Mario M. Pizarro

Baseline Methods for Bayesian Inference in Gumbel copula

Mario M. Pizarro[1,*]; Jacinto Martin[2]; Eva L. Sanjuán[3]; María I. Parra[4]

[1] Department of Mathematics, University of Extremadura, Badajoz, Spain, mariomp@unex.es
[2] Department of Mathematics, University of Extremadura, Badajoz, Spain, jrmartin@unex.es
[3] Department of Mathematics, University of Extremadura, Badajoz, Spain, etlopez@unex.es
[4] Department of Mathematics, University of Extremadura, Badajoz, Spain, mipa@unex.es

Abstract:

Standard methods usually employed to estimate the parameters of extreme value distributions use only a small part of the observed values. When block maxima values are considered, many data are discarded. This wasteful of information is more pronounced when we select the block maxima for two sets of correlated values. The use of copulas allows to model the structure of dependence between the observed sets. In addition, copulas define multivariate distributions with the desired univariate marginals in each situation. There are relationships between the parameters of different distributions and the parameters of the distribution of block maxima. Also, there are relationships between the Archimedean copula's parameter and Gumbel extreme value copula's parameter. The new Bayesian estimation methods are based on them. Baseline distribution method (BDM) determines the estimations of the base copula's parameters as well as of the univariate baseline distributions' parameters with all the data. Improved baseline distribution method (IBDM) aims to give more importance to the block maxima data than to the values of the base copula. It is performed by applying BDM to obtain a highly informative prior distribution for the parameters of distribution of extreme value given the relationship with baseline distribution's parameters. We compare empirically, through a broad simulation study, these new methods with Standard Bayesian analysis for uninformative prior distributions (Metropolis-Hastings method, MHM). Considering Joe's Archimedean copula with univariate marginal Normal distributions. Also, the Joe copula leading to extreme Gumbel copula and the Normal distribution leading to an extreme Gumbel distribution. When the block size is small, the results show that IBDM has intermediate behavior between MHM and BDM. When the block size is large enough, IBDM shows better behavior than BDM that remains stable.

Keywords:

Bayesian inference; highly informative prior; Archimedean copula; Gumbel copula; Extremes values distributions

1. Introduction:

Extreme Value Theory (EVT) is used to model and predict distributions that appear when we study extreme events. EVT is employed in several scientific fields as rare events appear in temperatures, precipitations, finances, … The usual data fitting techniques work well in the central area of the distribution, but poorly in the tail area due to the low number of extreme

observations. In EVT, there are two main approaches, block maxima method (BM) and peaks-over-threshold method (POT). The difference between them is based on the way each model classifies which observations are considered extreme events. For BM method, data are divided into blocks of equal size and the maximum is taken for each block. The challenge of this method lies in deciding the size of the blocks when they are not obvious. Meanwhile, in the case of the POT method, deciding the threshold value might be difficult. However, a common feature is that both strategies dispense with many available data.
Given a sequence of independent and identically distributed (i.i.d.) random variables $Y_1, \dots, Y_m$ with common distribution function F (baseline distribution), and given a fixed $k \in \mathbb{N}$ (block size), we define the block maxima as

$$X_i = \max_{(i-1)k<j\leq ik} Y_j \, , i = 1,2, \dots, n$$

Therefore, the whole set of observations, $m = k \times n$, are divided into n blocks of size k. Also, according to the Gnedenko (1943) and Fisher and Tippet (1928) theorems, the asymptotic distribution of block maxima of i.i.d. random variables can be approximated by the generalized extreme values (GEV) distribution, with distribution function

$$\mathrm{GEV}(\mathrm{x}; \xi, \mu, \sigma) = \exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\}$$

with $\xi, \mu \in \mathbb{R}, \sigma > 0$, and $1+\xi\frac{x-\mu}{\sigma} > 0$.
In addition, when $\xi = 0$, the right-hand side of previous equation is interpreted as

$$\mathrm{G}(\mathrm{x}; \mu, \sigma) = \exp\left\{-\,exp\,\left\{-\frac{x-\mu}{\sigma}\right\}\right\}$$

and it is called Gumbel distribution with parameters μ (location) and σ (scale).
Definition 1. We say that the distribution function F is in the domain of attraction of Gumbel distribution when there exist sequences $\{a_k\}$ and $\{b_k\}$, with $a_k > 0$, $b_k \in \mathbb{R}$ such that

$$\lim_{k\to\infty} F^k\,(a_k x + b_k) = \mathrm{G}(x), \mathrm{x} \in \mathbb{R}$$

