
DESCRIPTIVE STATISTICS FOR NONPARAMETRIC MODELS IV. SPREAD

by

P. J. BICKEL (1)*, AND E. L. LEHMANN (2)**

1. Ordering by spread

In the preceding paper of this series [1] (to which we refer as BL III) we studied the dispersion of a symmetric distribution about its center of symmetry. In the present paper we study a related aspect of dispersion, which does not require the assumption of symmetry but which even in the symmetric case does not coincide with the concept considered in BL III. Roughly speaking, instead of looking at dispersion relative to a fixed point, we now consider the spread of a random variable throughout its distribution. The difference is perhaps best explained in terms of an example.

Example 1. Let,

$$(1.1) \quad \begin{aligned} f_p(x) &= \frac{1}{2}p && \text{if } |x| \leq 1, \\ &= \frac{1}{2}(1-p) && \text{if } 1 < |x| \leq 2, \end{aligned}$$

for $0 < p < 1$. Let X have distribution F with density f_p , Y have distribution G with density f_{1-p} for $p > \frac{1}{2}$. Then it follows from (1.2) of BL III that $|Y|$ is stochastically larger than $|X|$ and hence Y is more dispersed about 0 than X according to the definition of BL III. This corresponds to our intuitive feeling that in a global sense G is more dispersed than F since it can be obtained by pushing some of the central mass of F into the tails. Yet, locally for $1 < |X| < 2$, G is more concentrated than F since it has a higher uniform density there.

The basic definition of BL III for calling a symmetric distribution G more dispersed about its center of symmetry than a distribution F about its center of symmetry is equivalent to

$$(1.2) \quad G^{-1}(v) - G^{-1}\left(\frac{1}{2}\right) \stackrel{\geq}{\leq} F^{-1}(v) - F^{-1}\left(\frac{1}{2}\right) \quad \text{as } v \gtrless \frac{1}{2}.$$

* This paper was prepared with support of U.S. Office of Naval Research Contract No. N00014-69-A-0200-1038/NR042-036.

** This research was prepared with support of National Science Foundation Grant No. GP-38485.

In the present paper, we shall call an arbitrary (i.e. not necessarily symmetric) distribution G *more spread out* than a distribution F if

$$(1.3) \quad G^{-1}(v) - G^{-1}(u) \geq F^{-1}(v) - F^{-1}(u) \quad \text{for all } 0 < u < v < 1,$$

where F^{-1} is defined by $F^{-1}(u) = \sup \{x : F(x) \leq u\}$, that is, if two percentage points of G are at least as far apart as the corresponding percentage points of F . A concept which contains the essence of this definition was introduced by Brown and Tukey (1946).

Note that the definition (1.3) reflects the greater concentration of F throughout, and it of course does not require symmetry, thus fulfilling the two desiderata mentioned above. Furthermore, a comparison of (1.3) with (1.2) shows that for symmetric distributions (1.2) is satisfied whenever (1.3) holds; that (1.3) is in fact more stringent follows from Example 1.

Two properties of (1.2) noted for symmetric distributions in BL III are seen to be implied for arbitrary distributions by (1.3), namely,

- (a) Any random variable is more spread out than a constant;
- (b) aX is more spread out than X if $a > 1$.

Note however, that the ordering is not invariant under monotone transformations: $G^{-1}(u) - F^{-1}(u) \uparrow$ does not imply $h[G^{-1}(u)] - h[F^{-1}(u)] \uparrow$.

In BL III it was noted that if F and G are symmetric about 0 with densities f and g satisfying: $g(x)/f(x)$ is increasing for $x > 0$, then (1.2) holds. Example 1 shows that these conditions are not enough to insure (1.3).

Example 2. Let X take on the values $a < b$ with probabilities p and q ($0 < p < 1$). Then

$$(1.4) \quad \begin{aligned} F^{-1}(v) - F^{-1}(u) &= 0 && \text{if } p \leq u \text{ or } v < p, \\ & && b - a \text{ if } u < p \leq v. \end{aligned}$$

From (1.4) it is easily seen that no continuous strictly increasing distribution can be either more or less spread out than F in the sense of (1.3). More generally, continuous distributions and discrete distributions are not comparable. This illustrates how strong a requirement (1.3) is. We shall show below that it is nevertheless satisfied in many cases. However, it is convenient first to give some alternative expressions for condition (1.3).

Suppose that F^{-1} and G^{-1} are differentiable. Dividing both sides of (1.3) by $v - u$, it is then clear that (1.3) implies

$$(1.5) \quad \frac{d}{du} [G^{-1}(u)] \geq \frac{d}{du} [F^{-1}(u)] \quad \text{for all } u$$

and conversely (1.5) implies (1.3). Thus, G is more spread out than F if G^{-1} is nowhere steeper than F^{-1} .

