University of Toronto
Department of Economics

Working Paper 726

Modally Focused, Likelihood Based, Inequality Measurement in Multivariate Ordered Categorical Paradigms: A Note.

By Gordon Anderson

June 27, 2022

Modally Focused, Likelihood Based, Inequality Measurement in Multivariate Ordered Categorical Paradigms: A Note.

Gordon Anderson

University of Toronto.

June 2022

Abstract.

The increasing use of multivariate ordered categorical data in the social sciences presents a challenge for those concerned with measuring inequality. The absence of cardinal measure and the ambiguities inherently associated with artificial attribution of scale to ordinal categories, precludes the use of standard distance-based inequality measures. However, these issues have been surmounted in the univariate world by employing notions of probabilistic distance (the increasing likelihood that some outcome between two categories will occur the bigger is the categorical gap between them) and measuring aggregate probabilistic distance from a Median category focus point. Unfortunately, in multivariate environments the median outcome is not uniquely defined, however the modal category is, thus providing a readymade reference point from which to measure probabilistic distance. In addition, as the most frequently observed outcome, its density value provides a natural measure of the extent of commonality or equality in the population, further rationalizing its use as a point of reference for inequality measurement. This note develops modally focused inequality measures for multivariate ordered categorical environments together with their asymptotic distributions for inference purposes and discusses their axiomatic foundations.

Introduction.

Following the foundational work of Mendelson (1987), as a means of ordering outcome inequalities in ordinal variable environments, the Median Preserving Spread (MPS) has become a popular workhorse in inequality and polarization measurement of ordered categorical outcomes (Allison and Foster 2004, Kobus 2015). It overcomes the scaling problems associated with measuring distance from a focus point in ordered categorical environments when categories have been artificially attributed cardinal value (Bond and Lang 2019) by quantifying distance in cumulated probabilistic terms. Furthermore, it obviates the need for dimension weighting in cardinally measured multidimensional frameworks. Unfortunately, applications are typically confined to the univariate paradigms since, in multivariate contexts, a unique median or quantile category from which to measure cumulated probabilistic distance is hard to conceptualize[1]. A straightforward alternative is to use the modal category as a focus which can, under appropriate circumstances, be readily conceptualized as a unique category or point in any multivariate measurement paradigm, with ordered categorical variates being a particular case of interest. Furthermore, since it is the most common outcome, the density value at the mode provides a very natural likelihood-oriented measure of the extent of commonality or equality of outcome inherent in a distribution and its complement readily seen as an elementary measure of the extent of inequality or lack of commonality of outcome.

Characterization of Modally Focused increased spread

The univariate paradigm.

Following Mendelson (1987), consider probability density functions f and g defined over K ordered categorical outcomes indexed accordingly as $k = 1,..,K$ where $k = 1$ corresponds to the lowest category and $k = K$ the highest and where $f \ \ (g)$ is represented by a K dimensioned vector $\underline{p}_f \ \ \left(\underline{p}_g\right)$ with typical element $p_{fk} \ \ \left(p_{gk}\right)$ where:

$p_{fk} \ \ \left(p_{gk}\right) = P(Outcome\ k|f) \ \ (P(Outcome\ k|g))$ with $\sum_{k=1}^{K} p_{fk} = 1$ and $\sum_{k=1}^{K} p_{gk} = 1$.

For a given outcome $k^* \in 1, K$ and outcomes $k = k^* + 1,..,K$, define the Upper Cumulants of f as $F_k^{U,k^*} = \sum_{i=k^*+1}^{k} p_{fi}$ (note for $k \leq k^*$, $F_k^{U,k^*} = 0$) and for outcomes $k = 1,..,k^* - 1$, define its Lower Cumulants as $F_k^{L,k^*} = \sum_{i=k}^{k^*-1} p_{fi}$ (note for $k \geq k^*$, $F_k^{L,k^*} = 0$). When $k > k^*$, F_k^{U,k^*} is the probability of an outcome between k^* and $k + 1$ occurring which is monotonically non decreasing in k, when $k < k^*$, F_k^{L,k^*} is the probability of an outcome between k^* and $k - 1$ occurring which is monotonically non-decreasing in $k^* - k$. Each record a sense of probabilistic distance of k from k^* in terms of the chance that an outcome will emerge between k and k^* which increases with $|k^* - k|$. Similarly defining G_k^{U,k^*},

