Adaptive Tests for Ordered Categorical Data

By Vance W. Berger and Anastasia Ivanova

Vance W. Berger is a Mathematical Statistician, Biometry Research Group, DCP, National Cancer Institute, Executive Plaza North, Bethesda, MD 20892-7354. (E-mail: vb78c@nih.gov).

Anastasia Ivanova is Assistant Professor, Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7400. (E-mail: aivanova@bios.unc.edu).

Abstract

Consider testing for independence against stochastic order in an ordered $2xJ$ contingency table, under product multinomial sampling. In applications one may wish to exploit prior information concerning the direction of the treatment effect, yet ultimately end up with a testing procedure with good frequentist properties. As such, a reasonable objective may be to simultaneously maximize power at a specified alternative and ensure reasonable power for all other alternatives of interest. For this objective, we find that none of the available testing approaches are completely satisfactory. We derive a new class of admissible adaptive tests, each of which strictly preserves the Type I error rate and strikes a balance between good global power and nearly optimal (envelope) power to detect a specific alternative of most interest. Prior knowledge of the direction of the treatment effect, the level of confidence in this prior information, and possibly the marginal totals might be used to select a test from this class.

KEY WORDS: Contingency table; exact conditional test; permutation test; adaptive test.

1. INTRODUCTION

When comparing two treatments on the basis of an ordered categorical variable, the data can be summarized as a $2xJ$ contingency table. For example, the objective tumor response data from 35 ovarian cancer patients treated with cisplatin-based combination chemotherapy and salvage platinum-based therapy (Chiara *et al*, 1993) are $(4, 7, 2, 2)$ and $(1, 6, 7, 6)$ for the patients with the treatment-free interval ≤ 12 months and > 12 months, respectively, where the categories are 'progressive disease', 'stable disease', 'partial response', and 'complete response'. Combining the first two categories into a single 'non-response' category, as is routinely done, yields counts $\boldsymbol{C}_1 = (11, 2, 2)$ and $\boldsymbol{C}_2 = (7, 7, 6)$ in the two groups. For simplicity, we treat the case $J = 3$, but our results apply more generally. It is common, in practice, to dispense with the specification of the alternative hypothesis, and proceed directly to the analysis. We prefer to match the analysis to the alternative hypothesis. After briefly presenting notation in Section 2 (details can be found in Berger, 1998, and Berger, Permutt, and Ivanova, 1998; henceforth BPI), we focus attention on stochastic order as the (composite) alternative hypothesis and an appropriate formalization of the superiority of one treatment to another (Cohen and Sackrowitz, 1998). Except for under pathological conditions on the margins, there is no monotone likelihood ratio or uniformly most powerful test, and there will be an entire class of admissible tests. In Section 3 we discuss linear rank tests based on assigning scores to the outcome levels. In Section 4 we discuss nonlinear rank tests such as the Smirnov, improved, convex hull, and $COM(\mathcal{L})$ Fisher tests. In Section 5 we discuss adaptive tests. We generalize, in Section 6, the adaptive test that Berger (1998) proposed for this problem to provide an entire class of exact, admissible, adaptive tests, each of which strikes a balance between good global power and optimal power to detect a specific alternative of most interest. In Section 7 we discuss using the margins to pick one test from this class. In Section 8 we assess the exact unconditional power of several of the aforementioned tests. In Section 9 we give some concluding remarks.

2. NOTATION AND FORMULATION

Consider product multinomial sampling, with n_1 and n_2 (each fixed by the design) patients treated with the control and active treatments, respectively. The vectors of cell probabilities (each summing to one) are $\boldsymbol{\pi}_1 = (\pi_{11}, \pi_{12}, \pi_{13})$ and $\boldsymbol{\pi}_2 = (\pi_{21}, \pi_{22}, \pi_{23})$, respectively, and the corresponding trinomial random vectors are $\boldsymbol{C}_1 = (C_{11}, C_{12}, C_{13})$ and $\boldsymbol{C}_2 = (C_{21}, C_{22}, C_{23})$, with $n_i = C_{i1} + C_{i2} + C_{i3}$, $i = 1,2$. The log odds ratios are calculated from $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ as $\theta_1 = \log\{(\pi_{11}\pi_{23})/(\pi_{21}\pi_{13})\}$ and $\theta_2 = \log\{(\pi_{12}\pi_{23})/(\pi_{22}\pi_{13})\}$. Let $T_j = C_{1j} + C_{2j}$, $j = 1,2,3$. As we condition on $\boldsymbol{T} = (T_1, T_2, T_3)$, the sample space Γ is the set of 2×3 contingency tables with non-negative integer-valued cell counts with row totals $\boldsymbol{n} = (n_1, n_2)$ and column totals $\boldsymbol{T}$. Given $\boldsymbol{T}, \boldsymbol{n}$, and $\boldsymbol{c} = (C_{11}, C_{12})$, we can reconstruct the entire 2×3 contingency table as $C_{13} = n_1 - C_{11} - C_{12}$ and $\boldsymbol{C}_2 = \boldsymbol{T} - \boldsymbol{C}_1$. Thus, we let $\boldsymbol{c}$ denote a point of Γ. Figure 1 displays C_{12} plotted against C_{11} for each of the 87 tables of Γ for our example, $\{(11,2,2);(7,7,6)\}$.

[Figure 1]

Observed table $(11,2)$ is circled. With $H(\boldsymbol{c}) = n_1!n_2!/\Pi_{i=1}^2 \Pi_{j=1}^3 C_{ij}!$, $\boldsymbol{\theta} = (\theta_1, \theta_2)$, $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2)$, and $K(\boldsymbol{T};\boldsymbol{\theta}) = 1/\sum_{c \in \Gamma} H(\boldsymbol{c}) \exp[\boldsymbol{\theta}'\boldsymbol{c}]$, our model follows the exponential family with density

$$P_{\pi}\{\boldsymbol{c}|\boldsymbol{T}\} = P_{\theta}\{\boldsymbol{c}|\boldsymbol{T}\} = K(\boldsymbol{T};\boldsymbol{\theta})H(\boldsymbol{c})\exp[\boldsymbol{\theta}'\boldsymbol{c}]. \tag{2.1}$$

As $P_{\pi}\{\boldsymbol{c}|\boldsymbol{T}\}$ (and hence the conditional power) depends on $\boldsymbol{\pi}$ only through $\boldsymbol{\theta}(\boldsymbol{\pi})$, $\boldsymbol{c}$ offers no information with which to distinguish $\boldsymbol{\pi}$ from $\boldsymbol{\pi}^*$ if $\boldsymbol{\theta}(\boldsymbol{\pi}) = \boldsymbol{\theta}(\boldsymbol{\pi}^*)$. The (conditional) hypotheses must then be formulated in terms of $\boldsymbol{\theta}$ to be identifiable (Berger, 1998). Because the common null hypothesis of equality $\boldsymbol{\pi}_1 = \boldsymbol{\pi}_2$ is equivalent to $\boldsymbol{\theta}(\boldsymbol{\pi}) = \boldsymbol{0}$, the (simple) null hypothesis is $H : \boldsymbol{\theta} = \boldsymbol{0}$. Let $\Delta_1 = \pi_{11} - \pi_{21}$, and $\Delta_2 = (\pi_{11} + \pi_{12}) - (\pi_{21} + \pi_{22}) = \pi_{23} - \pi_{13}$. We wish to test H against the one-sided alternative hypothesis that the active response distribution is stochastically larger than the control response distribution:

$$H_A' : \Delta_1 \geq 0,\ \Delta_2 \geq 0,\ \boldsymbol{\pi}_1 \neq \boldsymbol{\pi}_2.$$