Evaluating the derivatives in (1.5), we see that another form of the condition is

$$(1.6) \quad g[G^{-1}(u)] \leq f[F^{-1}(u)] \quad \text{for all } u.$$

This is a so-called tail ordering condition introduced by Doksum (1969).

We want to stress that in our opinion this ordering corresponds to spread as property (b) above indicates. Tailweight should be specified by a scale-free ordering such as van Zwet's [5] or Lawrence's [4] and measured by scale-free functionals such as the kurtosis.

Lemma 2.2 of Doksum [3] states that, for distributions symmetric about 0, G ordered with respect to F in Lawrence's sense and $g(0) \leq f(0)$ implies that G is more dispersed than F . This is consistent with our point of view. To see this note that $1/f(0)$ is a measure of scale and the condition $g(0) \leq f(0)$ added to the scale-free ordering of Lawrence creates a new ordering of spread which possesses properties (a) and (b).

We mention finally a form which for the sake of simplicity we shall state under the additional assumption that F and G are strictly increasing.

Theorem 1. *If F and G are strictly increasing then G is more spread out than F if and only if there exists a strictly increasing function h such that*

$$(i) \quad x < x' \text{ implies } h(x') - h(x) \geq x' - x$$

and

$$(ii) \quad \text{if } X \text{ has distribution } F, \text{ then } h(X) \text{ has distribution } G.$$

Proof. Suppose first that such a function exists. Then (1.3) follows from the relation $G^{-1}(u) = h[F^{-1}(u)]$. Conversely, if (1.3) holds, it is easily seen that the function $h(x) = G^{-1}[F(x)]$ has the desired property.

Theorem 1 shows that G being more spread out than F means that one can get from F to G by spreading all pairs of points further apart. Note also that Theorem 1 shows that our dispersion ordering depends on $G^{-1}F$ only.

Example 3. Let F be the uniform distribution on $(0, 1)$. Since the right-hand side of (1.6) is then 1 for all $0 < u < 1$, the condition for G with density g to be more spread out than F is that

$$(1.7) \quad g(x) \leq 1 \quad \text{for all } x.$$

Similarly, G is less spread out than F if and only if its support is an interval (a, b) of length < 1 and if in this interval $g(x) \geq 1$ for all x .

Example 4. Let $F = \Phi$ be the standard normal distribution. Then G is more spread out than Φ if and only if the normal probability plot of G , $y = \Phi^{-1} G(x)$ has slope ≥ 1 at all points.

Example 5. Let F be the double exponential distribution with density $\frac{1}{2} \exp(-|x|)$. Then

$$(1.8) \quad f[F^{-1}(u)] = u \quad \text{if } 0 < u < \frac{1}{2}, \\ 1 - u \quad \text{if } \frac{1}{2} < u < 1$$

and G is more spread out than F if

$$(1.9) \quad g(x) \leq G(x) \quad \text{for all } x < 0, \\ 1 - G(x) \quad x > 0.$$

Alternatively the failure rate of G must be at most 1 when x is interpreted as time running in either direction from the origin.

Example 6. Let F be the logistic distribution with density $e^{-x}(1 + e^{-x})^{-2}$. Then, G is more spread out than F if and only if $g(x) \leq G(x)(1 - G(x))$ for all x and less spread out if and only if the reverse inequality holds for all x .

From (1.9) it follows easily, for example, that a normal distribution with sufficiently small variance is less spread out than F , but that a normal distribution can never be more spread out than F no matter how large its variance. The situation is just the reverse in the case of a Cauchy distribution, which is more spread out than F if its scale is sufficiently large, but which can never be less spread out than F .

Finally, the logistic distribution provides an example of a distribution which is more spread out than F for sufficiently large scale and less spread out than F for sufficiently small scale.

An interesting connection between the dispersion ordering of BL III and the present ordering by spread is given by the following result.

Theorem 2. *If Y is more spread out than X and if Y', Y'' and X', X'' are independent copies of Y and X respectively, then $Y'' - Y'$ is more dispersed than $X'' - X'$.*

Proof. Let h be the function guaranteed by Theorem 1. Then $|Y'' - Y'| = |h(X'') - h(X')| \geq |X'' - X'|$ and hence $|Y'' - Y'|$ is stochastically larger than $|X'' - X'|$.

That the converse of Theorem 2 does not hold is shown by the following generalization of Example 1.

Example 7. Let $f = f_p, g = f_{p'}$ be defined by (1.1). Then if $p > p' \geq \frac{1}{2}$, it is easily seen that $X'' - X'$ is less dispersed than $Y'' - Y'$; on the other hand, G is not more spread out than F .

