STATISTICA, anno LXV, n. 2, 2005

THE MEANING OF KURTOSIS, THE INFLUENCE FUNCTION AND AN EARLY INTUITION BY L. FALESCHINI

A.M. Fiori, M. Zenga

1. INTRODUCTION

The oldest and most common measure of kurtosis is the standard fourth moment coefficient:

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = E\left(\frac{X-\mu}{\sigma}\right)^4 \tag{1}$$

which was introduced by K. Pearson (Pearson, 1905).

There has been much discussion over whether β_2 measures peakedness, tail weight or the "shoulders" of a distribution. A convenient way of grasping the meaning of β_2 is to ask how β_2 changes as observations are added to an existing distribution. To our knowledge, the author who first answered this question was Luigi Faleschini in 1948.

In this work we rediscover Faleschini's methodology and show how it can be used to assess the relative importance of tail heaviness and peakedness in determining β_2 (section 2). The implications of Faleschini's approach for a normal distribution are detailed in section 3. In section 4 we establish an unexpected connection between a partial derivative computed by Faleschini and the influence function of β_2, which appeared much later in statistical literature (Hampel, 1968 and 1974; Ruppert, 1987). Our conclusions are outlined in section 5, where a short parallel is drawn between Faleschini's methodology and Zenga's new approach to kurtosis (Zenga, 1996 and 2006).

2. FALESCHINI'S DERIVATIVE AND β_2

Faleschini (1948) proposed the following methodology. Consider a statistical variable X consisting of n real values, $x_1, x_2, ..., x_n$, sorted in increasing order ($x_1 <$

This paper reflects the common thinking of the authors, although, more specifically, sections 1, 3 and 5 are due to M. Zenga, sections 2 and 4 to A. M. Fiori.

$x_2 < ... < x_n$). Denote the corresponding frequencies by $f_1, f_2, ..., f_n$ and the total size of the distribution by $P = \sum_{r=1}^{n} f_r$. To observe how β_2 is changed by observations added at some particular point, say x_r, compute the partial derivative of β_2 with respect to the altered frequency f_r. According to the sign of $\partial\beta_2/\partial f_r$ it is easy to identify the ranges where adding new observations raises (lowers) β_2. As a consequence, the centre, flanks and tails of a distribution may be quantitatively specified and their relative importance in determining β_2 may be gauged from the absolute value of $\partial\beta_2/\partial f_r$.

Next is the computation of Faleschini's derivative. Referring to the binomial theorem:

$$(a-b)^k = \sum_{i=0}^{k} (-1)^i \binom{k}{i} a^{k-i} b^i \tag{2}$$

Faleschini used the following representation of a central moment of order s ($s = 1, 2, 3, 4, ...$):

$$\begin{aligned} \mu_s &= \frac{1}{P} \sum_{r=1}^{n} (x_r - \mu)^s \cdot f_r \\ &= \sum_{i=0}^{s} (-1)^i \binom{s}{i} m_{s-i} \cdot \mu^i \end{aligned} \tag{3}$$

where:

$$m_{s-i} = \frac{1}{P} \sum_{r=1}^{n} x_r^{\,s-i} \cdot f_r$$

is the moment of order $(s - i)$ from the origin and μ stands for the arithmetic mean ($\mu \equiv m_1$). Elementary rules of differentiation lead to partial derivatives of m_{s-i} and μ^i with respect to f_r:

$$\begin{aligned} \frac{\partial m_{s-i}}{\partial f_r} &= \frac{1}{\left(\sum_{r=1}^{n} f_r\right)^2} \left(x_r^{s-i} \cdot \sum_{r=1}^{n} f_r - 1 \cdot \sum_{r=1}^{n} x_r^{s-i} \cdot f_r \right) \\ &= \frac{1}{P}(x_r^{s-i} - m_{s-i}) \end{aligned} \tag{4}$$

$$\begin{aligned} \frac{\partial \mu^i}{\partial f_r} &= i \cdot \mu^{i-1} \cdot \frac{1}{\left(\sum_{r=1}^{n} f_r\right)^2} \left(x_r \cdot \sum_{r=1}^{n} f_r - 1 \cdot \sum_{r=1}^{n} x_r \cdot f_r \right) \\ &= \frac{1}{P} i \cdot \mu^{i-1} (x_r - \mu). \end{aligned} \tag{5}$$