This would imply the superiority of the active treatment. Unfortunately, $\boldsymbol{\theta}(\boldsymbol{\pi})$ provides insufficient information with which to determine if $\boldsymbol{\pi}$ satisfies H'_A, so no conditional alternative hypothesis is equivalent to H'_A. However, if $\boldsymbol{\pi}$ satisfies H'_A, then $\theta_1(\boldsymbol{\pi}) > 0$, and if $\theta_1 > 0$, then for any θ_2 there exists (Berger and Sackrowitz, 1997) $\boldsymbol{\pi}$ satisfying H'_A such that $\boldsymbol{\theta}(\boldsymbol{\pi}) = (\theta_1, \theta_2)$. As such, the treatment effect favors the active treatment whenever $\theta_1 > 0$, regardless of the value of θ_2, and we test H against $H_A : \theta_1 > 0$. One can also test for the superiority of the control ($\theta_1 < 0$). Let $\Omega_0 = \{\boldsymbol{\theta} | \theta = 0\}$, $\Omega_A = \{\boldsymbol{\theta} | \theta_1 > 0\}$, and $\Omega_C = \{\boldsymbol{\theta} | \theta_1 < 0\}$. The large unconditional indifference region, where neither group stochastically dominates the other, has been reduced by conditioning to the relatively small region $\Omega_I = \{\boldsymbol{\theta} | \theta_1 = 0, \theta_2 \neq 0\}$.

Let $\delta(\boldsymbol{\theta}) = 1 - \theta_2/\theta_1$ be the *direction* of the effect, with $\Omega_\nu = \{\theta \,|\delta(\theta\,) = \nu\}$. As θ_1 increases in both Δ_1 and Δ_2, while θ_2 ($\theta_1 - \theta_2$) increases in Δ_2 (Δ_1), and decreases in Δ_1 (Δ_2), the superiority of the active treatment to the control is due primarily to a shift from the middle to the best outcome ($\Delta_2 > \Delta_1$) if $\delta(\boldsymbol{\theta})$ is small, or from the worst the middle outcome ($\Delta_1 > \Delta_2$) if $\delta(\boldsymbol{\theta})$ is large. As $\delta(\boldsymbol{\theta})$ is unknown *a priori*, a test should be sensitive to departures from H_0 in any direction of $\Omega_A = \bigcup_{\nu \in \Re^1} \Omega_\nu$. A necessary condition for φ to be such an omnibus test is that its rejection region $R_\alpha(\varphi)$ contain $D[\Gamma]$, the set of directed extreme points of Γ (BPI, 1998). For reasonable α-levels omnibus tests exist (Sections 4 and 5.3). We exploit prior information about $\delta(\boldsymbol{\theta})$ to construct admissible, omnibus tests with especially good power in one direction, Ω_ν.

3. A NEW LOOK AT LINEAR RANK TESTS

Linear rank tests are based on numerical scores (ν_1, ν_2, ν_3), $\nu_1 < \nu_3$, assigned to the three outcome levels. With $\nu = (\nu_2 - \nu_1)/(\nu_3 - \nu_1)$, φ_ν uses test statistic $z_\nu(\boldsymbol{c}) = C_{11} + (1 - \nu)C_{12}$. Let $M_\nu(\boldsymbol{c}) = \{\boldsymbol{c}^* \in \Gamma \mid z_\nu(\boldsymbol{c}^*) \geq z_\nu(\boldsymbol{c})\}$ be the φ_ν extreme region of $\boldsymbol{c}$, with boundary $B_\nu(\boldsymbol{c})$ and $p_\nu(\boldsymbol{c}) = P_{\mathbf{0}}\{M_\nu(\boldsymbol{c})|\boldsymbol{T}\}$ the corresponding p-value. The level set (Frick, 2000, page 719) of $z_\nu(\boldsymbol{c})$ is $B_\nu(\boldsymbol{c}) \cap \Gamma$, with $o_\nu(\boldsymbol{c})$ its order (the number of points of Γ on $B_\nu(\boldsymbol{c})$). For $\boldsymbol{c} = (C_{11}, C_{12}) \in \Gamma$ and $\boldsymbol{c}^* = (C^*_{11}, C^*_{12}) \in \Gamma - \boldsymbol{c}$, $z_\nu(\boldsymbol{c}^*) = z_\nu(\boldsymbol{c})$ if and only if $\nu = 1 - (C_{11} - C^*_{11})/(C^*_{12} - C_{12})$, say

$v = v_{c,c^*}$. Let $V(\boldsymbol{c}) = \{v_1(\boldsymbol{c}), v_2(\boldsymbol{c}), ..., v_{K_c}(\boldsymbol{c})\}$ be the ordered set $\{v_{c,c^*} \mid |v_{c,c^*}| < \infty$, $\boldsymbol{c}^* \in \Gamma - \boldsymbol{c}\}$, and let $v_0(\boldsymbol{c}) = -\infty$ and $v_{K_c+1}(\boldsymbol{c}) = \infty$. For finite v, $o_v(\boldsymbol{c}) > 1$ if and only if $v \in V(\boldsymbol{c})$. Let $\varepsilon(\boldsymbol{c}) = \min_k[v_{k+1}(\boldsymbol{c}) - v_k(\boldsymbol{c})]/2$, $z_v^\perp(\boldsymbol{c}) = C_{12} + (v-1)C_{11}$, $B_v^+(\boldsymbol{c}) = \{\boldsymbol{c}^* \in B_v(\boldsymbol{c}) \cap \Gamma \mid z_v^\perp(\boldsymbol{c}^*) > z_v^\perp(\boldsymbol{c})\}$, and $B_v^-(\boldsymbol{c}) = \{\boldsymbol{c}^* \in B_v(\boldsymbol{c}) \cap \Gamma \mid z_v^\perp(\boldsymbol{c}^*) < z_v^\perp(\boldsymbol{c})\}$.

Lemma 1. Let $\boldsymbol{c} \in \Gamma$ and $k \in \{0, 1, ..., K_c\}$. If $|v_k(\boldsymbol{c}) \pm \varepsilon(\boldsymbol{c})| < \infty$, then $v_k(\boldsymbol{c}) \pm \varepsilon(\boldsymbol{c}) \notin V(\boldsymbol{c})$. If $v \in (v_k(\boldsymbol{c}), v_{k+1}(\boldsymbol{c}))$, then $M_v(\boldsymbol{c}) = M_{v_{k+1}(c)}(\boldsymbol{c}) - B^-_{v_{k+1}(c)}(\boldsymbol{c}) = M_{v_k(c)}(\boldsymbol{c}) - B^+_{v_k(c)}(\boldsymbol{c})$.

Proof. Increasing (decreasing) v by $\varepsilon(\boldsymbol{c})$ moves $B_v^-(\boldsymbol{c})$ ($B_v^+(\boldsymbol{c})$) into the interior of, and $B_v^+(\boldsymbol{c})$ ($B_v^-(\boldsymbol{c})$) completely out of, the new critical region, but if $v \in V(\boldsymbol{c})$, then no points of $\Gamma - M_v(\boldsymbol{c})$ are moved into the new critical region (Table 1). Hence, $o_{v-\varepsilon(c)}(\boldsymbol{c}) = o_{v+\varepsilon(c)}(\boldsymbol{c}) = 1$, and neither $v_k(\boldsymbol{c}) - \varepsilon(\boldsymbol{c})$ nor $v_k(\boldsymbol{c}) + \varepsilon(\boldsymbol{c})$ is in $V(\boldsymbol{c})$. If $v \notin V(\boldsymbol{c})$, say $v_k(\boldsymbol{c}) < v < v_{k+1}(\boldsymbol{c})$, then $o_v(\boldsymbol{c}) = 1$, so $B_v^+(\boldsymbol{c}) = B_v^-(\boldsymbol{c}) = \emptyset$ and $M_v(\boldsymbol{c})$ will not change when v varies within $(v_k(\boldsymbol{c}), v_{k+1}(\boldsymbol{c}))$. □

