

## Convex transformations: A new approach to skewness and kurtosis \*)

by W. R. van Zwet \*\*)

UDC 519.2

### Samenvatting

*In dit artikel worden een tweetal orde-relaties voor waarschijnlijkheidsverdelingen voorgesteld, die – beter dan de klassieke maten gebaseerd op derde en vierde momenten – aangeven wanneer een verdeling een grotere scheefheid of kurtosis bezit dan een andere verdeling. Voorts wordt een aantal karakterisering en toepassingen van deze orde-relaties behandeld. Bewijzen worden in dit artikel niet gegeven; deze zijn te vinden in een meer uitgebreide, aan dit onderwerp gewijde studie [5].*

### 1. Introduction

Every statistician will have at least an intuitive idea of what is meant by the concepts of „skewness” and „kurtosis” of a probability distribution and he will be aware of the fact that these should play an important role in applications. He will also probably feel vaguely dissatisfied with the existing measures for these concepts, i.e. the standardized third and fourth central moments, and indeed there are at least two perfectly good reasons for this uneasy feeling.

The first one is that, according to these measures, any pair of probability distributions that possess finite fourth moments may be compared as to skewness and kurtosis, whereas one feels that pairs of such distributions exist that are quite incomparable in these respects. The second reason is that, to the author's knowledge, very few interesting applications of any generality exist. It is fairly obvious that both disadvantages are closely related: the reason for the apparent lack of applications is precisely the fact that comparison of probability distributions on the basis of these measures is so often meaningless.

From the above it will be clear that at the root of the trouble lies the fact that these measures impose a simple ordering – i.e. an ordering where every pair of elements are comparable – on too large a class of probability distributions. Rather than restricting ourselves to considering smaller classes of distributions we shall try and find a more satisfactory approach by considering partial orderings – i.e. orderings where not every pair of distributions are necessarily comparable – to replace the classical measures.

\*) Rapport S 328 van de afdeling Mathematische Statistiek van het Mathematisch Centrum te Amsterdam; Lezing gehouden op de Statistische Dag 1964.

\*\*) Sous-chef voor mathematische statistiek van het Mathematisch Centrum te Amsterdam.

It will be shown in this paper that two partial order relations exist that seem to cover our intuitive ideas about skewness and kurtosis. These order relations will not only be seen to imply the ordering according to the classical measures, but also to be so much stronger than the classical orderings as to permit meaningful applications. No proofs will be given in this paper; they may be found in [5], where a more extensive study of the subject is made. Part of the material presented here and in [5] was previously discussed in [4].

## 2. Notation

Let  $\underline{x}$  be a non-degenerate real-valued random variable<sup>1)</sup> and let  $I$  be the smallest interval for which  $P(\underline{x} \in I) = 1$ . We define the distribution function  $F$  of  $\underline{x}$  by

$$F(x) = \frac{1}{2} P(\underline{x} < x) + \frac{1}{2} P(\underline{x} \leq x)$$

and the expectation and central moments of  $\underline{x}$  by

$$\mathcal{E}\underline{x} = \int_I x dF(x),$$

$$\sigma^2(\underline{x}) = \mu_2(\underline{x}) = \int_I (x - \mathcal{E}\underline{x})^2 dF(x), \text{ and}$$

$$\mu_k(\underline{x}) = \int_I (x - \mathcal{E}\underline{x})^k dF(x), \quad k = 3, 4, \dots,$$

where the right-hand sides denote STIELTJES integrals. We shall say that these expectations exist only if they are finite. The distribution given by  $F$  is said to be symmetrical about  $x_0 \in I$  if

$$F(x_0 - x) + F(x_0 + x) = 1 \quad \text{for all real } x.$$

Let  $\underline{x}_{1:n} \leq \underline{x}_{2:n} \leq \dots \leq \underline{x}_{n:n}$  denote an ordered sample of size  $n$  from the distribution  $F$ ;  $\underline{x}_{i:n}$  is called the  $i$ -th order statistic of a sample of size  $n$  from  $F$ . In the greater part of this paper we shall confine our attention to the class  $\mathcal{F}$  of distribution functions  $F$  satisfying

- (a)  $F$  is twice continuously differentiable on  $I$ ;
- (b)  $F'(x) > 0$  on  $I$ ;
- (c) There exist integers  $i$  and  $n$ ,  $1 \leq i \leq n$ , such that  $\mathcal{E}\underline{x}_{i:n}$  exists.