Combining (4) and (5) according to (3), we obtain the partial derivative of μ_s with respect to f_r :

$$\begin{aligned}\frac{\partial \mu_s}{\partial f_r} &= \sum_{i=0}^{s} (-1)^i \binom{s}{i} \frac{\partial (m_{s-i} \cdot \mu^i)}{\partial f_r} \\ &= \sum_{i=0}^{s} (-1)^i \binom{s}{i} \left\{ \frac{1}{P} (x_r^{s-i} - m_{s-i}) \cdot \mu^i + \frac{1}{P} i \cdot \mu^{i-1} (x_r - \mu) \cdot m_{s-i} \right\} \\ &= \frac{1}{P} \left\{ \sum_{i=0}^{s} (-1)^i \binom{s}{i} x_r^{s-i} \cdot \mu^i - \sum_{i=0}^{s} (-1)^i \binom{s}{i} m_{s-i} \cdot \mu^i \right. \\ &\quad \left. + (x_r - \mu) \cdot (-1) \cdot s \sum_{i=1}^{s} (-1)^{i-1} \frac{(s-1)!}{i(i-1)!(s-i)!} i \cdot \mu^{i-1} m_{s-i} \right\}.\end{aligned} \tag{6}$$

With the substitution $j = (i-1)$ the last summation is equal to:

$$\sum_{j=0}^{s-1} (-1)^j \binom{s-1}{j} \mu^j \, m_{(s-1)-j}$$

and (6) reduces, using (2) and (3), to the simple expression:

$$\frac{\partial \mu_s}{\partial f_r} = \frac{1}{P} \{ (x_r - \mu)^s - \mu_s - (x_r - \mu) \cdot s \cdot \mu_{s-1} \} \tag{7}$$

the leading term being the s-th power of the distance $(x_r - \mu)$.

Since β_2 is a ratio of central moments, a standard differentiation argument implies that:

$$\begin{aligned}\frac{\partial \beta_2}{\partial f_r} &= \frac{\partial \left(\frac{\mu_4}{\mu_2^2} \right)}{\partial f_r} = \frac{1}{\mu_2^4} \left(\frac{\partial \mu_4}{\partial f_r} \cdot \mu_2^2 - 2 \mu_2 \mu_4 \cdot \frac{\partial \mu_2}{\partial f_r} \right) \\ &= \frac{1}{P} \cdot \frac{1}{\mu_2^4} \{ [(x_r - \mu)^4 - \mu_4 - 4 \mu_3 (x_r - \mu)] \cdot \mu_2^2 - 2 \mu_2 \mu_4 [(x_r - \mu)^2 - \mu_2] \}.\end{aligned}$$

If the standardized quantity $(x_r - \mu)/\sqrt{\mu_2}$ is denoted by z_r and the skewness coefficient $\mu_3 / \mu_2^{3/2}$ by γ_1, the partial derivative of β_2 with respect to f_r may be rewritten as:

$$\frac{\partial \beta_2}{\partial f_r} = \frac{1}{P}(z_r^4 - \beta_2 - 4\gamma_1 z_r - 2\beta_2 z_r^2 + 2\beta_2)$$
$$= \frac{1}{P}[(z_r^2 - \beta_2)^2 - \beta_2(\beta_2 - 1) - 4\gamma_1 z_r]. \tag{8}$$

In the symmetric case ($\gamma_1 = 0$) closed-form solutions are available for the equation $\partial\beta_2/\partial f_r = 0$. Writing:

$$(z_r^2 - \beta_2)^2 = \beta_2(\beta_2 - 1)$$

we have:

$$z_r^2 = \beta_2 \pm \sqrt{\beta_2(\beta_2 - 1)}.$$

Hence:

$$z_r = \pm\sqrt{\beta_2 \pm \sqrt{\beta_2(\beta_2 - 1)}}$$

which is equivalent, for $\sigma = \sqrt{\mu_2}$, to the four solutions:

$$x_r^{(1,2,3,4)} = \mu \pm \sigma\sqrt{\beta_2 \pm \sqrt{\beta_2(\beta_2 - 1)}}.$$

