

ESTIMACIÓN BAYESIANA EN LA FAMILIA PARETO GENERALIZADA

RUBÉN SÁNCHEZ GÓMEZ *

Recibido/Received: 22 Feb 2006; Aceptado/Accepted: 13 Jun 2007

Resumen

La familia de distribuciones Pareto generalizada con parámetro de escala $\sigma > 0$ y de forma k , se ha utilizado para modelar excedencias sobre un umbral dado, no obstante la estimación paramétrica en esta familia presenta algunos problemas. En este trabajo se estudia el enfoque bayesiano para estimar los parámetros σ y k cuando no se tiene información a priori disponible y se discute el caso en que hay información previa. Se presenta un estudio de simulación para analizar el desempeño de la metodología bayesiana, usando distribuciones a priori no informativas y los métodos anteriormente propuestos en la literatura. Este estudio muestra que la estimación bayesiana supera en buena medida a métodos propuestos, en términos de sesgo y de raíz del error cuadrado medio. Las metodologías de estimación analizadas se aplican a conjuntos de datos reales.

Palabras clave: Familia Pareto generalizada, métodos de estimación, estudio Monte Carlo.

Abstract

The generalized Pareto family of distributions with scale parameter $\sigma > 0$ and k form, has been used for modeling surplus over a given threshold, even though the parametric estimation in this family has some problems. In this work we study the Bayesian approach for estimating parameters σ and k when no a priori information is available and we discuss the case when there is previous information. We present a simulation study in order to analyze the performance of the Bayesian methodology, employing non informative a priori distributions and the methods available in the literature. This study shows that the Bayesian estimation performs better than other proposed methods, in terms of bias and square root of the mean quadratic error. The estimation methodologies analyzed are applied to real data sets.

*Centro Universitario de los Lagos, Universidad de Guadalajara, Enrique Días de León S/N, Col. Paseos de la Montaña, C.P. 47460, Lagos de Moreno, Jalisco, México, Fax: +(52) 47.42.36.78, E-mail: rubensg@cencar.udg.mx, rsanchez@ccip.udg.mx.

Keywords: Generalized Pareto family, estimation methods, Monte Carlo study.

Mathematics Subject Classification: 58E17, 62C12.

1 Introducción

El análisis estadístico de magnitudes que exceden un nivel máximo permitido o umbral, se ha incrementado en años recientes, tanto por su desarrollo estadístico como por sus aplicaciones en procesos naturales de tipo Ambiental (Contaminación del aire, inundaciones, etc.), Climatológico (velocidad del viento, temperatura, precipitación, etc.) e Hidrológico (nivel del mar, nivel de lagos, flujo de ríos, etc.), en donde el propósito de investigación puede ser modelar los excesos a partir de un umbral alto, con fines de pronóstico, investigar tendencia en el tiempo o algún otro mecanismo de interés. Casos de aplicación se pueden encontrar en Hosking y Wallis (1987), Davison y Smith (1990), Castillo y Hadi (1997), Huang y Smith (1999), Hall y Tajvidi (2000) y Sánchez (2001) entre otras referencias.

La aproximación más simple para estos excesos sobre un umbral es la Distribución Pareto Generalizada (DPG), sugerida por Pickands (1975) y definida por

$$G(x; k; \sigma) = 1 - \left(1 - k \frac{x}{\sigma}\right)_+^{\frac{1}{k}},$$

en donde $\sigma > 0$, k es cualquier número real y $x_+ = \max(0, x)$. Nótese que el caso $k < 0$ corresponde a una reparametrización de la distribución Pareto estudiada ampliamente por Arnold (1983), para $k = 1$ se tiene una distribución uniforme en el intervalo $(0, \sigma)$, el caso $k = 0$ corresponde a la distribución exponencial en el límite cuando $k \rightarrow 0$ y para $k > 0$ se tiene una distribución $\left(\frac{\sigma}{k}\right) Beta(1, 1/k)$.

Una propiedad particularmente útil de esta distribución es la estabilidad del umbral, esto significa que si el modelo PG es válido para un umbral u , entonces debe ser válido para un umbral v , con $v > u$. Otra característica de la DPG es la función de vida media residual dada por $E(T - u | T > u) = (\sigma - ku)/(1 + k)$ cuando $k > -1$, $u > 0$ y $\sigma - ku > 0$. De aquí, si la gráfica del exceso medio sobre u , versus u es una línea recta con pendiente $-k/(1 + k)$ e intercepto $\sigma/(1 + k)$. Esto sugiere tanto un método gráfico para estimar k y σ , como una prueba de bondad de ajuste basada en la linealidad de la gráfica de $E(T - u | T > u)$ como función de u .

Se han propuesto varios métodos para estimar los parámetros en la DPG. Pickands (1975) sugiere algunos métodos considerando procedimientos para elegir el valor del umbral que define las excedencias, incluyendo metodologías no paramétricas, Bayesianas, basadas en estadísticos de orden y otras aproximaciones.

El método de estimación por Máxima Verosimilitud (MV) se ha considerado por varios autores y en particular, Smith (1984) observa que los estimadores por MV existen siempre y cuando $k < 1$ y son asintóticamente normales y eficientes solo si $k < \frac{1}{2}$. Por otro lado, Hosking y Wallis (1987) proveen una investigación detallada del proceso de estimación, destacando que la estimación por MV no presenta claramente su eficiencia aún con muestras de 500 observaciones. Grimshaw (1993) propone un algoritmo para

resolver las ecuaciones de verosimilitud señalando que si no se encuentra un mínimo local en

$$\{k < 0, \sigma > 0\} \cup \{0 < k \leq 1, \frac{\sigma}{k} > x_{n:n}\}, \quad x_{n:n} = \max\{x_1, \dots, x_n\},$$

no existen estimadores por MV. Por tanto, para tamaños de muestra pequeños (< 500) y en situaciones en que las estimaciones de k están fuera del intervalo $(\frac{-1}{2}, \frac{1}{2})$, que comunmente sucede si los datos son de cola pesada o en datos truncados, la estimación por MV presenta limitaciones.

