



# COLEGIO DE POSTGRADUADOS

---

---

INSTITUCIÓN DE ENSEÑANZA E INVESTIGACIÓN  
EN CIENCIAS AGRÍCOLAS

CAMPUS MONTECILLO

SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA  
ESTADÍSTICA

## ESTIMACIÓN DE LOS PARÁMETROS DE LA DISTRIBUCIÓN PARETO GENERALIZADA

ANNEL HURTADO JARAMILLO

T E S I S

PRESENTADA COMO REQUISITO PARCIAL PARA  
OBTENER EL GRADO DE:

DOCTORA EN CIENCIAS

MONTECILLO, TEXCOCO, EDO. DE MÉXICO  
2018

---

---

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

En adición al beneficio ético, moral y académico que he obtenido durante mis estudios en el Colegio de Postgraduados, el que suscribe Annel Hurtado Jaramillo, Alumno (a) de esta Institución, estoy de acuerdo en ser partícipe de las regalías económicas y/o académicas, de procedencia nacional e internacional, que se deriven del trabajo de investigación que realicé en esta institución, bajo la dirección del Profesor Javier Suárez Espinosa, por lo que otorgo los derechos de autor de mi tesis Estimación de los parámetros de la Distribución Pareto Generalizada

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

Montecillo, Mpio. de Texcoco, Edo. de México, a 17 de julio de 2018



Firma del  
Alumno (a)



Dr. Javier Suárez Espinosa  
Vo. Bo. del Consejero o Director de Tesis


La presente tesis titulada: **ESTIMACIÓN DE LOS PARÁMETROS DE LA DISTRIBUCIÓN PARETO GENERALIZADA**, realizada por la alumna: **ANNEL HURTADO JARAMILLO**, bajo la dirección del Consejo Particular indicado ha sido aprobada por el mismo y aceptada como requisito parcial para obtener el grado de:

## DOCTORA EN CIENCIAS

### SOCIOECONOMÍA, ESTADÍSTICA E INFORMÁTICA ESTADÍSTICA

#### CONSEJO PARTICULAR

CONSEJERO

  
Dr. Javier Suárez Espinoza

ASESOR

  
Dr. David Antonio Sotres Ramos

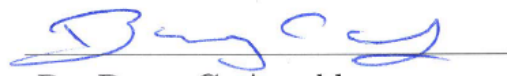
ASESOR

  
Dr. Gilberto Rendón Sánchez

ASESOR

  
Dr. Miguel Ángel Díaz Carreño

ASESOR

  
Dr. Barry C. Arnold

Montecillo, Texcoco, Estado de México,  
2018

# ESTIMACIÓN DE LOS PARÁMETROS DE LA DISTRIBUCIÓN PARETO GENERALIZADA

Annel Hurtado Jaramillo, DRA

Colegio de Postgraduados, 2018

## RESUMEN

La distribución Pareto Generalizada (GPD) ha sido extensamente usada dentro de la teoría de valores extremos para modelar valores que sobre pasan un umbral establecido. Para poder modelar este tipo de eventos es necesario estimar los parámetros de la GPD; sin embargo, los métodos de estimación existentes para esta distribución tienen algunas desventajas que pueden presentarse en situaciones prácticas. En este trabajo se propone un algoritmo basado en una malla, para obtener el estimador de mínima distancia (MD), utilizando una medida de distancia cuadrática entre la función de distribución teórica y la distribución empírica de la GPD. Un estudio de simulación muestra que la propuesta tiene un buen desempeño en el intervalo paramétrico considerado, y que no presenta los problemas que los estimadores clásicos y estimadores propuestos recientemente tienen. La propuesta para la estimación de los parámetros de la GPD puede ser una buena alternativa para el caso de otras distribuciones en las cuales los estimadores tradicionales presentan problemas.

**Palabras clave:** Mínima distancia; búsqueda en malla; función de distribución empírica.

# PARAMETER ESTIMATION FOR THE GENERALIZED PARETO DISTRIBUTION

Annel Hurtado Jaramillo, DRA

Colegio de Postgraduados, 2018

## ABSTRACT

The generalized Pareto distribution (GPD) has been widely used within the extreme value theory to model values that exceed an established threshold. In order to model this type of events, it is necessary to estimate the parameters of the GPD; however, the current estimation methods used for this distribution have some disadvantage which may appear in practical situations. In this paper we propose an algorithm, based on a grid, to obtain the minimum distance estimator (MD), using a quadratic distance measure between the theoretical distribution function and the empirical distribution of the GPD. Through a simulation study, we found that the proposal has good performance in the considered parametric interval, without showing the problems that classical estimators and recently proposed estimators have. The estimation proposal for the GPD could be a good alternative for the case of other distributions where the traditional estimators have troubles.

**Key words:** Minimum distance; grid search; Empirical distribution function.

## AGRADECIMIENTOS

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado para la realización de mis estudios de doctorado.

Al Colegio de Postgraduados por haberme brindado la gran oportunidad de continuar con mi preparación académica.

Expreso mi más profundo agradecimiento al Dr. Javier Suárez Espinosa por sus invaluable conocimientos, su tiempo y el gran apoyo que me brinda para la culminación de este trabajo.

Agradezco a los integrantes de mi Consejo Particular, los doctores David A. Sotres Ramos, Gilberto Rendón Sánchez, Miguel Ángel Díaz Carreño y Barry C. Arnold, por su tiempo y sus valiosos comentarios en la realización de este trabajo.

A los integrantes del personal administrativo del PISEI por su amabilidad y apoyo que siempre me brindaron.

## CONTENIDO

<b>RESUMEN</b>	<b>iv</b>
<b>ABSTRACT</b>	<b>v</b>
<b>AGRADECIMIENTOS</b>	<b>vi</b>
<b>LISTA DE TABLAS</b>	<b>viii</b>
<b>LISTA DE FIGURAS</b>	<b>x</b>
<b>INTRODUCCIÓN GENERAL</b>	<b>1</b>
Objetivos . . . . .	3
Revisión de Literatura . . . . .	3
<b>1. PARAMETER ESTIMATION FOR THE GENERALIZED PARETO DISTRIBUTION</b>	<b>6</b>
1.1. Abstract . . . . .	6
1.2. Resumen . . . . .	6
1.3. Introduction . . . . .	7
1.4. Generalized Pareto Distribution . . . . .	8
1.5. Parameter estimation of the GPD . . . . .	8

1.5.1. Numerical methods and proposal . . . . .	11
1.6. Comparison study . . . . .	12
1.7. Conclusions . . . . .	17
1.8. Algorithm . . . . .	17
<b>CONCLUSIONES GENERALES</b>	<b>20</b>
<b>REFERENCIAS</b>	<b>20</b>
<b>ANEXO</b>	<b>24</b>



## LISTA DE TABLAS

1.1. MSE of $\hat{\sigma}$ and $\hat{k}$ based on B=1,000 Monte Carlo replicates, n=50 for the estimation using the ML. . . . .	13
1.2. MSE of $\hat{\sigma}$ and $\hat{k}$ based on B=1,000 Monte Carlo replicates. Results obtained taking $(\sigma = \bar{x}, k = 0)$ as starting point. . . . .	14
1.3. MSE of $\hat{\sigma}$ and $\hat{k}$ based on B=1,000 Monte Carlo replicates. . . . .	15
1.4. MSE of $\hat{\sigma}$ and $\hat{k}$ based on B=1,000 Monte Carlo replicates. . . . .	16

## LISTA DE FIGURAS

1.	Clasificación de los riesgos . . . . .	1
1.1.	Comparison in terms of MSE of the estimations of $\hat{k}$ and $\hat{\sigma}$ , $n=100$ , according to data by Luceño (2006). . . . .	13
1.2.	Grid searching steps. . . . .	19

