A normed measure of variability among proportions

Mary P. COFFEY

Department of Mathematical Sciences, Oakland University, Rochester, MI 48309, USA

Marcia FEINGOLD

Division of Biostatistics and Research Epidemiology, Henry Ford Hospital, Detroit, MI 48202, USA

Judith BROMBERG

Department of Biostatistics, The University of Michigan, Ann Arbor, MI 48109, USA

Abstract: A measure of variability among a set of proportions is developed. There are no distributional assumptions, so the measure, H, is applicable in a wide variety of situations. H is scaled so that its range is zero (all proportions are equal) to one (maximum variability among the proportions, *given the weighted average*). The measure can be interpreted as a distance: the value of H indicates the position of the proportions relative to the possible extremes of no variability or maximum variability. Any set of constant weights can be applied to the proportions: the weights are used to compute the weighted average proportion and are also used to determine the extent to which each proportion affects the variability measure. Comparisons are made to other measures of variability and a numerical example is given.

Keywords: Homogeneity, Variability, Proportions, Cramer's V, Tschuprow's T, Φ^2, weights, multinomial, product binomial sampling.

1. Introduction

Consider experiments in which the responses are a set of s numbers, $p_1, \ldots, p_s$, where each p_i is between zero and one, inclusive. These p_i's could arise from any of a variety of situations, including product binomial sampling, multinomial sampling, binary repeated measures, and measures that are ratios of actual values to maximum values; the product binomial and multinomial situations are probably the most usual and useful. For simplicity, we call all of these p_i's "proportions". Our question is: what is the amount of variability among the proportions? We develop a descriptive measure of the disparity among proportions relative to the maximum possible disparity.

Our relative interest in each p_i is expressed by a set of weights, $w_1, \ldots, w_s$, where the weights are positive and sum to one. As we are not concerned with the direction of differences among the proportions, we use a weighted sum of squares, h, in the numerator of the variability measure: $h = \Sigma w_i (p_i - \bar{p})^2$, where $\bar{p} = \Sigma w_i p_i$. We then scale h by dividing it by its maximum value for the given set of weights and $\bar{p}$. Our variability measure H is defined by

$$H = \sqrt{h / \max\{ h \mid w_1, \ldots, w_s, \bar{p} \}} \, .$$

The scale factor incorporates the fact that the maximum possible variability among the proportions depends upon $\bar{p}$. For example, consider a situation in which we have three proportions ($s = 3$) and want to weight each equally (i.e., $w_1 = w_2 = w_3 = \frac{1}{3}$. Suppose that one set of proportions is $\{1, 1, 0.7\}$ with $\bar{p} = 0.9$, and a second set is $\{0.8, 0.8, 0.5\}$ with $\bar{p} = 0.7$. The differences between each p_i and the associated $\bar{p}$ are identical in the two sets. However, the p_i's in the first set are the most disparate possible given $\bar{p} = 0.9$; those in the second set are *not* the most extreme, given all possible sets of three proportions with $\bar{p} = 0.7$.

2. Examples of application

Example 1. Consider a multicenter clinical trial in which we record the proportion of patients successfully treated in each of five centers. These proportions are generated by binomial distributions. We are concerned about the variability of treatment efficacy across clinics. The extent of variability has policy implications. If there is high variability, then efforts should be directed towards determining why some clinics have substantially lower efficacy than others. Relative homogeneity indicates a more uniform success rate. Equal weights or weights proportional to the numbers of treated patients are appropriate for this situation.

Example 2. A marketing survey is done for a product, and consumers are asked to state their preferences from among a set of specified brands. The question of interest is, "Do consumers show any preferences among brands?" Low variability indicates consumer indifference among competing products. The numbers of consumers preferring each brand follow a multinomial model. Equal weights are a reasonable choice here, or, weights could be proportional to market share, so that smaller brands would not exert much influence on the measure.

Example 3. A subject is given a battery of tests all designed to measure level of anxiety. The tests are given different weights according to known reliability in the population, the time allowed for the test, or the psychologist's subjective estimate of the relative validity of each test for that subject. The measure of variability assesses the consistency of the subject's responses.

Example 4. A survey of Hispanics in the U.S. collected data on the use of Spanish and English, as well as data on social variables (age, education, length of residence in the U.S., etc.) Each respondent was classified into one of three language use categories. Several different linear discriminant models were developed to predict the language use category from the social variables. Models were evaluated and compared according to two criteria: the overall percentage of cases correctly classified, and the homogeneity of proportions of correct classifications across the three language categories. For this situation, equal weights for the language categories were used.

