Journal of the Korean Data & Information Science Society
2013, **24**(1), 171–178

http://dx.doi.org/10.7465/jkdi.2013.24.1.171
한국데이터정보과학회지

Noninformative priors for the shape parameter in the generalized Pareto distribution

Sang Gil Kang[1] · Dal Ho Kim[2] · Woo Dong Lee[3]

[1]Department of Computer and Data Information, Sangji University
[2]Department of Statistics, Kyungpook National University
[3]Department of Asset Management, Daegu Haany University

Received 26 October 2012, revised 9 December 2012, accepted 14 December 2012

Abstract

In this paper, we develop noninformative priors for the generalized Pareto distribution when the parameter of interest is the shape parameter. We developed the first order and the second order matching priors. We revealed that the second order matching prior does not exist. It turns out that the reference prior satisfies a first order matching criterion, but Jeffrey's prior is not a first order matching prior. Some simulation study is performed and a real example is given.

Keywords: Generalized Pareto distribution, matching prior, reference prior, shape parameter.

1. Introduction

The generalized Pareto distribution (GPD) introduced independently by Pickands (1975) and Balkema and de Haan (1974) as a limiting distribution for scaled excesses over a high threshold has become a central notion in the statistical analysis of extreme events (Smith, 1987; Davison and Smith, 1990; Embrechts *et al.*, 1997). Applications of the GPD to areas such as insurance, reliability, finance, meteorology and environment are widely spread out in the literature.

The GPD is a distribution that contains uniform, exponential and Pareto distribution as special cases. The probability density function (pdf) with scale parameter σ and shape parameter ξ is

$$f(x|\xi,\sigma) = \begin{cases} \sigma^{-1}(1+x\xi/\sigma)^{-(1+\xi)/\xi}, & \xi \neq 0, \\ \sigma^{-1}\exp(-x/\sigma), & \xi = 0, \end{cases} \tag{1.1}$$

where $\sigma > 0$. The shape parameter ξ plays an important role of determining the tail shape of the GPD. The support is $x > 0$ for $\xi \geq 0$ and $0 \leq \xi \leq -\sigma/\xi$, respectively. A random variable X with pdf (1.1) will be denoted by $X \sim \text{GPD}(\xi,\sigma)$.

[1] Associate professor, Department of Computer and Data Information, Sangji University, Wonju 220-702, Korea.
[2] Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea.
[3] Corresponding author: Professor, Department of Asset Management, Daegu Haany University, Kyungsan 712-715, Korea. E-mail: wdlee@dhu.ac.kr

Estimating upper quantiles in financial risk management is very important. Banks want to model and predict their loss in a certain period and use 99.9% value at risk (VaR) as a basis to compute the required capital. Simply, the 99.9% VaR implies the 99.9% quantile of the loss distribution. Financial loss data tends to have heavy upper tails that corresponds to the case $\xi > 0$ (see Ho, 2010). Thus we only consider GPD with $\xi > 0$ in this paper.

Estimation for GPD parameters are usually done via frequentist methods like maximum likelihood estimation (see Davison and Smith, 1990), probability weighted moments by Hosking and Wallis (1987) and elemental percentile method by Castillo and Hadi (1997).

In the context of estimating GPD parameters based on Bayesian methods, Arnold and Press (1989) explored the use of informative priors for Pareto distribution. Behrens *et al.* (2004) proposed prior elicitation following Coles and Powell (1996). de Zea Bermudez and Amaral Turkman (2003) proposed the use of vague priors. Castellanos and Cabras (2007) proposed using Jeffreys' prior as a default procedure when there is no prior information. Recently, Ho (2010) proposed the matching prior via prediction criterion for the purpose of estimating upper quantiles of GPD.

The present paper focuses on developing noninformative priors for the shape parameter of GPD(ξ, σ) distribution with $\xi > 0$. In the absence of sources of information or past data, Bayesian methods rely on the objective priors or the noninformative priors.

