STOCHASTIC ORDERING OF SPACINGS FROM DEPENDENT RANDOM VARIABLES

BY MOSHE SHAKED and Y. L. TONG

University of Arizona and University of Nebraska

Spacings (that is, the differences between successive order statistics) are useful in various applications in statistics. Many properties of the spacing are known when the spacings are constructed from a collection of independent identically distributed (i.i.d.) random variables. In this paper we study the spacings constructed from not necessarily i.i.d. random variables. We introduce models for which two sets of spacings, constructed from two sets of dependent random variables, can be stochastically ordered. Various examples will be given and applications for goodness-of-fit tests, tests for independence, density estimation and tests for outliers will be discussed.

1. Introduction. Let $X = (X_1, \ldots, X_n)$ denote an $n$-dimensional random vector and let

$$X^{(1)} \leq \ldots \leq X^{(n)}$$

be the ordered components (order statistics) of $X$. The nonnegative random variables

$$U_i = X^{(i+1)} - X^{(i)}, \quad i = 1, \ldots, n-1$$

are called the spacings and have various applications in statistics. For example, certain non-parametric test procedures depend on the maximum spacing or on linear combinations of spacings (see, e.g., Pyke (1965), Weiss (1965), Rao and Sethuraman (1970) and Kirmani and Alam (1974)); certain estimation and test procedures based on order statistics, such as those which depend on the range or midrange, involve spacings (David (1970), Ch. 6); and certain tests for slippage (Karlin and Truax (1960)) and outliers (Barnett and Lewis (1978), Ch. 3) also depend on spacings. For a comprehensive treatment of spacings see Pyke (1965, 1972).

In the literature the problem of spacings has been treated extensively under the assumption that $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables. In certain applications which involve a mixture of experiments, a (random) change of scale or a random shift in location may take place; then the random variables $X_1, \ldots, X_n$ are no longer independent. In this paper we study how the degree of dependence affects the distribution of the spacings. In the case when $X_1, \ldots, X_n$ are interchangeable, it follows from our main result that (in the model under consideration) the spacings vector $U = (U_1, \ldots, U_{n-1})$ becomes stochastically smaller if $X_1, \ldots, X_n$ are more positively dependent (that is, when $X_1, \ldots, X_n$ have more tendency to "hang together").

After stating the model and proving the main result in Section 2, we apply the result to an additive, a multiplicative and a ratio model. In Section 3, after combining results given in Shaked and Tong (1985), we obtain a partial ordering property for the spacings which correspond to a number of important multivariate distributions, such as the multivariate normal, multivariate stable, multivariate beta and the Dirichlet distribution. For all these distributions the corresponding spacings vector $U$ can be partially ordered through the degree of dependence of the components $X_1, \ldots, X_n$ of $X$.

\footnote{Research partially supported by the NSF Grant MCS-8200098.}

\footnote{Research partially supported by the NSF Grant MCS-8100775.}

\footnote{AMS 1980 subject classifications. Primary 60E15; secondary 62G30.}

Key words and phrases. Stochastic ordering of spacings, dependent random variables.
In Section 4 we give applications and study the monotonicity properties of certain well-known procedures concerning goodness-of-fit tests, tests for independence, density estimation and slippage tests for outliers, which all depend on spacings.

2. The Model. Let \( X = (X_1, \ldots, X_n) \) and \( Y = (Y_1, \ldots, Y_n) \) denote two \( n \)-dimensional random vectors and let

\[
X_{(1)} \leq \cdots \leq X_{(n)}, Y_{(1)} \leq \cdots \leq Y_{(n)}
\]

be their ordered components. Define the \((n-1)\)-dimensional spacings vectors by

\[
U = (U_1, \ldots, U_{n-1}) \text{ where } U_i = X_{(i+1)} - X_{(i)}, i = 1, \ldots, n-1.
\]

\[
V = (V_1, \ldots, V_{n-1}) \text{ where } V_i = Y_{(i+1)} - Y_{(i)}, i = 1, \ldots, n-1.
\]

The stochastic ordering of \( U \) and \( V \) will be developed under the following model:

**Model A.** There exist a random vector \( Z = (Z_1, \ldots, Z_n) \), random vectors (of any dimension) \( W_1, W_2, W_3, W_4 \) and Borel-measurable functions \( \phi, \delta_1 \) and \( \delta_2 \) such that

\[
(X_1, \ldots, X_n) \overset{d}{=} (Z_1 \phi(Z, W_1) + \delta_1(Z, W_3), \ldots, Z_n \phi(Z, W_1) + \delta_1(Z, W_3))
\]

\[
(Y_1, \ldots, Y_n) \overset{d}{=} (Z_1 \phi(Z, W_2) + \delta_2(Z, W_4), \ldots, Z_n \phi(Z, W_2) + \delta_2(Z, W_4)).
\]

Moreover, the following conditions are satisfied:

**A1.** \( Z_1, \ldots, Z_n \) are i.i.d., \( W_i \) is independent of \( Z \), \( i = 1, 2 \).

**A2.** \( \phi(z,w), \delta_1(z,w) \) and \( \delta_2(z,w) \) are permutation symmetric functions of \( z = (z_1, \ldots, z_n) \) for every fixed \( w \), and \( \phi > 0 \) for all \((z,w)\) whenever \( z \) is in the support of \( Z \) and \( w \) is in the support of \( W_1 \) or \( W_2 \).

**A3.** Either \( \phi(z,w) \) is nondecreasing (componentwise) in \( w \) for every \( z \) and \( W_1 \overset{sf}{=} W_2 \), or \( \phi(z,w) \) is nonincreasing in \( w \) for every \( z \) and \( W_1 \overset{st}{=} W_2 \).

**Theorem 1.** Assume that \( X \) and \( Y \) have the representation of Model A and that A1, A2 and A3 are satisfied. Then, for all \( k \) and all constants \( c_{ij}, i = 1, \ldots, n, j = 1, \ldots, k, \) such that \( \Sigma_{i=1}^n c_{ij} = 0 \), \( j = 1, \ldots, k, \)

\[
E\psi(\Sigma_{i=1}^n c_{ij} X_{(i)}), \ldots, \Sigma_{i=1}^n c_{ik} X_{(i)})
\]

\[
\geq E\psi(\Sigma_{i=1}^n c_{ij} Y_{(i)}), \ldots, \Sigma_{i=1}^n c_{ik} Y_{(i)})
\]

holds for all \( \psi \) which are componentwise nondecreasing such that the expectations exist. Consequently,

\[
U \overset{sf}{=} V.
\]

**Proof.** Let \( Z_{(1)} \leq \cdots \leq Z_{(n)} \) denote the order statistics of \( Z = (Z_1, \ldots, Z_n) \). Since \( \phi, \delta_1, \) and \( \delta_2 \) are permutation symmetric in \( z_1, \ldots, z_n \) for every fixed \( w \), we must have, a.s.,

\[
\phi(Z, W_j) = \phi(Z_{(1)}, \ldots, Z_{(n)}, W_j)
\]

\[
\delta_j(Z, W_{2+j}) = \delta_j(Z_{(1)}, \ldots, Z_{(n)}, W_{2+j})
\]

for \( j = 1, 2 \). This implies that

\[
Z_j \phi(Z, W_j) + \delta_j(Z, W_{j+2}) \leq Z_i \phi(Z, W_j) + \delta_j(Z, W_{j+2})
\]

holds if and only if \( Z_i \leq Z_j \). Consequently,

\[
(X_{(1)}, \ldots, X_{(n)}) \overset{d}{=} (Z_{(1)} \phi(Z, W_1) + \delta_1(Z, W_3), \ldots, Z_{(n)} \phi(Z, W_1) + \delta_1(Z, W_3))
\]

\[
= \phi(Z, W_1) (Z_{(1)}, \ldots, Z_{(n)}) + (\delta_1(Z, W_3), \ldots, \delta_1(Z, W_3)).
\]
Hence, for all \( c_{ij}, i = 1, \ldots, n, j = 1, \ldots, k \), such that \( \sum_{i=1}^{n} c_{ij} = 0 \),

\[
\left( \sum_{i=1}^{n} c_{ii} X_{(i)} \right) \ldots \left( \sum_{i=1}^{n} c_{ik} X_{(i)} \right) \equiv d \Phi(Z, W_1) \left( \sum_{i=1}^{n} c_{i1} Z_{(i)} \right) \ldots \left( \sum_{i=1}^{n} c_{ik} Z_{(i)} \right).
\]

Without loss of generality assume that \( \Phi \) is nondecreasing in \( w \) and that \( W_1 \geq W_2 \). Then

\[
E\Psi\left( \left| \sum_{i=1}^{n} c_{i1} X_{(i)} \right| \ldots \left| \sum_{i=1}^{n} c_{ik} X_{(i)} \right| \right) = E\Psi\left( \Phi(Z, W_1) \left| \sum_{i=1}^{n} c_{i1} Z_{(i)} \right| \ldots \left| \sum_{i=1}^{n} c_{ik} Z_{(i)} \right| \right) = E\xi_2(Z)
\]

where \( \xi_2(Z) \) is the conditional expectation of \( \psi \), over the distribution of \( W_1 \), given \( Z = z \). Let \( \xi_2(z) \) denote the similar conditional expectation of \( \psi \) over the distribution of \( W_2 \).