Hosking y Wallis (1987) consideran además la estimación por el Método de Momentos (**MOM**) y el Método de Probabilidades Ponderadas (**MPP**), desafortunadamente estos estimadores también presentan algunas limitaciones, ya que para $k \leq \frac{-1}{2}$ se tiene que $Var[X] = \infty$ y entonces, los estimadores por MOM y MPP no existen. Aún en los casos en que existen, pueden ser inconsistentes con la muestra observada cuando $\tilde{\sigma}/\tilde{k} < x_{n:n}$, en donde $x_{n:n}$ es el estadístico de orden mayor de la muestra.

Castillo y Hadi (1997) proponen otra alternativa para estimar parámetros y cuantiles de la DPG, con el Método de Elemental Percentil (**MEP**) que básicamente consiste en resolver el sistema de ecuaciones

$$F(x_{i:n}; k, \sigma) = p_{i:n}, \quad F(x_{j:n}; k, \sigma) = p_{j:n}$$

en donde $x_{i:n}$ y $x_{j:n}$ son dos estadísticos de orden distintos de una muestra aleatoria de tamaño n , de la distribución $F(x; k, \sigma)$ y $p_{i:n} = \frac{i+\gamma}{n+\beta}$ es el percentil correspondiente. Sin embargo, encuentran que el MEP desempeña un buen papel solo cuando $k < -1/2$ y $k > 1/2$, y observan que no es la mejor opción en el rango $-1/2 < k < 1/2$, comparando su desempeño con los métodos de MOM y MPP en términos de sesgo y raíz cuadrada del error cuadrado medio. De modo que, a pesar de sobreponer las desventajas observadas en los métodos de MOM y MPP, fuera de $(-1/2, 1/2)$ se tiene aún la incertidumbre de contar con un método confiable, lo que motiva la búsqueda de un método alternativo.

En la siguiente sección se propone la estimación bayesiana con distribuciones a priori no informativas y en la sección posterior se presentan los resultados que se obtienen en un estudio de simulación Montecarlo, comparando las técnicas propuestas en la literatura y la estimación bayesiana.

2 Análisis bayesiano con distribuciones a priori no informativas

Asumiendo que no hay información a priori disponible, en la literatura estadística se puede encontrar una serie de técnicas para derivar distribuciones a priori no informativas con aplicaciones en una gran variedad de modelos estadísticos (Robert, 1994; Lindsey, 1996; Yang y Berger, 1998). En este caso, basta seleccionar entre estas propuestas una alternativa apropiada a las características de la familia PG. Para tal efecto, supóngase que x_1, x_2, \dots, x_n es una muestra de una variable aleatoria distribuida PG con parámetro de forma k y parámetro de escala σ [o bien reparametrizando $\delta = \frac{1}{\sigma}$].

La función de verosimilitud para la muestra, se escribe como

$$\begin{aligned} L(k, \sigma | \text{datos}) &= \sigma^{-n} \prod_{i=1}^n (1 - kx_i/\sigma)_+^{\frac{1}{k}-1} \\ &= \delta^n \prod_{i=1}^n (1 - k\delta x_i)_+^{\frac{1}{k}-1}. \end{aligned}$$

De aquí, algunas distribuciones “priors” alternativas son las siguientes:

- Si el parámetro de escala es conocido, por la forma de la distribución PG se tendría que $k \in (-\infty, \frac{\sigma}{t_{n:n}})$ y aparentemente la opción más viable es suponer la distribución *priori localmente uniforme* $\pi(k) \propto 1$, popularizada por Laplace (1812),
- Considerando el modelo $f(x|k, \delta)$ y asumiendo similarmente que k es conocido, dado que $\delta > 0$, es bastante común utilizar $\pi(\delta) \propto \frac{1}{\delta}$.

De esta forma, analizando el caso más general en que ambos parámetros son desconocidos, en el presente trabajo se consideran las siguientes priors:

1. Priori localmente uniforme $\pi(k, \sigma) \propto 1$, de donde la distribución posteriori tiene la forma

$$\pi_1(k, \sigma | \text{datos}) \propto L(k, \sigma | \text{datos}) \pi(k, \sigma) = \sigma^{-n} \prod_{i=1}^n (1 - kx_i/\sigma)_+^{\frac{1}{k}-1}.$$

2. Priori $\log \delta$ localmente uniforme $\pi_{(k, \delta)} \propto \frac{1}{\delta}$, obteniendo

$$\pi_2(k, \sigma | \text{datos}) \propto \delta^{n-1} \prod_{i=1}^n (1 - k\delta x_i)_+^{\frac{1}{k}-1}.$$

Se puede observar que en cualquiera de estos casos es necesario integrar en forma numérica, por ello, en este trabajo se utiliza el método de cuadratura conocido como *regla compuesta de Simpson*, implementado en un programa Fortran 90. En todos los casos se asume la función de pérdida de error cuadrado medio, por lo que los estimadores bayesianos de k y δ son simplemente valores esperados de la distribución posterior dados los datos.