Let $p_{\min(v)}(\boldsymbol{c}) = p_v(\mathbf{c}) - \max(P_0\{B_v^-(\boldsymbol{c})\}, P_0\{B_v^+(\boldsymbol{c})\}) = \min(\lim_{u\nearrow v} p_u(\boldsymbol{c}), \lim_{u\searrow v} p_u(\boldsymbol{c}))$. Lemma 1 implies that $p_{\min(v)}(\boldsymbol{c}) = \min\{p_{v-\varepsilon(c)}(\boldsymbol{c}), p_{v+\varepsilon(c)}(\boldsymbol{c})\}$ is an actual p-value. As such, if $v \in v^*(\boldsymbol{c}) = \{v^* \mid p_{v^*}(\boldsymbol{c}) \leq p_v(\boldsymbol{c}) \text{ for all } v\}$, then $p_{v^*}(\boldsymbol{c}) \leq p_{\min(v)}(\boldsymbol{c})$ for all v. As the number of sets $M_v(\boldsymbol{c})$ is finite, the minimum p-value is attained, and $v^*(\boldsymbol{c}) \neq \emptyset$. If $v \in V(\boldsymbol{c})$, then $o_v(\boldsymbol{c}) > 1$, $B_v^-(\boldsymbol{c}) \cup B_v^+(\boldsymbol{c}) \neq \emptyset$, $p_{\min(v)}(\boldsymbol{c}) < p_v(\boldsymbol{c})$, and $v \notin v^*(\boldsymbol{c})$. Hence, $v^*(\boldsymbol{c}) \cap V(\boldsymbol{c}) = \emptyset$, and if $v \in v^*(\boldsymbol{c})$, then $v \notin V(\boldsymbol{c})$, say $v \in (v_k(\boldsymbol{c}), v_{k+1}(\boldsymbol{c}))$ for some k. By Lemma 1, $v^*(\boldsymbol{c})$ consists of one or several open intervals of the form $(v_k(\boldsymbol{c}), v_{k+1}(\boldsymbol{c}))$. In our example, $\{(11,2,2);(7,7,6)\}$, we have $\boldsymbol{c} = (11,2)$, $K_c = 42$, $\varepsilon(11,2) = 1/84$, and

$$V(\boldsymbol{c}) = \{-6,-5,-4,-3,-\tfrac{5}{2},-2,-\tfrac{5}{3},-\tfrac{3}{2},-\tfrac{4}{3},-\tfrac{5}{4},-\tfrac{6}{5},-1,-\tfrac{5}{6},-\tfrac{4}{5},-\tfrac{3}{4},-\tfrac{2}{3},-\tfrac{3}{5},-\tfrac{4}{7},-\tfrac{1}{2},$$
$$-\tfrac{3}{7},-\tfrac{2}{5},-\tfrac{1}{3},-\tfrac{2}{7},-\tfrac{1}{4},-\tfrac{1}{5},-\tfrac{1}{6},-\tfrac{1}{7},0,\tfrac{1}{7},\tfrac{1}{6},\tfrac{1}{5},\tfrac{1}{4},\tfrac{2}{7},\tfrac{1}{3},\tfrac{2}{5},\tfrac{1}{2},\tfrac{2}{3},1,\tfrac{3}{2},2,\tfrac{5}{2},3,4,5,6\}.$$

Figure 1 shows $M_{1/7}(11,2)$ by dark dots and $M_0(11,2) - M_{1/7}(11,2)$ by crosses. Because $(11,2)$ minimizes $z_{1/7}^\perp(11,2) = 7C_{12} - 6C_{11}$ over $B_{1/7}(11,2) \cap \Gamma$ (Table 1), $B_{1/7}^-(11,2) = \emptyset$ and

$p_{1/7}(11,2) = \lim_{u \nearrow 1/7} p_u(11,2) = 0.066$. Also $p_v(11,2) = 0.020$ for $v \in (1.0, 1.5) = v^*(11,2)$. If $v \in V(11,2)$, then $P_0\{B_v^-\} \leq P_0\{B_v^+\}$ for $v > 1.5$, while $P_0\{B_v^-\} \geq P_0\{B_v^+\}$ for $v < 1.0$.

[Table 1]

While φ_v is the locally most powerful (LMP) test (BPI, 1998) to detect $l\boldsymbol{\theta}$, for $l > 0$, this local optimality of $\varphi_{\delta(\theta)}$ is offset by potentially poor power on parts of $\Omega_A - \Omega_{\delta(\theta)}$. In fact, the φ_v critical region $R_\alpha(\varphi_v)$ will often fail to contain some points of $D[\Gamma]$, so the power of φ_v to detect $l\boldsymbol{\theta}$, for some $\boldsymbol{\theta} \in \Omega_A - \Omega_{\delta(\theta)}$, tends to zero as l gets large (BPI, 1998). Podgor, Gastwirth, and Mehta (1996) proposed the maximin efficiency robust test (MERT) in hopes of providing better power than linear rank tests. Ironically, the MERT is itself a linear rank test, and its rejection region may fail to contain $D[\Gamma]$, leading to poor power on parts of Ω_A and no power in the limit in some directions. Berger and Ivanova (2001) showed that at certain α-levels the most stringent linear rank test is φ_{v_S}, where v_S is such that the two points of $D[\Gamma]$ that are furthest (Euclidean distance) from each other are equated by $z_{v_S}(c)$. For $\{(11,2,2),(7,7,6)\}$, this gives $v_S = 0$, because Γ has two directed extreme points, $D[\Gamma] = \{(15,0);(6,9)\}$, and $z_0(15,0) = z_0(6,9)$.

4. NONLINEAR RANK TESTS

By allowing the boundary of $R_\alpha(\varphi)$ to curve, nonlinear rank tests often require smaller α-levels to ensure that $D[\Gamma] \subset R_\alpha(\varphi)$. However, this is not always the case. Berger and Ivanova (2001) provide an example in which the proportional odds and proportional hazards tests (McCullagh, 1980) are not nonlinear enough to be omnibus at reasonable α-levels.

4.1. The Smirnov test, φ_S

The Smirnov test, φ_S, uses as the test statistic the largest of 0, $D_1 = C_{11}/n_1 - C_{21}/n_2$, and $D_2 = (C_{11} + C_{12})/n_1 - (C_{21} + C_{22})/n_2$, and minimizes, among tests routinely available in standard statistical software packages (φ_S is a standard feature of StatXact), the α-level required for its rejection region to contain $D[\Gamma]$. However, φ_S is not admissible (Berger, 1998).

4.2. Improved tests

Permutt and Berger (2000) and Ivanova and Berger (2001) each proposed refinements of φ_S that break its ties. While such refinements are necessarily uniformly more powerful than φ_S (Rohmel and Mansmann, 1999, page 158), we reserve the term "improvement of φ" for a test whose exact (randomized) version is uniformly more powerful than the exact version of φ. By this definition, refinements are not necessarily improvements. Berger and Sackrowitz (1997) developed methodology for constructing admissible improvements of a given inadmissible test. In fact, by improving the "ignore-the-data" test, $\varphi_{ITD}(\boldsymbol{c}) = \alpha$ for all $\boldsymbol{c} \in \Gamma$, Berger and Sackrowitz (1997) constructed the first known test for this problem that is simultaneously admissible and unbiased. However, p-values from these improved tests cannot always be defined unambiguously because rejection regions at different α –levels need not be nested.