The domain of attraction of a distribution allows us to relate the parameters of different baseline distributions to the limit Gumbel distribution's parameters. Martin *et al.* (2020) shows the relationships for different baseline distributions. In particular, for the Normal distribution with mean μ_N and standard deviation σ_N, the relationships are

$$\begin{cases} \mu = \mu_N + \sigma_N\left[(2\ ln\ k)^{1/2} - \dfrac{ln\, ln\, k \ \ + \ ln\, 4\pi}{2(2\ ln\ k)^{1/2}}\right] \\ \sigma = \sigma_N(2\ ln\ k)^{-1/2} \end{cases}$$

Copula theory is increasingly used in multivariate models of extreme values. A copula C is a multivariate distribution function with uniform marginal univariates. However, according to the Sklar theorem (1959), copulas allow to obtain multivariate distributions with desired marginal univariates. An important feature of copulas is that they reflect the dependency structure between observations regardless of marginal distributions. Nelsen (2007) showed many different families of copulas, including Archimedean copulas and Extreme Value copulas (EVC).

Definition 2. According to Genest and Mackay (1996), we say that a 2-dimensional copula C is Archimedean with generator ϕ: $[0,1] \to [0,1]$ such that $\phi(1) = 0, \phi'(u) < 0$ and $\phi''(u) > 0$ if

$$C(u,v) = \begin{cases} \phi^{-1}\big(\phi\,(u) + \phi\,(v)\big) & if \quad \phi\,(u) + \phi(v) \leq \ \phi\,(0) \\ 0 & otherwise \end{cases}$$

Besides, Joe (1997) defined a 2-dimensional extreme value copula C^* as the copula that satisfies

$$C^*\big(u^k, v^k\big) = C^*(u,v)^k, \forall k \ > \ 0$$

In particular, the 2-dimensional Gumbel copula is an extreme value copula, and it is defined by its distribution function

$$C_G(u,v;\theta) = exp\left\{-\left[(-\,ln\, u)^{\theta} \ + \ (-\,ln\, v)^{\theta}\right]^{1/\theta}\right\}, u, v \in [0,1], \theta \geq 1$$

Also, notice that the Gumbel copula with parameter $\theta = 1$ represents independence between observations.

The use of copula theory in multivariate EVT is due to the following relationship. Given the sequences of 2 i.i.d. random variables $(Y_{11}, Y_{21}), \dots, (Y_{1m}, Y_{2m})$, the joint limit distribution of block maxima $(X_{1i}, X_{2i}), i = 1, \dots, n$ is such that

$$G(x_1, x_2) = C^*\big(G_1(x_1), G_2(x_2)\big)$$

where C^* is an extreme value copula, and G_1, G_2 are univariate extreme value distributions, that is, GEV distributions. As a result of this connection with EVT, the concept of domain of attraction also appears in copula theory. Hence, multivariate distributions of extreme values can be defined through extreme value copulas.

Definition 3. We say that the copula C is in the domain of attraction of an extreme value copula C^* when

$$\lim_{k\to\infty} C^k\left(u^{1/k}, v^{1/k}\right) = C^*(u, v), u, v \in [0,1]$$

In addition, there is a relationship between the Archimedean copula's parameters and the extreme value Gumbel copula's parameter. The following result allows us to obtain that relationship.

Proposition 1. If C is an Archimedean copula with generator ϕ_θ and there exists

$$\lim_{\omega\to 0} -\omega \frac{\phi_\theta'(1-\omega)}{\phi_\theta(1-\omega)} = \theta_1 \in [1, +\infty)$$

Then C is in the domain of attraction of an extreme value Gumbel copula with parameter θ_1.

In particular, we will consider two-dimensional Joe copula with parameter $\theta \geq 1$ and generator $\phi_\theta(\omega) = -\ln\big(1 - (1-\omega)^\theta\big)$, whose distribution function is

$$C_J(u, v;\ \theta) = 1 - \left[(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta\right]^{\frac{1}{\theta}}, u, v \in [0,1]$$

This copula is in the domain of attraction of an extreme value Gumbel copula with parameter θ.

In general, given a copula C with marginal univariates F_1, F_2, C is in the domain of attraction of an extreme value copula C^* with marginal distributions G_1, G_2 which are extreme value distributions. Also, F_1, F_2 are in the domain of attraction of G_1, G_2, respectively.

2. Methodology:

The inference methods used for extreme value distributions require conditions on the shape parameter ξ, because of the asymptotic theory on which they are based. This problem is saved with a Bayesian approach that provides the advantage of including additional information through the chosen prior distribution.