Owing to the inequality: $\beta_2 \geq 1$ (see Stuart and Ord, 1994) the above roots are always real and break up the range of the statistical variable X into five regions:

Region I: $x_1 \leq X \leq x_r^{(1)}$: $\frac{\partial \beta_2}{\partial f_r}$ *is positive*

Region II: $x_r^{(1)} < X < x_r^{(2)}$: $\frac{\partial \beta_2}{\partial f_r}$ *is negative*

Region III: $x_r^{(2)} \leq X \leq x_r^{(3)}$: $\frac{\partial \beta_2}{\partial f_r}$ *is positive*

Region IV: $x_r^{(3)} < X < x_r^{(4)}$: $\frac{\partial \beta_2}{\partial f_r}$ *is negative*

Region V: $x_r^{(4)} \leq X \leq x_n$: $\frac{\partial \beta_2}{\partial f_r}$ *is positive*.

New observations lower β_2 if added within regions II and IV, raise it if added within regions I, III and V. Regions II and IV can therefore be identified with the *flanks*, region III with the *centre*, regions I and V with the *tails*. Since $\partial\beta_2/\partial f_r$ is of order z_r^4, it grows very rapidly as x_r decreases below $x_r^{(1)}$ or increases beyond

$x_r^{(4)}$. Hence it is clear that β_2 is primarily a measure of tail behavior and only to a lesser extent of peakedness.

3. THE CASE OF THE NORMAL DISTRIBUTION

For any symmetric distribution with $\beta_2 = 3$ (including the normal as a special case), (8) reduces to:

$$\frac{\partial \beta_2}{\partial f_r} = \frac{1}{P}[(z_r^2 - 3)^2 - 6] \tag{9}$$

with roots:

$$x_r^{(1,4)} = \mu \pm 2,334\,\sigma$$
$$x_r^{(2,3)} = \mu \pm 0,742\,\sigma.$$

In the Gaussian model (Figure 1 - bottom) the centre ($|x_r - \mu| \leq 0,742\ \sigma$) accounts for nearly 55% of the data, the tails ($|x_r - \mu| \geq 2,334\ \sigma$) for less than 2%. The steepness of the function $\partial\beta_2/\partial f_r$ (Figure 1 - top) indicates however that tail outliers drastically affect the value of β_2.

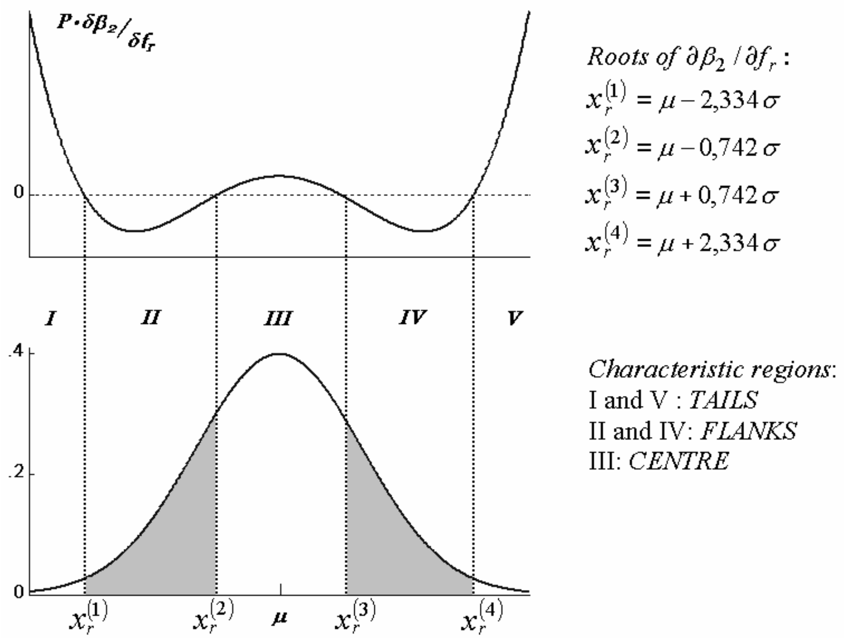

Figure 1 – The behavior of Faleschini's derivative for $\beta_2 = 3$ (top), with indication of the ranges corresponding to the centre, flanks and tails (bottom). Both charts are drawn for standard parameter values ($\mu = 0$, $\sigma = 1$), the lower represents a normal density function.

