

# Calculo de las constantes criticas
CC.Gumbel(n = length(data),alpha=0.05)

# Calculo del estimador V
n <- length(data)
xj_r <- numeric(length=n)

```

```

i <- 1
while(i <= n){
  j      <- i:n
  m_ord  <- sort(data)
  xj_r[i] <- m_ord[i]*sum(1/j)
  i <- i+1
}
m_ord
xj_r
pi^2/6*(mean(data) - sum(xj_r)/n)^2/var(data)

# Calculo del estimador de Anderson Darling
AD.Test <- function(x, alpha){
  library(evd)
  n <- length(x)
  x <- sort(x, decreasing = FALSE)
  f <- function(p){
    1/n*sum(x) - (sum(x*exp(-x/p)))/(sum(exp(-x/p))) - p
  }
  Pgumbel <- function(q, loc = 0, scale = 1){
    answer = exp(-exp(-(q-loc)/scale))
    answer[scale <= 0] = NaN
    answer
  }
  itSol <- uniroot(f, c(-1000,1000), tol=1e-07, maxiter=1e+07)
  theta <- as.numeric(itSol$root)

  if(theta>0){
    xi      <- -theta*log(sum(exp(-x/theta)/n))
    parafit <- numeric(2)
    parafit[1] <- xi
    parafit[2] <- theta
    names(parafit) <- c("localidad", "escala")
    pit     <- numeric(n)
    pit[1:n] <- Pgumbel(x[1:n], parafit[1], parafit[2])
  }
}

```

```
aux <- numeric(n)
i <- 1
while(i <=n){
aux[i] <- (2*i-1)*(log(pit[i]) + log(1-pit[n+1-i]))
i <- i+1
}
A2 <- -sum(aux)/n-n
AD <- A2*(1+0.2/sqrt(n))
return(list(parameter_estimation = parafit, alpha=alpha, A_Cuadrada = A2,
Anderson_Darling = AD))
}
else{
names(theta) <- c("escala")
AD <- NaN
return(list(parameter_estimation = theta,
alpha=alpha, Anderson_Darling = AD))
}
}
```

AD.Test(data,alpha=0.05)