Higher Order Cumulants and Tchebyshev-Markov Bounds for P-values in Distribution-Free Matched-Pairs Tests

Hari K. Iyer
*Colorado State University, Fort Collins, Colorado, USA*

Dominic F. Vecchia
*National Institute of Standards and Technology, Boulder, Colorado, USA*

Paul W. Mielke, Jr.
*Colorado State University, Fort Collins, USA*

Abstract: We consider a general class of Permutation Tests for Matched-Pairs (PTMP) which includes the Wilcoxon signed-ranks test, the sign test, the permutation version of the paired \( t \)-test, and many other known distribution-free procedures for paired comparisons. For some special cases, algorithms are available for efficient computation of exact significance probabilities for these tests. In general, however, only approximate tests are feasible in most applications. A common approximation is based on the first few exact moments of the test statistic. In this paper we describe a technique for deriving exact moments of any desired order for the permutation null distribution of a PTMP statistic. We use this technique to obtain explicit, efficient computational formulas for the first seven cumulants of the general PTMP statistic. For special cases of this statistic we give explicit formulas for computing cumulants of any order. Finally we demonstrate how these higher order cumulants can be used to bound the \( P \)-values of tests using Tchebyshev-Markov inequalities.

Key Words: Nonparametric tests; Paired \( t \)-test; Pearson distribution; Permutation test; Rank tests; Sign test; Tchebyshev-Markov inequalities; Wilcoxon signed-ranks test;

1. INTRODUCTION

Matched-pairs designs are among the most effective and widely applied statistical methods in scientific studies. Typically, matched-pairs data are analyzed using the ordinary paired \( t \)-test, or one of its nonparametric analogs such as the sign test and Wilcoxon signed-ranks test. The latter tests, which are based on permuted outcomes of a statistic, nevertheless share with the \( t \)-test the practical advantage of computational simplicity, owing to the choice of especially simple scores (like ranks) assigned to magnitudes of raw matched-pairs differences. The more general form of the Wilcoxon test, using an arbitrary monotonic trans-
formation of the absolute differences, presents well known computational challenges when samples sizes are large. As a result, a variety of approaches for both exact and approximate calculation of $P$-values have been proposed by various authors.

In this article we consider a very general class of Permutation Tests for Matched Pairs (PTMP) introduced by Mielke and Berry (1982). The class comprises, as special cases, all of the common nonparametric tests, the permutation version of the paired $t$-test, and many new procedures for matched-pairs analysis. We show how to compute exact lower and upper bounds for the $P$-value for any matched-pairs test that uses a PTMP statistic. Such bounds are based on generalized Tchebyshev-Markov inequalities that make use of the first $k$ moments or cumulants of a test statistic. The bounds can improve substantially as the number of available moments increases. As a first step we therefore derive the first seven exact cumulants for the general class of PTMP statistics. This generalizes the work of Mielke and Berry (1982), who presented the first three exact cumulants.

In principle, the method presented here can be used to obtain higher order cumulants if desired, but the computations become increasingly cumbersome. For special cases of the PTMP statistics, however, it is feasible to calculate any desired number of cumulants based on explicit formulas which we present.

Calculation of exact significance levels for permutation tests is impractical except for moderately-sized samples. Several authors have proposed efficient algorithms designed to extend the feasible sample size for computing exact $P$-values for some of the matched-pairs statistics that are special cases of the PTMP class. For instance, Pagano and Tritchler (1983) proposed a polynomial time algorithm which is applicable when the test statistics are linear in the observations or in functions of the observations, such as ranks. John and Robinson (1983) also described an algorithm for calculating exact significance levels for a subclass of PTMP statistics which includes the test based on the mean, linear rank tests, and locally most powerful permutation tests. Baker and Tilbury (1993) presented an algorithm for exact calculations of the permutation version of the usual matched-pairs $t$-test. Exact computations for robust tests based on the median were presented by Welch (1987) and, for those based on trimmed-means, were presented by Welch and Gutierrez (1988) and Spino and Pagano (1991).

When the sample size is too large for calculating exact significance levels of PTMP statistics, various approximate solutions are possible. Dwass (1957) suggested the use of a Monte-Carlo sampling approach, whereas Mielke and Berry (1982) calculate approximate $P$-values using a Pearson Type-III approximation based on the first three exact cumulants of the permutation distribution. Either of these approaches can be applied to any PTMP statistic. Robinson (1982) considered saddlepoint approximations for computing $P$-values for a subclass of the PTMP statistics. While the Monte-Carlo approach does provide "statistical" bounds for the $P$-value, neither the Pearson nor the saddlepoint method is able to provide an estimate of the error in the approximate $P$-value.

In Section 2 we define the general class of Permutation Tests for Matched-Pairs (PTMP), and describe a procedure for computing the exact cumulants of their null permutation distribution. It is shown that, by judiciously storing intermediate results of calculations, the cumulants can be computed in $O(n^2)$ operations, where $n$ is the number of matched-pairs in the sample. In fact, the same order of computations can be maintained when calculating any higher-order cumulant of the statistics, albeit with increasing storage requirements. Thus,
it may be feasible to compute cumulants of arbitrarily high orders if they are needed for a particular application. For certain special cases, including the classical procedures, very simple formulas for the seven cumulants are available. In particular, we present simpler expressions for cumulants of certain rank-based procedures, including two new simple rank tests considered by Mielke and Berry (1982). The latter procedures may be viewed as generalized versions of the Wilcoxon signed–ranks test. In Section 3 we apply these procedures to some examples and illustrate the performance of the Tchebyshev–Markov bounds for the significance probabilities by comparing them with exact values where feasible.

2. A CLASS OF MATCHED–PAIRS STATISTICS

Let \( (U_1, V_1), \ldots, (U_n, V_n) \) denote the \( n \) pairs of observable responses from \( n \) subjects. Define \( d_i = |U_i - V_i| \) and \( Z_i = \text{sign}(U_i - V_i) \). Then, under the null hypothesis of no treatment differences and conditional on \( d_1, \ldots, d_n \), the \( Z_i \) are independent and \( P(Z_i = 1) = P(Z_i = -1) = 1/2 \). In the permutation version of the usual matched–pairs \( t \)-test, putting \( X_i = d_i Z_i \), the test statistic is generally taken to be

\[
t(Z) = \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} d_i Z_i
\]

since, regardless of the outcome \( Z = (Z_1, \ldots, Z_n) \), there is a monotonic relationship between \( \sum_{i=1}^{n} X_i \) and the usual \( t \)-statistic

\[
\frac{X}{s_X}
\]

where

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

and

\[
s_X = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n(n-1)}}.
\]