Relating (9) to a normal distribution, Faleschini argued that "a symmetric curve with $\beta_2 > 3$ will be *iperbinomial* (i.e. *leptokurtic*) because some values at the centre and in the tails will have higher frequency than in the normal curve with the same mean and variance, and some values in the flanks will be underweighted compared to the corresponding normal curve. The situation is reversed for distributions with $\beta_2 < 3$" (Faleschini, 1948). Since $\beta_2 = 3$ is not sufficient to ensure the normal like behavior of a symmetric distribution (see Balanda and MacGillivray, 1988, or Kale and Sebastian, 1996) the above argument should be used with caution.

4. β_2 AND THE INFLUENCE FUNCTION APPROACH

Faleschini's derivative was nearly overlooked by statistical literature and a number of erroneous interpretations of β_2 appeared in the following decades (see Balanda and MacGillivray, 1988 or Zenga, 2006 for a comprehensive review). It was only in 1987 that Ruppert succeeded in "*quantitatively analyzing the meaning of kurtosis*" by means of Hampel's influence function (IF).

The influence function approach (Hampel, 1968 and 1974) focuses on statistical measures which may be expressed as (real-valued) functionals, i.e. mappings from a set of probability distributions to the real numbers. For a distribution F with finite fourth moment, the kurtosis coefficient (1) is the functional defined by:

$$\beta_2(F) = \frac{\mu_4(F)}{\mu_2^2(F)}$$

where:

$$\mu_s(F) = \int_{-\infty}^{\infty} [t - \mu(F)]^s dF(t) \qquad \text{for } s = 2, 4$$

and:

$$\mu(F) = \int_{-\infty}^{\infty} t \, dF(t) .$$

The IF of β_2 is computed as follows. Consider the cumulative distribution function:

$$G_x(t) = \begin{cases} 0 & \text{for } t < x \\ 1 & \text{for } t \geq x \end{cases} \tag{10}$$

and define the mixture:

$$F_{\varepsilon}(t) = (1-\varepsilon)F(t) + \varepsilon G_x(t) \qquad 0 \leq \varepsilon \leq 1. \tag{11}$$

"F_ε is nothing but F with contaminants at x and ε is the proportion of contamination" (Ruppert, 1987). Then,

$$\lim_{\varepsilon \downarrow 0} \frac{\beta_2(F_\varepsilon) - \beta_2(F)}{\varepsilon} \tag{12}$$

is the (one-sided) derivative of β_2 at F in the direction toward G_x and is called the influence function of β_2 at F and x [IF (x; F, β_2)]. Formula (12) has a heuristic interpretation as the change of β_2 caused by an infinitesimal contamination at the point x, standardized by the size of the contaminating mass. This is logically related to Faleschini's derivative and it is shown below that computing the IF of β_2 from its own definition (12) does indeed result in the same expression as $\partial\beta_2/\partial f_r$.

If $\mu(F_\varepsilon)$ denotes the mean of the contaminated distribution, it follows from (11) that:

$$\mu(F_\varepsilon) = (1-\varepsilon)\mu(F) + \varepsilon x \tag{13}$$

and the same argument generalizes to the central moments:

$$\mu_s(F_\varepsilon) = (1-\varepsilon)\int_{-\infty}^{\infty}(t-\mu(F_\varepsilon))^s\, dF(t) + \varepsilon(x-\mu(F_\varepsilon))^s \qquad \text{for integer } s.$$

Replacing $\mu(F_\varepsilon)$ by (13) we have:

$$\begin{aligned}\mu_s(F_\varepsilon) = (1-\varepsilon)&\int_{-\infty}^{\infty}\{[t-\mu(F)] - \varepsilon[x-\mu(F)]\}^s\, dF(t)\\ &+ \varepsilon[(1-\varepsilon)(x-\mu(F))]^s\end{aligned} \tag{14}$$

where the integrand may be expanded binomially by (2).

