DOI:10.2478/bile-2026-0009

Biometrical Letters
Vol. 0 (2026), No. 0, 1-20

A contrast-ratio representation of directional departure from marginal mean equality in ordinal square contingency tables

Shuji Ando

Department of Information Sciences, Faculty of Science and Technology,
Tokyo University of Science, Noda City, Chiba, 278-8510, Japan,
e-mail: shuji.ando@rs.tus.ac.jp

Summary

For ordinal square contingency tables with equally spaced category scores, we give an explicit contrast-ratio representation of the angular measure of departure from marginal mean equality introduced by Ando (2024). The bounded contrast D and the angular measure Ψ satisfy $\Psi = (4/\pi)\arctan(D)$ and therefore contain the same directional information. Ratio and half-log-ratio identities clarify the interpretation of the aggregated cumulative marginal quantities. A ratio-of-means formulation gives a compact asymptotic variance and two simple confidence intervals: a naive Wald interval and an atanh-transformed Wald interval. The transformed interval respects the parameter range whenever defined, but is unavailable at boundary estimates. Simulations examine departures at a fixed denominator and null scenarios with smaller denominators; performance for non-null departures combined with a small denominator is not established. Paired insomnia data illustrate reporting on equivalent scales and the coding-dependent interpretation of between-group comparisons. The contribution is improved representation, interpretation, and formulation of large-sample inference, rather than additional information or a general finite-sample guarantee.

Key words: cumulative marginal probability, directional measure, marginal mean equality, ordinal categorical data, ratio, square contingency table

1. Introduction

Square contingency tables arise from paired categorical observations with the same row and column categories. Classical models include symmetry and marginal homogeneity (MH), the latter requiring equal marginal distributions (Bowker, 1948; Stuart, 1955; Ireland, Ku and Kullback, 1969). For ordered categories, marginal mean equality (ME) provides a weaker summary of marginal change (Tomizawa, 1991; Agresti, 2010). Measures of departure complement goodness-of-fit tests by quantifying discrepancies from these models (Tomizawa, 1995; Tomizawa, Miyamoto and Ashihara, 2003; Ando, Fujimoto and Tomizawa, 2024).

Directional measures additionally distinguish opposite forms of marginal change (Ando, 2019, 2021, 2022; Iki, Nakano and Tomizawa, 2019). Ando (2024) proposed an angular measure of departure from ME using two aggregated cumulative marginal quantities. Its geometric construction appears two-dimensional, but the normalized coordinates sum to one. An explicit scalar representation can therefore clarify what the construction measures and how its values should be reported.

This paper characterizes that representation through a bounded contrast D. We show that the angular measure is an exact monotone transformation of D, and connect both to a ratio and a half-log-ratio. The contribution is a more transparent representation, interpretation, and formulation of standard large-sample inference for the existing measure, not a new measure carrying additional information. Reparameterization alone does not imply uniformly better finite-sample inference. We consider a naive Wald interval and an atanh-transformed interval as simple illustrations, rather than undertaking a comprehensive comparison of confidence-interval methods.

Section 2 gives the representation and its interpretation, section 3 develops estimation and inference, and sections 4 and 5 present simulations and an application. Section 6 sets out practical recommendations and limitations. Complete numerical tables are retained in Appendix A.

2. Contrast-ratio representation

2.1. Definition and exact relationship

Let $p_{ij} = \Pr(X = i, Y = j)$, $i, j = 1, \ldots, R$, denote the cell probabilities of an ordinal square table, where X and Y use the common, equally spaced scores $1, \ldots, R$. The probabilities are nonnegative and sum to one. Write $F_i^X = \Pr(X \leqslant i)$ and $F_i^Y = \Pr(Y \leqslant i)$, and define

$$A = \sum_{i=1}^{R-1} F_i^X = R - E(X), \qquad B = \sum_{i=1}^{R-1} F_i^Y = R - E(Y). \tag{1}$$

The ME model $E(X) = E(Y)$ is equivalent to $A = B$.

Throughout this paper, we assume $A+B > 0$. The excluded case $A+B = 0$ occurs only when $\Pr(X = R, Y = R) = 1$, for which neither the angular nor the contrast representation considered here is defined.

Ando (2024) defines $w_X = A/(A + B)$ and $w_Y = B/(A + B)$ and uses

$$\Psi = \frac{4}{\pi}\left[\arccos\left\{\frac{w_X}{\sqrt{w_X^2 + w_Y^2}}\right\} - \frac{\pi}{4}\right]. \tag{2}$$

Since $w_X + w_Y = 1$, the normalized pair has one degree of freedom. Moreover, $w_X/\sqrt{w_X^2 + w_Y^2} = A/\sqrt{A^2 + B^2}$, so the common normalization cancels from the angular construction. Consider the contrast

$$D = w_Y - w_X = \frac{B - A}{A + B}. \tag{3}$$

The following identities connect the existing angular measure to a bounded reporting scale, an interpretable ratio, and an unbounded working scale for inference.