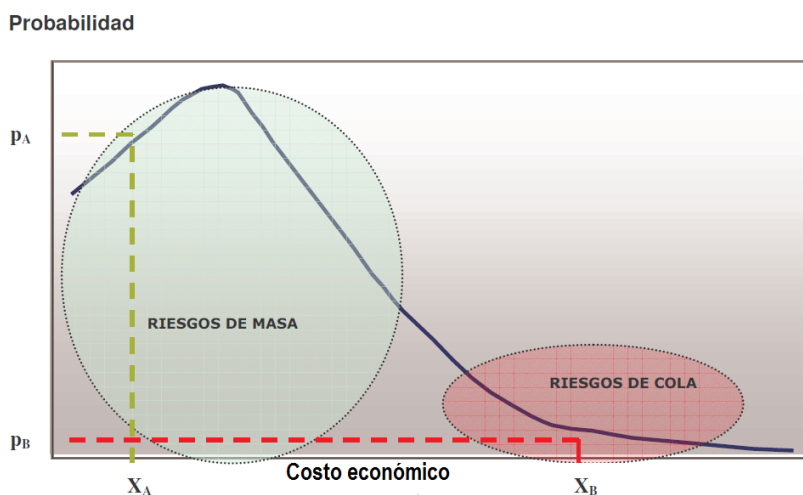
## INTRODUCCIÓN GENERAL

En los últimos años, debido al cambio climático las instituciones de seguros han sido un eslabón importante para el desarrollo de los países. La existencia de estas instituciones depende en gran medida del uso de metodologías estadísticas para la medición del riesgo de eventos que pueden implicar desequilibrios en su estado financiero.

Desde el punto de vista del seguro ([García-Pérez, 2007](#)) los riesgos se pueden clasificar en:

- Riesgos de masa
- Riesgos de cola

Los primeros son riesgos con consecuencias económicas bajas pero alta probabilidad de ocurrencia, los segundos son aquellos riesgos con baja frecuencia pero elevadas consecuencias económicas (ver figura 1). Las instituciones de seguros deben tener un monto



**Figura 1:** Clasificación de los riesgos

de recursos para hacer frente a alguna situación adversa en sus obligaciones y de esta

forma ser considerada “solvente” desde un punto de vista regulatorio. Luego entonces un elemento central para la solvencia de una aseguradora es el *riesgo*.

Para el análisis de los riesgos de masa son aplicables las técnicas estadísticas basadas en la Ley de los Grandes Números o el Teorema Central del Límite, mientras que para los riesgos de cola es necesaria la metodología de valores extremos.

En el campo actuarial los grandes siniestros son considerados como valores extremos, los cuales en estadística pueden ser modelados con distribuciones conocidas como distribuciones de cola pesada. Un enfoque tradicional para el análisis de valores extremos en una población dada se basa en la familia de la distribución generalizada de valores extremos (DGVE), dentro de esta familia se encuentran las distribuciones Gumbel, Frechet y Weibull. Otra distribución ampliamente utilizada para modelar este tipo de eventos es la distribución Pareto generalizada (GPD). La GPD es una distribución que modela valores que sobrepasan un umbral establecido conocidos como excedencias, para las empresas de seguros es importante modelar este tipo de eventos ya que éstos podrían poner en riesgo la estabilidad financiera de la entidad aseguradora. La GPD además de emplearse en el sector asegurador tiene numerosas aplicaciones en áreas como finanzas, hidrología y medio ambiente (Reiss y Thomas, 2007).

Para poder modelar eventos con distribuciones de densidad o de probabilidad generalmente es necesario estimar sus parámetros. Para el caso de la GPD se han empleado los métodos tradicionales; máxima verosimilitud (ML) ha sido considerado por Davison (1984) y (Smith, 1984) y el método de momentos (MOM) y momentos ponderados (PWM) por Hosking y Wallis (1987); sin embargo, estos métodos presentan algunos problemas. Por ejemplo, en el caso de ML, en ocasiones puede no existir la solución a las ecuaciones de derivadas parciales y cuando si existe solución se pueden presentar problemas de convergencia (Hosking y Wallis, 1987).

En el caso de los estimadores de MOM así como los de PWM se ha encontrado que estos son sensibles a la elección del umbral y algunas veces se pueden obtener estimaciones incorrectas (Dupuis, 1999), además de que estos estimadores pueden producir estimaciones fuera del espacio paramétrico (Deidda y Puliga, 2009). Debido a estos problemas se han propuesto diversas modificaciones de estos métodos (de Zea Bermudez y Kotz, 2010a), tales como: L-momentos, LH-momentos o L-momentos parciales.

De Zea-Bermudez y Kotz (2010a, 2010b) concluyen que es muy complicado recomendar

un método en particular. Lo anterior, es debido a que todos los métodos mencionados anteriormente tienen alguna desventaja que puede presentarse en alguna situación práctica. Además, los autores indican que los estudios de simulación obtenidos en la revisión hacen difícil la comparación de los métodos estudiados, ya que fueron obtenidos en condiciones diferentes; además, el desempeño de algunos métodos de estimación fue obtenido para muestras pequeñas y bajo la restricción del espacio paramétrico del parámetro de forma. Asimismo, ellos recomiendan realizar estudios de simulación para aquellos métodos de estimación que no se han estudiado, considerando la información que proveen.

Por lo anterior, en este trabajo de investigación se propone usar un algoritmo de búsqueda en malla para obtener el estimador de mínima distancia (MD). Dicha propuesta soluciona en gran medida los problemas de estimación antes mencionados.

## **Objetivos**

### **Objetivo general**

- Estimar los parámetros de la GPD usando el método de búsqueda en malla para la obtención del estimador de mínima distancia (MD).

### **Objetivos específicos**

- Proporcionar un algoritmo para la búsqueda en malla del estimador de la GPD.
- Realizar la comparación entre el método de MD bajo el algoritmo propuesto y algunos de los métodos de estimación sugeridos en la literatura.

## **Revisión de Literatura**

La GPD fue introducida por [Pickands \(1975\)](#) para modelar valores que sobrepasan un umbral establecido, esta distribución ha sido ampliamente utilizada para modelar datos de diversas áreas ya que con ella se pueden modelar colas de una amplia gama de distribuciones ([Lee, 2009](#)). Algunas de sus aplicaciones las podemos encontrar por ejemplo, para

modelar los niveles altos de ríos (Davison y Smith, 1990) y para modelar inundaciones (Reiss y Thomas, 2007). En el área de seguros y finanzas se ha empleado para ajustar de manera óptima siniestros extremos (García-Pérez, 2007) o para calcular el VaR (Arneric *et al.*, 2012, Jockovic, 2012).

Uno de los métodos que se ha empleado para estimar los parámetros de la GPD es el método de mínima distancia (MD), el cual tiene sus orígenes en Wolfowitz (1953, 1957). Los estimadores de MD son obtenidos minimizando la distancia entre la función de distribución empírica y la función de distribución de alguna familia paramétrica. Parr (1981) destaca que este método es atractivo debido a dos características principales: la primera es su fácil implementación, dado un conjunto de datos, un modelo paramétrico y una medida de distancia entre funciones de distribución, ya que solo se necesita programar una rutina de minimización para calcular el estimador. La segunda es que los estimadores de mínima distancia proporcionan resultados incluso si el modelo paramétrico supuesto es incorrecto. El método de MD se puede aplicar a muchas definiciones de distancia y tiene la propiedad de obtener estimadores consistentes.

Para la obtención del estimador de mínima distancia es necesario realizar la búsqueda del mínimo de una función, lo cual es conocido en otros contextos como optimización.

