The Annals of Statistics
1998, Vol. 26, No. 6, 2321–2338

DIRECTIONAL TESTS FOR ONE-SIDED ALTERNATIVES IN MULTIVARIATE MODELS[1]

BY ARTHUR COHEN AND HAROLD B. SACKROWITZ

Rutgers University

Consider one-sided testing problems for a multivariate exponential family model. Through conditioning or other considerations, the problem oftentimes reduces to testing a null hypothesis that the natural parameter is a zero vector against the alternative that the natural parameter lies in a closed convex cone $\mathscr{C}$. The problems include testing homogeneity of parameters, testing independence in contingency tables, testing stochastic ordering of distributions and many others. A test methodology is developed that directionalizes the usual test procedures such as likelihood ratio, chi square, Fisher, and so on. The methodology can be applied to families of tests where the family is indexed by a size parameter so as to enable nonrandomized testing by p-values. For discrete models, a refined family of tests provides a refined grid for better testing by p-values. The tests have essential monotonicity properties that are required for admissibility and for desirable power properties. Two examples are given.

1. Introduction A new general testing methodology is presented for testing one-sided alternatives in multivariate exponential family models. Since there rarely exist optimal tests for such complicated models, the literature usually contains many ad hoc tests, chi-square based tests and the likelihood ratio test (when feasible). In many instances these methods do yield reasonable procedures, but they are also often unsatisfactory. We uncover shortcomings of some existing approaches. One advantage of the new methodology, which can be applied to a wide variety of testing problems, is that it avoids the shortcomings of existing approaches. We show how to apply the general methodology to the particular problem of testing equality of multinomial distributions versus the stochastic order alternative. This problem motivated the study. The problem itself is an important one in medical studies as well as social science studies. Moses, Emerson, and Hosseini (1984) identified 27 instances of $2 \times J$ tables in a survey of articles in volume **306** (1982) of *The New England Journal of Medicine*. Cohen and Sackrowitz (CS) (1998a) argue that in many $2 \times J$ tables the notion of testing whether a treatment is "better" than a control should often be approached by testing a stochastic order alternative. We will see that, in spite of the fact that many tests have been proposed and used for this problem, none is without some shortcomings.

Received December 1996; revised August 1998.

[1] Supported in part by NSF Grant 96-18715 and NSA Grant MDA 904-98-1-0506.

AMS 1991 *subject classifications*. 62H15, 62H17.

Key words and phrases. Contingency tables, multivariate exponential family, stochastic order, Wilcoxon–Mann–Whitney test, likelihood ratio order, independence, order restricted inference, Fisher's test, peeling.

One large class of applications is to ANOVA models for testing the null hypothesis of homogeneity of parameters versus the alternative that the parameters lie in a closed convex cone. See CS (1994) and Robertson, Wright and Dykstra (RWD) (1988) for examples of such problems. In contingency table models there are a wide variety of one-sided testing problems. For example, in a two-way ordered table for the full multinomial model, one might test independence against all local log odds ratios nonnegative (often called trend). In the $2 \times J$ product multinomial model, one might test equality of multinomial distributions against the alternative that one distribution is stochastically larger than the other. In analogy to one-dimensional problems we refer to directional situations (such as the two above) as one-sided alternatives in contrast to non-directional or two-sided alternatives. In the first example, a two-sided alternative is "not independence." In the second example, a two-sided alternative is "distributions not equal."

Let $\boldsymbol{\nu}$ be the natural parameter of a multivariate exponential family. Let $\boldsymbol{\nu} \in \mathscr{N}$, where $\mathscr{N}$ denotes the natural parameter space. The approach focuses on the fundamental problem of testing the null hypothesis H_0: $\boldsymbol{\nu} = \mathbf{0}$ versus $H_1 - H_0$ where H_1: $\boldsymbol{\nu} \in \mathscr{C}$, and $\mathscr{C}$ is a closed convex cone and $\mathscr{C} \subset \mathscr{N}$. (This formally defines what we mean by a one-sided testing problem. A two-sided alternative is H_2: $\boldsymbol{\nu} \neq 0$.) This is, in a sense, the canonical form of these problems. We stress that most of the important applications do deal with composite null hypotheses. The simple null cases arise most often by conditioning and could arise by invariance considerations. Complete class and some admissibility results for this model appear, for example, in Eaton (1970) and Ledwina (1978).

The first step in our approach is to choose a "good" two-sided test procedure for testing H_0: $\boldsymbol{\nu} = \mathbf{0}$ versus H_2: $\boldsymbol{\nu} \neq \mathbf{0}$. In a two-sided problem there typically exists a number of satisfactory test procedures. Intuitively, our method serves to directionalize the two-sided test in such a way that the resulting test will be in a complete class, often be admissible and also satisfy any restrictions imposed by practical considerations of the original, usually unconditional problem.

In the discrete case, a refinement to the method is developed so that more levels of significance can be accommodated without randomization. In practice this results in more precise p-values, that is, a finer grid of p-values. This refinement of the methodology is effective in cases where the original test is two sided or even one sided. The refined test methodology is akin to the notion of "peeling." Peeling is discussed by Green (1985). Our refined methodology will be called directed vertex peeling (DVP) and will be described in the next section.

In Section 2 we make notions precise, develop the method and illustrate the method. Section 3 considers further properties and contains discussion about bounds and implementation. In Section 4 we discuss the method in two examples. The first example concerns a data set for testing equality of three binomial parameters against the alternative that they are decreasing. This is an example in which reasonable test procedures are already available.

Although the application of the new method is straightforward in this case, the example provides an illustration of the model and methodology.

The second example provides a new approach to the problem of testing equality of two multinomial distributions versus the alternative that the second distribution is stochastically larger than the first. Linear tests are often used. Graubard and Korn (1987) discuss the assignment of different weighting systems to be applied to the ordered categories. Using these weights, they suggest a discrete version of the two-sample *t*-test. The above process results in a linear test. Using rank scores as weights results in the Wilcoxon–Mann–Whitney (WMW) test. This latter test is very popular. See, for example, Ralphs and Zimmerman (1993) Emerson and Moses (1985), and Agresti and Finlay (1997). Linear tests will be shown to lack a robustness property. The justification for the effectiveness of our methodology in this example is based on theoretical considerations, practical (intuitive) considerations, and extensive numerical work. In Section 4 we will discuss justifications. An extensive numerical study on this problem appears in CS (1998b). This latter paper contains relevant numerical power and expected p-value computations and also gives a detailed illustration of the methodology for testing stochastic ordering in a 2×3 table.

2. Methodology

We begin with some facts about cones and cone orders. A convex cone is a subset $\mathscr{C} \subseteq \mathbb{R}^k$ such that if $\mathbf{x}, \mathbf{y} \in \mathscr{C}$, $\lambda_1\mathbf{x} + \lambda_2\mathbf{y} \in \mathscr{C}$ for all $\lambda_1, \lambda_2 \geq 0$. The cone $\mathscr{C}$ is pointed if whenever $\mathbf{x} \in \mathscr{C}$, $\mathbf{x} \neq \mathbf{0}$, then $-\mathbf{x} \notin \mathscr{C}$. A closed convex cone $\mathscr{C}$ induces a preordering less than or equal to $[\mathscr{C}]$ as follows: $\mathbf{x} \leq [\mathscr{C}]\mathbf{y}$ if and only if $\mathbf{y} - \mathbf{x} \in \mathscr{C}$ [Marshall and Olkin (1970)]. A function $W(\mathbf{x})$ is nondecreasing with respect to (w.r.t.) the cone $\mathscr{C}$ or is said to be cone order monotone w.r.t. $\mathscr{C}$ (COM $[\mathscr{C}]$) if $W(\mathbf{x}) \leq W(\mathbf{y})$ whenever $\mathbf{x} \leq [\mathscr{C}]\mathbf{y}$. Note that if $\mathscr{C} \subset \mathscr{C}'$ and $W(\mathbf{x})$ is COM $[\mathscr{C}']$, then $W(\mathbf{x})$ is COM $[\mathscr{C}]$, so that increasing the cone size decreases the class of cone order monotone functions.