3 Un estudio comparativo

Cuando no se tiene información a priori disponible, se puede comparar el desempeño de los estimadores de Bayes con los estimadores clásicos en términos de su error cuadrado medio. Además, considerando el hecho de que necesariamente estos estimadores bayesianos estarán sesgados, es de interés investigar también el desempeño de los métodos bajo estudio en términos de sesgo. Por esto, en esta sección se presentan los resultados obtenidos en

un estudio de simulación paralelo a los trabajos presentados por Hosking y Wallis (1987) y por Castillo y Hadi (1997), comparando el desempeño de los métodos propuestos en la literatura y la estimación Bayesiana, utilizando las distribuciones a priori no informativas antes mencionadas.

Además, como el método de MV es eficiente solo para muestras grandes (> 500), que pueden presentar problemas al resolver las ecuaciones de verosimilitud en forma numérica para algunas muestras y como no se desempeña mejor que los métodos de MOM y MPP para muestras pequeñas (Hosking y Wallis, 1987), se excluye en este trabajo (paralelamente al trabajo de Castillo y Hadi).

3.1 El experimento de simulación

De esta forma, el experimento de simulación se ejecuta bajo las siguientes condiciones: Se consideran los valores $\{-2.0, -1.0, -0.4, -0.2, 0, 0.2, 0.4, 1.0, 2.0\}$ en el parámetro de forma k y dado que los resultados son invariantes respecto al parámetro de escala (Hosking y Wallis, 1987), se fija $\sigma = 1$ y examinando muestras de tamaños $n = \{15, 25, 50\}$, los resultados se basan en 1000 corridas de simulación.

Para generar variables pseudo-aleatorias $PG(k, \sigma)$, se utiliza el método de la transformada inversa que coincide con la transformación Box-Cox definida por

$$g(u, k, \sigma) = \begin{cases} -\sigma \log(u) & k = 0 \\ -\sigma \frac{u^k - 1}{k} & k \neq 0 \end{cases} \quad (1)$$

en donde $u \sim U[0, 1]$. En este caso, los números pseudo-aleatorios utilizados se obtienen con el método congruencial mixto y para cada combinación de k y n se estiman los parámetros utilizando MOM, MPP, MEP y la estimación bayesiana con:

1. Priori localmente uniforme,
2. Priori $\log \delta$ localmente uniforme.

3.2 Resultados del estudio de simulación

De los resultados generados en el experimento de simulación, en las tablas 1 y 2 se presenta el sesgo y la Raíz del Error Cuadrado Medio (**RECM**) respectivamente, para las estimaciones del parámetro de forma k y en las tablas 2 y 4 se muestran los sesgos y RECM para los estimadores del parámetro de escala σ . Se puede observar en estas tablas que para todos los métodos, el sesgo decrece en magnitudes conforme se incrementa el tamaño de muestra, dando evidencia de consistencia en los estimadores, en tanto que la RECM también decrece.

Como una medida para comparar la bondad de ajuste global Castillo y Hadi (1997) presentan el error absoluto promedio escalado (**EAPE**)

$$EAPE = n^{-1} \sum_{i=1}^n \frac{|x_{i:n} - \hat{x}_{i:n}|}{x_{n:n} - x_{1:n}}$$

en donde $\hat{x}_{i:n} = \hat{\sigma}[1 - (1 - p_{i:n})\hat{k}]/\hat{k}$. Aunque sea extraño que usaran esta medida en particular y no alguna otra medida de distancia más común, en la tabla 3 se muestran los EAPE obtenidos en este experimento de simulación.

Como una medida de distancia alternativa comparando la bondad de ajuste, en la Tabla 4 se puede observar la distancia Kullback-Leibler (**K-L**) entre el modelo exacto

$$f(x|k, \sigma) = \frac{1}{\sigma} \left(1 - \frac{kx}{\sigma}\right)_+^{(\frac{1}{k}-1)},$$

del cual se generan los valores simulados, y el modelo estimado

$$g(x|\hat{k}, \hat{\sigma}) = \frac{1}{\hat{\sigma}} \left(1 - \frac{kx}{\hat{\sigma}}\right)_+^{(\frac{1}{\hat{k}}-1)},$$

generado con las estimaciones de k y σ a partir de las muestras con los métodos de MOM, MPP, MEP y estimación Bayesiana con las prioris no informativas.

