Estimation of the scale parameter in Burr distribution

Erika Hönschová

honschova@fmph.uniba.sk

Department of Applied Mathematics and Statistics, Comenius University, Bratislava

ABSTRACT

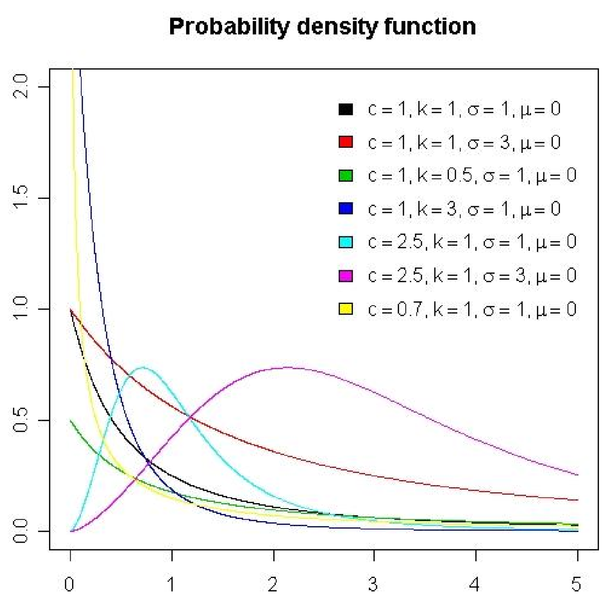
The poster presents an asymptotically normally distributed L-estimate of the scale parameter of the Burr distribution.

BURR DISTRIBUTION

Let X be a random variable with distribution function belonging to the location-scale family of the Burr distribution given by

$$F(x,\mu,\sigma,k,c) = 1 - \left(1 + \left(\frac{x-\mu}{\sigma}\right)^c\right)^{-k} \quad \text{for} \quad x \geq \mu,$$

where $k > 0, c > 0, \sigma > 0$, $\mu \in \mathcal{R}$. The Burr distribution has been applied in studies of household income, insurance risk, reliability analysis etc, e.g. by Tadikamalla [3], Embrechts and Schmidli [1], McDonald [2].



ESTIMATION OF THE SCALE PARAMETER

We will use L-estimates in the form:

$$L_n = \sum_{i=1}^{n} c_{ni} X_n^{(i)}, \quad c_{ni} = \int_{(i-1)/n}^{i/n} J(u)du,$$

where $X_n^{(1)} \leq X_n^{(2)} \ldots \leq X_n^{(n)}$ are the order statistics and $J(u)$ is a weights-generating function. Under various set of conditions imposed on the distribution function of the random sample and the weights-generating function, the asymptotic representation, given by the formula

$$\tilde{L}_n = \nu + \frac{1}{n}\sum_{i=1}^{n}\psi(X_i) + \mathcal{O}_P(\frac{1}{n}), \qquad (1)$$

where

$$\nu = \int_0^1 J(u)F^{-1}(u)\,du, \;\; \psi(x) = \int_{-\infty}^{+\infty} J(F(y))F(y)\,dy - \int_x^{+\infty} J(F(y))\,dy,$$

holds. Let $X_1, \ldots X_n$ be a random sample from the Burr distribution. The parameters μ, k, c are assumed to be known. To estimate the scale parameter define

$$\phi(x) = \frac{1}{\mathcal{I}(\sigma)} \frac{\partial \ln\left(f(x,\mu,\sigma,k,c)\right)}{\partial\sigma}$$

where $\mathcal{I}(\sigma)$ denotes the Fisher information. Put

$$J(u) = \phi'(F^{-1}(u)) = \frac{(k+1)(k+2)}{k}(1-u)^{\frac{2}{k}}((1-u)^{-\frac{1}{k}}-1)^{\frac{c-1}{c}}$$

and define the estimate by

$$\hat{\sigma}_n = \sum_{i=1}^{n} c_{ni}\left(X_n^{(i)} - \mu\right). \qquad (2)$$

For $c > \frac{1}{2}, k > 0$ the asymptotic representation (1) holds, with $\nu = \sigma$. Since $\int \psi(x)f(x)dx = 0$, by means of the central limit theorem it is easy to prove, that

$$\sqrt{n}(\hat{\sigma}_n - \sigma) \to N(0,V)$$

in distribution, where $V = \frac{2+k}{kc^2}\sigma^2$. The function $J(u)$ was computed in such a way, that $V = \frac{1}{\mathcal{I}(\sigma)}$, therefore the estimate $\hat{\sigma}_n$ is also **asymptotically efficient**.