Entre los métodos que se usan para encontrar los valores que minimizan una función se tienen el método Nelder-Mead, los métodos cuasi-Newton como el algoritmo BFGS (publicado simultáneamente en 1970 por Broyden, Fletcher, Goldfarb y Shanno), entre otros.

El algoritmo de Nelder-Mead se basa en un método geométrico de programación no lineal que usa poliedros de dimensión  $n+1$  en cuyos vértices se tiene la aproximación al óptimo global. Para el caso de la GPD, si el valor inicial que se le proporciona a la función, para el caso  $k < 0$ , es menor al valor verdadero del parámetro, entonces en algunos casos el algoritmo no podrá encontrar el óptimo.

Por otro lado, para poder implementarse el algoritmo BFGS se requiere del gradiente de la función a minimizar y de un punto de inicio, cuando el gradiente no se proporciona el método usa diferencias finitas para estimarlo. Cuando se emplea este algoritmo para hallar estimaciones de los parámetros de la GPD se pueden presentar algunos problemas:

- Si la muestra generada proviene de una GPD con parámetros ( $\sigma > 0$ ,  $k > 0$ ) y en

el punto inicial  $(\sigma_0, k_0)$  que se proporciona,  $k_0 > k$ , entonces existen problemas de convergencia.

- Si la muestra de la cual se desea estimar los parámetros viene de una GPD con  $(\sigma > 0, k < 0)$  y el punto inicial proporcionado  $(\sigma_0, k_0)$  hace que  $k_0 x / \sigma_0 > 1$ , entonces al igual que en el caso anterior se tienen problemas de convergencia.

Debido a los problemas que se presentan en este tipo de algoritmos se propone usar la técnica de búsqueda en malla para obtener las estimaciones por el método de MD de los parámetros de la GPD. Esta técnica consiste en dividir las dimensiones de  $x$  y  $y$  en pequeños incrementos para crear una malla; después la función se evalúa en cada nodo (entre más densa es la malla la probabilidad de localizar el punto óptimo es mayor) y finalmente se obtiene el valor que minimiza la función. Esta técnica funciona en discontinuidades y funciones no diferenciables.

# CAPÍTULO 1. PARAMETER ESTIMATION FOR THE GENERALIZED PARETO DISTRIBUTION

## 1.1 Abstract

The generalized Pareto distribution (GPD) has been widely used within the extreme value theory to model values that exceed an established threshold. In order to model this type of events, it is necessary to estimate the parameters of the GPD; however, the current estimation methods for this distribution have some disadvantage which may appear in practical situations. In this paper we propose an algorithm, based on a grid, to obtain the minimum distance estimator (MD), using a quadratic distance measure between the theoretical distribution function and the empirical distribution of the GPD. Through a simulation study, the proposal has good performance in the parametric interval considered, without showing the problems that classical estimators and recently proposed estimators have.

Key Words: Minimum distance, grid search, empirical distribution function.

## 1.2 Resumen

La distribución Pareto Generalizada (GPD) ha sido extensamente usada dentro de la teoría de valores extremos para modelar valores que sobre pasan un umbral establecido. Para poder modelar este tipo de eventos es necesario estimar los parámetros de la GPD; sin embargo, los métodos de estimación existentes para esta distribución tienen algunas desventajas que pueden presentarse en situaciones prácticas. En este trabajo se propone un algoritmo basado en una malla, para obtener el estimador de mínima distancia (MD), utilizando una medida de distancia cuadrática entre la función de distribución teórica y la distribución empírica de la GPD. Un estudio de simulación muestra que la propuesta tiene un buen desempeño en el intervalo paramétrico considerado, y que no presenta los problemas que los estimadores clásicos y estimadores propuestos recientemente tienen.



Palabras clave: Mínima distancia, búsqueda en malla, función de distribución empírica.

### 1.3 Introduction

The extreme value theory is a branch of statistics that deals with the study of probabilistic and statistical issues related with very high or very low values in sequences of random variables. In this area, it is common to model values that exceed an established threshold, known as exceedances. For example, in the case of insurances, it is very important to model values that exceed a determined value (exceedances above the threshold), since this type of events could risk the stability and financial profitability of an insurance firm.

The generalized Pareto distribution (Pickands, 1975) has been widely used to model exceedances above a threshold; for example, it has been used to model water flow exceedances in rivers (Davison y Smith, 1990), wave heights (Castillo y Hadi, 1997), daily maximum temperature data (Plavcová y Kyselý, 2011), insurance claims (Embrechts *et al.*, 1997), and to calculate the risk value in finances (Jockovic, 2012), those quoted by De Zea-Bermudez and Kotz (2010a) among others.

Nevertheless, users who have modeled with the generalized Pareto distribution (GPD) have often had problems estimating its parameters, and since the estimation of probabilities of this type of events depends on these estimates, the problem of estimation becomes very relevant.

The importance of the estimation issue in the GPD can be showed in two articles by De Zea-Bermudez and Kotz (2010a, 2010b), who expose an extensive discussion of the methods used to estimate the parameters of the GPD. Among these, the followings can be mentioned: the method of moments (MOM), the probability-weighted moments method (PWM), and generalizations of these methods like L-moments, LH-moments, and partial L-moments; also, the maximum likelihood method, and Bayesian methods, among others.

De Zea-Bermudez and Kotz (2010a, 2010b) conclude that it is very complicated to give preference to any particular method. This is because all the studied methods have some disadvantages that can show up in a practical situation. In addition, the authors indicate that the simulation studies obtained in the discussion make difficult to compare the methods used, as they were obtained under different conditions; moreover, the performance of some estimation methods was obtained for small samples and under res-

triction of the shape parametric space. They recommend to carry out simulation studies for those estimation methods that have not been studied, considering the information they provide.

It is clear that an estimation method with good performance and that works well under very diverse practical situations would be desirable. In this research, we propose an algorithm, based on a grid, to obtain an estimator by the minimum distance (MD) method and we present a simulation study to measure its performance, comparing it against classic estimators and recently proposed estimators. In section 1.4 a brief description of the GPD and its characteristics is presented. Section 1.5 includes the estimation methods for the GPD used in this research. Section 1.6 contains a simulation study to measure the performance of the estimators. The conclusions are presented in Section 1.7.

## 1.4 Generalized Pareto Distribution

A random variable  $X$  is said to have a generalized Pareto distribution if its distribution function is given by

$$F(x; k, \sigma) = 1 - \left(1 - \frac{kx}{\sigma}\right)^{1/k}, \quad (1.1)$$

where  $\sigma > 0$  and  $k \in \mathbb{R}$  are scale and shape parameters, respectively; the support of the function are the real positives ( $x > 0$ ) when  $k \leq 0$  and  $0 \leq x \leq \sigma/k$  for  $k > 0$ .

One important characteristic of this distribution is that if  $k$  tends to 0, the GPD is reduced to the exponential distribution with a mean  $\sigma$ , while, when  $k = 1$ , the GPD becomes a uniform distribution in the range  $(0, \sigma)$ , for the case  $k < 0$ , we have a Pareto II distribution ([Arnold, 2015](#)).

## 1.5 Parameter estimation of the GPD

*Maximum Likelihood (ML) Method.*

It is a widely adopted standard estimation technique that can be applied to diverse distributions. In the case of the GPD, several authors have proven that the ML estimator is the best estimator in the presence of large samples, but in small samples ( $\leq 100$ ),

convergence problems arise and the method can be outmatched by other techniques (Hosking y Wallis, 1987). The log-likelihood function for the GPD is given by

$$\ln L(k, \sigma; x) = -n \ln \sigma + \left( \frac{1}{k} - 1 \right) \sum_{i=1}^n \ln \left( 1 - \frac{kx_i}{\sigma} \right). \quad (1.2)$$

The method consists of maximizing the equation 1.2, solving the equations of partial derivatives with respect to each unknown parameter, then it is necessary to use numerical methods; however, as mentioned by Hosking y Wallis (1987), the algorithms used to obtain the ML estimations could present convergence problems, even in large sized samples. Another feature to point out is that when  $k > 1$  and  $x_i \rightarrow \sigma/k$ , the ML estimators do not exist, since  $\lim L(k, \sigma; X) = \infty$  (Choulakian y Stephens, 2001).