Proposition 2.1. *For $A, B \geqslant 0$ with $A+B > 0$, the contrast in equation* (3) *and the angular measure in equation* (2) *satisfy*

$$\Psi = \frac{4}{\pi}\arctan(D), \qquad D = \tan\left(\frac{\pi}{4}\Psi\right). \tag{4}$$

Thus, they have the same sign and induce the same directional ordering. If $A, B > 0$, then, with $\rho = B/A$,

$$D = \frac{\rho - 1}{\rho + 1}, \qquad \rho = \frac{1 + D}{1 - D}, \tag{5}$$

$$\eta = \operatorname{atanh}(D) = \frac{1}{2}\log\rho = \frac{1}{2}\log\frac{B}{A}. \tag{6}$$

Proof. Let $\theta = \arccos\{A/\sqrt{A^2 + B^2}\} \in [0, \pi/2]$. Then

$$\tan(\theta - \pi/4) = \frac{\sin\theta - \cos\theta}{\cos\theta + \sin\theta} = \frac{B - A}{A + B} = D.$$

Since $\theta - \pi/4 \in [-\pi/4, \pi/4]$, it equals $\arctan(D)$, including when $A = 0$ or $B = 0$. Substitution gives equation (4); monotonicity gives the ordering statement. For positive A and B, dividing the numerator and denominator in equation (3) by A and rearranging gives equations (5) and (6). □

2.2. Interpretation and ordinal coding

The contrast lies in $[-1, 1]$, equals zero exactly under ME, and changes sign on interchanging X and Y. Its endpoints require $B = 0 < A$ or $A = 0 < B$, respectively. For example, $D = 0.20$ gives $B/A = 1.50$: the aggregated column cumulative marginal quantity is 1.5 times its row counterpart, not an improvement probability ratio.

From equation (1),

$$D = \frac{E(X) - E(Y)}{2R - E(X) - E(Y)}. \tag{7}$$

Thus, D has the sign of the marginal mean difference. Its denominator

depends on the marginal distributions and is estimated from the data; D is not the proportion of the maximum possible mean shift.

Remark 2.1 (Dependence on the direction of ordinal coding)**.** *For $X^* = R + 1 - X$ and $Y^* = R + 1 - Y$, the reversed contrast is defined when $E(X) + E(Y) > 2$ and then equals*

$$D^* = \frac{E(Y) - E(X)}{E(X) + E(Y) - 2}. \tag{8}$$

When both contrasts are defined, the sign reverses, but the magnitude and rankings of tables by magnitude need not be preserved. The same dependence is inherited by Ψ. The coding direction should therefore be fixed on substantive grounds and held common across comparisons.

3. Estimation and inference

3.1. Estimator and asymptotic variance

Suppose (X_n, Y_n), $n = 1, \ldots, N$, are independent copies of (X, Y). Let f_{ij} be the cell counts, with $\widehat{p}_{ij} = f_{ij}/N$. Replacing the probabilities in equation (1) by their empirical counterparts gives $\widehat{A} = R - \overline{X}$ and $\widehat{B} = R - \overline{Y}$. Define $U = X - Y$ and $V = 2R - X - Y$. Whenever $\overline{V} > 0$,

$$\widehat{D} = \frac{\widehat{B} - \widehat{A}}{\widehat{A} + \widehat{B}} = \frac{\overline{U}}{\overline{V}}, \qquad D = \frac{E(U)}{E(V)}. \tag{9}$$

Proposition 3.1. *Under multinomial sampling, if $\mu_V = E(V) > 0$ and $\mathrm{Var}(U - DV) > 0$, then*

$$\sqrt{N}(\widehat{D} - D) \xrightarrow{d} N(0, \sigma_D^2), \qquad \sigma_D^2 = \frac{\mathrm{Var}(U - DV)}{\mu_V^2}. \tag{10}$$

Proof. On the event $\overline{V} > 0$,

$$\sqrt{N}(\widehat{D} - D) = \frac{1}{\overline{V}} \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (U_n - DV_n).$$

The summands have mean zero and finite variance because the scores are bounded. The central limit theorem, $\overline{V} \xrightarrow{p} \mu_V > 0$, and Slutsky's theorem give the result. The event $\overline{V} > 0$ has probability tending to one. □

Since $E(U - DV) = 0$, a consistent plug-in variance and the estimated standard error are

$$\widehat{\sigma}_D^2 = \frac{\sum_{i=1}^{R} \sum_{j=1}^{R} \widehat{p}_{ij} \{(i-j) - \widehat{D}(2R - i - j)\}^2}{(\widehat{A} + \widehat{B})^2}, \qquad \widehat{\mathrm{SE}}(\widehat{D}) = \frac{\widehat{\sigma}_D}{\sqrt{N}}. \tag{11}$$

Although D depends only on the marginal means, its sampling variance depends on the joint distribution of the paired observations. The condition $\mathrm{Var}(U - DV) > 0$ excludes degenerate tables, such as $\Pr(X = Y) = 1$. A zero plug-in variance in a particular sample can yield a degenerate interval and is not evidence of exact coverage.

The delta method also gives

$$\sigma_\Psi^2 = \left(\frac{4}{\pi}\right)^2 \frac{\sigma_D^2}{(1 + D^2)^2}, \qquad \sigma_\eta^2 = \frac{\sigma_D^2}{(1 - D^2)^2} \quad (A, B > 0). \tag{12}$$

3.2. Confidence intervals and availability

A conventional approximate $100(1 - \alpha)\%$ confidence interval is

$$CI_{\mathrm{Wald}} = \widehat{D} \ \pm z_{1-\alpha/2} \widehat{\mathrm{SE}}(\widehat{D}). \tag{13}$$

This naive Wald interval is not constrained to $[-1, 1]$. If $\widehat{A}, \widehat{B} > 0$, define

$$\widehat{\eta} = \mathrm{atanh}(\widehat{D}), \qquad \widehat{\mathrm{SE}}(\widehat{\eta}) = \frac{\widehat{\mathrm{SE}}(\widehat{D})}{1 - \widehat{D}^2}.$$