The (positive) dual of a cone $\mathscr{C}$ is defined as

$$\mathscr{C}^* = \{\boldsymbol{\theta}\colon \mathbf{v}'\boldsymbol{\theta} \geq 0,\ \text{all } \mathbf{v} \in \mathscr{C}\}.$$

Throughout this section the model considered concerns a random $k \times 1$ vector, $\mathbf{X}$, having an exponential family density expressed as

$$h(\mathbf{x})\beta(\boldsymbol{\nu})\exp(\mathbf{x}'\boldsymbol{\nu}). \tag{2.1}$$

Note that the sample space is denoted by $\mathscr{X}$. We test $H_0\colon \boldsymbol{\nu} = \mathbf{0}$ versus $H_1 - H_0$, where $H_1\colon \boldsymbol{\nu} \in \mathscr{C}$ and $\mathscr{C}$ is a closed convex cone.

First we note that a complete class of tests for H_0 versus $H_1 - H_0$ consists of tests which are COM $[\mathscr{C}^*]$ and which have convex acceptance regions. (In the discrete case, randomization is permitted only on boundary points of the acceptance region.) This statement follows from the same argument given by Cohen, Sackrowitz and Samuel-Cahn [(1995), Theorem 3.1] which is based on a theorem of Eaton (1970).

Of course there exist many admissible tests, but admissibility alone does not ensure a desirable test procedure. Other considerations due to the nature of a particular problem may suggest additional compelling restrictions on the choice of test procedures. As will be seen in Example 4.2, such an effect can manifest itself in the requirement that a test be COM $[\mathscr{K}]$ for some $\mathscr{K} \supseteq \mathscr{C}^*$.

The new test method begins with the selection of a test,

$$\phi_A(\mathbf{x}) = \begin{cases} 0, & \text{if } \mathbf{x} \in A, \\ 1, & \text{otherwise,} \end{cases}$$

which is appropriate for testing H_0: $\boldsymbol{\nu} = \mathbf{0}$ versus the "two-sided" alternative H_2: $\boldsymbol{\nu} \neq \mathbf{0}$. We point out that "good" two-sided tests based on principles such as likelihood ratio, chi-square and Fisher are often available. However, at a minimum, A, the acceptance region of the chosen two-sided test, is taken to be convex.

For any pointed convex cone $\mathscr{K} \supseteq \mathscr{C}^*$ we define a new, "one-sided," test for

$$H_0\colon \boldsymbol{\nu} = \mathbf{0} \text{ versus } H_1 - H_0 \text{ where } H_1\colon \boldsymbol{\nu} \in \mathscr{C}, \text{ by}$$

$$\phi_{\tilde{A}}(\mathbf{x}) = \begin{cases} 0, & \text{if } \mathbf{x} \in \tilde{A}, \\ 1, & \text{otherwise,} \end{cases} \tag{2.2}$$

$$\text{where } \tilde{A} = A - \mathscr{K} = \{\mathbf{x}\colon \mathbf{x} = \mathbf{a} - \mathbf{b},\ \mathbf{a} \in A,\ \mathbf{b} \in \mathscr{K}\}.$$

In Figure 1 we illustrate the construction of $\tilde{A}$, given a particular A and a particular $\mathscr{K}$. In words, $\tilde{A}$ consists of all points that can be expressed as $\mathbf{a} - \mathbf{b}$ where $\mathbf{a} \in A$ and $\mathbf{b}$ lies in $\mathscr{K}$. The reader can simply check in Figure 1 that any $\mathbf{x} \notin \tilde{A}$ cannot be expressed as $\mathbf{a} - \mathbf{b}$. One can visualize $\tilde{A}$ as the union over $a \in A$ of the translated cones $a - \mathscr{K}$.

The one-sided test defined in (2.2) has desirable properties. It is easy to verify that the test in (2.2) has a convex acceptance region and is COM $[\mathscr{C}^*]$. In fact, the test in (2.2) is COM $[\mathscr{K}]$, but since $\mathscr{K} \supset \mathscr{C}^*$, any test which is COM $[\mathscr{K}]$ is COM $[\mathscr{C}^*]$. Thus we are assured that the test in (2.2) lies in the complete class.

We remark that tests defined by (2.2) are precisely the projected tests of Cohen, Kemperman and Sackrowitz (1994) for the specialized models assumed in that paper. In that paper, projected tests of two-sided Schur convex tests were shown to have desirable properties. Thus tests here represent an extension of projected tests to a richer class of models.

At this point we assume that the sample space $\mathscr{X}$ is finite. It is desirable to be able to report p-values or to perform tests by using p-values. In order to report p-values we focus on nonrandomized tests and further note that one must have a family of tests with nested acceptance regions, each corresponding to a different size test. See Lehmann [(1986), page 170].

Thus far we have been concerned with a single "good" two-sided test and indicated how to convert it to a one-sided test. We can also consider a nested family of "good" two-sided tests which will be converted to a nested family of one-sided tests. The family of two-sided acceptance regions can often be determined by a test statistic $T(\mathbf{X})$ as follows:

$$A_\alpha = \{\mathbf{x}\colon \mathbf{x} \in \mathscr{X},\ T(\mathbf{x}) < C_\alpha\} \tag{2.3}$$

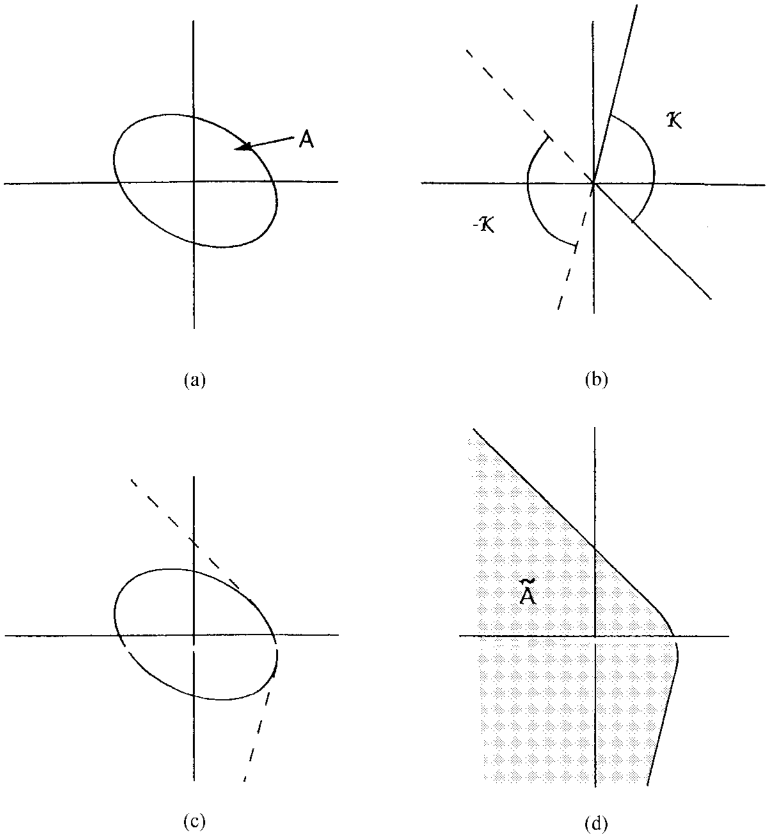


FIG. 1. *Directionalizing a convex set A in the direction of the cone* $\mathscr{K}$. (a) *the convex set A*; (b) *the cone* $\mathscr{K}$ (*and the cone* $-\mathscr{K}$); (c) *finding the points, a, −b*; (d) *the resulting set A.*

as α, the size of the test, varies from 0 to 1. Because of discreteness and limitations to nonrandomized tests, not all values of $\alpha \in [0, 1]$ are used. Examples of $T(\mathbf{X})$'s are Pearson's chi-square, the likelihood ratio test (LRT) statistic, or Fisher, that is, $T(\mathbf{X}) = 1/h(\mathbf{X})$ where $h(\cdot)$ was defined for the multivariate exponential family distribution in (2.1). Fisher's test assigns points to the rejection region according to the probability of points under H_0. That is, points with smallest probability under H_0 enter the rejection region first. See Pagano and Halvorsen (1981) for a Fisher exact test in an $I \times J$ contingency table. We will always take $T(\mathbf{X})$ to be a convex function since this will ensure that A_α is convex.