*Method of Moments (MOM) and Probability-Weighted Moments (PWM).*

Hosking y Wallis (1987) proposed the MOM and the PWM for GPD parameters, the estimators of moments of  $k$  and  $\sigma$  are given by

$$\hat{k} = \frac{1}{2} \left( \frac{\bar{x}^2}{s^2} - 1 \right) \quad y \quad \hat{\sigma} = \frac{\bar{x}}{2} \left( 1 + \frac{\bar{x}^2}{s^2} \right)$$

where  $\bar{x}$  and  $s^2$  are the sample mean and variance, respectively.

On the other hand, the estimators of PWM of the parameters of the GPD are

$$\hat{k} = \frac{a_0}{a_0 - 2a_1} - 2 \quad y \quad \hat{\sigma} = \frac{2a_0a_1}{a_0 - 2a_1} \quad (1.3)$$

where

$$a_s = \frac{1}{n} \sum_{i=1}^n x_{i:n} (1 - p_{i:n})^s, \quad (1.4)$$

with  $s = 0$  and  $1$ . Landwehr *et al.* (1979) recommend using  $p_{i:n} = (1 + \gamma) / (n + c)$ , with  $\gamma = -0.35$  and  $c = 0$ .

However, in the case of the GPD, these estimators only exist when  $k > -1/2$  (Hosking y Wallis, 1987). Besides, the MOM has shown to be sensitive to the presence of atypical values (Deidda y Puliga, 2009).

### *Minimum Distance Method*

Another method that has been used to obtain the estimation of the parameters of the GPD is the minimum distance method (MD). The origin of the MD method goes back to Wolfowitz (1953, 1957). In these articles, Wolfowitz developed the method as a means to obtain consistent estimators when classic methods fail. The idea upon which this method is based on is presented here.

In a classic estimation problem we have a random sample,  $X_1, \dots, X_n$ , with distribution  $F(x, \boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  is unknown and we wish to estimate it. In this case, Wolfowitz' MD method starts by defining a distance measurement among two distribution functions

$$d(F(\cdot, \boldsymbol{\theta}) - F_n(\cdot)),$$

where  $F_n(x)$  is the empirical distribution function defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(x - x_i)$$

where  $I$  is the indicator function, defined as

$$\begin{aligned} I(x) &= 1 && \text{si } x - x_i \geq 0 \\ &= 0 && \text{en otro caso} \end{aligned}$$

After this, the minimum distance among all distances is taken, this is

$$d_{min}(F(\mathbf{x}, \boldsymbol{\theta}^*) - F_n(\mathbf{x})) = \min_{\boldsymbol{\theta} \in \Theta} d(F(\cdot, \boldsymbol{\theta}) - F_n(\cdot))$$

and we take  $\boldsymbol{\theta}^*$  as the MD estimator for  $\boldsymbol{\theta}$ . The method can be applied to different distance measures.

The MD method has been used by Luceño (2006) and Chen *et al.* (2016). Luceño (2006) uses the classic Kolmogorov-Smirnov (KS), Cramer-Von Mises (CM), and Anderson-Darling (AD) statistics, which are used to carry out goodness of fit tests in homogeneous populations. Moreover, he introduces modified statistics like the Anderson Darling right-tail (ADR) and left-tail (ADL), as well as the second order Anderson Darling statistics (AD2R and AD2L). Luceño, in his simulation study, shows that the MD method outperforms classic estimation methods in terms of mean square error. However, the estimations

obtained from these distance measures have occasionally been found not to be satisfactory when the theoretical distribution implies a shape parameter (Huber, 1972, Parr y Schucany, 1980), as is the case of the GPD.

Meanwhile, Chen *et al.* (2016) base their study on the idea of the residual analysis of a regression, they use a robust estimator, the  $M$  estimator (MDM) to find a  $\boldsymbol{\theta}$  that satisfies

$$\tilde{\boldsymbol{\theta}}_n = \operatorname{argmin} \frac{1}{n} \sum_{i=1}^n \rho[r_i(\boldsymbol{\theta})] \quad (1.5)$$

where  $\rho(\cdot)$  is the Tukey biweight function (Yohai y Zamar, 1988). This algorithm depends on its starting point.

Besides the MDM estimator, they also propose a weighted  $M$  estimator (MDW), and define it as

$$\tilde{\boldsymbol{\theta}}_n^* = \operatorname{argmin} \frac{1}{n} \sum_{i=1}^n \rho[r_i^*(\boldsymbol{\theta})] \quad (1.6)$$

where  $r_i^*(\boldsymbol{\theta}) = r_i(\boldsymbol{\theta})/w_i$  and the  $w_i$ 's are the weights, given in turn by  $w_i(\tilde{\boldsymbol{\theta}}_n) = \sqrt{F_{\tilde{\boldsymbol{\theta}}_n}(x_i)(1 - F_{\tilde{\boldsymbol{\theta}}_n}(x_i))}$  where  $\tilde{\boldsymbol{\theta}}_n$  is the MDM estimator obtained in 1.5.

As it can be seen, the methods mentioned above have some disadvantages that can come up in a practical situation.

### 1.5.1 Numerical methods and proposal

To obtain the maximum likelihood estimator and the minimum distance estimator, it is necessary, as the name indicates, to search for the maximum or minimum of a function. This is known in other contexts as optimization. Among the methods that search for the values that minimize a function, we can mention the Nelder-Mead method, as well as quasi-Newton methods like the BFGS algorithm (published simultaneously by Broyden, Fletcher, Goldfarb, and Shanno in 1970).

The Nelder-Mead algorithm is based on a non-linear programming geometric method that uses dimension polyhedrons  $n+1$  in whose vertices is the approximation to the global optimum. If the starting value given to the function, in the case  $k < 0$ , is lower than  $k$ , in some cases the algorithm will not be able to find the optimum. Moreover, in order to

the BFGS algorithm to be implemented, the gradient of the function to be minimized and a starting point are required . When the gradient is not provided, the method uses finite differences to estimate it. When this algorithm is used to find estimations of the parameters of the GPD, some problems can arise:

- If a sample generated comes from a GPD with parameters  $\sigma > 0$  and  $k > 0$ , and the starting point,  $(\sigma_0, k_0)$ , is such as  $k_0 > k$ , then there will be convergence problems.
- If a sample comes from a GPD with  $\sigma > 0$  and  $k < 0$ , and the starting point provided  $(\sigma_0, k_0)$  makes  $k_0x/\sigma_0 > 1$ , then like in the previous case, there will be convergence problems.

Due to these problems that arise from this type of algorithms, in this paper we propose to use the grid search technique to obtain the estimation of the parameters of the GPD through the MD method. The grid search technique works in discontinuities and non-differentiable functions. It consists of dividing the  $x$  and  $y$  dimensions into small increments to create a grid; the function is evaluated in each node: the denser the grid, the higher the probability of locating the optimum point.

We found that by using the grid search technique there are no convergence problems and we can do the estimations of  $\sigma$  and  $k$  in all the studied parametric space. The description of the proposed methodology is found in Algorithm 1 and Figure 1.2 (see Appendix).

## 1.6 Comparison study

In this section the performance of the proposed algorithm is presented and compare it against other methods suggested in the literature.

