Statistics & Probability Letters 8 (1989) 297–299
North-Holland
September 1989

SHARP INEQUALITIES BETWEEN SKEWNESS AND KURTOSIS

Vijay K. ROHATGI and Gábor J. SZÉKELY
Bowling Green State University, Bowling Green, OH 43403-0221, USA

Received September 1988
Revised December 1988

Abstract: Denote by α and β the skewness and kurtosis respectively of a distribution with finite fourth moment. We show that $\alpha^2 \leqslant \beta + 2$. For unimodal distributions we prove that $\alpha^2 \leqslant \beta + \frac{6}{5}$, and for infinitely divisible distributions $\alpha^2 \leqslant \beta$.

Keywords: skewness, kurtosis, mixtures of distributions, unimodal distributions, infinitely divisible distributions.

1. Introduction

Denote by m the expectation of a distribution function F and by m_n, $n = 1, 2, \ldots$, the central moment of order n. Then the variance of F is $\sigma^2 = m_2$, the skewness is $\alpha = m_3/\sigma^3$, and the kurtosis is given by $\beta = m_4/\sigma^4 - 3$. A direct application of the Cauchy–Schwartz inequality gives

$$m_3^2 \leqslant m_4 m_2,$$

so that

$$\alpha^2 \leqslant \beta + 3.$$

In this note first we show that the constant 3 on the right-hand side can be replaced by 2 and this is the best constant. Further improvements are then obtained by restricting F

2. Results

We shall first prove the following result.

Theorem 1. *The inequality*

$$m_3^2 \leqslant m_4 m_2 - m_2^3 \tag{1}$$

or, equivalently

$$\alpha^2 \leqslant \beta + 2 \tag{2}$$

holds with equality if and only if the distribution F is concentrated on at most two points.

Remark 1. According to Balanda and MacGillivray (1988, p. 112) the relationship between the skewness and kurtosis of a distribution "receives little attention because of the common practice of restricting the discussion of kurtosis to symmetric distributions only" and (p. 119) "... the notion of kurtosis in asymmetric distributions and its relationship with skewness need further attention".

Remark 2. Skewness may also be measured by $\alpha^* = (m - \mu)/\sigma$ where μ is the median of F. The inequality $|\alpha^*| \leqslant 1$ was proved by Hotelling and Solomons (1932) and improved by Majinder (1962).

Proof of Theorem 1. A discrete version of Theorem 1 was proved by Wilkins (1944) who attributes it to Pearson (1916). Wilkins also notes that the equality in (1) (or (2)) holds if and only if F is concentrated on two points.

To prove (1) in general we assume without loss of generality that $m = 0$. Then (1) can be rewritten as

$$(\mathscr{E}X^3)^2 \leqslant \mathscr{E}X^4 \mathscr{E}X^2 - (\mathscr{E}X^2)^3. \tag{3}$$

Set $u = \mathscr{E}X^4$, $v = \mathscr{E}X^3$ and $w = \mathscr{E}X^2$. Since $z = uw$

$-w^3-v^2$ is concave the theorem will be proved if we can show that:

(i) for all two point distributions with zero mean equality holds in (3);

(ii) all finite distributions with zero mean are (convex) mixtures of distributions concentrated on at most two points. (All distributions can be approximated by finite distributions. Since z is strictly concave the equality can hold only for distributions concentrated on at most two points.)

If a distribution is concentrated on two points a and b, $a \leqslant 0 \leqslant b$, and has zero expectation then it takes the value a with probability $b/(b-a)$ and the value b with probability $-a/(b-a)$. It is easy to check that for all such distributions we have equality in (3).

Now suppose that the support of F is

$$a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n \quad (n \geqslant 3)$$

and the value a_i is taken with probability p_i, $i=1,2,\ldots,n$. Then $\Sigma_{i=1}^n p_i = 1$ and since $m=0$, $\Sigma_{i=1}^n a_i p_i = 0$. We proceed by induction. Replace a_{n-1} and a_n by

$$a'_{n-1} = \frac{p_{n-1}}{p_{n-1}+p_n} a_{n-1} + \frac{p_n}{p_{n-1}+p_n} a_n$$

and suppose a'_{n-1} is taken with probability $p_{n-1} + p_n$. Then this new distribution concentrated on $a_1,\ldots,a_{n-2}$, a'_{n-1} also has zero expectation and proceeding inductively we can suppose that it is a mixture of two point distributions with zero expectation. But then F is a mixture of distributions with zero expectation and concentrated on at most 3 points (a_i, a_{n-1}, a_n for some $i=1,2,\ldots,n-2$). Now we only need to decompose all three point distributions with zero expectation as a mixture of two two-point distributions with zero expectation, and this is an easy exercise. □