2. Measures of spread

The axioms for a measure of spread coincide with those for a measure of dispersion given in BL III except that in (1.9) of that paper dispersion ordering is replaced by the present spread ordering and that minor differences result from the dropping of

the assumption of symmetry. For the sake of completeness, we shall now restate the full set of axioms.

A *measure of spread* is a functional $\Delta(F)$ (also denoted by $\Delta(X)$ where X is a random variable with distribution F) defined over a sufficiently large class of distributions which is closed under changes of location and scale. We shall require Δ to be nonnegative and to satisfy

$$(2.1) \quad \Delta(aX) = |a| \Delta(X) \quad \text{for } a > 0,$$

$$(2.2) \quad \Delta(X + b) = \Delta(X) \quad \text{for all } b,$$

and

$$(2.3) \quad \Delta(-X) = \Delta(X).$$

As before, these conditions imply that

$$(2.4) \quad \Delta(c) = 0 \quad \text{for any constant } c.$$

The converse: $\Delta(X) = 0$ implies $X = c$, is false for measures of spread as well as for measures of dispersion. Let Δ be the measure given in (2.12) below and F assign mass $\frac{1}{2}$ each to the points ± 1 and $\frac{1}{2}$ to 0. Then $\Delta(F) = 0$. A nonnegative functional satisfying (2.1), (2.2) and (2.3) will be called a measure of spread if it satisfies in addition

$$(2.5) \quad \Delta(F) \leq \Delta(G) \quad \text{whenever } G \text{ is more spread out than } F.$$

Note that if $\Delta(F)$ is a measure of spread, so is $\kappa \Delta(F)$ for any $\kappa > 0$.

A large and interesting class of measures of spread is obtained when the following theorem is applied to some of the results of BL III.

Theorem 3. *Let $\tau(X)$ be a measure of dispersion in the sense of BL III and let X', X'' be two independent copies of X . Then*

$$(2.6) \quad \Delta(F) = \tau(X'' - X')$$

is a measure of spread.

Proof. That Δ satisfies (2.1)–(2.3) is obvious from the fact that τ satisfies the corresponding conditions. To prove (2.5), suppose that G is more spread out than F . Then $X'' - X'$ and $Y'' - Y'$ are symmetric about 0 and it follows from Theorem 2 that $Y'' - Y'$ is more dispersed than $X'' - X'$ and hence from the fact that τ is a measure of dispersion that $\Delta(F) = \tau(X'' - X') \leq \tau(Y'' - Y') = \Delta(G)$ as was to be proved.

Example 8. As a first example let $\tau(F)$ be the standard deviation of X , which was seen in BL III to be a measure of dispersion. Then

$$\Delta(F) = \{E(X'' - X')^2\}^{1/2} = \sqrt{(2)} [\tau(F)].$$

It follows that $\Delta(F)$ is a measure of spread, and hence also that

$$(2.7) \quad \Delta(F) = SD(F)$$

is a measure of spread although no longer restricted to symmetric F .

An obvious generalization, obtained by starting with the p th power deviations considered in BL III are the measures

$$(2.8) \quad \Delta_p(F) = \{E |X'' - X'|^p\}^{1/p}.$$

Example 9. Similarly, by starting with $\tau(F) = \text{med } |X|$, we find that

$$(2.9) \quad \Delta(F) = \text{med } |X'' - X'|$$

is a measure of spread.

Example 10. A class of examples not having the above structure is given by

$$(2.10) \quad \Delta(F) = F^{-1}(t) - F^{-1}(1 - t) \quad \text{for any } t > \frac{1}{2},$$

which obviously satisfies (2.1)–(2.3) and (2.5). The same is true of the more general class

$$(2.11) \quad \Delta(F) = \left[\int_{1/2}^1 [F^{-1}(t) - F^{-1}(1 - t)]^\gamma dA(t) \right]^{1/\gamma}$$

where A is any finite measure on $(\frac{1}{2}, 1)$.

A case of particular interest is $t = \frac{3}{4}$, and hence

$$(2.12) \quad \Delta(F) = F^{-1}(\frac{3}{4}) - F^{-1}(\frac{1}{4})$$

the interquartile range of F . In BL III, this was seen to be a measure of dispersion for symmetric distribution; it now follows that it is a measure of spread for arbitrary distributions. More generally, the class of measures (2.11) when restricted to symmetric distributions for suitable A coincides with the class of dispersion measures (1.10) of BL III. Note, however, that even for $\gamma = 2$ and $A = \text{Lebesgue measure}$, the measure (2.11) is not a multiple of the SD when F is asymmetric.

That not every measure of spread, when restricted to symmetric distributions, reduces to a measure of dispersion is shown by the following example.