| Estimado | n | Método | k | | | | | | | | |
|----------------|-----|---|-----------|--------|--------|--------|--------|--------|--------|--------|--------|
| | | | -2.00 | -1.00 | -.40 | -.20 | 0 | .20 | .40 | 1.00 | 2.00 |
| \hat{k} | 15 | MOM | 1.600 | 0.716 | 0.303 | 0.208 | 0.119 | 0.080 | 0.066 | 0.123 | 0.460 |
| | | MPP | 1.262 | 0.534 | 0.283 | 0.258 | 0.198 | 0.202 | 0.213 | 0.305 | 0.547 |
| | | MEP | -0.523 | -0.278 | -0.152 | -0.118 | -0.112 | -0.089 | -0.074 | -0.002 | -0.278 |
| | | $\pi_{(k,\delta)} \propto 1$ | -0.414 | -0.269 | -0.177 | -0.139 | -0.132 | -0.108 | -0.086 | 0.003 | -0.209 |
| | | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | -0.291 | -0.167 | -0.079 | -0.044 | -0.034 | -0.010 | 0.016 | 0.117 | -0.132 |
| | 25 | MOM | 1.564 | 0.645 | 0.228 | 0.153 | 0.087 | 0.060 | 0.041 | 0.049 | 0.241 |
| | | MPP | 1.175 | 0.405 | 0.171 | 0.163 | 0.130 | 0.130 | 0.13 | 0.159 | 0.315 |
| | | MEP | -0.405 | -0.242 | -0.131 | -0.087 | -0.065 | -0.039 | -0.035 | 0.010 | -0.304 |
| | | $\pi_{(k,\delta)} \propto 1$ | -0.271 | -0.172 | -0.134 | -0.091 | -0.084 | -0.058 | -0.054 | 0.006 | -0.176 |
| | | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | -0.188 | -0.109 | -0.076 | -0.036 | -0.029 | -0.003 | 0.001 | 0.073 | -0.133 |
| | 50 | MOM | 1.531 | 0.598 | 0.178 | 0.083 | 0.045 | 0.041 | 0.023 | 0.027 | 0.100 |
| | | MPP | 1.097 | 0.324 | 0.110 | 0.065 | 0.064 | 0.077 | 0.067 | 0.081 | 0.143 |
| | | MEP | -0.315 | -0.163 | -0.089 | -0.084 | -0.061 | -0.027 | -0.026 | -0.065 | -0.285 |
| | | $\pi_{(k,\delta)} \propto 1$ | -0.103 | -0.086 | -0.057 | -0.074 | -0.056 | -0.031 | -0.031 | -0.008 | -0.134 |
| | | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | -0.054 | -0.049 | -0.028 | -0.045 | -0.030 | -0.005 | -0.007 | 0.022 | -0.119 |
| $\hat{\sigma}$ | 15 | MOM | 2886.985 | 15.802 | 0.404 | 0.197 | 0.109 | 0.056 | 0.062 | 0.096 | 0.208 |
| | | MPP | 487.098 | 3.170 | 0.32 | 0.244 | 0.187 | 0.158 | 0.169 | 0.187 | 0.227 |
| | | MEP | 0.066 | -0.001 | 0.015 | 0.001 | 0.000 | -0.020 | -0.005 | 0.021 | -0.072 |
| | | $\pi_{(k,\delta)} \propto 1$ | -0.044 | -0.061 | -0.057 | -0.069 | -0.061 | -0.077 | -0.063 | -0.035 | -0.133 |
| | | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | 0.203 | 0.115 | 0.089 | 0.063 | 0.065 | 0.036 | 0.044 | 0.050 | -0.096 |
| | 25 | MOM | 16224.217 | 13.820 | 0.321 | 0.157 | 0.089 | 0.055 | 0.03 | 0.041 | 0.107 |
| | | MPP | 1631.259 | 1.783 | 0.195 | 0.160 | 0.132 | 0.114 | 0.093 | 0.096 | 0.129 |
| | | MEP | 0.031 | -0.043 | -0.008 | 0.001 | 0.008 | 0.009 | -0.006 | 0.015 | -0.083 |
| | | $\pi_{(k,\delta)} \propto 1$ | -0.061 | -0.081 | -0.072 | -0.061 | -0.054 | -0.051 | -0.063 | -0.029 | -0.107 |
| | | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | 0.102 | 0.032 | 0.022 | 0.026 | 0.027 | 0.023 | 0.002 | 0.027 | -0.085 |
| | 50 | MOM | 45893.874 | 4.880 | 0.267 | 0.096 | 0.048 | 0.038 | 0.023 | 0.021 | 0.045 |
| | | MPP | 2300.663 | 0.583 | 0.132 | 0.070 | 0.067 | 0.069 | 0.055 | 0.048 | 0.058 |
| | | MEP | -0.055 | -0.032 | 0.001 | -0.026 | -0.013 | -0.001 | -0.006 | -0.028 | -0.083 |
| | | $\pi_{(k,\delta)} \propto 1$ | -0.058 | -0.058 | -0.031 | -0.059 | -0.048 | -0.032 | -0.035 | -0.024 | -0.074 |
| | | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | 0.037 | 0.012 | 0.022 | -0.011 | -0.005 | 0.006 | -0.002 | 0.002 | -0.066 |

Tabla 1: Sesgo de los estimadores de k y σ , para distintos tamaños de muestra n y distintos valores simulados de k .

Esta medida de pérdida de información permite comparar el desempeño de las metodologías bajo estudio, modelando $f(x|k, \sigma)$ con $g(x|\hat{k}, \hat{\sigma})$. De esta forma, la metodología

que se desempeña mejor, será aquella que presente menor distancia K-L. Cabe mencionar que similarmente al caso de sesgo y RECM, el EAPE (tabla 5) y la distancia K-L (tabla 6) decrecen conforme se incrementa el tamaño de muestra.

Se puede observar además de consistencia que para todos los métodos y todos los tamaños de muestra, la tendencia del sesgo y de la RECM para los estimadores de k y σ es decreciente en magnitud conforme el valor de k se aproxima al intervalo $(0, \frac{1}{2})$ y crece conforme k toma valores más alejados de este intervalo. Con esto, en parte se confirma la conclusión de Castillo y Hadi, quienes perciben solo tendencia decreciente en la magnitud del sesgo conforme k crece a partir de -2 .

Se confirma también la conclusión observada por Hosking y Wallis (1987) y reafirmada por Castillo y Hadi (1997); el desempeño de los métodos depende del valor del parámetro k .