Application of (14) for $s = 2$ yields the variance of the contaminated distribution:

$$\begin{aligned}\mu_2(F_\varepsilon) &= (1-\varepsilon)\{\mu_2(F) + \varepsilon^2(x-\mu(F))^2\} + \varepsilon(1-\varepsilon)^2(x-\mu(F))^2\\ &= \mu_2(F) + \varepsilon\{(x-\mu(F))^2 - \mu_2(F)\} - \varepsilon^2(x-\mu(F))^2\end{aligned} \tag{15}$$

and for $s = 4$, after some computation and rearrangement, the fourth central moment obtains:

$$\begin{aligned}\mu_4(F_\varepsilon) = \mu_4(F) + \varepsilon\,\{(x-\mu(F))^4 - \mu_4(F) - 4\mu_3(F)(x-\mu(F))\} \\ -\varepsilon^2(x-\mu(F))\{4(x-\mu(F))^3 - 4\mu_3(F) - 6\mu_2(F)(x-\mu(F))\} \\ +6\varepsilon^3(x-\mu(F))^2\{(x-\mu(F))^2 - \mu_2(F)\} \\ -3\varepsilon^4(x-\mu(F))^4.\end{aligned} \tag{16}$$

We are now in a position to compute the influence function of β_2:

$$\begin{aligned} IF(x;F,\beta_2) &= \lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\left\{\frac{\mu_4(F_\varepsilon)}{\mu_2^2(F_\varepsilon)} - \frac{\mu_4(F)}{\mu_2^2(F)}\right\} \\ &= \lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\frac{\mu_2^2(F)\cdot\mu_4(F_\varepsilon) - \mu_4(F)\cdot\mu_2^2(F_\varepsilon)}{\mu_2^2(F)\cdot\mu_2^2(F_\varepsilon)}.\end{aligned}$$

Using (15) and (16) we obtain, as $\varepsilon\downarrow 0$,

$$\begin{aligned} IF(x;F,\beta_2) = \frac{1}{\mu_2^4(F)}\{\mu_2^2(F)\cdot[(x-\mu(F))^4 - \mu_4(F) - 4\mu_3(F)\cdot(x-\mu(F))] \\ -2\mu_4(F)\cdot\mu_2(F)\,[(x-\mu(F))^2 - \mu_2(F)]\}\end{aligned}$$

and replacing $(x-\mu(F))/\sqrt{\mu_2(F)}$ by z and $\mu_3(F)/[\mu_2(F)]^{3/2}$ by γ_1 it reduces to:

$$IF(x;F,\beta_2) = (z^2-\beta_2)^2 - \beta_2(\beta_2-1) - 4\gamma_1 z \tag{17}$$

which is nothing but Faleschini's derivative (8) multiplied by P (the size of the distribution) and rewritten in a continuous random variable notation (z_r is substituted for z). Although Faleschini considered β_2 as a moment ratio and not as a functional of a probability distribution F, he had already answered the basic question addressed by Ruppert: how does β_2 change if we throw in an additional observation at some point on the real line? The partial derivative $\partial\beta_2/\partial f_r$ is indeed the first intuitive notion of the influence function of β_2 and provides the "*quantitative understanding of kurtosis*" invoked by Ruppert (1987).

It is worth mentioning that Ruppert did not compute (17) but a slight variant called the symmetric influence function (SIF)[1]. He started with a symmetric distribution F with $\mu(F) = 0$ and contaminated it at $\pm x$ with equal probability, so as to leave the position of $\mu(F)$ unchanged. The resulting SIF of β_2 differs from (17) only in the last term (which obviously disappears), hence it provides the same information as the IF of β_2 when the underlying distribution is itself symmetric (we refer the reader to Fiori (2005) for further details).

[1] It is interesting to note that Ruppert himself referred the SIF of β_2 to an earlier intuition by Darlington (1970).