New methods to estimate Gumbel distribution's parameters as block maxima distribution are shown in Martin *et al.* (2020). These methods are compared to the Metropolis-Hastings method (MHM) that uses non-informative prior distributions. The proposed methods are Baseline Distribution method (BDM) and Improved Baseline Distribution method (IBDM). BDM estimates the parameters of the baseline distribution, and IBDM uses highly informative prior distributions to estimate the parameters of the Gumbel distribution of block maxima. Also, IBDM uses the relationship between the parameters of baseline distribution and the parameters of Gumbel distribution of block maxima to construct the prior distribution. All methods are based on the MCMC techniques, concretely a Metropolis-Hastings (MH) algorithm.

IBDM is a method that gives more weight to block maxima data than to baseline data. It is quite similar to a BDM when the block size is large, therefore there is little block maxima data. However, when the block size becomes smaller (more block maxima data), IBDM approaches MHM.

This work presents two-dimensional cases of the BDM and IBDM. Given Proposition 1, when the base copula is Archimedean, there exists a relationship between the parameters of the base copula with the Gumbel extreme value copula's parameter.
In MHM and BDM, to estimate the parameters of the baseline and Gumbel marginal distributions, the prior distributions proposed in Martin *et al.* (2020) are considered. In addition, an $U(1{,}1000)$ is settled for the parameters of the base copula and Gumbel copula.
Improved Baseline Distribution Method: Prior distributions are considered for Gumbel copula's parameters and Gumbel marginal distributions' parameters. Then, we take advantage of the relationships between parameters of baseline distributions with extreme value distributions' parameters. Prior distributions of the parameters are obtained by applying the transformations to posterior distribution given by BDM.

3. Result:

Joe copula's parameter $\theta = 2$ is considered to be a base copula with marginals $Y_1 \sim N(0,1)$ and $Y_2 \sim N(0,\sigma)$, with $\sigma = 2^{-1}, 2$. Proposition 1 indicates that Joe Archimedean copula is in the domain of attraction of the Gumbel extreme value copula with parameter $\theta_1 = 2$. In addition, Normal distributions are in the domain of attraction of the Gumbel distribution of block maxima.
We made a simulation study for the two cases considered. For each case, we generated $m_{ij} = k_j \times n_i$ values of the base copula (with the marginals indicated), where

- k_j is the block size, $k_j = 10^j, j = 1,2$; and
- n_i is the number of block maxima, $n_i = 2^i, i = 3,4,\dots,7$.

Besides, each sequence is replicated $M = 100$ times.
MHM generally provides high skewed estimations for the posterior distributions than IBDM. Figure 1 shows that for the Gumbel copula parameter and the scale parameters of Gumbel marginal, MHM presents a right skew compared to IBDM, when the number of block maxima is small. However, when there are many block maxima values, MHM and IBDM provide highly concentrated estimations. In addition, IBDM has less variability than MHM (Figure 2). Both methods present similar estimations for the location parameters of Gumbel distribution.