In view of the fact that

\[
\sum_{i<j}(X_i - X_j)^2 = n \sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2
\]

and that \( \sum_{i=1}^{n} X_i^2 \) is invariant to the particular outcome of \( Z_1, \ldots, Z_n \), it follows that there is a monotonic decreasing relationship between \( \sum_{i<j}(X_i - X_j)^2 \) and \( (\sum_{i=1}^{n} X_i)^2 \). Consequently, for a two-sided test of the null hypothesis of no treatment differences, one could use as a test statistic

\[
\delta = \delta(Z) = \frac{1}{\binom{n}{2}} \sum_{i<j}(X_i - X_j)^2 = \frac{1}{\binom{n}{2}} \sum_{i<j}(d_i Z_i - d_j Z_j)^2,
\]

the average of all pairwise squared differences among \( X_1, \ldots, X_n \). In this case, small values of \( \delta \) would correspond to large values of \( (\sum_{i=1}^{n} X_i)^2 \) and would indicate a real difference between the two treatments. The quantity \( \delta \) is obviously a measure of spread of the values \( X_1, \ldots, X_n \). In fact, identity (1) demonstrates that \( \delta \) is proportional to the variance of \( X_1, \ldots, X_n \). Hence,
the permutation version of the \( t \)-test of the null hypothesis of no treatment differences can be carried out based on the permutation distribution of \( \delta \) in (2) with small realized values of \( \delta \) indicating rejection of the null hypothesis.

Intuitively, a small value for the realized \( \delta \), relative to all possible realizations of \( \delta(Z) \) based on all possible sign changes, would suggest that the differences between observations within each pair are “more alike” in the realized data than in most of the permuted outcomes. This in turn would lead us to believe that the observed superiority, if any, of one treatment over another, is inconsistent with the hypothesis of random assignment.

When viewed in this manner, it becomes clear that any meaningful way of quantifying the spread of \( X_1, \ldots, X_n \) would yield a sensible test statistic. In particular we may use any measure of separation \( \Delta(X_i, X_j) \) in place of the specific choice \((X_i - X_j)^2\), and we would still obtain an intuitively meaningful test statistic for the null hypothesis of no treatment differences. The measure \( \Delta(X_i, X_j) \) must satisfy the following key properties:

1. \( \Delta(X_i, X_i) = 0 \)
2. \( \Delta(X_i, X_j) = \Delta(X_j, X_i) \)
3. \( \Delta(X_i, X_j) = \Delta(-X_i, -X_j) \).

These properties are automatically satisfied for the specification \( \Delta(X_i, X_j) = (X_i - X_j)^2 \). Note that \( \Delta \) is not required to be a metric. In fact, the choice \( \Delta(X_i, X_j) = (X_i - X_j)^2 \) does not yield a metric as the triangle inequality is not satisfied. Yet, the specific test obtained with this particular choice of \( \Delta \) gives us the usual permutation version of the \( t \)-test, which is known to be optimal, at least for Normal populations.

In their definition of PTMP statistics, Mielke and Berry (1982) used the particular choice

\[
\Delta(X_i, X_j) = |X_i - X_j|^\nu
\]  

(3)

where \( \nu > 0 \) is called the distance exponent. The choice \( \nu = 2 \) corresponds to the usual \( t \)-statistic. The choice \( \nu = 1 \) makes \( \Delta \) a metric, and Monte Carlo power comparisons for some of these tests involving nonnormal distributions were reported by Mielke and Berry (1982).

One may further generalize the statistic \( \delta \) by replacing \( d_1, \ldots, d_n \) by suitably transformed positive values \( a_1, \ldots, a_n \), often called “scores.” Letting \( X_i = a_i Z_i \) and defining

\[
\delta(Z) = \frac{1}{\binom{n}{2}} \sum_{i<j} \Delta(X_i, X_j)
\]  

(4)

we are led to a very general class of matched-pairs test statistics. We will refer to these as PTMP (Permutation Tests for Matched Pairs) statistics.

Mielke and Berry (1982) considered the following systems of scores:

(a) \( a_i = 1/2 \).

(b) \( a_i = r_i \), where \( r_1, \ldots, r_n \) are the ranks of \( d_1, \ldots, d_n \).

(c) \( a_i = r_i^2 \), where \( r_i \) are the ranks as in (b).
They pointed out that choices (a) and (b) above, with \( \Delta(X_i, X_j) = (X_i - X_j)^2 \), respectively correspond to two-sided versions of the sign-test and the Wilcoxon's signed-ranks test, and that the choice \( a_i = r_i \), \( \Delta(X_i, X_j) = |X_i - X_j| \) leads to a new class of tests which they demonstrated to be never much worse than, but at times much better than, either the sign test or the Wilcoxon's signed-ranks test for a variety of nonnormal distributions. Brockwell and Mielke (1985) presented asymptotic distributions of \( \delta(Z) \) for four special cases discussed by Mielke and Berry (1982).

Critical regions for tests based on the PTMP statistics given in (1) correspond to small values of \( \delta(Z) \). When \( n \) is small, the significance level associated with the realized value of \( \delta(Z) \) may be calculated exactly by enumerating the permutation distribution, but for \( n \) large, the significance level is often approximated using moment methods or Monte-Carlo methods. Mielke and Berry (1982) presented formulas for computing the first three moments of \( \delta(Z) \). To obtain significance levels they used a three-moment approximation for the null permutation distribution of \( \delta(Z) \) by a Pearson type III density function. Suppose \( \delta_0 \) is the observed value of \( \delta(Z) \), and \( \mu_\delta \), \( \sigma_\delta \), and \( \gamma_\delta \) are the mean, standard deviation, and skewness coefficient of the distribution of \( \delta(Z) \), respectively. Then the \( P \)-value for the matched-pairs test is approximated by

\[
p = P\{\delta(Z) \leq \delta_0\} \approx \int_{-\infty}^{T_0} f(y)dy
\]

where

\[
f(y) = \frac{(-2/\gamma)^{4/\gamma^2}}{\Gamma(4/\gamma^2)} \left[ -\frac{(2 + y\gamma)/\gamma}{(4-\gamma^2)/\gamma^2} e^{-2(y+\gamma)/\gamma^2} ,
\right.
\]

\[-\infty < y < -2/\gamma, \quad T_0 = (\delta_0 - \mu_\delta)/\sigma_\delta, \quad \gamma = \gamma_\delta \leq -0.001 \]

(a \( P \)-value approximation based on the normal distribution is reported if \( \gamma_\delta > -0.001 \)). Computational algorithms for both the exact \( P \)-value and Pearson approximation are presented in Berry and Mielke (1985).

2.1 Exact Moments

Here we discuss a general approach which, in principle, may be used for obtaining cumulants of any order for PTMP statistics. In particular, we present formulas for calculating the first seven cumulants of \( \delta \). The technique used to obtain the formulas is illustrated for the fourth cumulant; higher order cumulants may be obtained in a similar fashion, the only difference being the complexity of the expressions involved.

