European Journal of Operational Research 221 (2012) 205–212

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Minimizing the area of a Pareto confidence region

Arturo J. Fernández *

Departamento de Estadística e Investigación Operativa, Universidad de La Laguna, 38271 La Laguna, Tenerife, Spain

ARTICLE INFO

Article history:
Received 31 January 2011
Accepted 4 March 2012
Available online 13 March 2012

Keywords:
Constrained optimization
Lagrangian method
Nonlinear programming
Pareto distribution
Right and double censoring
Trimmed data

ABSTRACT

A constrained optimization problem is formulated and solved in order to determine the smallest confidence region for the parameters of the Pareto distribution in a proposed family of sets. The objective function is the area of the region, whereas the constraints are related to the required confidence level. Explicit expressions for the area and confidence level of a given region are first deduced. An efficient procedure based on minimizing the corresponding Lagrangian function is then presented to solve the nonlinear programming problem. The process is valid when some of the smallest and largest observations have been discarded or censored, i.e., both single (right or left) and double censoring are allowed. The optimal Pareto confidence region is derived by simultaneously solving three (four) nonlinear equations in the right (double) censoring case. In most practical situations, Newton's method with the balanced set as the starting point only needs a few iterations to find the global solution. In general, the reduction in area of the optimal Pareto region with respect to the balanced set is considerable if the sample size, n, is small or moderately large, which is usual in practice. This reduction is sometimes impressive when n is quite small and the censoring degree is fairly high. Two numerical examples regarding component lifetimes and fire claims are included for illustrative and comparative purposes.

1. Introduction

An extensive collection of optimization techniques are widely used for solving quantitative problems in disciplines that include engineering, economics, biology, and business, among others. Optimization algorithms are also commonly employed to attain the best statistical methods. Papers by Fernández (2005, 2010a, 2011), Balamurali and Jun (2007), Arizono et al. (2008), Lu and Tsai (2009), Pérez-González and Fernández (2009), Zhu et al. (2010), Abellán et al. (2011), Cha and Finkelstein (2011), Fernández et al. (2002, 2011), Kim et al. (2011), Li and Glazebrook (2011), Selim and Al-Zu'bi (2011), Chun (2012), Fernández and Pérez-González (2012a, 2012b), Trindade and Ambrósio (2012) and Wu et al. (2012) are just a sample.

Essentially, the determination of the smallest confidence region in a given family of sets with a specified confidence level is a constrained optimization problem. A class of joint confidence regions for the two parameters of the Pareto distribution is first presented in this paper. The analyst should choose the smallest region in the class as often as possible for the best option. An optimization procedure is therefore described to select the minimum-area region in the proposed class of sets with a prefixed confidence level.

The Pareto distribution was introduced by the Italian-born Swiss economist, sociologist and philosopher Vilfredo Pareto to describe the distribution of personal income and wealth. This probability law provides a versatile population model which has a wide variety of applications in many practical fields, including service times in queuing systems, insurance risk studies, property values, stock price fluctuations, business failures, city population sizes, areas burnt in forest fires, migrations, sizes of firms, availability of natural resources, errors clustering in communications circuits and word frequencies.

The Pareto surname is also attached to other valuable concepts, such as Pareto efficiency or optimality, Pareto analysis, Pareto principle, and Pareto interpolation. At first, the Pareto principle (also known as the 80/20 rule) referred to the observation that 80% of wealth belonged to only 20% of the population. This principle indicates that, in general, the relationship between effects and causes is not balanced.

The Pareto principle, which is based on the Pareto distribution but is only slightly related to Pareto efficiency, has been found helpful in a wide range of areas such as manufacturing, management, decision-making, planning and human resources. In terms of preferences, a choice is defined as Pareto optimal if there is no alternative that everyone will regard as at least as good, and which at least one person will regard as better; see, e.g., the recent articles by Lindroth et al. (2010), Nikulin and Mäkelä (2010), Song and Kusiak (2010), He and Khouja (2011), Laumanns and Zenklusen (2011), Wu et al. (2011) and Zio and Bazzo (2011).