En general, el MPP se desempeña mejor para $-0.4 \leq k < 0$ y para $0 < k \leq 0.4$ el MOM es mejor. Además, el MEP se desempeña mejor para $k < -0.4$ y $k > 0.4$ para muestras de tamaño $n = 15, 25$ y compite con el MPP y MOM para muestras de tamaño $n = 50, 100, 200$, mientras que para el rango $-0.4 \leq k \leq 0.4$ el MEP tiene un desempeño regular.

| Estimado | n | Método | k | | | | | | | | |
|---|-----|---|------------|---------|-------|-------|-------|-------|-------|-------|-------|
| | | | -2.00 | -1.00 | -.40 | -.20 | 0 | .20 | .40 | 1.00 | 2.00 |
| \hat{k} | 15 | MOM | 1.602 | 0.733 | 0.383 | 0.321 | 0.3 | 0.311 | 0.406 | 0.75 | 2.199 |
| | | MPP | 1.275 | 0.607 | 0.441 | 0.428 | 0.412 | 0.439 | 0.521 | 0.776 | 1.515 |
| | | MEP | 1.284 | 0.825 | 0.539 | 0.454 | 0.44 | 0.374 | 0.374 | 0.43 | 0.657 |
| | | $\pi_{(k,\delta)} \propto 1$ | 0.905 | 0.638 | 0.47 | 0.395 | 0.393 | 0.325 | 0.318 | 0.346 | 0.363 |
| | | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | 0.864 | 0.620 | 0.462 | 0.392 | 0.397 | 0.329 | 0.335 | 0.382 | 0.316 |
| | 25 | MOM | 1.565 | 0.654 | 0.282 | 0.248 | 0.221 | 0.244 | 0.284 | 0.492 | 1.131 |
| | | MPP | 1.182 | 0.456 | 0.298 | 0.311 | 0.292 | 0.318 | 0.354 | 0.511 | 0.953 |
| | | MEP | 1.053 | 0.629 | 0.392 | 0.343 | 0.287 | 0.264 | 0.249 | 0.288 | 0.539 |
| | | $\pi_{(k,\delta)} \propto 1$ | 0.681 | 0.437 | 0.33 | 0.294 | 0.261 | 0.239 | 0.226 | 0.254 | 0.305 |
| | | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | 0.656 | 0.419 | 0.315 | 0.289 | 0.257 | 0.241 | 0.229 | 0.278 | 0.274 |
| | 50 | MOM | 1.532 | 0.602 | 0.221 | 0.146 | 0.15 | 0.164 | 0.188 | 0.329 | 0.698 |
| | | MPP | 1.1 | 0.362 | 0.219 | 0.169 | 0.189 | 0.212 | 0.229 | 0.339 | 0.613 |
| MEP | | 0.817 | 0.481 | 0.3 | 0.223 | 0.195 | 0.165 | 0.154 | 0.145 | 0.446 | |
| $\pi_{(k,\delta)} \propto 1$ | | 0.427 | 0.308 | 0.227 | 0.189 | 0.174 | 0.162 | 0.147 | 0.185 | 0.211 | |
| $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | | 0.418 | 0.299 | 0.221 | 0.18 | 0.167 | 0.16 | 0.145 | 0.19 | 0.194 | |
| $\hat{\sigma}$ | 15 | MOM | 35694.802 | 180.516 | 0.681 | 0.449 | 0.408 | 0.373 | 0.437 | 0.54 | 0.972 |
| | | MPP | 5976.301 | 30.265 | 0.606 | 0.533 | 0.507 | 0.468 | 0.523 | 0.558 | 0.684 |
| | | MEP | 0.921 | 0.591 | 0.477 | 0.43 | 0.441 | 0.373 | 0.379 | 0.353 | 0.305 |
| | | $\pi_{(k,\delta)} \propto 1$ | 0.373 | 0.315 | 0.292 | 0.265 | 0.298 | 0.260 | 0.267 | 0.264 | 0.196 |
| | | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | 0.622 | 0.447 | 0.384 | 0.336 | 0.366 | 0.301 | 0.314 | 0.291 | 0.171 |
| | 25 | MOM | 313059.703 | 292.89 | 0.504 | 0.34 | 0.304 | 0.32 | 0.299 | 0.352 | 0.481 |
| | | MPP | 31356.655 | 29.403 | 0.414 | 0.385 | 0.365 | 0.376 | 0.345 | 0.363 | 0.42 |
| | | MEP | 0.708 | 0.431 | 0.353 | 0.327 | 0.308 | 0.310 | 0.269 | 0.250 | 0.232 |
| | | $\pi_{(k,\delta)} \propto 1$ | 0.302 | 0.277 | 0.259 | 0.250 | 0.245 | 0.248 | 0.219 | 0.210 | 0.165 |
| | | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | 0.425 | 0.329 | 0.292 | 0.280 | 0.272 | 0.270 | 0.230 | 0.225 | 0.149 |
| | 50 | MOM | 998604.69 | 24.793 | 0.382 | 0.233 | 0.208 | 0.213 | 0.221 | 0.24 | 0.299 |
| | | MPP | 49954.525 | 1.388 | 0.291 | 0.24 | 0.238 | 0.247 | 0.246 | 0.246 | 0.272 |
| MEP | | 0.458 | 0.323 | 0.26 | 0.224 | 0.214 | 0.207 | 0.201 | 0.149 | 0.184 | |
| $\pi_{(k,\delta)} \propto 1$ | | 0.288 | 0.241 | 0.226 | 0.206 | 0.194 | 0.195 | 0.188 | 0.17 | 0.113 | |
| $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | | 0.332 | 0.259 | 0.24 | 0.209 | 0.197 | 0.199 | 0.191 | 0.173 | 0.104 | |

Tabla 2: RECM de los estimadores de k y σ , para distintos valores simulados de k y n .

Sin embargo, tomando en cuenta el desempeño de la estimación Bayesiana, se mejora en gran parte el desempeño de las tres metodologías antes mencionadas con cualquiera de las prioris señaladas en la sección anterior.