3. Properties of H

The minimum value of H is zero; this value occurs when all the p_i's are equal. The maximum value of H is one and occurs when the p_i's are as varied as possible, given $\bar{p}$. H can achieve its maximum and minimum values for all combinations of the number of categories, sample size, weights, and value of $\bar{p}$.

There is a clear interpretation of H as the fraction of the relative distance from perfect homogeneity to the largest possible scatter of values given the average. When all the deviations $p_i - \bar{p}$ are multiplied by the same constant c ($0 < c \leqslant 1/H$), the new value of H is c times its former value. For example, consider the set of proportions $\{1, 1, 0.7\}$, where each proportion receives equal weight. Then $\bar{p} = 0.9$ and $H = 1$, as the proportions are as widespread as possible given the value of $\bar{p}$. However, if the proportions were $\{0.98, 0.98, 0.74\}$ (i.e., each proportion is only 80% of its former distance from $\bar{p}$), then $H = 0.8$. In this sense, H can be interpreted as a relative distance between the situations of equal proportions and maximum variability among the proportions conditional on $\bar{p}$.

The properties of H given in the previous two paragraphs are independent of the distribution from which the proportions were drawn. Although we usually tend to think of proportions as arising from discrete data, H can also be used for any continuous distribution bounded by 0 and 1.

H possesses desirable symmetry properties. If the proportions represent distances from one of the bounds of the underlying distribution (1 or 0), the value of H would not change if the measurements were changed to represent distances from the opposite boundary. For product binomial sampling, the value of H is the same regardless of which of the two outcomes is labeled a success. For multinomial sampling, H is invariant with respect to category ordering.

Asymptotic standard errors for H depend on the sampling scheme giving rise to the proportions, the types of weights that are used, and whether or not all of the population probabilities are equal. Using properties of Pearson's X^2, we have obtained asymptotic standard errors for multinomial sampling with equal weights, and for product binomial sampling with weights proportional to sample sizes. Expressions for asymptotic standard errors for arbitrary weights for these two sampling schemes are not readily obtainable.

Table 1

Asymptotic expected value and standard error of H under multinomial sampling, for equal weights and equal population probabilities

s	$\sqrt{n}\,E_0(H)$	$\sqrt{n}\,SE_0(H)$
2	0.798	0.603
3	0.886	0.463
4	0.921	0.389
5	0.940	0.341
6	0.952	0.308
7	0.959	0.282
8	0.965	0.262
9	0.969	0.246
10	0.973	0.232

For the following discussion, let Π_i represent the underlying population probability for the ith proportion. X^2 denotes Pearson's statistic for testing the hypothesis of equal probabilities. A zero subscript on variances and standard errors refers to the case of equal Π_i's, while a subscript of 1 refers to the case in which not all the population probabilities are equal.

In the multinomial case, the s proportions necessarily sum to one. With equal weights, $H = \sqrt{X^2/[(s-1)n]}$, where n is the total sample size. Using properties of the chi distribution, we have

$$\mathrm{Var}_0(H) \sim \frac{1}{n} - \frac{2}{(s-1)n}\,\frac{\Gamma^2[s/2]}{\Gamma^2[(s-1)/2]}$$

when all s population probabilities are equal. Table 1 shows the asymptotic expectations and standard errors for this case. The asymptotic variance of H when the Π_i's differ (but still assuming equal weights) can be obtained by combining equation (30.49) in Kendall and Stuart [3] for $\mathrm{Var}(X^2)$ with Taylor series methods:

$$\mathrm{Var}_1(H) \sim \frac{(n-1)s^2\left[\Sigma\Pi_i^3 - (\Sigma\Pi_i^2)^2\right]}{(s-1)nX_\Pi^2},$$

where X_Π^2 is the value of X^2 calculated using the population probabilities. When we replace the unknown population parameters by their sample estimates, we have

$$\mathrm{Var}_1(H) \sim s^2\frac{(n-1)\left[\Sigma p_i^3 - (\Sigma p_i^2)^2\right]}{(s-1)nX^2}.$$