Since \( \Phi \) is a nondecreasing function of \( w \) it follows that \( \psi(\Phi(z,w)|\sum_{i=1}^{n} c_{i1} z_{(i)}|, \ldots, \phi(z,w)|\sum_{i=1}^{n} c_{ik} z_{(i)}|) \) is also a nondecreasing function of \( w \) for every fixed \( z \). Thus, \( W_1 \geq W_2 \) implies that

\[ \xi_1(Z) \geq \xi_2(Z) \text{ a.s.} \]

The proof is now completed by applying the equality

\[ E\Psi(\left| \sum_{i=1}^{n} c_{i1} Y_{(i)} \right| \ldots \left| \sum_{i=1}^{n} c_{ik} Y_{(i)} \right|) = E\xi_2(Z). \]

The last statement of the theorem follows by defining \( k = n-1, c_{i+1,i} = 1, c_{i,i} = -1, c_{i,j} = 0, j \neq i, i+1 \) where \( i = 1, \ldots, n-1 \).

In the following we consider special forms of Model A.

(a) (An additive model). Assume that there exist constants \( a \geq 0, b \) and \( d \) (\( b \) and \( d \) have the same sign) and independent random variables \( Z_1, \ldots, Z_n \) and \( W \) such that

\[
(X_1, \ldots, X_n) \equiv (aZ_1 + bW, \ldots, aZ_n + bW)
\]

\[
(Y_1, \ldots, Y_n) \equiv (cZ_1 + dW, \ldots, cZ_n + dW).
\]

Note that without loss of generality we can assume that \( b \geq 0 \) and \( d \geq 0 \) because otherwise one can replace \( W \) by \( -W \).

Letting \( W_1 \) and \( W_2 \) be degenerate at \( a > 0 \) and \( c > 0 \), respectively, letting \( W_3 \equiv W_4 \equiv W \), setting \( \delta_1(z,w) = bw \) and \( \delta_2(z,w) = dw \) and assuming \( a \geq c \), it is easy to see that \( X \) and \( Y \) of (2.2) have the representation of Model A and satisfy A1, A2 and A3 provided \( Z_1, \ldots, Z_n \) are i.i.d.

In some applications (see Shaked and Tong (1985)) \( X \) and \( Y \) have the same marginals, that is,

\[
X_i \equiv Y_i, i = 1, \ldots, n.
\]

In other applications the following condition, which is weaker than (2.3), holds:

\[
\sum_{i=1}^{n} E X_i = \sum_{i=1}^{n} E Y_i
\]

For example, (2.4) holds if \( EZ_1 = \ldots = EZ_n = EW = 0 \) or if \( EZ_1 = \ldots = EZ_n = EW \) and \( a + b = c + d \).

Let \( U_i = X_{i(i+1)} - X_{(i)} \) and \( V_i = Y_{i(i+1)} - Y_{(i)} \), \( i = 1, \ldots, n-1 \). From Theorem 1 it follows that if \( X \) and \( Y \) satisfy (2.2) with \( a \geq c \) then

\[ (U_1, \ldots, U_{n-1}) \geq (V_1, \ldots, V_{n-1}). \]

Shaked and Tong (1985) considered the condition

\[
(EX_{(1)}, \ldots, EX_{(n)}) \geq (EY_{(1)}, \ldots, EY_{(n)}).
\]

that is,

\[ \sum_{i=1}^{k} E X_{(i)} \leq \sum_{i=1}^{k} E Y_{(i)}, k = 1, \ldots, n-1 \]
They denoted the relation (2.6) by \( X \rho Y \) and discussed some applications. Following their arguments it follows that if \( X \) and \( Y \) satisfy (2.5) and (2.4) then \( X \rho Y \). Thus

**Proposition 1.** If \( X \) and \( Y \) satisfy (2.2) with \( a \geq c \) and (2.4), then \( X \rho Y \).

The special case where \( X \) and \( Y \) have the representation

\[
(X_1, \ldots, X_n) \overset{d}{=} (1-p)^{\lambda_1}(Z_1, \ldots, Z_n) + \rho^{\lambda_2}(W_1, \ldots, W),
\]
\[
(Y_1, \ldots, Y_n) \overset{d}{=} (1-\eta)^{\lambda_1}(Z_1, \ldots, Z_n) + \eta^{\lambda_2}(W_1, \ldots, W),
\]

where \( Z_1, \ldots, Z_n \) and \( W \) are as in (2.2), \( 0 \leq p < \eta \leq 1 \) and \( \alpha > 0 \), is Model 4.1 in Shaked and Tong (1985).

