The asymptotic relative efficiency of mixed statistical tests — **Source link** ⬈

E. Feustal, L. Davisson

Institutions: California Institute of Technology

Published on: 01 Apr 1967 - IEEE Transactions on Information Theory (IEEE)

Topics: Sign test and Wilcoxon signed-rank test

Related papers:

- On adaptive nonparametric detection systems using dependent samples

- On nonparametric signal detectors

- On a Theorem of Pitman

- Asymptotic Efficiency of Certain Locally Most Powerful Rank Tests

- Nonparametric detection

The Asymptotic Relative Efficiency
of Mixed Statistical Tests

EDWARD A. FEUSTAL, MEMBER, IEEE, AND LEE D. DAVISSON, MEMBER, IEEE

Abstract—Mixed statistical tests are described. It is shown that these tests have a much higher efficiency than conventionally used statistics such as the sign test and polarity coincidence correlation without the high operational complexity of the Wilcoxon, Mann-Whitney, Kendall τ, or Fisher-Yates: Terry-Hoeffding tests.

I. INTRODUCTION

IN THE PAST ten years important results were obtained in the theory of distribution-free or non-parametric tests. Engineering interest was focused on the use of these statistics in detection theory, particularly in the sonar problem. The nonparametric detector is appropriate when insensitivity to change in input statistics, a fixed false alarm rate independent of input statistics, and sampled-data processing of input are desired.

In applications such as sonar where very large samples are available, the asymptotic relative efficiency (ARE) is the usual figure of merit for a detector. An additional consideration is the complexity of implementation. Simple detection tests such as the sign test and polarity coincidence correlator have very low efficiencies, but are easily implemented requiring $O(N)$ operations on the N observations. Complex detection tests which require ranking such as the Wilcoxon and Mann-Whitney have high efficiencies but require $O(N^2)$ operations.

A class of statistical tests is proposed in this paper that has relatively high efficiency with intermediate complexity. This class is called the class of "mixed statistical tests." A mixed statistic is a threshold test which is derived as follows. The N observations from each of k channels (samples) are divided into p groups of m observations. An intermediate statistic on each of the m observations appropriate for k channels reduces kN observations to p values. The p values are summed to form the final value which is compared with the threshold.

The advantage of this procedure is that the asymptotic relative efficiency as a function of m approaches its maximum value rapidly with increasing m and only requires $O(m^2p) = O(mN)$ operations. Hence recent advances in digital techniques allow real-time implementation of these more efficient statistical techniques.

In this paper the ARE of mixed statistics which use the Wilcoxon, Mann-Whitney, Kendall, and Fisher-Yates: Terry-Hoeffding procedures as intermediate statistics are investigated, and digital implementations are developed. These schemes allow increases of 30 percent to 200 percent in efficiency over sign tests and related techniques under the assumption of Gaussian noise input.

II. ASYMPTOTIC RELATIVE EFFICIENCY AND EFFICACY

The tests considered in this paper are based on the Neyman-Pearson criteria. Under these criteria, the probability of detection β is maximized while the false alarm rate is constrained to be less than a fixed level α. Thus the optimum test is generally uniquely determined if the probability distributions of signal and noise are known. Nonparametric tests are suboptimum tests which are applicable when only some general properties of the signal and noise are known and are not uniquely determined. Hence a further criterion is necessary to order the performance of such detectors. The selected criterion is Pitman's asymptotic relative efficiency.[1] It is particularly appropriate in the case of weak signal detection. If the hypothesis is taken to be noise alone and the alternative to be signal plus noise, the asymptotic relative efficiency applies when the alternative approaches the hypothesis—that is, the signal-to-noise ratio approaches zero. Specifically, it relates the number of samples required to maintain the false alarm rate below α while the probability of detection is β in one detector to the number required for a second in the limiting case:

$$\text{ARE } (d_1, d_2) = \lim_{\substack{K \to H \\ N_1, N_2 \to \infty}} \frac{N_1(\alpha, \beta, K)}{N_2(\alpha, \beta, K)} \qquad (1)$$

where

α is the probability of false alarm not to be exceeded;
β is the probability of detection to be maintained;
N_i is the minimum number of observations required for detector d_i.

It is assumed that both d_1 and d_2 are consistent.

The tests of interest are threshold tests. A statistic S^N based on the N observations is compared with a threshold $C(N)$ which has been determined in an a priori fashion from the required value of α. In most previous electrical engineering work, functions on S^N, $C(N)$ were used to calculate the ARE in an extremely laborious fashion. Statisticians have used a concept known as efficacy.[2]

It is a measure of efficiency for one test which may be directly compared to that of another to obtain the ARE:

$$\text{ARE}\,(d_1,\,d_2) \;=\; \frac{\mathcal{E}_1}{\mathcal{E}_2} \tag{2}$$

where $\mathcal{E}_i$ is the efficacy of the ith test. The advantage of efficacy is ease of calculation with complicated tests.

There are two forms of efficacy, one for the one-channel case and one for the two-channel case. The two-channel test calculates a statistic on the sets $\{x_i/i = 1, \cdots, m\}$ and $\{y_i/i = 1, \cdots, n\}$ where the first set is from the X channel and the second is from the Y channel. Depending on the detection problem, Y may be a reference channel or an independent signal channel.

The two-channel form follows:

$$\mathcal{E} \;=\; \lim_{N\to\infty} \left\{ \frac{E_K(S^N) - E_H(S^N)}{\Delta_N \,\text{Var}_H^{1/2}\,(S^N)} \right\}^2 \frac{1}{\gamma(1-\gamma)N} \tag{3}$$

where it is assumed that 1) the central limit theorem applies to the statistic S^N, 2) $\Delta_N = kN^{-\frac{1}{2}} \propto E_K(S^N) - E_H(S^N)$, and 3) $0 < \gamma < 1$ is the proportion of observations in the X channel (usually $\frac{1}{2}$), or alternatively:

$$\mathcal{E} \;=\; \lim_{\substack{N\to\infty \\ \Delta_N\to 0}} \left\{ \frac{\dfrac{\partial E_K\,(S^N)}{\partial \Delta_N}}{\text{Var}_{H_o}^{1/2}\,(S^N)} \right\}^2 \frac{1}{\gamma(1-\gamma)N}. \tag{4}$$

The one-channel form is very similar:

$$\mathcal{E}(S^N) \;=\; \lim_{N\to\infty} \left\{ \frac{\dfrac{\partial E_K\,(S^N)}{\partial \Delta_N}}{\text{Var}_{H_o}^{1/2}\,S^N} \right\}^2 \frac{1}{N} \tag{5}$$

where all but condition 3) apply. Alternatively, this latter form may be written:[1]

$$\mathcal{E}(S^N) \;=\; \lim_{N\to\infty} \frac{[\mu_{S^N}(kN^{-1/2}) - \mu_{S^N}(0)]^2}{k^2\,\text{Var}_H\,(S^N)}. \tag{6}$$

where $\mu_S{}^m(\theta)$ is the mean S^N for signal θ.