We next consider product binomial sampling. Inference about H with weights proportional to sample sizes is fairly straightforward. Pearson's X^2 provides a test of the hypothesis that the population value of H is zero. For unequal population probabilities, we examine the numerator and denominator of H

separately. The numerator of H is $\sqrt{(\bar{p}\bar{q}/N)X^2}$, where N is the total sample size (in the s by 2 contingency table). Combining Pearson's approximation for the estimated variance of X^2 in the case of unequal Π_i's with fixed marginal frequencies (Kendall and Stuart [3, eqn. 33.70]) with a Taylor series argument, the asymptotic variance for the numerator of H is

$$\text{Var}_1(\sqrt{h}) \sim \frac{\bar{p}\bar{q}}{X^2}\left(\frac{1}{N}\sum_{i,j}\frac{\left(n_{ij}-\dfrac{n_{i.}n_{.j}}{N}\right)^2}{\left(\dfrac{n_{i.}n_{.j}}{N}\right)^2}+\frac{X^2}{N}\left(1-\frac{X^2}{N}\right)\right),$$

where $n_{.j}$ is the marginal total corresponding to the jth column ($j=1,2$), $n_{i.}$ is the row total ($i=1,\dots,s$), and n_{ij} is the count in the ij^{th} cell. Given $\bar{p}$ and the weights, the denominator of H is a constant, so

$$\text{Var}_1(H) \sim \text{Var}_1(\sqrt{h})/\max(h\mid w_1,\dots,w_s,\bar{p}).$$

4. The scale factor

The maximum value of h for a given $\bar{p}$ and set of weights occurs when at most one of the proportions is not equal to one or zero. The maximum value of h is always $\leqslant \bar{p}(1-\bar{p})$; specifically, max $h=\bar{p}(1-\bar{p})-w_r r(1-r)$, where w_r is one of the set of s weights and r is a function of $\bar{p}$ and the set of weights (see Appendix 1). The maximum value of h cannot be computed explicitly except in special circumstances. For the case of equal weights, the maximum value of h is $\bar{p}(1-\bar{p})-r(1-r)/s$, where $r=s\bar{p}-[s\bar{p}]$. For the multinomial with equal weights, $\bar{p}=1/s$ and the maximum value of h is $(s-1)/s^2$. A Fortran program to compute H and the maximum h, for any $\bar{p}$ and any vector of weights w, is given in Appendix 2.

Table 2
Maximum value of h when the weights are equal

s	$\bar{p}$				
	0.1, 0.9	0.2, 0.8	0.3, 0.7	0.4, 0.6	0.5
2	0.010	0.040	0.090	0.160	0.250
3	0.020	0.080	0.180	0.187	0.167
4	0.030	0.120	0.170	0.180	0.250
5	0.040	0.160	0.160	0.240	0.200
6	0.050	0.133	0.183	0.200	0.250
7	0.060	0.126	0.197	0.217	0.214
8	0.070	0.130	0.180	0.220	0.250
9	0.080	0.142	0.187	0.213	0.222
10	0.090	0.160	0.210	0.240	0.250

Table 3
Maximum value of h when weights are equally spaced

s	$\bar{p}$				
	0.1, 0.9	0.2, 0.8	0.3, 0.7	0.4, 0.6	0.5
2	0.020	0.080	0.180	0.180	0.125
3	0.050	0.130	0.180	0.200	0.250
4	0.090	0.160	0.210	0.240	0.250
5	0.065	0.160	0.193	0.240	0.233
6	0.086	0.152	0.200	0.229	0.238
7	0.084	0.151	0.201	0.234	0.250
8	0.083	0.156	0.206	0.233	0.250
9	0.084	0.160	0.204	0.240	0.244
10	0.085	0.160	0.205	0.240	0.245

Tables 2 through 4 show how the maximum value of h varies for $\bar{p} = 0.1(0.1)0.9$, and some particular values of w. In Table 2 the weights are equal. (For the multinomial model with equal weights, $\bar{p} = 1/s$ and the maximum value of h is $(s - 1)/s^2$.) In Table 3 the weights are equally spaced, $w_i = 2i/[s(s + 1)]$ ($i = 1, \ldots, s$), illustrating situations in which there is a variety of different weights. In Table 4, all weights are equal except for one that is twice as large as the others, as might be the case if one measurement were particularly important. These tables are illustrative only, and should not be used for extrapolating to other weights or values of $\bar{p}$: note the lack of monotonicity both with s and with $\bar{p}$.

The following procedure can be used to obtain bounds on the maximum value of h. The upper bound for any $\bar{p}$, any w, and any number of categories, s, is $\bar{p}(1 - \bar{p})$; the lower bound is $\bar{p}(1 - \bar{p}) - \max w_i/4$. For weights corresponding to those in both Tables 3 and 4, the lower bound is $\bar{p}(1 - \bar{p}) - 1/(2s + 2)$. Note that this lower bound for h can be considerably smaller than $\bar{p}(1 - \bar{p})$, especially for values of $\bar{p}$ close to 0 or 1, but approaches $\bar{p}(1 - \bar{p})$ as s increases.