For notational simplicity, we consider the centered and scaled statistic \( \gamma \) given by

\[
\gamma = n(n-1)[\delta - E(\delta)] \tag{5}
\]

where

\[
E(\delta) = \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (a_{ij} + b_{ij})
\]

with \( a_{ij} \) and \( b_{ij} \) defined by

\[
b_{ij} = \Delta(a_i, a_j) \quad \text{and} \quad a_{ij} = \begin{cases} \Delta(a_i, a_j) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}
\]
Letting $\mu_m(Y)$ and $\kappa_m(Y)$ denote, respectively, the $m$th central moment and $m$th cumulant of the random variable $Y$, we note that $\mu_1(\gamma) = \kappa_1(\gamma) = 0$ and the remaining moments and cumulants of $\gamma$ are related to those of $\delta$ by

$$\mu_m(\gamma) = n^m(n-1)^m \mu_m(\delta) \quad m = 2, 3, \ldots$$

and

$$\kappa_m(\gamma) = n^m(n-1)^m \kappa_m(\delta) \quad m = 2, 3, \ldots$$

Hence it suffices to calculate the cumulants of $\gamma$.

The second and third cumulants of $\gamma$ follow from Mielke and Berry (1982):

$$\kappa_2(\gamma) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (a_{ij} - b_{ij})^2$$

(6)

$$\kappa_3(\gamma) = -\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} (a_{ij} - b_{ij})(a_{ik} - b_{ik})(a_{jk} - b_{jk})$$

(7)

The general procedure for obtaining higher cumulants will be illustrated below by displaying the details involved in deriving the fourth moment of $\gamma$. The following key identity significantly simplifies the calculation of higher moments of $\gamma$ (and thus higher moments of $\delta$):

$$\gamma = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} Z_i Z_j$$

where

$$c_{ij} = \frac{1}{2} (b_{ij} - a_{ij}).$$

(8)

Since $\mu_1(\gamma) = 0$, it follows that the fourth central moment of $\gamma$ is given by

$$\mu_4(\gamma) = \sum_{i_1, i_2, i_3, i_4, i_5, i_6, i_7, i_8} c_{i_1, i_2, i_3, i_4, i_5, i_6, i_7, i_8} E(Z_{i_1} Z_{i_2} Z_{i_3} Z_{i_4} Z_{i_5} Z_{i_6} Z_{i_7} Z_{i_8})$$

Recalling that $Z_1, \ldots, Z_n$ are independently distributed under the null hypothesis, with $P(Z_i = 1) = P(Z_i = -1) = 1/2$, we observe that $E(Z_i^m) = 0$ for all odd integers $m$. Thus, the only nonzero expectations in the summation above are those in which each subscript appears an even number of times. Since $E(Z_i^m) = 1$ for all even integers $m$ we obtain, after some algebra

$$\mu_4(\gamma) = 8 \sum_{i,j} c_{ij}^4 + 48 \sum_{i,j,k} c_{ij}^2 c_{ik}^2 + 12 \sum_{i,j,k,l} c_{ij}^2 c_{kl}^2 + 48 \sum_{i,j,k,l} c_{ij} c_{ik} c_{jl} c_{kl}$$

(9)

where the symbol $\sum^\#$ denotes a restricted multiple summation, where no two indices (each running from 1 to $n$) are permitted to be equal.

2.2 Efficient Computational Formulas

The expression (9) for $\mu_4(\gamma)$ is not well-suited for computer implementation, but efficient computational formulas may be derived by straightforward algebraic manipulations of the
restricted sums. By using an approach demonstrated by Siemiettycki (1978) and also by Vecchia and Iyer (1989), one may obtain the more efficient formula

\[ \mu_4(\gamma) = 32 \sum_i \sum_j c_{ij}^4 - 96 \sum_i \left( \sum_j c_{ij}^2 \right)^2 + 12 \left( \sum_i \sum_j c_{ij}^2 \right)^2 + 48 \sum_i \sum_j \left( \sum_k c_{ik} c_{jk} \right)^2 \]

Combining this expression with equation (6) we see that the fourth cumulant of \( \gamma \) is given by

\[ \kappa_4(\gamma) = \mu_4(\gamma) - 3\mu_2^2(\gamma) = 32 \sum_i \sum_j c_{ij}^4 - 96 \sum_i \left( \sum_j c_{ij}^2 \right)^2 + 48 \sum_i \sum_j \left( \sum_k c_{ik} c_{jk} \right)^2 \] (10)

As is often true, the expression for the cumulant is simpler than that for the corresponding moment. For this reason, the remaining discussion is restricted to presentation of the cumulants of \( \gamma \).

Formulas for cumulants of \( \gamma \) may be simplified further for improved computational speed, albeit at the expense of additional storage. For the \( n \times n \) matrix \( C = (c_{ij}) \) based on equation (8), we define the matrices

\[ G = (g_{ij}) = C^2 \quad \text{and} \quad H = (h_{ij}) = G^2 \]

The expression (10) for the fourth cumulant of \( \gamma \), for example, may now be written as

\[ \kappa_4(\gamma) = 32 \sum_i \sum_j c_{ij}^4 - 96 \sum_i g_{ii}^2 + 48 \sum_i h_{ii} \]

Expressions for the first seven cumulants of \( \gamma \) are listed below. The formulas were verified by comparing the results to those obtained by complete enumeration of the permutation distribution for several small data sets.
\[ \kappa_1(\gamma) = 0 \]
\[ \kappa_2(\gamma) = 2 \sum_{i=1}^{n} g_{ii} \]
\[ \kappa_3(\gamma) = 8 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} g_{ij} \]
\[ \kappa_4(\gamma) = 32 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^4 - 96 \sum_{i=1}^{n} g_{ii}^2 + 48 \sum_{i=1}^{n} h_{ii} \]
\[ \kappa_5(\gamma) = 1280 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^3 g_{ij} - 1920 \sum_{i=1}^{n} g_{ii} \sum_{j=1}^{n} c_{ij} g_{ij} + 384 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} h_{ij} \]
\[ \kappa_6(\gamma) = 8192 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^6 - 30720 \sum_{i=1}^{n} g_{ii} \sum_{j=1}^{n} c_{ij}^4 - 7680 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^2 \sum_{k=1}^{n} c_{ik} c_{jk}^2 \]
\[ + 15360 \sum_{i=1}^{n} g_{ii}^3 + 15360 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^3 \sum_{k=1}^{n} c_{ik} g_{jk} + 11520 \sum_{i=1}^{n} g_{ii} \sum_{j=1}^{n} c_{ij}^2 g_{jj} \]
\[ + 23040 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^2 g_{ij}^2 - 23040 \sum_{i=1}^{n} g_{ii} h_{ii} - 11520 \sum_{i=1}^{n} \left( \sum_{j=1}^{n} c_{ij} g_{ij} \right)^2 \]
\[ + 3840 \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} h_{ij} \]
\[ \kappa_7(\gamma) = 688128 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^5 g_{ij} + 573440 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} \sum_{k=1}^{n} c_{ik}^3 c_{jk}^3 \]
\[ - 430080 \sum_{i=1}^{n} \left( \sum_{j=1}^{n} c_{ij}^4 \right) \left( \sum_{j=1}^{n} c_{ij} g_{ij} \right) - 1720320 \sum_{i=1}^{n} g_{ii} \sum_{j=1}^{n} c_{ij}^3 g_{ij} \]
\[ - 215040 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^3 \sum_{k=1}^{n} c_{ik} c_{jk} g_{kk} - 645120 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} g_{ij} \sum_{k=1}^{n} c_{ik}^2 c_{jk}^2 \]
\[ + 645120 \sum_{i=1}^{n} g_{ii}^2 \sum_{j=1}^{n} c_{ij} g_{ij} + 215040 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^3 h_{ij} \]
\[ + 322560 \sum_{i=1}^{n} g_{ii} \sum_{j=1}^{n} c_{ij}^2 \sum_{k=1}^{n} c_{jk} g_{jk} + 645120 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^2 g_{ij} \sum_{k=1}^{n} c_{ik} g_{jk} \]
\[ + 322560 \sum_{i=1}^{n} g_{ii} \sum_{j=1}^{n} c_{ij} g_{ij} g_{jj} + 215040 \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}^3 g_{ij} \]
\[ - 322560 \sum_{i=1}^{n} g_{ii} \sum_{j=1}^{n} g_{ij} \sum_{k=1}^{n} c_{jk} g_{ik} - 322560 \sum_{i=1}^{n} h_{ii} \sum_{j=1}^{n} c_{ij} g_{ij} \]
\[ + 46080 \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij} \sum_{k=1}^{n} c_{ik} g_{jk} \]

