The Annals of Statistics
1996, Vol. 24, No. 5, 2250–2265

ASYMPTOTICS FOR THE MAXIMUM OF HALF-NORMAL PLOTS[1]

BY LARRY GOLDSTEIN AND LOUIS GORDON

University of Southern California and Filoli Information Systems

We study approximations to the distribution of the largest Walsh–Fourier coefficient obtained by transforming i.i.d. observations from a non-normal parent distribution. The study is inspired by the use of half-normal plots in screening studies and in the analysis of molecular sequence data. In the latter context, we indicate why continuity corrections are needed.

1. Introduction and summary. The full 2^ν factorial design of ν factors each at two levels has played a prominent role in the development of data-analytic and mathematical statistical methods. In this paper, we solve by classical methods a problem of natural data-analytic interest. The problem has substantial mathematical interest as well because it seems remarkably resistant to more modern approaches to solution. Specifically, we study the approximate distribution of the maximum estimated effect resulting from the analysis of variance of a full 2^ν factorial design when errors are i.i.d. with mean 0, and all main effects and interactions are also 0.

The design matrix for the full factorial design can be written in terms of the Kronecker product $\otimes$ as follows. Define the 2×2 matrix

$$M_1 = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Inductively, define $M_\nu = M_1 \otimes M_{\nu-1}$ for $\nu > 1$. The columns $U_1, \ldots, U_{2^\nu}$ can be interpreted as the main effects and interactions for an analysis of variance. The columns are typically labeled by those factors taken at their higher level. The first four columns comprise the vector of all 1's (the common mean) followed by three vectors whose entries repeat respectively in blocks of -1 and $+1$ (the first factor's main effect), in blocks of $-1, -1, +1, +1$ (the second factor's main effect) and in blocks of $+1, -1, -1, +1$ (the second-order interaction of the first two main effects).

Fisher (1942) observes that it is possible to map the subsets $A \subset \{1, \ldots, \nu\}$

Received October 1992; revised September 1995.
[1]Research supported in part by NSF Grant DMS-90-05833.
AMS 1991 *subject classifications*. Primary 60G70; secondary 62K15.
Key words and phrases. Half-normal plot, extremes, Fisher's test.

onto the columns U_A of M_ν in such a fashion that the Abelian group $(2^{\{1,\ldots,\nu\}}, \cdot)$ under the operation symmetric difference is mapped isomorphically onto the Abelian group $(\{U_A\}, \cdot)$ under the operation elementwise multiplication. Alternatively, by identifying columns with successive segments of the unit interval each having length $2^{-\nu}$, we obtain the specification of the first 2^ν Walsh functions in canonical order. The full sequence of Walsh functions constitutes one of the classical orthonormal bases for square-integrable functions on the unit interval. They share many properties with the sines and cosines, including their identity as a representation of a commutative group. See Dym and McKean (1972), Chapter 4, for properties of representations of finite commutative groups and Stoffer (1991) for a comprehensive survey of the statistical literature involving Walsh functions. Another data-analytic problem which strongly uses the group structure of subsets is the subsampling method of Hartigan (1969).

Daniel (1959) suggests using graphical methods in conjunction with the 2^ν factorial design to screen for significant main effects. In modern jargon, given a column vector X of $n = 2^\nu$ observations, Daniel suggests using a q–q plot of the set of absolute values of Walsh–Fourier coefficients $\{|U_j^T X n^{-1/2}|\}$ and gives critical values for declaring that those coefficients with largest absolute value correspond to effects with nonzero mean. Zahn (1975a, 1975b) refines these critical values. Both Zahn's and Daniel's calculations presuppose that measurement errors are independently normally distributed.

The use of Walsh–Fourier coefficients has recently been suggested in the context of molecular sequence analysis. See Tavaré and Giddings (1988). Stoffer, Tyler and McDougall (1992) also suggest the use of such analysis, and mention that a Fisher's test for Walsh–Fourier coefficients would be useful. In both papers, the authors code molecular sequence data by using indicator functions to indicate the identity of each letter in the alphabet (typically $\{A, C, G, T\}$) making up the sequence—either using a separate indicator for the presence of each letter of the alphabet or using a set of indicators coding characteristics, for example, purine versus pyrimidine.

Much is known about the distribution of the Fourier coefficients of i.i.d. variates. Freedman and Lane (1980) study properties of the empirical distribution function of the Fourier coefficients of i.i.d. non-Gaussian white noise. The metric they use, however, gives no information about extremely large or small coefficients. In this paper we solve the analogous question for the largest Walsh–Fourier coefficient of independent white noise that they pose for the largest Fourier coefficient.

We show below that, under suitable regularity conditions on the parent distribution, the distributions of the maximum and of the maximum absolute value of the Walsh–Fourier coefficients of i.i.d. data are well approximated by distributions induced by i.i.d. standard normal data. Specifically, for $M_\nu = (U_1, \ldots, U_{2^\nu})$ and for Φ the standard normal distribution function, we obtain the following result.

THEOREM 1. *Let $X_1, \ldots, X_n$ be $n = 2^\nu$ i.i.d. variates distributed as F with mean 0, variance 1 and moment generating function $\psi(\cdot)$, finite in some neighborhood of 0. Given $0 < \varepsilon < 1/2$ and $\delta > 0$, there exists a constant $\gamma = \gamma(F, \varepsilon, \delta)$ depending only on F, ε and δ for which*

$$(1) \qquad \left| P\left\{ \max_{1 \le j \le n} |U_j^T X| \le t n^{1/2} \right\} - e^{-2n(1-\Phi(t))} \right| < \gamma(F, \varepsilon, \delta) n^{\delta - \varepsilon},$$

whenever $t > \sqrt{2\log(n) - (3 - 2\varepsilon)\log\log(n)}$. The distribution of the maximum of the Walsh–Fourier coefficients is approximated by $e^{-n(1-\Phi(t))}$, on the same interval by an error bound of the identical form.

If, in addition, F is absolutely continuous, or is supported on a lattice, the order of approximation in the same range of test values t can be improved to $O(n^{\delta - 2\varepsilon})$. In the case of a lattice distribution, the test value $tn^{1/2}$ must be a midpoint between two of the lattice points in the coarsest lattice supporting the distribution.

