International Journal of Statistics and Management System
Vol. 4 No. 2 (July-December, 2019)

Generalized inferences for the common scale parameter of several Pareto populations*

Sumith Gunasekera† and Malwane M. A. Ananda‡

Received: 13th August 2018 Revised: 24th December 2018 Accepted: 10th March 2019

Abstract

A problem of interest in this article is statistical inferences concerning the common scale parameter of several Pareto distributions. Using the generalized p-value approach, exact confidence intervals and the exact tests for testing the common scale parameter are given. Examples are given in order to illustrate our procedures. A limited simulation study is given to demonstrate the performance of the proposed procedures.

1 Introduction

In this paper, we consider k ($k \geq 2$) independent Pareto distributions with an unknown common scale parameter θ (sometimes referred to as the "location parameter" and also as the "truncation parameter") and unknown possibly unequal shape parameters α_i's ($i = 1, 2, ..., k$). Using the generalized variable approach (Tsui and Weerahandi [8]), we construct an exact test for testing θ. Furthermore, using the generalized confidence interval (Weerahandi [11]), we construct an exact confidence interval for θ as well. A limited simulation study was carried out to compare the performance of these generalized procedures with the approximate procedures based on the large sample method as well as with the other test procedures based on the combination of p-values.

Key words and phrases: Generalized p-value, generalized tests, Pareto distribution, common scale, parameter, size of the test.

†*Mailing Address:* Department of Mathematics, University of Tennessee-Chattanooga, Chattanooga, TN 37403 USA. *E–mail:* Sumith-Gunasekera@utc.edu

‡*Mailing Address:* Department of Mathematical Sciences, University of Nevada-Las Vegas, Las Vegas, NV 89154 USA *E–mail:* malwane.ananda@unlv.edu

In many statistical applications involving nuisance parameters, the conventional statistical methods do not provide exact solutions. As a result, even with small sample sizes practitioners often resort to asymptotic methods which are known to perform very poorly with the small samples. Tsui and Weerahandi [8] generalized the conventional definition of p-value so that the above mentioned problems can be easily resolved. Therefore, generalized p-value approach based on exact probability statements rather than on asymptotic approximations performs better than the classical p-value approach based on approximate procedures.

Generalized inferential methods have now been successfully applied to obtain exact tests in a variety of statistical models (for applications in *regressions*: Weerahandi [9] and many others and; for applications in *mixed models:* Weerahandi [10] and many others; for *one-way ANOVA*: Weerahandi [12]; for *two-way ANOVA:* Ananda and Weerahandi [2]; for *ANCOVA*: Ananda [1]).

According to a number of simulation studies, when compared, tests and confidence intervals obtained by using the generalized approach have been found to outperform the approximate procedures both in size and power. For a complete coverage and applications of these generalized tests and confidence intervals, the reader is referred to Weerahandi [13, 14].

2 Generalized Variable Approach. The two-parameter Pareto distribution with the shape parameter α and the scale parameter θ has a cumulative distribution function given by

$$F(x) = \begin{cases} 1 - \left(\frac{\theta}{x}\right)^{\alpha} & \text{if } x \geq \theta \\ 0 & \text{if } x < \theta, \end{cases} \tag{2.1}$$

where θ, $\alpha > 0$ and $x \in [\theta, \infty)$.

Suppose $\mathbf{X}^m = (X_1, X_2, \ldots, X_j, \ldots, X_m)$, $j = 1, 2, \ldots, m$ be a random sample of size m from (2.1). Quant [7] showed that the maximum likelihood estimators of θ and α, respectively, are

$$\widehat{\theta} = \min_{1 \leq j \leq m} X_j = X_{(1)} \qquad \text{and} \qquad \widehat{\alpha} = mY^{-1}, \tag{2.2}$$

where $Y = \sum_{j=1}^{m} \ln(X_j / X_{(1)})$, and Malik [6] derived the distributions of maximum likelihood estimators $\widehat{\theta}$ and $\widehat{\alpha}$ that are given by

$$\widehat{\theta} \sim Pa\,(m\alpha, \theta\,) \text{ and } \widehat{\alpha} \sim \Gamma^{-1}(m-1,\ m\alpha). \tag{2.3}$$

where $\Gamma^{-1}(c,d)$ is the inverse gamma distribution with shape parameter c and scale parameter d and $Pa\,(\alpha,\theta\,)$ is the Pareto distribution with shape parameter α and scale parameter θ.