III. MIXED STATISTICAL TESTS

The sample mean test, a well-known parametric test which is optimum for independent Gaussian observations of known variance, requires about 22 000 samples for $\alpha = 0.001$, $\beta = 0.95$ and signal-to-noise ratio (SNR) $= -30$ dB. For this large number of observations, the central limit theorem should apply. It also shows the reason why only simple tests have been used in the past. The amount of computation for a sign test is $O(N)$ while the amount of computation for the Wilcoxon is $O(N^2)$. Yet the ARE of the sign test for normal alternatives is only about 0.63, whereas the ARE of the Wilcoxon for normal alternatives is about 0.95. If the Wilcoxon is used, a saving of about 30 percent in the sampling time results, but the processing time becomes so long as

[1] Suggested by a reviewer.

to preclude real-time application. Herein lies the justification for mixed statistical tests.

A mixed statistic is the sum of p observations of an intermediate statistic S_I^m operating on independent groups of m independent data points taken from each of k simultaneous channels or "samples." The subscript I denotes "intermediate" and subsequently stands for the exact intermediate statistic used (S for sign test, etc.). $N = mp$ is the total number of data points for each sample, and will be used to calculate the ARE with respect to any other k-sample test for fixed false alarm rate and probability of detection. The work involved is $O(m^2p)$. The question is, "What is the asymptotic relative efficiency of the mixed statistic as a function of m?" The answer is that the ARE rises sharply with increasing m, usually within the range $6 \le m \le 15$ to between 80 and 90 percent of the efficiency of the highly efficient rank statistic.

An intuitive motivation for this procedure is the idea that nonparametric techniques are transformations which convert arbitrary distributions into known ones, usually very close to Gaussian distributions for 20 or 30 observations. It is known that the optimum method of combining independent Gaussian samples for detection is to add them. The intermediate statistic serves to convert data from an arbitrary distribution to "almost" Gaussian data which are then summed and detected. The fact that the ARE of the mixed statistical test is sharply monotonic with m confirms this intuitive approach.

The discussion to follow first investigates the one-sample case and then the two-sample case. The k-sample case will be discussed at a later date. Three types of statistics will be investigated: the Wilcoxon (and subsequently the similar Mann–Whitney and Kruskal–Wallis), the Kendall τ, and the Fisher–Yates: Terry Hoeffding (FYTH) (and subsequently the k-sample expected rank-correlation test of Rushfort and Daly).

A simple, one-sample, distribution-free test that has been throughly investigated[3] is the sign test:

$$S_S^N \;=\; \sum_{i=1}^{N} \text{sgn}\,x_i \tag{7}$$

or equivalently,

$$S_S^N \;=\; \sum_{i=1}^{N} u(x_i) \tag{8}$$

$$u(x_i) \;=\; \begin{cases} 1 & x_i \ge 0 \\ 0 & \text{otherwise}. \end{cases} \tag{9}$$

The ARE of this detector in location testing with respect to the student's t-test for Gaussian alternatives is $2/\pi$. The Wilcoxon, one-sample detector is a general extension of this detector and calculates the statistic:[4]

$$S_W^N \;=\; \sum_{i=1}^{N}\sum_{j=i}^{N} \text{sgn}\,(x_i + x_j) \quad \text{or} \quad S_W^N \;=\; \sum_{i=1}^{N}\sum_{j=i}^{N} u(x_i + x_j). \tag{10}$$

The ARE of this detector with respect to the student's t-test for Gaussian alternatives is $3/\pi$.

The following question arises: If p observations of the output of a Wilcoxon operating on m data points are summed and a decision is based on the statistic[2]

$$S_W^N = \sum_{i=0}^{p-1} S_W^m(i) \tag{11}$$

where

$$S_W^m(i) = \sum_{i=1}^{m} \sum_{k=i}^{m} u(x_{im+j} + x_{im+k}) \tag{12}$$

what is the ARE with respect to the optimum parameter test for Gaussian alternatives? First the efficacies for mixed statistics from one channel will be calculated. For sufficiently small θ:

$$\mu_{S^m}(a\theta) = a\mu_{S^m}(\theta). \tag{13}$$

For a sequence of alternatives $\theta = k'N^{-1/2}$, the efficacy of the mixed test based on

$$S^N = \sum_{i=0}^{p-1} S_I^m(i), \qquad N = mp \tag{14}$$

is

$$\mathcal{E}(S^N) = \lim_{N\to\infty} \frac{[\mu_{S^N}(k'N^{-1/2}) - \mu_{S^N}(0)]^2}{k'^2 \, \mathrm{Var}_H (S^N)} \tag{15}$$

$$= \lim_{N\to\infty} \frac{p^2[\mu_{S^m}(k'N^{-1/2}) - \mu_{S^m}(0)]^2}{pk'^2 \, \mathrm{Var}_H S^m} \tag{16}$$

$$= \lim_{N\to\infty} p \, \frac{k'^2 N^{-1}}{k^2 m^{-1}} \frac{[\mu_{S^m}(km^{-1/2}) - \mu_{S^m}(0)]^2}{k'^2 \, \mathrm{Var}_H (S^m)} \tag{17}$$

$$= \mathcal{E}(S_I^m). \tag{18}$$

In precisely the same manner, the efficacy of the two-channel mixed statistic is shown to be

$$\mathcal{E}(S^N) = \mathcal{E}(S_I^m). \tag{19}$$

Returning to the question, we find that

$$\mathrm{ARE} = \frac{24(m - 1 + \sqrt{2})^2}{4\pi(2m + 1)(m + 1)}$$

$$= 2/\pi \qquad m = 1 \tag{20}$$

$$\to 3/\pi \qquad m \to \infty.$$

Most interesting is the functional curve of the ARE versus m plotted in Fig. 1. Starting at 0.637 and rising to 0.955 at infinity, by $m = 6$ the ARE has reached a value of 0.863.