Note that the bounds for the test values required in Theorem 1 cover all but the extreme left tails. For example, the approximate distribution $e^{-n(1-\Phi(t))}$ for the maximum Walsh–Fourier coefficient has median $(2\log n - \log\log n - \log(4\pi) + o(1))^{1/2}$; it has mass $\exp(-(1 + o(1))\sqrt{\log n})$ to the left of $(2\log(n) - 2\log\log(n))^{1/2}$, so that our result yields a large-sample approximation to the entire distribution. An indication that the quality of approximation might be better in the upper tail than in the lower tail is that $e^{-n(1-\Phi(t))}$ puts mass e^{-n} at $-\infty$.

Using Mill's ratio, one can show that for all values of t the approximation (1) is of order $o(\log^{-r} n)$ for any fixed positive r. As a consequence of the exact order calculations of Hall (1979), one cannot obtain a uniform bound of this order when replacing $e^{-n(1-\Phi(t))}$ by the usual double exponential (extreme value) approximation.

In Section 2 we use a Monte Carlo experiment to explore the finite-sample applicability of Theorem 1. In Section 3 we prove the theorem using the Bonferroni inequalities. Finally, in Section 4 we provide a few technical remarks related to the proof.

2. A simulation experiment. In this section we describe the results of a simulation experiment which provide some insight into the utility of our asymptotic approximations. Presented in Figures 1 and 2 are the results of 10,000 simulations each for $20 = 4 \times 5$ choices of parent distribution F and sample size n. In Figure 1 we compare our asymptotic approximation with the observed distribution of the maximum Walsh–Fourier coefficient in the Walsh transform of n observations taken i.i.d. with distribution F. In Figure 2 we make the analogous comparison for the distribution of the maximum absolute value among Walsh–Fourier coefficients computed for the same data as were used to prepare Figure 1.

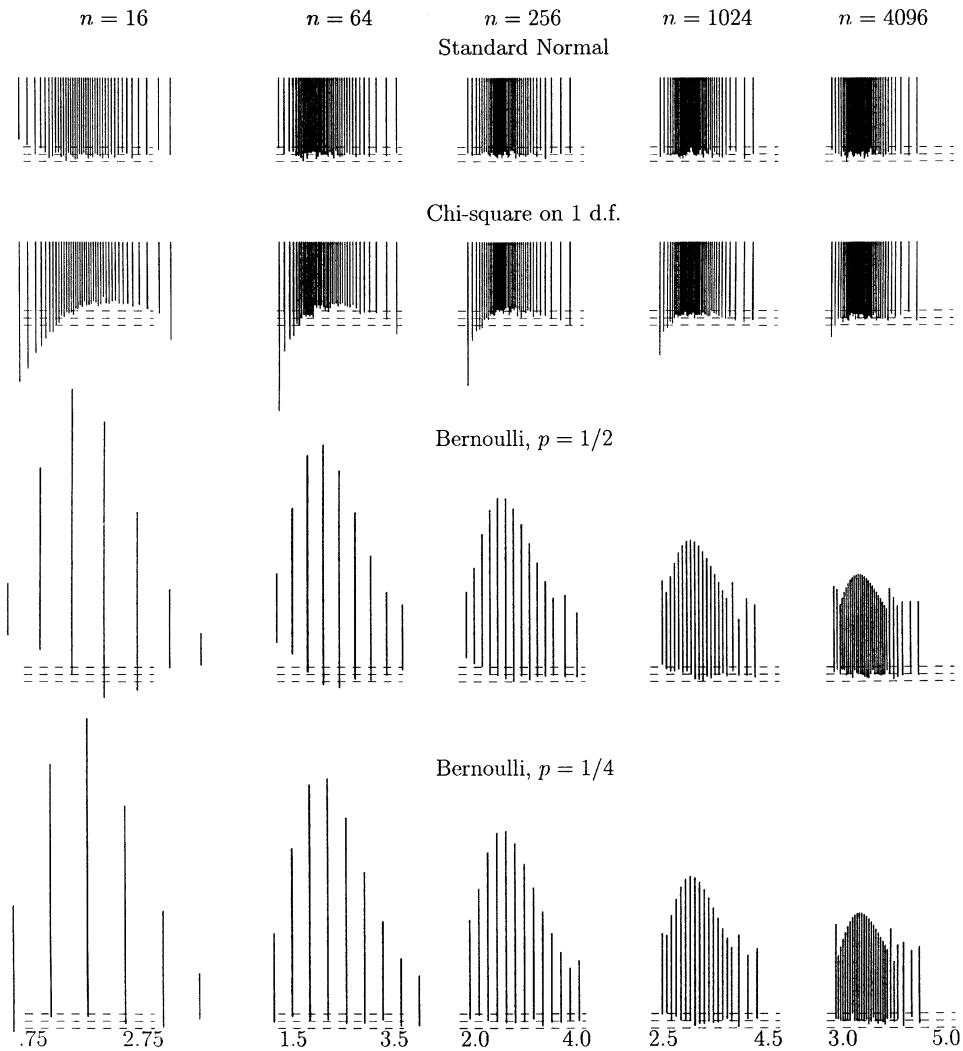

FIG. 1. *Hanging rootogram for approximate distribution* $e^{-n(1-\Phi(t))}$ *for the maximum Walsh–Fourier coefficient of n observations,* 10,000 *simulations.*

All distributions are standardized to facilitate comparison. Choices for distribution function are (1) standard normal; (2) $(\chi_1^2 - 1)/\sqrt{2}$, to illustrate the effects of heavy tails; (3) (Bernoulli(1/2) − 1/2)/$\sqrt{1/4}$, chosen as a symmetric lattice distribution which might be used in molecular sequence analysis to model the purine versus pyrimidine diochotomy; and (4) (Bernoulli(1/4) − 1/4)/$\sqrt{3/16}$, chosen as a model for coding the location of one of the four base pairs comprising molecular sequence data. Sample sizes are chosen from $n \in \{4^2, 4^3, 4^4, 4^5, 4^6\}$. All computations were performed