For a given fixed cone $\mathscr{K}$, the family (2.3) gives rise to the nested family of one-sided tests

$$\tilde{A}_\alpha = A_\alpha - \mathscr{K} = \{\mathbf{x}\colon \mathbf{x} = \mathbf{a} - \mathbf{b},\ \mathbf{a} \in A_\alpha,\ \mathbf{b} \in \mathscr{K}\}. \tag{2.4}$$

(Note that $\tilde{A}_\alpha$ does not have size α even though A_α does.) Let $\mathscr{G}(T(\cdot), \mathscr{K})$ denote this family of directed tests determined by the test statistic $T(\cdot)$ and the cone $\mathscr{K}$.

A larger nested family that contains the nested family of (2.4) can be constructed. The larger family is more desirable since the grid of p-values for the larger family is finer and contains the grid of p-values for the smaller family. This enables more powerful testing through p-values.

The motivation for the forthcoming test construction is fourfold.

1. We wish to utilize the notion of peeling off sample points to assign to the rejection region. We want to peel off points first which, in a sense, reflect the largest departure from the null hypothesis in the *direction* of the alternative hypothesis.
2. Our resulting acceptance region must be convex and the resulting test must be COM $[\mathscr{K}]$ so as to qualify for admissibility and desirability.
3. The construction must produce a nested family of tests which *contains* the nested family $\mathscr{S}(T(\cdot), \mathscr{K})$.
4. Suppose a test procedure depends on a statistic, $T^*(\mathbf{x})$ say, and there were several $\mathbf{x}$ which yielded the same value of T^* (i.e., ties). When the same tie occurs for many sample points, an overly conservative p-value ensues. Our directed vertex peeling method, to be described in the next paragraph, represents a substantial improvement, since many such ties are broken in a manner consistent with (1) above. An example of a situation where the same tie exists for many sample points appears in CS (1992) where another method of tie breaking is presented.

To describe the construction to achieve (1), (2), (3) and (4), we need some terminology. If R is a convex set, we say $\mathbf{x}_0 \in R$ is a $\mathscr{K}$ directed point of R if

$$\mathscr{K}(\mathbf{x}_0) \cap R = \{\mathbf{x}_0\}, \tag{2.5}$$

where $\mathscr{K}(\mathbf{x}_0) = \{x_0\} + \mathscr{K}$. [$\mathscr{K}(\mathbf{x}_0)$ is the cone $\mathscr{K}$ with vertex translated to $\mathbf{x}_0$.]

The family of nested tests are represented by acceptance regions of increasing size α (probability of Type I error) constructed in a stepwise fashion. The first acceptance region is $B_1 = \mathscr{X}$. This corresponds to a test of size $\alpha = 0$. Let $V_1 = \{\text{vertices of the convex hull of } \mathscr{X}\}$. Let $V_1^* = \{\mathbf{x}\colon \mathbf{x} \in V_1,\ \mathscr{K}(\mathbf{x}) \cap B_1 = \mathbf{x}\}$, that is, V_1^* is the collection of $\mathscr{K}$ directed vertices of the convex hull of B_1. Lemma 2.1 below will imply that V_1^* is not empty. Define B_2 by

$$B_2 = B_1 \cap \left\{\mathbf{x}\colon \mathbf{x} \in V_1^*, T(\mathbf{x}) = \max_{\mathbf{v} \in V_1^*} T(\mathbf{v})\right\}^c. \tag{2.6}$$

Thus B_2 is obtained from B_1 by removing the $\mathscr{K}$ directed vertices of the convex hull of B_1 having the largest value of the T statistic. We continue in this fashion, defining

$$\begin{aligned} V_i &= \{\text{vertices of the convex hull of } B_i\}, \\ V_i^* &= \{\text{directed vertices of } V_i\} \quad \text{and} \\ B_{i+1} &= B_i \cap \left\{\mathbf{x}\colon \mathbf{x} \in V_i^*,\ T(\mathbf{x}) = \max_{\mathbf{v} \in V_i^*} T(\mathbf{v})\right\}^c. \end{aligned} \tag{2.7}$$

TABLE 1
Probability distribution of (X_1, X_2)

		x_1			
		0	**1**	**2**	**3**
	2	0.025	0.050	0.100	0.025
x_2	1	0.100	0.200	0.200	0.100
	0	0.025	0.050	0.100	0.025

We continue until all points have been removed. We will refer to this iterative procedure as a directed vertex peeling (DVP) process. Peeling is discussed in Green (1985). The number of steps required, call it M, will depend on the number of points in $\mathscr{X}$ and on the number of ties encountered when maximizing the T statistic among directed vertices. Thus we have a collection of acceptance regions $\mathscr{X} = B_1 \supseteq B_2 \supseteq \cdots \supseteq B_{M+1} = \varnothing$. The corresponding tests are $\phi_i(\mathbf{x}) = 1 - I_{B_i}(\mathbf{x})$. By the nature of the construction, the B_i are convex and the $\phi_i(\mathbf{x})$ are COM $[\mathscr{K}]$. We will denote the family of tests generated in this fashion [based on a given $T(\cdot)$ and $\mathscr{K}$] by $\mathscr{F}(T(\cdot), \mathscr{K})$.

In terms of achieving the goals (1), (2), (3) and (4), we note that (2.6) and (2.7) are defined so that (1) is achieved with T evaluated at directed vertices serving as a measure of distance from H_0. We have already noted that (2) is achieved, and in the next section we will show that $\mathscr{S}(T(\cdot), \mathscr{K}) \subseteq \mathscr{F}(T(\cdot), \mathscr{K})$. In connection with (4) the number of sample points giving rise to the same value of T now result in fewer ties because many of these sample points are not directed vertices. At this point we will give a very simple numerical illustration of the construction of the family $\mathscr{F}(T(\cdot), \mathscr{K})$. For a more realistic and detailed illustration on determining a p-value using this method, we refer the reader to CS (1998b).

Simple illustration of DVP method. Let the random variables X_1, X_2 take on values in $\mathscr{X} = \{(x_1, x_2)\colon x_1 = 0, 1, 2, 3;\ x_2 = 0, 1, 2\}$ with probabilities given in Table 1. Let $\mathscr{K}$ be the first quadrant cone, that is, $\mathscr{K} = \{(x_1, x_2)\colon x_1 \geq 0, x_2 \geq 0\}$. As test statistic we take $T(\mathbf{x}) = (x_1 - 2)^2 + (x_2 - 1)^2$. Table 2 gives the value of T at each point of $\mathscr{X}$. The DVP process is summarized in Table 3. A graphical representation is given in Figure 2.