We consider Bayesian priors such that the resulting credible intervals for the shape parameter have coverage probabilities equivalent to their frequentist counterparts. This kind of prior is called the probability matching prior. From the work of Welch and Peers (1963), the probability matching prior was proposed and developed by many authors such as Stein (1985), Tibshirani (1989), Mukerjee and Dey (1993), Datta and Ghosh (1995, 1996) and Mukerjee and Ghosh (1997). Although this matching can be justified only asymptotically, our simulation results indicate that this is indeed achieved for small or moderate sample sizes as well.

Bernardo (1979) introduced the reference priors which maximizes the Kullback-Leibler divergence between the prior and the posterior. This means the reference priors give the least information to the posterior. Ghosh and Mukerjee (1992) and Berger and Bernardo (1989,1992) developed a general algorithm to derive a reference prior by splitting the parameters into several groups according to their order of inferential importance. This approach is very successful in various practical problems (Kang, 2011; Kim *et al.*, 2009). Quite often reference priors satisfy the matching criterion described earlier.

The outline of the remaining sections is as follows. In Section 2, we develop the first order probability matching priors for the shape parameter. Next we derive Fisher information matrix, and also derive the reference priors for the shape parameter. It turns out that the second order matching prior does not exist. The one-at-a-time reference prior satisfies the first order matching criterion, but Jeffreys' prior does not a first order matching prior. In Section 3, We provide that the propriety of the posterior distribution for a general class of prior distributions which include the reference priors as well as first order matching prior. In Section 4, simulated frequentist coverage probabilities under the proposed priors using simulation and real data are given.

2. The noninformative priors

2.1. The probability matching priors

For a prior π, let $\theta_1^{1-\alpha}(\pi;\mathbf{X})$ denote the $(1-\alpha)$th percentile of the posterior distribution of θ_1, that is,

$$P^{\pi}[\theta_1 \le \theta_1^{1-\alpha}(\pi;\mathbf{X})|\mathbf{X}] = 1-\alpha, \tag{2.1}$$

where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_t)^T$ and θ_1 is the parameter of interest. We want to find priors π for which

$$P[\theta_1 \le \theta_1^{1-\alpha}(\pi;\mathbf{X})|\boldsymbol{\theta}] = 1-\alpha + o(n^{-u}). \tag{2.2}$$

for some $u > 0$, as n goes to infinity. Priors π satisfying (2.2) are called matching priors. If $u = 1/2$, then π is referred to as a first order matching prior, while if $u = 1$, π is referred to as a second order matching prior.

In order to find such matching priors π, let

$$\theta_1 = \xi \text{ and } \theta_2 = (1+\xi)\sigma.$$

With this parametrization, the likelihood function of parameters (θ_1, θ_2) for the model (1.1) is given by

$$L(\theta_1,\theta_2) \propto \frac{1+\theta_1}{\theta_2}\left(1+\frac{\theta_1(1+\theta_1)}{\theta_2}x\right)^{-\frac{1+\theta_1}{\theta_1}}. \tag{2.3}$$

Based on (2.3), the Fisher information matrix is given by

$$\mathbf{I}(\theta_1,\theta_2) = \begin{pmatrix} \frac{1}{(1+\theta_1)^2} & 0 \\ 0 & \frac{1}{(1+2\theta_1)\theta_2^2} \end{pmatrix}. \tag{2.4}$$

From the above Fisher information matrix $\mathbf{I}$, θ_1 is orthogonal to θ_2 in the sense of Cox and Reid(1987). Following Tibshirani(1989), the class of the first order probability matching prior is characterized by

$$\pi_m^{(1)}(\theta_1,\theta_2) \propto (1+\theta_1)^{-1} d(\theta_2), \tag{2.5}$$

where $d(\theta_2) > 0$ is an arbitrary function differentiable in its argument.