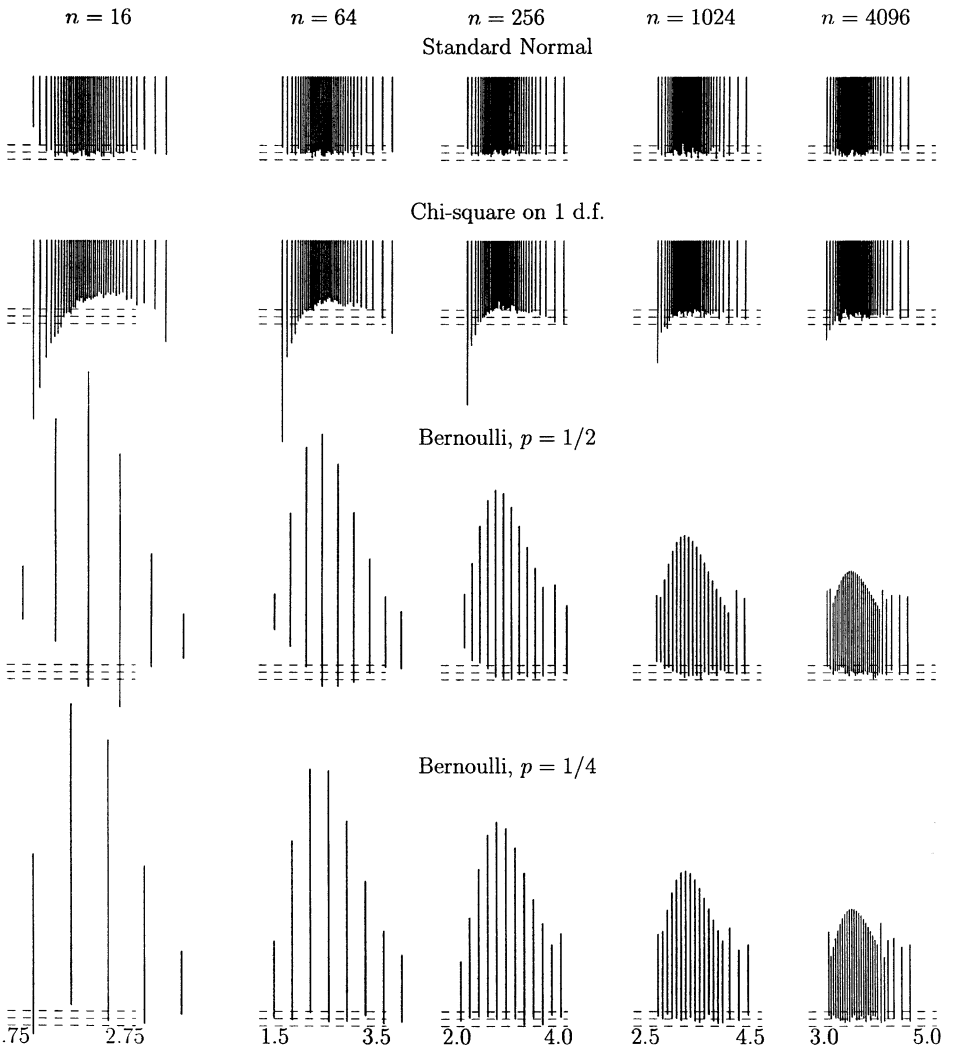

FIG. 2. *Hanging rootogram for approximate distribution* $e^{-2n(1-\Phi(t))}$ *for the maximum absolute value Walsh–Fourier coefficient of n observations*, 10,000 *simulations.*

using Matlab, version 3.5i, running on Sun Sparcstations. See MathWorks (1989).

For continuous parent distributions approximate quantiles are computed by inverting the approximate distribution to obtain the αth quantile $q_\alpha = \Phi^{-1}(1 + \log(\alpha)/n)$ for the maximum coefficient and $q_\alpha = \Phi^{-1}(1 + \log(\alpha)/2n)$ for the maximum of the absolute values, for $\alpha = j/40$ and $1 \le j \le 40$. For discrete distributions the approximate quantiles are rounded to the nearest midpoint of the lattice supporting the distribution.

Forty bins are thereby determined, and a hanging rootogram of these forty values is plotted for each combination of parent distribution and sample size. See Tukey (1965) for a discussion of rootograms. Vertical lines are plotted at the midpoint of each histogram bin. The top of each vertical line is located at the square root of expected content of the bin, using $e^{-n(1-\Phi(t))}$ or $e^{-2n(1-\Phi(t))}$ as the asymptotic approximation for the true distribution function. (Extreme bins are plotted at the same horizontal distance as that between the immediately adjacent two bins.) The length of each vertical line is the square root of the empirical histogram. If the approximation were perfect, the bottoms of the vertical lines would all lie horizontally at the level 0, so that deviations from the horizontal provide an indication of lack of fit.

Because the square root is a variance stabilizing transformation for the Poisson, we indicate by three parallel dotted lines the zero level and levels at $\pm 3\sigma$, in units corresponding to sampling variation. The horizontal lengths of the dotted lines correspond to intervals of length 2 on comparable scales with conveniently round endpoints, to facilitate comparisons of distributional scale or location for varying sample sizes n. Endpoints for the intervals are displayed at the bottom set of rootograms. For example, all rootograms in the fifth column display bins whose centers lie between 3.0 and 5.0.

A number of features of the graphs are immediately apparent. Note that effects of granularity of the data are evident for small values of n. In this case, there are relatively few possible values which the maximum can take. For example, when $n = 16$ the possible values that the Walsh–Fourier coefficients can take are quarter-integers. When one standardizes in the Bernoulli$(1/2)$ case to unit variance, one obtains possible values at half-integers; as a check, there are indeed exactly four bins spanning the interval of length 2. In the Bernoulli$(1/4)$ case, standardization using a smaller variance makes the bins more widely spread. Note also that there is a compression in range as one increases sample size. This is consistent with the behavior in the Gaussian case, where the approximate variance of extreme values decreases at rate $O(1/\log n)$.

The approximation is, of course, nearly perfect in the Gaussian case displayed first, because the Walsh transform is obtained by an orthogonal rotation. What little lack of fit there is can be attributed to approximating the exact distribution of the form $(1 - \Phi(t))^n$ or $(2\Phi(t) - 1)^n$ by functions of the form $e^{-m(1-\Phi(t))}$ for $m = n, 2n$. Fit for all other distributions improves with increasing sample size, and is perhaps acceptable by the time $n = 256$. The approximation is better in the upper tail than it is in the lower tail, as might be expected by the restricted values of t required for the best rates of convergence in Theorem 1. The approximation appears equally good whether one is approximating the maximum coefficient or the maximum coefficient in absolute value.