Now, let $\{X_{ij}\}, i=1,2,\dots,k;\ j=1,2,\dots,m_i$ be independently distributed with X_{ij}'s *i.i.d.* with common p.d.f. for a given i^{th} Pareto population.

$$f(x_i)=\frac{\alpha_i\theta^{\alpha_i}}{x_i^{(\alpha_i+1)}}I_{[x_i\,\geq\,\theta]},\quad \theta,\alpha_i>0,\forall i,$$

where I denoting the usual indicator function, and θ denotes the common unknown scale parameter and α_i's are unknown and possibly unequal shape parameters. Elfessi and Jin [5] showed that

$$\widehat{\theta}=T=\min_{1\leq i\leq k}X_{i(1)}\qquad\text{and}\qquad\widehat{\alpha}_i=A_i=m_iY_i^{-1},\tag{2.4}$$

where $Y_i=\sum_{j=1}^{m_i}\ln(X_{ij}/X_{i(1)})$ for $i=1,2,...,k$. Furthermore, Elfessi and Jin [5] showed that

$$T\sim Pa\,(\alpha^*,\theta)\qquad\text{and}\qquad A_i\sim\Gamma^{-1}(m_i-1,\ m_i\alpha_i),\tag{2.5}$$

where $\alpha^*=\sum_{i=1}^{k}m_i\alpha_i$ and $i=1,2,\dots,k$.
Therefore,

$$2\alpha^*\ln(T/\theta)=V\sim\varkappa_2^2\qquad\text{and}\qquad 2m_i\alpha_iA_i^{-1}\ =W_i\sim\varkappa^2_{2m_i-2},\tag{2.6}$$

For a single Pareto distribution with common scale parameter θ and shape parameter α, Arnold [3] described the confidence intervals – for θ, when α is known; for α, when θ is known, and the joint confidence region for θ and α. Using certain classical independent tests that are based on the combination of probabilities: namely, the Tippet, the Fisher, the inverse normal, and the logit, Baklizi [4] constructed the confidence intervals for θ.

2.1 Statistical Testing of hypothesis for θ. Let us get started testing the hypothesis:

$$H_0:\,\theta\leq\theta_0\quad\text{Vs.}\quad H_a:\,\theta>\theta_0\,,\tag{2.7}$$

where θ_0 is a known quantity.

Suppose $\mathbf{X}_i^{m_i}=(X_{i1},X_{i2},\dots,X_{im_i})$ is a random sample of size m_i from a truncated Pareto populations $Pa\,(\alpha_i,\theta)$, $i=1,2,\dots,k$, where θ denotes the common unknown scale parameter and $\alpha_1,\alpha_2,\dots,\alpha_k$ are unknown and possibly unequal shape parameters. Furthermore, suppose $\mathbf{x}_i^{m_i}=(x_{i1},x_{i2},...,x_{im_i})$ is the observed sample of the i^{th} population.

Now, from (2.6), the generalized pivot for estimating θ and α_i are, respectively, given by

$$R_\theta = te^{-V/(\sum_{i=1}^k W_i a_i)} \quad \text{and} \quad R_{\alpha_i} = 0.5 W_i a_i / m_i, \tag{2.8}$$

where a_i is the observed value of A_i, or simply the estimate of α_i and t is the observed value of T, or simply the estimate of θ.

Now, consider the potential generalized test variable for testing

$$H_0^I : \theta \le \theta_0 \quad \text{Vs.} \quad H_a^I : \theta > \theta_0 \,,$$

where θ_0 is a known quantity, defined by

$$T_\theta = T(\mathbf{X}; \mathbf{x}, \zeta) = R_\theta - \theta = te^{-V/(2\sum_{i=1}^k W_i a_i)} - \theta \,, \tag{2.9}$$

where $\zeta = (\theta, \delta)$ is a vector of unknown parameters, θ being the parameter of interest and δ is a vector of nuisance parameters