8
2.3 Special Cases

The scores $a_1, a_2, \ldots, a_n$ used in the definition of the PTMP statistic in equation (4) of Section 2 may be taken to be the observed absolute values of matched-pairs differences, i.e., $d_1, d_2, \ldots, d_n$, or appropriate transformations of the differences. Using $\Delta(X_i, X_j) = |X_i - X_j|^v$, Mielke and Berry (1982) discussed four special cases of the PTMP statistic by choosing particular values of the distance exponent $v$ and transformations of observed matched-pairs differences, and presented some Monte-Carlo power comparisons. Letting $r_1, r_2, \ldots, r_n$ denote the rank order statistics of the observed absolute values of $n$ matched-pairs differences, the special cases they considered are: (1) $a_i = 1/2$, $v > 0$ arbitrary, (2) $a_i = r_i$, $v = 2$, (3) $a_i = r_i$, $v = 1$, and (4) $a_i = r_i^2$, $v = 1$. The first case is easily shown to be equivalent to the two-sided sign test, while the second case is equivalent to the two-sided Wilcoxon signed-ranks test. In each of these four special cases the cumulants of $\gamma$ simplify to polynomials in the sample size $n$. Recalling that $\kappa_1(\gamma) = 0$, the next six cumulants of $\gamma$ for the special cases are given in the Appendix.

For the particular choice $\Delta(X_i, X_j) = |X_i - X_j|^2$, cumulants of $\gamma$ of any order can be easily calculated by first relating $\gamma$ to the generalized version of the Wilcoxon signed-ranks statistic. We define the generalized Wilcoxon statistic

$$W(Z) = \sum_{i=1}^{n} a_i Z_i$$

where the $a_i$'s are arbitrary scores rather than the ranks of $d_1, d_2, \ldots, d_n$. By symmetry of the permutation distribution, all odd order cumulants of $W$ are zero, and even order cumulants are given by

$$\kappa_{2j}(W) = \frac{2^{2j}(2^{2j} - 1)B_{2j}}{2j} \sum_{i=1}^{n} a_i^{2j} \quad \text{for} \quad j = 1, 2, \ldots$$

where the $B_{2j}$ are the even-order Bernoulli numbers. Hence, moments of $W$ are readily computed and, by observing that

$$\gamma(Z) = 2 \sum_{i=1}^{n} a_i^2 - 2 \, W^2(Z),$$

moments of $\gamma$ of any order can also be easily calculated. In the examples below, we have used as many as 20 exact moments of $\gamma$ to calculate bounds for $P$-values when $v = 2$.

3. USE OF TCHEBYSHEV–MARKOV BOUNDS

Moments of a random variable can often be used to obtain bounds on certain probabilities involving the random variable. Markov’s inequality and Tchebyshev’s inequality are two well-known examples. Generalizations of these inequalities have been studied by several authors; for instance, see Shohat and Tamarkin (1943), Royden (1953), and Zelen (1954), Godwin (1955), and Mallows (1956). Zelen (1954) provided explicit expressions for generalized Tchebychef–Markov bounds for the distribution function of a random variable in terms of its first four moments. He also stated the general result for obtaining bounds for the distribution function based on higher order moments.

When it is known that a probability distribution has a support which is contained in a finite interval, sharper bounds for probabilities are available than when the support is
specified as the entire real line. In our application, it can be shown that the support of the permutation distribution of $\gamma$ is contained in a finite interval $[A, B]$. For the distance function $\Delta(X_i, X_j) = |X_i - X_j|^v$, in particular, specific values of $A$ and $B$ that are used to obtain bounds in our examples are:

$$A = \begin{cases} -2 \sum_{i=1}^{n} (n-i) a_i & \text{if } v = 1 \\ 2 \left[ \sum_{i=1}^{n} a_i^2 - \left( \sum_{i=1}^{n} a_i \right)^2 \right] & \text{if } v = 2 \end{cases}$$

$$B = \begin{cases} 2 \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (-1)^{j-i} (a_j - a_i) - (-1)^n \sum_{i=1}^{n-1} a_i}{2 \sum_{i=1}^{n} a_i^2} & \text{if } v = 1 \\ 2 \sum_{i=1}^{n} a_i^2 & \text{if } v = 2 \end{cases}$$

where, without loss of generality, we assume that the scores are ordered so that $0 \leq a_1 \leq a_2 \leq \ldots \leq a_n$. In fact, these formulas provide exact bounds on the support, with the one exception that the maximum value $B$ for $v = 2$ is conservative unless $W(Z) = \sum_{i=1}^{n} a_i Z_i = 0$ for some choice of the vector of signs $Z$.

Below, we present several examples for which we compute generalized Tchebyshev-Markov bounds for $P$-values using the first seven exact cumulants and the bounds $[A, B]$ shown above. The particular version of Tchebyshev-Markov inequalities we use are from Zelen (1954, page 380), with $A$ and $B$ as defined above. The first example uses data collected at the National Institute of Standards and Technology (NIST). The second example uses data considered by Welch (1987). The third example uses 13 published data sets to illustrate the performance of Tchebyshev-Markov Bounds and the Pearson approximation relative to exact $P$-values. For these 13 analyses we also show exact $P$-value bounds based on 20 cumulants when $v = 2$.

### 3.1 Silicon Wafer Wiring Configurations Example

As part of the certification process for a standard reference material (SRM) by the National Institute of Standards and Technology (NIST), resistivity measurements were made on doped silicon wafers using two different wiring configurations, 1 and 2. The wafers were cut from a single crystal and 6 repeat measurements (6 days) were made for each wafer using both wiring configurations. One of the objectives of the study was to test the significance of the difference between wiring configurations 1 and 2. Data in Table 1 are the observed differences between resistivity measurements on the wafers obtained using the two wiring configurations.

