Statistica & Applicazioni - Vol. VI, n. 1, 2008, pp. 57-63

A NOTE ON THE BIPOLAR MEAN: IS IT A SINGLE MEAN?

Alessandro Zini*

SUMMARY

In this note a generalisation of the bipolar mean, recently introduced in the literature by Maffenini and Zenga (2005), is presented and discussed. It seems interesting in those contexts, where the categories of an ordinal variable are to be thought as "non equidistant", like in some problems in the customer satisfaction analyses. It is pointed out that the particular bipolar mean chosen for a problem can modify the "ordering" among the frequency distributions with fixed sum N.

Keywords: *internality, non equidistant categories, skewed (asymmetric) distribution, stochastic dominance, transfers of frequency units.*

1. INTRODUCTION: A MEAN VALUE FOR ORDINAL VARIABLES, THE BIPOLAR MEAN

The definition of some mean value for qualitative ordinal variables is an old question in statistics. Now, an interesting proposal due to Maffenini and Zenga (2005) is presented.

Let us consider a distribution of a qualitative ordinal variable:

$$B = \begin{cases} b_1 \prec \cdots \prec b_j \prec \cdots \prec b_s \\ n_1 \quad \cdots \quad n_j \quad \cdots \quad n_s \end{cases} \qquad N = \sum_{j=1}^{s} n_j,$$

and its cumulative and retro-cumulative frequencies, respectively:

$$C_j = \sum_{i=1}^{j} n_i \quad \text{and} \quad R_j = \sum_{i=j}^{s} n_i.$$

In the sequel, β will be the set of the $\binom{N+s-1}{N}$ frequency distributions of B with fixed sum N (see, Feller (1968)), while β^* the set of the $(N \cdot s - N + 1)$ frequency distributions of B with fixed sum N, which are concentrated either on one category or on two consecutive categories.

Definition 1 *(Stochastic Dominance)*
Let us consider two distributions of β, $(n_1, \cdots, n_s)$ and $(n'_1, \cdots, n'_s)$. $(n'_1, \cdots, n'_s)$ is said to <u>dominate</u> $(n_1, \cdots, n_s)$, by writing

* Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali - Università degli Studi di Milano-Bicocca. - P.zza dell'Ateneo Nuovo, 1 - 20126 MILANO (e-mail: alessandro.-zini@unimib.it).

$$(n_1, \cdots, n_s) \overset{W}{\prec} (n_1, \cdots, n_s)$$

if

$$R'_j \geq R_j \quad \text{(or, equivalently,} \quad C'_j \leq C_j),$$

for all $j = 1,\ 2, \cdots,\ s$, with at least one strict inequality.

It may be shown that the function

$$\bar{S} = \frac{1}{N}\sum_{j=1}^{s} R_j = \frac{1}{N}\sum_{j=1}^{s}\sum_{i=j}^{s} n_i \tag{1}$$

may be written also as follows:

$$\bar{S} = \frac{1}{N}\sum_{j=1}^{s} j \cdot n_j,$$

from which, (1) may ''formally'' be seen as the arithmetic mean of the indexes of the labels of the (ordered) categories of B.

Moreover, $\bar{S}$ is coherent with the ordering $\overset{W}{\prec}$, i.e.:

$$(n_1, \cdots, n_s) \overset{W}{\prec} (n'_1, \cdots, n'_s) \quad \Rightarrow \quad \bar{S}(n_1, \cdots, n_s) < \bar{S}(n'_1, \cdots, n'_s).$$

Definition 2 *(Bipolar Mean)*
Let $\bar{s}$ be the value of $\bar{S}$ with respect to $(n_1, \cdots, n_s)$. $\bar{b} = \bar{b}(n_1, \cdots, n_s)$ is called bipolar mean BM of the distribution $(n_1, \cdots, n_s)$. In particular:

1. if $\bar{s}$ is an integer $r = 1, \cdots, s$, then BM is the distribution which degenerates on the category b_r (i.e., the N statistical units have category b_r);
2. if $r < \bar{s} < r + 1$, $r = 1, \cdots, s - 1$, then

$$BM = \begin{cases} \quad b_r & \quad b_{r+1} \\ n_r = N \cdot (r + 1 - \bar{s}) & n_{r+1} = N(\bar{s} - r) \end{cases}. \tag{2}$$

Instead of synthesizing a distribution by (only) one category, as median and mode do, through the bipolar mean a distribution in β is synthesized by a distribution concentrated on one or two consecutive categories (i.e., on a distribution in β^*): so, the instrument the measurements are made by is much more precise (Maffenini and Zenga (2005), Corollary 1 p. 12).