The probability density function and cumulative distribution function of a random variable X having a *Pareto*(α, τ) law with

* Tel.: +34 922 318179; fax: +34 922 318170.
E-mail address: ajfernan@ull.es

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2012.03.007

shape (or inequality) parameter $\alpha > 0$ and precision parameter $\tau > 0$ are given, respectively, by

$$f(x;\alpha,\tau) = \tau\alpha(\tau x)^{-\alpha-1} \quad \text{and} \quad F(x;\alpha,\tau) = 1-(\tau x)^{-\alpha}, \quad x \geqslant 1/\tau.$$

The construction of confidence regions for unknown parameters based on the available experimental data is of considerable interest and practical significance in many empirical studies. In particular, the determination of a joint confidence region for the Pareto parameters is often needed. Assuming a frequentist perspective, Chen (1996) presented a joint confidence region for (α, τ) based on complete or right censored samples. This method, however, cannot be extended to the double censoring situation. Adopting a Bayesian viewpoint, Fernández (2006a, 2008a) derived joint highest posterior density credibility regions for α and τ using trimmed samples and multiply censored data, respectively. Recently, Wu (2008) has proposed a frequentist confidence region for (α, τ) under double censoring, which improves Chen's method in terms of a smaller area in the complete and right censoring cases.

This paper presents a class of confidence regions for (α, τ) based on a doubly censored sample, which includes the region considered by Wu (2008) as the balanced case. Since, in choosing a confidence region, it is usually advantageous to minimize its size, a constrained optimization problem is formulated and solved in order to determine the smallest confidence region in the proposed class with the desired confidence level. Our approach is valid when certain proportions of the smallest and largest observations have been eliminated or censored. In many statistical studies, some extreme data may not be recorded due to restrictions on data collection, experimental difficulties or some other extraordinary reasons. Several extreme values are also discarded when the observations are poorly known or the presence of outliers is suspected; see Fernández (2004, 2006b, 2008b, 2009, 2010b, 2010c), and references therein.

The structure of the remainder of this article is as follows. Given a doubly censored sample from the *Pareto*(α, τ) distribution, the next section presents the likelihood function, the maximum likelihood estimators of the Pareto parameters and also several distributional results related to the minimal sufficient statistic for (α, τ). A family of joint confidence regions for α and τ is defined in Section 3. The confidence level of a region is also deduced in closed-form. Section 4 provides an explicit expression for the area of a Pareto confidence region. A nonlinear programming problem is formulated and solved in Section 5 to find the smallest Pareto region in the right and double censoring cases. The optimal solution is obtained by using the Lagrangian method. Section 6 offers a comparison of the areas of optimal and balanced confidence regions, whereas two illustrative examples are given in Section 7. The paper concludes with several remarks.

2. Pareto experimental data

Consider a random sample of size n from a *Pareto*(α, τ) population with unknown parameters α and τ, and let $X_r, \ldots, X_s$ be the ordered observations remaining when the $(r-1)$ smallest and $(n-s)$ largest sample values have been discarded or censored, where $1 \leqslant r \leqslant s \leqslant n$.

Given the observed data $\mathbf{X} = (X_r, \ldots, X_s)$, the likelihood function of (α, τ) presented in Fernández (2006a) can be written as

$$L(\alpha,\tau|\mathbf{X}) = \frac{n!\{1-(X_r\tau)^{-\alpha}\}^{r-1}\{U\tau^{n-r+1}\}^{-\alpha}}{(r-1)!(n-s)!\alpha^{r-s-1}\prod_{i=r}^{s}X_i}, \qquad \alpha > 0, \quad \tau \geqslant \frac{1}{X_r},$$

where $U = X_s^{n-s}\prod_{i=r}^{s}X_i$. In general, the sample evidence is contained in the minimal sufficient statistic (X_r, U).

Hereafter, it will be assumed that $r < s$ and $X_r < X_s$. In such a case, the corresponding maximum likelihood estimators of α and τ are readily found to be

$$\hat{\alpha} = \frac{s-r+1}{W} \quad \text{and} \quad \hat{\tau} = \frac{1}{X_r}\left(\frac{n}{n-r+1}\right)^{W/(s-r+1)},$$

where $W = \log(U) - (n-r+1)\log(X_r)$. Obviously, $\hat{\alpha}$ and $\hat{\tau}$ are jointly minimal sufficient for (α, τ). Similarly, (X_r, W) is also minimal sufficient for (α, τ). Moreover, X_r and W are independent because W can be expressed as

$$W = \sum_{i=r}^{s-1}(n-i+1)\log(X_{i+1}/X_i)$$

and $\log(X_r)$ and $\log(X_{i+1}/X_i)$, $i = r, \ldots, s-1$, are independent.