4.3. Convex hull tests, φ_{CH}

Berger (1998) established the one-to-one correspondence between the class of convex hull type tests and the minimal complete class of admissible tests. The convex hull test (BPI, 1998), φ_{CH}, is the simplest member of this class, and is qualitatively similar to the improvements of both φ_S and φ_{ITD}, while minimizing, among all families of tests, the α-level required for its rejection region to contain $D[\Gamma]$. In addition, φ_{CH} is based on a test statistic, so rejection regions at different α-levels are nested, and p-values are provided. As such, φ_{CH} is about as good a test as there is for the conditional problem, which is as close as one can get to the unconditional problem when dealing with $\boldsymbol{\theta}$ instead of $\boldsymbol{\pi}$. Specifically, admissible (unbiased) tests for the conditional problem are conditionally admissible (unbiased) for the unconditional problem (Berger and Sackrowitz, 1997). However, the mapping from $\boldsymbol{\pi}$ to $\boldsymbol{\theta}$ is nonlinear, and small corners of $\boldsymbol{\pi}$-space (neighborhoods of structural zeros) correspond to large regions of $\boldsymbol{\theta}$-space.

By giving each direction $\delta(\boldsymbol{\theta})$ equal consideration, φ_{CH} accommodates these small corners of $\boldsymbol{\pi}$–space as much as the large regions of $\boldsymbol{\pi}$-space that are of most unconditional interest. As such, φ_{CH} may not be ideal when viewed unconditionally. Cohen and Sackrowitz (1998) proposed another member of the convex hull class, the $COM(\mathcal{L})$ Fisher test, or $\varphi_{COM(\mathcal{L})}$, constructed recursively by adding to the critical region those directed extreme points of the current acceptance region that are least likely under H_0. Because the test statistics of $\varphi_{COM(\mathcal{L})}$ and φ_{CH} are defined not algebraically but relationally, in terms of the position of $\boldsymbol{c}$ relative to other points of Γ, the rejection regions need to be constructed recursively, layer by layer.

5. ADAPTIVE TESTS

Hogg (1974, page 917) and Edgington (1995, pages 371-373) defined adaptive tests as tests with data-based test statistics (this is distinct from another definition used, e.g., by Rukhin and Mak, 1992). Gross (1981, Section 5) suggested that such an "analysis based on...data-dependent scores may yield procedures that compare favorably to fixed-score procedures...", and Gastwirth (1985) stated "when the MERT for a particular problem has a low r^2, adaptive procedures are needed". Partition Γ into regions sharing a common test statistic. Donegani (1991) and Good (1994, page 122) suggested conditioning on the region. Because the region need not be even nearly ancillary, such conditioning may entail a loss of power, so we prefer comparing the value of the test statistics across regions. The intuitive objection to "comparing apples to oranges" notwithstanding, such an approach is "good" or "bad" only to the extent to which it produces a "good" or "bad" test. We will find this approach to result in tests with excellent properties.

5.1. Adaptive tests for this problem

Without knowing θ *a priori*, we do not know where to maximize the power. We could estimate $\delta(\boldsymbol{\theta})$ from $\boldsymbol{c}$, say as $\hat{\delta}_c$, perhaps using maximum likelihood, and use the LMP test $\varphi_{\hat{\delta}_c}$. While the p-value of $\varphi_{\hat{\delta}_c}$ evaluated at observed outcome $\boldsymbol{c}$, $p_{\hat{\delta}_c}(\boldsymbol{c})$, will be stochastically too small to serve as a valid p-value, $p_{\hat{\delta}_c}(\boldsymbol{c})$ can be used as a *test statistic*, to be compared to its null distribution (Rohmel and Mansmann, 1999, page 165). Variation in $\boldsymbol{c}$ is reflected in $p_{\hat{\delta}_c}(\boldsymbol{c})$ through *both* the argument and the subscript. Another possible test statistic would be $z_{\hat{\delta}_c}(\boldsymbol{c})$, suitably normalized (see Section 5.2). Using either $p_{\hat{\delta}_c}(\boldsymbol{c})$ or $z_{\hat{\delta}_c}(\boldsymbol{c})$ as a test statistic, any estimator $\hat{\delta}_c$ of $\delta(\theta)$ induces an adaptive test, with regions $\Gamma_\nu = \hat{\delta}^{-1}(\nu) = \{\boldsymbol{c} \in \Gamma | \hat{\delta}_c = \nu\}$.

5.2. The Smirnov test and other binary adaptive tests for which $\Gamma_\nu = \emptyset$ for $\nu \notin \{0,1\}$

While the nonlinear tests described in Section 4 are not typically defined by an adaptive mechanism, the Smirnov test φ_S can be defined as a binary adaptive test, with $\Gamma_\nu = \emptyset$ for $\nu \notin \{0,1\}$. Specifically, let $\Gamma_0 = \{\boldsymbol{c} \in \Gamma \mid C_{12} > n_1 T_2/(n_1+n_2)\}$ and $\Gamma_1 = \Gamma - \Gamma_0$. On Γ_ν, φ_S uses the φ_ν test statistic $z_\nu(\boldsymbol{c})$, with $C_{11}+C_{12}$ $(\nu = 0)$ and C_{11} $(\nu = 1)$ normalized to D_2 and D_1 (from Section 4.2), respectively, to facilitate the comparison of points from Γ_1 $(D_1 > D_2)$ to those from Γ_0 $(D_2 \geq D_1)$. Other binary adaptive tests include defining Γ_0 and Γ_1 by whichever of φ_0 and φ_1 yields a smaller p-value [i.e., $\Gamma_0 = \{\boldsymbol{c} \in \Gamma : p_0(\boldsymbol{c}) < p_1(\boldsymbol{c})\}$] or a larger χ^2.

5.3. Berger's (1998) adaptive test, φ_A

To judge the extremity of outcome $\boldsymbol{c}$ by how small a p-value it can yield when all LMP tests are applied, use $p_{\nu^*(c)}(\boldsymbol{c}) = \min_{-\infty \leq \nu \leq \infty} p_\nu(\boldsymbol{c})$ as the test statistic. That is, estimate $\delta(\theta)$ non-uniquely as $\hat{\delta}_c = \nu$ for any value $\nu \in \nu^*(\boldsymbol{c})$, so $\Gamma_\nu = \{\boldsymbol{c} \in \Gamma | \nu \in \nu^*(\boldsymbol{c})\}$ are the regions. As the critical region of φ_A is $R_\alpha(\varphi_A) = \cup_{\nu \in \Re^1} R_{\alpha^*(\nu)}(\varphi_\nu)$ for some set of $\alpha^*(\nu) < \alpha$, φ_A is intuitively similar to union-intersection tests (Roy, 1953; Marden, 1991). Despite being constructed non-recursively, φ_A is a convex hull type test (Berger, 1998), and hence φ_A is admissible. Also, φ_A tends to be an omnibus test, because $D[\Gamma] \subset R_\alpha(\varphi_A)$ for reasonable α-levels.

6. ACCOMMODATING A FAVORED ALTERNATIVE

We have seen that φ_ν is LMP on Ω_ν, while φ_A is a good omnibus test. Suppose that we want the best of both, and believe *a priori* that $\delta(\theta) = \delta_P$. Let $\tau \geq 0$ be a measure of strength in the belief that $\delta(\theta) = \delta_P$. The dual objectives are ensuring nearly LMP power on Ω_{δ_P} and reasonable power on $\Omega_A - \Omega_{\delta_P}$, with relative importance dictated by τ. One might use φ_{δ_P} for large τ, or φ_A for small τ, but none of the aforementioned tests would suffice for intermediate values of τ. We bridge this gap by starting with φ_A and then penalizing those $\boldsymbol{c}$ whose minimizing LMP p-value is obtained by v far from δ_P. To this end, let

$$A(\delta_P, \tau, \boldsymbol{c}) = \min_{-\infty \leq v \leq \infty} [p_{\min(v)}(\boldsymbol{c})(1 + |\delta_P - v|)^\tau],$$

and let $\varphi_{\delta_P,\tau,\alpha}$ ($\varphi_{\delta_P,\tau}$ when the α –level is clear) be the level-α adaptive test based on test statistic $A(\delta_P, \tau, \boldsymbol{c})$. Because $A(\delta_P, 0, \boldsymbol{c}) = p_{v^*(c)}(\boldsymbol{c})$, $\varphi_{\delta_P,0} = \varphi_A$ for any δ_P. Let $v_{[\delta_P,\tau]}(\boldsymbol{c}) = \{v \mid p_{\min(v)}(\boldsymbol{c})(1 + |\delta_P - v|)^\tau = A(\delta_P, \tau, \boldsymbol{c})\}$. Clearly $p_{\min(v)}(\boldsymbol{c})(1 + |\delta_P - v|)^\tau \leq 1$ if $v \in v_{[\delta_P,\tau]}(\boldsymbol{c})$. Lemmas 2-4 confine $v_{[\delta_P,\tau]}(\boldsymbol{c})$ to a finite subset of an interval that shrinks to δ_P as τ gets large.