3. Proof of Theorem 1. Recall that $X_1, \ldots, X_n$ is an i.i.d. sequence with distribution $F(\cdot)$, having mean 0, variance 1 and moment generating function $\psi(\cdot)$. The sample size $n = 2^\nu$ is some large power of 2. The distribution of the

symmetrized random variables $X_1 - X_2$ is denoted by $G(\cdot)$. Denote by F^{*m}, G^{*m} and Φ^{*m} the m-fold convolutions of F, G and Φ, respectively.

Our first two lemmas are concerned with the combinatorial structure of the various linear combinations whose maximum we seek. The complex dependency structure—illustrated in Figure 3 and discussed after the proof is complete—is the reason we have resorted to the Bonferroni inequalities so effectively used by Watson (1954), as opposed to more modern methods of Poisson approximation. See our remarks in Section 4.

LEMMA 2. *Let columns $V_1, \ldots, V_\kappa$ of M_ν generate a subgroup of exactly 2^κ elements under the operation elementwise multiplication of columns. The set of linear combinations $\{V_1^T X, \ldots, V_\kappa^T X\}$ then has representation in terms of the 2^κ subsets $A \subset \{1, \ldots, \kappa\}$, indexing the 2^κ i.i.d. random variables $Z_A \sim F^{*n/2^\kappa}$, where $n = 2^\nu$. Specifically, for all $1 \le j \le \kappa$, we may write $V_j^T X = \sum_{A \subset \{1,\ldots,\kappa\}} b_j(A) Z_A$ and $b_j(A) \in \{-1, 1\}$. All 2^κ possible combinations of ± 1 appear among the 2^κ vectors $(b_1(A), \ldots, b_\kappa(A))$.*

PROOF. By hypothesis, the $n \times 2^\kappa$ matrix V formed of the 2^κ columns generated by all possible elementwise products of the V_j has full rank, because the columns of M_ν are orthogonal. Each entry in V is determined by the set of ± 1's in the rows of the κ columns corresponding to the original V_j's.

	row element															
column transpose of $M_4 = [U_1, \ldots, U_{16}]$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
U_1^T	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
U_2^T	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+
U_3^T	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+
U_4^T	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+
U_5^T	−	−	−	−	+	+	+	+	−	−	−	−	+	+	+	+
Lemma 2: independent linear combinations																
2, 3, 5 set	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+
	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+
	−	−	−	−	+	+	+	+	−	−	−	−	+	+	+	+
$\sum_{j \in \{2,3,5\}} b_j(A)$	−3	−1	−1	1	−1	1	1	3	−3	−1	−1	1	−1	1	1	3
Lemma 3: dependent linear combinations																
2, 3, 4 set	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+
	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+
	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+
$\sum_{j \in \{2,3,4\}} b_j(A)$	−1	−1	−1	3	−1	−1	−1	3	−1	−1	−1	3	−1	−1	−1	3

FIG. 3. *The combinatorial structure of the coefficients $n = 16$ and $k = 3$.*

Because column rank equals row rank, there must be at least 2^κ distinct rows in V. Hence each of the 2^κ possible combinations of κ 1's and -1's must appear as rows in the generating κ columns.

Note that the rows of M_ν also comprise a group under the operation elementwise multiplication. (Indeed, the transpose of every column U_j appears as some row read in inverse order, as can be seen by induction on ν.) For each $A \subset \{1, \ldots, \kappa\}$, let S_A denote the set of *row* indices having -1 entries in V_j for those $j \in A$ and $+1$ entries in columns V_j for $j \in A^c \cap \{1, \ldots, \kappa\}$. For example, $S_\varnothing$ indexes those rows having entry $+1$ in all the κ original columns V_j, and $S_{\{1, \ldots, \kappa\}}$ indexes rows having all -1's in the distinguished κ columns. Choose and fix any $A \neq \varnothing$. We already know that $S_A \neq \varnothing$, so choose some distinguished row R^* indexed by an element of S_A. Because the rows form a group, elementwise multiplication by R^* carries rows indexed by $S_\varnothing$ one-to-one into the set of rows indexed by S_A. Similarly, elementwise multiplication by R^* carries the rows indexed by S_A one-to-one into rows indexed by $S_\varnothing$. Hence each S_A has $2^{\nu-\kappa} = n/2^\kappa$ rows.

For $A \subset \{1, \ldots, \kappa\}$, write $Z_A = \sum_{i \in S_A} X_i$, a sum of $n/2^\kappa$ observations X_i. Note that $Z_A \sim F^{*n/2^\kappa}$. By construction, all X_i summing to Z_A share the same coefficient in all linear combinations $V_j^T X$ for $j = 1, \ldots, \kappa$. Let $b_j(A) \in \{-1, 1\}$ be that coefficient, so that $V_j^T X = \sum_{A \subset \{1, \ldots, \kappa\}} b_j(A) Z_A$. We have previously seen that each S_A is nonempty so that all combinations of ± 1's appear as coefficients. □

Note that Lemma 2 does not apply to any subset of columns containing the first column composed of all 1's. The next lemma is applicable to any subset of columns.

LEMMA 3. *Given columns $V_1, \ldots, V_k$ of M_ν, there exist constants $b_j(A) \in \{-1, 1\}$ and i.i.d. random variables $Z_A \sim F^{*n/2^k}$, both indexed by $A \subset \{1, \ldots, k\}$ such that the k random variables $V_j^T X$ have representation $V_j^T X = \sum_{A \subset \{1, \ldots, k\}} b_j(A) Z_A$.*

PROOF. If the k columns V_j generate a subgroup of size 2^k, Lemma 2 immediately is applicable. Otherwise the subgroup has size 2^κ with $\kappa < k$.