Table 4
Maximum value of h when one weight is twice as large as the others

s	$\bar{p}$				
	0.1, 0.9	0.2, 0.8	0.3, 0.7	0.4, 0.6	0.5
2	0.020	0.080	0.180	0.180	0.125
3	0.030	0.120	0.170	0.180	0.250
4	0.040	0.160	0.160	0.240	0.200
5	0.050	0.133	0.183	0.200	0.250
6	0.060	0.126	0.197	0.217	0.214
7	0.070	0.130	0.180	0.220	0.250
8	0.080	0.142	0.187	0.213	0.222
9	0.090	0.160	0.210	0.240	0.250
10	0.082	0.145	0.191	0.218	0.227

5. Related measures

The variability measure can be applied in different situations: s by 2 contingency tables, multinomial classifications, binary analogs of the randomized complete block design, and repeated testing situations in which the proportions are the fractions of achieved scores to maximum scores. In the literature, there are global measures of variability for a set of proportions, as well as measures and tests that are appropriate for specific sampling situations.

The coefficient of variation (CV) is frequently used in social science literature as a measure of variability relative to the mean. Both the CV and the S measure, a normed version of the CV developed by Martin and Gray [5], can be calculated in any situation with s numbers between 0 and 1, inclusive. When applied to proportions,

$$CV = \sqrt{\left(\sum (p_i - \bar{p})^2 / (s - 1)\right)} / \bar{p}$$

and can exceed one. Martin and Gray derived

$$S = CV / \sqrt{s}$$

as a normed coefficient of variation–that is bounded by one. However, S achieves its upper bound of 1 only when exactly one proportion is 1 and all the others are zero.

For s by 2 contingency tables, classical goodness of fit statistics can be used to test the hypotheses of equal success probabilities across s levels of the row variable. Association measures based on the Pearson chi-square statistic and the proportional-reduction-in-error measures can be used to describe the variability of s proportions. These measures can be considered as competitors of H only when the weights in H are proportional to the row margins. In s by 2 tables, the mean square contingency coefficient Φ^2 and Goodman and Kruskal's tau for predicting the column category are identical: Cramer's $V = \Phi$ (and $\sqrt{\tau}$). While the largest value of these measures is 1, this value is impossible to achieve for many sets of marginal totals. However, H can obtain its upper bound of 1 regardless of the marginal frequencies in a particular table.

Let $n_{i.}$ denote the row total in the ith row, $\bar{p}$ the overall proportion in the first column and $N = \Sigma n_{i.}$. Then Φ can be written as

$$\Phi = \sqrt{\left(\sum (n_{i.}/N)(p_i - \bar{p})^2\right) / \bar{p}(1 - \bar{p})}$$

while H with weights proportional to row margins is

$$H = \sqrt{\left[\sum (n_{i.}/N)(p_i - \bar{p})^2\right] / (\bar{p}(1 - \bar{p}) - w_r r(1 - r))}.$$

Therefore, H is greater than or equal to Φ, with equality when $\max[h \mid w_1, \ldots, w_s, \bar{p}] = \bar{p}(1 - \bar{p})$. That is, Φ can reach its upper limit of one only when the sum of a subset of the weights equals $\bar{p}$ (for equal weights, this means that $s\bar{p}$ is an integer).

Consider, for example, the set of proportions $\{1, 1, 1, 0.5, 0\}$ based on equal sample sizes. The average proportion is 0.7. For that value of $\bar{p}$, the proportions are as disparate as possible, yielding an H value of 1.0, while Cramer's $V = 0.87$, Pearson's P (the coefficient of contingency) = 0.66, and Tschuprow's $T = 0.62$. If the proportions were $\{0.85, 0.85, 0.85, 0.60, 0.35\}$, which is exactly halfway between the most disparate position and complete uniformity (given $\bar{p} = 0.7$), $H = 0.50$, Cramer's $V = 0.44$, Pearson's $P = 0.40$, and Tschuprow's $T = 0.31$. A particular disadvantage of the S measure for s by 2 tables is its sensitivity to the assignment of zeros and ones to success and failure (Kish [4]). For the example given above where $\bar{p} = 0.7$, $S = 0.29$. If the set of proportions were $\{0, 0, 0, 0.5, 1\}$ (the complement of the first example), then $\bar{p} = 0.3$ and the values of H, V, P, and T are unchanged. However, S increases from 0.29 to 0.67.