Writing $L_\eta = \widehat{\eta} - z_{1-\alpha/2}\widehat{\mathrm{SE}}(\widehat{\eta})$ and $U_\eta = \widehat{\eta} + z_{1-\alpha/2}\widehat{\mathrm{SE}}(\widehat{\eta})$ gives

$$CI_{\mathrm{atanh}} = [\tanh(L_\eta), \tanh(U_\eta)]. \tag{14}$$

This interval lies in $(-1, 1)$ whenever defined, but range preservation does not guarantee nominal finite-sample coverage. If $\widehat{A} + \widehat{B} = 0$, the contrast is undefined. Conditional on a positive denominator, $\widehat{D} = 1$ exactly when all observed $X_n = R$, and $\widehat{D} = -1$ exactly when all observed $Y_n = R$. In either boundary case the atanh interval is unavailable. Thus, boundary estimates arise from a degenerate observed margin, not simply from zero cell counts. No continuity correction or specialized boundary inference is introduced.

Symmetric Wald intervals constructed separately on the D and Ψ scales need not coincide after transformation. In contrast, monotonically transforming the endpoints of one interval preserves its coverage event exactly. Section 6 uses this distinction to give a consistent reporting recommendation.

4. Simulation study

4.1. Designs and evaluation

The simulations examine basic finite-sample behavior of the estimator and the two intervals, not a comprehensive class of interval procedures. Both designs use $R = 4$, $N = 30, 50, 100, 200$, and main-diagonal effects $\gamma = 0, 1, 2$. The principal design uses

$$r_i(\beta) = \frac{\exp(\beta i)}{\sum_{k=1}^{R} \exp(\beta k)}, \qquad c_i(\beta) = r_{R+1-i}(\beta), \tag{15}$$

with β chosen numerically to satisfy

$$E_\beta(X) = \frac{R + 1 + (R-1)d}{2}, \qquad d = 0,\ \pm 0.2,\ \pm 0.5,\ \pm 0.8. \tag{16}$$

This gives $D = d$ and fixes $A + B = R - 1 = 3$. The joint probabilities are

$$p_{ij} \propto a_i b_j \exp\{\gamma I(i = j)\}, \tag{17}$$

where positive scaling factors a_i, b_j are obtained by iterative proportional fitting to retain the specified margins. Here γ controls the main-diagonal effect, not a general association parameter; $\gamma = 0$ gives independence. Realized diagonal probabilities depend on the margins and are listed in Table A1.

The sensitivity design instead sets $c_i(\beta) = r_i(\beta)$, so $D = 0$ throughout. Values $\beta = 0, 1, 2$ give $A + B = 3.0000, 1.0147, 0.3104$, respectively, using the same N and γ values. The $\beta = 0$ setting repeats the principal null probability tables with independent Monte Carlo replication. These designs separately investigate departures at a fixed denominator and smaller denominators under the null; they do not investigate non-null D combined with a small denominator.

For each scenario, $M = 10{,}000$ multinomial samples were generated. With $M = 10{,}000$, the Monte Carlo standard error of an estimated coverage probability near 0.95 is approximately 0.0022.

Bias and RMSE are calculated over replications with $\overline{V} > 0$. Let $E_W = \{\overline{V} > 0\}$ and $E_T = \{\widehat{A} > 0, \widehat{B} > 0\}$ denote availability for the naive and transformed formulas. Reported coverage and mean width are conditional on the corresponding event. Thus, the two conditional coverages do not use identical subsets when boundary estimates occur. We also record the frequency of out-of-range naive intervals, boundary estimates, and zero denominators. For the transformed interval, overall coverage counts unavailable intervals as failures to cover:

$$C_T^{\text{all}} = \Pr(E_T,\ D \in CI_{\text{atanh}}) = \Pr(E_T)\Pr(D \in CI_{\text{atanh}} \mid E_T). \tag{18}$$

4.2. Results

In the principal design, the absolute bias did not exceed 0.0040 across the 84 scenarios, and RMSE decreased with N (Table A2). Changing d to $-d$

transposes the probability table, so the distributions of $\widehat{D}$ are exact mirror images. Table 1 therefore displays nonnegative d; Table A3 retains both signs and all interval widths.

For $|D| = 0.8$ and $N = 30$, naive Wald coverage was approximately 0.904–0.911, whereas conditional atanh coverage was 0.939–0.945. Differences were generally smaller for moderate departures and larger samples. Naive intervals extended outside $[-1, 1]$ in approximately 5.4–5.9% of samples with $|D| = 0.8$ and $N = 30$. No zero denominators occurred in the principal design; boundary estimates occurred in at most 0.06% of replications, giving atanh availability of at least 99.94%.

Table 1. Principal-design 95% coverage probabilities for $D \geqslant 0$. Each entry is naive Wald/atanh coverage, conditional on the respective availability event. The denominator is $A + B = 3$ in every scenario. Both signs and all widths are retained in Table A3.