In the latter case, assume without loss that the columns $V_1, \ldots, V_\kappa$ generate the subgroup. From Lemma 2, the first κ columns' linear combinations have representations $V_j^T X = \sum_{A \subset \{1, \ldots, \kappa\}} b_j^*(A) Z_A^*$, where the Z_A^* are i.i.d. as $F^{*n/2^\kappa}$, and $b_j^*(A) \in \{-1, 1\}$. The representation is easily extended to all k columns, because columns $\kappa + 1, \ldots, k$ are obtained by elementwise multiplication of entries in columns 1 through κ. Now arbitrarily allocate the $n/2^\kappa$ summands making up each Z_A^* into $2^{k-\kappa}$ equal-sized groups and define $b_j(A) = b_j^*(A \cap \{1, \ldots, \kappa\})$ for $A \subset \{1, \ldots, k\}$ to obtain the desired representation. □

Figure 3 illustrates the idea of the proofs which formalism may have obscured. Consider the case $k = 3$ and $n = 2^4$. We show in Figure 3 the first

five columns of M_4, U_1 through U_5, where only the signs of the entries in the Hadamard matrix M_4 need be displayed.

The $\kappa = 3$ columns U_2, U_3, U_5 are independent under the group operation elementwise multiplication, so we conclude from the proof of Lemma 2 that we may partition the $n = 2^4$ rows into $2^\kappa = 8$ subsets, each having $n/2^\kappa$ rows. Note that each of the 2^κ possible patterns of $+1$ and -1 appears in exactly two rows determined by columns U_2, U_3 and U_5, as claimed in the lemma. For the sake of comparison, columns U_2, U_3 and U_4 are dependent in the group sense because elementwise multiplication of any two columns yields the third. Hence Lemma 3 is applicable in this case. Because U_2 and U_3 determine the entries of U_4, we may partition the rows into four classes sharing common entries in columns U_2, U_3 and U_4 with each class having four rows, but not every combination of ± 1's now appear.

We next provide a useful bound on the distance between normal distributions with different means and variances. We denote the standard normal density by $\phi(x) = \Phi'(x)$. We denote the normal density having mean μ and variance σ^2 by $\phi(x|\,\mu, \sigma^2)$, with associated distribution function $\Phi(x|\,\mu, \sigma^2)$.

LEMMA 4. *Assume* $\sigma_2^2 \geq \sigma_1^2$. *Then*

$$\int_{-\infty}^{\infty} \left|\phi\left(x|\,\mu_2, \sigma_2^2\right) - \phi\left(x|\,\mu_1, \sigma_1^2\right)\right| dx \leq \frac{|\mu_2 - \mu_1|}{\sigma_2} + \left(\sigma_2^2/\sigma_1^2 - 1\right).$$

PROOF. The left-hand side of the inequality is the total variation distance between two normal distributions. Write $\delta = (\mu_2 - \mu_1)/\sigma_2$ and $\theta = \sigma_2/\sigma_1 \geq 1$. The desired integral is twice

$$\begin{aligned}
&\max_{x_1, x_2}\left|\Phi\left(\frac{x_2 - \mu_2}{\sigma_2}\right) - \Phi\left(\frac{x_1 - \mu_2}{\sigma_2}\right) - \left[\Phi\left(\frac{x_2 - \mu_1}{\sigma_1}\right) - \Phi\left(\frac{x_1 - \mu_1}{\sigma_1}\right)\right]\right| \\
&\quad = \max_{x_1, x_2}\left|\Phi\left(\frac{x_2}{\theta} - \delta\right) - \Phi\left(\frac{x_1}{\theta} - \delta\right) - \left[\Phi(x_2) - \Phi(x_1)\right]\right| \\
&\quad \leq \max_{y_1, y_2}\left|\Phi(y_2 - \delta) - \Phi(y_1 - \delta) - \left[\Phi(y_2) - \Phi(y_1)\right]\right| \\
&\qquad + \max_{x_1, x_2}\left|\Phi\left(\frac{x_2}{\theta}\right) - \Phi\left(\frac{x_1}{\theta}\right) - \left[\Phi(x_2) - \Phi(x_1)\right]\right|,
\end{aligned}$$

where we have added and subtracted $\Phi(x_2/\theta) - \Phi(x_1/\theta)$ and then used the triangle inequality. The first summand in the upper bound is itself bounded by $\Phi(|\delta|/2) - \Phi(-|\delta|/2) < |\delta|/\sqrt{2\pi}$. If $\theta = 1$ the former bound suffices. If not, $\theta > 1$, in which case the second summand is bounded above by

$$2\left|\Phi\left(\theta\sqrt{\frac{\log(\theta^2)}{\theta^2 - 1}}\right) - \Phi\left(\sqrt{\frac{\log(\theta^2)}{\theta^2 - 1}}\right)\right| < \theta - 1 < \frac{1}{2}(\theta + 1)(\theta - 1)$$

by the mean value theorem and the bound $\log(\theta) < \theta - 1$. □

We use the notation $O_e(x)$ for a quantity which, when divided by x, is bounded by 1 in absolute value. The notation is convenient in a number of the following arguments where we require uniformity of bounds in normal approximation simultaneously across a number of distributions.

LEMMA 5. *Denote by P_F the probability measure in which the X's have common distribution F and by P_Φ the probability measure in which the X's have common standard normal distribution Φ. Assume F has mean 0, variance 1 and moment generating function $\psi(\cdot)$, finite in some neighborhood of 0. There then exists a constant $\gamma = \gamma(F)$ such that, for any $k \geq 1$ columns $V_1, \ldots, V_k$ of M_ν and any $r_1, \ldots, r_k$ elements of $\{-1, 1\}$,*

$$\left| P_F\left\{ \min_{j \leq k} r_j V_j^T X > tn^{1/2} \right\} - P_\Phi\left\{ \min_{j \leq k} r_j V_j^T X > tn^{1/2} \right\} \right| < \frac{k^2 t^3 \gamma^{k+1} \psi^k(t)}{n^{1/2}},$$

whenever $\gamma k^2 t^3 n^{-1/2} < 1 < t$.

PROOF. Consider the random variable $S = \sum_{j=1}^k r_j V_j^T X$. Use Lemma 3 to represent the sum

$$S = \sum_{A \subset \{1, \ldots, k\}} \sum_{j=1}^k r_j b_j(A) Z_A = \sum_{A \subset \{1, \ldots, k\}} \sum_{j=1}^k c_j(A) Z_A,$$

where $c_j(A) \in \{-1, 1\}$ and the Z_A are i.i.d. with distribution $F^{*n/2^k}$ under P_F-measure or with distribution $\Phi^{*n/2^k}$ under P_Φ-measure. Under either measure, $(n/2^k)\sum_A(\sum_{j=1}^k c_j(A))^2$ is the variance of S, a sum of k uncorrelated variates, each having variance n. Hence $\sum_A(\sum_{j=1}^k c_j(A))^2 = k2^k$.