In the multinomial situation, classical goodness of fit statistics can be used to test the hypothesis of equal probabilities. If the s category weights are equal, there are two measures equal to H. In this situation, $S = H$. The measure of variation (MV) (Nachmias and Rosenblum [6]), and an identical measure named the Index of Qualitative Variation (Ott, Larson, and Mendenhall [7]) measure the variation in proportions for s different categories: $MV = 1 - H^2$. H, $1 - MV$, and all of the other measures given above are zero when all of the proportions are equal.

Cochran's Q (Cochran [2]) tests the hypothesis of equal probabilities for the randomized block design. It uses more information than is contained in the treatment success proportions, and so cannot be compared directly to H.

6. Numerical example

The faculty of the University of Michigan was surveyed recently to ascertain their perceptions of merit criteria and review procedures (University of Michigan Committee on the Economic Status of the Faculty [8]). Questionnaires were sent to all regular faculty members during the 1985–1986 academic year.

To assess the performance of the merit system, the Board of Regents was interested not only in the overall responses to questions, but also in the variation in responses for different faculty subgroups. Results were reported separately by rank, by ethnic background, and by school. The question of interest for each such breakdown was: how much variation is there? While the proportions in each response category should be given, summary descriptive measures of variation are helpful, as a great deal of data was contained in the report. Descriptive measures are more appropriate than inference procedures in this case, as the entire faculty was surveyed.

We will focus on the responses to one particular question – "How much information have you been given about the criteria employed in the review process in your department/unit?" The proportions were reported for the combined category of "none or little". The summary of the report stated that "There are substantial differences...among different groups of faculty." We use H with

Table 5
Proportion of faculty reporting no or little information about review process criteria

By Rank $H = 0.16$

Rank	Number	Percent
Professor	782	34.8
Associate	338	46.4
Assistant	382	50.8

By Ethnic Background $H = 0.25$

Ethnic Group	Number	Percent
Asian	53	62.2
Black	47	55.3
Hispanic	10	40.0
Native American	21	66.7
White	1355	39.7

By School $H = 0.50$

School	Number	Percent
Architecture	26	52.9
Art	21	33.3
Business Admin	59	40.6
Dentistry	95	72.6
Education	37	44.7
Law	16	88.9
Engineering	158	28.5
LSA	417	31.9
Medicine	373	60.8
Music	56	31.6
Natural Resources	21	9.5
Nursing	48	31.3
Pharmacy	19	5.3
Social Work	29	3.3
Public Health	73	26.0

equal weights to describe the variation in proportions across any one mode of classification. With the use of equal weights, each category contributes equally to H. The use of weights proportional to sample size would yield a measure that is dominated by categories with a large number of faculty; in particular, the results for whites and the school of Literature, Science, and the Arts would heavily influence the measures computed across ethnic groups and schools, respectively.

Table 5 presents the responses to the question on information about review criteria for the faculty members broken down by rank, ethnic group, and school. The values of H for these three classification modes are 0.16, 0.25, and 0.50, respectively. In other words, the proportion of faculty with little or no information about review criteria varies little with rank or ethnic background, but varies more across schools. The school proportions are midway between the situations

of equality and the maximum possible variability given the average proportion. The variability of the proportions for various ethnic backgrounds is $\frac{1}{4}$ the maximum possible, given $\bar{p}$.

Unlike other measures, H can achieve its upper bound of 1 regardless of the table margins; as a result, it is more easily interpreted than measures based on X^2. This feature is also useful to compare relative variability across different tables. On the other hand, the magnitude of X^2 and the size of the p-value measure only statistical significance, which may be of no interest, as in this situation where the amount of dispersion is the primary concern.