---

[2] S_W^N has been used both for the mixed statistic and for the original to save subscripts. The meaning should be clear from the context.
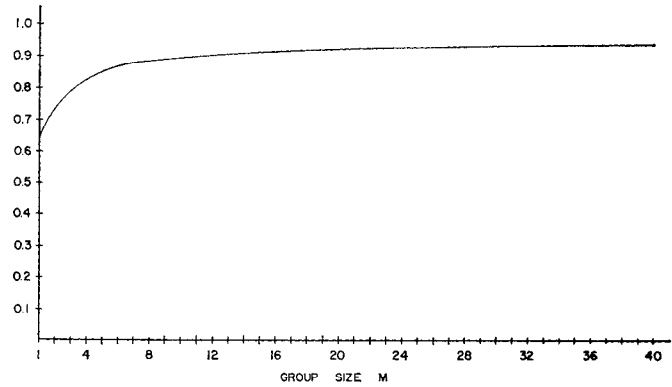

Fig. 1. Asymptotic relative efficiency versus group size for mixed Wilcoxon.

TABLE I

Type	ARE	N	Work
1) Sample Mean Test	1.000	22 400	$0(2.2 \times 10^4)$
2) Wilcoxon Test	0.955	23 500	$0(5.5 \times 10^8)$
3) 6-point Mixed Wilcoxon Test	0.863	26 000	$0(1.6 \times 10^5)$
4) Sign Test	0.637	35 000	$0(3.5 \times 10^4)$

Consider what this means in a practical situation. Let N be the number of data points for student's test required for α and β for:

$$H : x_i \sim N(0, \sigma)$$

$$K : x_i \sim N(\mu, \sigma) \tag{21}$$

$$N = (\Phi^{-1}(1 - \alpha) - \Phi^{-1}(1 - \beta))^2 \frac{\sigma_x^2}{\mu_x^2}$$

for $\alpha = 0.001$, $\beta = 0.95$, $A^2 = 22.4$

$$N = 22.4 \, \frac{\text{NOISE POWER}}{\text{SIGNAL POWER}}. \tag{22}$$

Suppose the above ratio is 30 dB. Table I then summarizes the average number of samples required for four tests for Gaussian alternatives and indicates the order of magnitude of the work involved. Of the non-parametric tests, the 6-point mixed Wilcoxon is clearly superior to the sign test in efficiency and to the Wilcoxon test in the amount of work that is required.

When digital hardware is utilized to implement the mixed Wilcoxon, a real-time processor is possible for typical rates of 2 MHz. A single $(m - 1)$-tap delay line and $\frac{1}{2}m(m + 1)$ threshold devices plus a binary adder are sufficient. A detector for $m = 3$ is shown in Fig. 2. Since the output of the binary counter can be taken in parallel, no analog-to-digital conversion is necessary before transmission. In addition, the information rate for transmission to a central processing point is reduced in the same manner as in the sign detector, a desirable feature for sonobouy detection.
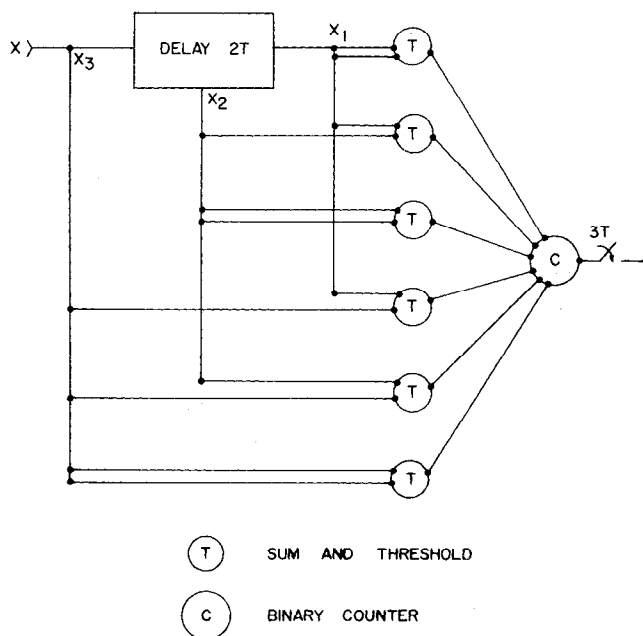
(T)   SUM AND THRESHOLD

(C)   BINARY COUNTER

Fig. 2.

IV. Two-Sample Tests

While one-sample tests are very useful in giving information about mean location changes (or more generally when the signal is of any known form), they are not very useful when the signal is unknown. If two samples with independent noises are available, the presence of a common signal can be detected by testing for correlation or some other statistical test on dependence. A common method is polarity coincidence correlation (PCC).[5]-[8] The basis statistic for PCC is:

$$S^N_{PCC} = \sum_{i=1}^{m} \text{sgn } x_i \text{ sgn } y_i \qquad (23)$$

or alternatively

$$S^N_{PCC} = \frac{1}{2} \sum_{i=1}^{m} u(x_i y_i) \qquad (24)$$

where the difference is a shift in the mean. Although such processing is extremely simple, the ARE is about $(2/\pi)^2$ with respect to the parametric detector which measures the sample correlation for Gaussian inputs, and the ARE with respect to the ideal Neyman–Pearson scheme is $\frac{1}{2}(2/\pi)^2$.[5] Several other statistics are of a similar type. The Kendall τ[9] is one such example:

$$S^N_\tau = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{2}\{1 + \text{sgn } (x_i - x_j) \text{ sgn } (y_i - y_j)\}. \qquad (25)$$

The ARE for this test with respect to the correlator is $9/\pi^2$. Then the ARE of the Kendall detector with respect to the PCC is 9/4. This is a large improvement, but is accompanied by a good deal of additional work.