TABLE 2
Values of $T(x_1, x_2) = (x_1 - 2)^2 + (x_2 - 1)^2$ *on the sample space*

		x_1			
		0	**1**	**2**	**3**
	2	5	2	1	2
x_2	1	4	1	0	1
	0	5	2	1	2

TABLE 3
DVP method, with $T = (x_1 - 2)^2 + (x_2 - 1)^2$, *applied to the model given in Tables* 1 *and* 2

Step $\boldsymbol{i}$	$\boldsymbol{B_i}$	$\boldsymbol{V_i}$	$\boldsymbol{V_i^*}$	**Action**	**Cumulative probability of removed points**[1]
1	$(0,0)(0,1)(0,2)$ $(1,0)(1,1)(1,2)$ $(2,0)(2,1)(2,2)$ $(3,0)(3,1)(3,2)$	$(0,0)(0,2)(3,0)$ $(3,2)$	$(3,2)$	remove (3,2) with $T = 2$	0.025
2	$(0,0)(0,1)(0,2)$ $(1,0)(1,1)(1,2)$ $(2,0)(2,1)(2,2)$ $(3,0)(3,1)$	$(0,0)(0,2)(3,0)$ $(2,2)(3,1)$	$(2,2)$ $(3,1)$	remove (2,2) and (3,1) with $T = 1$	0.225
3	$(0,0)(0,1)(0,2)$ $(1,0)(1,1)(1,2)$ $(2,0)(2,1)(3,0)$	$(0,0)(0,2)(3,0)$ $(1,2)$	$(3,0)$ $(1,2)$	remove (3,0) and (1,2) with $T = 2$	0.300
4	$(0,0)(0,1)(0,2)$ $(1,0)(1,1)(2,0)$ $(2,1)$	$(0,0)(0,2)(2,0)$ $(2,1)$	$(0,2)$ $(2,1)$	remove (0,2) with $T = 5$	0.325
5	$(0,0)(0,1)(1,0)$ $(1,1)(2,0)(2,1)$	$(0,0)(0,1)(2,0)$ $(2,1)$	$(2,1)$	remove (2,1) with $T = 0$	0.525
6	$(0,0)(0,1)(1,0)$ $(1,1)(2,0)$	$(0,0)(0,1)(2,0)$ $(1,1)$	$(2,0)$ $(1,1)$	remove (2,0) and (1,1) with $T = 1$	0.825
7	$(0,0)(0,1)(1,0)$	$(0,0)(0,1)(1,0)$	$(1,0)$ $(0,1)$	remove (0,1) with $T = 4$	0.875
8	$(0,0)(1,0)$	$(0,0)(1,0)$	$(1,0)$	remove (1,0) with $T = 2$	0.975
9	$(0,0)$	$(0,0)$	$(0,0)$	remove (0,0) with $T = 5$	1.000

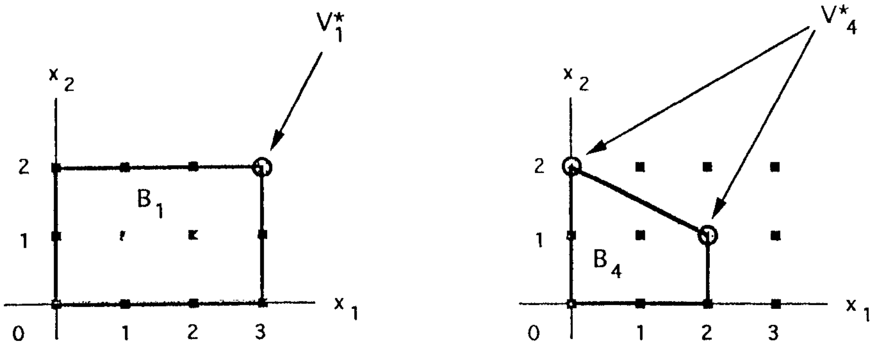
[1] The cumulative probability of a removed point is equal to the p-value associated with that point if it is observed.

As an example of p-value determination for an observed sample point, suppose we observe $(0,2)$. The p-value is 0.325. Note that T for this sample point is 5, the largest possible value of T in the sample space, yet $(0,2)$ does not enter the rejection region by the DVP process until the fourth stage.

In Table 3 note that the column labeled B_i lists a decreasing sequence of acceptance regions for the DVP method using the statistic $T(\mathbf{x}) = (x_1 - 2)^2 + (x_2 - 1)^2$. By contrast, the coarse method given by (2.2) using the same statistic $T(\mathbf{x})$ gives a decreasing sequence of four acceptance regions as follows:

B_1 corresponding to the whole space with $T = 2$;
B_2 corresponding to the whole space except the point $(3,2)$ with $T = 1$;
B_5 corresponding to $T = 0$;

and finally the empty set.



step 1 step 4

FIG. 2. *Graphical representation of Steps* 1 *and* 4 *of Table* 3.

3. Further properties and remarks.

We start this section by demonstrating that $\mathscr{G}(T(\cdot),\mathscr{K}) \subseteq \mathscr{F}(T(\cdot),\mathscr{K})$. That is, we will show that a test determined by $\{(2.2),(2.3)\}$ is precisely a test determined by some B_j, $j = 1,\dots,M+1$. To do so will require the following two lemmas. (For the proofs of the lemmas and the theorem, see the Appendix.)

LEMMA 3.1. *Let* $\mathscr{K}$ *be a pointed convex cone and* D *a convex polygon. Then there exists a vertex*, $\mathbf{v}_0 \in D$, *such that* $\mathbf{v}_0 + \mathbf{b} \notin D$ *for all* $\mathbf{b} \in \mathscr{K}$, $\mathbf{b} \neq \mathbf{0}$.

LEMMA 3.2. *Let* Ω *be any convex COM* $[\mathscr{K}]$ *set* [*i.e.*, $I_{\Omega^c}(\mathbf{x})$ *is a COM* $[\mathscr{K}]$ *function*]. *If* $\mathbf{x}_0 \notin \Omega$, *then there exists a hyperplane* $\boldsymbol{\gamma}'(\mathbf{x}-\mathbf{x}_0) = 0$, *through* $\mathbf{x}_0$ *such that* $\boldsymbol{\gamma}'(\boldsymbol{\omega}-\mathbf{x}) < 0$ *for all* $\boldsymbol{\omega} \in \Omega$. *Furthermore, any such* $\boldsymbol{\gamma} \in \mathscr{K}^*$, *the dual of* $\mathscr{K}$.

THEOREM 3.3. *If* $T(\mathbf{x})$ *is a convex function and* $\mathscr{K}$ *is a closed, convex, pointed cone, then*

$$\mathscr{G}(T(\cdot),\mathscr{K}) \subseteq \mathscr{F}(T(\cdot),\mathscr{K}).$$

Theorem 3.3 implies that the set of p-values for tests whose acceptance regions are of the form (2.4) is contained in the set of p-values corresponding to the COM $[\mathscr{K}]$ tests based on (2.7). The p-value associated with a particular observed sample point $\mathbf{x}_0$ can be obtained by using the same iterative procedure used to construct the B_j's and stopping as soon as $\mathbf{x}_0$ is the point which is removed. If $j(\mathbf{x}) = \min\{j\colon \mathbf{x} \notin B_j\}$, then the p-value associated with $\mathbf{x}_0$ is equal to $p_0(B^c_{j(\mathbf{x}_0)})$, the probability of $B^c_{j(\mathbf{x}_0)}$ under H_0.

The level of difficulty in implementing tests based on this procedure depends on the sample size, the dimension of the cone and the nature of the cone. As is the case with most exact tests in discrete situations, sometimes a hand calculator will be sufficient and other times the problem will not be feasible even by computer. Typically, computer assistance is desirable. If the dimension of the cone is 3 or more, computer assistance is virtually a necessity.

From a computing standpoint, the challenge is to develop algorithms which can quickly and efficiently identify the directed vertices of the convex hulls of the B_i's.

Many common models, such as those in Section 4, permit some simplification of the programming process. The important features of such models are (1) the existence of a linear transformation $\mathbf{z} = A\mathbf{x}$ on the sample space so that COM $[\mathscr{K}]$ in the $\mathbf{x}$-variables becomes equivalent to nondecreasing in the $\mathbf{z}$-variables, and (2) the $\mathbf{z}$-variables only assume integer values. When working with the $\mathbf{z}$'s, the appropriate cone is just the first quadrant, that is, $\{\mathbf{z}\colon z_i \geq 0,\ \text{all } i\}$.

In CS (1998b) two numerical examples are studied in detail. In both cases the DVP process was carried out using the computer.

Crude bounds on the p-value, which do not necessitate identifying directed vertices and are thus much more easily computed, are sometimes adequate for testing purposes.