According to Fernández (2006a), the random variable $Y_1 = 2\alpha W$ follows a chi-square distribution with $2s-2r$ degrees of freedom, denoted by χ^2_{2s-2r}. Moreover, $(\tau X_r)^{-\alpha} \sim \mathcal{B}eta(n-r+1, r)$, which implies that $Y_2 = (n-r+1)\{(\tau X_r)^{\alpha}-1\}/r$ has a F-distribution with $2r$ and $2(n-r+1)$ degrees of freedom, designated by $\mathcal{F}_{2r,2(n-r+1)}$. Furthermore, as X_r and W are independent, it is clear that the pivotal quantities $Y_1 \sim \chi^2_{2s-2r}$ and $Y_2 \sim \mathcal{F}_{2r,2(n-r+1)}$ are also statistically independent.

3. A class of Pareto confidence regions

Confidence regions for (α, τ) can be constructed based on the pivotal vector (Y_1, Y_2) when $r < s$ and $X_r < X_s$ (i.e., when $W > 0$).

Assume that, for $0 \leqslant \beta < 1$, $\chi^2_{2s-2r;\beta}$ and $\mathcal{F}_{2r,2(n-r+1);\beta}$ denote the β-quantiles of the χ^2_{2s-2r} and $\mathcal{F}_{2r,2(n-r+1)}$ distributions, respectively, and $\chi^2_{2s-2r;1} = \mathcal{F}_{2r,2(n-r+1);1} = +\infty$. Then, as $Y_1 \sim \chi^2_{2s-2r}$ and $Y_2 \sim \mathcal{F}_{2r,2(n-r+1)}$, it turns out that

$$\Pr\left(\chi^2_{2s-2r;p_1} < Y_1 < \chi^2_{2s-2r;p_2}\right) = p_2 - p_1$$

and

$$\Pr(\mathcal{F}_{2r,2(n-r+1);q_1} < Y_2 < \mathcal{F}_{2r,2(n-r+1);q_2}) = q_2 - q_1$$

for $0 \leqslant p_1 < p_2 \leqslant 1$ and $0 \leqslant q_1 < q_2 \leqslant 1$. In view of the fact that Y_1 and Y_2 are independent, it follows immediately that

$$\Pr\left(\chi^2_{2s-2r;p_1} < Y_1 < \chi^2_{2s-2r;p_2}, \quad \mathcal{F}_{2r,2(n-r+1);q_1} < Y_2 < \mathcal{F}_{2r,2(n-r+1);q_2}\right) = \varepsilon,$$

where $\varepsilon = (p_2 - p_1)(q_2 - q_1)$. Consequently, the set

$$S(p_1,p_2,q_1,q_2) = \left\{(\alpha,\tau) : a_1 < \alpha < a_2, \frac{b_1^{-1/\alpha}}{X_r} < \tau < \frac{b_2^{-1/\alpha}}{X_r}\right\},$$

where

$$a_i = \frac{\chi^2_{2s-2r;p_i}}{2W}, \quad \text{and} \quad b_i = \left\{1+\frac{r\mathcal{F}_{2r,2(n-r+1);q_i}}{n-r+1}\right\}^{-1}, \quad i = 1,2, \qquad (1)$$

is a $100\varepsilon\%$ confidence region for (α, τ) because the confidence level $\Pr\{(\alpha, \tau) \in S(p_1, p_2, q_1, q_2)\} = \varepsilon$.

Clearly, given $\varepsilon \in (0, 1)$, the family of sets

$$C_\varepsilon = \{S(p_1,p_2,q_1,q_2) : (p_2-p_1)(q_2-q_1) = \varepsilon, \quad 0 \leqslant p_1 < p_2 \leqslant 1, \quad 0 \leqslant q_1 < q_2 \leqslant 1\}$$

constitutes a class of $100\varepsilon\%$ confidence regions for (α, τ). Alternatively, if $S(p_1, p_2, q_1, q_2)$ is now denoted by $R(a_1, a_2, b_1, b_2)$, the family C_ε could also be defined as

$$C_\varepsilon = \{R(a_1,a_2,b_1,b_2) : G(a_1,a_2)H(b_1,b_2) = \varepsilon, \quad 0 \leqslant a_1 < a_2 \leqslant +\infty, \quad 0 \leqslant b_2 < b_1 \leqslant 1\},$$

where

$$G(a_1,a_2) = \Pr(2Wa_1 < Y_1 < 2Wa_2)$$

and

$$H(b_1,b_2) = \Pr(b_2 < (\tau X_r)^{-\alpha} < b_1)$$