

Influence Function Approach to Sensitiveness of Kurtosis Indexes

Approccio alla Sensibilità degli Indici di Curtosi
Basato sulla Funzione di Influenza

Francesco Campobasso

Dip. Scienze Statistiche Carlo Cecchi, Università di Bari, fracampo@dss.uniba.it

Riassunto. Obiettivo di questo lavoro è confrontare le risposte di due diversi indici di disnormalità, quello di Pearson e quello attribuito a Gini, rispetto alle frequenze di ciascuna modalità. I risultati, ottenuti analizzando sia caratteri continui sia caratteri discreti, consentono di mettere *meglio* in evidenza il diverso comportamento degli indici.

Keywords: Kurtosis Indexes, Influence Function.

1. Introduction

Fiori and Zenga (2005) analysed the behaviour of the kurtosis index introduced by Pearson, starting from a Faleschini's work (1948) and arriving to identify its answer with respect to the frequencies of each modality. The same result was obtained by using the influence function suggested by Hampel (1974) and by Ruppert (1987). The aim of this paper is to consider the index attributed to Gini and compare it with Pearson's one.

2. The sensitiveness of kurtosis index attributed to Gini

The kurtosis index attributed to Gini is given by $G = \sigma^2 / \delta^2$, where σ is the standard deviation and δ is the mean deviation of the distribution. It is well known that for the normal distribution $G = \pi/2$ and that values of G smaller (larger) than $\pi/2$ indicate platikurtosis (leptokurtosis).

As the partial derivatives of σ^2 and δ with respect to the frequency n_r ($r=1,2,\dots,s$) are $\partial\sigma^2/\partial n_r = [(x_r - \mu)^2 - \sigma^2]/N$ and $\partial\delta/\partial n_r = [|x_r - \mu| - \delta]/N$, it follows that the partial derivative of the index G with respect to the frequency n_r is $\partial G/\partial n_r = \{ [(x_r - \mu)^2 - \sigma^2]\delta^2 - 2\delta[|x_r - \mu| - \delta]\sigma^2 \} / N\delta^4$. By substituting $z_r = (x_r - \mu)/\sigma$, setting the derivative equal to zero and dividing for G , we obtain $z_r^2 - 2\sqrt{G}|z_r| + 1 = 0$, whose roots are $z_r = \pm\sqrt{G} \pm \sqrt{G-1}$. In the case of the standard normal distribution, as $G = \pi/2$, the range of the derivative is divided in the following sections: $(-\infty, -2,009)$, $(-2,009, -0,498)$, $(-0,498, 0,498)$, $(0,498, 2,009)$, $(2,009, \infty)$.

3. The influence function in the continuous case

Let's consider – only for simplicity – a variable T having a cumulative distribution function $F(t)$, such that $E(T)=0$ and $\text{var}(T) = \sigma^2$. Let's imagine we contaminate the variable

T at the point x through a quantity ε ($0 < \varepsilon < 1$). The cumulative distribution function of the contaminated variable is $F_\varepsilon(t) = (1 - \varepsilon)F(t) + \varepsilon H(t)$, where $H(t)$ equals 0 for $t < x$ and 1 for $t \geq x$.

This device enables us to obtain the influence function of G : $IF(G) = \lim_{\varepsilon \rightarrow 0} [\sigma_\varepsilon^2 / \delta_\varepsilon^2 - \sigma^2 / \delta^2] / \varepsilon$, where σ_ε^2 and δ_ε are the variance and the mean deviation of the contaminated variable respectively.

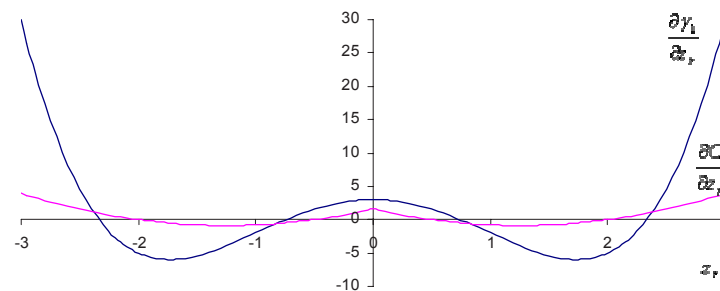
Fiori and Zenga (2005) proved that $\sigma_\varepsilon^2 = \sigma^2 + (x^2 - \sigma^2)\varepsilon - x^2\varepsilon^2$; by performing some computations we prove that $\delta_\varepsilon = 2(1 - \varepsilon) \int_{-\infty}^x F(t)dt + \varepsilon(1 - \varepsilon)(|x| - x)$.

Confining ourselves to the case of symmetrical variables and considering the standardized value $z = x / \sigma$, setting the influence function equal to zero and dividing for σ^2 , we obtain the same equation as the one obtained by Faleschini's procedure.

4. Final remarks

Fig.1 shows the behaving of the derivatives of G and γ_1 with respect to z_r in the case of the standard normal distribution. The comparison of the two derivatives shows that G is generally less sensitive to contaminations than γ_1 . Indeed, when attention is focused on the middle part of the distribution, the areas where a contamination determines an increase of kurtosis are narrower for the G index; on the other hand, when attention is focused on the tails of the distribution, the areas where a contamination determines an increase of kurtosis are larger for the G index. The conjoint use of the two indexes enables a better evaluation of the nonnormality.

Figure 1: Derivatives of G and γ_1 for the standard normal distribution.



References

- Faleschini L. (1948) Su alcune proprietà dei momenti impiegati nello studio della variabilità, asimmetria e curtosi, *Statistica*, 8, 503-513.
- Fiori A.M., Zenga M. (2005) L'interpretazione della curtosi e la sua curva di influenza in un'intuizione di L. Faleschini, *Statistica*, anno LXV, 2005, 2, 135-144.
- Hampel F.R. (1974) The influence curve and its role in robust estimation, *Journal of the American Statistical Association*, 69, 383-393.
- Ruppert D. (1987) What is kurtosis? An influence function approach, in *The American Statistician*, 41, 1, 1-5.