Let $I_j = I_{\{r_j V_j^T X > t\sqrt{n}\}}$. Let $c_A = \sum_{j=1}^k c_j(A)$. Observe that $|c_A| \leq k$. Let s be a constant whose value will be chosen later and write $S = \sum_A c_A Z_A$. We wish to approximate

$$\begin{aligned} &E_F\left\{ \prod_{j=1}^k I_j \right\} \\ &\quad = \int \cdots \int_{S > ktn^{1/2}} \left[\prod_{j=1}^k I_{\{\sum_A c_j(A) z_A > t\sqrt{n}\}} \right] \prod_{A \subset \{1, \ldots, k\}} F^{*n/2^k}(dz_A) \\ &\quad = \left(\prod_A \psi^{n/2^k}(sc_A) \right) \exp(-ktn^{1/2}s) \int \cdots \int_{S > ktn^{1/2}} \left[\prod_{j=1}^k I_j \right] \\ &\qquad \times \exp(-s(S - ktn^{1/2})) \prod_A \frac{\exp(sc_A z_A) F^{*n/2^k}(dz_A)}{\psi^{n/2^k}(sc_A)}. \end{aligned} \tag{2}$$

Because F has a moment generating function, we may choose a constant $\gamma > 1$ sufficiently large such that

$$
\begin{aligned}
\log(\psi(s)) &= \frac{s^2}{2} + O_e(\gamma s^3), \\
\frac{d}{ds}\log(\psi(s)) &= s + O_e(\gamma s^2), \\
\frac{d^2}{ds^2}\log(\psi(s)) &= 1 + O_e(\gamma s), \\
\frac{d^4}{ds^4}\log(\psi(s)) &= O_e(\gamma),
\end{aligned} \tag{3}
$$

whenever $\gamma|s| < 1$.

The shifted probability distribution $\exp(sc_A z)F(dz)/\psi(sc_A)$ has cumulant generating function $\log(\psi(sc_A + \zeta)) - \log(\psi(sc_A))$. Recall that $|c_A| \le k$. Hence $\gamma k|s| < 1$ implies the shifted distribution has mean $\mu_A(s) = c_A s + O_e(\gamma c_A^2 s^2)$, variance $\sigma_A^2(s) = 1 + O_e(\gamma c_A s)$ and third centered absolute moment bounded by γ.

If $F = \Phi$, then $\mu_A(s) = c_A s$ and $\sigma_A^2(s) = 1$ exactly. Recall that $\sum_A c_A^2 = k2^k$. Use (2), (3) and the Berry–Esseen theorem of Feller (1971), Chapter 16, applied 2^k times, once for each of the subsets A indexing the Z_A, to conclude

$$
\begin{aligned}
E_F\left\{\prod_{j=1}^k I_j\right\} = \exp\left(n2^{-k}\sum_A\left[\frac{c_A^2 s^2}{2} + O_e\bigl(\gamma k c_A^2 s^3\bigr)\right] - ktn^{1/2}s\right) \\
\times\left(\int\cdots\int_{S>ktn^{1/2}}\exp\bigl(-s(S-ktn^{1/2})\bigr)\right. \\
\times\prod_{j=1}^k I_j\prod_A \Phi^{*n/2^k}\bigl(dz_A|\,\mu_A(s),\sigma_A^2(s)\bigr) \\
\left.+\sum_A O_e\left(3\gamma\left(\frac{n}{2^k}\right)^{-1/2}\right)\right),
\end{aligned}
$$

whenever $\gamma k|s| < 1$. Set $s = tn^{-1/2}$ and again use $\sum_A c_A^2 = k2^k$ to obtain

$$
\begin{aligned}
E_F\left\{\prod_{j=1}^k I_j\right\} = \exp\bigl(-kt^2/2\bigr)\left(\int\cdots\int_{S>ktn^{1/2}}\exp\bigl(-s(S-ktn^{1/2})\bigr)\right. \\
\times\prod_{j=1}^k I_j\prod_A \Phi^{*n/2^k}\bigl(dz_A|\,\mu_A(s),\sigma_A^2(s)\bigr) \\
\left.+O_e\bigl(3\gamma k^2 t^3 n^{-1/2}\bigr) + O_e\bigl(3\gamma 2^{3k/2}n^{-1/2}\bigr)\right),
\end{aligned}
$$

whenever $\gamma k^2 t^3 n^{-1/2} < 1$. If $\gamma tkn^{-1/2} < 1/2$ as well, we may use Lemma 4 to obtain

$$
\begin{aligned}
E_F\left\{\prod_{j=1}^{k} I_j\right\} = \exp(-kt^2/2)\int\cdots\int_{S>ktn^{1/2}} \exp(-s(S-ktn^{1/2})) \\
\times \prod_{j=1}^{k} I_j \prod_A \Phi^{*n/2^k}(dz_A|c_A s, 1) \\
+ \exp(-kt^2/2)\big[O_e(2\gamma kt^2 2^k n^{-1/2}) \\
+ O_e(3\gamma k^2 t^3 n^{-1/2}) \\
+ O_e(3\gamma 2^{3k/2} n^{-1/2})\big],
\end{aligned}
\tag{4}
$$

the first error term due to slightly changing 2^k normal means and variances, the middle term due to approximating the cumulant generating function by a quadratic and the last term due to the Berry–Esseen theorem. The first expression on the right-hand side of (4) is exactly $E_\Phi\{\prod_{j=1}^k I_j\}$, proving the assertion, by increasing the size of γ to accommodate the normalizing factor in $\phi(\cdot)$ and the terms with exponent k in the error bounds. □

We have proved the preceding lemma under fairly general conditions on the parent distribution F and with no restrictions on the identity of the k Walsh–Fourier coefficients of interest. We can obtain substantially better bounds with slightly stronger restrictions on the distribution and on the particular k coefficients selected.