Acknowledgments

References

[1] Y.M.M. Bishop, S.E. Fienberg and P.W. Holland, *Discrete Multivariate Analysis: Theory and Practice.* (MIT Press, Cambridge, Mass., 1975).
[2] W. Cochran, The comparison of percentages in matched samples, *Biometrika* **37** (1950) 256–266.
[3] M.G. Kendall and A. Stuart. The Advanced Theory of Statistics, Fourth Edition. (Macmillan Publishing CO., Inc., New York, 1979).
[4] L. Kish (1965). *Survey Sampling.* New York: Wiley.
[5] J.D. Martin and L.N. Gray, Measurement of relative variation: Sociological examples, *American Sociological Review* **36** (1971) 496–502.
[6] D. Nachmias and D.H. Rosenbloom, Measuring bureaucratic representation and integration, *Public Administration Review* **6** (1973) 590–597.
[7] L. Ott, R.F. Larson and W. Mendenhall, *Statistics: A Tool for the Social Sciences, 3rd ed* (Duxbury press, Boston, 1983).
[8] University of Michigan, The Committee on the Economic Status of the Faculty, *Report on the 1985/86 Faculty Salary and Compensation Survey. Part Two: Merit Criteria and Review Procedures*, Mimeographed (1987).

Appendix 1. Derivation of the maximum value of *h*

Theorem. *Consider a set of values $p' = (p_1, \ldots, p_s)$, where all of the p_i's are between zero and one, inclusive. Associated with these proportions are a set of weights $w' = (w_1, \ldots, w_s)$, where the weights are positive and sum to one. Define the weighted average proportion as $\bar{p} = \sum w_i p_i$. Then the maximum value of $h = \sum w_i (p_i - \bar{p})^2$, for a given value of $\bar{p}$, occurs when at most one p_i is not equal to zero or one.*

The maximum value of h for a general set of weights is $\bar{p}(1 - \bar{p}) - w_r r(1 - r)$ for an appropriate choice of w_r and r. For the case of equal weights, the maximum value of h is $\bar{p}(1 - \bar{p}) - r(1 - r)/s$, where $r = s\bar{p} - [s\bar{p}]$.

Part 1: Consider any set of proportions p_1 with more than one element different from zero or one. We can always find another set of proportions p_2 with the same value of $\bar{p}$ such that $h(p_2) > h(p_1)$.

Assume, WOLOG, that the first two elements of p_1 are r_1 and r_2, where $0 < r_i < 1$. We want to choose λ_1 and λ_2 so that $p_2' = (\lambda_1, \lambda_2, 0, \dots, 0) + p_1'$, subject to the constraints

$$w_1\lambda_1 + w_2\lambda_2 = 0, \quad \text{as } \bar{p} \text{ is fixed}, \tag{1}$$

$$0 \leqslant r_i + \lambda_1 \leqslant 1, \quad \text{for } i = 1, 2, \tag{2}$$

and

$$h(p_2) - h(p_1) > 0. \tag{3}$$

as $h(p) = \Sigma w_i p_i^2 - \bar{p}^2$, we have, from (3),

$$2(w_1 r_1 \lambda_1 + w_2 r_2 \lambda_2) + w_1 \lambda_1^2 + w_2 \lambda_2^2 > 0. \tag{4}$$

We solve (1) for λ_2 and substitute this into (4) to get

$$2\lambda_1(r_1 - r_2) + \lambda_1^2(1 + w_1/w_2) > 0. \tag{5}$$

Assume, WOLOG, that $w_1 \geqslant w_2$. If $r_1 \geqslant r_2$, let $\lambda_1 = (w_2/w_1)\min(1 - r_1, r_2)$ and, due to (1), we have $\lambda_2 = -\min(1 - r_1, r_2)$. This satisfies (2), and λ_1 is positive, so (5) holds. If $r_1 < r_2$, let $\lambda_1 = -(w_2/w_1)\min(1 - r_2, r_1)$ and $\lambda_2 = \min(1 - r_2, r_1)$. Again, (2) and (5) hold.

Part 2: If p_0 is any point in $\mathbb{R}^s$ with at most one component different from 0 or 1, then h has a local maximum at p_0.

WOLOG for this part of the proof, let $p_0' = (1, \dots, 1, 0, \dots, 0, r) = (1_n' \mid 0_m' \mid r)$, where $0 \leqslant r < 1$. To show that the directional derivative of h at p_0, with respect to any vector δ, is negative, let δ be partitioned similarly: $\delta' = (\delta_n' \mid \delta_m' \mid \delta_r)$. Let w be the set of weights: $w' = (w_n' \mid w_m' \mid w_r)$. The directional derivative of h with respect to δ evaluated at p_0 is

$$h(p_0') \cdot \delta = w_n' \cdot \delta_n(1 - \bar{p}) + w_m' \cdot \delta_m(0 - \bar{p}) + w_r \delta_r(r - \bar{p}). \tag{6}$$