True D	γ	$N = 30$	$N = 50$	$N = 100$	$N = 200$
0	0	0.939/0.946	0.942/0.945	0.948/0.950	0.948/0.949
0	1	0.941/0.945	0.947/0.950	0.950/0.951	0.950/0.950
0	2	0.949/0.951	0.948/0.949	0.946/0.948	0.944/0.945
0.2	0	0.940/0.945	0.945/0.948	0.949/0.951	0.950/0.951
0.2	1	0.940/0.945	0.941/0.946	0.946/0.947	0.948/0.949
0.2	2	0.932/0.936	0.937/0.940	0.943/0.943	0.946/0.946
0.5	0	0.932/0.945	0.936/0.945	0.945/0.946	0.945/0.947
0.5	1	0.929/0.941	0.938/0.945	0.943/0.947	0.943/0.944
0.5	2	0.933/0.940	0.938/0.944	0.946/0.950	0.951/0.953
0.8	0	0.907/0.942	0.925/0.949	0.937/0.949	0.944/0.946
0.8	1	0.906/0.941	0.924/0.946	0.936/0.947	0.945/0.952
0.8	2	0.905/0.940	0.923/0.947	0.933/0.947	0.942/0.950

In the null sensitivity design, estimation became less precise as the margins became concentrated and $A + B$ decreased. For $N = 30$, the RMSE ranges across γ were 0.0612–0.0989, 0.1394–0.2036, and 0.2893–0.3682 for $A + B = 3.000, 1.015, 0.310$, respectively. Table 2 retains all sensitivity scenarios; full width and out-of-range diagnostics are in Table A4.

For $A + B = 0.310$ and $N = 30$, naive conditional coverage was 0.865–

0.898 and atanh conditional coverage was 0.955–0.973. The latter included overcoverage relative to 0.95. Counting unavailable transformed intervals as noncoverage gave overall atanh coverage of 0.948, 0.932, and 0.940 for $\gamma = 0, 1, 2$, respectively. These overall values must be distinguished from the conditional summaries in the table. Boundary estimates occurred in 2.01–2.50% of replications and zero denominators in 0.01–0.25%; naive intervals were out of range in 9.98–31.07% of valid samples. By $N = 200$, conditional coverages in this smallest-denominator setting were 0.939–0.942 and 0.946–0.952 for the naive and transformed intervals, respectively.

These findings concern the settings investigated. The fixed-denominator design does not assess small-denominator inference away from the null, and the small-denominator design assesses coverage only at $D = 0$. The results neither establish performance for non-null D combined with a small denominator nor justify choosing the transformed interval on the basis of the denominator alone. Conditional coverage must be considered together with interval availability and overall coverage; a higher coverage probability is not necessarily closer to the nominal level.

5. Application to insomnia data

5.1. Data

We use paired ordinal data from a randomized, double-blind trial of an active hypnotic drug versus placebo, reported by Francom, Chuang-Stein and Landis (1989) and reproduced by Agresti (2019, Table 9.6). These data have also illustrated related departure measures (Iki, Nakano and Tomizawa, 2019; Ando, Fujimoto and Tomizawa, 2024). Time to falling asleep was categorized as less than 20 minutes (1), 20 to less than 30 (2), 30 to less than 60 (3), and 60 or more (4). Table 3 uses baseline as the row variable X and follow-up as the column variable Y. Positive D therefore indicates a lower mean category score at follow-up under this coding, not a change measured directly in minutes.

Table 2. Null sensitivity analysis ($D = 0$). W/T denotes naive Wald/atanh coverage, conditional on the respective availability event. The last two columns are percentages of replications with a defined boundary contrast and with an undefined contrast, respectively. Table A4 includes all widths and out-of-range diagnostics.

β	$A + B$	γ	N	W/T	$\widehat{D} = \pm 1$	$\overline{V} = 0$
0	3.000	0	30	0.936/0.941	0.00	0.00
0	3.000	0	50	0.949/0.951	0.00	0.00
0	3.000	0	100	0.948/0.949	0.00	0.00
0	3.000	0	200	0.950/0.950	0.00	0.00
0	3.000	1	30	0.945/0.949	0.00	0.00
0	3.000	1	50	0.944/0.946	0.00	0.00
0	3.000	1	100	0.946/0.948	0.00	0.00
0	3.000	1	200	0.952/0.953	0.00	0.00
0	3.000	2	30	0.948/0.950	0.00	0.00
0	3.000	2	50	0.951/0.952	0.00	0.00
0	3.000	2	100	0.948/0.949	0.00	0.00
0	3.000	2	200	0.950/0.951	0.00	0.00
1	1.015	0	30	0.917/0.942	0.00	0.00
1	1.015	0	50	0.930/0.946	0.00	0.00
1	1.015	0	100	0.944/0.949	0.00	0.00
1	1.015	0	200	0.943/0.947	0.00	0.00
1	1.015	1	30	0.917/0.936	0.00	0.00
1	1.015	1	50	0.934/0.943	0.00	0.00
1	1.015	1	100	0.941/0.948	0.00	0.00
1	1.015	1	200	0.944/0.946	0.00	0.00
1	1.015	2	30	0.926/0.939	0.00	0.00
1	1.015	2	50	0.936/0.944	0.00	0.00
1	1.015	2	100	0.943/0.946	0.00	0.00
1	1.015	2	200	0.948/0.949	0.00	0.00
2	0.310	0	30	0.865/0.973	2.50	0.01
2	0.310	0	50	0.902/0.955	0.10	0.00
2	0.310	0	100	0.928/0.951	0.00	0.00
2	0.310	0	200	0.939/0.952	0.00	0.00
2	0.310	1	30	0.866/0.955	2.34	0.05
2	0.310	1	50	0.908/0.949	0.15	0.00
2	0.310	1	100	0.929/0.944	0.00	0.00
2	0.310	1	200	0.942/0.951	0.00	0.00
2	0.310	2	30	0.898/0.962	2.01	0.25
2	0.310	2	50	0.912/0.943	0.13	0.00
2	0.310	2	100	0.931/0.944	0.00	0.00
2	0.310	2	200	0.939/0.946	0.00	0.00

Table 3. Square contingency tables of time to falling asleep at follow-up, stratified by time to falling asleep at baseline, for the active and placebo groups. Source: Agresti (2019, Table 9.6); data originally reported by Francom, Chuang-Stein and Landis (1989).