Note that in this special case we can actually have a stronger statement concerning the distribution of \( U \) and \( V \) corresponding to \( X \) and \( Y \). That is,

\[
(U_1, \ldots, U_{n-1}) \overset{d}{=} (1-p)^{\lambda_1}(Z_{(2)}-Z_{(1)}, \ldots, Z_{(n)}-Z_{(n-1)})
\]

(2.8)

\[
(V_1, \ldots, V_{n-1}) \overset{d}{=} (1-\eta)^{\lambda_1}(Z_{(2)}-Z_{(1)}, \ldots, Z_{(n)}-Z_{(n-1)}).
\]

Thus we have

\[
(U_1, \ldots, U_{n-1}) \overset{d}{=} \left( \frac{1-p}{1-\eta} \right)^{\lambda_1}(V_1, \ldots, V_{n-1}).
\]

Consequently

\[
\sum_{i=1}^{n-1} \lambda U_i \overset{d}{=} \left( \frac{1-p}{1-\eta} \right)^{\lambda_1} \sum_{i=1}^{n-1} V_i
\]

for all \( \lambda_1, \ldots, \lambda_{n-1} \).

(b) (A multiplicative model). Assume \( X \) and \( Y \) have the representation

\[
(X_1, \ldots, X_n) \overset{d}{=} W_1(Z_1, \ldots, Z_n) + (W_3, \ldots, W_4),
\]
\[
(Y_1, \ldots, Y_n) \overset{d}{=} W_2(Z_1, \ldots, Z_n) + (W_4, \ldots, W_4),
\]

where \( W_1 \) and \( Z \) are independent, \( W_2 \) and \( Z \) are independent and \( Z_1, \ldots, Z_n \) are i.i.d. If \( W_1 \) and \( W_2 \) are nonnegative a.s. and \( W_1 \overset{st}{=} W_2 \) then this is a special case of Model A and Theorem 1 applies.

(c) (A ratio model). In certain situations \( X_1, \ldots, X_n \) have the representation

\[
(X_1, \ldots, X_n) \overset{d}{=} (Z_1 / (\Sigma_{i=1}^n h(Z_i) + W_1), \ldots, Z_n / (\Sigma_{i=1}^n h(Z_i) + W_1)),
\]

where \( Z_1, \ldots, Z_n \) and \( W_1 \) are independent and \( Z_1, \ldots, Z_n \) are i.i.d. [When \( Z_i \geq 0, W_1 > 0 \) a.s. and \( h \) is the identity function, then

\[
(X_1, \ldots, X_n) \overset{d}{=} (Z_1 / (\Sigma_{i=1}^n Z_i + W_1), \ldots, Z_n / (\Sigma_{i=1}^n Z_i + W_1)).
\]

In this model if

\[
(Y_1, \ldots, Y_n) \overset{d}{=} (Z_1 / (\Sigma_{i=1}^n h(Z_i) + W_2), \ldots, Z_n / (\Sigma_{i=1}^n h(Z_i) + W_2)),
\]

\( h \geq 0 \) and \( W_2 > 0 \) a.s., then \( W_2 \overset{st}{=} W_1 \) implies \( U \overset{st}{=} V \). This follows from Theorem 1 because in Model A one can take

\[
\phi(z, w) = [\Sigma_{i=1}^n h(Z_i) + w]^{-1}
\]
\[
\delta_1(z, w) = \delta_2(z, w) = 0.
\]

We note in passing that it is easy to show that \( U \overset{st}{=} V \) also for spacings vectors constructed from \( X \) and \( Y \) which satisfy the following model:
Model B. There exist a random vector \( Z = (Z_1, \ldots, Z_n) \), a random vector \( W \) and Borel-measurable functions \( \phi_1, \phi_2, \delta_1 \) and \( \delta_2 \) such that
\[
(X_1, \ldots, X_n) \overset{d}{=} (Z_1\phi_1(Z,W) + \delta_1(Z,W), \ldots, Z_n\phi_1(Z,W) + \delta_1(Z,W)),
\]
\[
(Y_1, \ldots, Y_n) \overset{d}{=} (Z_1\phi_2(Z,W) + \delta_2(Z,W), \ldots, Z_n\phi_2(Z,W) + \delta_2(Z,W)).
\]
Moreover, the following conditions are satisfied:

B1. \( Z_1, \ldots, Z_n \) are i.i.d. and \( W \) and \( Z \) are independent.