<table>
<thead>
<tr>
<th>Wafer No.</th>
<th>Differences (ohm.cms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Differences (ohm.cms)</td>
</tr>
<tr>
<td>17</td>
<td>-0.0108 -0.0111 -0.0062 0.0020 0.0018 0.0002</td>
</tr>
<tr>
<td>39</td>
<td>-0.0089 -0.0315c -0.0040 -0.0022 -0.0012 -0.0034</td>
</tr>
<tr>
<td>63</td>
<td>-0.0016 -0.0111 -0.0059 -0.0078 -0.0007 0.0006</td>
</tr>
<tr>
<td>103</td>
<td>-0.0050 -0.0140 -0.0048 0.0018 0.0016 0.0044</td>
</tr>
<tr>
<td>125</td>
<td>-0.0056 -0.0155 -0.0010 -0.0014 0.0003 -0.0017</td>
</tr>
</tbody>
</table>

*Measure differences between two wiring configurations.

Data are from Crystal 21565, using Probe No. 2062.

*Value was deemed an outlier by the investigators.

Suppose the individual resistivity measurements follow the model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk}$$
where $\mu$ is a constant, $\alpha_i$ is the (fixed) effect of wafer $i$, $\beta_j$ is the (fixed) effect of day $j$, $\tau_k$ is the effect of wiring configuration $k$, and $\epsilon_{ijk}$ is random error. Suppose also that the random errors are independently and identically distributed with mean zero. Then the differences

$$Y_{ij1} - Y_{ij2} = \tau_1 - \tau_2 + \epsilon_{ij1} - \epsilon_{ij2}$$

have a symmetric distribution, and the mean of this distribution is zero under the null hypothesis. Moreover, the differences are independent of one another and the assumptions required for the PTMP test to be valid are satisfied.

We computed the exact $P$-values for a two-sided test of no wiring configuration effects using different choices of distance functions and scores. We also computed lower and upper generalized Tchebyshhev bounds for the $P$-values using the first 7 exact moments of the PTMP statistics. These are given in Table 2. In each case the distance function used was of the form given in (3) and we used two different values for $v$, namely $v = 1$ and $v = 2$.

<table>
<thead>
<tr>
<th>Distance Exponent</th>
<th>Score, $a_i$</th>
<th>Exact $P$-value</th>
<th>Approx. Pearson $P$-value</th>
<th>Generalized Tchebyshhev Bounds</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v = 2$</td>
<td>$a_i = d_i$</td>
<td>$137506/2^{30} \approx 0.000128$</td>
<td>0.000392</td>
<td>0.000000</td>
<td>0.000668</td>
<td></td>
</tr>
<tr>
<td>$v = 1$</td>
<td>$a_i = d_i$</td>
<td>$182810/2^{30} \approx 0.000170$</td>
<td>0.000242</td>
<td>0.000000</td>
<td>0.000916</td>
<td></td>
</tr>
<tr>
<td>$v = 2$</td>
<td>$a_i = \text{rank}(d_i)$</td>
<td>$379144/2^{30} \approx 0.000353$</td>
<td>0.000431</td>
<td>0.000001</td>
<td>0.001677</td>
<td></td>
</tr>
<tr>
<td>$v = 1$</td>
<td>$a_i = \text{rank}(d_i)$</td>
<td>$596056/2^{30} \approx 0.000555$</td>
<td>0.000614</td>
<td>0.000001</td>
<td>0.002733</td>
<td></td>
</tr>
</tbody>
</table>

In this example, using the bounds alone, we could conclude that there are significant differences between the two wiring configurations. Also, the Pearson approximations for the $P$-values appear to be satisfactory.

**Example 2: Telephone Line Fault Rates Example**

Welch (1987) proposed a rerandomized-medians test for matched-pairs data as an alternative to the test based on the observed mean difference. He illustrated the robustness of this test to outliers using data from a matched-pairs experiment for testing a method of reducing faults on telephone lines. The data are fault rates for 14 matched-pairs of areas and are listed in Table 3.

<table>
<thead>
<tr>
<th>Pair Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test rate $r_T$</td>
<td>676</td>
<td>206</td>
<td>230</td>
<td>256</td>
<td>280</td>
<td>433</td>
<td>337</td>
<td>466</td>
<td>497</td>
<td>512</td>
<td>794</td>
<td>428</td>
<td>452</td>
<td>512</td>
</tr>
<tr>
<td>Control rate $r_C$</td>
<td>88</td>
<td>570</td>
<td>605</td>
<td>617</td>
<td>653</td>
<td>2913</td>
<td>924</td>
<td>286</td>
<td>1098</td>
<td>982</td>
<td>2346</td>
<td>232</td>
<td>615</td>
<td>519</td>
</tr>
<tr>
<td>$(1/r_T - 1/r_C) \times 10^6$</td>
<td>-988</td>
<td>310</td>
<td>269</td>
<td>229</td>
<td>204</td>
<td>197</td>
<td>189</td>
<td>-135</td>
<td>110</td>
<td>93</td>
<td>83</td>
<td>-78</td>
<td>59</td>
<td>3</td>
</tr>
</tbody>
</table>

In his analysis of these data, Welch (1987) used a reciprocal transformation of the fault rates since the unit-treatment additivity assumption appeared to be satisfied on this scale. He obtained a one-sided $P$-value of 0.0607 using his rerandomized-medians test, 0.3796 using the rerandomized-means test, and 0.3292 using the normal theory $t$-test. As pointed out by him, the rerandomized-means and the normal theory $t$-test show extreme sensitivity to the first pair, which appears to be an outlier.
We reanalyze these data using various PTMP statistics and, for consistency, we too use the reciprocal scale; see row 3 of Table 3. In each case we computed the exact $P$-value for a two-sided test of no difference between the test and the control treatments using the distance function given in (3). We used two different values for $v$, namely $v = 1$ and $v = 2$, and two different choices of scores. For each case, we computed lower and upper generalized Tchebyshev bounds for the $P$-values using the first seven exact moments of the PTMP statistics. The results are given in Table 4.