As we are holding $\bar{p}$ constant, $w' \cdot \delta = 0$. Therefore, (6) reduces to

$$h(p_0') \cdot \delta = w_n' \cdot \delta_n + w_r \delta_r r. \tag{7}$$

Note that all of the elements of δ_n must be non-positive, all of the elements of δ_m non-negative, and $-r \leqslant \delta_r \leqslant 1 - r$. Also, if δ_r is positive, at least one element of δ_n must be negative; if δ_r is negative, at least one element of δ_m must be positive. If δ_r is zero, then both δ_n and δ_m have non-zero elements. Therefore, if δ_r is non-positive, (7) must be negative. If δ_r is positive, then, as $w' \cdot \delta = 0$, we have

$$w_n' \cdot \delta_n = -w_m' \cdot \delta_m - w_r \delta_r \leqslant -w_r \delta_r < -w_r \delta_r r.$$

Applying this result to (7), we see that the directional derivative is always negative.

Part 3: The maximum value of h is $\bar{p}(1 - \bar{p}) - w_r r(1 - r)$, where w_r and r are defined in the following paragraph.

If all of the weights are equal, let $n = [s\bar{p}]$ and $r = s\bar{p} - [s\bar{p}]$. Then

$$h = (1/s)(n + r^2) - \bar{p}^2 = \bar{p}(1 - \bar{p}) - r(1 - r)/s.$$

If the weights are not equal, then several possibilities must be explored to find which vectors of the type p_0 (vectors with at most one value not equal to 0 or 1) maximize h. If there is any subset of weights such that their sum, W, equals $\bar{p}$, then let p_0 have ones in the positions corresponding to the weights in W and zeros elsewhere. In this case, $h = \Sigma w_i p_i^2 - \bar{p}^2 = \bar{p}(1 - \bar{p})$. Otherwise, let W be the sum of any subset of the weights such that $\bar{p} - W$ is less than at least one of the remaining weights not included in W. Let w_r be the smallest of those remaining weights, and define $r = (\bar{p} - W)/w_r$. Then h is maximized for that particular subset of weights included in W by letting p_0 have ones corresponding to the weights in W, r in the position corresponding to w_r, and zeros elsewhere. In this case