For  $F \in \mathcal{F}$  the inverse function  $G$  is uniquely defined on  $(0,1)$  by<sup>2)</sup>

$$GF(x) = x \quad \text{for } x \in I.$$

<sup>1)</sup> We denote random variables by underlining their symbols.

<sup>2)</sup> We shall usually not use brackets to denote composite functions and write  $GF$  and  $GF(x)$  rather than  $G(F(\cdot))$  and  $G(F(x))$ .

We shall also be concerned with the subclass  $\mathcal{S} \subset \mathcal{F}$  of symmetric distributions in  $\mathcal{F}$ .

When we consider simultaneously two random variables,  $x$  and  $x^*$ , with distribution functions  $F$  and  $F^*$ , we shall adopt similar conventions and notations with regard to  $x^*$  and  $F^*$ , and write:  $I^*$ ,  $x_{i;n}^*$  and  $G^*$ .

A real-valued function  $\varphi$  defined on  $I$  is said to be convex on  $I$  if for all  $x_1, x_2 \in I$  and  $0 \leq \lambda \leq 1$

$$\varphi(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda\varphi(x_1) + (1 - \lambda)\varphi(x_2),$$

i.e. the graph of  $\varphi$  lies below any chord. We note that this definition implies continuity of  $\varphi$  on  $I$ , except perhaps at its endpoints, if these exist. A real-valued function  $\varphi$  on  $I$  is said to be antisymmetrical and concave-convex on  $I$  about  $x_0 \in I$ , if for all  $x_0 - x \in I, x_0 + x \in I$ ,

$$\varphi(x_0 - x) + \varphi(x_0 + x) = 2\varphi(x_0),$$

and if  $\varphi$  is concave for  $x \leq x_0$  and convex for  $x \geq x_0, x \in I$ ;  $x_0$  will be called a central point of  $\varphi$ .

### 3. Convex and concave-convex transformations

Suppose that  $\varphi$  is non-decreasing and convex on  $I$  and consider the random variables  $x$  and  $\varphi(x)$ . Apart from an overall linear change of scale such a transformation of the random variable  $x$  to the random variable  $\varphi(x)$  effects a contraction of the lower part of the scale of measurement and an extension of the upper part. As, moreover, this deformation increases towards both ends of the scale, the transformation from  $x$  to  $\varphi(x)$  produces what one intuitively feels to be an increased skewness to the right. The following theorem holds:

#### Theorem 3.1

If  $\varphi$  is a non-decreasing convex function on  $I$ , which is not constant on  $I$ , and if  $\mu_{2k+1}(x)$  and  $\mu_{2k+1}(\varphi(x))$  exist, then

$$\frac{\mu_{2k+1}(x)}{\sigma^{2k+1}(x)} \leq \frac{\mu_{2k+1}(\varphi(x))}{\sigma^{2k+1}(\varphi(x))}, \quad \text{for } k = 1, 2, \dots$$

It is intuitively equally appealing that a non-decreasing, antisymmetric and concave-convex transformation of a symmetrically distributed random variable should lead to an increased kurtosis of the distribution. We have:

#### Theorem 3.2

Let  $\varphi$  be a non-decreasing, antisymmetrical, concave-convex function on  $I$ ,

which is not constant on  $I$ , and let the distribution given by  $F$  be symmetrical about  $x_0$ , where  $x_0$  denotes a central point of  $\varphi$ . Then, if  $\mathcal{E}\varphi^{2k}(x)$  exists,

$$\frac{\mu_{2k}(x)}{\sigma^{2k}(x)} \leq \frac{\mu_{2k}(\varphi(x))}{\sigma^{2k}(\varphi(x))}, \text{ for } k = 2, 3, \dots$$

#### 4. Two weak-order relations

In the remaining part of this paper we shall confine our attention to distribution functions  $F \in \mathcal{F}$ ; part of the results, however, remain valid without this restriction.

Returning to the theorems of section 3 we remark that they obviously continue to hold if one replaces  $\varphi(x)$  by any other random variable with the same distribution, i.e. they hold for any  $x^*$  with distribution function  $F^*$  satisfying

$$F^*\varphi(x) = P(x^* \leq \varphi(x)) = P(\varphi(x) \leq \varphi(x)) = P(x \leq x) = F(x),$$

or 
$$\varphi(x) = G^*F(x) \quad \text{on } I.$$

We therefore define the following order relations on  $\mathcal{F}$  and  $\mathcal{S}$  respectively:

##### Definition 4.1

If  $F, F^* \in \mathcal{F}$ , then  $F \underset{c}{<} F^*$  (or equivalently  $F^* \underset{c}{>} F$ ) if and only if  $G^*F$  is convex on  $I$ .