**Table 4. $P$-Values and Bounds for Telephone-Faults Experiment**

<table>
<thead>
<tr>
<th>Distance Exponent</th>
<th>Score, $a_i$</th>
<th>Exact $P$-value</th>
<th>Approx. Pearson $P$-value</th>
<th>Generalized Tchebyshev Bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v = 2$</td>
<td>$a_i = d_i$</td>
<td>$12440/2^{14} \approx 0.759277$</td>
<td>0.783536</td>
<td>0.517998 0.964129</td>
</tr>
<tr>
<td>$v = 1$</td>
<td>$a_i = d_i$</td>
<td>$964/2^{14} \approx 0.058838$</td>
<td>0.057961</td>
<td>0.013988 0.144742</td>
</tr>
<tr>
<td>$v = 2$</td>
<td>$a_i = \text{rank}(d_i)$</td>
<td>$1286/2^{14} \approx 0.078491$</td>
<td>0.069188</td>
<td>0.019029 0.166392</td>
</tr>
<tr>
<td>$v = 1$</td>
<td>$a_i = \text{rank}(d_i)$</td>
<td>$884/2^{14} \approx 0.053955$</td>
<td>0.050335</td>
<td>0.012160 0.127162</td>
</tr>
</tbody>
</table>

Our two-sided exact $P$-value for $v = 2$ using the raw scores is twice the one-sided $P$-value of Welch's rerandomized-means test which illustrates the equivalence of the two tests. It is interesting to note that using $v = 1$ seems to have made the test robust to the outlying pair. In fact, this test gives a two-sided $P$-value of 0.0588 whereas Welch's one-sided $P$-value using the rerandomized-medians test is 0.0607. The ranks based PTMP tests with $v = 1$ as well as $v = 2$ also appear robust to the outlier, as one might reasonably expect.

In this example the upper bounds for the $P$-values are not as sharp as one might like. The Pearson approximation, however, seems to be sufficiently accurate for all practical purposes.
3.3 Selected Examples from the Literature

Finally, we apply the PTMP procedures to 14 data sets described in Table 5. [Some of these data sets have been used by Welch and Gutierrez (1988) and Spino and Pagano (1991) to illustrate their algorithms for computing exact P-values of permutation tests based trimmed means.] For each data set, Table 6 provides the P-value for the t test, as well as the exact permutation P-value and corresponding three-moment Pearson approximation for PTMP tests based on both squared distance \((v = 2)\) and ordinary distance \((v = 1)\). To illustrate the performance of Tchebyshev–Markov bounds relative to exact P-values, Table 6 also provides bounds based on 7 moments for both PTMP tests \((v = 1, 2)\) in the first row of each example and, in the second row, P-value bounds when 20 moments are used.

Table 5. Example Data Sets

<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>n</th>
<th>Comparing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypnosis</td>
<td>Hollander and Wolfe (1973, page 45)</td>
<td>6</td>
<td>Changes in hypnotic susceptibility on the Stanford Profile Scales with hypnotic training</td>
</tr>
<tr>
<td>Dogs</td>
<td>Hollander and Wolfe (1973, page 32)</td>
<td>7</td>
<td>Changes in amount of insulin in dogs with stimulation of the vagus nerve</td>
</tr>
<tr>
<td>Depression</td>
<td>Hollander and Wolfe (1973, page 29)</td>
<td>9</td>
<td>Changes in depression scale factor after initiation of tranquilizer therapy</td>
</tr>
<tr>
<td>Marijuana</td>
<td>Weil, Zinberg, and Nelsen (1968)</td>
<td>9</td>
<td>Changes in performance on the Digit Symbolization Test for native subjects smoking high-marijuana and placebo cigarettes for 15 minutes</td>
</tr>
<tr>
<td>Dents</td>
<td>Hahn and Nelson (1970, page 95)</td>
<td>10</td>
<td>Depths of dents in metals measured by a standard device and a new measuring device</td>
</tr>
<tr>
<td>Shoes</td>
<td>Ryan et al. (1985, pp. 101–104)</td>
<td>10</td>
<td>Wear on two materials for the soles of boys' shoes</td>
</tr>
<tr>
<td>Faults</td>
<td>Welch (1957)</td>
<td>14</td>
<td>Reciprocals of fault rates for test and control telephone equipment</td>
</tr>
<tr>
<td>Plants</td>
<td>Fisher (1966, chap. 3)</td>
<td>15</td>
<td>Heights of self- and cross-fertilized Zea Mays plants</td>
</tr>
<tr>
<td>IQ</td>
<td>Lehmann (1975, page 149)</td>
<td>24</td>
<td>Gain in IQ for two groups of students</td>
</tr>
<tr>
<td>Grapefruit</td>
<td>Croxton et al. (1967, page 564)</td>
<td>25</td>
<td>Percentages of solids in shaded and exposed grapefruit halves</td>
</tr>
<tr>
<td>Chicks</td>
<td>Hollander and Wolfe (1973, page 41)</td>
<td>25</td>
<td>Responsivity of chick embryos to light stimulus</td>
</tr>
<tr>
<td>Presidents</td>
<td>Sommers (1996)</td>
<td>30</td>
<td>Heights of winners and losers in presidential elections from 1856 to 1992</td>
</tr>
<tr>
<td>Rain</td>
<td>Battan (1986)</td>
<td>37</td>
<td>Rainfall for seeded and nonseeded days in the Second Arizona Cloud-Seeding Experiment for 1959–1960</td>
</tr>
</tbody>
</table>
Table 6. Comparison of \( P \)-Values with Tchebyshev–Markov Bounds

<table>
<thead>
<tr>
<th>Data set</th>
<th>n</th>
<th>( t )-test</th>
<th>Pearson Exact</th>
<th>Lower</th>
<th>Upper</th>
<th>Pearson Exact</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hypnosis</td>
<td>6</td>
<td>.0583</td>
<td>.060684 .125000</td>
<td>.053825 .158247</td>
<td>.078440 .135040</td>
<td>.034515</td>
<td>.160570</td>
<td></td>
</tr>
<tr>
<td>2. Dogs</td>
<td>7</td>
<td>.0211</td>
<td>.060389 .140625</td>
<td>.008265 .089870</td>
<td>.012166 .021616</td>
<td>.015931</td>
<td>.019978</td>
<td></td>
</tr>
<tr>
<td>3. Depression</td>
<td>9</td>
<td>.0163</td>
<td>.061872 .27346</td>
<td>.002072 .040794</td>
<td>.011446 .031601</td>
<td>.015917</td>
<td>.027364</td>
<td></td>
</tr>
<tr>
<td>4. Marijuana</td>
<td>10</td>
<td>.0101</td>
<td>.067886 .092750</td>
<td>.022913 .185407</td>
<td>.077883 .058983</td>
<td>.042672</td>
<td>.160578</td>
<td></td>
</tr>
<tr>
<td>5. Dents</td>
<td>10</td>
<td>.0075</td>
<td>.064242 .078125</td>
<td>.025655 .161938</td>
<td>.082887 .089444</td>
<td>.028633</td>
<td>.187036</td>
<td></td>
</tr>
<tr>
<td>6. Fires</td>
<td>10</td>
<td>.0111</td>
<td>.063328 .056250</td>
<td>.074251 .089549</td>
<td>.074251 .089549</td>
<td>.074251</td>
<td>.089549</td>
<td></td>
</tr>
<tr>
<td>7. Shoes</td>
<td>10</td>
<td>.0085</td>
<td>.069768 .013872</td>
<td>.002505 .022309</td>
<td>.009890 .013672</td>
<td>.002427</td>
<td>.032525</td>
<td></td>
</tr>
<tr>
<td>9. Plants</td>
<td>15</td>
<td>.0497</td>
<td>.047660 .052675</td>
<td>.012483 .120309</td>
<td>.021240 .021240</td>
<td>.021240</td>
<td>.021240</td>
<td></td>
</tr>
<tr>
<td>10. IQ</td>
<td>24</td>
<td>.0000</td>
<td>.821313 .000000</td>
<td>.000000 .000000</td>
<td>.000000 .000000</td>
<td>.000000</td>
<td>.000000</td>
<td></td>
</tr>
<tr>
<td>11. Grapefruit</td>
<td>25</td>
<td>.0042</td>
<td>.004593 .004444</td>
<td>.000261 .014216</td>
<td>.005344 .005344</td>
<td>.005344</td>
<td>.005344</td>
<td></td>
</tr>
<tr>
<td>12. Chicks</td>
<td>25</td>
<td>.0000</td>
<td>.000054 .000010</td>
<td>.000000 .000000</td>
<td>.000000 .000000</td>
<td>.000000</td>
<td>.000000</td>
<td></td>
</tr>
<tr>
<td>13. Rain</td>
<td>27</td>
<td>.3777</td>
<td>.427017 .051055</td>
<td>.180937 .658091</td>
<td>.454417 .441023</td>
<td>.146037</td>
<td>.834552</td>
<td></td>
</tr>
<tr>
<td>14. Presidents</td>
<td>20</td>
<td>.1650</td>
<td>.186862 .174218</td>
<td>.052019 .345818</td>
<td>.241520 .237703</td>
<td>.079036</td>
<td>.464218</td>
<td></td>
</tr>
</tbody>
</table>

