ISSN - 0103 - 2577

A Note on "Estimating the Fraction of Population in an Income Bracket Using Pareto Distribution"

ELISABETE CORRÊA LEME

Nº 127

NOTAS DO ICMSC

São Carlos (SP)
Out. 1992

A Note on "Estimating the fraction of population in an income bracket using Pareto distribution".

Elisabete Corrêa Leme
Departamento de Ciências de Computação e Estatística
Instituto de Ciências Matemáticas de São Carlos
Universidade de São Paulo
13.560, S. Carlos, SP., Brasil

Summary

Dealing with the problem of getting a minimum variance unbiased (m.v.u.) estimate for the survival function, $R(x)$, of a Pareto distribution, Shanmugam [1987] has adopted an approach based in transformation of random variables which he defends to be much simpler than the one considered by Lwin [1972]. In his work, section 3, case 1, where he presents the comparation between the precisions of maximum likelihood (m.l.) estimate, $\hat{R}(x)$, and his estimate $\tilde{R}(x)$, we have verified a slip in the variance of $\tilde{R}(x)$ and its estimate. Although this have not affect the quality of the results there exhibited, we believe that their adjustments, presented in this paper, would bring more simplicity on some expressions (like (3.5) and (3.7)) and more facility in management with posterior results.

Keywords: Pareto distribution, maximum likelihood-minimum variance unbiased estimates, survival function.

1 Introduction

Denoting by X the person's income in a population, Vilfredo Pareto [1897] showed that, if

$$R(x) = P(X \geq x) \tag{1.1}$$

then, at least for large values of x,

$$ln\ R(x) \cong \propto (lu\beta - lnx)$$

where β and α are parameters, β meaning the minimum income and α being representative of distribution shape and known as "constant of Pareto".

A random variable X is called a classic Pareto type if its probability density function is

$$f(x) = \frac{\alpha\beta^{\alpha}}{x^{\alpha+1}}\ , x \geq \beta > 0, \alpha > 0 \tag{1.2}$$

If $(X_1, X_2, \ldots, X_n)$ is a random sample of incomes from density (1.2) then it is an established result that the m.l. estimate for $R(x)$, when α is known, is given by

$$\hat{R}(x) = \left(\frac{S_2}{x}\right)^{\alpha} \tag{1.3}$$

where $S_2 = min(X_1, X_2, \ldots, X_n)$, and that the variance of $\hat{R}(x)$ is given by

$$Var\ \hat{R}(x) = \frac{n}{(n-2)(n-1)^2}\left(\frac{\beta}{x}\right)^{2\alpha} \tag{1.4}$$

2 About the Shanmugam estimate

Among the (m.v.u.) estimates which have been proposed for $R(x)$ in the especific case when α is the parameter known, the one established by Shanmugam through the Rao-Blackwell theorem and built on $S_2 = min(X_1, \ldots, X_n)$, is given by

$$\tilde{R}(x) = \left(1 - \frac{1}{n}\right)\left(\frac{S_2}{x}\right)^{\alpha} \tag{2.1}$$

In order to determine the variance of estimate $\tilde{R}(x)$, we take into account the expressions (2.1), (1.3) and (1.4). So, we obtain

$$\begin{aligned} Var\ \tilde{R}(x) &= \left(1-\tfrac{1}{n}\right)^2 Var\ \hat{R}(x) \\ &= \left(1-\tfrac{1}{n}\right)^2 \tfrac{n}{(n-2)\ (n-1)^2} \left(\tfrac{\beta}{x}\right)^{2\alpha} \end{aligned}$$

And hence

$$Var\ \tilde{R}(x) = \frac{1}{n(n-2)} \left(\frac{\beta}{x}\right)^{2\alpha}, n > 2 \tag{2.2}$$

Now, denoting by $u(s_2, n)$ the m.v.u. estimate of $Var\ \tilde{R}(x)$ and making use of the expectation inversion technique, the expression for $u(s_2, n)$ is obtained as follows.

$$\int_\beta^\infty u(s_2, n) g(s_2, n) ds_2 = \frac{1}{n(n-2)} \left(\frac{\beta}{x}\right)^{2\alpha} \tag{2.3}$$

That is,

$$\int_\beta^\infty u(s_2, n)\ \frac{n\alpha\beta^{n\alpha}}{s_2^{n\alpha+1}}\ \frac{n(n-2)ds_2}{\left(\frac{\beta}{x}\right)^{2\alpha}} = 1$$

Then,

$$\int_\beta^\infty u(s_2, n) n^2 \left(\frac{\beta}{s_2}\right)^{2\alpha} \left(\frac{x}{\beta}\right)^{2\alpha} \frac{(n-2)\alpha\beta^{(n-2)\alpha}}{s_2^{(n-2)\alpha+1}}\ ds_2 = 1$$

And then

$$\int_\beta^\infty u(s_2, n) n^2 \left(\frac{\beta}{s_2}\right)^{2\alpha} \left(\frac{x}{\beta}\right)^{2\alpha} g(s_2, n-2) ds_2 = 1$$

Hence

$$\int_\beta^\infty u(s_2, n) n^2 \left(\frac{x}{s_2}\right)^{2\alpha} g(s_2, n-2) ds_2 - 1 \equiv 0$$