If one wants to perform a test at size α, then a computed lower bound that exceeds α means that the null hypothesis should not be rejected. On the other hand, if a computed upper bound is less than or equal to α, then the null hypothesis is rejected. A lower bound, more easily obtained by the computer, is

$$p_0(\mathscr{K}(\mathbf{x}_0)). \tag{3.1}$$

A crude upper bound is $[1 - p_0((-\mathscr{K})(\mathbf{x}_0))]$ where

$$(-\mathscr{K})(\mathbf{x}_0) = \big\{\mathbf{x}\colon \mathbf{x} \in \mathscr{X},\ \mathbf{x} = \mathbf{x}_0 - \mathbf{b},\ \text{all } \mathbf{b} \in \mathscr{K}\big\}. \tag{3.2}$$

A better upper bound can be obtained as follows. For each $\mathbf{x} \in \mathscr{X}$, let $M(\mathbf{x})$ be the smallest value T attains in $\mathscr{K}(\mathbf{x})$. That is, $M(\mathbf{x}) = \min_{\mathbf{u}\in\mathscr{K}(\mathbf{x})} T(\mathbf{u})$. For an observed $\mathbf{x}_0$, compute $M(\mathbf{x}_0)$. For every other point $\mathbf{x} \notin (-\mathscr{K})(\mathbf{x}_0) = \{\mathbf{x}_0\} - \mathscr{K}$, $M(\mathbf{x})$ is compared to $M(\mathbf{x}_0)$. If $W = \{\mathbf{x}\colon \mathbf{x} \notin (-\mathscr{K})(\mathbf{x}_0),\ M(\mathbf{x}) \geq M(\mathbf{x}_0)\}$, then $p_0(W)$ is an upper bound on the p-value. The number of comparisons needed can be greatly reduced by recognizing that $\mathbf{x} \in \mathscr{K}(\mathbf{x}_0)$ implies $\mathbf{x} \in W$. Also if $\mathbf{x} \notin W$, then $\mathbf{x} - \mathbf{b} \notin W$ all $\mathbf{b} \in \mathscr{K}$.

4. Examples.

4.1. *Testing equality of binomial parameters versus the simple order alternative.* This problem has been studied by Cohen, Perlman, and Sackrowitz (1990) and recently by Agresti and Coull (1996). The model is U_i, $i = 1, \dots, J$ are independent binomial variables with parameters (n_i, p_i). (For simplicity we take $n_i = n$.) Test $H_0\colon p_1 = \cdots = p_J$ versus $H_1 - H_0$ where $H_1\colon p_1 \geq p_2 \geq \cdots \geq p_J$. Another way to express H_1 is $H_1\colon \{\mathbf{p} : \Gamma\mathbf{p} \geq 0\}$, where

$$\Gamma^{(J-1)\times J} = \begin{pmatrix} 1 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -1 & \cdots & \cdots & 0 \\ \vdots & & & & & \\ 0 & \cdots & \cdots & 0 & 1 & -1 \end{pmatrix}.$$

The entire problem could be equivalently stated in terms of the natural parameters $\log(p_i/(1-p_i))$. The joint density of $\mathbf{U} = (U_1, \ldots, U_J)'$ is a multivariate exponential family with natural parameter vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_J)'$, $\mu_i = \log[p_i/(1-p_i)]$ and can be written as

$$f_{\mathbf{U},\boldsymbol{\mu}}(\mathbf{u}, \boldsymbol{\mu}) = \beta(\boldsymbol{\mu}) h(\mathbf{u}) \exp(u'\mu). \tag{4.1}$$

A complete sufficient statistic under H_0 is $Y = \sum_{i=1}^{J} U_i$. Let $B^{J\times J} = \binom{\Gamma}{1,1,\ldots,1}$, and let $\boldsymbol{\nu} = B\boldsymbol{\mu}$, $\mathbf{X} = (B^{-1})'\mathbf{U}$. Then from (4.1), we have that the conditional distribution of $\mathbf{X}^{(1)}$, the first $(J-1)$ components of $\mathbf{X}$, given $X_n = Y = y$ [of the form (2.1)] is

$$f_{\mathbf{X}^{(1)}|Y=y}(\mathbf{x}^{(1)}|y) = h_y(\mathbf{x}^{(1)}) C_y(\boldsymbol{\nu}^{(1)}) \exp(\mathbf{x}^{(1)'}\boldsymbol{\nu}^{(1)}). \tag{4.2}$$

The null hypothesis is transformed to H_0: $\boldsymbol{\nu}^{(1)} = \mathbf{0}$ and the alternative is $H_1 - H_0$, where H_1: $\nu_i \geq 0$, $i = 1, 2, \ldots, J-1$. We may write H_1: $\boldsymbol{\nu} \in Q$ with Q the first quadrant cone and dual $Q^* = Q$. In terms of the original variables U_i, $i = 1, 2, \ldots, J$, the X_i's are partial sums. That is, $X_i = \sum_{j=1}^{i} U_j$, $i = 1, 2, \ldots, J-1$. See CS (1994), Section 9E2.

As a numerical example, consider the data given in Baker, Detsky, Wesson (1982) and discussed in Moses, Emerson and Hosseini (1984) on the nutritional status of 48 hospitalized patients and the occurrence of infections among the patients. The data is summarized in Table 4. The conditional p-value given the first row total, using the test method outlined in (2.6) and (2.7) with $T(\mathbf{x}^{(1)})$ the Fisher statistic, that is, $T(\mathbf{x}^{(1)}) = 1/h_y(\mathbf{x}^{(1)})$, is 0.00064. For this model our test is admissible. Our test methodology for this discrete model can potentially be an improvement on others suggested for this model. Other methods do not offer the refined grid of p-values that our method offers because our method breaks most ties.

4.2. *Testing equality of two multinomial distributions versus stochastic ordering.*

4.2a. *Model.* Let X_{ij}, $i = 1, 2$, $j = 1, 2, \ldots, J$ be cell frequencies of a $2 \times J$ table where it is assumed that $\mathbf{X}_i = (X_{i1}, \ldots, X_{iJ})'$ are independent multinomial vectors with cell probabilities $\mathbf{p}_i = (p_{i1}, \ldots, p_{iJ})'$, such

TABLE 4

	Nutritional status			
Outcome	**Normal**	**Mild malnutrition**	**Severe malnutrition**	**Total**
Infection	21	4	5	30
No infection	4	3	11	18
Total	25	7	16	

that $\sum_{j=1}^{J} X_{ij} = n_i$, $\sum_{j=1}^{J} p_{ij} = 1$. Test H_0: $\mathbf{p}_1 = \mathbf{p}_2$ versus $H_1 - H_0$ where H_1: $\sum_{j=1}^{k} p_{1j} \geq \sum_{j=1}^{k} p_{2j}$, $k = 1, 2, \ldots, J-1$. The alternative means that the second multinomial distribution is stochastically larger than the first. Under H_0, $\mathbf{m} = (X_{.1}, \ldots, X_{.J})$, where $X_{.j} = \sum_{i=1}^{2} X_{ij}$, is a complete sufficient statistic. If we define $\mathbf{X}^{(1)} = (X_{11}, \ldots, X_{1(J-1)})'$, then the conditional distribution of $\mathbf{X}^{(1)}|\mathbf{m}$ is of the form

$$f_{\mathbf{X}^{(1)}|\mathbf{m}}(\mathbf{x}^{(1)}|\mathbf{m}) = \beta_{\mathbf{m}}(\boldsymbol{\nu}) h_{\mathbf{m}}(\mathbf{x}^{(1)}) \exp(\mathbf{x}^{(1)'}\boldsymbol{\nu}), \tag{4.3}$$

with $\nu_j = \ln p_{1j} p_{2J} / p_{2j} p_{1J}$, $j = 1, \ldots, J-1$ as natural parameters. The ν_j are log odds ratios formed by the jth and Jth columns of the $2 \times J$ table of cell probabilities. Whereas $\boldsymbol{\nu}$ is a function of $(\mathbf{p}_1, \mathbf{p}_2)$, the function is not one–one. It is many to one. Nevertheless, the null hypothesis H_0: $\mathbf{p}_1 = \mathbf{p}_2$ is equivalent to H_0: $\boldsymbol{\nu} = \mathbf{0}$. The alternative $H_1 - H_0$ is not equivalent to $H_1^* - H_0$ where H_1^*: $\nu_1 \geq 0$. However, Berger and Sackrowitz (1997) show that given any $\boldsymbol{\nu}$ such that $\nu_1 > 0$, there exists some $(\mathbf{p}_1, \mathbf{p}_2)$ which satisfies the stochastic order alternative. Therefore, it is desirable to have a test with reasonable power for all $\boldsymbol{\nu}$ such that $\nu_1 > 0$.