B2. \( \phi_1(z,w), \phi_2(z,w), \delta_1(z,w), \) and \( \delta_2(z,w) \) are permutations symmetric functions of \( Z_1, \ldots, Z_n \) for every fixed \( w \), and \( \phi_1 \geq \phi_2 > 0 \) over the support of \( (Z,W) \).

The main difference between Models A and B is that in Model B we have two functions \( \phi_1 \) and \( \phi_2 \) compared to the single function \( \phi \) of Model A. But, in Model B, \( \phi_1 \) and \( \phi_2 \) are not required to be monotone.

We end this section by showing that Model 4.2 (unlike Model 4.1) of Shaked and Tong (1985), which involves positive dependence by mixture, does not necessarily imply the basic relation (2.1).

A random vector \( Y \) is called positively dependent by mixture (PDM) if there exists a random vector \( W \) such that, given \( W = w \), \( Y_1, \ldots, Y_n \) are conditionally i.i.d. Shaked (1977) and Shaked and Tong (1985) showed that, in some respects, a PDM random vector \( Y \) is more positively dependent than a random vector \( X \) of i.i.d. random variables where \( X \) and \( Y \) have the same marginals. One can expect that the spacings vectors \( U \) and \( V \), corresponding to \( X \) and \( Y \) satisfy \( U \preceq V \). The following example shows that this is not necessarily the case.

**Example.** Let \( Y_1 \) and \( Y_2 \) have the joint probabilities

<table>
<thead>
<tr>
<th>( y_1 )</th>
<th>( y_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \frac{1}{6} )</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>( \frac{1}{6} )</td>
</tr>
</tbody>
</table>

and let \( X_1 \) and \( X_2 \) be i.i.d. such that \( X_i \overset{d}{=} Y_i, i = 1, 2 \), that is \( P[X_1 = 1] = P[X_1 = 2] = P[X_1 = 3] = \frac{1}{3} \). Then \( P[V_1 \geq 2] = P[Y_2 = 0] = \frac{1}{3} \) whereas \( P[U_1 \geq 2] = P[X_2 = 2] = \frac{3}{4} \). Hence it is not true that \( U \overset{st}{=} V \) although \( (Y_1, Y_2) \) is PDM.

It follows that if \( X \) and \( Y \) satisfy Model 4.2 of Shaked and Tong (1985) then it is not necessarily true that the corresponding spacings satisfy \( U \overset{st}{=} V \).

3. Examples. In this section we describe some examples of well-known distributions for which the results of Section 2 apply.

(a) **Exchangeable normal variables.** Let \( X \) be a multivariate normal random vector with means \( \mu_x \), variances \( \sigma_x^2 \) and correlations \( \rho \); let \( Y \) be another multivariate normal random vector with means \( \mu_y \), variances \( \sigma_y^2 \) and correlations \( \eta \). If \( 0 \leq \rho < \eta \leq 1 \) then \( U \overset{st}{=} V \).

This follows from (2.7) with \( \alpha = 2 \) where \( Z_1, \ldots, Z_n \) and \( W \) are i.i.d. normal random variables with mean 0 and variance \( \sigma^2 \). Note that adding \( \mu_y \) to all \( X_i \)'s and \( \mu_x \) to all \( Y_i \)'s does not change the distributions of \( U \) and \( V \).

(b) **Multivariate Cauchy and stable variables.** It is shown in Shaked and Tong (1985) that
some exchangeable stable random vectors $X$ and $Y$ have the representation (2.7). Hence Theorem 1 applies.