The class of prior given in (2.5) can be narrowed down to the second order probability matching priors as given in Mukerjee and Ghosh (1997). A second order probability matching prior is of the form (2.5), and also d must satisfy an additional differential equation of Mukerjee and Ghosh (1997), namely

$$\frac{1}{6} d(\theta_2)\frac{\partial}{\partial\theta_1}\{I_{11}^{-\frac{3}{2}} L_{1,1,1}\} + \frac{\partial}{\partial\theta_2}\{I_{11}^{-\frac{1}{2}} L_{112} I^{22} d(\theta_2)\} = 0, \tag{2.6}$$

where

$$L_{1,1,1} = E\left[\left(\frac{\partial \log L}{\partial\theta_1}\right)^3\right] = \frac{2(5+3\theta_1)}{(1+\theta_1)^3(1+3\theta_1)},$$

$$L_{112} = E\left[\frac{\partial^3 \log L}{\partial\theta_1^2\partial\theta_2}\right] = \frac{2}{(1+\theta_1)(1+2\theta_2)(1+3\theta_1)\theta_2}, I_{11} = \frac{1}{(1+\theta_1)^2}, I^{22} = (1+2\theta_1)\theta_2^2.$$

Then (2.6) simplifies to

$$\frac{1+3\theta_1}{12}\frac{\partial}{\partial\theta_1}\left\{\frac{2(5+3\theta_1)}{1+3\theta_1}\right\}+\frac{1}{d(\theta_2)}\frac{\partial}{\partial\theta_2}\left\{\theta_2 d(\theta_2)\right\}=0. \tag{2.7}$$

Note that the first term in the left hand side of (2.7) depends only on the model, whereas the second term involves the prior. Also, the first term depends only on θ_1 and the second only on θ_2. There can be no solution to (2.7) unless the first term is a constant. The first term of (2.7) is not a constant. Therefore this rules out the existence of the second order matching priors.

2.2. The reference priors

Reference priors introduced by Bernardo (1979), and extended further by Berger and Bernardo (1992) have become very popular over the years for the development of noninformative priors. In this Section, we derive the reference priors for different groups of ordering of (θ_1, θ_2). Then due to the orthogonality of the parameters, following Datta and Ghosh (1995), choosing rectangular compacts for each θ_1, θ_2 when θ_1 is the parameter of interest, the reference priors are given by as follows.

If θ_1 is the parameter of interest, then the reference prior distributions for different groups of ordering of (θ_1, θ_2) are:

Group ordering	Reference prior	
$\{(\theta_1,\theta_2)\}$,	$\pi_J \propto (1+\theta_1)^{-1}(1+2\theta_1)^{-\frac{1}{2}}\theta_2^{-1}$,	(2.8)
$\{\theta_1,\theta_2\}$,	$\pi_r \propto (1+\theta_1)^{-1}\theta_2^{-1}$.	(2.9)

Remark 2.1 In the above reference priors, the one-at-a-time reference prior, π_r, satisfy the first order matching criterion. And Jeffreys' prior, π_J, is not a first order matching prior.

Notice that the matching priors (2.5) include many different matching priors because of the arbitrary selection of the function g. And for some functions, there does not seem to be any improvement in the coverage probabilities with these posteriors. So we consider a particular first order matching prior where g is θ_2^{-1} in matching priors (2.5). This prior is given by

$$\pi_m(\theta_1,\theta_2) \quad \propto \quad (1+\theta_1)^{-1}\theta_2^{-1} \tag{2.10}$$

and is the one-at-a-time reference prior.

Remark 2.2 In the original parametrization (ξ, σ), Jeffreys' prior and the reference prior are given by

$$\pi_J(\xi,\sigma) \quad \propto \quad (1+\xi)^{-1}\sigma_1^{-1}, \tag{2.11}$$

$$\pi_r(\xi,\sigma) \quad \propto \quad (1+\xi)^{-1}(1+2\xi)^{-\frac{1}{2}}\sigma_1^{-1}. \tag{2.12}$$

3. Propriety of the posterior distribution

We investigate the propriety of posteriors for a general class of priors which include Jeffreys' prior (2.11) and the reference prior (2.12). We consider the class of priors

$$\pi(\xi, \sigma) \propto (1+\xi)^{-a}(1+2\xi)^{-b}\sigma^{-1}, \tag{3.1}$$

where $a > 0$ and $b \geq 0$. The following general theorem can be proved.