Remark 3. Theorem 1 can also be proved using the obvious inequality

$$0 \leqslant \mathscr{E}\left\{(X-a)^2(X-b)^2\right\}. \tag{4}$$

Indeed if a random variable X is concentrated only on two points a and b and it is given that $\mathscr{E}X=0$, $\operatorname{var}(X)=1$ and $\mathscr{E}X^4 = m_4$ then

$$a = \frac{\sqrt{m_4-1} - \sqrt{m_4+3}}{2} \quad \text{and}$$

$$b = \frac{\sqrt{m_4-1} + \sqrt{m_4+3}}{2}.$$

Substituting these values in (4) we get

$$0 \leqslant m_4 - 1 - m_3\sqrt{m_4-1}$$

which is (1). Equality holds if and only if X takes at most two values.

An obvious generalization can be derived from the inequality

$$0 \leqslant \mathscr{E}\prod_{i=1}^{n}(X-a_i)^2 \tag{5}$$

where $a_1,\ldots,a_n$ are arbitrary real numbers. Suppose X takes these values with respective probabilities $p_1,\ldots,p_{n-1}$, $1-p_1,\ldots,p_{n-1}$. Thus the degree of freedom of the system is $2n-1$ and we may impose $2n-1$ conditions, say $2n-1$ moments are given. Then a_i, $i=1,\ldots,n$ can be expressed in terms of these $2n-1$ moments and from (5) we get an inequality between $m_1,\ldots,m_{2n}$ which turns into an equality if and only if X takes at most n values. All these inequalities, however, do not seem to be as interesting as the case $n=2$ discussed above.

Theorem 2. *For unimodal distributions the inequality*

$$\alpha^2 \leqslant \beta + \tfrac{6}{5} \tag{6}$$

holds with equality for uniform distributions.

For infinitely divisible distributions

$$\alpha^2 \leqslant \beta \tag{7}$$

with equality for normal and Poisson distributions.

Proof. A classical characterization of unimodal distributions due to A. Khintchine (see, for example, Feller, 1971, p. 158) states that all unimodal distributions are mixtures of uniform ones. Thus we just have to check (6) for uniform distributions. Indeed for uniform distribution on (a, b)

we have $m_2 = \frac{1}{12}(b-a)^2$, $m_3 = 0$, and $m_4 = \frac{1}{80}(b-a)^4$ so that

$$m_3^2 = 0 = m_4 m_2 - c m_2^3$$

gives $c = \frac{9}{5}$ and (6) follows.

To prove (7) we apply the Lévy–Khintchine representation of an infinitely divisible distribution to conclude (Feller, 1971, p. 566) that all infinitely divisible distributions can be approximated by translations of convolutions of normal and general Poisson distributions. For the normal and Poisson distributions it is easy to check that

$$m_3^2 = m_4 m_2 - 3m_2^3. \tag{8}$$

Indeed for a normal distribution $m_3 = 0$, $m_4 = 3m_2^2$ and for a Poisson distribution $m_2 = m_3 = \lambda$, $m_4 = \lambda + 3\lambda^2$ so (8) holds in either case.

Recall now that for independent (centered) random variables X and Y,

$$m_3(X+Y) = m_3(X) + m_3(Y).$$

and

$$m_4(X+Y) = m_4(X) + m_4(Y) + 6m_2(X)m_2(Y).$$

Applying these identities we see easily that (7) remains valid for convolutions of normal and general Poisson distributions.

As an example, we note that for the gamma density function

$$f(x) = x^{\lambda-1}\,\mathrm{e}^{-x}/\Gamma(\lambda), \quad x > 0,$$

$m_2 = \lambda$, $m_3 = 2\lambda$, $m_4 = 3\lambda^2 + 6\lambda$ so that strict inequality holds in (7). □

References

Balanda, K.P. and H.L. MacGillivray (1988), Kurtosis: A critical review, *Amer. Statist.* **42**, 111–119.

Feller, W. (1971), *An Introduction to Probability Theory and its Applications, Vol. II* (Wiley, New York).

Hotelling, H. and L.M. Solomons (1932), The limits of a measure of skewness, *Ann. Math. Statist.* **3**, 141–142.

Majinder, K.N. (1962), Improved bounds on a measure of skewness, *Ann. Math. Statist.* **33**, 1192–1194.

Pearson, K. (1916), Mathematical contributions to the theory of evolution, XIX; second supplement to a memoir on skew variation. *Philos. Trans. Roy. Soc. London Ser. A* **216**, 432.

Wilkins, J.E. (1944), A note on skewness and kurtosis, *Ann. Math. Statist.* **15**, 333–335.