The success of the mixed method for the sign test gives rise to the hope for a similar result here. Calculations

show that for a test using

$$S^N_\tau = \sum_{i=0}^{p-1} S^m_\tau(i) \qquad (26)$$

where

$$S^m_\tau(i) = \sum_{j=1}^{m} \sum_{k=j+1}^{m} \frac{1}{2}\{1 + \text{sgn } (x_{im+j} - x_{im+k})$$
$$\cdot \text{sgn } (y_{im+j} - y_{im+k})\} \qquad (27)$$

the ARE with respect to the correlator for Gaussian alternatives is:

$$\text{ARE } (d_1, d_2) = \left[\frac{9(m-1)}{2(2m+5)}\right]\left(\frac{2}{\pi}\right)^2. \qquad (28)$$

The functional curve of the ARE of PCC compared to the m-data point, summed Kendall test is interesting (see Fig. 3). Starting at a value of 1/2 for $m = 2$, it increases slowly, reaching 27/26 at $m = 4$ and 1.62 at $m = 10$. At $m = 29$, it is 2. For $m = 10$, only 5/8 as many samples are required for the Kendall mixed statistic as for the PCC detector. One possible implementation is shown in Fig. 4. $m(m - 1)$ threshold devices are required and $m(m - 1)/2$ exclusive-OR devices. All can be constructed from diodes and simple differential amplifier circuitry, which is very commonly microminaturized.

In some instances it may be desirable to utilize a test for randomness instead of one for independence. In this case the two-sample Mann–Whitney[10],[11] appears to be the "best" distribution-free test. It is well known that the Mann–Whitney has an ARE of $3/\pi$ with respect to the two-sample student's t test. Again we are interested in the sum of p observations of the Mann–Whitney operating on two samples of m data points:

$$S^m_{MW}(k) = \sum_{i=1}^{m} \sum_{j=1}^{m} \text{sgn } (x_{km+i} - y_{km+j}) \qquad (29)$$

$$S^N_{MW} = \sum_{k=0}^{p-1} S^m_{MW}(k) \qquad (30)$$

The ARE for this case is: $3/\pi(2m/2m + 1)$, $m \geq 1$ (see Fig. 5). By the time $m = 5$, the ARE is about 91 percent of the full value or 0.864 with respect to the Mann–Whitney. For any continuous distribution, the worst possible ARE for the Mann–Whitney or the Wilcoxon has been shown to be 0.864.[12] The preceding analysis remains valid for these distributions providing the factor $3/\pi$ is replaced by 0.864. This means that for any continuous distribution the ARE with respect to student's t test should not fall below $(0.864)^2$ for the five-data-point mixed statistic. An implementation for this statistic involves more threshold units than the others mentioned previously, requiring m^2 threshold gates in addition to the binary adder. An implementation for $m = 3$ is given in Fig. 6.
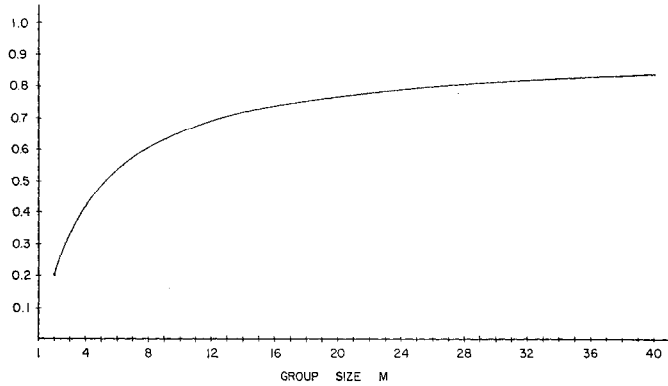
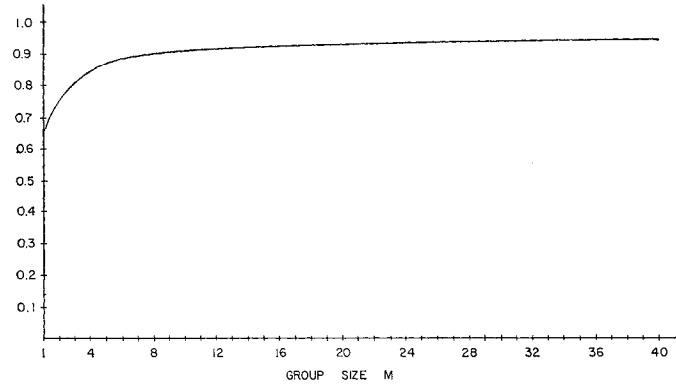Fig. 3. Asymptotic relative efficiency versus group size
for mixed Kendall.

Fig. 5. Asymptotic relative efficiency versus group size
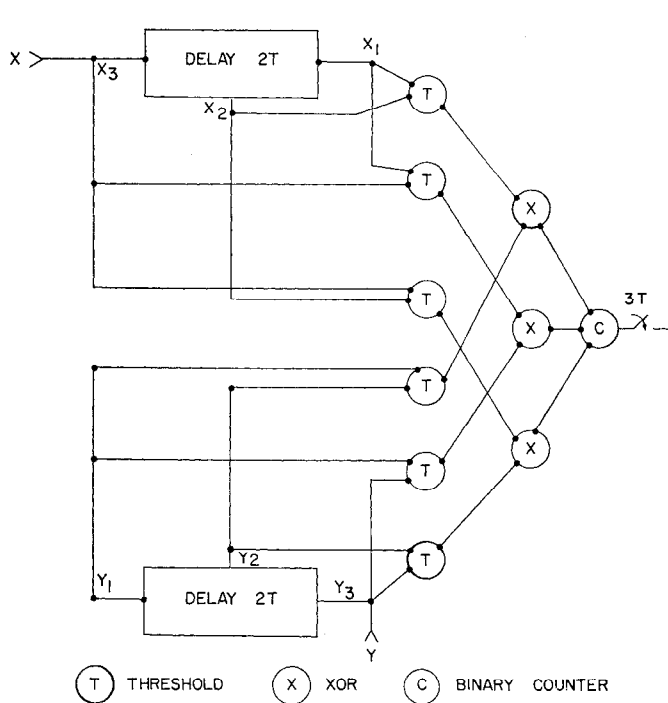for mixed Mann-Whitney.

Fig. 4.

Fig. 6.