##### Definition 4.2

If  $F, F^* \in \mathcal{S}$ , then  $F \underset{s}{<} F^*$  (or equivalently  $F^* \underset{s}{>} F$ ) if and only if  $G^*F$  is convex for  $x > x_0$ ,  $x \in I$ , where  $x_0$  denotes the point of symmetry of  $F$ .

We shall say in this case that  $F$   $c$ -precedes or  $s$ -precedes  $F^*$ , or that  $F^*$   $c$ -follows or  $s$ -follows  $F$ , and that the two are  $c$ -comparable or  $s$ -comparable. We shall also speak of  $c$ -ordering,  $s$ -ordering,  $c$ -comparison,  $s$ -comparison, etc., where the letters  $c$  and  $s$  stand for convex and symmetrical. According to the above the meaning of these definitions is clear:  $F \underset{c}{<} F^*$  if and only if a random variable with distribution  $F$  may be transformed into one with distribution  $F^*$  by an increasing and convex transformation; for symmetrical distributions,  $F \underset{s}{<} F^*$  if and only if this can be done by an increasing, antisymmetrical, concave-convex transformation. From the theorems of the preceding section the implications are also obvious: we have every right to say that  $F \underset{c}{<} F^*$  implies that  $F^*$  has greater skewness to the right than  $F$ , whereas for symmetrical distributions  $F \underset{s}{<} F^*$  implies that  $F^*$  has greater kurtosis than  $F$ .



Since it is easily seen that both order relations are reflexive ( $F < F$ ) and transitive ( $F < F^*$ ,  $F^* < F^{**}$  implies  $F < F^{**}$ ) they are weak orderings. If one defines an equivalence relation  $\sim$  by

**Definition 4.3**

If  $F, F^* \in \mathcal{F}$ , then  $F \sim F^*$  if and only if  $F(x) = F^*(ax + b)$  for some constants  $a > 0$  and  $b$ ,

it is also easy to show that  $F \sim F^*$  if and only if  $F <_c F^*$  and  $F^* <_c F$ ; for  $F, F^* \in \mathcal{S}$  one finds that  $F \sim F^*$  if and only if  $F <_s F^*$  and  $F^* <_s F$ . Hence by passing to the collections  $\overline{\mathcal{F}}$  and  $\overline{\mathcal{S}}$  of equivalence classes one may define partial orderings ( $F < F^*$ ,  $F^* < F$  implies  $F = F^*$ ) on  $\overline{\mathcal{F}}$  and  $\overline{\mathcal{S}}$  by ordering equivalence classes according to the  $c$ - and  $s$ -ordering of their representatives.

In statistical parlance the above asserts that  $c$ - and  $s$ -ordering are both independent of location and scale parameters. The classes  $\overline{\mathcal{F}}$  and  $\overline{\mathcal{S}}$  are the classes of types of laws belonging to  $\mathcal{F}$  and  $\mathcal{S}$ . We may consequently restrict our attention to  $c$ - and  $s$ -comparison of standardized distribution functions.

Here we give only two examples of  $c$ - and  $s$ -ordering. The gamma distributions may be shown to be  $c$ -following one another with decreasing values of the parameter, whereas the symmetric beta distributions  $s$ -follow one another with increasing values of the parameter. Further examples may be found in [5].

**5. Characterization theorems**

In this section we give two theorems that provide a number of characterizations of the order relations  $<_c$  and  $<_s$  in terms of inequalities for expected values and odd moments of order statistics.

**Theorem 5.1**

Let  $R$  be a dense subset of  $(0,1)$ . Then for  $F, F^* \in \mathcal{F}$  the following statements are equivalent:

- (1)  $F <_c F^*$ ;
- (2)  $F(\mathcal{E}x_{i:n}) \leq F^*(\mathcal{E}x^*_{i:n})$  for all  $n = 1, 2, \dots$  and  $i = 1, 2, \dots, n$ , for which  $\mathcal{E}x_{i:n}$  and  $\mathcal{E}x^*_{i:n}$  exist;
- (3)  $\frac{\mu_{2k+1}(x_{i:n})}{\sigma^{2k+1}(x_{i:n})} \leq \frac{\mu_{2k+1}(x^*_{i:n})}{\sigma^{2k+1}(x^*_{i:n})}$  for all  $k = 1, 2, \dots, n = 1, 2, \dots$ , and  $i = 1, 2, \dots, n$ , for which  $\mu_{2k+1}(x_{i:n})$  and  $\mu_{2k+1}(x^*_{i:n})$  exist;