---

[1] For example, consider a discrete uniform ordered categorical bivariate distribution $f\left(x_i, y_j\right)$ with K categories in each dimension, each ordered by their respective subscripts so that $i' < i'' \leftrightarrow x_{i'} <_P x_{i''}$ and similarly, $j' < j'' \leftrightarrow y_{j'} <_P y_{j''}$ and $f\left(x_i, y_j\right) = 1/K^2 \ \forall\, i, j = 1,..,K$ so that $F\left(x_i, y_j\right) = (i * j)/K^2$. There is an array of pairs $\left(x_i^*, y_j^*\right)$ for which $F\left(x_{i-1}^*, y_j^*\right) < 0.5\ \&\ 0.5 < F\left(x_i^*, y_j^*\right)$ or $F\left(x_i^*, y_{j-1}^*\right) < 0.5\ \&\ 0.5 < F\left(x_i^*, y_j^*\right)$

G_k^{L,k^*}, the Upper and Lower Cumulants of g, then g constitutes an increasing spread of f with respect to outcome k^* when:

$$G_k^{L,k^*} \geq F_k^{L,k^*} \forall\, k = 1,..,k^* - 1 \text{ and } G_k^{U,k^*} \geq F_k^{U,k^*} \forall\, k = k^* + 1,..,K \text{ with } > \text{somewhere}. \qquad [1]$$

The Mendelson (1987) condition [1] amounts to a first order stochastic dominance condition on the "downward looking" below k^* conditional distributions (i.e. imagine the category orderings below k^* were reversed) and the "upward looking" above k^* conditional distributions where f dominates g in each context. Intuitively, with respect to k^* inequality in g distribution is greater than inequality in f distribution with respect to k^* when the chance of a below k^* outcome and the chance of an above k^* outcome are both at least as great in g as they are in f with strictly greater than in at least one case[2].

Given the absence of cardinal measure, setting k^* as the "median" category and using these notions has been the basis of inequality and polarization measurement in univariate ordered categorical paradigms (Allison and Foster 2004, Kobus 2015). However, the median outcome could well be an unlikely event and, if inequality is construed as the antithesis of complete commonality in the population, it would not serve as a good focus point for an inequality measure.

The Modal Preserving Spread.

Define the Modal outcome of distribution f as outcome k^* *such that* $p_{fk^*} = \max_k p_{fk}$. Determining k^* by seeking that category for which $\hat{p}_{fk^*} = \max_k \hat{p}_{fk}$ where $\hat{p}_{fk}, k = 1,..,K$ are the maximum likelihood estimates of category densities, renders k^* as the maximum likelihood estimate of the category most likely to command unanimity of membership. Since the smallest possible value of p_{fk^*} is $\frac{1}{K} + \varepsilon$ where ε is an arbitrarily small positive value, $\frac{1}{K} < p_{fk^*} \leq 1$, and when p_{fk^*} is viewed as the chance that the whole population resides in outcome k^*, $LC(f) = (Kp_{fk^*} - 1)/(K-1)$ is a very natural likelihood based measure or index on the unit interval of the extent of commonality or equality of outcome in the distribution, so that its complement, $II(f) = 1 - LC(f) = K(1 - p_{fk^*})/(K-1)$ provides an intuitive likelihood based measure of inequality of outcome[3]. Unfortunately, it is not responsive to variation in spread in the rest of the distribution in the sense that a marginal shift in mass from k' to k'' where k', $k'' \neq k^*$ would leave it unaltered unless the shift rendered k'' the new modal outcome. To capture this, the concept of a modal preserving spread needs to be considered. Basically g constitutes a Modal Preserving Spread of f if [1] holds and k^*remains the modal outcome of g i.e. $p_{gk^*} = \max_k p_{gk}$.