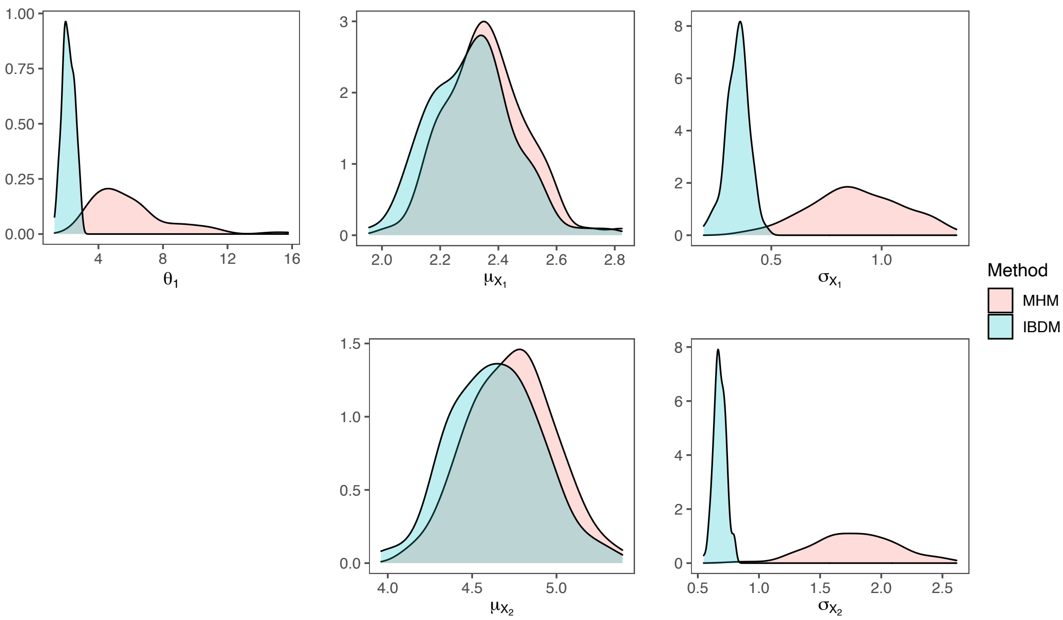

Figure 1. Probability density functions for M estimations of the block maxima parameters θ_1 (top left), μ (centre) and σ (right), obtained for the methods MHM and IBDM, with $k = 100$ and $n = 8$, from marginals $Y_1 \sim N(0,1)$ and $Y_2 \sim N(0,2)$.

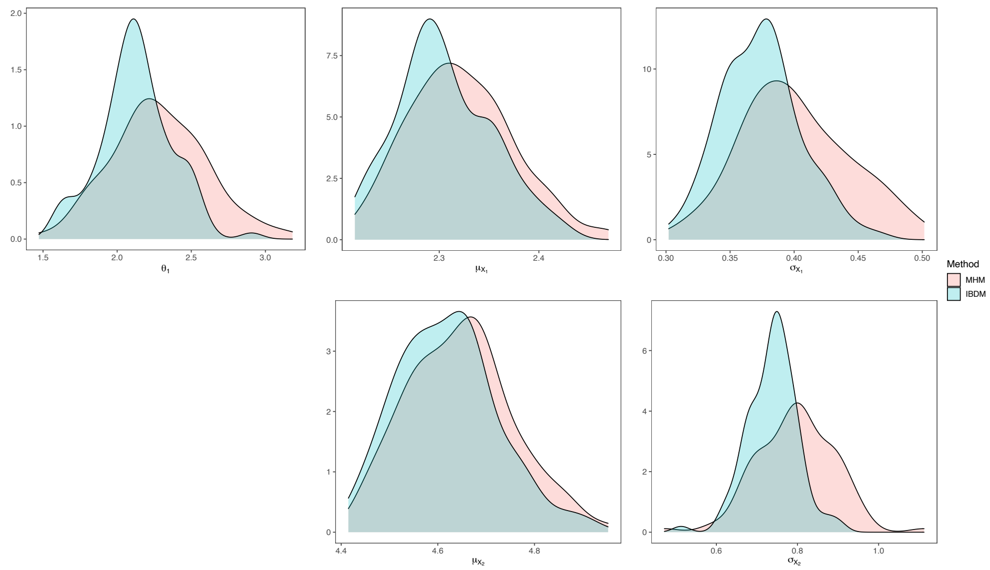

Figure 2. Probability density functions for M estimations of the block maxima parameters θ_1 (top left), μ (centre) and σ (right), obtained for the methods MHM and IBDM, with $k = 100$ and $n = 64$, from marginals $Y_1 \sim N(0,1)$ and $Y_2 \sim N(0,2)$.

To compute measures of error in order to evaluate the quality of estimations, we compared estimated distribution functions (C^*) with real ones (C^k) through their mean absolute distance (AD). As analytical computation is not possible, we made a Monte-Carlo computation employing sample size $s = 10^4$. Then,

$$AD_j = \frac{1}{s} \sum_{i=1}^{s} |C^*\left(x_{1i}, x_{2i};\ \hat{v}_j\right) - C^k(x_{1i}, x_{2i};\ v)|$$

with $j = 1, \ldots, M$, where M is the number of samples. Also, $C^*(x_{1i}, x_{2i};\ \hat{v})$ denotes the estimated Joe copula with Normal marginal distributions, for the baseline parameter $\hat{v}$. And, for k big enough, $C^*(x_{1i}, x_{2i};\ \hat{v})$ is estimated Gumbel copula, with Gumbel marginal distributions, where $\hat{v}$ denotes block maxima parameters. Then, we employed Mean Absolute Error (MAE), defined as

$$MAE = \frac{1}{M} \sum_{j=1}^{M} AD_j$$

Figure 3 shows MAE for the three methods. When block size is small (top panels), BDM is the best method. However, for bigger values of k (bottom panels), IBDM provide better results than BDM. This method provides more uniform error values for different values of number of block maxima (n). When n is big enough, MHM and IBDM offer similar values. Also, when the block size is small, IBDM provides error measures which are intermediate between MHM and BDM.