(4) If  $i$  and  $n$  tend to infinity in such a way that  $\lim \frac{i}{n} = r, r \in R$ , then

$$\lim n (F^* (\mathcal{E} x_{i;n}^*) - F (\mathcal{E} x_{i;n})) \geq 0;$$

(5) If  $i$  and  $n$  tend to infinity in such a way that  $\lim \frac{i}{n} = r, r \in R$ , then for at least one value of  $k = 1, 2, \dots$

$$\lim \sqrt{n} \left( \frac{\mu_{2k+1} (x_{i;n}^*)}{\sigma^{2k+1} (x_{i;n}^*)} - \frac{\mu_{2k+1} (x_{i;n})}{\sigma^{2k+1} (x_{i;n})} \right) \geq 0.$$

Two remarks should be made about this theorem. The first one is that for a given distribution  $F$  convexity would seem to be a rather heavy requirement to prove the inequalities of theorem 3.1. The equivalence of statements (1) and (5) of theorem 5.1 shows, however, that if these inequalities are to hold even for a single value of  $k$  and for the class of distributions of large sample order statistics from a given distribution, then convexity is necessary as well as sufficient. The second remark is that the equivalence of statements (2) and (4) and of (3) and (5) enable us to derive small sample inequalities from their large sample counterparts.

### Theorem 5.2

Let  $R$  be a dense subset of  $(\frac{1}{2}, 1)$ . Then for  $F, F^* \in \mathcal{S}$  the following statements are equivalent:

(1)  $F <_s F^*$ ;

(2)  $F (\mathcal{E} x_{i;n}) \leq F^* (\mathcal{E} x_{i;n}^*)$  for all  $n = 1, 2, \dots$  and  $\frac{n+1}{2} \leq i \leq n$ , for which  $\mathcal{E} x_{i;n}^*$  exists;

(3) If  $i$  and  $n$  tend to infinity in such a way that  $\lim \frac{i}{n} = r, r \in R$ , then

$$\lim n (F^* (\mathcal{E} x_{i;n}^*) - F (\mathcal{E} x_{i;n})) \geq 0;$$

(4) If  $i$  and  $n$  tend to infinity in such a way that  $\lim \frac{i}{n} = r, \frac{1}{2} < r < 1$ , then for all  $k = 1, 2, \dots$

$$\lim \sqrt{n} \left( \frac{\mu_{2k+1} (x_{i;n}^*)}{\sigma^{2k+1} (x_{i;n}^*)} - \frac{\mu_{2k+1} (x_{i;n})}{\sigma^{2k+1} (x_{i;n})} \right) \geq 0;$$

(5) Statement (4) is valid for all  $r \in R$  and at least one value of  $k = 1, 2, \dots$

We note that small sample inequalities concerning odd moments are lacking; the corresponding large sample result is given in statement (4). For  $i \leq \frac{n+1}{2}$ ,  $0 < r < \frac{1}{2}$  and  $R$  dense in  $(0, \frac{1}{2})$ , the inequalities of theorem 5.2 are of course reversed.

For large classes of distributions  $F \in \mathcal{S}$  small sample inequalities between  $F(x_{i:n})$  and quantities of the type  $\frac{i-\alpha}{n+1-2\alpha}$  may be obtained by  $s$ -comparison with a class of distribution functions for which the inverse functions  $G$  are incomplete beta functions. For these results we refer to [5], where one may also find still another characterization of the order relations  $<_c$  and  $<_s$  in terms of a measure of skewness based on the median.

## 6. Applications

Although it has been made clear that the relations  $<_c$  and  $<_s$  may be taken to indicate increasing skewness and kurtosis, we still have to demonstrate that these relations meet with more success in applications than the classical measures based on third and fourth moments. To this end three examples of comparison of distributions will be considered where skewness or kurtosis obviously play an important role.

The first example is taken from a paper by J. L. HODGES jr. and E. L. LEHMANN [2]. They discuss the relative asymptotic efficiency  $e_{W:N}(F)$  of WILCOXON's two sample test  $W$  to the normal scores test  $N$ , for the case where the underlying distribution is of type  $F$ . Numerical evidence leads them to suppose that  $e_{W:N}$  will increase as the tails of the underlying distribution grow heavier. Application of the relation  $<_s$  to a formula for  $e_{W:N}(F)$  given in [2] immediately yields the desired result:

Under certain regularity conditions,  $F, F^* \in \mathcal{S}$  and  $F <_s F^*$  implies  $e_{W:N}(F) \leq e_{W:N}(F^*)$ .