En particular, con la priori $\log \delta$ localmente uniforme se puede observar en general menor sesgo y menor RECM para todos los tamaños de muestra. Aunque su desempeño decrece un poco para $k = -2, -1$ supera los resultados obtenidos con el MEP propuesto por Castillo y Hadi. De hecho, el desempeño de la estimación Bayesiana usando la priori localmente uniforme o la priori $\log \delta$ localmente uniforme es mejor para $n = 15, 25, 50$.

| n | Método | k | | | | | | | | |
|-----|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | | -2.00 | -1.00 | -.40 | -.20 | 0 | .20 | .40 | 1.00 | 2.00 |
| 15 | MOM | 0.086 | 0.073 | 0.060 | 0.055 | 0.054 | 0.052 | 0.053 | 0.056 | 0.068 |
| | MPP | 0.075 | 0.068 | 0.059 | 0.057 | 0.057 | 0.056 | 0.057 | 0.059 | 0.064 |
| | MEP | 0.026 | 0.031 | 0.038 | 0.039 | 0.043 | 0.045 | 0.048 | 0.052 | 0.073 |
| | $\pi_{(k,\delta)} \propto 1$ | 0.104 | 0.061 | 0.050 | 0.049 | 0.052 | 0.053 | 0.058 | 0.071 | 0.094 |
| | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | 0.089 | 0.055 | 0.047 | 0.046 | 0.049 | 0.05 | 0.055 | 0.067 | 0.09 |
| 25 | MOM | 0.062 | 0.054 | 0.043 | 0.039 | 0.038 | 0.037 | 0.038 | 0.044 | 0.051 |
| | MPP | 0.051 | 0.047 | 0.041 | 0.039 | 0.039 | 0.039 | 0.040 | 0.045 | 0.047 |
| | MEP | 0.019 | 0.023 | 0.029 | 0.031 | 0.033 | 0.035 | 0.038 | 0.042 | 0.06 |
| | $\pi_{(k,\delta)} \propto 1$ | 0.076 | 0.038 | 0.034 | 0.034 | 0.036 | 0.038 | 0.042 | 0.053 | 0.068 |
| | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | 0.067 | 0.036 | 0.032 | 0.032 | 0.034 | 0.036 | 0.039 | 0.051 | 0.066 |
| 50 | MOM | 0.037 | 0.034 | 0.026 | 0.024 | 0.024 | 0.024 | 0.025 | 0.029 | 0.034 |
| | MPP | 0.029 | 0.027 | 0.024 | 0.024 | 0.024 | 0.025 | 0.026 | 0.03 | 0.032 |
| | MEP | 0.010 | 0.015 | 0.020 | 0.021 | 0.023 | 0.025 | 0.026 | 0.033 | 0.043 |
| | $\pi_{(k,\delta)} \propto 1$ | 0.034 | 0.022 | 0.020 | 0.021 | 0.023 | 0.025 | 0.027 | 0.035 | 0.044 |
| | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | 0.032 | 0.021 | 0.02 | 0.020 | 0.022 | 0.024 | 0.026 | 0.034 | 0.043 |

Tabla 3: Datos simulados: Error absoluto promedio escalado para varios valores de k y n .

Finalmente, de los resultados obtenidos en la distancia K-L (tabla 4), se puede observar que en general, todas las metodologías presentan una distancia pequeña en el rango $-0.4 \leq k \leq 0.4$, mientras que para $k = 1, 2$, la pérdida de información crece rápidamente.

| n | Método | k | | | | | | | | |
|-----|---|-------|-------|-------|-------|-------|-------|-------|---------|-------------------|
| | | -2.00 | -1.00 | -.40 | -.20 | 0 | .20 | .40 | 1.00 | 2.00 |
| 15 | MOM | 2.367 | 0.608 | 0.438 | 0.685 | 1.179 | 0.584 | 0.394 | 144.817 | 6262471021609.200 |
| | MPP | 1.549 | 0.365 | 0.613 | 0.961 | 1.392 | 0.825 | 0.548 | 128.530 | 4652855158683.580 |
| | MEP | 0.250 | 0.135 | 0.263 | 0.385 | 0.640 | 0.451 | 0.295 | 119.953 | 4199563618322.130 |
| | $\pi_{(k,\delta)} \propto 1$ | 0.111 | 0.100 | 0.272 | 0.397 | 0.747 | 0.510 | 0.373 | 74.288 | 1791702598871.520 |
| | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | 0.179 | 0.142 | 0.303 | 0.479 | 0.843 | 0.539 | 0.374 | 62.338 | 1524786766050.54 |
| 25 | MOM | 2.746 | 0.518 | 0.150 | 0.291 | 1.007 | 0.459 | 0.349 | 103.378 | 4039186768928.850 |
| | MPP | 1.535 | 0.213 | 0.169 | 0.394 | 1.147 | 0.656 | 0.488 | 91.110 | 3187097007511.960 |
| | MEP | 0.196 | 0.092 | 0.101 | 0.169 | 0.575 | 0.342 | 0.254 | 72.206 | 3147387027766.320 |
| | $\pi_{(k,\delta)} \propto 1$ | 0.101 | 0.065 | 0.098 | 0.156 | 0.714 | 0.360 | 0.315 | 53.106 | 1220879094214.370 |
| | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | 0.138 | 0.082 | 0.103 | 0.191 | 0.797 | 0.402 | 0.323 | 47.882 | 1039977826753.370 |
| 50 | MOM | 3.226 | 0.527 | 0.064 | 0.033 | 0.914 | 0.312 | 0.223 | 66.690 | 2710155914544.310 |
| | MPP | 1.465 | 0.153 | 0.046 | 0.038 | 1.067 | 0.469 | 0.346 | 63.391 | 2269867203148.490 |
| | MEP | 0.125 | 0.067 | 0.042 | 0.037 | 0.405 | 0.206 | 0.128 | 34.571 | 2413407805084.580 |
| | $\pi_{(k,\delta)} \propto 1$ | 0.093 | 0.053 | 0.034 | 0.028 | 0.674 | 0.225 | 0.148 | 38.045 | 707004878948.454 |
| | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | 0.110 | 0.063 | 0.037 | 0.033 | 0.724 | 0.245 | 0.156 | 34.993 | 627510119545.924 |

Tabla 4: Datos simulados: Distancia Kullback-Leibler para los valores de k y n .