LEMMA 6. *If, in addition to the hypotheses of Lemma* 5, *F is either a lattice distribution or is absolutely continuous and if the k columns $V_1, \ldots, V_k$ are independent in the group sense of elementwise multiplication, then there exists a constant γ depending only on the distribution function F for which*

$$
\left|P_F\left\{\min_{j\le k} r_j V_j^T X > tn^{1/2}\right\} - P_\Phi\left\{\min_{j\le k} r_j V_j^T X > tn^{1/2}\right\}\right| < \frac{k^3 t^4 \gamma^{k+1}\phi^k(t)}{n},
$$

whenever $\gamma k^3 t^4 n^{-1/2} < 1 < t$. If F is concentrated on the lattice $a + bj$ for integer j and maximal span b, then t must also be taken to be of the form $a + b(j + 1/2)$.

PROOF. Recall that $G(\cdot)$ is the distribution of the difference of two i.i.d. variates distributed as F. Denote its moment generating function by $\psi_G(s) =$

$\psi(s)\psi(-s)$, so that as in (3) there exists a constant γ for which

$$\begin{aligned}
\log(\psi_G(s)) &= s^2 + O_e(\gamma s^4),\\
\frac{d}{ds}\log(\psi_G(s)) &= 2s + O_e(\gamma s^3),\\
\frac{d^2}{ds^2}\log(\psi_G(s)) &= 2 + O_e(\gamma s^2),\\
\frac{d^3}{ds^3}\log(\psi_G(s)) &= O_e(\gamma s),\\
\frac{d^4}{ds^4}\log(\psi_G(s)) &= O_e(\gamma),
\end{aligned} \tag{5}$$

whenever $\gamma|s| < 1$. Now use Lemma 2 to write $V_j^T X = \sum_{A \subset \{1,\dots,k-1\}} c_j(A) W_A$, where the W_A are 2^{k-1} i.i.d. variates with common distribution $G^{*n/2^{k-1}}$. We now follow the proof of Lemma 5, save that we use (the proofs of) Theorems 4.1 and 4.2 of Feller (1971), Chapter 16, instead of the Berry–Esseen theorem. The results themselves are not applicable because we require uniformity among the shifted distributions of Z_A. Needed also in the case of densities is a uniform version of the Riemann–Lebesgue lemma. One can modify the proof of Feller (1971), Section 15.4, Lemma 3.

The symmetrization afforded by Lemma 2 with consequent improved expansions (5) allows us to replace the Berry–Esseen bounds by uniform bounds of order $O(t/n)$. Arguing as in the derivation of (4) yields the existence of a positive constant $\gamma = \gamma(F)$ for which

$$\begin{aligned}
E_F\left\{\prod_{j=1}^k I_j\right\} &= E_\Phi\left\{\prod_{j=1}^k I_j\right\}\\
&\quad + \exp(-kt^2/2)\left[O_e\left(\frac{\gamma k t^3 2^k}{n}\right) + O_e\left(\frac{\gamma k^3 t^4}{n}\right) + O_e\left(\frac{\gamma 2^{3k/2} t}{n}\right)\right].
\end{aligned}$$

□

We now complete the proof of Theorem 1. We may use the finiteness of the moment generating function and Markov's inequality to restrict our attention to the case $t < (4\log n)^{1/2}$, which we assume for the rest of the section.

The principle of inclusion and exclusion yields

$$\begin{aligned}
&P_F\left\{\max_{j\le n}|U_j^T X| \le tn^{1/2}\right\}\\
&\quad = 1 + \sum_{k=1}^n (-1)^k \sum_{\substack{A\subset\{1,\dots,n\}\\ |A|=k}} \sum_{r_\cdot\in\{-1,1\}^A} P_F\left\{\min_{i\in A} r_i U_i^T X > tn^{1/2}\right\},
\end{aligned} \tag{6}$$

where $|A|$ denotes the cardinality of the subset A. The identity also holds when $F = \Phi$, in which case the orthogonality of the columns of M_ν allows us to write the left-hand side of (6) as $(1 - 2[1 - \Phi(t)])^n$.

The Bonferroni inequalities, as used in Watson (1954), let us truncate the inclusion–exclusion series (6) at any k, obtaining either an upper or lower bound, as determined by the sign of the first dropped term. Truncate one of the two inclusion–exclusion series for P_F or for P_Φ at κ and the other at $\kappa - 1$. Then apply Lemma 5 to obtain

$$
\begin{aligned}
&\left| P_F\left\{ \max_{j \le n} |U_j^T X| \le t n^{1/2} \right\} - P_\Phi\left\{ \max_{j \le n} |U_j^T X| \le t n^{1/2} \right\} \right| \\
&\quad = \left| P_F\left\{ \max_{j \le n} |U_j^T X| \le t n^{1/2} \right\} - \left(1 - 2[1 - \Phi(t)]\right)^n \right| \\
&\quad < \binom{n}{\kappa} 2^\kappa (1 - \Phi(t))^\kappa + \sum_{k=1}^{\kappa} \gamma \binom{n}{k} 2^k (\gamma \phi(t))^k \frac{k^2 t^3}{n^{1/2}} \\
&\quad < \left(\frac{2en(1 - \Phi(t))}{\kappa} \right)^\kappa + \sum_{k=1}^{\kappa} \gamma \left[\frac{2n\gamma e \phi(t)}{k} \right]^k \frac{k^2 t^3}{n^{1/2}}
\end{aligned}
\tag{7}
$$

for $1 \le t \le (4 \log n)^{1/2}$ and n sufficiently large. In passing from the first to the second inequality, we have used that Stirling's formula is actually a lower bound, namely $k! \ge \sqrt{2\pi} k^{k+1/2} e^{-k}$ for $k \ge 0$. See, for example, Whittaker and Watson (1927), Section 12.4.

Let $0 < \varepsilon < 1/2$ be given and let $\delta > 0$ be arbitrarily small. Let $\kappa = \lfloor c \log(n)/\log\log(n) \rfloor$ for some positive constant c yet to be determined and take $t \ge (2\log(n) - (3 - 2\varepsilon)\log\log(n))^{1/2}$ for n sufficiently large. We now use that $s(1 - \Phi(s))/\phi(s) \to 1$ as $s \to \infty$. The first bounding term in (7) is $O(\exp((\delta - \varepsilon c)\log n)$. Because $\varepsilon \le 1/2$, the summands in the second bounding term are eventually increasing in $k \le \kappa$. The κth summand is of order $O(\exp((\delta + (1/2 - \varepsilon)c - 1/2)\log n))$; so also is the entire second sum of $O(\log n)$ terms. Both bounds are uniform in t for t in the specified range, and their sum decays fastest when $c = 1$.