Theorem 3.1 The posterior distribution of (ξ, σ) under the prior π, (3.1), is proper if $n + a + b - 2 > 0$.

Proof: Note that the joint posterior for ξ and σ given $\mathbf{x}$ is

$$\pi(\xi, \sigma|\mathbf{x}) \quad \propto \quad \sigma^{-n-1}(1+\xi)^{-a}(1+2\xi)^{-b}\prod_{i=1}^{n}\left(1+\frac{\xi}{\sigma}x_i\right)^{-\frac{1+\xi}{\xi}}. \tag{3.2}$$

Then we obtain

$$\pi(\xi, \sigma|\mathbf{x}) \quad \leq \quad \sigma^{-n-1}(1+\xi)^{-a}(1+2\xi)^{-b}\left(1+\frac{\xi}{\sigma}z\right)^{-n\frac{1+\xi}{\xi}} \equiv \pi'(\xi, \sigma|\mathbf{x}), \tag{3.3}$$

where $z = \min\{x_1, \cdots, x_n\}$. Integrating with respect to σ from (3.3), then we get

$$\begin{aligned}
\pi'(\xi|\mathbf{x}) &\propto \xi^{-n}(1+\xi)^{-a}(1+2\xi)^{-b}\frac{\Gamma\left[\frac{n}{\xi}\right]}{\Gamma\left[n+\frac{n}{\xi}\right]} \\
&\propto \frac{\xi^{-n}(1+\xi)^{-a}(1+2\xi)^{-b}}{\frac{n}{\xi}\prod_{i=1}^{n-1}(i+\frac{n}{\xi})} \\
&\leq c_1(n+\xi)^{-(n-1)}(1+\xi)^{-a}(1+2\xi)^{-b} \\
&\leq c_2(1+\xi)^{-(n+a+b-1)} < \infty, \qquad (3.4)
\end{aligned}$$

if $n + a + b - 2 > 0$. Here c_1 and c_2 are a constants. This completes the proof. □

Under the prior (3.1), the marginal posterior density of ξ is given by

$$\pi(\xi|\mathbf{x}) \quad \propto \quad \int_0^\infty \sigma^{-n-1}(1+\xi)^{-a}(1+2\xi)^{-b}\prod_{i=1}^{n}\left(1+\frac{\xi}{\sigma}x_i\right)^{-\frac{1+\xi}{\xi}} d\sigma. \tag{3.5}$$

Note that actually, normalizing constant for the marginal density of ξ required one dimensional integration. Therefore we have the marginal posterior density of ξ, and so it is easy to compute the marginal moment of ξ. In Section 4, we investigate the frequentist coverage probabilities for the π_J and π_r respectively.

4. Numerical studies

We evaluate the frequentist coverage probability by investigating the credible interval of the marginal posterior density of ξ under the noninformative prior π given in (3.1) for several configurations ξ, σ and n. That is to say, the frequentist coverage of a $(1-\alpha)th$ posterior quantile should be close to $1-\alpha$. This is done numerically. Table 4.1 gives numerical values of the frequentist coverage probabilities of 0.05 (0.95) posterior quantiles for the our prior. The computation of these numerical values is based on the following algorithm for any fixed true (ξ, σ) and any prespecified probability value α. Here α is 0.05 (0.95). Let $\xi^{\pi}(\alpha|\mathbf{X})$ be the posterior α-quantile of ξ given $\mathbf{X}$. That is, $F(\xi^{\pi}(\alpha|\mathbf{X})|\mathbf{X}) = \alpha$, where $F(\cdot|\mathbf{X})$ is the marginal posterior distribution of ξ. Then the frequentist coverage probability of this one sided credible interval of ξ is

$$P_{(\xi,\sigma)}(\alpha; \xi) \quad = \quad P_{(\xi,\sigma)}(0 < \xi \leq \xi^{\pi}(\alpha|\mathbf{X})). \tag{4.1}$$

The computed $P_{(\xi,\sigma)}(\alpha;\xi)$ when $\alpha = 0.05(0.95)$ are shown in Table 4.1. In particular, for fixed n and (ξ, σ), we take $10,000$ independent random samples of $\mathbf{X} = (X_1, \cdots, X_n)$ from the generalized Pareto population.