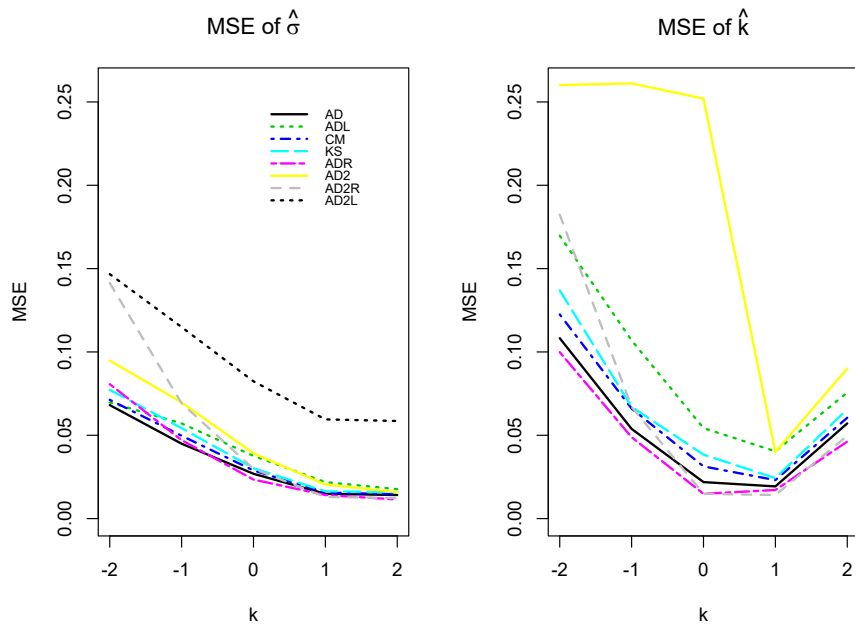
A simulation study is done under the following conditions: the following  $k$  values are taken  $k = \{-2, -1, -0.4, -0.2, 0.001, 0.2, 0.4, 1, 2\}$  (they are the most common values observed in practical situations) and three sample sizes are considered  $n = \{15, 50, 100\}$ . Given that the results do not vary with respect to the scale parameter (Hosking y Wallis, 1987),  $\sigma = 1$  is set. The results are based on 1000 Monte Carlo replicates.

For each combination of  $\sigma$ ,  $n$  and  $k$ , the estimations of the parameters  $k$  and  $\sigma$  are done using the MD method under Algorithm 1 (see Appendix), and their performance is compared in terms of the Mean Squared Error (MSE) against the AD, ADR, MOM, PWM, MDM, and MDW methods. The ML method is not included in the comparison, since as Table 1.1 shows, the convergence percentage for values of  $k = \{0.2, 0.4\}$ , is under 30 % when  $n = 50$ .

**Table 1.1:** MSE of  $\hat{\sigma}$  and  $\hat{k}$  based on B=1,000 Monte Carlo replicates, n=50 for the estimation using the ML.

Estimator	$k$								
	-2	-1	-0.4	-0.2	0.001	0.2	0.4	1	2
$\hat{k}$	31.498	0.085	0.048	0.036	0.033	0.029	0.027	-	-
$\hat{\sigma}$	0.160	0.099	0.070	0.057	0.060	0.061	0.034	-	-
	(100 %)	(100 %)	(100 %)	(100 %)	(93.2 %)	(26.7 %)	(7.8 %)	-	-

From the statistics proposed by Luceño (2006), only AD and ADR are used for the comparison study, because in general these statistics have a better performance in comparing against KS, CM, ADL, AD2R and AD2L statistics (see Figure 1.1).



**Figure 1.1:** Comparison in terms of MSE of the estimations of  $\hat{k}$  and  $\hat{\sigma}$ , n=100, according to data by Luceño (2006).

For calculating AD and ADR, the `gpdmgf` function, which can be found in the POT

library of the R software, was used. For evaluating this function and find  $\hat{\sigma}$  and  $\hat{k}$ , a starting point is necessary. If the starting point is near the real value of  $\sigma$  and  $k$ , these estimators have a good performance; however, when the starting point is far from the real values, there are erroneous estimations of the parameters and convergence problems, especially for values of  $k \leq -1$  (see Table 1.2); i. e. these methods (AD and ADR) show good performance only when *a priori* knowledge of the parameters is given. The predetermined starting values assigned by the function `gpdmgf` to the AD and ADR methods are  $(\sigma = \bar{x}, k = 0)$ .

**Table 1.2:** MSE of  $\hat{\sigma}$  and  $\hat{k}$  based on B=1,000 Monte Carlo replicates. Results obtained taking  $(\sigma = \bar{x}, k = 0)$  as starting point.

Estimator	n	Method	k								
			-2	-1	-0.4	-0.2	0.001	0.2	0.4	1	2
$\hat{k}$	15	AD	1.810	0.395	0.247	0.214	0.206	0.186	0.192	0.252	0.558
$\hat{\sigma}$			0.771	0.383	0.262	0.217	0.216	0.174	0.160	0.135	0.120
Convergence			(100 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)
$\hat{k}$	15	ADR	0.706	0.349	0.205	0.170	0.146	0.140	0.147	0.191	0.438
$\hat{\sigma}$			2.835E+04	0.523	0.289	0.245	0.214	0.184	0.169	0.129	0.101
Convergence			(99.9 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)
$\hat{k}$	50	AD	2.312	0.181	0.063	0.050	0.042	0.036	0.033	0.046	0.114
$\hat{\sigma}$			1.816E+10	1.668E06	0.096	0.058	0.052	0.045	0.042	0.033	0.027
Convergence			(67.4 %)	(93.4 %)	(99.9 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)
$\hat{k}$	50	ADR	1.382	0.145	0.047	0.041	0.032	0.028	0.026	0.036	0.104
$\hat{\sigma}$			6.930E+10	3.036E06	0.069	0.056	0.047	0.042	0.037	0.029	0.026
Convergence			(70 %)	(93.6 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)
$\hat{k}$	100	AD	3.062	0.320	0.030	0.023	0.019	0.016	0.015	0.021	0.049
$\hat{\sigma}$			7.78E+19	1.081E+07	0.056	0.030	0.027	0.021	0.019	0.016	0.012
Convergence			(24.2 %)	(72 %)	(99.5 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)
$\hat{k}$	100	ADR	3.095	0.293	0.027	0.020	0.016	0.012	0.010	0.015	0.044
$\hat{\sigma}$			5.83E+12	9.242E+06	0.059	0.029	0.022	0.019	0.015	0.012	0.011
Convergence			(23.4 %)	(74.1 %)	(99.1 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)	(100 %)

Table 1.3 shows the comparison in terms of MSE of the MD method, using the methodology proposed by Algorithm 1 versus the AD, ADR, MOM, PWM, MDM, and MDW methods.

For the  $k$  case, with a sample size of  $n = 15$ , we find that the best estimation method in terms of MSE is PWM in the  $-1 \leq k \leq -0.2$  interval, while the MOM method has the lowest MSE in the  $-0.2 < k < 0.4$  interval. For  $-2 \leq k < -1$  and  $0.4 \leq k \leq 2$ , the best estimator is ADR. When  $n = 50$ , the MOM and PWM methods are still the best in the mentioned intervals for  $n = 15$ ; however, we can notice that for  $k < -1$ , the ADR method shows very high MSE values, and the same behavior can be seen when  $n = 100$  and  $k < -0.4$ . This shows that ADR is sensitive to the starting values that it is assigned,



and that when these values are far from the real parameter value, the estimations are not good. Table 1.3 shows that if  $n = 15$  and  $50$ , the AD method has a higher MSE for  $k < -1$  and that the MOM and PWM methods have a very high MSE when  $k < -1$  and  $k > 1$  for  $n = 50$  or  $100$ . The estimations carried out through the MD, MDM, and MDW methods have a similar performance when  $n = 50$  or  $100$ . From these methods, MDW appears to be the best.