(c) **Multivariate Dirichlet and beta variables.** Let $X$ have the distribution defined by

$$(X_1, \ldots, X_n) = \left(\frac{Z_1}{\Sigma_{i=1}^n Z_i + W}, \ldots, \frac{Z_n}{\Sigma_{i=1}^n Z_i + W}\right)$$

where $Z_1, \ldots, Z_n$ are i.i.d. gamma random variables (for Dirichlet) or i.i.d. chi-squared random variables (for multivariate beta), $W$ is a gamma random variable or a chi-squared random variable (with the same shape parameter but possibly with different scale parameter) and $Z$ and $W$ are independent. In this case, as is shown in Section 2, a partial ordering of the spacings can be obtained via the value of the shape parameter of $W$.

Note that in this case $X_1, \ldots, X_n$ are not positively dependent. Actually they are negatively correlated. The result of Theorem 1 can be interpreted here by saying that the less negatively dependent are $X_1, \ldots, X_n$ the smaller stochastically are the corresponding spacings.

4. Applications

4.1. **Goodness of fit tests.** Let $Z_1, \ldots, Z_n$ be random variables, let $Z_{(1)}, \ldots, Z_{(n)}$ be the corresponding order statistics and let $U_i = Z_{(i+1)} - Z_{(i)}, i = 1, \ldots, n-1,$ be the corresponding spacings. Statistics like the largest spacing, the $k$th smallest spacing, partial sums of ordered spacings, etc., have been used in statistical literature to construct tests of goodness of fit and related hypotheses (see, e.g., Rao and Sobel (1980) and references there). Pyke (1965) discussed statistics of the form $\sum_{i=1}^{n-1} g(U_i)$ where $g$ is some monotone function. A general form for all the statistics mentioned above is $\sum_{i=1}^{n-1} g_i(U_i)$ where the $g_i's$ are all monotone in the same direction (see Weiss (1957)).

In most applications the $Z_i's$ are i.i.d. with a common distribution $F$, say, and one is interested to test $H_0:F=F_0$ where $F_0$ is a given distribution (which may or may not depend on some unknown parameters). The hypothesis $H_0$ is then rejected if $\sum_{i=1}^{n-1} g_i(U_i)$ is large; tables of critical values have been prepared for various choices of the $g_i's$.

In some practical situations it may happen that the assumption of independence of the observations is not valid. For example, a random shift of all the observations combined with a change of scale may transform the $Z_i's$ into dependent random variables with the same (or with different) marginals as the original $Z_i's$ (we denote these dependent random variables then by $Y_i's$). For example, assume that the $Z_i's$ are i.i.d. normal random variables with mean $\mu$ and variance $\sigma^2$. Define

$$Y_i = (1-p)^{1/2}Z_i + p^{1/2}Z, i = 1, \ldots, n.$$ 

where $Z$ is a random shift, independent of $Z_1, \ldots, Z_n$ and having a normal distribution with mean $\lambda$ and variance $\sigma^2$. Then each $Y_i$ is a normal random variable with mean $(1-p)^{1/2} \mu + p^{1/2} \lambda$ and variance $\sigma^2$, but now the $Y_i's$ are not independent. In most applications $\mu = \lambda = 0$ (the condition $\lambda = 0$ means that, on the average, the random shift is zero), so that the $Y_i's$ have the same common marginal distribution as the original $Z_i's$, but they are not independent. It follows from Theorem 1, then, that if one uses the test statistic $\sum_{i=1}^{n-1} g_i(U_i)$, where the $g_i's$ are nondecreasing, then one has a smaller probability of rejection of $H_0$ than intended. Actually, Theorem 1 shows that the more dependent the $Y_i's$ are, the smaller is the probability of rejection of $H_0$. The opposite is true if the $g_i's$ are nonincreasing.

Of course the above analysis is valid whenever the $Y_i's$ are distributed according to any multivariate stable distribution (see Section 3).

Relation (2.10) is particularly useful in this setting. To see this let $p = 0$ in (2.7) and
STOCHASTIC ORDERING OF SPACINGS

\[ \eta \in (0,1). \] Then \((X_1, \ldots, X_n) = (Z_1, \ldots, Z_n),\) i.e. \(X_1, \ldots, X_n\) are i.i.d. whereas \(Y_1, \ldots, Y_n\) are not independent. The corresponding spacings \(U\) and \(V\), constructed from \(X\) and \(Y\), respectively, satisfy (2.8) and hence (2.9) and (2.10). If \(\eta\) is known and if the test statistic is \(\sum_{i=1}^{n} \lambda_i U_i\) then the critical values for testing \(H_0\) (mentioned above) can be obtained from existing tables by multiplying the tabulated values by \((1-\eta)^{1/\alpha}\) [recall that here we take \(p = 0\)].