Lemma 2. For any δ_P, $\tau > 0$, $v_* \in v_{[\delta_P,\tau]}(\boldsymbol{c})$, and $v^* \in v^*(\boldsymbol{c})$, $|\delta_P - v_*| \leq |\delta_P - v^*|$.

Proof. If there exist $v^* \in v^*(\boldsymbol{c})$ and $v_* \in v_{[\delta_P,\tau]}(\boldsymbol{c})$ such that $|\delta_P - v^*| < |\delta_P - v_*|$, then $p_{v^*}(\boldsymbol{c})(1 + |\delta_P - v^*|)^\tau < p_{\min(v_*)}(\boldsymbol{c})(1 + |\delta_P - v_*|)^\tau$, and v_* cannot be in $v_{[\delta_P,\tau]}(\boldsymbol{c})$. □

Lemma 3. For any δ_P, $\tau > 0$, and $\boldsymbol{c} \in \Gamma$, $v_{[\delta_P,\tau]}(\boldsymbol{c}) \subset V(\boldsymbol{c}) \cup \delta_P$.

Proof. Assume that there exists $v \neq \delta_P$ in $v_{[\delta_P,\tau]}(\boldsymbol{c}) - V(\boldsymbol{c})$, say $v_k(\boldsymbol{c}) < v < v_{k+1}(\boldsymbol{c})$. Let $v^* = v_k(\boldsymbol{c})$ if $\delta_P \leq v_k(\boldsymbol{c})$, $v^* = \delta_P$ if $v_k(\boldsymbol{c}) < \delta_P < v_{k+1}(\boldsymbol{c})$, or $v^* = v_{k+1}(\boldsymbol{c})$ if $v_{k+1}(\boldsymbol{c}) \leq \delta_P$. Now $v^* \subset V(\boldsymbol{c}) \cup \delta_P$ and $p_{\min(v)}(\boldsymbol{c})(1 + |\delta_P - v|)^\tau > p_{\min(v^*)}(\boldsymbol{c})(1 + |\delta_P - v^*|)^\tau$. □

Lemma 4. For any δ_P and $\boldsymbol{c} \in \Gamma$, $v_{[\delta_P,\tau]}(\boldsymbol{c}) = \{\delta_P\}$ for sufficiently large τ.

Proof. Let $D_c(\delta_P) = \min_{v \in V(c)} |\delta_P - v|$, and for any $\tau > 0$, let $v \in v_{[\delta_P,\tau]}(\boldsymbol{c}) - \delta_P$. By Lemma 3 $v \in V(\boldsymbol{c}) - \delta_P$, so $|\delta_P - v| \geq D_c(\delta_P) > 0$, and for $\tau > -\ln(p_{\min(\delta_P)}(\boldsymbol{c}))/\ln(1 + D_c(\delta_P))$ we have

$p_{\min(v)}(\boldsymbol{c})(1+|\delta_P - v|)^\tau \geq p_{\min(v)}(\boldsymbol{c})(1+|D_c(\delta_P)|)^\tau > 1$, a contradiction to $v \in v_{[\delta_P,\tau]}(\boldsymbol{c})$. □

By Lemma 4, $\varphi_{\delta_P,\infty}$ induces the same ordering on Γ as φ_{δ_P} does, thereby optimizing power on Ω_{δ_P}. Yet the $\varphi_{\delta_P,\infty}$ test statistic is $p_{\min(\delta_P)}(\boldsymbol{c})$, and not necessarily $p_{\delta_P}(\boldsymbol{c})$, so $\varphi_{\delta_P,\infty}$ is a refinement of φ_{δ_P} (Section 4.2). In fact, $p_v(11,2) \leq p_{v,\infty}(11,2) \leq p_{\min(v)}(11,2)$ for all v (Table 1), and $p_{0.5}(11,2) = 0.0385$, but $M_{0.5,\infty}(11,2)$ excludes $(10,4)$, so $\varphi_{0.5,\infty}$ attains statistical significance at $\alpha = 0.025$ (one-sided) with $p_{0.5,\infty}(11,2) = 0.0249$. We now establish the admissibility of $\varphi_{\delta_P,\tau,\alpha}$.

Theorem 1. For any triple $\delta_P \in \mathfrak{R}^1$, $\tau \geq 0$, and $\alpha \in [0,1]$, $\varphi_{\delta_P,\tau,\alpha}$ is admissible.

Proof. By Theorem 3.3 of Berger (1998), it suffices to show that for any $B \subset \Gamma$, if $\boldsymbol{c}^*$ minimizes $A(\delta_P,\tau,\boldsymbol{c})$ over B, then $\boldsymbol{c}^* \in D[B]$. If $\boldsymbol{c}^* \notin D[B]$, then $\boldsymbol{c}^*$ cannot, for any v, uniquely minimize p_v over B, and for every v there exists $\boldsymbol{c} \in B - \boldsymbol{c}^*$ such that $p_v(\boldsymbol{c}) \leq p_v(\boldsymbol{c}^*)$. If $v \notin V(\boldsymbol{c}^*)$, then $o_v(\boldsymbol{c}^*) = 1$, so $p_v(\boldsymbol{c}) \neq p_v(\boldsymbol{c}^*)$, and $p_v(\boldsymbol{c}) \leq p_v(\boldsymbol{c}^*) - \min_{c\in\Gamma} P_0\{\boldsymbol{c}|\Gamma\}$. Let $v_1 \in v_{[\delta_P,\tau]}(\boldsymbol{c}^*)$. By the continuity in v of the function $(1+|\delta_P - v|)^\tau$, we can, for any $\varepsilon > 0$, choose $v_2 \notin V(\boldsymbol{c}^*)$ suitably close to v_1 to satisfy $p_{v_2}(\boldsymbol{c}^*) = p_{\min(v_1)}(\boldsymbol{c}^*)$, and, thus,

$$A(\delta_P,\tau,\boldsymbol{c}) = \min_{-\infty\leq v\leq\infty} [p_{\min(v)}(\boldsymbol{c})(1+|\delta_P - v|)^\tau] \leq p_{v_2}(\boldsymbol{c})(1+|\delta_P - v_2|)^\tau$$

$$\leq [p_{v_2}(\boldsymbol{c}^*) - \min_{c\in\Gamma} P_0\{\boldsymbol{c}|\Gamma\}](1+|\delta_P - v_2|)^\tau = [p_{\min(v_1)}(\boldsymbol{c}^*) - \min_{c\in\Gamma} P_0\{\boldsymbol{c}|\Gamma\}](1+|\delta_P - v_2|)^\tau$$

$$< A(\delta_P,\tau,\boldsymbol{c}^*) - \min_{c\in\Gamma} P_0\{\boldsymbol{c}|\Gamma\}(1+|\delta_P - v_2|)^\tau + \varepsilon < A(\delta_P,\tau,\boldsymbol{c}^*),$$

the last inequality holding for $\varepsilon < \min_{c\in\Gamma} P_0\{\boldsymbol{c}|\Gamma\}$. This is a contradiction. □

Unless $|\delta_P - v_S|$ (Section 3) is small, the larger τ is, the less $\varphi_{\delta_P,\tau}$ focuses on omnibus power. Hence, the α-level required for $R_\alpha(\varphi_{\delta_P,\tau,\alpha})$ to contain $D[\Gamma]$ tends to increase in τ.