Analogous to the normal scores test for one sample is the Fisher–Yates: Terry–Hoeffding test (FYTH) for two samples.[13] Suggested by Fisher and Yates, it remained for Terry[14] to show the optimal properties. Chernoff and Savage[15] have shown that for hypotheses and alternatives with absolutely continuous densities the ARE of the FYTH against student's two-sample t must always be equal or greater than 1 with the equality holding for the case of normal alternatives. Capon[16] has shown that the test is the locally most powerful rank test for Gaussian alternatives. With all these advantages, one asks: "Why has it not been implemented?" A glance at the statistic is sufficient to show why. The FYTH statistic is:

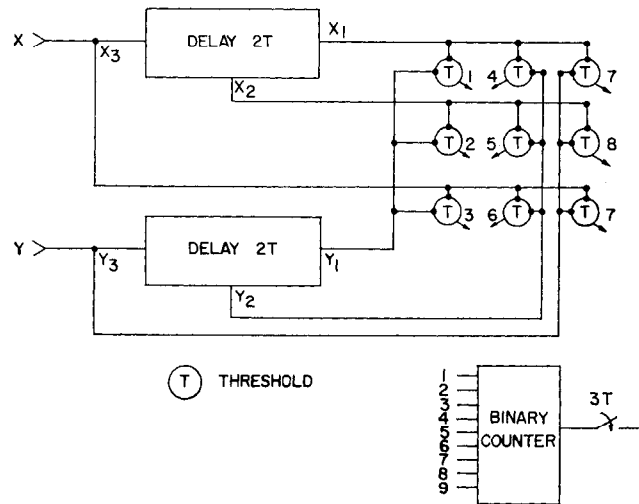$$S_{\text{FYTH}}^{N} = \frac{1}{N} \sum_{i=1}^{2N} a_i z_{i,s} \qquad (31)$$

where a_i is the expected value of the ith smallest order statistic drawn from the absolute normal population of zero mean and unit variance. $z_{i,s}$ is the 1(0) function which has value 1 if the ith smallest observation is drawn from the first sample. It is immediately clear that for large N, the implementation is a horrendous problem; N coefficients must be provided, $2N$ items must be ranked, $O(4N^2)$, the proper coefficients must be selected, and then a sum must be computed.

As expected, the mixed statistic proves itself considerably easier to implement. The test will consist of the sum of p observations of the statistic operating on m data points of two samples. In Appendix I, the ARE of this mixed statistic is calculated with respect to the two-sample student's t test as:

$$\text{ARE } (d_1, d_2) = \frac{2m}{2m - 1} \sum_{s=1}^{2m} \{E(s, 2m)\}^2 \qquad (32)$$

where

$$E(s, 2m) = a_s.$$

Unfortunately, there is no simple interpretation of this expression. Ruben has shown[17] that the resulting value

may be approximated by hyperspherical simplices to within 0.1 percent for $m = 100$. In Fig. 7, the ARE is plotted. Observe that between 15 and 20 samples are required for a high ARE. Lehmann and Hodges[18] have discussed the advantages of the Mann–Whitney as compared to the FYTH and have concluded that for probability distributions with heavy tails the Mann–Whitney has a higher efficacy.

Because the mixed statistic requires only a small number of comparisons for a relatively high efficiency, an implementation, as in Fig. 8, is considerably more complicated than the Mann–Whitney or the Kendall τ since pure digital hardware cannot be used, but the greater efficiency that it possesses may warrant its use. The system then requires $(2m - 1)m$ threshold gates, m binary adders of size $(2m - 1)m$, a diode selector matrix for selecting coefficients, $2m$ coefficient generators, an analog summer, and an A-D converter for transmission. By using a commutator and hold circuitry, one adder and a simpler selector matrix may be used.

APPENDIX I

ONE-SAMPLE TESTS

Let x_i, $i = 1, \cdots, N$ be independent samples of a random process. To test the hypothesis-alternative pair:

$$H : x_i = n_i \qquad \rbrace \; n_i \text{ Gaussian, mean 0, variance } \sigma^2$$
$$\text{vs. } K : x_i = \mu_0 + n_i \rbrace \; \mu_0 \text{ is a positive constant.}$$

The optimum Neyman–Pearson detector calculates

$$\sum_{i=1}^{N} x_i = S^N. \tag{33}$$

Let N_G be the number of samples required for false alarm rate α and probability of detection β. In the limit as $N_G \to \infty$, when $\mu_0 \to 0$ and since $\text{Var}_K (S^N) = \text{Var}_H (S^N)$,

$$\Phi^{-1}(1 - \alpha) - \Phi^{-1}(1 - \beta) = A = \frac{\sqrt{N}_G \mu_0}{\sigma} \tag{34}$$

$$N_G = \frac{\sigma^2}{\mu_0^2} A^2 \tag{35}$$

where Φ^{-1} is the inverse to the normal distribution function of zero mean and unit variance. For the nonparametric sign detector

$$S_S^N = \sum_{i=1}^{N} \text{sgn } x_i \tag{36}$$

under the same H and K:

$$E_K(S_S^N) = N_S(p - q) = N(1 - 2\Phi(-\mu_i/\sigma_i)) \tag{37}$$

where $p = Pr(x_i > 0)$ and $q = 1 - p$

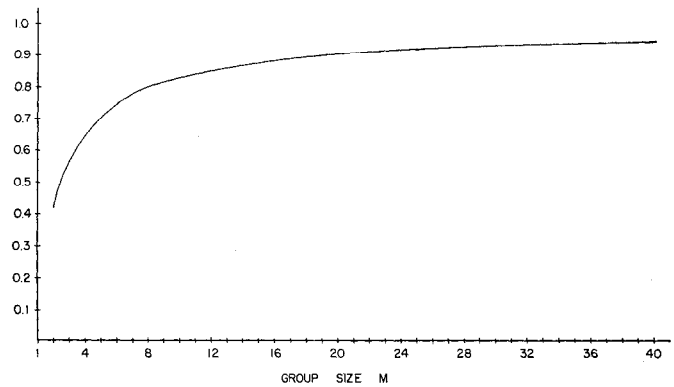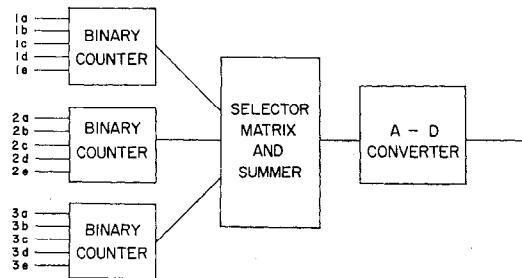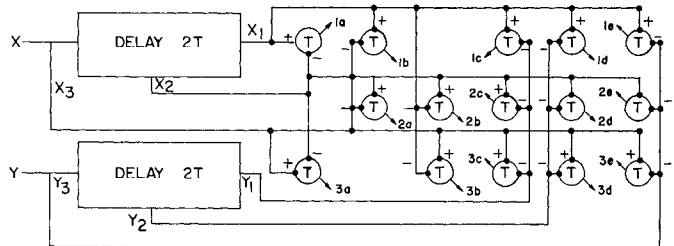$$\text{Var}_H (S_S^N) = N_S(\tfrac{1}{4}) \tag{38}$$