We remark in passing that comments which are similar to the above apply to any statistic which is a monotone function of the spacings. In particular, they apply to the statistics discussed in del Pino (1979) which are monotone functions of the \(k\)-spacings: \(Z_{(k+1)} - Z_{(1)}\), \(\ldots\), \(Z_{(n)} - Z_{(n-k)}\).

4.2. Tests for independence. The discussion in Section 4.1 shows that if observations are dependent in the sense Model A instead of being independent then the significance levels of many tests, which use these observations, are different than the desired ones. In particular, if the \(g_i\)'s are nondecreasing then the probability of rejection decreases as the observations become more positively dependent.

One way to interpret this discussion is to observe that the statistics \(\sum_{i=1}^{n} g_i(U_i)\) actually yield unbiased tests for the hypothesis which claims that the random variables are independent versus the alternative which claims, for example, that they are positive dependent in the sense (2.2) with \(a \geq c\) and \(d \geq b = 0\).

We note in passing that the resulting tests are not necessarily optimal in any sense. We do not try to derive here any optimality property for any test. We remark, however, that one advantage of the above tests is that existing tables of critical values of statistics of the form \(\sum_{i=1}^{n} g_i(U_i)\) and existing results about their asymptotic distributions (see Pyke (1965) and references there) may be applied for testing the hypothesis of independence mentioned above.

4.3. Empirical distributions and quantile function estimates. Let \(Z_1, \ldots, Z_n\) be identically distributed random variables with a common distribution \(F\). Let \(Z_{(1)} \leq \ldots \leq Z_{(n)}\) be the corresponding order statistics. The empirical distribution function, \(\hat{F}\), is (denoting \(Z_{(n+1)} = \infty\))

\[ \hat{F}(z) = 0 \text{ if } z < Z_{(1)} \]
\[ = i/n \text{ if } z \in (Z_{(i)}, Z_{(i+1)}), \quad i = 1, \ldots, n, \]

that is, \(\hat{F}\) is constant on intervals whose length are the spacings associated with the \(Z_i\)'s.

In most applications the \(Z_i\)'s are independent and then \(\hat{F}\) is an estimate of \(F\) whose properties are well known. However, in some applications the \(Z_i\)'s are dependent in the sense of Model A [then we denote them by \(Y_1, \ldots, Y_n\)] although marginally \(Y_i \overset{M}{=} Z_i\). In that case \(\hat{F}\) is not necessarily an unbiased estimator of \(F\). By Theorem 1, the more positively dependent the \(Y_i\)'s are, the shorter (stochastically) the corresponding spacings are and thus, the shorter (stochastically) the range of the support of \(\hat{F}\) is. Geometrically, if \(X\) and \(Y\) satisfy (2.1) then the graph of the \(\hat{F}\) based on the \(Y_i\)'s will be (stochastically) steeper than the graph of the \(\hat{F}\) based on the \(X_i\)'s.

Thus, various statistics which are functions of \(\hat{F}\) can be compared stochastically. This is the case if these statistics are nondecreasing functions of the underlying spacings. For example, various measures of dispersion (such as the range, the interquartile range, the sample variance, etc.) computed from \(\hat{F}\) based on the \(X_i\)'s are stochastically larger than the same computed from \(\hat{F}\) based on the \(Y_i\)'s.
The inverse of $F$, 

$$Q(u) = \inf\{x : F(x) \geq u\}$$

and its density $q$ (if it exists), are called, respectively, the quantile function and the quantile-density function. Parzen (1979) discusses various estimators $\hat{Q}$ and $\hat{q}$ of $Q$ and $q$.

Graphically, one of the estimators, $\hat{Q}$, is obtained by “inverting” $\hat{F}$ (that is, flipping the graph of $\hat{F}$ around the main diagonal of the bivariate plane). Parzen (1979) also suggests various “sensible” estimates of $q$ which are obtained by differentiating “smooth” versions of $\hat{Q}$. For example, one estimator of $q$ is given by

$$\hat{q}(u) = n(Z_{(i)} - Z_{(i-1)}) \text{ for } u \in \left(\frac{i-1}{n}, \frac{i}{n}\right), i = 2, \ldots, n.$$  

The comments about the influence of positive dependence on $\hat{Q}$ and $\hat{q}$ are similar to the ones made above about $\hat{F}$. Various monotone functionals of $\hat{Q}$ are discussed in Parzen (1979). Thus, one can stochastically compare various statistics based on a $\hat{Q}$ which was constructed from $X_i$’s to similar statistics based on a $\hat{Q}$ which was constructed from $Y_i$’s where $X$ and $Y$ satisfy (2.1).