The observed value of T_θ is $t_{\theta_{\text{obs}}} = T(\mathbf{x}; \mathbf{x}, \zeta) = 0$. It is clear that when θ is specified, T_θ has probability distribution that is free of nuisance parameters. Furthermore, when $\mathbf{x}$ and nuisance parameters are fixed, the cdf of T_θ is monotonically decreasing function of θ for any given t_θ. Therefore, T_θ is a generalized test variable (Weerahandi [12]) that can be used to test the given hypothesis. Thus, the generalized p-value, sometimes referred to as the generalized observed level of significance or generalized significance level, for testing $H_0^I : \theta \le \theta_0$ Vs. $H_a^I : \theta > \theta_0$ is given by

$$p\text{-value} = Pr(T_\theta < t_{\theta_{\text{obs}}} \mid \theta = \theta_0),$$

$$p\text{-value} = Pr(te^{-V/(\sum_{i=1}^k W_i a_i)} < \theta_0). \tag{2.10}$$

Similarly, generalized p-value for testing $H_0^{II} : \theta \ge \theta_0$ Vs. $H_a^{II} : \theta < \theta_0$ is given by

$$p\text{-value} = Pr(T_\theta > t_{\theta_{\text{obs}}} \mid \theta = \theta_0),$$

$$p\text{-value} = Pr(te^{-V/(\sum_{i=1}^k W_i a_i)} > \theta_0). \tag{2.11}$$

Then, the generalized p-value for testing $H_0^{III} : \theta = \theta_0$ Vs. $H_a^{III} : \theta \ne \theta_0$ is given by

$$p\text{-value} = 2 \min[Pr(T_\theta < t_{\theta_{\text{obs}}} \mid \theta = \theta_0),\ Pr(T_\theta > t_{\theta_{\text{obs}}} \mid \theta = \theta_0)],$$

$$p\text{-value} = 2\min[Pr(te^{-V/(\sum_{i=1}^{k} W_i a_i)} < \theta_0), Pr(te^{-V/(\sum_{i=1}^{k} W_i a_i)} < \theta_0)]. \quad (2.12)$$

Therefore, p-value can be computed by numerical integration with respect to V and W_i, which are independent random variables with known probability density functions. Furthermore, p-value can also be evaluated by the Monte Carlo method in which large numbers of random numbers from the chi-squared distributions with degree-of-freedom 2 as well as with the degree-of-freedom $(2m_i - 2)$ for $i = 1, 2, \cdots, k$ are generated and the fraction of random numbers pairs for which $R_\theta < \theta_0$ is determined. This is a Monte Carlo estimate of the generalized p-value for testing $\theta \le \theta_0$ Vs. $\theta > \theta_0$. Similarly, the Monte Carlo estimates of generalized p-value's for testing $\theta \ge \theta_0$ Vs. $\theta < \theta_0$ and $\theta = \theta_0$ Vs. $\theta \ne \theta_0$ can also be computed.

2.2 A $100(1-\gamma)\%$ Confidence Interval for θ. Since the value of R_θ is θ and the distribution of R_θ is independent of any unknown parameters, R_θ is a generalized pivotal quantity for θ. Therefore R_θ is the generalized pivotal quantity for constructing $100(1-\gamma)\%$ confidence interval for θ where γ is the confidence coefficient (Weerahandi 1993).

Computing Algorithm:

1. Compute a_i for $i = 1, 2, \ldots, k$ and t

2. Generate n (e.g. 75 000) random variates from each $W_i \sim \varkappa^2_{2m_i-2}$ and $V \sim \varkappa^2_2$ for $i = 1, 2, ..., k$

3. Compute R_θ.

4. Rank this array of R_θ's from smaller to larger.

The 100γ-th percentile of R_θ's, $R_\theta(\gamma)$, is the lower bound of the one-sided $100(1-\gamma)\%$ confidence interval, and $(R_\theta(\gamma/2),\ R_\theta(1-\gamma/2))$ is a two-sided $100(1-\gamma)\%$ confidence interval.

For actual coverage probabilities (empirical confidence level), repeat the above process for N (e.g. 50 000) times and calculate the fraction of times θ falls within the calculated (empirical) generalized confidence intervals.

3 Example.

Example 1: Comparison of proposed procedure with the classical approach based on large sample method.