Fig. 7. Asymptotic relative efficiency versus group size for mixed Fisher-Yates: Terry-Hoeffding.

(T)   GATED   THRESHOLD

Fig. 8.

when

$$\mu_0 \to 0 \qquad E_k(S_S^N) \Rightarrow \frac{N}{\sqrt{2\pi}} \frac{\mu_0}{\sigma} \tag{39}$$

and

$$N_S = \frac{2}{\pi} \frac{\sigma^2}{\mu_0^2} A \tag{40}$$

$$\text{ARE } (d_S, d_G) = \frac{N_s}{N_G} = \frac{2}{\pi}. \tag{41}$$

For the Wilcoxon one-sample detector:

$$\sum_{i=1}^{N} \sum_{j=i}^{N} u(x_i + x_j) = S_W^N \tag{42}$$

$$E_K(S_W^N) = Np_2 + \frac{N(N - 1)}{2} p_1 \tag{43}$$

$$\text{Var}_H (S_W^N) = \frac{N(2N + 1)(N + 1)}{24} \tag{44}$$

where: $p_2 = Pr(x_i > 0)$ and $p_1 = Pr(x_i + x_j > 0)$.

Making a series expansion yields

$$E_K(S_W^N) - E_H(S_W^N) = \frac{N(N - 1 + \sqrt{2})\mu_0}{\sqrt{4\pi}\sigma} \quad \text{for } \mu_0/\sigma \text{ small.} \tag{45}$$

Then for the mixed statistic based on the Wilcoxon

$$S_W^N = \sum_{i=1}^{p} (S_W^m - E(S_W^m/\mu_0 = 0)) \tag{46}$$

and for $pm = N$,

$$E(S_W^N) = N \frac{(m - 1 + \sqrt{2})}{\sqrt{4\pi}\sigma_i} \mu_i \tag{47}$$

$$\text{Var } (S_W^N) = N \frac{(2m + 1)(m + 1)}{24} \tag{48}$$

$$\text{ARE } (d_G, d_W) = \frac{24(m - 1 + \sqrt{2})^2}{(2m + 1)(m + 1)4\pi} \tag{49}$$

which agrees as $m \rightarrow \infty$: ARE $(d_G, d_W) = 3/\pi$ and for $m = 1$: ARE $(d_G, d_W) = 2/\pi$, the result for the sign detector.

APPENDIX II

TWO-SAMPLE TESTS

Here two parametric test results are required: the number of observations for the student's t, two-sample test, and the number required for the correlator. The hypothesis-alternative pair for the first detector, d_{st}, is

$H : \{x_i\}, \{y_i\}$ are statistically independent with bivariate normal distribution

$$F(x) = F(y)$$

$K : \{x_i\}, \{y_i\}$ are statistically independent with

$$F(x) = F(y - \theta), \quad \theta = \text{positive constant.}$$

The student's t statistic is[12]:

$$\frac{t^2}{2m - 2} = \frac{(\bar{x} - \bar{y})^2 \frac{m_1 m_2}{2m}}{m_1 s_1^2 + m_2 s_2^2}. \tag{50}$$

This is the optimum parametric test for determining location shift with a Gaussian model of unknown variance where m_1 is the number of observations in the first sample, m_2 in the second, $m_1 + m_2 = m$ and s_1^2 and s_2^2 are the respective sample variances. In this case s_1 and s_2 are known and $m_1 = m_2 = m$ so

$$\frac{t^2}{2m - 2} = \frac{1}{2} \frac{(\bar{x} - \bar{y})^2}{(s_1^2 + s_2^2)} \tag{51}$$

if $s_1^2 = s_2^2 = \sigma^2$

$$= \frac{2}{m^2\sigma^2} \left(\sum_{i=1}^{m} (x_i - y_i) \right)^2. \tag{52}$$

This tends to normality with increasing m with mean θ and variance $2\sigma^2/m$.

The second test d_c is a correlation test which for known variance and

$H_2 : \{x_i\}, \{y_i\}$ statistically independent Gaussian

$$\sigma_x = \sigma_y = \sigma_i, \quad \mu_x = \mu_y = 0, \quad \sigma_{xy} = 0$$

$K_2 : \{x_i\}, \{y_i\}$ correlated Gaussians

$$\sigma_x = \sigma_y = \sigma_i, \quad \mu_x = \mu_y = 0, \quad \sigma_{xy}^2 = \rho\sigma_x\sigma_y$$

is

$$S_c^m = \sum_{i=1}^{m} \frac{x_i y_i}{\sigma_i^2}. \tag{53}$$

Then

$$E_{K_2}(S_c^m) = \rho m \tag{54}$$

$$\text{Var}_{H_2} [S_c^m] = m \tag{55}$$

and the limiting distribution for $S_c^m/k \sim N(m\rho, m)$.
The Kendall τ statistic is:

$$S_\tau^m = \frac{2}{m(m - 1)} \sum_{i<j}^{m}\sum h_{ij} \tag{56}$$

where

$$h_{ij} = \frac{1}{2}\{1 - \text{sgn } (x_i - x_j) \text{ sgn } (y_i - y_j)\} \tag{57}$$

$$E_{K_2}(S_\tau^m) = \left[1 - \frac{2}{\pi} \sin^{-1} \rho \right] \tag{58}$$

and

$$|E_{K_2}(S_\tau^m) - E_{H_2}(S_\tau^m)| \simeq \frac{2}{\pi} \rho \quad \text{for small } \rho. \tag{59}$$