In Table 4.1, we can observe that the reference prior π_r meets well the target coverage probabilities than Jeffreys' prior π_J. Also note that the results of table are not much sensitive to the change of the values of (ξ, σ). Thus we recommend to use the matching prior and the reference prior.

Table 4.1 Frequentist coverage probability of 0.05 (0.95) posterior quantiles of ξ

ξ	σ	n	π_J	π_r
0.5	0.5	10	0.028 (1.000)	0.043 (1.000)
		20	0.036 (0.995)	0.050 (0.999)
		30	0.038 (0.971)	0.049 (0.983)
		40	0.040 (0.952)	0.050 (0.968)
	1.0	10	0.031 (1.000)	0.046 (1.000)
		20	0.038 (0.996)	0.051 (0.999)
		30	0.041 (0.970)	0.050 (0.986)
		40	0.039 (0.952)	0.051 (0.969)
	3.0	10	0.032 (1.000)	0.046 (1.000)
		20	0.036 (0.995)	0.050 (0.999)
		30	0.037 (0.970)	0.048 (0.984)
		40	0.042 (0.955)	0.051 (0.971)
1.0	0.5	10	0.032 (0.977)	0.046 (0.997)
		20	0.036 (0.929)	0.047 (0.954)
		30	0.039 (0.927)	0.047 (0.948)
		40	0.041 (0.933)	0.051 (0.948)
	1.0	10	0.038 (0.977)	0.053 (0.997)
		20	0.038 (0.929)	0.050 (0.956)
		30	0.039 (0.927)	0.050 (0.946)
		40	0.040 (0.934)	0.049 (0.948)
	3.0	10	0.035 (0.979)	0.047 (0.998)
		20	0.039 (0.931)	0.050 (0.955)
		30	0.041 (0.927)	0.051 (0.946)
		40	0.040 (0.934)	0.050 (0.949)
3.0	0.	10	0.034 (0.912)	0.048 (0.943)
		20	0.039 (0.929)	0.050 (0.949)
		30	0.040 (0.931)	0.048 (0.947)
		40	0.042 (0.937)	0.050 (0.950)
	1.0	10	0.037 (0.912)	0.051 (0.943)
		20	0.036 (0.928)	0.048 (0.945)
		30	0.042 (0.935)	0.050 (0.947)
		40	0.041 (0.938)	0.049 (0.949)
	3.0	10	0.033 (0.908)	0.048 (0.941)
		20	0.042 (0.930)	0.051 (0.949)
		30	0.040 (0.939)	0.050 (0.952)
		40	0.043 (0.940)	0.051 (0.951)

Example 4.1 This example taken from Giles *et al.* (2011), and involves American Insurance Association data relating insurance losses in excess of 5,000,000 (in 1981 dollars) due to major hurricanes between 1949 and 1980. The data are provided by Hogg and Klugman (1983), and we have subtracted 5 million from each of their sample values, so allowing for the scale of the data reported by those authors, our first datum is 1766.0, etc.

The maximum likelihood estimator (MLE), the Bayes estimates based on Jeffreys' prior and the reference prior for the shape parameter are given by

$$\hat{\xi}_{MLE} = 0.86436,\ \hat{\xi}_J = 0.92161 \text{ and } \hat{\xi}_r = 0.96769,$$

respectively. These estimates for the shape parameter are the similar results, but the estimate based on the reference prior is slightly larger than other estimates.

5. Concluding remarks

In the generalized Pareto distribution, we have found a prior which is the first and the second order matching prior, and reference priors for the shape parameter. We revealed the second order matching prior does not exist. It turns out that the reference prior satisfies a first order matching criterion, but Jeffreys' prior is not a first order matching prior. As illustrated in our numerical study, the reference prior seems to be the best appropriate results than Jeffreys' prior in the sense of asymptotic frequentist coverage property.

References

Arnold, B. C. and Press, S. J. (1989). Bayesian estimation and prediction for Pareto data. *Journal of the American Statistical Association*, **84**, 1079-1084.