7. MARGIN-BASED SELECTION OF δ_P AND τ

It may turn out that there is no solid prior information with which to select δ_P or τ. Graubard and Korn (1987) suggested that $\varphi_{0.5}$ be used in the absence of a reason to use a different test. While all linear rank tests, including $\varphi_{0.5}$, may have poor overall power profiles in some cases (BPI, 1998; Berger and Ivanova, 2001), we do feel that it may be reasonable to focus power on $\Omega_{0.5}$, by using $\varphi_{0.5,\tau}$. Only if one uses $\tau = \infty$ is one betting everything on the belief that $\delta(\boldsymbol{\theta}) = 0.5$, but even in this case $\varphi_{0.5,\tau}$ is still preferable to $\varphi_{0.5}$, because $\varphi_{0.5,\infty}$ is a *refinement* of $\varphi_{0.5}$. If δ_P and α are both fixed, but one is unsure of the value of τ to use, then one could use the margins ($\boldsymbol{n}$ and $\boldsymbol{T}$, summarized by Γ) to select τ. Specifically, use the largest τ that allows $R_\alpha(\varphi_{\delta_P,\tau,\alpha})$ to contain $D[\Gamma]$. If a range of α-levels would be considered, say $0.01 \leq \alpha \leq 0.1$, then use the smallest α-level in selecting τ. Restricting attention to the integer values of τ, and using $\delta_P = 0.5$, we note that for $\{(11,2,2),(7,7,6)\}$, $D[\Gamma] = \{(6,9);(15,0)\}$ is contained by $R_{0.01}(\varphi_{0.5,18})$, $R_{0.025}(\varphi_{0.5,20})$, $R_{0.05}(\varphi_{0.5,22})$, and $R_{0.1}(\varphi_{0.5,24})$; but none of $R_{0.01}(\varphi_{0.5,19})$, $R_{0.025}(\varphi_{0.5,21})$, $R_{0.05}(\varphi_{0.5,23})$, or $R_{0.1}(\varphi_{0.5,25})$ contain $(6,9)$. Consequently, we would use $\varphi_{0.5,18}$.

8. COMPARISONS OF TESTS

We compare the exact unconditional power of $\varphi_{0.0}$, $\varphi_{0.5}$, $\varphi_{1.0}$, φ_S, φ_{CH}, $\varphi_{COM(\mathcal{L})}$, and some adaptive tests, considering all possible 2×3 tables with row margins $n_1 = n_2 = 10$. Figure 2 presents Γ-plots. Because Γ is not fixed in this computation, we consider only adaptive tests for which neither δ_P nor τ depends on Γ. We fix $\boldsymbol{\pi}_1 = (0.3, 0.4, 0.3)$ and consider 23 different vectors $\boldsymbol{\pi}_2$ such that $\boldsymbol{\pi}_1$ stochastically dominates $\boldsymbol{\pi}_2$. We are interested in maximizing the power for each of these 23 scenarios, while preserving the type I error rate for the 24th, $\boldsymbol{\pi}_1 = \boldsymbol{\pi}_2$. For each pair $(\boldsymbol{\pi}_1, \boldsymbol{\pi}_2)$ we obtain θ and $\delta(\boldsymbol{\theta}) = 1 - \theta_2/\theta_1$, the optimal score for the linear rank test. Bold entries represent the best power, for given $\delta(\boldsymbol{\theta})$, among the tests we consider. Because the linear rank tests $\varphi_{0.0}$, $\varphi_{0.5}$, and $\varphi_{1.0}$ are excessively conservative, per the bottom row of Table 2, they are dominated (at $\alpha = 0.05$) by their corresponding adaptive tests $\varphi_{0.0,100}$, $\varphi_{0.5,100}$, and $\varphi_{1.0,100}$. This is not surprising and will be the case quite generally. In addition, φ_A dominates $\varphi_{COM(\mathcal{L})}$. Notice that $\varphi_{0.5,1}$ comes close to dominating each other omnibus test (φ_A, $\varphi_{COM(\mathcal{L})}$, φ_S, and φ_{CH}). In fact, only where $\delta(\theta) \leq -2$ is φ_A or $\varphi_{COM(\mathcal{L})}$ more powerful than $\varphi_{0.5,1}$.

[Figure 2], [Table 2]

9. DISCUSSION

In an effort to improve the comparison of two treatments on the basis of ordered categorical data, we defined a new class of adaptive tests. We showed each of these tests to be admissible, while providing unambiguous p-values and a non-iterative construction. There is nothing particular about ordered trinomial distributions that makes this problem especially amenable to treatment with our adaptive approach. For any hypothesis testing problem with a composite alternative hypothesis, one can enumerate the alternatives and the corresponding LMP test for each. One can then apply each of these LMP tests to a given outcome, and find the smallest of the resulting p-values. Using this minimized LMP p-value as a test statistic produces a test analogous to φ_A, and reduces to the uniformly most powerful test if one exists. One can then bridge the gap between φ_A and the LMP tests as we have done, with adaptive tests tailored to fit a favored direction. We would expect this approach to yield good tests in a variety of contexts.

REFERENCES

Berger, V. W. (1998), "Admissibility of Exact Conditional Tests of Stochastic Order," *Journal of Statistical Planning and Inference*, 66, 39-50.

Berger, V. W., and Ivanova A. (2001), "The Bias of Linear Rank Tests when Testing for Stochastic Order in Ordered Categorical Data," *Journal of Statistical Planning and Inference*, in press.

Berger, V. W., Permutt, T., and Ivanova A. (1998), "The Convex Hull Test for Ordered

Categorical Data," *Biometrics*, 54, 1541-1550.

Berger, V. W., and Sackrowitz, H. (1997), "Improving Tests for Superior Treatments in Contingency Tables," *Journal of American Statistical Association*, 92, 438, 700-705.

Chiara, S., Compora, E., Merlini, L., Simoni, C., Iskra, L., Odicino, F., Ragini, N., Conte, P. F., Rosso, R. (1993), "Recurrent Ovarian Carcinoma: Salvage Treatment with Platinum in Patients Responding to First-line Platinum-based Regiments," *European Journal of Cancer*, 29A, 652.

Cohen, A., and Sackrowitz, H. (1998), "Directional Tests for One-sided Alternatives in Multivariate Models," *Annals of Statistics*, 26, 6, 2321-2338.

Donegani, M. (1991), "An Adaptive and Powerful Randomization Test," *Biometrika*, 78, 4, 930-933.

Edgington, E. S. (1995), *Randomization Tests* (third ed.), Marcel Dekker, New York.

Frick, H. (2000), "Undominated p-values and Property *C* for Unconditional One-Sided Two-Sample Binomial Tests," *Biometrical Journal*, 42, 6, 715-728.

Gastwirth, J. L. (1985), "The Use of Maximum Efficiency Robust Tests in Combining Contingency Tables and Survival Analysis," *The Journal of the American Statistical Association* 80, 390, 380-384.

Good, P. (1994), *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*, Springer-Verlag, New York.

Gross, S. T. (1981), "On Asymptotic Power and Efficiency of Tests of Independence in Contingency Tables with Ordered Classifications," *Journal of American Statistical Association*, 76, 376, 935–941.