2. THE GENERALISED BIPOLAR MEAN

Instead of $\bar{S}$, the following function is considered:

$$\bar{K} = \varphi^{-1}\left\{\frac{1}{N}\sum_{j=1}^{s} \varphi(j) \cdot n_j\right\} = \varphi^{-1}[M(\varphi)], \tag{3}$$

where $\varphi(.)$ is monotonic.

By the internality property of the arithmetic mean it follows that:

$$\min \varphi(j) \leq M(\varphi) \leq \max \varphi(j),$$

and, as $\varphi(.)$ is monotonic:

$$1 \leq \bar{K} \leq s.$$

Taking into account **Definition 2**, it is important, from the point of view of the bipolar mean, which values are taken by (3) when considering the distributions belonging to β^*: these values will be indicated by $\bar{k}^*$.

By definition, any distribution of B in β^* has the following form:

$$B = \begin{cases} b_r, & b_{r+1} \\ n_r & n_{r+1} = N - n_r \end{cases}, \qquad n_r = 0,\ 1, \cdots, N \quad r = 1,\ 2, \cdots, s-1,$$

and the corresponding value of (3) is:

$$\bar{k}* = \varphi^{-1}\left\{\frac{n_r\varphi(r) + (N - n_r)\varphi(r+1)}{N}\right\}. \tag{4}$$

Solving for n_r, (4) yields:

$$n_r = N\frac{\varphi(r+1) - \varphi(\bar{k}^*)}{\varphi(r+1) - \varphi(r)}, \tag{5}$$

which is an integer number, for all the distributions in β^*.

Moreover, (5) is a decreasing function of $\bar{k}^*$.

Thus:

1. the relation between the values $\bar{k}^*$ and the distributions in β^* is one to one;
2. each $\bar{k}^*$ yields an integer value for n_r.

Another aspect must be inspected in order to generalise the bipolar mean. Computing in (3) the value $\bar{k} \in [r, r+1]$, for some integer r, two cases may occur:

1. $\bar{k} = \bar{k}^*$. This may happen even if $\bar{k}$ is not integer; e.g. when B has the following distribution

   $$B = \begin{cases} b_1 & b_2 & b_3 & b_4 & b_5 \\ 1 & 3 & 5 & 1 & 0 \end{cases},$$

   and $\varphi(x) = \log x$,

   then $\bar{k} = \sqrt[10]{1^1 2^3 3^5 4^1 5^0} = \sqrt{6}$, $\bar{k}* = \sqrt[10]{2^5 3^5} = \sqrt{6}$.
2. $\bar{k} \neq \bar{k}^*$. In this case, one must choose the value $\bar{k}^*$ nearest to $\bar{k}$, and the related distribution (the generalised bipolar mean) in β^*.

In both cases, the generalised bipolar mean concentrates the frequencies either on one or on two (consecutive) categories whose frequencies are given by (5). Now, it is time to introduce the following

Definition 3 *(Generalised Bipolar Mean)*
Let $\bar{k}$ be the value of $\bar{K}$ with respect to $(n_1, \cdots, n_s)$ in (3). $\bar{g} = \bar{g}(n_1, \cdots, n_s)$ is called generalised bipolar mean *GBM* of the distribution $(n_1, \cdots, n_s)$. In particular:

1. if $\bar{k}$ is an integer r then *GBM* is the distribution which degenerates on the category b_r (i.e., the N statistical units possess category b_r), $r = 1, \cdots, s$;
   if $r < \bar{k} < r + 1$, $r = 1, \cdots, s - 1$, then

$$GBM = \begin{cases} \qquad b_r & \qquad b_{r+1} \\ n_r = N\frac{\varphi(r+1)-\varphi(\bar{k}*)}{\varphi(r+1)-\varphi(r)} & n_{r+1} = N\frac{\varphi(\bar{k}*)-\varphi(r)}{\varphi(r+1)-\varphi(r)} \end{cases}, \tag{6}$$

2. where $\bar{k}^*$ is the nearest value with respect to $\bar{k}$, and n_r is an integer number.