Balkema, A. A. and de Haan, L. (1974). Residual lifetime at great age. *The Annals of Probability*, **2**, 792-804.

Behrens, C., Lopes, H. F. and Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*, **4**, 227-244.

Berger, J. O. and Bernardo, J. M. (1989). Estimating a product of means : Bayesian analysis with reference priors. *Journal of the American Statistical Association*, **84**, 200-207.

Berger, J. O. and Bernardo, J. M. (1992). On the development of reference priors (with discussion). In *Bayesian Statistics IV*, edited by J. M. Bernardo et al., Oxford University Press, Oxford, 35-60.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *Journal of Royal Statistical Society B*, **41**, 113-147.

Castellanos, M. E. and Cabras, A. (2007). A default Bayesian procedures for the generalized Pareto distribution. *Journal of Statistical Planning and Inference*, **137**, 473-483.

Castillo, E. and Hadi, A. (1997). Fitting the generalized Pareto distribution to data. *Journal of the American Statistical Association*, **92**, 1609-1620.

Coles, S. G. and Powell, E. A. (1996). Bayesian methods in extreme value modelling: A review and new developments. *International Statistical Review*, **64**, 119-136.

Cox, D. R. and Reid, N. (1987). Parametric orthogonality and approximate conditional inference (with discussion). *Journal of Royal Statistical Society B*, **49**, 1-39.

Datta, G. S. and Ghosh, M. (1995). Some remarks on noninformative priors. *Journal of the American Statistical Association*, **90**, 1357-1363.

Datta, G. S. and Ghosh, M. (1996). On the invariance of noninformative priors. *The Annals of Statistics*, **24**, 141-159.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of Royal Statistical Society B*, **52**, 393-442.

de Zea Bermudez, P. and Amaral Turkman, M. A. (2003). Bayesian approach to parameter estimation of the generalized Pareto distribution. *Test*, **12**, 259-277.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling extremal events for insurance and finance*, Springer, Berlin.

Giles, D. E., Feng, H. and Godwin, R. T. (2011). *On the bias of the maximum likelihood estimators for the two-parameter Lomax distribution*, Econometrics Working Paper EWP1104, Department of Economics, University of Victoria, Canada.

Ghosh, J. K. and Mukerjee, R. (1992). Noninformative priors (with discussion). In *Bayesian Statistics IV*, edited by J. M. Bernardo et al., Oxford University, 195-210.

Ho, K. (2010). A matching prior for extreme quantile estimation of the generalized Pareto distribution. *Journal of Statistical Planning and Inference*, **140**, 1513-1518.

Hogg, R. V. and Klugman, S. A. (1983). On estimation of long-tailed skewed distributions with actuarial applications. *Journal of Econometrics*, **23**, 91-102.

Hosking, J. R. M. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, **29**, 339-349.

Kang, S. G. (2011). Noninformative priors for the common mean in log-normal distributions. *Journal of the Korean Data & Information Science Society*, **22**, 1241-1250.

Kim, D. H., Kang, S. G. and Lee, W. D. (2009). Noninformative priors for Pareto distribution. *Journal of the Korean Data & Information Science Society*, **20**, 1213-1223.

Mukerjee, R. and Dey, D. K. (1993). Frequentist validity of posterior quantiles in the presence of a nuisance parameter : Higher order asymptotics. *Biometrika*, **80**, 499-505.

Mukerjee, R. and Ghosh, M. (1997). Second order probability matching priors. *Biometrika*, **84**, 970-975.

Pickands, J. (1975). Statistical infernces using extreme order statistics. *The Annals of Statistics*, **3**, 119-131.

Smith, R. L. (1987). Estimating tails of probability distributions. *The Annals of Statistics*, **15**, 1174-1207.

Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution. *Sequential Methods in Statistics*, Banach Center Publications, **16**, 485-514.

Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika*, **76**, 604-608.

Welch, B. L. and Peers, H. W. (1963). On formulae for confidence points based on integrals of weighted likelihood. *Journal of Royal Statistical Society B*, **25**, 318-329.