This example deals with the Pareto distributions $X_i \sim Pa(\alpha_i, \theta)$ where $i = 1, 2, 3$

generated by the following population parameters: $\theta = 100$, $\alpha_1 = 0.5$, $\alpha_2 = 1.0$, $\alpha_3 = 1.5$ and sample sizes $m_1 = m_2 = m_3 = 10$. The data generated from these distributions are:

$X_1 \sim Pa\,(\alpha_1, \theta)$: $182.4447, 766.6342, 149.9515, 183.5521, 131.3459,$
$184.8249, 403.8077, 314.5954, 1264.0143, 116.9585$

$X_2 \sim Pa\,(\alpha_2, \theta)$: $815.0133, 113.2192, 216.6859, 266.3277, 255.2327,$
$354.8153, 640.5599, 417.5773, 109.8015, 167.6198$

$X_3 \sim Pa\,(\alpha_3, \theta)$: $102.8793, 142.2166, 101.4941, 104.4409, 247.1254,$
$316.8746, 213.758, 227.4824, 164.4707, 335.9244$

Assuming that all of the above parameters are unknown, the lower and upper bound of the one-sided 90% generalized empirical confidence interval for θ calculated from this data, respectively, are 98.06473 and 106.2982 while a two-sided 90% generalized empirical confidence interval for θ is given by (96.86647, 109.1483).

Furthermore, suppose we need to test $\theta \leq 100$ Vs. $\theta > 100$ using this data. Then the generalized p-value and the classical p-value, respectively, are 0.00252 and 0.11636. If we are to test $\theta \leq 95$ Vs. $\theta > 95$, the generalized p-value and the classical p-value, respectively, are given by 0.00144 and 0.10049 while generalized p-value and the classical p-value for testing $\theta \leq 98$ Vs. $\theta > 98$ are, respectively, given by 0.00192 and 0.27830.

Table 1: Actual type I error rates for testing H_0^I when nominal level $\gamma = 0.1$.

Parameters:	k	θ	$\alpha = (\alpha_1, \alpha_2, \alpha_3)$	Generalized	Classical
	$k = 3$	$\theta = 100$	$\alpha = (0.5, 1.0, 1.5)$	0.052	0.040
	$k = 3$	$\theta = 100$	$\alpha = (2.0, 2.5, 3.0)$	0.046	0.273
	$k = 3$	$\theta = 100$	$\alpha = (3.5, 4.0, 4.5)$	0.048	0.438
	$k = 3$	$\theta = 500$	$\alpha = (0.5, 1.0, 1.5)$	0.044	0.000
	$k = 3$	$\theta = 500$	$\alpha = (2.0, 2.5, 3.0)$	0.049	0.007
	$k = 3$	$\theta = 500$	$\alpha = (3.5, 4.0, 4.5)$	0.049	0.043

Table 1 shows the classical and generalized empirical (actual) type I error rates (size of the test) for testing $H_0^I : \theta \leq 100$ Vs. $H_a^I : \theta > 100$ and for testing $H_0^I : \theta \leq 500$ Vs. $H_a^I : \theta > 500$ when nominal (intended) type I error rate is at 0.1. All results are based on 50,000 replications.

Table 2 shows the empirical probability coverage for the generalized method and classical method when the intended nominal confidence level is at $\gamma = 0.1$. The simulation is based on 75 000 replications.

Table 2: Probability coverages for 90% two-sided confidence intervals

Parameters:	k	θ	$\sigma = (\sigma_1, \sigma_2, \sigma_3)$	Generalized	Classical
	$k = 3$	$\theta = 100$	$\alpha = (0.5, 1.0, 1.5)$	0.90	0.99
	$k = 3$	$\theta = 100$	$\alpha = (2.0, 2.5, 3.0)$	0.82	0.86
	$k = 3$	$\theta = 100$	$\alpha = (3.5, 4.0, 4.5)$	0.88	0.64
	$k = 3$	$\theta = 500$	$\alpha = (0.5, 1.0, 1.5)$	0.80	0.97
	$k = 3$	$\theta = 500$	$\alpha = (2.0, 2.5, 3.0)$	0.92	0.99
	$k = 3$	$\theta = 500$	$\alpha = (3.5, 4.0, 4.5)$	0.86	0.98

Overall, the coverage probability of the generalized confidence interval is more satisfactory compared to the classical method.

Example 2: **Comparison of proposed procedure with the classical approach based on the inverse normal method**.