**Table 1.3:** MSE of  $\hat{\sigma}$  and  $\hat{k}$  based on B=1,000 Monte Carlo replicates.

Estimator	$n$	Method	$k$									
			-2	-1	-0.4	-0.2	0	0.2	0.4	1	2	
$\hat{k}$	15	AD	1.810	0.395	0.247	0.214	0.206	0.186	0.192	0.252	0.558	
		ADR	0.706	0.349	0.205	0.170	0.146	0.140	0.147	0.191	0.438	
		MOM	2.592	0.530	0.137	0.114	0.103	0.114	0.186	0.719	5.373	
		PWM	1.506	0.268	0.125	0.113	0.117	0.137	0.186	0.377	1.225	
		MDM	0.921	0.461	0.297	0.276	0.239	0.198	0.187	0.230	0.534	
		MDW	1.018	0.640	0.446	0.522	0.380	0.444	0.382	0.501	0.739	
		MD	0.865	0.580	0.454	0.424	0.408	0.401	0.391	0.434	0.639	
	50	AD	2.312	0.181	0.063	0.050	0.042	0.036	0.033	0.046	0.114	
		ADR	1.382	0.145	0.047	0.041	0.032	0.028	0.026	0.0366	0.104	
		MOM	2.350	0.358	0.041	0.028	0.023	0.023	0.032	0.114	0.467	
		PWM	1.180	0.112	0.037	0.030	0.028	0.037	0.044	0.102	0.319	
		MDM	0.271	0.145	0.089	0.072	0.063	0.048	0.041	0.052	0.120	
		MDW	0.222	0.141	0.069	0.057	0.055	0.048	0.043	0.055	0.126	
		MD	0.264	0.145	0.098	0.087	0.079	0.074	0.072	0.082	0.144	
	100	AD	3.062	0.320	0.030	0.023	0.019	0.016	0.015	0.021	0.049	
		ADR	3.095	0.293	0.027	0.0198	0.016	0.012	0.010	0.015	0.044	
		MOM	2.299	0.311	0.025	0.015	0.010	0.010	0.015	0.051	0.219	
		PWM	8.717	3.115	0.589	0.155	0.014	0.183	0.672	4.076	16.271	
		MDM	0.121	0.066	0.046	0.037	0.031	0.023	0.020	0.024	0.061	
		MDW	0.106	0.054	0.031	0.026	0.021	0.018	0.016	0.022	0.059	
		MD	0.128	0.063	0.042	0.037	0.033	0.031	0.035	0.035	0.066	
	$\hat{\sigma}$	15	AD	0.771	0.383	0.262	0.217	0.216	0.174	0.160	0.135	0.120
			ADR	2.834E+04	0.523	0.289	0.245	0.214	0.184	0.169	0.129	0.101
			MOM	4.616E+09	354.242	0.456	0.228	0.210	0.179	0.228	0.381	1.034
PWM			2.264E+09	14.922	0.262	0.234	0.181	0.206	0.202	0.212	0.262	
MDM			1.254	0.501	0.314	0.260	0.222	0.181	0.172	0.133	0.117	
MDW			0.692	0.322	0.297	0.250	0.210	0.194	0.182	0.146	0.132	
MD			0.837	0.401	0.301	0.281	0.255	0.232	0.211	0.171	0.131	
50		AD	1.816E+10	1.669E+03	0.096	0.058	0.052	0.045	0.042	0.033	0.027	
		ADR	6.930E+10	3.036E+03	0.069	0.056	0.047	0.042	0.037	0.029	0.026	
		MOM	9.000E+23	1822.305	0.143	0.058	0.043	0.040	0.043	0.060	0.084	
		PWM	1.097E+09	0.821	0.061	0.054	0.048	0.055	0.049	0.055	0.062	
		MDM	0.148	0.112	0.079	0.070	0.054	0.049	0.043	0.034	0.028	
		MDW	0.127	0.086	0.063	0.055	0.054	0.048	0.041	0.034	0.029	
		MD	0.156	0.102	0.078	0.071	0.065	0.060	0.055	0.044	0.033	
100		AD	7.782E+19	1.081E+04	0.056	0.030	0.027	0.021	0.019	0.016	0.012	
		ADR	5.831E+12	9.243E+03	0.059	0.029	0.022	0.019	0.015	0.012	0.011	
		MOM	2.359E+13	2.923E+06	0.069	0.029	0.022	0.018	0.021	0.028	0.040	
		PWM	1.654E+07	0.261	0.028	0.025	0.023	0.024	0.023	0.027	0.030	
		MDM	0.073	0.047	0.037	0.035	0.028	0.025	0.022	0.017	0.015	
		MDW	0.063	0.040	0.029	0.026	0.023	0.023	0.020	0.016	0.014	
		MD	0.076	0.045	0.036	0.033	0.030	0.027	0.027	0.021	0.016	

With respect to the estimations for  $\sigma$ , if  $n = 15$ , the AD, MD, and MDW methods have a

similar performance in all the parametric space considered, while for this sample size and  $k < -1$ , the ADR method has a higher MSE than the mentioned methods. For  $n = 50$  or  $100$  and  $k < -0.4$ , the AD method begins to have very high MSE values, like ADR. In all the sample sizes considered, if  $k < -0.4$ , the best estimator for  $\sigma$  is MDW, while for this same interval, the MOM and MPW estimators show high MSE values. We can also note that for the AD, ADR, MD, MDM, and MDW methods, the MSE decreases as the value of  $k$  increases.

If the comparison between estimation methods is done using the **minimax** criterion (see Table 1.4), we can notice that if the starting point is near the real parameter value and the sample size considered is  $n = 15$ , the best estimator for  $k$  is ADR, since it is the estimation method that has the minimum value of the maximum values of MSE presented in the estimation. The second best is AD, while if  $n = 50$  or  $100$ , the second best are AD and MDM. However, when the starting point is far from the real value of the parameters, the second best estimator is MD, outperformed only by ADR when  $n = 15$  and MDW when  $n = 50$ . For a sample size of  $n = 100$ , the best estimator is MDW while MD is the third best. For the case of  $\hat{\sigma}$ , we have that if  $n = 15$  or  $50$ , the best estimator, using the minimax criterion, is MDW, followed by AD (when the starting point is near the real value). If  $n = 100$ , the best estimator is AD, followed by ADR.

**Tabla 1.4:** MSE of  $\hat{\sigma}$  and  $\hat{k}$  based on B=1,000 Monte Carlo replicates.

Estimator	Method	$n = 15$	$n = 50$	$n = 100$	
$\hat{k}$	MD	0.865	0.264	0.128	
	MDM	0.921	0.271	0.121	
	MDW	1.018	0.222	0.106	
	When starting point is near to the real value				
	AD	0.740	0.222	0.106	
	ADR	0.689	0.203	0.101	
	When starting point is far from the real value				
	AD	1.810	2.312	3.062	
	ADR	0.706	1.382	3.095	
	MD	0.837	0.156	0.128	
	MDM	1.254	0.148	0.121	
	MDW	0.692	0.127	0.106	
$\hat{\sigma}$	When starting point is near to the real value				
	AD	0.793	0.144	0.068	
	ADR	1.293	0.190	0.101	
	When starting point is far from the real value				
	AD	0.771	1.816E+10	7.78E+19	
	ADR	2.835E+04	6.930E+10	5.83E+12	

In general, both for the estimation of  $k$  and  $\sigma$ , the MD method is competent in all sample sizes, with no convergence problems or estimations outside the parametric space considered. The MD method, unlike the AD, ADR, MDM, and MDW methods, does not require a specific starting point to carry out the algorithm.