	Follow-up				
Baseline	(1)	(2)	(3)	(4)	Total
(a) Active group					
Less than 20 (1)	7	4	1	0	12
From 20 to less than 30 (2)	11	5	2	2	20
From 30 to less than 60 (3)	13	23	3	1	40
60 or more (4)	9	17	13	8	47
Total	40	49	19	11	119
(b) Placebo group					
Less than 20 (1)	7	4	2	1	14
From 20 to less than 30 (2)	14	5	1	0	20
From 30 to less than 60 (3)	6	9	18	2	35
60 or more (4)	4	11	14	22	51
Total	31	29	35	25	120

5.2. Results

The estimated baseline and follow-up means were 3.025 and 2.008 for active treatment, and 3.025 and 2.450 for placebo. Table 4 presents the contrasts, equivalent angular values and ratios, and both intervals. It also displays $\widehat{A}+\widehat{B}$ to make the different normalizations explicit. The active-group ratio $\widehat{B}/\widehat{A} = 2.043$ compares aggregated cumulative marginal quantities; it does not mean that the probability of improvement doubled or that sleep latency was halved.

Table 4. Angular and contrast-ratio representations for the insomnia data. SE denotes the estimated standard error of $\widehat{D}$. The denominator $\widehat{s} = \widehat{A}+\widehat{B}$ is displayed to clarify the group-specific normalization.

Group	$\widehat{s}$	$\widehat{D}$ (SE)	$\widehat{\Psi}$	$\widehat{\rho}$	Naive Wald 95% CI	atanh Wald 95% CI
Active	2.966	0.343 (0.039)	0.420	2.043	(0.267, 0.419)	(0.265, 0.416)
Placebo	2.525	0.228 (0.039)	0.285	1.590	(0.151, 0.304)	(0.150, 0.303)

Both contrasts were positive, and the active-group estimate was larger. For independent groups, the estimated standard error of $\widehat{D}_A - \widehat{D}_P$ is $\{\widehat{\mathrm{SE}}(\widehat{D}_A)^2 + \widehat{\mathrm{SE}}(\widehat{D}_P)^2\}^{1/2}$. Here the difference was 0.115, its standard error was 0.055, and its approximate 95% Wald interval was $(0.007, 0.223)$. This is a comparison of normalized directional contrasts under the common coding $1, \ldots, 4$. In particular, writing $\Delta_g = E(X_g) - E(Y_g)$ and $s_g = 2R - E(X_g) - E(Y_g)$ for group g,

$$D_A - D_P = \frac{\Delta_A}{s_A} - \frac{\Delta_P}{s_P}. \tag{19}$$

It is not the difference in mean category changes $\Delta_A - \Delta_P$, nor a treatment effect measured in minutes. Each normalization is determined by its group-specific marginal distributions and estimated from that group's data. The category intervals are also unequal in width, so their assigned scores must not be interpreted as elapsed minutes.

Reversing the category direction changes the sign and generally the magnitude of each contrast, as shown in Remark 2.1. Thus the between-group comparison is coding-dependent and is included as an illustration on the specified reporting scale, not as a direct treatment effect on the original ordinal scale or as a replacement for the trial's prespecified treatment-effect analysis.

6. Discussion

The contrast-ratio representation exposes the one-dimensional structure of the angular measure of Ando (2024). It preserves exactly the same directional information while making interpretation and standard large-sample calculations more explicit. The contribution is this characterization and its use, not a new ordering, increased discriminatory information, or a general claim of improved finite-sample inference.

For practical reporting, D may serve as the primary bounded summary, with the category scores and direction stated explicitly. The ratio B/A may

be added when a multiplicative interpretation of the aggregated cumulative marginal quantities is useful. The half-log-ratio $\operatorname{atanh}(D)$ is mainly a working scale for inference; its confidence limits are ordinarily transformed back to the reporting scale. If the transformed interval for D is $[L, U] \subset (-1, 1)$, consistent intervals on the ratio and angular scales are

$$\left[\frac{1+L}{1-L}, \frac{1+U}{1-U}\right] \quad \text{and} \quad \left[\frac{4}{\pi}\arctan(L), \frac{4}{\pi}\arctan(U)\right], \tag{20}$$

respectively. These are the same coverage event expressed on different scales, not separate confirmations. Boundary estimates and unavailable intervals should be disclosed rather than treated as resolved by transformation.

Interpretation remains tied to the fixed category scores and to normalization by the aggregated cumulative marginal quantities. In particular, neither a value of D nor a between-group difference in D is a percentage improvement or a raw mean change. Reversing the ordinal coding can change magnitudes and rankings. Unequally spaced scores would require modifying the marginal-mean identities in equation (1). Although the parameter uses only the marginal means, inference must retain the dependence between the paired observations.