3. APPLICATIONS AND CONCLUSIONS

Let us consider in (3) the important class of "power" means, obtained by choosing as $\varphi(.)$ the following function:

$$\varphi(j) = \begin{cases} j^\gamma & \gamma \neq 0 \\ \log(j) & \gamma = 0 \end{cases},$$

yielding:

$$\bar{K}_\gamma = \begin{cases} \left(\frac{1}{N}\sum_{j=1}^{s} j^\gamma n_j\right)^{\frac{1}{\gamma}}, & \gamma \neq 0 \\ \prod_{j=1}^{s} j^{\frac{n_j}{N}}, & \gamma = 0 \end{cases}. \tag{7}$$

For a well known property of power means, the inequalities $\bar{K}_0 \leq \bar{K}_1 \leq \bar{K}_2$ are fulfilled.

Here few non trivial examples are presented.

EXAMPLE 1 Let Δ_r be the distribution which concentrates all the N statistical units on the category b_r.

TABLE 1. - *Some distributions of B, $N = 20$ $s = 5$*

j	A	B	C	D	E	F	G	H	I	L
1	10	8	5	4	0	0	0	0	0	0
2	0	1	3	4	10	8	8	11	12	0
3	0	0	0	0	0	1	0	1	0	20
4	0	0	0	0	10	11	12	8	8	0
5	10	11	12	12	0	0	0	0	0	0
$\bar{k}_0$	2.236	2.509	2.914	3.017	2.828	2.988	3.031	2.693	2.639	3
$\bar{k}_1$	3	3.25	3.55	3.6	3	3.15	3.2	2.85	2.8	3
$\bar{k}_2$	3.606	3.619	3.981	4	3.162	3.294	3.347	3.008	2.967	3

Referring to TABLE 1, distributions A and E are symmetric, so the bipolar mean induced by $\bar{K}_1$ yields the distribution Δ_3= L. In this context, for symmetric distributions, $\bar{K}_0$ takes values less than 3, so the respective distributions induced by $\bar{K}_0$ are dominated by Δ_3, while $\bar{K}_2$ takes values greater than 3, so the distributions induced by $\bar{K}_2$ dominate Δ_3.

So, indicating with $\underset{*}{\prec}$ the ordering between the distributions induced by a *GBM,* according to $\bar{K}_1$, $I \underset{*}{\prec} K \underset{*}{\prec} L \sim E \sim A \underset{*}{\prec} F \underset{*}{\prec} G \underset{*}{\prec} B \underset{*}{\prec} C \underset{*}{\prec} D$, according to $\bar{K}_0$, $A \underset{*}{\prec} B \underset{*}{\prec} I \underset{*}{\prec} H \underset{*}{\prec} E \underset{*}{\prec} C \underset{*}{\prec} F \underset{*}{\prec} L \underset{*}{\prec} D \underset{*}{\prec} G$, and according to $\bar{K}_2$, $I \underset{*}{\prec} L \underset{*}{\prec} H \underset{*}{\prec} E \underset{*}{\prec} F \underset{*}{\prec} G \underset{*}{\prec} A \underset{*}{\prec} B \underset{*}{\prec} C \underset{*}{\prec} D$.

In conclusion, *the choice of the mean influences the ordering among the distributions*.

EXAMPLE 2 $N = 50$ customers give their judgements about the good/service X, according to $s = 5$ categories (TABLE 2).

It is known that:

1. if a $\bar{K}$ takes an integer value r for a distribution, it means that this distribution is equivalent to Δ_r in the ordering.
2. the *BM* (induced by $\bar{K}_1$) makes the distributions, which are symmetric about b_r, equivalent to Δ_r.