### 4.4. Tests for outliers.

For $i = 1, \ldots, n$, let $Z_i$ be a normal random variable with mean $\mu_i$ and variance $\sigma^2$. Consider the null hypothesis $H_0: \mu_1 = \ldots = \mu_n$ and the following possible alternatives which state that one or $k$ of the $Z_i$’s are outliers:

- $A: \mu_1 = \ldots = \mu_{i-1} = \mu_{i+1} = \ldots = \mu_n < \mu_i$ for some $i \in 1, \ldots, n$ (one of the $Z_i$’s is an outlier caused by a slippage to the right),
- $A': \mu_1 = \ldots = \mu_{i-1} = \mu_{i+1} = \ldots = \mu_n > \mu_i$ for some $i \in 1, \ldots, n$ (one of the $Z_i$’s slipped to the left),
- $A'': \mu_1 = \ldots = \mu_{i-1} = \mu_{i+1} = \ldots = \mu_n \neq \mu_i$ for some $i \in 1, \ldots, n$ (one of the $Z_i$’s is an outlier),

$B_k: n-k \mu_i$’s are equal to an unknown $\mu$ and the other $\mu_i$’s are larger than $\mu$ (there are $k$ outliers caused by slippages to the right). Similarly $B'_k$ and $B''_k$ can be defined.

Various tests have been proposed for testing these and similar alternatives when the $Z_i$’s are assumed to be independent (Barnett and Lewis (1978, pp. 89–115)). For example, if $\sigma^2$ is known then one can test $A$ [respectively, $B_k$] by rejecting $H_0$ if

$$T_A = \sigma^{-1}\phi_1(Z) = \sigma^{-1}(Z_{(n)} - \bar{Z}) > c$$

for some $c$ [respectively,

$$T_{B_k} = \sigma^{-1}\phi_k(Z) = \sigma^{-1}(Z_{(n)} + \ldots + Z_{(n-k+1)} - k\bar{Z}) > c$$

for some $c$];

here $\bar{Z} = n^{-1}(Z_1 + \ldots + Z_n)$. Similarly, $A'$ [respectively $B'_k$] can be tested by rejecting $H_0$ when

$$T_{A'} = \sigma^{-1}\phi'_1(Z) = \sigma^{-1}(\bar{Z} - Z_{(1)}) > c$$

for some $c$ [respectively,

$$T_{B'_k} = \sigma^{-1}\phi'_k(Z) = \sigma^{-1}(k\bar{Z} - Z_{(1)} - \ldots - Z_{(k)}) > c$$

for some $c$].

The two-sided alternative $A''$ may be tested by rejecting $H_0$ when

$$T_{A''} = \sigma^{-1}\psi(Z) = \sigma^{-1}\max (Z_{(n)} - \bar{Z}, \bar{Z} - Z_{(1)}) > c$$

for some $c$ and $B''_k$ may be tested by rejecting $H_0$ when

$$T_{B''_k} = \sigma^{-2}\varphi(Z) = \sigma^{-2}\sum_{i=1}^{n} (Z_i - \bar{Z})^2 > c$$

for some $c$.

(see Dixon (1950, p. 490)) or by rejecting $H_0$ when
\[ \tilde{T}^{\alpha}_{B} = \sigma^{-1} \phi(X) = \sigma^{-1} (Z(\alpha) - Z(1)) > c \text{ for some } c. \]

If \( \sigma^2 \) is unknown, but an independent estimate \( S_0^2 \) of \( \sigma^2 \) is available, then one can test the various alternatives by replacing \( \sigma \) by \( S_0 \) in the above statistics (see details in Barnett and Lewis (1978, pp. 89–115)).

Note that all the above test statistics are nondecreasing functions of the spacings (for example \( Z_{(n)} \tilde{Z} = n^{-1} \sum_{i=1}^{n-1} (Z_{(i)} - Z_{(i)}) = n^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} U_j \)). It follows from Theorem 1 that if the observations are not independent but instead that a random shift and a rescaling in the sense Model A have been applied to the \( Z_i \)'s [denote them then by \( Y_i \)'s] leaving the marginals unchanged, then the significance level of each of the above tests may be smaller than the desired one.

Of course, the same analysis applies also to \( Z_i \)'s which have distributions other than normal (see Section 3 for examples).

REFERENCES