4. Discussion and Conclusion:

1. In EVT, the estimation of the parameters of the distribution is one of the most common problems, because data are usually scarce. This is more remarkable when we have multivariate situations. In this work, we considered the case when block maxima bivariate distribution is a Gumbel copula, and we developed two Bayesian methods, BDM and IBDM, to estimate the posterior distribution of the parameters.
2. BDM and IBDM methods make use of all the available data from the base copula, not only block maxima values.
3. We obtained that posterior distributions of the parameters for IBDM are more concentrated and less skewed than the ones offered through MHM.

4. In general, the results obtained show that BDM and IBDM offer lower measures of error. MHM shows the worst results, especially when extreme data are scarce.
5. When block size gets bigger values, IBDM is the best method, while BDM provides more uniform results for different values of the number of block maxima.

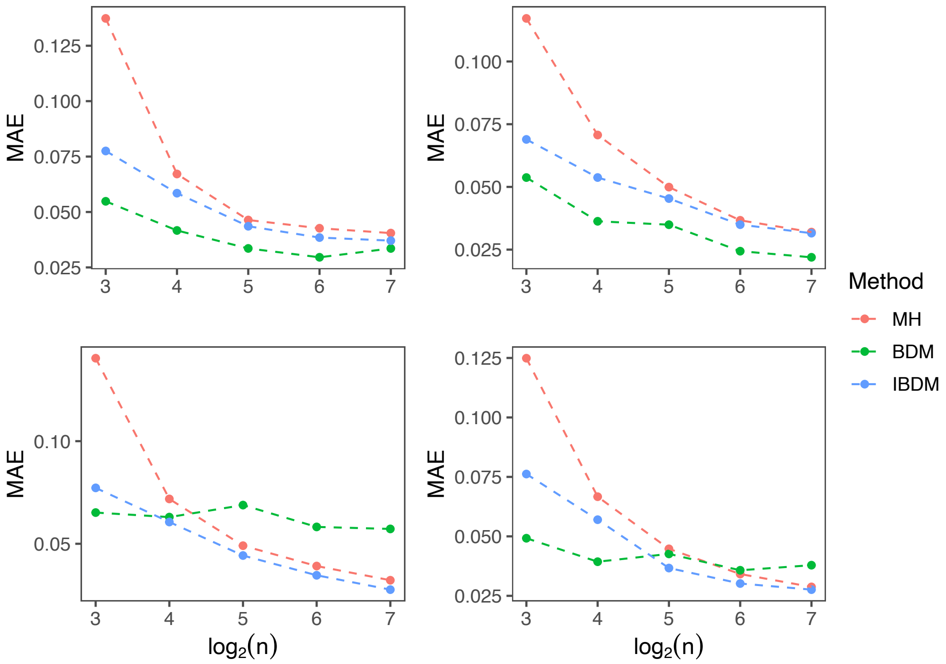

Figure 3. Mean absolute error (MAE) for the three methods MHM (red), BDM (green) and IBDM (blue) with $k = 10$ (top), 100 (bottom), and different values of n, from marginal distributions $Y_1 \sim N(0,1)$ and $Y_2 \sim N(0,1/2)$ (left) and $Y_1 \sim N(0,1)$ and $Y_2 \sim N(0,2)$ (right).

Funding: This work was partially supported by Junta de Extremadura, Consejería de Economía e Infraestructuras FEDER Funds IB16063, Project MTM2017-86875-C3-2-R from Ministerio de Economía, Industria y Competitividad de España, and GR18108 project from Junta de Extremadura.

References:
1. Gnedenko, B. (1943). Sur la distribution limite du terme máximum d'une serie aleatoire. *Ann. Math.*, 44, 423-453.
2. Fisher, R.A.; Tippet, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*; Cambridge University Press: Cambridge, UK; Volume 24, pp. 180-190.
3. Martin, J.; Parra, M.I.; Pizarro, M.M.; Sanjuán, E.L. (2020). Baseline Methods for Bayesian Inference in Gumbel Distribution. *Entropy*, 22, 1267. https://doi.org/10.3390/e22111267
4. Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8, 229-331.
5. Nelsen, R.B. (2007). An introduction to copulas. *Springer Science & Business Media*.
6. Genest, C. and Mackay, J. (1996). The joy of copulas: Bivariate distributions with uniform marginals, *American Statistician*, 40, 280-283.
7. Joe, H. (1997). Multivariate Models and Dependence Concepts, *Monographs on Statistics and Applied Probability*, 73, Chapman & Hall, London