This can be readily checked by considering $UAMBI(f,g) = \frac{\sum_{k=1}^{K}((G_k^{U,k^*} - F_k^{U,k^*}) + (G_k^{L,k^*} - F_k^{L,k^*}))}{\sum_{k=1}^{K}\left(\left|(G_k^{U,k^*} - F_k^{U,k^*})\right| + \left|(G_k^{L,k^*} - F_k^{L,k^*})\right|\right)}$, when $UAMBI(f,g) = 1$, distribution g constitutes an unambiguous Modal Preserving Spread of distribution f. Furthermore, given dispersion from the focus point k^* is maximized when k^*/K mass is allocated to the lowest outcome and $\frac{(K-k^*)}{K}$ is allocated to the highest outcome:

---

[2] This construct is similar to notions of left and right distributional separation developed in Anderson (2004).

[3] Indeed, in the unordered categorical world IC and II provide equally useful indices of commonality and inequality.

$$0 \le IMPS(g,f) = \frac{\sum_{k=1}^{K}((G_k^{U,k^*} - F_k^{U,k^*}) + (G_k^{L,k^*} - F_k^{L,k^*}))}{\left(\frac{k^*\sum_{i=1}^{k^*-1} i}{K} + \frac{(K-k^*)\sum_{i=k^*+1}^{K}(i-k^*)}{K} - \sum_{k=1}^{K}\left(F_k^{U,k^*} + F_k^{L,k^*}\right)\right)} \le 1 \qquad [2]$$

provides an index measure on the unit interval of the extent of increased Modally Focused relative spread or inequality associated with a move from f to g. Suppose f^e was the distribution of a completely equal society with all agents enjoying outcome k^*, then $p_{fk^*} = 1$ and $p_{fk} = 0 \ \forall \ k \neq k^*$ so that $F_k^{U,k^*} = 0$ and $F_k^{L,k^*} = 0 \ \forall \ k$, then [2] becomes:

$$IMPS(g,f^e) = \frac{\sum_{k=1}^{K}(G_k^{U,k^*} + G_k^{L,k^*})}{\left(\frac{k^*\sum_{i=1}^{k^*-1} i}{K} + \frac{(K-k^*)\sum_{i=k^*+1}^{K}(i-k^*)}{K}\right)} = MFI(g) \qquad [3]$$

[3] corresponds to a measure of the extent of inequality inherent in the ordered categorical distribution g relative to a state of complete equality at the category most likely to command unanimous membership and provides a measure, $MFI(g)$, of the Modally Focused Inequality inherent in distribution g. Let the k^* Focussed Probabilistic Distance vector $\underline{G}^{PD,k^*}$, recording the chance of being in the collection of categories successively further distanced from k^*, be given by:

$$\underline{G}^{PD,k^*} = \begin{bmatrix} G_1^L \\ . \\ G_{k^*-1}^L \\ 0 \\ G_{k^*+1}^U \\ . \\ G_{K-k^*}^U \end{bmatrix}$$

Note that the Probabilistic Distance function is an increasing function of the categorical distance from the i^* category which does not depend upon arbitrary attribution of value to a category in the form of a scale. Letting $\varphi(K,k^*) = \left(\frac{k^*\sum_{i=1}^{k^*-1} i}{K} + \frac{(K-k^*)\sum_{i=,k^*+1}^{K}(i-k^*)}{K}\right)$ and given a K dimensioned unit vector d, $MFI(g)$ may be written as:

$$MFI(g) = \frac{1}{\varphi(K,k^*)} d' \underline{G}^{PD,k^*} \qquad [4]$$

Higher Order Considerations.