In fact, the complete class of tests for the conditional problem of testing H_0 versus $H_1^* - H_0$ serves as a complete class for the unconditional problem of testing H_0 versus $H_1 - H_0$. Furthermore, any test which is conditionally admissible for testing H_0 versus $H_1^* - H_0$ is shown in CS (1998a) to be unconditionally admissible for testing H_0 versus $H_1 - H_0$. Conditional testing in contingency tables is common. Of course, what is most desirable is a test with good conditional properties and good unconditional properties. Our methodology is applied conditionally, but clearly such an approach yields a test that also qualifies as an unconditional test.

H_1 may be expressed as H_1: $\boldsymbol{\nu} \in \mathscr{C}$ where $\mathscr{C}$ is the cone $\{\boldsymbol{\nu}: (1, 0, \ldots, 0)\boldsymbol{\nu} \geq 0\}$. Thus the dual cone $\mathscr{C}^*$ consists only of the single ray $(\lambda, 0, \ldots, 0)$, all $\lambda \geq 0$. Hence to be COM $[\mathscr{C}^*]$ (and in the complete class) a test need only be monotone in X_{11}. Most tests proposed for this problem, including the LRT and WMW tests are COM $[\mathscr{C}^*]$. Unfortunately, merely satisfying this modest condition does not guarantee a desirable test procedure.

It is also important to view the problem from a practical standpoint. We begin with Tables 5 and 6, which present two possible data sets consistent with the marginal totals appearing in Patefield (1982). After examining the tables, some statisticians would be more inclined to assert stochastic order

TABLE 5

	Worse	**Same**	**Better**	**Total**
Control	5	11	1	17
Treatment	3	8	4	15
Total	8	19	5	

TABLE 6

	Worse	**Same**	**Better**	**Total**
Control	0	16	1	17
Treatment	8	3	4	15
Total	8	19	5	

for Table 5 than for Table 6. Therefore, they would want that the p-value associated with Table 5 should be less than that for Table 6.

Nguyen and Sampson (1987) study the related problem of testing independence versus positive quadrant dependence in $R \times C$ contingency tables. Although they assume the full multinomial model, their problem reduces to testing equality of multinomials versus stochastic ordering when $R = 2$, the row totals are fixed and the cell probabilities are normalized, that is, $(p_{ij}/p_{i.})$. They also, in essence, argue that intuitively any test procedure for their problem should be COM $[\mathscr{L}]$. In their terminology, they would say that Table 5 is more concordant than Table 6.

In fact, intuitively, in our opinion, to be recommended for use in practice, a procedure should have the property that its p-value will decrease if any X_{1j} increases while some $X_{1j'}$ decreases for $j < j'$, holding all margins fixed. Using our current terminology, this is equivalent to a procedure being COM $[\mathscr{L}]$ where $\mathscr{L}$ is the cone with generators $(1, -1, 0, \dots)$, $(0, 1, -1, 0, \dots), \dots, (0, \dots, 0, 1, -1), (0, \dots, 0, 1)$. That is, the test function should be monotone in each of these directions.

The cone $\mathscr{L}$ is chosen because it contains $\mathscr{C}^*$ and because it is the smallest cone containing $\mathscr{C}^*$ that preserves the concordance property. Tests which are COM $[\mathscr{L}]$ are said to be concordant monotone.

4.2b. *Recommended tests.* The unconditional LRT is studied by Grove (1980) and Bhattacharya and Dykstra (1994). Agresti and Coull (1998) consider the unconditional LRT statistic and perform the test using that statistic, conditionally. That is, they use simulation to determine a conditional critical value. The large sample distribution of the LRT statistic is studied in Bhattacharya and Dykstra (1994) and recently by Wang (1996), Section 3, who notes difficulties with the computational aspects of the LRT. Agresti and Coull (1998) avoid some of the difficulties. They do not perform the conditional LRT. (Thus far, the conditional maximum likelihood estimators (MLE) have only been found in the 2×2 table by Kou and Ying (1996).) Yet the test used by Agresti and Coull would appear to be reasonable for the conditional problem.

These likelihood ratio based tests are *not* COM $[\mathscr{L}]$ although they are COM $[\mathscr{C}^*]$. Among its capabilities, the program of Agresti and Coull (1998) can be used to find the LRT conditional p-values as 0.169 for the data of Table 5 and a p-value of 0.019 for the data of Table 6.

The LRT statistic, T_{01}, as given by Bhattacharya and Dykstra (1994), page 243, rejects for large values. In this example, the LRT statistic is 2.777

for Table 5 and is 22.652 for Table 6. This ordering of values for the LRT statistic is consistent with the p-values found. Since, in our opinion, the COM [$\mathscr{L}$] property is intuitively appealing we prefer to recommend a test with this property. Of course, the LRT for composite hypotheses also has intuitive support of a different type and some statisticians often find that intuition compelling.

As noted earlier, the WMW test is the most popular of linear tests suggested for this problem. Let $\boldsymbol{\gamma}$ be a $(J-1) \times 1$ vector of constants. For a particular choice of $\boldsymbol{\gamma}$, a linear test rejects if $\boldsymbol{\gamma}'\mathbf{x}$ is large. It is easy to see that such tests are conditionally Bayes with respect to prior distributions, which concentrate all their mass on the single ray $\{\lambda\boldsymbol{\gamma}\colon \lambda \geq 0\}$. Thus, despite its intuitive appeal, we are lead by the narrowness of the prior to suspect a lack of robustness for such tests; that is, an inability of the WMW and other linear tests to detect many important alternatives that are not “near” the direction $\boldsymbol{\gamma}$.

The test procedures we recommend are based on statistics such as chi-square, LRT and Fisher that use the DVP process for the cone $\mathscr{L}$. For example, if $T(\mathbf{X}) = 1/h(\mathbf{X})$, that is, Fisher’s statistic, then we do recommend its use in the DVP process, along with the cone $\mathscr{L}$. We call this the COM [$\mathscr{L}$] Fisher test.

We refer to CS (1998b) for a numerical study which compares the WMW test with the COM [$\mathscr{L}$] Fisher test. Conditional power and conditional expected p-values are computed for a 2×3 table and a 2×4 table. The numerical results regarding WMW are consistent with the lack of robustness we would expect from a test which is Bayes against a prior putting all its mass in one direction. The study indicates that the power and expected p-value indices are *comparable* for alternative points for which WMW is expected to do well. For all other alternatives, the COM [$\mathscr{L}$] Fisher test is decidedly better. The superiority of COM [$\mathscr{L}$] Fisher over WMW was observed numerically in other 2×3 and 2×4 tables.

When trying to establish the relative worth of procedures, we must keep in mind that both the COM and WMW tests are admissible (as are most tests which have been proposed for use). Thus neither test can be better for all alternatives. Therefore, in comparing two tests, the issues are (1) how often one power is greater than the other, (2) the magnitude of the differences in power and (3) if the tests are sensitive to alternatives of interest.

In Table 7 we present some values of the conditional power functions of the COM [$\mathscr{L}$] Fisher and WMW tests studied in CS (1998b). The marginal totals are those of Patefield (1982) and used in our Tables 5 and 6. The level of the COM [$\mathscr{L}$] Fisher test is 0.043 and that of the WMW test is 0.047. The information in Table 7 is extremely helpful in trying to understand the subtleties of the problem. We focus on only a few parameter points, but the behavior seen here is not unusual and is repeated many times throughout the parameter space.