The simulations assess only two simple intervals. The principal design fixes $A + B = R - 1$, whereas the small-denominator sensitivity design is restricted to $D = 0$. Its conditional coverage results, including overcoverage in some settings, and its overall coverage results do not establish performance for non-null contrasts combined with a small denominator. They do not support a general preference based on the denominator alone. Range preservation is an algebraic property, not a coverage guarantee, and a heavily concentrated observed margin may make the atanh interval unavailable. A systematic study of the combined non-null/small-denominator regime and specialized inference for boundary or degenerate samples remain separate topics for further work.

Acknowledgements

The author would like to thank the anonymous reviewers and the editors for their comments and suggestions to improve this paper.

References

Agresti A. (2010): Analysis of Ordinal Categorical Data, 2nd ed. Wiley, Hoboken.

Agresti A. (2019): An Introduction to Categorical Data Analysis, 3rd ed. Wiley, Hoboken.

Ando S. (2019): A bivariate index for visually measuring marginal inhomogeneity in square tables. International Journal of Statistics and Probability 8: 58–65.

Ando S. (2021): An index to simultaneously analyze the degree and directionality of departure from global marginal homogeneity in square contingency tables. Journal of the Korean Statistical Society 50: 997–1008.

Ando S. (2022): Directional measure for analyzing the degree of deviance from generalized marginal mean equality model in square contingency tables. Sankhyā B 84: 708–721.

Ando S. (2024): Measure of deviancy from marginal mean equality based on cumulative marginal probabilities in square contingency tables. Metrika 87: 1037–1048.

Ando S., Fujimoto K., Tomizawa S. (2024): A measure of departure from marginal homogeneity using continuation odds for square contingency tables with ordered categories. Sankhyā B 86: 311–332.

Bowker A.H. (1948): A test for symmetry in contingency tables. Journal of the American Statistical Association 43: 572–574.

Francom S.F., Chuang-Stein C., Landis J.R. (1989): A log-linear model for ordinal data to characterize differential change among treatments. Statistics in Medicine 8: 571–582.

Iki K., Nakano H., Tomizawa S. (2019): A directional measure for marginal homogeneity in square contingency tables with ordered categories. Biometrical Letters 56: 1–11.

Ireland C.T., Ku H.H., Kullback S. (1969): Symmetry and marginal homogeneity of an $r \times r$ contingency table. Journal of the American Statistical Association 64: 1323–1341.

Stuart A. (1955): A test for homogeneity of the marginal distributions in a two-way classification. Biometrika 42: 412–416.

Tomizawa S. (1991): Decomposing the marginal homogeneity model into two models for square contingency tables with ordered categories. Calcutta Statistical Association Bulletin 41: 201–207.

Tomizawa S. (1995): Measures of departure from marginal homogeneity for contingency tables with nominal categories. Journal of the Royal Statistical Society Series D: The Statistician 44: 425–439.

Tomizawa S., Miyamoto N., Ashihara N. (2003): Measure of departure from marginal homogeneity for square contingency tables having ordered categories. Behaviormetrika 30: 173–193.

Appendix A. Complete simulation tables

Tables A1–A4 give the complete numerical results for the designs in section 4.1. Table A1 lists the probability-table characteristics, Table A2 gives bias and RMSE, and Tables A3 and A4 give interval results for the principal and null sensitivity designs, respectively. Both signs of the principal contrast are retained. Tables 1 and 2 present reduced versions for the main discussion.

Table A1. Simulation probability-table characteristics.

True D	β	γ	$\Pr(X = Y)$	Diagonal-effect setting
−0.8	−1.4340	0	0.0316	independence
−0.8	−1.4340	1	0.0509	diagonal effect
−0.8	−1.4340	2	0.0723	stronger diagonal effect
−0.5	−0.6740	0	0.1464	independence
−0.5	−0.6740	1	0.2458	diagonal effect
−0.5	−0.6740	2	0.3257	stronger diagonal effect
−0.2	−0.2440	0	0.2322	independence
−0.2	−0.2440	1	0.4292	diagonal effect
−0.2	−0.2440	2	0.6076	stronger diagonal effect
0	0.0000	0	0.2500	independence
0	0.0000	1	0.4754	diagonal effect
0	0.0000	2	0.7112	stronger diagonal effect
0.2	0.2440	0	0.2322	independence
0.2	0.2440	1	0.4292	diagonal effect
0.2	0.2440	2	0.6076	stronger diagonal effect
0.5	0.6740	0	0.1464	independence
0.5	0.6740	1	0.2458	diagonal effect
0.5	0.6740	2	0.3257	stronger diagonal effect
0.8	1.4340	0	0.0316	independence
0.8	1.4340	1	0.0509	diagonal effect
0.8	1.4340	2	0.0723	stronger diagonal effect

Table A2. Bias and RMSE of $\widehat{D}$. Each entry gives bias with RMSE in parentheses.