It is possible to construct an inequality index which weighs more heavily mass at the extremes of the distribution. This higher order analysis reflects greater concern for distance from the modal point of commonality. Following the Gravel, Magdalou and Moyes (2021) exploitation of Hammond (1976) transfers, define G_k^{HU,k^*}, the Upper Cumulated Cumulants of g as $G_k^{HU,k^*} = \sum_{i=k^*+1}^{k} G_i^{U,k^*}$ (note for $k \le k^*$, $G_k^{HU,k^*} = 0$) and, for outcomes $k = 1,..,k^*-1$, define G_k^{HL,k^*} its Lower Cumulated Cumulants as $G_k^{HL,k^*} = \sum_{i=k}^{k^*-1} G_i^{L,k^*}$ (note for $k \ge k^*$, $G_k^{HL,k^*} = 0$). Let the k^* Focused Higher Order Probabilistic Distance vector $\underline{G}^{HPD,k^*}$, recording the chance of being in the collection of categories successively further distanced from k^*, be given by:

$$\underline{G}^{HPD,k^*} = \begin{bmatrix} G_1^{HL} \\ . \\ G_{k^*-1}^{HL} \\ 0 \\ G_{k^*+1}^{HU} \\ . \\ G_{K-k^*}^{HU} \end{bmatrix}$$

Then for a suitably redefined scaling function $\varphi^H(K, k^*)$, a higher order modally focused inequality measure $MFIH(g)$, may be written as:

$$MFIH(g) = \frac{1}{\varphi^H(K,k^*)} d'\underline{G}^{HPD,k^*} \qquad [5]$$

Axiomatic Issues.

The axiomatic development of indices has been popular in inequality measurement (Sen 1997) and in that regard, II, MFI and $MFIH$ indices can each be shown to satisfy axioms of continuity, scale independence, normalization and coherence. All the indices are continuous in the probability measure p_k, are scale independent by definition (any scale attributed to the categories does not appear in the formulae) and normalized, i.e. confined to the unit interval. In the case of coherence, the inequality measure should diminish when LC, the likelihood of complete commonality or equality (recall $LC = (Kp_{k^*} - 1)/(K - 1)$), increases. As II is the complement of LC, it automatically diminishes as LC increases. With regard to MFI and $MFIH$, both are monotonic increasing functions of p_k, $k \neq k^*$ and, since $p_{k^*} = (LC(K-1)+1)/K$, for any $k \neq k^*$,and $p_k = 1 - p_{k^*} - \sum_{\substack{i=1 \\ i\neq k^*, i\neq k}}^{I} p_i$, it will be the case that: $p_k = 1 - (LC(K-1)+1)/K - \sum_{\substack{i=1 \\ i\neq k^*, i\neq k}}^{I} p_i$ so that both MFI and $MFIH$ will be diminishing functions of the likelihood of complete commonality or equality. Yalonetzky (2021) recently proposed a consistency property whereby, in the case of an inequality index, its value should be unaffected when the categorical ordering is reversed. Since reversing the ordering of g doesn't change its modal category and the upper cumulants become the lower cumulants and the lower cumulants become the upper cumulants, this simply causes G_i^L (G_i^{HL}) for $i = 1,..,k^* - 1$ to be renamed G_i^U (G_i^{HU}) $i = 1,..,k^* - 1$ and G_i^U (G_i^{HU}) for $i = k^* + 1..K$ to be renamed G_i^L (G_i^{HL}) $i = k^* + 1 \ldots, K$. With all computed values unaltered, $II(g)$, [4] and [5] would be unaffected by the reversal and the consistency property satisfied.

Inference.