Comparing entry 1 with entry 2 or entry 3 with entry 4 of Table 7, we find the behavior of the power function of the WMW test disappointing. Surely for each pair of alternatives, the first presents a stronger case of stochastic

TABLE 7

Selected alternatives for the model of Tables 5 and 6

	True proportions			**Log odds ratios**		**Power**	
	p_{11}	p_{12}	p_{13}				
	p_{21}	p_{22}	p_{23}				
	Worst	···	**Best**	v_1	v_2	**COM [$\mathscr{L}$]**	**WMW**
Control	0.4	0.5	0.1	2.77	1.39	0.646	0.738
Treatment	0.1	0.5	0.4				
Control	0.4	0.5	0.1	3.47	3.69	0.825	0.240
Treatment	0.1	0.1	0.8				
Control	0.2	0.5	0.3	1.39	1.20	0.472	0.353
Treatment	0.1	0.3	0.6				
Control	0.2	0.5	0.3	1.67	0.29	0.569	0.147
Treatment	0.1	0.1	0.8				
Control	0.2	0.7	0.1	1.61	2.46	0.546	0.155
Treatment	0.2	0.3	0.5				
Control	0.3	0.1	0.6	1.39	0.29	0.481	0.554
Treatment	0.1	0.1	0.8				
Control	0.3	0.2	0.5	1.28	−0.22	0.511	0.570
Treatment	0.1	0.3	0.6				
Control	0.4	0.4	0.2	1.39	−0.56	0.678	0.617
Treatment	0.1	0.7	0.2				
Control	0.3	0.6	0.1	1.79	3.58	0.700	0.004
Treatment	0.3	0.1	0.6				
Control	0.4	0.5	0.1	0.69	0.92	0.258	0.160
Treatment	0.4	0.4	0.2				
Control	0.6	0.2	0.2	1.10	0.69	0.375	0.351
Treatment	0.4	0.2	0.4				
Control	0.6	0.3	0.1	3.40	1.32	0.929	0.954
Treatment	0.1	0.4	0.5				
Control	0.6	0.3	0.1	3.04	3.04	0.859	0.556
Treatment	0.2	0.1	0.7				

order. Yet the power of the WMW test decreases dramatically as one goes from the less preferred alternative to the more preferred alternative. The table also reveals the level of sensitivity of the WMW test to certain types of alternatives.

It is also of interest to note that, in this example, the WMW test rejects for large values of $51X_{11} + 24X_{12}$. Therefore, the WMW is conditionally Bayes with respect to a prior putting all its mass on the set $\{\boldsymbol{\nu}\colon \nu_1 = 51\nu_2/24\}$. This property is approximately true for the first entry in Table 7.

In summary, the COM [$\mathscr{L}$] Fisher test is concordant monotone, is robust in the sense of being sensitive to most stochastic order alternatives (a property not enjoyed by WMW), has comparable conditional power and expected p-value properties in areas where WMW does well and superior properties elsewhere and finally, is an admissible test.

REMARK 4.1. Some commentaries above are qualitative judgments based on examples and based on some mathematical results. A clearly preferred test may not exist in some circumstances, depending on the size of the conditional sample space, as sizes of tests can be arbitrarily chosen. In fact, if the size is small enough, all the tests may be the same. Based on our understanding of this problem and based on some additional numerical work, our opinion is that our test procedures will be preferred whenever test sizes are not too small relative to the sample sizes.

APPENDIX

PROOF OF LEMMA 3.1. Since $\mathscr{K}$ is pointed, there exists a $\mathbf{p}$ such that $\mathbf{p}'\mathbf{b} > 0$ for all $\mathbf{b} \in \mathscr{K}$, $\mathbf{b} \neq \mathbf{0}$. Let $\mathbf{v}_0$ be a vertex of D such that $\mathbf{p}'\mathbf{v}_0 = \sup_{\mathbf{v}\in D} \mathbf{p}'\mathbf{v}$ [i.e., $\mathbf{p}'(\mathbf{x}-\mathbf{v}_0) = 0$ is tangent to D through $\mathbf{v}_0$]. Thus $\mathbf{p}'(\mathbf{v}-\mathbf{v}_0) \leq 0$, all $\mathbf{v} \in D$. Next consider, for $\mathbf{b} \in \mathscr{K}$, $\mathbf{b} \neq \mathbf{0}$, $\mathbf{p}'((\mathbf{v}_0+\mathbf{b})-\mathbf{v}_0) = \mathbf{p}'\mathbf{b} > 0$ and so $\mathbf{v}_0+\mathbf{b} \notin D$.

PROOF OF LEMMA 3.2. Since Ω is convex and $\mathbf{x}_0 \notin \Omega$, the separating hyperplane theorem [e.g., Ferguson (1967), page 73] implies the existence of a $\boldsymbol{\gamma}$ such that $\boldsymbol{\gamma}'(\boldsymbol{\omega} - \mathbf{x}_0) < 0$, all $\boldsymbol{\omega} \in \Omega$. We claim $\boldsymbol{\gamma} \in \mathscr{K}^*$. If not, there exists a $\mathbf{b}^* \in \mathscr{K}$ such that $\boldsymbol{\gamma}'\mathbf{b}^* < 0$. Let $\boldsymbol{\omega}_0 \in \Omega$. Then since Ω is a COM $[\mathscr{K}]$ set, $\boldsymbol{\omega}_0 - \lambda\mathbf{b}^* \in \Omega$ for all $\lambda \geq 0$. Thus $0 > \boldsymbol{\gamma}'(\boldsymbol{\omega}_0 - \lambda\mathbf{b}^* - \mathbf{x}_0) = \boldsymbol{\gamma}'(\boldsymbol{\omega}_0 - \mathbf{x}_0) - \lambda\boldsymbol{\gamma}'\mathbf{b}^*$, which is a contradiction for large λ.

PROOF OF THEOREM 3.3. It suffices to show $\tilde{A}_\alpha \cap \mathscr{X} = B_j$ for some $j = 1, 2, \ldots, M+1$.

Obviously $\tilde{A}_\alpha \cap \mathscr{X} \subseteq \mathscr{X} = B_1$. Thus, to establish the result, it suffices to show that whenever $B_i \supseteq \tilde{A}_\alpha \cap \mathscr{X}$ but $B_i \neq \tilde{A}_\alpha \cap \mathscr{X}$, then $B_{i+1} \supseteq \tilde{A}_\alpha \cap \mathscr{X}$.

By definition, any $\mathscr{K}$ directed point of $\tilde{A}_\alpha$ must belong to A_α. In particular, any $\mathscr{K}$ directed vertex of B_i that lies in $\tilde{A}_\alpha$ must lie in A_α. Also, by definition $T(\mathbf{x}_1) > T(\mathbf{x}_2)$ for all $\mathbf{x}_1 \in \tilde{A}_\alpha^c$, $\mathbf{x}_2 \in A_\alpha$. Thus if $\mathbf{v}_1, \mathbf{v}_2$ are $\mathscr{K}$-directed vertices of B_i such that $\mathbf{v}_1 \in \tilde{A}_\alpha^c$ and $\mathbf{v}_2 \in \tilde{A}_\alpha$, then $T(\mathbf{v}_1) > T(\mathbf{v}_2)$. In such a case, in going from B_i to B_{i+1}, no point of $\tilde{A}_\alpha$ would be removed. Hence, under the assumption that $B_i \supseteq \tilde{A}_\alpha \cap \mathscr{X}$, but $B_i \neq \tilde{A}_\alpha \cap \mathscr{X}$, it suffices to show that there exists a $\mathscr{K}$ directed vertex of B_i which does not belong to $\tilde{A}_\alpha$.