5. CONCLUSIONS

We have shown that Luigi Faleschini devised an influence function approach to kurtosis in 1948, i.e. twenty years before the IF was available in Hampel's thesis (1968) and nearly forty years before the publication of Ruppert's SIF for β_2 (1987). By Faleschini's derivative $\partial\beta_2/\partial f_r$ we have illustrated that peakedness and tailedness are both components of kurtosis, and β_2 is clearly dominated by tail weight.

An alternative to Faleschini's idea of adding observations at a point in a frequency distribution is to consider transfers between different units for a fixed distribution size. This technique is central to income inequality studies and appears to have first been used in kurtosis measurement by Zenga (1996, 2006). By transformations that increase (decrease) kurtosis in an observed population Zenga obtained a "two-faced" variant of the Lorenz curve which he termed the "kurtosis diagram".

Besides allowing for a partial ordering of distributions by kurtosis, the kurtosis diagram suggests how to construct new kurtosis measures with desirable properties. Interested readers are referred to the appropriate references.

Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali Università degli Studi di Milano-Bicocca

ANNA MARIA FIORI
MICHELE ZENGA

REFERENCES

K.P. BALANDA, H.L. MacGILLIVRAY (1988), *Kurtosis: a critical review*, "The American Statistician", 42, n. 2, pp. 111-119.

R.B. DARLINGTON (1970), *Is kurtosis really peakedness?* "The American Statistician", 24, n. 2, pp. 19-22.

L. FALESCHINI (1948), *Su alcune proprietà dei momenti impiegati nello studio della variabilità, asimmetria e curtosi*, "Statistica", 8, pp. 503-513.

A.M. FIORI (2005), *La curtosi: nuovi aspetti metodologici e prospettive inferenziali*, Tesi di Dottorato in Statistica Metodologica ed Applicata, Università degli Studi di Milano-Bicocca, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali.

F.R. HAMPEL (1968), *Contributions to the theory of robust estimation.* Ph.D. Dissertation, University of California, Berkeley, Department of Statistics.

F.R. HAMPEL (1974), *The influence curve and its role in robust estimation*, "Journal of the American Statistical Association", 69, pp. 383-393.

B.K. KALE, G. SEBASTIAN (1996), *On a class of nonnormal symmetric distributions with a kurtosis of three*, in H.N. NAGARAJA, P.K. SEN, D.F. MORRISON (eds.), *Statistical theory and applications: papers in honor of Herbert J. David*, Springer-Verlag, New York.

K. PEARSON (1905), *Skew variation, a rejoinder*, "Biometrika", 4, n. 1/2, pp. 169-212.

M. POLISICCHIO, M. ZENGA (1997), *Kurtosis diagram for continuous variables*, "Metron", 55, pp. 21-41.

A. POLLASTRI (1999), *Analysis of the kurtosis in a skew distribution*, "Metron", 57, pp. 131-146.

D. RUPPERT (1987), *What is kurtosis? An influence function approach*, "The American Statistician", 41, n. 1, pp. 1-5.

A. STUART, J.K. ORD (1994), *Kendall's advanced theory of statistics* (*Sixth Edition*), *vol. I*: *Distribution Theory*, Edward Arnold, London.
M. ZENGA (1996), *La curtosi*, "Statistica", 56, pp. 87-102.
M. ZENGA (2006), *Kurtosis*, in S. KOTZ, C.B. READ, N. BALAKRISHNAN AND B. VIDAKOVIC (eds.), *Encyclopedia of Statistical Sciences* (*second edition*), John Wiley and Sons, New York.

RIASSUNTO

L'interpretazione della curtosi e la sua curva di influenza in un'intuizione di L. Faleschini

In questo lavoro si è mostrato come Luigi Faleschini, in un articolo apparso su *Statistica* nel 1948, abbia per primo intuito la derivazione della curva di influenza del coefficiente di curtosi.

SUMMARY

The meaning of kurtosis, the influence function and an early intuition by L. Faleschini

In this work we have shown that an onverlooked paper by Luigi Faleschini (1948) contains the first intuition of the influence function approach to kurtosis.