* Bounds in the second row (\( v = 2 \)) are based on 20 cumulants.

If one were to conduct tests with \( \alpha = 0.05 \), significant differences between treatments would be declared for data sets 3, 7, 11, and 12 when using \( v = 2 \). \( P \)-value bounds based on 7 moments would yield the same conclusion in each of these cases as would bounds based on 20 moments. The bounds based on 7 moments yield inconclusive results for data sets 2, 4, 5, and 9, because the value 0.05 is between the lower and the upper bound for the \( P \)-value. However, when 20 moments are used, the only inconclusive result occurs for data set 9.

If \( v = 1 \) is used, then treatments would be declared significantly different at \( \alpha = 0.05 \) for data sets 3, 7, 9, 11, and 12, and the same conclusion would be reached using the \( P \)-value bounds based on 7 moments in every case except for data set 9. Even for data set 9, the upper bound for the \( P \)-value only slightly misses the value 0.05. Inconclusive results are obtained for data sets 1, 4, 5, 8, and 9. In the case of \( v = 1 \), formulas for moments higher than order 7 are unavailable at the present time but we expect that the bounds will get better with more moments.

The Pearson approximation performed adequately in practically every case. The only data set for which the Pearson approximation would reject \( H_0 \) at \( \alpha = 0.05 \), but the exact test would not, is data set 9. Also, it is interesting to note that, in a few instances, the Pearson approximation gives a \( P \)-value that is outside the range provided by the Tchebyshev–Markov bounds; for instance, examine the \( P \)-value bounds using 20 moments, for data sets 1, 3, 12.

4. CONCLUDING REMARKS

Distribution-free tests based on a permutation or a randomization argument are not as widely used as they perhaps ought to be. One reason for this is the computational difficulty that is generally associated with the calculation of exact \( P \)-values for such tests. Standard approximation techniques such as saddle-point approximations or moment–based Pearson type approximations have the drawback that the error associated with these approximations have no general computable bounds. So, in any given problem, it is difficult to know how close or how far the approximate \( P \)-value is from the exact \( P \)-value. One generally has
to rely on past experience with such approximations. Monte-Carlo methods pose a similar problem in that they only provide probabilistic bounds for $P$-values. Although this may be satisfactory for some applications, there is the additional problem that different analysts could get different results.

In this paper we have demonstrated that, when a sufficient number of exact moments of a test statistic are available, useful exact bounds for $P$-values can be computed and decisions made using these bounds. The method is based on generalized Tchebyshev-Markov inequalities which have been available since nearly the beginning of this century. However, putting them to practical use has only become feasible due to the availability of high speed computing and specialized computer packages. We used MAPLE V to carry out many of the algebraic manipulations required while deriving the cumulant formulas. We have presented efficient computational formulas for the first seven exact cumulants of a general class of matched-pairs permutation statistics. Well known tests are obtained as special cases of the general class, including the sign test, Wilcoxon signed-ranks test, and the permutation version of the paired $t$-test. We demonstrated via examples that higher-order cumulants allow us to obtain useful bounds for the significance probability in practical problems.

Although we demonstrate the use of the Tchebyshev-Markov inequalities in the context of matched-pairs tests, their potential use is much more general. Our experience also shows that the bounds can improve significantly as more and more moments are used to compute them.

ACKNOWLEDGMENTS

References to commercial products in this article do not imply endorsement, nor do they imply that these products are the best available for the purpose. This work is a contribution of the U. S. government and is not subject to copyright in the United States. The research was supported in part by the National Science Foundation, Grant No. DMS9105745.
APPENDIX: CUMULANTS OF SPECIAL CASES

Mielke and Berry (1982) discussed four special cases of (1) by choosing particular values of the distance exponent \( v \) and transformations of observed matched-pairs differences. Letting \( r_1, r_2, \ldots, r_n \) denote the rank order statistics of the observed absolute values of \( n \) matched-pairs differences, the special cases they considered are: (1) \( a_i = 1/2, \ v > 0 \) arbitrary, (2) \( a_i = r_i, \ v = 2 \), (3) \( a_i = r_i, \ v = 1 \), and (4) \( a_i = r_i^2, \ v = 1 \).

In each of these cases the cumulants of \( \gamma \) simplify to polynomials in the sample size \( n \). The first case is easily shown to be equivalent to the two-sided version of the sign test, while Case 2 produces the two-sided version of the Wilcoxon signed-ranks test. Recalling that \( \kappa_1(\gamma) = 0 \), the next six cumulants of \( \gamma \) for the special cases are given below.