Table 3 shows the comparison of the expected length of the generalized confidence interval for θ with the confidence lengths found in Baklizi [4] that are constructed using certain classical independent tests given by the Tippett's method, the Fisher's method, the inverse normal method and the logit method which are based on the combination of p-values. This is done on the basis of simulation. To keep the consistency with the specifications of parameters, sample sizes, and number of simulations found in Baklizi [4], we consider $k = 2$ and take $(m_1, m_2) = (10, 5)$, $(10, 10)$, $(10, 15)$, $\theta = 100$, $\alpha_1 = 1$, $\alpha_2 = 0.5$, 1 and $\gamma = 0.05, 0.1$. Note that, since the inverse normal method outperforms other methods in Baklizi [4], comparison is done between the confidence lengths based on inverse normal method and the proposed generalized method.

Table 3: Comparison of expected lengths of $100(1-\gamma)\%$ confidence intervals based on the inverse normal method and the generalized variable method.

γ	α_1	α_2	m_1	m_2	Generalized	Inverse normal
0.10	1	0.5	10	5	0.0398	0.2386
				10	0.0249	0.1861
		1.0		5	0.0338	0.1942
				10	0.0212	0.1372
0.05		0.5		5	0.0400	0.3101
				10	0.0251	0.2380
		1.0		5	0.0341	0.2474
				10	0.0216	0.1762

The inverse normal method:

Consider testing the hypothesis $H_0^i : \theta \geq \theta_0$ Vs. $H_a^i : \theta < \theta_0$ based on the i^{th} sample. The p-values of the individual tests are $P_i = \Pr(F_{2,2(m_i-1)} > f_i)$, where f_i is the observed value of F_i.Thus, $P_i = \left(1 + a_i \ln\left(\frac{t_i}{\theta}\right)\right)^{-m_i+1} \sim U(0,1)$, Here $F_i = \frac{V_i/2}{W_i/2m_i-2} = (m_i-1)A_i \ln(T_i/\theta) \sim F_{2,2(m_i-1)}$ is derived from each random variables W_i $(i=1,2)$ defined as $W_i = 2m_i\alpha_i A_i^{-1} \sim \varkappa^2_{2m_i-2}$ and V_i $(i=1,2)$ defined as $V_i = 2m_i\alpha_i \ln(T_i/\theta) \sim \varkappa^2_2$.

Numerical studies shows that generalized approach for hypothesis testing and confidence interval estimation of θ generally satisfactory and better than available some other methods for making inferences of θ.

References

[1] Ananda, M. M. A. (1998). Bayesian and non-Bayesian solutions to analysis of covariance models under heteroscedasticity. *Journal of Econometrics.* **86** 177–192.

[2] Ananda, M. M. A. and Weerahandi, S. (1997). Two-way Anova with unequal cell frequencies and unequal variances. *Statistica Sinica.* **7** 631–646.

[3] Arnold, B. C. (1983). *Pareto Distributions.* International Co-operative Publishing House, Maryland, USA.

[4] Baklizi, A. S. (2002). Confidence Lower Bounds for the Common Truncation Parameter of Several Pareto Distributions using Combined Independent Tests. *Far East Journal of Theoretical Statistics.* **8** 115–121.

[5] Elfessi, A. and Jin, C. (1996). On Robust Estimation of the Common Scale Parameter of Several Pareto Distributions. *Statistics and Probability Letters.* **29** 345–352.

[6] Malik, H. J. (1970). Estimation of the Parameters of the Pareto Distribution. *Metrika.* **15** 126–132.

[7] Quant, R. E. (1966). Old and new methods of estimation and the Pareto distribution. *Metrika.* **10** 55–82.

[8] Tsui, K. and Weerahandi, S. (1989). Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association.* **84** 602–607.

[9] Weerahandi, S. (1987). Testing regression equality with unequal variances. *Econometrica.* **55** 1211–1215.

[10] Weerahandi, S. (1991). Testing variance components in mixed models with generalized p-values. *Journal of the American Statistical Association.* **86** 151–153.

[11] Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association.* **88** 899–905. correction in **89**, 726.

[12] Weerahandi, S. (1995). ANOVA under unequal variances. *Biometrics.* **51** 589-599.

[13] Weerahandi, S. (1995). *Exact Stastical Methods for Data Analysis.* Springer–Verlag, New York.

[14] Weerahandi, S. (2004). *Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models.* John Wiley, New Jersey.