To accomplish this, let $\mathbf{x}_0$ be any point such that $\mathbf{x}_0 \notin \tilde{A}_\alpha$ and $\mathbf{x}_0 \in B_i$. By Lemma 2.2, there exists a $\boldsymbol{\gamma} \in \mathscr{K}^*$ such that $\boldsymbol{\gamma}'(\mathbf{x} - \mathbf{x}_0) < 0$ for all $\mathbf{x} \in \tilde{A}_\alpha$. Let R_γ be such that

$$R_\gamma = \sup_{\mathbf{x}\in B_i} \boldsymbol{\gamma}'(\mathbf{x} - \mathbf{x}_0).$$

The hyperplane $\boldsymbol{\gamma}'(\mathbf{x}-\mathbf{x}_0) = R_\gamma$ (which is a tangent hyperplane to the polygon B_i) is such that $\boldsymbol{\gamma}'(\mathbf{x} - \mathbf{x}_0) \leq R_\gamma$, all $\mathbf{x} \in B_i$. Furthermore $D_\gamma = \{\mathbf{x}\colon \boldsymbol{\gamma}'(\mathbf{x} - \mathbf{x}_0) = R_\gamma\} \cap B_i$ is a polygon whose vertices are a subset of the vertices of B_i. By Lemma 2.2, there exists a vertex $\mathbf{v}_0$ of D_γ (and hence of B_i) such that $\mathbf{v}_0 + \mathbf{b} \notin D_\gamma$ for all $\mathbf{b} \in \mathscr{K}$, $\mathbf{b} \neq 0$.

Next we show that $\mathbf{v}_0$ is the vertex we seek. Certainly $\mathbf{v}_0 \notin \tilde{A}_\alpha$ as $\boldsymbol{\gamma}'(\mathbf{v}_0 - \mathbf{x}_0) = R_\gamma \geq 0$. Next consider $\mathbf{v}_0 + \mathbf{b}$ for $\mathbf{b} \in \mathscr{K}$. Since $\mathbf{v}_0 + \mathbf{b} \notin D_\gamma$, $\mathbf{v}_0 + \mathbf{b}$ can belong to B_i only if $\boldsymbol{\gamma}'(\mathbf{v}_0+\mathbf{b}-\mathbf{x}_0) < R_\gamma$. But $\boldsymbol{\gamma}'(\mathbf{v}_0+\mathbf{b}-\mathbf{x}_0) = \boldsymbol{\gamma}'(\mathbf{v}_0-\mathbf{x}_0)+\boldsymbol{\gamma}'\mathbf{b} = R_\gamma + \boldsymbol{\gamma}'\mathbf{b} \geq R_\gamma$ as $\boldsymbol{\gamma} \in \mathscr{K}^*$, $\mathbf{b} \in \mathscr{K}$. This completes the proof. □

REFERENCES

AGRESTI, A. and COULL, B. A. (1996). Order restricted tests for stratified comparisons of binomial proportions. *Biometrics* **52** 1103–1111.

AGRESTI, A. and COULL, B. A. (1998). Order restricted inference using odds ratios for monotone trend alternatives in contingency tables. *Comput. Statist. Data Analysis.* **28**, 139–155.

AGRESTI, A. and FINLAY, B. (1997). *Statistical Methods in the Social Sciences*, 3rd ed. Prentice Hall, Upper Saddle River, NJ.

BAKER, J. P., DETSKY, A. S., WESSON, D. E., WOLMAN, S. L., STEWART, S., WHITEWELL, J., LANGER, B. and JEEJEEBHOG, K. N. (1982). Nutritional assessment: a comparison of clinical judgment and objective measurements. *New England J. Medicine* **306** 969–972.

BERGER, V. and SACKROWITZ, H. B. (1997). Improving tests of stochastic order. *J. Amer. Statist. Assoc.* **92** 700–705.

BHATTACHARYA, B. and DYKSTRA (1994). Statistical inference for stochastic ordering. In *Stochastic Orders and Their Applications* (M. Shaked and J. G. Shantikumar, eds.) 221–249. Academic Press, Boston.

COHEN, A., KEMPERMAN, J. H. B. and SACKROWITZ, H. B. (1994). Projected tests for order restricted alternatives. *Ann. Statist.* **22** 1539–1546.

COHEN, A., PERLMAN, M. D. and SACKROWITZ, H. B. (1990). Unbiasedness of tests of homogeneity when alternatives are ordered. In *Proceedings of the Symposium on Dependence in Statistics and Probability* (H. W. Block, A. R. Sampson and T. H. Swirts, eds.) 135–146. IMS, Hayward, CA.

COHEN, A. and SACKROWITZ, H. B. (1992). An evaluation of some tests of trend in contingency tables. *J. Amer. Statist. Assoc.* **87** 470–475.

COHEN, A. and SACKROWITZ, H. B. (1994). Association and unbiased tests in statistics. In *Stochastic Orders and Their Applications* (M. Shaked and J. G. Shantikumar, eds.) 251–274. Academic Press, Boston.

COHEN, A. and SACKROWITZ, H. B. (1998a). Testing whether treatment is better than control with ordered categorical data: definitions and complete class theorems. Unpublished manuscript.

COHEN, A. and SACKROWITZ, H. B. (1998b). Testing whether treatment is better than control with ordered categorical data: an evaluation of new methodology. Unpublished manuscript.

COHEN, A., SACKROWITZ, H. B. and SAMUEL-CAHN, E. (1995). Constructing tests for normal order-restricted inference. *Bernoulli* **1** 321–333.

EATON, M. L. (1970). A complete class theorem for multidimensional one-sided alternatives. *Ann. Math. Statist.* **41** 1884–1888.

EMERSON, J. D. and MOSES, L. E. (1985). A note on the Wilcoxon–Mann–Whitney test for $2 \times k$ ordered tables. *Biometrics* **41** 303–309.

FERGUSON, T. S. (1967). *Mathematical Statistics, A Decision Theoretic Approach*. Academic Press, New York.

GRAUBARD, B. I. and KORN, E. I. (1987). Choice of column scores for testing independence in ordered $2 \times K$ contingency tables. *Biometrics* **43** 471–476.

GREEN, P. J. (1985). Peeling data. In *Encyclopedia of Statistical Sciences* **6** 660–664. Wiley, New York.

GROVE, D. M. (1980). A test of independence against a class of ordered alternatives in a $2 \times C$ contingency table. *J. Amer. Statist. Assoc.* **75** 454–459.

KOU, S. G. and YING, Z. (1996). Asymptotics for a 2×2 table with fixed margins. *Statist. Sinica* **6** 809–829.

LEDWINA, T. (1978). Admissible tests for exponential families with finite support. *Math. Operationsforsch. Statist. Ser. Statist.* **9** 105–118.

LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*. Wiley, New York.

MARSHALL, A. W. and OLKIN, I. (1970). *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York.

MOSES, L. E., EMERSON, J. D. and HOSSEINI, H. (1984). Analyzing data for ordered categories. *New England J. Medicine* **311** 442–448.

NGUYEN, T. T. and SAMPSON, A. R. (1987). Testing for positive quadrant dependence in ordinal contingency tables. *Naval Res. Logist.* **34** 859–877.

PAGANO, M. and HALVORSEN, K. T. (1981). An algorithm for finding the exact significance levels of r by c contingency tables. *J. Amer. Statist. Assoc.* **76** 931–934.

PATEFIELD, W. M. (1982). Exact tests for trends for ordered contingency tables. *J. Roy. Statist. Soc. Ser. C* **31** 32–43.

RAHLFS, V. W. and ZIMMERMAN, H. (1993). Scores: Ordinal data with few categories—how they should be analyzed. *Drug Inform. J.* **27** 1227–1240.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, New York.

WANG, Y. (1996). A likelihood ratio test against stochastic ordering in several populations. *J. Amer. Statist. Assoc.* **91** 1676–1683.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
HILL CENTER, BUSCH CAMPUS
110 FRELINGHUYSEN ROAD
PISCATAWAY, NEW JERSEY 08854-8019
E-MAIL: artcohen@rci.rutgers.edu