We have thus far shown that the P_F-distribution of $\max_{j \le n} |U_j^T X|$ is well approximated by $(1 - 2[1 - \Phi(t)])^n$ over the specified range of t, over which $1 - \Phi(t) = O(\log(n)/n)$. Finally, use the inequality $nx^2 e^{-nx} > e^{-nx} - (1 - x)^n > 0$ for $|x| < 1$ to show that $(1 - 2[1 - \Phi(t)])^n$ is well approximated by $e^{-2n(1 - \Phi(t))}$ over the necessary set of t's. For a proof of the inequality, see Whittaker and Watson (1927), Section 12.21.

In case F is lattice or is absolutely continuous, use Lemmas 5 and 6 to show the second sum in (7) is bounded by $O(\exp(\delta + (1/2 - \varepsilon)c - 1)\log n))$. In this case the best rate is obtained when $c = 2$. Hence Theorem 1 is proved for $\max_{j \le n} |U_j^T X|$. The proof for the distribution of the maximum Walsh–Fourier coefficient is essentially identical.

4. Remarks.

1. We have actually proved substantially more than Theorem 1. The identical argument shows that observing the location of extreme observations among finitely many disjoint subsets of the $2 \cdot 2^\nu$ determinations of Walsh func-

tion and sign behave asymptotically independently with probability of observation approximately $e^{-m(1-\Phi(t))}$, where m is the cardinality of a subset in question. We suspect, but have not proved, that the location of the maximum among the $2 \cdot 2^{\nu}$ possibilities is uniformly distributed in the limit.

2. The natural context in which to raise our problem is that of Poisson approximation as developed in Aldous (1989), with formal computations carried out by some variant of the Chen–Stein method. We have been unable to give a proof of our result using either the local methods discussed in Arratia, Goldstein and Gordon (1990) or the coupling methods of Barbour, Holst and Janson (1992). The difficulty in the former case is caused by our inability to define a local neighborhood of dependence. The difficulty in the latter case is caused by dependence that is neither exchangeable nor monotone. All difficulties are exemplified in Figure 3. By summing and considering very skew distributions, we see that the joint distribution of $U_2^T X, U_3^T X, U_5^T X$ is not in general the same as the joint distribution of $U_2^T X, U_3^T X, U_4^T X$.
3. Our method of proof combines an insight into the combinatorial structure of the Walsh–Fourier coefficients with the large deviation asymptotics of Bahadur and Ranga Rao (1960). Of possible technical interest is that we can use the large deviation style of calculation with approximate centering in contrast to a more straightforward application of the theory requiring exact centering.
4. There is a more extensive literature to the analogous problem of finding the maximum Fourier coefficient of non-Gaussian i.i.d. observations. Besides Freedman and Lane (1980), see Walker (1965), Turkman and Walker (1984) and Gersho, Gopinath and Odlyzko (1979).

REFERENCES

ALDOUS, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. Springer, New York.

ARRATIA, R., GOLDSTEIN, L. and GORDON, L. (1990). Poisson approximation and the Chen–Stein method. *Statist. Sci.* **5** 403–424.

BAHADUR, R. R. and RANGA RAO, R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation*. Oxford Univ. Press.

DANIEL, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics* **1** 311–342.

DYM, H. and MCKEAN, H. P. (1972). *Fourier Series and Integrals*. Academic Press, New York.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

FISHER, R. A. (1942). The theory of confounding in factorial experiments in relation to the theory of groups. *Ann. Eugenics* **11** 341–353.

FREEDMAN, D. and LANE, D. (1980). The empirical distribution of Fourier coefficients. *Ann. Statist.* **8** 1244–1251.

GERSHO, A., GOPINATH, B. and ODLYZKO, A. M. (1979). Coefficient inaccuracy in transversal filtering. *Bell System Tech. J.* **58** 2301–2316.

HALL, P. (1979). On the rate of convergence of normal extremes. *J. Appl. Probab.* **16** 433–479.
HARTIGAN, J. A. (1969). Using subsample values as typical values. *J. Amer. Statist. Assoc.* **64** 1303–1317.
MATHWORKS (1989). *Pro-Matlab User's Guide*. The MathWorks, South Natick, MA.
STOFFER, D. S. (1991). Walsh–Fourier analysis and its statistical applications. *J. Amer. Statist. Assoc.* **86** 461–479.
STOFFER, D. S., TYLER, D. E. and MCDOUGALL, A. J. (1992). Spectral analysis for categorical time series: scaling and the spectral envelope. Preprint.
TAVARÉ, S. and GIDDINGS, B. W. (1988). Some statistical aspects of the primary structure of nucleotide sequences. In *Mathematical Methods for DNA Sequences* (M. S. Waterman, ed.) 117–132. CRC Press, Boca Raton, FL.
TUKEY, J. (1965). The future of processes of data analysis. In *Proceedings of the Tenth Conference Design of Experiments in Army Research Development and Testing* 691–729. Army Research Office, Durham, NC. [Also in *The Collected Works of John W. Tukey, Volume IV: Philosophy and Principles of Data Analysis, 1965–1986*. Wadsworth Advanced Books and Software, Monterey, CA, 1986.]
TURKMAN, K. F. and WALKER, A. M. (1984). On the asymptotic distributions of maxima of trigonometric polynomials with random coefficients. *Adv. in Appl. Probab.* **16** 819–842.
WALKER, A. M. (1965). Some asymptotic results for the periodogram of a stationary time series. *J. Austral. Math. Soc.* **5** 107–128.
WATSON, G. S. (1954). Extreme values in samples from m-dependent stationary stochastic sequences. *Ann. Math. Statist.* **25** 798–800.
WHITTAKER, E. T. and WATSON, G. N. (1927). *A Course of Modern Analysis*, 4th ed. Cambridge Univ. Press.
ZAHN, D. A. (1975a). Modifications of and revised critical values for the half-normal plot. *Technometrics* **17** 189–200.
ZAHN, D. A. (1975b). An empirical study of the half-normal plot. *Technometrics* **17** 201–212.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CALIFORNIA 90089-1113

FILOLI INFORMATION SYSTEMS
777 CALIFORNIA AVE.
PALO ALTO, CALIFORNIA 94304-1102