To facilitate inference, note that $\underline{G}^{PD,i^*}$ can be obtained from the probability density vector by pre-multiplying $\underline{p}_g$ by the K dimensioned square cumulation matrix C_{k^*} with typical element $c_{i,j}$ $i,j = 1,..,I$ where for $i,j < k^*$, $c_{i,j} = 1$ $when\ j \geq i\ and\ 0$ otherwise, and for $i,j > k^*$, $c_{i,j} = 1$ $when\ j \leq i\ and\ 0$ otherwise, all other elements of the matrix are 0. Thus: $\underline{G}^{PD,k^*} = C_{k^*}\underline{p}_g$ Where, as an example, for I = 6 and $k^* = 3$, C_{k^*} is of the form:

$$C_{k^*} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

Inference in this context is straightforward. Following Rao (2009) under independent random sampling, $\underline{\widehat{p_g}}$, the estimator of the vector of outcome probabilities $\underline{p_g}$ is multivariate normal:

$$\sqrt{n}\left(\underline{\widehat{p_g}} - \underline{p_g}\right) \sim N\left(\underline{0}, V_g\right)$$

where:

$$V_g = \begin{bmatrix} p_{1,g} & 0 & 0 & . & 0 \\ 0 & p_{2,g} & 0 & . & 0 \\ 0 & 0 & p_{3,g} & . & 0 \\ . & . & . & . & 0 \\ 0 & 0 & 0 & . & p_{\mathrm{K},g} \end{bmatrix} - \begin{bmatrix} p_{1,g} \\ p_{2,g} \\ . \\ . \\ p_{\mathrm{K},g} \end{bmatrix} \begin{bmatrix} p_{1,g} & p_{2,g} & . & . & p_{K,g} \end{bmatrix}$$

And $\underline{\hat{G}_g}^{PD,k^*} = C_{k^*}\underline{\widehat{p_g}}$ will be such that:

$$\sqrt{n}\left(\underline{\hat{G}_g}^{PD,k^*} - \underline{G_g}^{PD,k^*}\right) \sim N\left(\underline{0}, C_{k^*}V_g C_{k^*}{}'\right)$$

So that $\widehat{MFI}(g)$, estimates of $MFI(g)$ will be such that:

$$\sqrt{n}\left(\widehat{MFI}(g) - MFI(g)\right) \sim \sqrt{n}\frac{1}{\varphi(K,k^*)} d'\left(\underline{\hat{G}_g}^{PD,k^*} - \underline{G_g}^{PD,k^*}\right) \sim N\left(0, \frac{1}{\varphi(K,k^*)^2} d' C_{k^*}V_g C_{k^*}{}' d\right)$$

Similarly:

$$\sqrt{n}\left(\widehat{MFIH}(g) - MFIH(g)\right) \sim \sqrt{n}\frac{1}{\varphi^H(K,k^*)} d'\left(\underline{\hat{G}_g}^{HPD,k^*} - \underline{G_g}^{HPD,k^*}\right) \sim N\left(0, \frac{1}{\varphi^H(K,k^*)^2} d' C_{k^*}{}^2 V_g (C_{k^*}{}^2)' d\right)$$

The Multivariate Paradigm.

Consider the bivariate categorical case where both dimensions are ordered with $p_{f,i,j} \geq 0$: $i = 1,..,I, j = 1,..,J$ $\sum_{i=1}^{I}\sum_{j=1}^{J} p_{f,i,j} = 1$ with the ordering again following the dimension indexing, cumulative and counter cumulative density functions are well defined with $F_{i,j} = \sum_{k=1}^{i}\sum_{l=1}^{j} p_{f,k,l}$ $for\ i = 1,..,I, j = 1,..,J.$

In the modal case where k^* coordinates are $\{i^*, j^*\}$ so that $\max_{i,j} p_{f,i,j} = p_{f,i^*,j^*}$:

$$\text{Let } p_{f,i^*,j}^{**} = p_{f,i^*,j}\ j = 1,..,J\ and\ \ p_{f,i,j^*}^{**} = p_{f,i,j^*}\ i = 1,..,I$$

$$F_{i,j}^{**} = F_{i+1,j}^{**} + p_{f,i,j}\ \forall\ i < i^*\ and\ F_{i,j}^{**} = F_{i,j}^{**} + p_{f,i,j}\ \forall\ i > i^*, \forall j = 1,..,J$$

$$F_{i,j}^{\mathrm{L}k^*} = F_{i,j+1}^{\mathrm{L}k^*} + F_{i,j}^{**}\ \forall\ j < j^*\ and\ F_{i,j+1}^{\mathrm{U}k^*} = F_{i,j}^{\mathrm{U}k^*} + F_{i,j}^{**}\ \forall\ j > j^*, i = 1,..,I$$