Graubard, B. I. and Korn, E. L. (1987), "Choice of Column Scores for Testing Independence in Ordered 2xk Contingency Tables," *Biometrics*, 43, 471-476.

Hogg, R. V. (1974), "Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory," *Journal of American Statistical Association*, 69, 348, 909-923.

Ivanova, A., and Berger, V. W. (2001), "Drawbacks of Integer Scoring of Ordered Categorical Data," *Biometrics* (in press).

Marden, J. I. (1991), "Sensitive and Sturdy p-values," *Annals of Statistics*, 19, 2, 918-934.

McCullagh, P. (1980), "Regression methods for ordinal data," *Journal of Royal Statistical Society* B, 42, 2, 109-142.

Permutt, T., and Berger, V. W. (2000), "Rank Tests in Ordered 2xk Contingency Tables," *Communications in Statistics, Theory and Methods*, 29, 5, 989-1003.

Podgor, M. J., Gastwirth, J. L. and Mehta, C. R. (1996), "Efficiency Robust Tests of Independence in Contingency Tables with Ordered Categories," *Statistics in Medicine*, 15, 2095-2105.

Rohmel, J. and Mansmann, U. (1999), "Unconditional Non-Asymptotic One-Sided Tests for Independent Binomial Proportions when the Interest Lies in Showing Non-Inferiority and/or Superiority," *Biometrical Journal*, 41, 2, 149-170.

Roy, S. N. (1953), "On a Heuristic Method of Test Construction and Its Use in Multivariate Analysis," *Annals of Statistics*, 24, 220-238.

Rukhin, A. L., and Mak, K. S. (1992), "Adaptive Test Statistics and Bahadur Efficiency," *Statistica Sinica*, 2, 541-552.

Table 1. Linear rank tests, $\nu \in [0,2]$ for $\{(11,2,2);(7,7,6)\}$.

ν	$o_\nu(11,2)$	Endpoints of: B_ν^+	B_ν^-	p_ν (minimum is underlined)	p_ν^-	p_ν^+	$P_0\{B_\nu^+\}$	$P_0\{B_\nu^-\}$	$p_{\nu,\infty}$	M_ν-$M_{\nu,\infty}$
$\nu \in (-1/7,0)$	1			0.2262	0.2262	0.2262			0.2262	
$\nu = 0$	10	(4,9)-(10,3)	(12,1)-(13,0)	0.2277	0.2262	$\underline{0.0661}$	0.1615	0.0015	0.0726	(7,6)-(10,3)
$\nu \in (0,1/7)$	1			0.0661	0.0661	0.0661			0.0661	
$\nu = 1/7$	2	(5,9)		0.0661	0.0661	$\underline{0.0661}$	$2.1*10^{-5}$		0.0661	
$\nu \in (1/7,1/6)$	1			0.0661	0.0661	0.0661			0.0661	
$\nu = 1/6$	2	(6,8)		0.0661	0.0661	$\underline{0.0657}$	0.0004		0.0661	
$\nu \in (1/6,1/5)$	1			0.0657	0.0657	0.0657			0.0657	
$\nu = 1/5$	2	(7,7)		0.0657	0.0657	$\underline{0.0629}$	0.0028		0.0657	
$\nu \in (1/5,1/4)$	1			0.0629	0.0629	0.0629			0.0629	
$\nu = 1/4$	2	(8,6)		0.0629	0.0629	$\underline{0.0538}$	0.0091		0.0629	
$\nu \in (1/4,2/7)$	1			0.0538	0.0538	0.0538			0.0538	
$\nu = 2/7$	2	(6,9)		0.0538	0.0538	$\underline{0.0538}$	$5.7*10^{-6}$		0.0538	
$\nu \in (2/7,1/3)$	1			0.0538	0.0538	0.0538			0.0538	
$\nu = 1/3$	3	(7,8)-(9,5)		0.0538	0.0538	$\underline{0.0387}$	0.0152		0.0387	(9,5)
$\nu \in (1/3,2/5)$	1			0.0387	0.0387	0.0387			0.0387	
$\nu = 2/5$	2	(8,7)		0.0387	0.0387	$\underline{0.0382}$	0.0005		0.0387	
$\nu \in (2/5,1/2)$	1			0.0382	0.0382	0.0382			0.0382	
$\nu = 1/2$	4	(9,6)-(10,4)	(12,0)	0.0385	0.0382	$\underline{0.0237}$	0.0148	0.0003	0.0249	(10,4)
$\nu \in (1/2,2/3)$	1			0.0237	0.0237	0.0237			0.0237	
$\nu = 2/3$	2	(10,5)		0.0237	0.0237	$\underline{0.0220}$	0.0017		0.0237	
$\nu \in (2/3,1)$	1			0.0220	0.0220	0.0220			0.0220	
$\nu = 1$	5	(11,4)-(11,3)	(11,1)-(11,0)	0.0276	0.0220	**$\underline{0.0198}$**	0.0078	0.0056	0.0276	
$\nu \in (1,3/2)$	1			**0.0198**	**0.0198**	**0.0198**			0.0198	
$\nu = 3/2$	2		(10,0)	0.0205	**$\underline{0.0198}$**	0.0205		0.0008	0.0205	
$\nu \in (3/2,2)$	1			0.0205	0.0205	0.0205			0.0205	
$\nu = 2$	4	(12,3)	(10,1)-(9,0)	0.0294	$\underline{0.0205}$	0.0289	0.0005	0.0089	0.0294	
$\nu \in (2,5/2)$	1			0.0289	0.0289	0.0289			0.0289	

Note that all the values are calculated at (11,2); $p_{v,\infty}$ and $M_{v,\infty}$ are the p-value and extreme region, respectively, of the adaptive test based on v and $\tau = \infty$.

Table 2. Exact unconditional power of the conservative (nonrandomized) versions of linear rank tests (φ_0, φ_1, $\varphi_{0.5}$), adaptive tests ($\varphi_{0,100}$, $\varphi_{1,100}$, $\varphi_{0.5,100}$, $\varphi_{0.5,1}$), omnibus adaptive test φ_A, the $\varphi_{COM(\mathcal{L})}$ test, Smirnov test φ_S, and convex hull test φ_{CH}, with $\alpha \leq 0.05$, and ten observations per row. Bold entries represent best power among these tests for given θ.