Example 11. Let X take on the values $-1, 0, 1$ with probabilities $p/2, 1 - p, p/2$ respectively. Then

$$E\{|X'' - X'|^r\} = 2p(1 - p) + 2^{r-1}p^2$$

and it is easily seen that for $r < 1$, the left-hand side is not an increasing function of p . Since $|X|$ is stochastically increasing with p , this means that $[E\{|X'' - X'|^r\}]^{1/r}$ is not a measure of dispersion for symmetric distributions although it is a measure of spread.

3. Choice of measure

The choice of a measure of spread, as was the case for measures of location and of dispersion will be based largely on the accuracy with which it can be estimated, and the appropriate measure of accuracy is the same of that discussed in BL III for dispersion, the *standardized asymptotic variance*, i.e. the asymptotic variance of the estimator of $\Delta(F)$ divided by $\Delta^2(F)$. As in the earlier paper one would again try to find measures which behave satisfactorily (a) in relation to robustness, either by being robust or at least more robust than $\Delta_1(F)$; and (b) possessing good efficiency relative to $\Delta_1(F)$, ideally for all F , but if this cannot be achieved at least for the type of F likely to occur in practice.

We have not carried out this program in the present case. The work of BL III suggests that in terms of the indicated properties no completely satisfactory measure is likely to exist. It suggests further that the following two types of measures may be reasonable compromise solutions.

(i) *The p th power measures (2.8)*. While not robust, for $p < 2$ these measures are presumably more robust than $\Delta(F)$, which corresponds to the case $p = 2$. The natural estimator of $\Delta(F)$ is $\hat{\Delta}(F)$ or the asymptotically equivalent statistic

$$(3.1) \quad \hat{\Delta}_p = \left\{ \frac{1}{\binom{n}{2}} \sum_{i < j} |X_j - X_i|^p \right\}^{1/p}.$$

Since this is a U -statistic, it follows from the work of Hoeffding that $\sqrt{(n)} \cdot [\hat{\Delta} - \Delta(F)]$ is asymptotically normal. It seems plausible to conjecture that for values of $p \geq 1.5$ the asymptotic efficiency of $\hat{\Delta}_p$ relative to the standard deviation $\hat{\Delta}_2$ will be reasonably high at least for typical distributions.

(ii) *Trimmed standard deviations*. There are two possible versions of trimmed standard deviations in this context.

(a) We can take $\gamma = 2$ and Λ the uniform distribution on $(\frac{1}{2}, 1 - \beta)$ in (2.11).

(b) We can consider $\tau(X - X', \alpha, \beta)$ where $\tau(X, \alpha, \beta)$ is given by (3.1) of BL III.

(As usual we let X represent its distribution.)

It is easy to see that if $0 < \alpha < 1 - \beta < 1$ both of these measures are robust. Their estimates are given by (essentially),

$$(a) \quad [n(\frac{1}{2} - \beta)]^{-1/2} \left[\sum_{k=n/2}^{n(1-\beta)} [X_{(k)} - X_{(n-k+1)}]^2 \right]^{1/2}$$

$$(b) \quad \left[\binom{n}{2} (1 - \beta - \alpha) \right]^{-1/2} \left[\sum_{k=\binom{n}{2}}^{\binom{n}{2}(1-\beta)} (X - X')_{(k)}^2 \right]^{1/2}$$

where $X_{(1)} \leq \dots \leq X_{(n)}$ are the order statistics of X_1, \dots, X_n and $(X - X')_{(1)} \leq \dots \leq (X - X')_{(n)}^{(2)}$ are the order statistics of the pseudo sample $X_i - X_j$, $i < j$.

Asymptotic normality with variance of the order $1/n$ of these estimates is evident only for case (a) when $\gamma = 1$. It seems plausible that this property holds in general and that the resulting efficiencies with respect to the S.D. have the same general numerical features as we found for their analogues in studying dispersion. However, we do not pursue this. We do not know a fortiori which of the measures (a) or (b) is preferable and leave these interesting questions open.

References

- [1] BICKEL, P. J. - LEHMAN, E. L. (1976). Descriptive statistics for nonparametric models III. Dispersion. *Ann. Statist.*, **4**, 1139—1158.
- [2] BROWN, G. - TUKEY, J. W. (1946). Some distributions of sample means. *Ann. Math. Statist.* **7**, 1—12.
- [3] DOKSUM, K. (1969). Starshaped transformations and the power of rank tests. *Ann. Math. Statist.*, **40**, 1167—1176.
- [4] LAWRENCE, M. J. (1975). Inequalities of s -ordered distributions. *Ann. Statist.*, **3**, 413—428.
- [5] ZWET, VAN, W. R. (1964). Convex transformations of random variables. *Math. Centrum*, Amsterdam.

(1), (2) UNIVERSITY OF CALIFORNIA, BERKELEY, CALIFORNIA, U.S.A.

Received October 1975