The second example concerns a paper by H. HOTELLING [3] where the behaviour of STUDENT's test under non-standard conditions is studied. Let  $x_1, x_2, \dots, x_n$  be a random sample from a distribution  $F \in \mathcal{F}$ , for which either  $\mu = \mathcal{E}x$  exists, or  $F \in \mathcal{S}$ ; in the latter case we define  $\mu$  by  $F(\mu) = \frac{1}{2}$ . Furthermore let

$$t_n = \frac{\bar{x} - \mu}{s} \sqrt{n},$$

where  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  and  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ .

The probability that  $t_n$  will exceed a constant value  $t$  will be denoted by  $P(t_n > t | F)$  and we define

$$R_n(F) = \lim_{t \rightarrow \infty} \frac{P(t_n > t | F)}{P(t_n > t | \Phi)}$$

where  $\Phi$  denotes the normal distribution function.

Suppose that, assuming the underlying distribution to be normal, one carries out STUDENT'S right-sided test for the hypothesis  $\mu \leq \mu_0$ , whereas in fact  $F$  is not normal at all. Then obviously  $R_n(F)$  denotes the limit of the ratio of the actual size and the assumed size of the test as both these sizes tend to zero. It may therefore serve to provide a rough idea of what to expect when the assumption of normality is violated.

For  $n = 3$  numerical values found by HOTELLING for some symmetrical distributions seem to indicate – paradoxically enough at first sight – that  $R_n(F)$  decreases as the tails of  $F$  become heavier. Making use of an expression for  $R_n(F)$  given in [3] one easily shows this idea to be correct for  $s$ -ordered symmetric distributions, whereas a similar result may be proved for  $c$ -ordered distributions. In fact we have:

If  $F, F^* \in \mathcal{F}$ , and if either  $\mathcal{E}x, \mathcal{E}x^*$  exist and  $F <_c F^*$ , or  $F, F^* \in \mathcal{S}$  and  $F <_s F^*$ , then  $R_n(F) \geq R_n(F^*)$  for  $n = 2, 3, \dots$

Finally we discuss the relative efficiency of sample median to sample mean in estimating the point of symmetry of a symmetric distribution. Let  $x_1, x_2, \dots, x_n$  denote a random sample from a distribution  $F \in \mathcal{S}$  with finite variance  $\sigma^2(x)$ , and suppose one wishes to estimate  $\mathcal{E}x$ . Two unbiased estimates that are generally used are the sample median

$$\frac{x_{n+1}}{2} : n$$

and the sample mean

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i$$

where we have supposed  $n$  to be odd. The choice between them should depend on the ratio of their (small sample) efficiencies

$$r_n(F) = \frac{\text{eff}\left(\frac{x_{n+1}}{2} : n\right)}{\text{eff}(\bar{x}_n)} = \frac{\sigma^2(x)}{n\sigma^2\left(\frac{x_{n+1}}{2} : n\right)}$$

The following result is easily obtained:

For distributions  $F, F^* \in \mathcal{S}$  having finite variances,  $F <_s F^*$  implies  $r_n(F) \leq r_n(F^*)$  for  $n = 1, 3, 5, \dots$



---

This result supports the statement by G. W. BROWN and J. W. TUKEY [1] that „it is probable that the relative efficiencies of mean and median are greatly affected by the length of the tail”.

### References

- [1] G. W. BROWN, J. W. TUKEY, Some distributions of sample means, *Ann. Math. Statist.* **17** (1946), 1-12.
- [2] J. L. HODGES jr., E. L. LEHMANN, Comparison of normal scores and Wilcoxon tests, *Proc. 4th Berkeley Symp.* (1961) Vol. 1, 307-317.
- [3] H. HOTELLING, The behavior of some standard statistical tests under nonstandard conditions, *Proc. 4th Berkeley Symp.* (1961) Vol. 1, 319-359.
- [4] W. R. VAN ZWET, Two weak-order relations for distribution functions, Report S 303 (VP 17) Dept. of Math. Stat. (1962), Mathematisch Centrum, Amsterdam.
- [5] W. R. VAN ZWET, Convex transformations of random variables, *Math. Centre Tracts* 7 (1964), Mathematisch Centrum, Amsterdam.