Este comportamiento no es del todo extraño, ya que la densidad crece a infinito con-

forme $x \rightarrow \frac{k}{\sigma}$ para $k > 1$. Sin embargo, aún en estas condiciones se puede observar que la mejor opción es la estimación Bayesiana con la priori $\log \delta$ localmente uniforme. En particular, para $k \leq 0.4$ el MEP y la estimación Bayesiana presentan el mejor desempeño, sobretodo para muestras de tamaño $n \leq 50$.

En resumen, la estimación Bayesiana con cualquiera de las prioris propuestas en este trabajo compite con los métodos sugeridos en la literatura para todos los valores de k , con tamaños de muestra $n = 15, 25$, mientras que para $n = 50$, la estimación Bayesiana presenta el mejor desempeño y por ende, es la mejor alternativa entre las metodologías revisadas en este trabajo. De los resultados obtenidos en EAPE (tabla 3), se puede observar que para $k < -0.4$ el MEP presenta los valores mas pequeños, mientras que la estimación Bayesiana con la priori $\log \delta$ localmente uniforme presenta valores muy próximos a los del MEP. En el rango $-0.4 \leq k \leq 0.4$ el MEP presenta los EAPE mas pequeños para tamaños de muestra $n = 15, 25$, mientras que la estimación Bayesiana con priori $\log \delta$ localmente uniforme presenta valores EAPE muy cercanos. Para $n = 50$ la estimación Bayesiana con la priori $\log \delta$ localmente uniforme presenta los valores EAPE más pequeños y el desempeño del MEP decrece.

4 Un caso de aplicación: datos de Bilbao

El primer ejemplo citado por Castillo y Hadi consiste en medidas de nivel de las olas en Bilbao, España, con registros obtenidos con sensores instalados en una boya en enero de 1997, dentro del Programa de Clima Marítimo de CEDEX.

Verificando el supuesto PG, en la Figura 1 se puede observar el ajuste lineal de la función de vida media residual estimada versus el umbral. En este caso se puede observar que los datos presentan una fuerte relación lineal ($R^2 > 0.9$), por lo que el supuesto PG parece razonable.

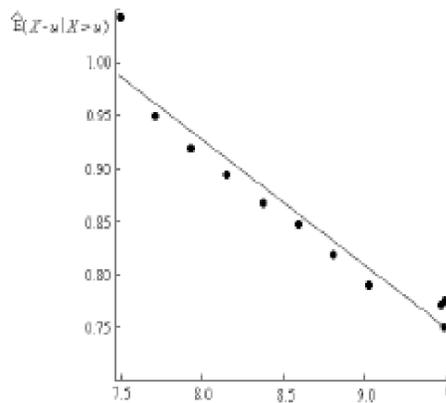


Figura 1:

Castillo y Hadi obtienen los estimadores de k y σ para distintos valores de umbral

con el fin de monitorear el efecto del umbral en los resultados obtenidos aplicando los métodos de MOM, MPP y MEP. Por lo tanto, retomando este problema en la tabla 5 se pueden observar los valores estimados y sus errores estándar de los parámetros k y σ , agregando ahora la estimación Bayesiana con las prioris localmente uniforme y $\log \delta$ localmente uniforme. Nótese que en todos los casos la estimación Bayesiana presenta el error estándar mas bajo para todos los valores del umbral.

| u | m | \hat{k} | | | | | $\hat{\sigma}$ | | | | |
|-----|-----|------------------|------------------|------------------|------------------------------|---|------------------|------------------|------------------|------------------------------|---|
| | | MOM | MPP | MEP | $\pi_{(k,\delta)} \propto 1$ | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | MOM | MPP | MEP | $\pi_{(k,\delta)} \propto 1$ | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ |
| 7.0 | 179 | 1.049 (0.158) | 1.092 (0.149) | 0.815 (0.071) | 0.777 (0.085) | 0.785 (0.088) | 2.749 (0.304) | 2.806 (0.289) | 2.406 (0.196) | 2.319 (0.037) | 2.338 (0.036) |
| 7.5 | 154 | 0.611 (0.114) | 0.627 (0.121) | 0.685 (0.079) | 0.722 (0.041) | 0.725 (0.049) | 1.632 (0.186) | 1.649 (0.190) | 1.698 (0.175) | 1.775 (0.056) | 1.782 (0.057) |
| 8.0 | 106 | 0.647 (0.158) | 0.661 (0.168) | 0.742 (0.099) | 0.737 (0.083) | 0.746 (0.087) | 1.385 (0.209) | 1.397 (0.215) | 1.456 (0.192) | 1.439 (0.122) | 1.453 (0.123) |
| 8.5 | 69 | 0.723 (0.229) | 0.749 (0.240) | 0.814 (0.152) | 0.823 (0.133) | 0.841 (0.140) | 1.130 (0.228) | 1.147 (0.232) | 1.176 (0.197) | 1.182 (0.112) | 1.204 (0.114) |

Tabla 5: Datos de olas de Bilbao: parámetros estimados por los 5 métodos y distintos valores del umbral (u); m es el número de excedencias sobre u , los errores estándar se presentan entre paréntesis.