## 1.7 Conclusions

The simulation study shows that there is no one estimation method whose performance is uniformly better for the parametric space studied. The performance of the proposal is among the three best methods studied, with the advantage that it has no convergence problems nor does it require a specific starting point *a priori* to get a good performance.

## 1.8 Algorithm

**Algorithm:** Obtention of the MD estimators through grid searching

Input: $n$ : sample size

$\sigma$ : value of the scale parameter of the GPD

$k$ : value of the shape parameter of the GPD

$B$ : number of Monte Carlo replicates

Output: $\hat{\sigma}$ : estimation of the  $\sigma$  parameter

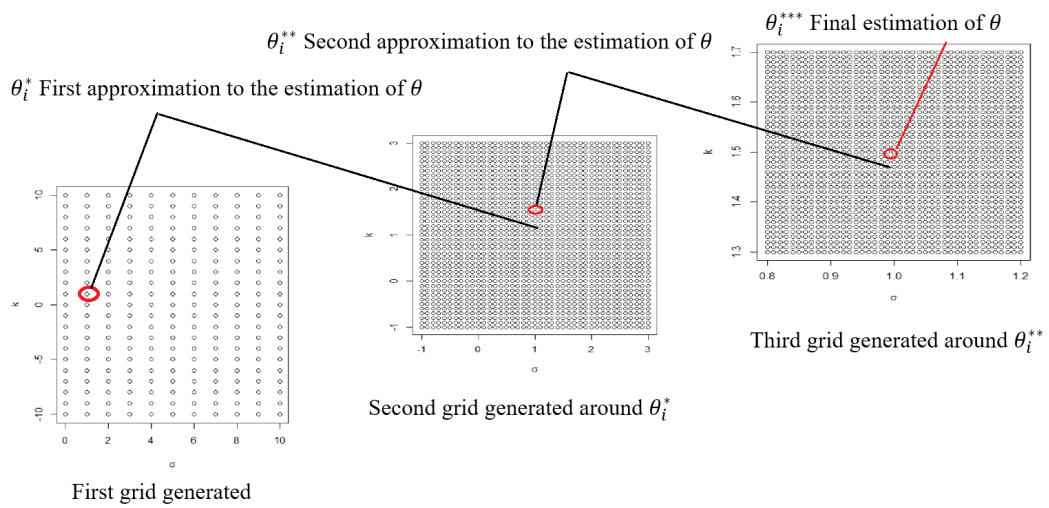
$\hat{k}$ : estimation of the  $k$  parameter.

1. for  $i \leftarrow 1$  to  $B$  do
2. Generate a sample size  $n$  of the GPD,  $X \sim \text{GPD}(\sigma, k)$
3. Obtain the empirical distribution function of the sample,  $F_n(\mathbf{x})$ .
4. Create a grid that contemplates the parametric space studied through all the possible ordered pairs of  $\sigma_1 \times k_1$ , where  $\sigma_1 = \{0, 1, \dots, 10\}$  and  $k_1 = \{-5, -4, \dots, 5\}$ . Each ordered pair is called a node and is denoted as  $\theta_i$  for  $i = 1, 2, \dots, 231$ .
5. for  $i \leftarrow 1$  to 231
- 6.

$$z_i = \sum_{j=1}^n (F(x_j; \theta_i) - F_n(x_j))^2$$

where  $F(x_j; \boldsymbol{\theta}_i)$  is the distribution function presented in equation 1.1 ( $z$  is a convex function and it is not difficult to prove for the cases  $\sigma = k$  and  $\sigma = c$  where  $c$  is a constant ( $c \in \mathbb{R}^+$ ) (see annex 1.8)).

7. Order  $z_i$  and take  $\boldsymbol{\theta}_i^*$  which make  $z_i$  be minimum.  $\boldsymbol{\theta}_i^*$  is taken as the first approximation to the estimation of  $\boldsymbol{\theta} = (\sigma, k)$ .
8. Generate a new grid around  $\boldsymbol{\theta}_i^*$  through  $\sigma_2 \times k_2$ , where  $\sigma_2$  is the set of all the values from  $\boldsymbol{\theta}_i^*[1] - 2$  to  $\boldsymbol{\theta}_i^*[1] + 2$  separated by length values of 0.1, and  $k_2$  is the set of values from  $\boldsymbol{\theta}_i^*[2] - 2$  to  $\boldsymbol{\theta}_i^*[2] + 2$  separated by length values of 0.1.  $\boldsymbol{\theta}_i^*[1]$  represents the first approximation to the estimation of  $\sigma$  and  $\boldsymbol{\theta}_i^*[2]$  the first approximation to the respective estimation for  $k$ .
9. for  $i \leftarrow 1$  to 1681 **do**
10. Repeat step 6 taking as  $\boldsymbol{\theta}_i$  the nodes from the grid obtained in step 8.
11. Repeat step 7 and take  $\boldsymbol{\theta}_i^{**}$  that make  $z_i$  be minimum.  $\boldsymbol{\theta}_i^{**}$  is taken as the second approximation to the estimation of  $\boldsymbol{\theta} = (\sigma, k)$ .
12. Generate a new grid around  $\boldsymbol{\theta}_i^{**}$  through  $\sigma_3 \times k_3$ , where  $\sigma_3$  is the set of all the values from  $\boldsymbol{\theta}_i^{**}[1] - 0.2$  to  $\boldsymbol{\theta}_i^{**}[1] + 0.2$  separated by length intervals of 0.01, and  $k_3$  is the set of values from  $\boldsymbol{\theta}_i^{**}[2] - 0.2$  to  $\boldsymbol{\theta}_i^{**}[2] + 0.2$  separated by length intervals of 0.01.  $\boldsymbol{\theta}_i^{**}[1]$  represents the second approximation to the estimation of  $\sigma$  and  $\boldsymbol{\theta}_i^{**}[2]$  the second approximation to the respective estimation for  $k$ .
13. for  $i \leftarrow 1$  to 1681 **do**
14. Repeat step 6 taking as  $\boldsymbol{\theta}_i$  the nodes from the grid obtained in step 12.
15. Repeat step 7 and take  $\boldsymbol{\theta}_i^{***}$  that make  $z_i$  be minimum as the final estimation of  $\boldsymbol{\theta} = (\sigma, k)$ . return  $\hat{\sigma}$  and  $\hat{k}$



**Figure 1.2:** Grid searching steps.

## CONCLUSIONES GENERALES

El estudio de simulación muestra que no existe un estimador cuyo desempeño sea uniformemente mejor para el espacio paramétrico estudiado. El desempeño del estimador de MD usando el algoritmo propuesto se encuentra entre los tres mejores estimadores estudiados, con las ventajas de que no presenta problemas de convergencia ni requiere de un punto de inicio a priori para su buen desempeño. La propuesta para la estimación de los parámetros de la GPD puede ser una buena alternativa para el caso de otras distribuciones en las cuales los estimadores tradicionales presentan problemas.