If it is tedious to compute the score c_{ni} , we may use the approximation $\tilde{c}_{ni} = \frac{1}{n}J(\frac{i}{n+1})$. If μ is not known we may estimate it by $\hat{\mu} = X_n^{(1)}$ and define the scale estimate by $\hat{\hat{\sigma}}_n = \sum_{i=1}^{n} c_{ni}\left(X_n^{(i)} - \hat{\mu}\right)$. For $\frac{1}{2} < c < 2$ $\sqrt{n}\left(\hat{\hat{\sigma}}_n - \sigma\right) \to N(0, \frac{2+k}{kc^2}\sigma^2)$ in distribution.

COMPARISON WITH ESTIMATE DERIVED BY VÄNNMAN

The estimation of the location and scale parameter in the spetial case $c = 1$ (known as Pareto distribution) was studied in [4], [5]. In [5] the author derived the BLUE, based on order statistic, when μ, k are known but only for $k > 2$. If $\frac{2}{n} < k \leq 2$, the estimate based on first m order statistics is derived, but under the condition $m < n + 1 - 2/k$. The asymptotic distribution of the estimate is unknown. The estimat derived in [5] with $m = n - \left\lfloor\frac{2}{k}\right\rfloor$ is given by:

$$\tilde{\sigma}_m = \frac{1}{T_m}((k+1)\sum_{i=1}^{m-1} B_i X_n^{(i)} + ((n-m+1)\,k-1)\,B_k X_n^{(k)}$$

$$+ (T_m + 2 - nk)\,\mu),$$

where

$$T_m = \frac{nk - 2 - ((n-m)\,k-2)\,B_m}{k+2}$$

$$B_i = \left(1 - \frac{2}{k\,(n-i+1)}\right) B_{i-1}, \;\; B_0 = 1.$$

This estimate is unbiased, with variance $V(\tilde{\sigma}_m) = \frac{\sigma^2}{T_m}$. The estimat defined by (2) is for $c = 1$ in the form

$$\hat{\sigma}_n = \sum_{i=1}^{n} c_{ni}(X_n^{(i)} - \mu), \; c_{ni} = (k+1)\left[\left(\frac{n-i+1}{n}\right)^{\frac{2}{k}+1} - \left(\frac{n-i}{n}\right)^{\frac{2}{k}+1}\right].$$

In the following table are this two estimates compared on the basis of 5000 simulations of samples from the Pareto distribution with $\mu = 0$, $\sigma = 1$. The 25% and 75% sample quantiles are computed for each sample of estimates.

n=20				
k	3	1	0.5	0.1
$(Q1(\hat{\sigma}_n), Q3(\hat{\sigma}_n))$	(0.84, 1.25)	(0.82, 1.42)	(0.8, 1.64)	(0.86, 3.17)
$(Q1(\tilde{\sigma}_m), Q3(\tilde{\sigma}_m))$	(0.79, 1.17)	(0.7, 1.2)	(0.6, 1.22)	*

n=50				
k	3	1	0.5	0.1
$(Q1(\hat{\sigma}_n), Q3(\hat{\sigma}_n))$	(0.9, 1.14)	(0.88, 1.21)	(0.86, 1.31)	(0.8, 1.46)
$(Q1(\tilde{\sigma}_m), Q3(\tilde{\sigma}_m))$	(0.87, 1.12)	(0.82, 1.14)	(0.77, 1.17)	(0.5, 1.24)

* For $n = 20$, $k = 0.1$ is the estimate $\tilde{\sigma}_m$ not defined.

The estimate $\hat{\sigma}_n$ is useful mostly if k is small, and the sample size isn't large, because then the estimate $\tilde{\sigma}_m$ is not defined. We may also use it if $c \neq 1$, what allows to tune the frequency curve of the Burr distribution better to the observed data, because the observer may choose the values of both c and k to carry out the concerned data analysis.

References

[1] Embrechts P. and Schmidli H.(1994) *Modelling of extremal events in insurance and finance.* Mathematical Methods of Operations Research **39** 1–34.

[2] McDonald. J. B. (1984) *Some Generalized Functions for the Size Distribution of Income.* Econometrica **53**, 647–663.

[3] Tadikamalla, P. R. (1980) *A Look at the Burr and Related Distributions.* International Statistical Review **48**, 337–344

[4] Kulldorf, G. and Vännman K. (1973) *Estimation of the Location and Scale Parameters of a Pareto Distribution by Linear Functions of Order Statistics.* Journal of the American Statistical Association. **68**, 218–227.

[5] Vännman K. (1976) *Estimators Based on Order Statistics from a Pareto Distribution.* Journal of the American Statistical Association. **71**, 704–708.

Acknowledgement. The poster was supported by grant VEGA 1/3016/06.