$\delta(\theta)$	θ	φ_0	$\varphi_{0,100}$	φ_1	$\varphi_{1,100}$	$\varphi_{0.5}$	$\varphi_{0.5,100}$	$\varphi_{0.5,1}$	φ_A	$\varphi_{COM(\mathcal{L})}$	φ_S	φ_{CH}
-∞	(2.20, ∞)	0.825	**0.895**	0.142	0.569	0.682	0.794	0.865	0.874	0.866	0.820	0.657
-∞	(1.39, ∞)	0.623	**0.782**	0.053	0.289	0.399	0.547	0.715	0.752	0.741	0.604	0.495
-∞	(0.85, ∞)	0.420	**0.643**	0.018	0.130	0.201	0.325	0.567	0.635	0.622	0.389	0.410
-2.00	(0.69, 2.08)	0.247	**0.370**	0.018	0.089	0.137	0.208	0.308	0.327	0.319	0.225	0.201
-1.35	(0.51, 1.20)	0.126	**0.195**	0.019	0.062	0.085	0.122	0.158	0.158	0.154	0.114	0.100
-1.00	(0.29, 0.58)	0.054	**0.093**	0.019	0.045	0.049	0.067	0.077	0.075	0.072	0.051	0.052
-0.78	(1.25, 2.23)	0.418	**0.516**	0.054	0.214	0.300	0.394	0.441	0.434	0.429	0.402	0.270
-0.26	(1.10, 1.39)	0.247	**0.317**	0.054	0.161	0.211	0.267	0.267	0.246	0.243	0.240	0.161
-0.14	(2.08, 2.37)	0.622	**0.684**	0.143	0.455	0.562	0.641	0.624	0.589	0.584	0.619	0.409
0.00	(2.08, 2.08)	0.569	**0.632**	0.148	0.437	0.536	0.604	0.573	0.532	0.526	0.568	0.369
0.13	(0.92, 0.80)	0.126	**0.181**	0.054	0.124	0.138	0.171	0.158	0.140	0.137	0.130	0.107
0.21	(1.95, 1.54)	0.419	**0.492**	0.144	0.366	0.438	0.491	0.446	0.399	0.391	0.427	0.286
0.30	(1.89, 1.33)	0.354	0.432	0.144	0.340	0.396	**0.443**	0.395	0.349	0.339	0.367	0.260
0.40	(1.83, 1.11)	0.287	0.369	0.144	0.313	0.349	**0.392**	0.343	0.300	0.289	0.306	0.238
0.45	(1.80, 0.98)	0.248	0.333	0.144	0.297	0.321	**0.361**	0.314	0.273	0.262	0.272	0.226
0.50	(1.76, 0.88)	0.220	0.305	0.144	0.286	0.300	**0.338**	0.293	0.255	0.243	0.247	0.218
0.58	(0.69, 0.29)	0.054	0.095	0.054	0.103	0.084	**0.106**	0.099	0.089	0.086	0.069	0.078
0.68	(1.61, 0.51)	0.127	0.210	0.144	0.251	0.220	**0.257**	0.224	0.197	0.187	0.167	0.191
0.80	(1.52, 0.31)	0.089	0.165	0.144	**0.238**	0.182	0.219	0.197	0.177	0.168	0.136	0.178
0.90	(1.54, 0.15)	0.067	0.142	0.157	**0.249**	0.164	0.204	0.195	0.179	0.170	0.125	0.182
0.95	(1.39, 0.07)	0.054	0.120	0.144	**0.230**	0.140	0.179	0.177	0.165	0.157	0.111	0.164
1.00	(∞, 1.67)	0.629	0.735	0.360	0.744	0.779	**0.788**	0.722	0.681	0.647	0.657	0.544
1.00	(∞,1.14)	0.425	0.583	0.360	0.641	0.636	**0.661**	0.588	0.542	0.502	0.482	0.475
1.00	(∞, 0.69)	0.252	0.431	0.361	**0.560**	0.488	0.534	0.471	0.430	0.396	0.342	0.436
1.00	(∞, 0.29)	0.128	0.292	0.361	**0.509**	0.349	0.418	0.390	0.360	0.335	0.248	0.401
1.00	(∞,-0.12)	0.055	0.177	0.361	**0.495**	0.233	0.321	0.362	0.344	0.323	0.201	0.363
1.00	(∞, -0.56)	0.019	0.093	0.360	**0.522**	0.145	0.245	0.399	0.389	0.360	0.192	0.335
1.37	(1.10, -0.41)	0.019	0.060	0.144	**0.239**	0.083	0.126	0.178	0.174	0.166	0.090	0.143
1.55	(0.41, -0.22)	0.019	0.044	0.054	**0.101**	0.047	0.065	0.079	0.077	0.075	0.043	0.061
	(0.00, 0.00)	0.019	0.039	0.019	0.039	0.026	0.035	0.042	0.041	0.040	0.023	0.030

Figure 1. Permutation sample space for $\{(11,2,2);(7,7,6)\}$.

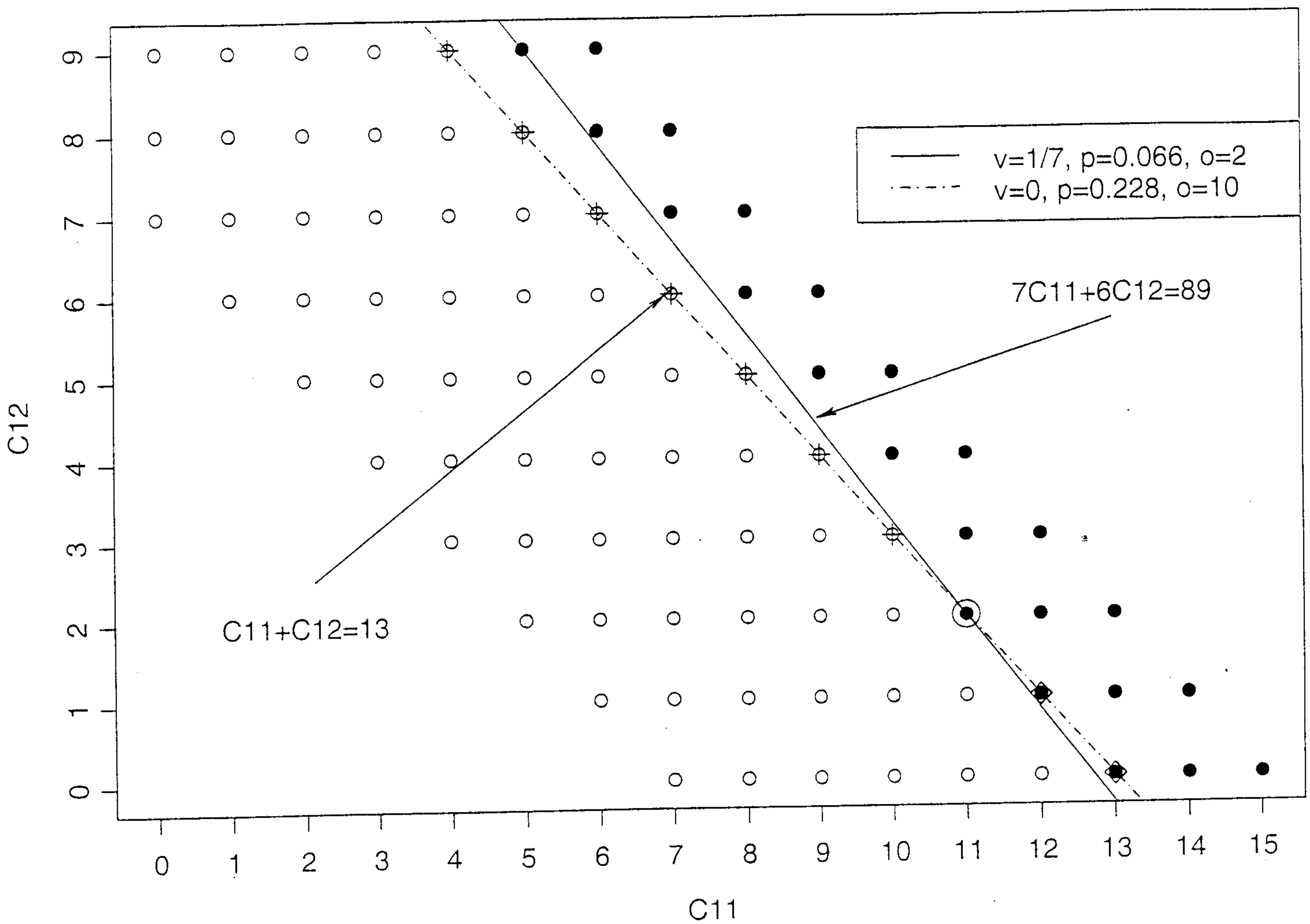

Figure 2. Rejection regions and p-values for several tests for $\{(11,2,2);(7,7,6)\}$.

Linear rank test $\varphi_{0.5}$, and adaptive tests $\varphi_{0.5,1}$, $\varphi_{0.5,3}$, $\varphi_{0.5,100}$, the omnibus adaptive test, φ_A, the Smirnov test φ_S, and the convex hull test φ_{CH}, and the $\varphi_{COM(\mathcal{L})}$ test.