True D	γ	$N = 30$	$N = 50$	$N = 100$	$N = 200$
−0.8	0	−0.0027 (0.0659)	−0.0019 (0.0505)	−0.0005 (0.0357)	−0.0003 (0.0255)
−0.8	1	−0.0029 (0.0652)	−0.0002 (0.0513)	−0.0009 (0.0365)	−0.0005 (0.0253)
−0.8	2	−0.0013 (0.0667)	−0.0006 (0.0517)	−0.0006 (0.0363)	0.0002 (0.0254)
−0.5	0	−0.0028 (0.0910)	−0.0013 (0.0692)	−0.0010 (0.0497)	0.0000 (0.0352)
−0.5	1	−0.0035 (0.0872)	−0.0019 (0.0676)	−0.0005 (0.0479)	−0.0010 (0.0338)
−0.5	2	−0.0033 (0.0881)	−0.0008 (0.0683)	−0.0016 (0.0480)	−0.0006 (0.0340)
−0.2	0	−0.0030 (0.0973)	−0.0014 (0.0741)	−0.0005 (0.0524)	−0.0005 (0.0369)
−0.2	1	−0.0030 (0.0855)	−0.0013 (0.0650)	−0.0007 (0.0461)	−0.0002 (0.0328)
−0.2	2	−0.0029 (0.0746)	−0.0020 (0.0562)	−0.0002 (0.0403)	−0.0003 (0.0285)
0	0	−0.0009 (0.0969)	−0.0003 (0.0755)	−0.0008 (0.0527)	−0.0002 (0.0375)
0	1	0.0006 (0.0819)	−0.0008 (0.0623)	0.0002 (0.0442)	−0.0006 (0.0312)
0	2	0.0000 (0.0608)	−0.0002 (0.0471)	0.0002 (0.0334)	0.0000 (0.0235)
0.2	0	0.0009 (0.0955)	0.0004 (0.0744)	−0.0002 (0.0526)	0.0001 (0.0370)
0.2	1	0.0028 (0.0845)	0.0009 (0.0657)	0.0006 (0.0461)	0.0002 (0.0326)
0.2	2	0.0029 (0.0744)	0.0019 (0.0573)	0.0011 (0.0408)	0.0004 (0.0284)
0.5	0	0.0026 (0.0901)	0.0018 (0.0702)	0.0010 (0.0492)	0.0001 (0.0349)
0.5	1	0.0040 (0.0879)	0.0013 (0.0688)	0.0008 (0.0479)	0.0002 (0.0343)
0.5	2	0.0029 (0.0880)	0.0026 (0.0686)	0.0015 (0.0482)	0.0004 (0.0336)
0.8	0	0.0021 (0.0657)	0.0018 (0.0505)	0.0006 (0.0359)	0.0003 (0.0256)
0.8	1	0.0017 (0.0664)	0.0012 (0.0509)	0.0005 (0.0365)	0.0005 (0.0254)
0.8	2	0.0022 (0.0672)	0.0017 (0.0516)	0.0011 (0.0364)	0.0000 (0.0259)

Table A3. Coverage probabilities and average widths of the naive Wald and atanh-transformed Wald confidence intervals. Each entry is shown as "Wald coverage/atanh coverage [Wald width/atanh width]". Coverage of the atanh-transformed interval is conditional on the interval being defined. Naive Wald summaries condition on $\overline{V} > 0$; no zero denominators occurred in this principal design.

True D	γ	$N = 30$	$N = 50$	$N = 100$	$N = 200$
−0.8	0	0.904/0.941 [0.245/0.256]	0.924/0.948 [0.194/0.199]	0.941/0.950 [0.140/0.141]	0.946/0.950 [0.099/0.100]
−0.8	1	0.911/0.945 [0.246/0.258]	0.925/0.943 [0.195/0.200]	0.934/0.949 [0.140/0.142]	0.949/0.955 [0.100/0.100]
−0.8	2	0.906/0.939 [0.249/0.261]	0.924/0.945 [0.197/0.202]	0.940/0.951 [0.142/0.143]	0.949/0.953 [0.101/0.102]
−0.5	0	0.931/0.941 [0.346/0.345]	0.943/0.947 [0.270/0.269]	0.942/0.945 [0.192/0.192]	0.949/0.949 [0.136/0.136]
−0.5	1	0.935/0.944 [0.336/0.335]	0.940/0.947 [0.263/0.262]	0.946/0.947 [0.187/0.187]	0.949/0.952 [0.133/0.133]
−0.5	2	0.930/0.942 [0.336/0.335]	0.938/0.946 [0.263/0.263]	0.947/0.951 [0.187/0.187]	0.948/0.949 [0.133/0.133]
−0.2	0	0.937/0.944 [0.371/0.367]	0.946/0.949 [0.289/0.287]	0.948/0.949 [0.205/0.204]	0.949/0.950 [0.145/0.145]
−0.2	1	0.935/0.941 [0.325/0.322]	0.943/0.947 [0.252/0.251]	0.946/0.947 [0.180/0.179]	0.947/0.947 [0.127/0.127]
−0.2	2	0.928/0.931 [0.283/0.281]	0.942/0.944 [0.221/0.220]	0.942/0.943 [0.157/0.156]	0.945/0.945 [0.111/0.111]
0	0	0.939/0.946 [0.373/0.369]	0.942/0.945 [0.290/0.288]	0.948/0.950 [0.206/0.205]	0.948/0.949 [0.146/0.146]
0	1	0.941/0.945 [0.312/0.309]	0.947/0.950 [0.244/0.242]	0.950/0.951 [0.172/0.172]	0.950/0.950 [0.122/0.122]
0	2	0.949/0.951 [0.230/0.229]	0.948/0.949 [0.180/0.179]	0.946/0.948 [0.128/0.128]	0.944/0.945 [0.090/0.090]
0.2	0	0.940/0.945 [0.371/0.367]	0.945/0.948 [0.289/0.287]	0.949/0.951 [0.205/0.204]	0.950/0.951 [0.145/0.145]
0.2	1	0.940/0.945 [0.324/0.322]	0.941/0.946 [0.252/0.251]	0.946/0.947 [0.179/0.179]	0.948/0.949 [0.127/0.127]
0.2	2	0.932/0.936 [0.284/0.282]	0.937/0.940 [0.221/0.220]	0.943/0.943 [0.157/0.157]	0.946/0.946 [0.111/0.111]
0.5	0	0.932/0.945 [0.345/0.344]	0.936/0.945 [0.270/0.269]	0.945/0.946 [0.193/0.192]	0.945/0.947 [0.136/0.136]
0.5	1	0.929/0.941 [0.336/0.335]	0.938/0.945 [0.263/0.263]	0.943/0.947 [0.187/0.187]	0.943/0.944 [0.133/0.133]
0.5	2	0.933/0.940 [0.335/0.335]	0.938/0.944 [0.263/0.262]	0.946/0.950 [0.187/0.187]	0.951/0.953 [0.133/0.133]
0.8	0	0.907/0.942 [0.245/0.256]	0.925/0.949 [0.194/0.199]	0.937/0.949 [0.139/0.141]	0.944/0.946 [0.099/0.100]
0.8	1	0.906/0.941 [0.247/0.258]	0.924/0.946 [0.195/0.200]	0.936/0.947 [0.140/0.142]	0.945/0.952 [0.100/0.101]
0.8	2	0.905/0.940 [0.249/0.260]	0.923/0.947 [0.197/0.202]	0.933/0.947 [0.141/0.143]	0.942/0.950 [0.101/0.102]

