Johnson descriptive statistics

Zdeněk Fabián

[1] Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 2, 18200 Prague 8

Keywords: robust variance, Johnson variance, ICORS 2005.

1 Introduction

There is a large amount of regular distributions, the mean and/or variance of which do not exist. A question is, what measures of central tendency and dispersion of such distributions can them replace.

As a measure of central tendency of unimodal distributions with support $\mathcal{S} = \mathcal{R}$ was proposed in [1] a mode. To find a measure of central tendency of distributions with support $\mathcal{S} \neq \mathcal{R}$, which need not be unimodal themselves, it was proposed in the same paper the following procedure:

Let $\eta : \mathcal{R} \to \mathcal{S}$ be a strictly monotone and continuously differentiable mapping. By [1], any F with support $\mathcal{S} \neq \mathcal{R}$ defines a distribution function $G_\eta = F\eta^{-1}$, an $\eta-$prototype of F. If the mode $y^*(G_\eta)$ exists, the point

$$x_\eta = \eta^{-1}(y^*(G_\eta)) \tag{1}$$

can be taken as a measure of central tendency of F. The existence and location of the $\eta-$point (1) of distribution F depend on the mappings η. To avoid this ambiguity we restricted ourselves [2] to the Johnson's mappings [3],

$$\eta(x) = \begin{cases} \log(x-a) & \text{if } \mathcal{S} = (a, \infty) \\ \log \dfrac{(x-a)}{(b-x)} & \text{if } \mathcal{S} = (a, b) \\ \log(b-x) & \text{if } \mathcal{S} = (-\infty, b) \end{cases} \tag{2}$$

2 Johnson location

In the first part of the submitted paper we show that the parametric space belonging to $\mathcal{S} \neq \mathcal{R}$ can be reparametrized in such a way that the "Johnson's $\eta-$point" of a parametric distribution with support $\mathcal{S}$ becomes a parameter. Let us call it Johnson location, although it is not a location in the usual sense. Some of currently used parametric distribution on $\mathcal{R} \neq \mathcal{S}$ has this parameter as their actual parameter, others not. In the latter cases, the Johnson point of the distribution is to be found and expressed by means of the actual parameters.

3 Johnson variance

We define a Johnson variance σ_η of distribution F as the reciprocal value of the Fisher information about the Johnson location and show that the couple (x_η, σ_η) characterize central tendency and spread of distribution F and that its sample version, $(\hat{x}_{\eta,n}, \hat{\sigma}_{\eta,n})$, satisfactorily describes the data taken as independent realizations drawn from F.

4 Johnson statistics

It appeared that, on the basis of the behavior of the Johnson variance, statistical distributions can be divided into three groups.

(i) Distributions with σ_η identical with or proportional to the variance.

(ii) Distributions, variance of which do not exist or exist only within a limited range of parameters (heavy tailed families). The Johnson locations of heavy-tailed distributions exist and a couple $(\hat{x}_{\eta,n}, \hat{\sigma}_{\eta,n})$ appears to be a robust characteristic of the data even if the estimates were obtained by the maximum likelihood method.

(iii) Distributions, variances of which exist and differs from the Johnson variances. The sample variance gives a better description of the variability of the data if the value of the scale parameter of the distribution is small, but if it is large, a better alternative is a robust version of the sample Johnson variance.

5 Literature

References

Z. Fabián (2001). Induced cores and their use in robust parametric estimation. *Commun. in Statist., Theory Meth.*, 30, 537–556.

Z. Fabián and I. Vajda (2003). Core functions and core divergences of regular distributions. *Kybernetika,* 39, 1, 29–42.

N.L. Johnson (1949). Systems of frequency curves generated by methods of translations. *Biometrika,* 36, 149-176.