Kendall[9] gives the variance of S_τ^m as:

$$\text{Var}_{H_2} S_\tau^m = \frac{2(2m + 5)}{9m(m - 1)}. \tag{60}$$

Let $N = pm$. The statistic

$$S_\tau^N = \frac{1}{p} \sum_{i=1}^{p} S_\tau^m(i). \tag{61}$$

Then

$$E(S_\tau^N) \Rightarrow \frac{2\rho}{\pi} \quad \text{for small } \rho \tag{62}$$

and

$$\text{Var } (S_\tau^N) = \frac{2(2m + 5)}{9N(m - 1)}. \tag{63}$$

Then

$$\text{ARE } (d_c, d_\tau) = \frac{4 \cdot 9(m - 1)}{(\pi)^2(2m + 5)} = \left(\frac{2}{\pi}\right)^2 \frac{9(m - 1)}{2(2m + 5)} \tag{64}$$

which for

$$m \rightarrow \infty \Rightarrow \frac{9}{\pi^2}.$$

The statistic for PCC is

$$S_{\text{PCC}}^N = \sum_{i=1}^N \left(u(x_i y_i) - \tfrac{1}{2} \right) \tag{65}$$

$$E_{K_2}(S_{\text{PCC}}^N) = \frac{N}{\pi} \sin^{-1} \rho \tag{66}$$

$$\text{Var}_{H_2}(S_{\text{PCC}}^N) = \frac{N}{2}. \tag{67}$$

Then the ARE $(d_c, d_{\text{PCC}}) = (2/\pi)^2$. $\tag{68}$

If, instead of a test for dependence, a test for homogeneity is desired, i.e., H_1 vs. K_1, the Mann–Whitney or Fisher–Yates: Terry–Hoeffding tests are used. Here it is expected that one channel will contain noise alone. The Mann–Whitney test statistic for m samples in one channel and n in the other is:

$$S_{MW}^{m,n} = \sum_{i=1}^m \sum_{j=1}^n u(x_i - y_j) \tag{69}$$

has:

$$\text{Var}(S_{MW}^{m,n}) = \frac{mn}{12}(m + n + 1) \tag{70}$$

and incremental expectation

$$E_{K_1}(S_{MW}^{m,n}) - E_{H_1}(S_{MW}^{m,n}) = \frac{nm}{2\sqrt{\pi}} \frac{\mu}{\sigma}. \tag{71}$$

The sum $N = pm$, and

$$S_{MW}^N = \sum_{i=1}^p S_{MW}^{m,n} \tag{72}$$

For $n = m$:

$$\frac{N_t}{N_{MW}} = \left\{ \frac{\dfrac{\sqrt{2}\sigma}{\mu}}{\dfrac{\sqrt{\pi}\sqrt{2m+1}\sqrt{m}}{\sqrt{3m}} \dfrac{\sigma}{\mu}} \right\}^2 = \frac{3}{\pi} \frac{2m}{2m+1} \tag{73}$$

$$\text{ARE}(d_{t^2}, d_{MW}^m) = \left(\frac{3}{\pi} \right)\left(\frac{2m}{2m+1} \right). \tag{74}$$

Alternatively, the Fisher–Yates: Terry–Hoeffding statistic may be used. This statistic is for equal sample sizes:

$$C_1^m = \frac{1}{m} \sum_{i=1}^m E[x_i, 2m]z_i \tag{75}$$

where $E[x_i, 2m]$ is the expected value of the ith smallest observation of $2m$ drawn from a normal population of zero mean and unit variance and z_i is the indicator function having the value 1 if the ith smallest observation is from the first sample.

Capon[16] has shown that the efficacy of this detector using K_1 and H_1 is simply:

$$E^* = \inf G_{\varphi_0} \tag{76}$$

and that the variance is:

$$\text{Var}(C_1^m) = \frac{(1-\gamma)}{\gamma M} \inf G_{\varphi_0} \tag{77}$$

where

$$\inf G_{\varphi_0} = E\left\{ \left(\frac{\partial}{\partial \theta} \ln g_\varphi(\theta) \Big|_{\theta=0} \right)^2 \right\} \tag{78}$$

and $\gamma M = m$, $N = mp$. Then let

$$e = \sum_{j=1}^p C_{1j}^m. \tag{79}$$

$$\text{Var}(e) = p \, \text{Var}(C_1^m) \tag{80}$$

$$E(e) = p E(C_1^m) \tag{81}$$

$$E_\varphi^* = \lim_{M \to \infty} \left\{ \frac{\frac{\partial}{\partial \Delta} E(e)}{\text{Var}^{1/2}(e)} \right\}^2 \frac{1}{\gamma(1-\gamma)N} \tag{82}$$

$$\Delta = kM^{-1/2} \tag{83}$$

$$\chi = kN^{-1/2} \tag{84}$$

$$\Delta = p^{1/2}\chi \tag{85}$$

$$\frac{\partial}{\partial \Delta} = \frac{\partial}{\partial \chi} \cdot \frac{dx}{d\Delta} \tag{86}$$

$$= \frac{\partial}{\partial \chi} p^{-1/2} \tag{87}$$

$$= \lim_{M \to \infty} \left[\frac{p^{1/2} \frac{\partial}{\partial \chi} E(C_1^m)}{p^{1/2} \text{Var}^{1/2}(C_1^m)} \right]^2 \frac{1}{\gamma(1-\gamma)N} \tag{88}$$

$$= \lim_{N \to \infty} \left[\frac{\frac{\partial}{\partial \chi} E(C_1^m)}{\text{Var}^{1/2}(C_1^m)} \right]^2 \frac{1}{\gamma(1-\gamma)N} \tag{89}$$

$$= \inf G_{\varphi_0}. \tag{90}$$

From Kendall:[12]

$$\text{Var}_{H_0}(C_1^m) = \frac{(1-\gamma)}{\gamma(M)(M-1)} \sum_{S=1}^M \{E(S, M)\}^2 \tag{91}$$

$$\inf G_{\varphi_0} = \frac{M}{M-1} \left\{ \frac{1}{M} \sum_{S=1}^M \{E(S, M)\}^2 \right\}. \tag{92}$$

From Ruben:[17]

$$\bar{G} = \frac{1}{M} \sum_{i=1}^M [E(i, M)]^2$$
$$\simeq \frac{M-1}{M+1}\alpha_1 + \frac{(M-1)(M-2)(M-3)}{(M+1)(M+2)(M+3)}\alpha_3 \cdots \tag{93}$$

$$E_e^* \simeq \frac{2m}{2m-1} \bar{G} \tag{94}$$

$$\text{ARE}(d_{t^2}, d_{\text{FYTH}}) \simeq \frac{2m}{2m-1} \bar{G}. \tag{95}$$

REFERENCES

[1] E. J. G. Pitman, Lecture notes on nonparameteric statistical inference, Columbia University, New York, N. Y., Spring 1948.