## REFERENCIAS

- Arneric, J., Lolic, I. y Galietic, J. (2012). Threshold Parameter of the Expected Losses. *Croatian Operational Research Review*, 3, 270–279.
- Arnold, B. C. (2015). *Pareto Distributions*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman and Hall/CRC, second edition edición.
- Castillo, E. y Hadi, A. S. (1997). Fitting the Generalized Pareto Distribution to Data. *JASA, Theory and Methods*, 92, 440, 1609–1620.
- Chen, P., Ye, Z.-S. y Zhao, X. (2016). Minimum Distance Estimation for the Generalized Pareto Distribution. *Technometrics*, 0, 0, 1–14.
- Choulakian, V. y Stephens, M. (2001). Goodness of fit tests for the generalized Pareto distribution. *Technometrics*, 43, 4, 478–484.
- Davison, A. C. (1984). *Statistical Extremes and Applications*, cap. Modelling Excesses over High Thresholds, with an Application, 461–482. Springer.
- Davison, A. C. y Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Statist. Soc.*, 52, 237–254.
- de ZeaBermudez, P. y Kotz, S. (2010a). Parameter estimation of the generalized Pareto distribution - Part I. *Journal of Statistical Planning and Inference*, 140, 1353–1373.
- de Zea-Bermudez, P. y Kotz, S. (2010b). Parameter estimation of the generalized Pareto distribution - Part II. *Journal of Statistical Planning and Inference*, 140, 1374–1388.
- Deidda, R. y Puliga, M. (2009). Performances of some parameter estimators of the generalized Pareto distribution over rounded-off samples. *Physics and Chemistry of the Earth*, 626–634.
- Dupuis, D. (1999). Exceedances over High Thresholds: A Guide to Threshold Selection. *Extremes*, 1, 3, 251261.
- Embrechts, P., Kluppelberg, C. y Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer.
- García-Pérez, A. (2007). *Una aportación al análisis de solvencia: la teoría del valor extremo..* Tesis Doctoral, Universidad de Alcalá.

- Hosking, J. y Wallis, J. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*.
- Huber, P. J. (1972). The 1972 Wald Lecture Robust Statistics: A Review. *The Annals of Mathematical Statistics*, 43, 4, 1041–1067.
- Jockovic, J. (2012). Quantile estimation for the generalized Pareto distribution with application to finance. *Yugoslav Journal of Operations Research*, 22, 2, 297–3111.
- Landwehr, J. M., Matalas, N. y Wallis, J. (1979). Probability weighted moments compared with some traditional techniques for estimating Gumbel parameters and quantiles. *Water Resour. Res.*, 15, 1055–1064.
- Lee, W. C. (2009). Applying generalized pareto distribution to the risk management of commerce fire insurance. None.
- Luceño, A. (2006). Fitting the generalized Pareto distribution to data using maximum goodness of fit estimators. *Computational Statistics and Data Analysis*, 51, 904–917.
- Parr, W. C. (1981). Minimum distance estimation: a bibliography. *Communications in Statistics - Theory and Methods*, 10, 12, 1205 – 1224.
- Parr, W. C. y Schucany, W. R. (1980). Minimum Distance and Robust Estimation. *Journal of the American Statistical Association*, 75, 371, 616–624.
- Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3, 1, 119–131.
- Plavcová, E. y Kyselý, J. (2011). Evaluation of daily temperatures in Central Europe and their links to large-scale circulation in an ensemble of regional climate models. *Tellus A*, 63, 5, 1052 – 1054.
- Reiss, R.-D. y Thomas, M. (2007). *Statistical analysis of extreme values. With applications to insurance, finance, hydrology and other fields..* Birkhauser, tercera edición.
- Smith, R. L. (1984). *Statistical Extremes and Applications*, cap. Threshold Methods for Sample Extremes, 621–638. Springer.
- Wolfowitz, J. (1953). Estimation by the Minimum Distance Method. *Ann. Inst. Stat. Math.*, 5, 9–23.
- Wolfowitz, J. (1957). The minimum distance method. *The Annals of Mathematical Statistics*, 28, 1, 75–88.



Yohai, V. J. y Zamar, R. H. (1988). High Breakdown-Point Estimates of Regression by Means of the Minimization of an Efficient Scale. *Journal of the American Statistical Association*, 83, 402, 406 – 4013.

## ANEXO

**Demostración, para dos casos, de que la función objetivo del método de mínima distancia es una función convexa.**

- **Caso 1.**  $\sigma = k$

Bajo este caso de 1.1 se tiene que

$$F(x; k) = 1 - (1 - x)^{1/k}, \quad (7)$$

Por lo que podemos definir una medida de distancia entre dos funciones de distribución como

$$g(k) = \sum_{i=1}^n \left[ 1 - (1 - x_i)^{1/k} - F_n(x) \right]^2$$

donde  $F_n(x)$  es la función de distribución empírica.

Bajo el método de estimación de mínima distancia surge la siguiente pregunta ¿es  $g(k)$  una función convexa?, para dar respuesta usaremos la siguiente definición.

**Definición:** Sea  $f$  una función dos veces diferenciable y sea  $\text{dom}(f) = \mathbb{R}$ , si  $f$  tiene un dominio convexo, entonces  $f$  es convexa si

$$f''(x) \geq 0 \quad \forall x \in \text{dom}(f)$$

De este modo se tiene que probar que

$$g''(k) \geq 0 \quad \forall k \in \text{dom}(g)$$

por lo cual, obteniendo la segunda derivada de  $g(k)$  se tiene

$$\begin{aligned}
g''(k) &= \underbrace{\sum_{i=1}^n \frac{2(\log(1-x_i))^2(1-x_i)^{2/k}}{k^4}}_a \\
&\quad - \underbrace{\sum_{i=1}^n \frac{4\log(1-x_i)(1-x_i)^{1/k}(F(x_i; k) - F_n(x_i))}{k^3}}_b \\
&\quad - \underbrace{\sum_{i=1}^n \frac{2(\log(1-x_i))^2(1-x_i)^{1/k}(F(x_i; k) - F_n(x_i))}{k^4}}_c
\end{aligned}$$

Ahora, por la Ley Fuerte de los Grandes Números

$$\lim_{n \rightarrow \infty} F_n(x_i) = F(x_i; k),$$

por lo que  $b$  y  $c$  tienden a ser cero cuando  $n \rightarrow \infty$ , de este modo

$$g''(k) = \sum_{i=1}^n \frac{2(\log(1-x_i))^2(1-x_i)^{2/k}}{k^4} \geq 0,$$

al ser una suma de cuadrados. Así, podemos concluir que  $g(k)$  es una función convexa.

- **Caso 2.**  $\sigma = c$ , donde  $c$  es real positivo.

En este caso

$$F(x; k) = 1 - \left( \frac{1 - kx}{c} \right)^{1/k}, \quad (8)$$

Por lo que

$$g(k) = \sum_{i=1}^n \left[ 1 - \left( \frac{1 - kx_i}{c} \right)^{1/k} - F_n(x_i) \right]^2$$

Ahora, obteniendo  $g''(k)$  se tiene

$$\begin{aligned}
g''(k) = & \underbrace{\sum_{i=1}^n 2 \left( \frac{1-kx_i}{c} \right)^{2/k} \left( -\frac{\log\left(\frac{1-kx_i}{c}\right)}{k^2} - \frac{x_i}{k(1-kx_i)} \right)^2}_{a^*} \\
& - \underbrace{\sum_{i=1}^n 2 \left( \frac{1-kx_i}{c} \right)^{1/k} (F(x_i; k) - Fn(x_i)) \left( -\frac{\log\left(\frac{1-kx_i}{c}\right)}{k^2} - \frac{x_i}{k(1-kx_i)} \right)^2}_{b^*} \\
& - \underbrace{\sum_{i=1}^n 2 \left( \frac{1-kx_i}{c} \right)^{1/k} (F(x_i; k) - Fn(x_i)) \left( 2\frac{\log\left(\frac{1-kx_i}{c}\right)}{k^3} + \frac{2x_i}{k^2(1-kx_i)} - \frac{x_i^2}{k(1-kx_i)^2} \right)}_{c^*}
\end{aligned}$$

por la Ley Fuerte de los Grandes Números  $b^*$  y  $c^*$  tienden a ser cero cuando  $n \rightarrow \infty$ , de este modo

$$g''(k) \geq 0,$$

Así, podemos concluir que para  $\sigma = c$ ,  $g(k)$  es una función convexa.