**A.1 Case 1: \( a_i = 1/2, \ v > 0 \) Arbitrary**

Suppose \( a_i = 1/2 \) for \( i = 1, \ldots, n \). We observe that \( \gamma \) does not depend on the distance exponent \( v \) in this case, and that the test is equivalent to the two-sided version of the sign test. The first seven cumulants of \( \gamma \) are

\[
\begin{align*}
\kappa_2(\gamma) &= \frac{n(n-1)}{2} \\
\kappa_3(\gamma) &= -n(n-1)(n-2) \\
\kappa_4(\gamma) &= n(n-1)(3n^2 - 15n + 17) \\
\kappa_5(\gamma) &= -4n(n-1)(n-2)(3n^2 - 21n + 31) \\
\kappa_6(\gamma) &= 4n(n-1)(15n^4 - 210n^3 + 990n^2 - 1950n + 1382) \\
\kappa_7(\gamma) &= -8n(n-1)(n-2)(45n^4 - 810n^3 + 4830n^2 - 12030n + 10922)
\end{align*}
\]

**A.2 Case 2: \( a_i = r_i, \ v = 2 \)**

When the \( a_i \) are the rank order statistics and we choose a squared distance statistic in equation (1), the resulting test is equivalent to the two-sided version of the Wilcoxon signed-ranks test (see Mielke and Berry (1982) for the identity relating \( \delta \) and the Wilcoxon statistic). The simplified cumulants of \( \gamma \) for this case, assuming there are no ties, are given by

\[
\begin{align*}
\kappa_2(\gamma) &= \frac{2n(n-1)(2n+1)(2n-1)(5n+6)(n+1)}{45} \\
\kappa_3(\gamma) &= -\frac{8}{945}n(n-1)(n-2)(2n+1)(2n-1)(2n-3)(n+1) \\
&\quad \cdot (35n^2 + 91n + 60) \\
\kappa_4(\gamma) &= \frac{16}{4725}n(n-1)(2n+1)(2n-1)(n+1) \\
&\quad \cdot (700n^7 - 3360n^6 - 2207n^5 + 20220n^4 - 2345n^3 \\
&\quad - 48540n^2 + 7872n + 42840) \\
\kappa_5(\gamma) &= -\frac{128}{31185}n(n-1)(n-2)(2n+1)(2n-1)(2n-3)(n+1)
\end{align*}
\]

16
\[ \kappa_8(\gamma) = \frac{256}{42567525} \cdot \frac{n(n-1)(2n+1)(2n-1)(n+1)}{14014000 n^{13} - 252252000 n^{12} + 1294693400 n^{11} - 747674928 n^{10} - 13283494185 n^9 + 31287579750 n^8 + 50484194150 n^7 - 21714111404 n^6 - 76771266845 n^5 + 741076663950 n^4 + 4437435600 n^3 - 1314428115168 n^2 + 67765913280 n + 952975679040) \]

\[ \kappa_7(\gamma) = -\frac{1024}{6081075} \frac{n(n-1)(n-2)(2n+1)(2n-1)(2n-3)(n+1)}{200200 n^{14} - 47647600 n^{13} + 32052000 n^{12} - 391919528 n^{11} - 409948545 n^{10} + 14524752045 n^9 + 19609270970 n^8 - 144110460494 n^7 - 40798154075 n^6 + 804420884225 n^5 + 133949507520 n^4 - 2655708902328 n^3 - 1153765704240 n^2 + 4022239142880 n + 3306115612800} \]

A.3. Case 3: \( a_i = r_i, \quad v = 1 \)

Rank tests based on the statistic (1), but using distance \((v = 1)\) rather than squared distance, may be thought of as generalized Wilcoxon tests. In their power study, Mielke and Berry (1982) concluded that the choice \( v = 1 \) often produced better power characteristics than corresponding squared distance statistics. The simplified cumulants of \( \gamma \) for simple rank scores and absolute distance, again assuming there are no ties, are given by

\[ \kappa_2(\gamma) = \frac{n^2(n-1)(n+1)}{3} \]

\[ \kappa_3(\gamma) = -\frac{2n(n-2)(4n+3)(n+1)(n-1)^2}{15} \]

\[ \kappa_4(\gamma) = \frac{8n(n-1)(n+1)(17n^5 - 102n^4 + 13n^3 + 87n^2 - 221n + 24)}{105} \]

\[ \kappa_5(\gamma) = -\frac{4}{945} \frac{n(n-1)(n-2)(n+1)}{992 n^6 - 9176 n^5 + 19885 n^4 + 5030 n^3 - 51957 n^2 + 17586 n + 23760} \]

\[ \kappa_6(\gamma) = \frac{16}{10395} \frac{n(n-1)(n+1)}{11056 n^9 - 199008 n^8 + 1229097 n^7 - 3202008 n^6 + 1676610 n^5 + 8541768 n^4 - 14633827 n^3 - 712752 n^2 + 13266144 n - 3930120} \]

\[ \kappa_7(\gamma) = -\frac{32}{135135} \frac{n(n-1)(n-2)(n+1)}{1} \]
\[ \cdot \left( 349504 n^{10} - 8475472 n^9 + 71003896 n^8 - 267058999 n^7 \\
+ 348210422 n^6 + 640942288 n^5 - 2372007121 n^4 + 906774769 n^3 \\
+ 3464110995 n^2 - 2287811970 n - 1100064672 \right) \]

A.4. Case 4: \( a_i = r_i^2 \), \( v = 1 \)

The final case is based on \( v = 1 \) and squared-rank scores, \( a_i = r_i^2 \). The reader is referred to Mielke and Berry (1982) for a discussion of situations where this statistic was found to have promising power characteristics.

\[
\kappa_2(\gamma) = \frac{n^2 (n-1) (n+1) (2n^2-3)}{15} \\
\kappa_3(\gamma) = -\frac{n (n-1) (n-2) (n+1) (14n^3 - 17n^2 - 36n + 3) (n^2 - 2)}{105} \\
\kappa_4(\gamma) = \frac{8}{51975} n(n-1)(n+1) \cdot \left( 1323n^9 - 11304n^8 + 20111n^7 + 46116n^6 - 155845n^5 - 22986n^4 \\
+ 354434n^3 - 93771n^2 - 255483n + 61560 \right) \\
\kappa_5(\gamma) = -\frac{4}{2837835} n(n-1)(n-2)(n+1) \cdot \left( 294840n^{11} - 4127076n^{10} + 14998336n^9 + 7232432n^8 - 143919740n^7 \\
+ 150645068n^6 + 524580826n^5 - 796599907n^4 - 943959600n^3 \\
+ 1304674563n^2 + 740377098n - 411490800 \right) \\
\kappa_6(\gamma) = \frac{8}{241215975} n(n-1)(n+1) \cdot \left( 31975398n^{16} - 817832232n^{14} + 7329714641n^{13} - 27239172120n^{12} \\
+ 3785001937n^{11} + 296819276916n^{10} - 720155773257n^9 \\
- 793731145500n^8 + 5026436516467n^7 - 1756852251876n^6 \\
- 15217607068324n^5 + 13189901557020n^4 + 21113130247458n^3 \\
- 22030494862608n^2 - 9798994483200n + 7898269402800 \right) \\
\kappa_7(\gamma) = -\frac{16}{5892561675} n(n-1)(n-2)(n+1) \cdot \left( 1195978446n^{17} - 42197104989n^{16} + 536001017882n^{15} \\
- 3160915012466n^{14} + 6392832285577n^{13} + 25784167445855n^{12} \\
- 167524748399949n^{11} + 134608636200525n^{10} + 122116872315293n^9 \\
- 2735760136596938n^8 - 419581113171643n^7 + 1594289590993813n^6 \\
+ 5722608851767764n^5 - 45464416282230498n^4 - 484805106932130n^3 \\
+ 59044104217790298n^2 - 1806408975149820n - 1848967429408560 \right) \]
REFERENCES