Again, when p_{f,i^*,j^*} is viewed as the likelihood that the whole population resides in outcome $\{i^*, j^*\}$, $IC(f) = \left(IJp_{f,i^*,j^*} - 1\right)/(IJ - 1)$ is a very natural measure or index on the unit interval of the commonality or equality of outcome in the distribution, so that its complement, $II(f) = IJ\left(1 - p_{f,i^*,j^*}\right)/(IJ - 1)$ provides an intuitive likelihood based measure of inequality of outcome and it is an equally useful index of such in unordered categorical paradigms.

The corresponding 2-dimensional version of [3] is given by:

$$MFI(g) = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\left(G_{i,j}^{U,k^*} + G_{i,j}^{L,k^*}\right)}{\left(\frac{j^*i^*\sum_{i=1}^{i^*-1}\sum_{j=1}^{j^*} ij}{IJ} + \frac{(IJ - j^*i^*)\sum_{i=i^*+1}^{I}\sum_{j=j^*+1}^{J}(ij - i^*j^*)}{IJ}\right)}$$

Appropriately vectorized versions of the I x J matrices $G_{.,.}^{U,k^*}$ and $G_{.,.}^{L,k^*}$ and their corresponding IJ square cumulation matrix $C_{i^*j^*}$ can be constructed to form the i^*, j^* Focused Probabilistic Distance vector $\underline{G}^{PD,i^*,j^*}$, recording the chance of being in the collection of categories successively further distanced from i^*, j^*. Then, given an IJ dimensioned unit vector d, $MFI(g)$ may be written as:

$$MFI(g) = \frac{1}{\varphi(IJ, i^*, j^*)} d'\underline{G}^{PD,i^*,j^*} \qquad [4]$$

Such that:

$$\sqrt{n}\left(\widehat{MFI}(g) - MFI(g)\right) \sim \sqrt{n}\frac{1}{\varphi(K, i^*, j^*)} d'^{\left(\underline{\hat{G}}_g^{PD,i^*,j^*} - \underline{G}_g^{PD,i^*,j^*}\right)}$$

$$\sim N\left(0, \frac{1}{\varphi(K, i^*, j^*)^2} d'C_{i^*,j^*}V_g C_{i^*,j^*}'d\right)$$

Extension to higher dimensioned outcomes and their corresponding asymptotic distributions is straightforward (though somewhat tedious!).

References.

Anderson G.J. “Toward an Empirical Analysis of Polarization” (2004) Journal of Econometrics 122 1-26

Anderson G.J. and T.W. Leo. (2021) “Sufficient conditions for j’th order stochastic dominance for discrete cardinal variables, and their formulae,” Economics Letters, Volume 209, 110144.

Allison, R. A. and J. E. Foster (2004). Measuring health inequality using qualitative data. Journal of Health Economics 23, 505- 24.

Bond T.N. and K. Lang (2019) "The Sad Truth about Happiness Scales," Journal of Political Economy 127, no. 4: 1629-1640.

Gravel, N., Magdalou, B., Moyes, P., (2021). Ranking distributions of an ordinal variable. Economic Theory 71, 33–80.

Hammond, P.J. (1976) “Equity, Arrow’s Conditions and Rawls’s Difference Principle” Econometrica 44 793-803

Kobus, M. (2015). Polarisation measurement for ordinal data. *Journal of Economic Inequality 13(2),* 275- 97.

Mendelson, H. (1987). Quantile- preserving spread. *Journal of Economic Theory* ***42,*** 334-51.

Sen A.K. (1997) On Economic Inequality. Expanded edition with a substantial annexe by James E. Foster and Amartya Sen. Clarendon Press Oxford.

Yalonetzky G. (2021) “Consistent Inequality Measurement for ordinal variables”. Mimeo.