Table A4. Sensitivity analysis for null scenarios with equal margins and varying $A + B$. Coverage and width for the naive Wald interval are conditional on $\widehat{D}$ being defined; coverage and width for the atanh-transformed interval are conditional on that interval being available. The last three columns are percentages.

β	$A+B$	γ	N	Wald coverage	atanh coverage	Wald width	atanh width	Wald CI outside	$\widehat{D}=\pm1$	$\overline{V}=0$
0	3.000	0	30	0.936	0.941	0.373	0.369	0.00	0.00	0.00
0	3.000	0	50	0.949	0.951	0.291	0.289	0.00	0.00	0.00
0	3.000	0	100	0.948	0.949	0.206	0.205	0.00	0.00	0.00
0	3.000	0	200	0.950	0.950	0.146	0.146	0.00	0.00	0.00
0	3.000	1	30	0.945	0.949	0.312	0.309	0.00	0.00	0.00
0	3.000	1	50	0.944	0.946	0.243	0.242	0.00	0.00	0.00
0	3.000	1	100	0.946	0.948	0.172	0.172	0.00	0.00	0.00
0	3.000	1	200	0.952	0.953	0.122	0.122	0.00	0.00	0.00
0	3.000	2	30	0.948	0.950	0.231	0.229	0.00	0.00	0.00
0	3.000	2	50	0.951	0.952	0.179	0.179	0.00	0.00	0.00
0	3.000	2	100	0.948	0.949	0.127	0.127	0.00	0.00	0.00
0	3.000	2	200	0.950	0.951	0.090	0.090	0.00	0.00	0.00
1	1.015	0	30	0.917	0.942	0.759	0.724	0.08	0.00	0.00
1	1.015	0	50	0.930	0.946	0.596	0.578	0.00	0.00	0.00
1	1.015	0	100	0.944	0.949	0.425	0.419	0.00	0.00	0.00
1	1.015	0	200	0.943	0.947	0.302	0.300	0.00	0.00	0.00
1	1.015	1	30	0.917	0.936	0.653	0.629	0.10	0.00	0.00
1	1.015	1	50	0.934	0.943	0.511	0.500	0.00	0.00	0.00
1	1.015	1	100	0.941	0.948	0.365	0.361	0.00	0.00	0.00
1	1.015	1	200	0.944	0.946	0.259	0.258	0.00	0.00	0.00
1	1.015	2	30	0.926	0.939	0.501	0.488	0.01	0.00	0.00
1	1.015	2	50	0.936	0.944	0.396	0.390	0.00	0.00	0.00
1	1.015	2	100	0.943	0.946	0.283	0.281	0.00	0.00	0.00
1	1.015	2	200	0.948	0.949	0.201	0.200	0.00	0.00	0.00
2	0.310	0	30	0.865	0.973	1.239	1.130	31.07	2.50	0.01
2	0.310	0	50	0.902	0.955	1.011	0.937	7.13	0.10	0.00
2	0.310	0	100	0.928	0.951	0.734	0.703	0.05	0.00	0.00
2	0.310	0	200	0.939	0.952	0.524	0.513	0.00	0.00	0.00
2	0.310	1	30	0.866	0.955	1.099	1.012	21.40	2.34	0.05
2	0.310	1	50	0.908	0.949	0.896	0.839	4.56	0.15	0.00
2	0.310	1	100	0.929	0.944	0.647	0.625	0.04	0.00	0.00
2	0.310	1	200	0.942	0.951	0.462	0.453	0.00	0.00	0.00
2	0.310	2	30	0.898	0.962	0.835	0.786	9.98	2.01	0.25
2	0.310	2	50	0.912	0.943	0.700	0.667	2.00	0.13	0.00
2	0.310	2	100	0.931	0.944	0.512	0.500	0.01	0.00	0.00
2	0.310	2	200	0.939	0.946	0.366	0.362	0.00	0.00	0.00

For $\beta = 2$ and $N = 30$, the overall atanh coverage values reported in section 4.2, counting unavailable intervals as noncoverage, are 0.948, 0.932, and 0.940 for $\gamma = 0, 1, 2$, respectively. These differ from the conditional coverages in Table A4. The sensitivity design assesses only $D = 0$.