And so

$$\int_{\beta}^{\infty}[u(s_2,n)n^2\left(\frac{x}{s_2}\right)^{2\alpha}-1]g(s_2,n-2)ds_2 \equiv 0$$

We notice that

$$u(s_2,n)n^2\left(\frac{x}{s_2}\right)^{2\alpha}-1\equiv 0$$

and hence

$$u(s_2,n)=\frac{1}{n^2}\left(\frac{s_2}{x}\right)^{2\alpha} \tag{2.4}$$

To obtain a $100(1-\delta)\%$ confidence interval for the fraction $R(x)$ based on $u(s_2,n)$, we note that,

$$\frac{\widetilde{R}(x)-R(x)}{\sqrt{u(s_2,n)}} = (n-1)-n\left(\frac{\beta}{s_2}\right)^{\alpha} \tag{2.5}$$

Then, the expression

$$P\left(l_1 \leq \frac{\widetilde{R}(x)-R(x)}{\sqrt{u(s_2,n)}} \leq l_2\right) = 1-\delta$$

where l_1 and l_2 are lower and upper limits, reduces to

$$P\left[(n-1)-l_2 \leq n\left(\frac{\beta}{s_2}\right)^{\alpha} \leq (n-1)-l_1\right] = 1-\delta \tag{2.6}$$

Or then

$$P\left[\left(\frac{n}{n-1-l_1}\right)^{1/\alpha} \leq \frac{s_2}{\beta} \leq \left(\frac{n}{n-1-l_2}\right)^{1/\alpha}\right] = 1-\delta$$

As $n\alpha > 0$, we still obtain

$$P\left\{n\alpha\left[\left(\tfrac{n}{n-1-l_1}\right)^{1/\alpha}-1\right] \leq n\alpha\left(\tfrac{s_2}{\beta}-1\right) \leq \right.$$
$$\left. \leq n\alpha\left[\left(\tfrac{n}{n-1-l_2}\right)^{1/\alpha}-1\right]\right\} = 1-\delta \tag{2.7}$$

Since $\left[n\alpha\left(\frac{s_2}{\beta}-1\right)\right]$ is distributed as F_{γ_1,γ_2} with $\gamma_1 = 2$ and $\gamma_2 = 2n\alpha$ degrees of freedom, then the expression (2.7) provides for the limits l_1 and l_2 the values

$$l_1 = (n-1) \quad - \quad \frac{n}{\left(\frac{F_{2,2n\alpha,1-\delta/2}}{n\alpha}+1\right)^{\alpha}}$$

$$l_2 = (n-1) \quad - \quad \frac{n}{\left(\frac{F_{2,2n\alpha,\delta/2}}{n\alpha}+1\right)^{\alpha}} \tag{2.8}$$

Rewriting (2.6) as

$$P\left[\left(\frac{n-1-l_2}{n}\right)\left(\frac{s_2}{x}\right)^{\alpha} \le \left(\frac{\beta}{x}\right)^{\alpha} \le \left(\frac{n-1-l_1}{n}\right)\left(\frac{s_2}{x}\right)^{\alpha}\right] = 1-\delta$$

and making use of (2.8), we obtain

$$P\left[\left\{\frac{s_2 n\alpha}{x\left(n\alpha+F_{2,2n\alpha,1-\delta/2}\right)}\right\}^{\alpha} \le R(x) \le \left\{\frac{s_2 n\alpha}{x\left(n\alpha+F_{2,2n\alpha,\delta/2}\right)}\right\}^{\alpha}\right] = 1-\delta$$

which is the expression (3.11) acquired by Shanmugam for a $100(1-\delta)\%$ confidence interval for the fraction $R(x)$.

References

[Pareto, V.] (1897) - Cours d'Economie Politique Lausanne and Paris: Rouge and Cie.

[Shanmugam, R.] (1987) - Estimating the fraction of population in an income bracket using Pareto distribution - Revista Brasileira de Probabilidade e Estatística, 1, 139-156

N O T A S D O I C M S C

Nº 126/92 - MORABITO, R.N.; ARENALES, M.N. - Staged and constrained two-dimensional cutting problems: a new approach

Nº 125/92 - RUAS, M.A.S. - On the equisingularity of families of corank 1 generic germs

Nº 124/92 - BIASI, C.; MENDES, M.D.C. - Generalization of Fibonacci sequence and application

Nº 123/92 - MOREIRA, E.S.; TRINDADE JR., O.; BATAGIN JR.; F. - Processamento paralelo atraves da distribuicao em ambientes SunOS

Nº 122/92 - MOCHIDA,D.K.H.; FUSTER, M.C.R.; RUAS, M.A.S. - Geometric characterization of the singularities of height functions on surfaces in 4 - spaces

Nº 121/92 - ACHCAR, J.A. - Some aspects of reparametrization for inferences of K-Out-of-M system reliability in the independent exponenetial case

Nº 120/92 - RODRIGUES, J. - Weighted balanced loss function for the exponential mean time to failure

Nº 119/92 - ACHCAR, J.A.; ROSALES, O.L.A. - Use of Bayesian methods in the analysis of two-factor experiments under an inverse Gaussian distribution

Nº 118/92 - NUNES, M.G.V. - Planning explanations from natural deduction based reasoning

Nº 117/92 - BIASI, C.; MOTTA, W.; OSAMU, S. - A remark on the separation by immersions in codimension-1