$$h = W + w_r r^2 - \bar{p}^2 = (W + w_r r) - w_r r + w_r r^2 - \bar{p}^2$$

$$= \bar{p}(1 - \bar{p}) - w_r r(1 - r).$$

Therefore, W must be chosen in such a way as to minimize $w_r r(1 - r)$. □

Appendix 2. A Fortran program to calculate *H*

```
      IMPLICIT REAL*8 (A – H,O – Z)
      INTEGER S
      COMMON EQUALW,PBAR,W(20),P(20),S
C
C     REQUEST AND RECEIVE INPUT
C     S: NUMBER OF PROPORTIONS, WHERE S IS GE 2 AND S LE 20.
C     P: VECTOR OF PROPORTIONS, WHERE P(I) IS GE 0 AND P(I) LE
C        1.
C     W: VECTOR OF WEIGHTS, WHERE THE SUM OF THE W(I)
C        EQUALS 1.
C
   1000 CALL INPUTT
C
C   COMPUTE H
C   XNUM: A SUBROUTINE THAT COMPUTES h, THE NUMERATOR
C         OF H
C     XMAX: A SUBROUTINE THAT COMPUTES MAX(h), THE DE-
C           NOMINATOR OF H,
C     GIVEN THE WEIGHTS AND PBAR
C
      CALL XNUM(HNUM)
      CALL XMAX(HMAX)
```

```
      H = DSQRT(HNUM/HMAX)
      WRITE (6,6000) H,HMAX
      GO TO 1000
C
 6000 FORMAT ('0H = ', F6.3, ' MAX(h) = ', F6.3)
      END


      SUBROUTINE INPUTT
      IMPLICIT REAL*8(A − H,O − Z)
      INTEGER S
      COMMON EQUALW,PBAR,W(20),P(20),S
C
 1000 WRITE (6,6100)
      READ (5,5100,END = 9900) S, (P(I),I = 1,S)
      IF (S .LT. 1) STOP
      IF (S .LT. 2 .OR. S .GT. 20) GO TO 9100
C
      PBAR = 0D0
      PMAX = 1D0
      PMIN = 0D0
      DO 1100 I = 1,S
         W(I) = 0D0
         PMAX = DMAX1(P(I),1D0)
 1100 PMIN = DMIN1(P(I),0D0)
C
      WRITE (6,6200)
      READ (5,5200,END = 1200)(W(I),I = 1,S)
      WSUM = 0D0
      EQUALW = 0D0
 1200 DO 1300 I = 1,S
         WSUM = WSUM + W(I)
 1300 EQUALW = DMAX1(W(I),0D0)
      IF (EQUALW .GT. 0D0) GO TO 1500
C
C EQUAL WEIGHTS ARE THE DEFAULT
C
      DO 1400 I = 1,S
 1400    W(I) = 1D0/S
      WSUM = 1D0
C
 1500 DO 1600 I = 1,S
 1600    PBAR = PBAR + W(I)*P(I)
      WRITE (6,6300) S, PBAR, (P(I),I = 1,S)
      WRITE (6,6400) (W(J),J = 1,S)
      IF (PMIN .LT. 0D0 .OR. PMAX .GT. 1D0) GO TO 9200
```

```
      IF (DABS(WSUM − 1D0) .GT. 1D − 5) GO TO 9300
      RETURN
 9900 STOP
C
C   ERROR PROCESSING
C
 9100 WRITE (6,6910) S
      GO TO 1000
 9200 WRITE (6,6920) (P(I),I = 1,S)
      GO TO 1000
 9300 WRITE (6,6930) WSUM
      GO TO 1000
C
 5100 FORMAT (I5, 20F10.5)
 5200 FORMAT (20F10.5)
 6100 FORMAT ('0Enter S and P(1),...,P(S) separated by commas',
     + ' or ⟨CR⟩ to stop.')
 6200 FORMAT ('If the weights are unequal, enter them.',
     + ' Else press ⟨CR⟩.')
 6300 FORMAT ('0S = ', I2, ' Mean P = ', F5.3/
     + '0Proportions:', 10F7.3/13X,10F7.3)
 6400 FORMAT ('0Weights: ',10F7.3/13X,10F7.3)
 6910 FORMAT ('0***S = ', I2, ' is out of range.',
     + ' S must be between 2 and 20.')
 6920 FORMAT ('0***Each proportion must be between zero and one,',
     + ' inclusive. At least one is outside that of that interval.')
 6930 FORMAT ('0***The weights sum to ', F6.3, '. They must sum to 1.')
      END


      SUBROUTINE XNUM(HNUM)
      IMPLICIT REAL*8(A − H,O − Z)
      INTEGER S
      COMMON EQUALW,PBAR,W(20),P(20),S
C
      HNUM = −(PBAR**2)
      DO 100 I = 1,S
 100  HNUM = HNUM + W(I)*P(I)**2
      RETURN
      END


      SUBROUTINE XMAX(HMAX)
      IMPLICIT REAL*8 (A − H,O − Z)
      REAL*8 Q(20)
      INTEGER S, SLESS1, IX(20)
```

```
      COMMON EQUALW,PBAR,W(20),P(20),S
C
      IF (EQUALW .GT. 0D0) GO TO 100
C
C  EQUAL WEIGHTS
C
      N = S*PBAR
      R = S*PBAR - N
      HMAX = PBAR*(1D0 - PBAR) - R*(1D0 - R)/S
      RETURN
C
C  UNEQUAL WEIGHTS
C
100   SLESS1 = S - 1
         NPERMU = 2**SLESS1
         HMAX = 0D0
         PBARSQ = PBAR**2
C
C         COMPUTE ALL PERMUTATIONS OF S - 1 ZEROS AND ONES
C
         DO 500 I1 = 1,S
            IS = 0
            DO 200 I2 = 1,S
               IF (I2 .EQ. I1) GO TO 200
               IS = IS + 1
                  IX(IS) = I2
200         CONTINUE
C
C      FOR EACH PERMUTATION CALCULATE h AND FIND MAX(h)
C
            DO 400 I4 = 1, NPERMU
               H = - PBARSQ
               REMAIN = PBAR
               NUMBER = I4 - 1
               DO 300 I3 = 1, SLESS1
                  NDX = IX(I3)
                  Q(NDX) = IABS(MOD(NUMBER,2))
                  WQ = W(NDX)*Q(NDX)
                  IF (WQ .GT. REMAIN) GO TO 400
                  H = H + WQ*Q(NDX)
                  REMAIN = REMAIN - WQ
300            NUMBER = NUMBER/2
                  Q(I1) = REMAIN/W(I1)
               IF (Q(I1) .LT. 0D0 .OR. Q(I1) .GT. 1D0) GO TO 400
400         HMAX = DMAX1(HMAX,H)
500      CONTINUE
            RETURN
            END
```