Por otro lado, en la tabla 6 se muestran los errores absolutos escalados (**EAPE**) y una estimación de la distancia Kullback-Leibler mediante el Criterio de Información de Akaike de segundo orden (**AIC_C**)

$$AIC_C = -2 \log(L(\hat{\theta})) + 2K \left(\frac{n}{n - K - 1} \right) \quad (2)$$

en donde n es el tamaño de la muestra, K es el número de parámetros que intervienen en el modelo y $L(\hat{\theta})$ es la log-verosimilitud evaluada en el vector de parámetros estimado.

Nótese que similarmente al caso del error estándar, los valores mas pequeños del error absoluto promedio escalado y la estimación de la distancia Kullback-Leibler corresponden a la estimación Bayesiana.

| u | m | Error absoluto promedio escalado | | | | | Información Kullback-Leibler | | | | |
|-----|-----|----------------------------------|-------|-------|------------------------------|---|------------------------------|---------|---------|------------------------------|---|
| | | MOM | MPP | MEP | $\pi_{(k,\delta)} \propto 1$ | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ | MOM | MPP | MEP | $\pi_{(k,\delta)} \propto 1$ | $\pi_{(k,\delta)} \propto \frac{1}{\delta}$ |
| 7.0 | 179 | 0.038 | 0.039 | 0.031 | 0.024 | 0.025 | 348.630 | 342.740 | 383.830 | 385.250 | 384.970 |
| 7.5 | 154 | 0.020 | 0.021 | 0.023 | 0.020 | 0.020 | 272.060 | 271.097 | 267.729 | 267.118 | 267.102 |
| 8.0 | 106 | 0.024 | 0.025 | 0.026 | 0.018 | 0.018 | 145.670 | 144.948 | 140.511 | 140.416 | 140.203 |
| 8.5 | 69 | 0.031 | 0.032 | 0.032 | 0.021 | 0.021 | 57.615 | 56.554 | 53.122 | 52.757 | 52.473 |

Tabla 6: Datos de olas de Bilbao: Error absoluto promedio escalado e información Kullback-Leibler estimada; m es el número de excedencias sobre el umbral u .

5 Conclusiones

El objetivo principal en este trabajo fue determinar una metodología confiable para estimar los parámetros de la DPG en todo el soporte paramétrico natural ($\sigma > 0, k \in \mathbb{R}$), ya

que las metodologías propuestas en la literatura pueden presentar problemas. Se sugiere la estimación Bayesiana con prioris no informativas (localmente uniforme y $\log \delta$ localmente uniforme), encontrando en los resultados de simulación que el mejor desempeño en términos de sesgo y de RECM, corresponde al método de Bayes con la función de pérdida de error cuadrado medio, por lo que las estimaciones de los parámetros son simplemente los valores esperados de k y σ en la distribución a posteriori.

Por la forma de la PG, es necesario integrar en forma numérica y dado que la distribución posteriori es propia, no se detecta problema alguno integrando con el método de cuadratura conocido como regla compuesta de Simpson. Por tanto, se pueden obtener estimaciones de cualquier cantidad posterior de interés como regiones de confianza, factor de Bayes o cuantiles posteriores.

Utilizando el EAPE y distancia K-L como una medida de bondad de ajuste entre los métodos analizados, se observa nuevamente en los resultados del experimento Monte Carlo que las estimaciones con metodología Bayesiana presenta un mejor ajuste a los datos con la distribución a priori no informativa $\log \delta$ localmente uniforme. Se aplicaron los métodos disponibles en la literatura y la estimación Bayesiana a tres conjuntos de datos reales, observando que los estimadores con el método de Bayes presentan el menor error estándar. Por último, la sugerencia final es utilizar la metodología Bayesiana con la priori $\log \delta$ localmente uniforme para estimar los parámetros en la DPG cuando no se tiene información a priori disponible.

Referencias

- [1] Arnold, B.; Press, S.J. (1983) "Bayesian inference for Pareto populations", *Journal of Econometrics* **21**: 287–306.
- [2] Castillo, E.; Hadi, A.S. (1997) "Fitting the generalized Pareto distribution to data", *JASA* **92**(440): 1609–1620.
- [3] Davison, A.C.; Smith, R.L. (1990) "Models for exceedances over high thresholds", (with discussion), *JRSS, series B* **52**: 393–442.
- [4] Grimshaw, S.D. (1993) "Computing maximum likelihood estimates for the generalized Pareto distribution", *Technometrics* **35**: 185–191.
- [5] Hosking, J.R.M.; Wallis, J.R. (1987) "Parameter and quantile estimation for the generalized Pareto distribution", *Technometrics* **29**(3): 339–349.
- [6] Huang, L.; Smith, R.L. (1999) "Meteorologically-dependent trends in urban ozone", *Environmetrics* **10**: 103–118.
- [7] Kullback, S. (1959) *Information Theory and Statistics*. Dover Publications, U.S.A.
- [8] Lindsey, J.K. (1996) *Parametric Statistical Inference*. Clarendon Press, Oxford.
- [9] Pickands, J. (1975) "Statistical Inference using extreme order statistics", *Annals of Statistics* **3**: 119–131.

- [10] Robert, C.P. (1994) *The Bayesian Choice. A Decision-Theoretic Motivation*. Springer, New York.
- [11] Sánchez-Gómez, R. (2001) *Análisis de Tendencia en Excedencias sobre un Umbral Alto, con Aplicación en Ozono Urbano*. Tesis de Doctorado, Colegio de Postgraduados, México.
- [12] Smith, R.L. (1984) “Discussion on a model fitting analysis of daily rainfall data”, *J. R. Statistic. Soc., series A* 147: 1–34.
- [13] Yang, R.; Berger, J.O. (1998) “A catalog of noninformative priors”, Technical Report, Purdue University.