[2] G. E. Noether, "On a theorem of Pitman," *Ann. Math. Statistics*, vol. 26, pp. 64–68, 1955.

[3] J. E. Walsh, *Handbook of Nonparametric Statistics*, vols. I and II. Princeton, N. J.: Van Nostrand, 1962, 1965.

[4] J. W. Carlyle and J. B. Thomas, "On nonparametric signal detectors," *IEEE Trans. on Information Theory*, vol. IT-10, pp. 146–152, April 1964.

[5] S. S. Wolff, J. B. Thomas, and T. R. Williams, "The polarity-coincidence correlator; a nonparametric detection device," *IRE Trans. on Information Theory*, vol. IT-8, pp. 5–9, January 1962.

[6] M. Kanefsy and J. B. Thomas, "A comparison of some detectors for non-Gaussian inputs," Commun. Lab., Dept. of Elec. Engrg., Princeton University, Princeton, N. J. Tech. Rept. 3, February 1964.

[7] ——, "On detection by polarity coincidence correlation using dependent samples," Commun. Lab., Dept. of Elec. Engrg., Princeton University, Princeton, N. J., Tech. Rept. 6, May 1964.

[8] M. Kanefsky, "Detection of weak signals with polarity coincidence arrays," *Proc. 1966 Internat'l Conf. on Information Theory;* also published in *IEEE Trans. on Information Theory*, vol. IT-12, pp. 260–268, April 1966.

[9] M. G. Kendall, *Rank Correlation Methods*, 3rd ed. New York: Hafner, 1963.

[10] H. B. Mann and D. R. Whitney, "On a test of whether one or two random variables is stochastically larger than the other," *Ann. Math. Statistics*, vol. 18, p. 50, 1947.

[11] J. Capon, "Nonparametric methods for the detection of signals in noise," Columbia University, New York, N. Y., Tech. Rept. T−1/N, March 1959.

[12] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 2. New York: Hafner, 1961.

[13] D. A. S. Fraser, *Nonparameteric Methods in Statistics*. New York: Wiley, 1957.

[14] M. E. Terry, "Some rank-order tests which are most powerful against specific alternatives," *Ann. Math. Statistics*, vol. 23, pp. 346–366, 1952.

[15] H. Chernoff and I. R. Savage, "Asymptotic normality and efficiency of certain nonparametric test statistics," *Ann. Math. Statistics*, vol. 29, pp. 972–994, 1958.

[16] J. Capon, "Asymptotic efficiency of certain locally most powerful rank tests," *Ann. Math. Statistics*, vol. 32, pp. 88–100, 1961.

[17] H. Ruben, "On the sum of squares of normal scores," *Biometrics*, vol. 43, pp. 456–458, 1956.

[18] E. L. Lehman and J. L. Hodges, Jr., "Comparison of normal scores and Wilcoxon tests," *4th Berkeley Symp. on Mathematical Statistics and Probability*. Berkeley, Calif.: University of California Press, 1961, pp. 316–317.

Differentiation of Karhunen-Loève Expansion and Application to Optimum Reception of Sure Signals in Noise

T. T. KADOTA

Abstract—The first part of this paper is concerned with differentiation of the Karhunen–Loève expansion of a stochastic process. In particular, we establish that the expansion series can be differentiated term by term while retaining the same sense of convergence, if the covariance $R(s, t)$ has a continuous second partial derivative and the sample function $x(t)$ is almost surely differentiable. The result can be generalized to the case of higher-order differentiation. Namely, if $(\partial^{2n}/\partial s^n \partial t^n) R(s, t)$ is continuous and $x(t)$ has the nth derivative $x^{(n)}(t)$ almost surely, then the series can be differentiated term by term n times, and the resultant series converges in the stochastic mean to $x^{(n)}(t)$ uniformly in t.

In the second half, the above result is applied to the problem of optimum reception of binary signals in Gaussian noise. Suppose the binary sure signals are $m_1(t)$ and $m_2(t)$ and the noise covariance is $R(s, t)$. Then we prove the well-known conjecture that the optimum receiver correlates the observable waveform with the solution $g(t)$ of the integral equation $\int R(s, t)g(s) \, ds = m_2(t) - m_1(t)$ *even if* the solution contains δ-functions and their derivatives. This result can be generalized to the case of M-ary sure signals.

I. DIFFERENTIATION OF KARHUNEN–LOÈVE EXPANSION

SUPPOSE we have a second-order stochastic process $\{x(t), -T \le t \le T\}$ with zero mean and a continuous covariance $R(s, t)$. Then, according to the Karhunen–Loève theorem,[1] $x(t)$ can be expanded in terms of the orthonormalized eigenfunctions $\psi_i(t)$, $i = 0, 1, 2, \cdots$, of R as follows:

$$x(t) = \sum_i \xi_i(x) \psi_i(t), \tag{1}$$

where the coefficients are defined almost surely by

$$\xi_i(x) = \int_{-T}^{T} x(t) \psi_i(t) \, dt, \qquad i = 0, 1, 2, \cdots, \tag{2}$$

and the series converges in the stochastic mean to $x(t)$ uniformly in t, namely,

$$\lim_{n \to \infty} E\left\{ \left| x(t) - \sum_{i=0}^{n} \xi_i(x) \psi_i(t) \right|^2 \right\} = 0$$

uniformly with respect to t, $-T \le t \le T$.