Instead, according to point 2, it is important to note that, in general, a *GBM* has a different behaviour: for certain *GBM*s, distributions which are symmetric about b_r follow Δ_r in the stochastic ordering induced by the *GMB*, while for other *GBM*s the distributions which are symmetric about b_r are followed by Δ_r in the stochastic ordering induced by the *GMB*.

TABLE 2. - *Some distributions of B $N = 50$, and related distributions in β^**

	M	N	O	P	Q
very unsatisfactory unsatisfactory neutral satisfactory very satisfactory	13 12 10 8 7	7 8 10 12 13	4 7 10 13 16	16 13 10 7 4	24 0 2 0 24
$\bar{k}_0$	2.300	2.950	3.294	2.060	2.263
$\bar{k}_1$	2.680	3.320	3.800	2.200	3.000
$\bar{k}_2$	3.013	3.594	3.821	2.721	3.583
$\bar{k}_0$: distribution in β^*	$n_2 = 33$ $n_3 = 17$	$n_2 = 2$ $n_3 = 48$	$n_3 = 34$ $n_4 = 16$	$n_2 = 46$ $n_3 = 4$	$n_2 = 35$ $n_3 = 15$
$\bar{k}_1$: distribution in $\beta*$	$n_2 = 16$ $n_3 = 34$	$n_3 = 34$ $n_4 = 16$	$n_3 = 10$ $n_4 = 40$	$n_2 = 40$ $n_3 = 10$	$n_3 = 50$ $n_4 = 0$
$\bar{k}_2$: distribution in $\beta*$	$n_3 = 49$ $n_4 = 1$	$n_3 = 22$ $n_4 = 28$	$n_3 = 10$ $n_4 = 40$	$n_2 = 16$ $n_3 = 34$	$n_3 = 23$ $n_4 = 27$

This fact is shown in TABLE 2, where Δ_3 is equivalent to Q according to $\bar{K}_1$, follows Q according to $\bar{K}_0$, and is followed by Q according to $\bar{K}_2$.

Every customer may be a market maker with respect to new potential customers, according to the nature of X (please, note that information about X is available to the researcher, e.g. the elasticity of the demand curve provides a dimension of the market).

It is realistic to assume that customers judging X as ''neutral'' are not market makers: thus, the possible distributions of the judgements are to be compared with Δ_3. According to the nature of X, customers which are unsatisfied may ''compensate'' as market makers those who are satisfied, so that symmetric distributions of judgements may be thought equivalent to Δ_3, as the (''arithmetic'') bipolar mean does. In other situations, instead, this does not occur: either ''unsatisfied'' customers may inform other potential customers and discourage them to buy it, or ''satisfied'' customers may favour an increase in the number of new customers.

In the former case, it is natural to assume that symmetric distributions about b_3 are dominated by Δ_3, while in the latter case, it is natural to assume that symmetric distributions about b_3 dominate Δ_3, in the stochastic ordering induced by the *GMB*.

Moreover, it is possible to show that, for positive transfers of frequency units, when $\gamma < 1$ in (7), the distance between consecutive values of the mean increases, while when $\gamma > 1$, the distance between consecutive values of the mean decreases.

Then, in the former case, a *GBM* induced by (7) with $\gamma < 1$ will be chosen, while in the latter case a *GBM* induced by (7) with $\gamma > 1$ will be chosen.

The distributions of TABLE 2 satisfy the stochastic order of **Definition 1**, except for distribution Q, which is symmetric about b_3:

$$P \overset{W}{\prec} M \overset{W}{\prec} N \overset{W}{\prec} O$$

The *GBM*s are coherent with this ordering, and it is interesting to examine their different behaviour in evaluating distribution Q:

- according to $\bar{K}_1$, $P \underset{*}{\prec} M \underset{*}{\prec} Q \sim \Delta_3 \underset{*}{\prec} N \underset{*}{\prec} O$,
- according to $\bar{K}_0$, $P \underset{*}{\prec} Q \underset{*}{\prec} M \underset{*}{\prec} N \underset{*}{\prec} \Delta_3 \underset{*}{\prec} O$,
- according to $\bar{K}_2$, $P \underset{*}{\prec} \Delta_3 \underset{*}{\prec} M \underset{*}{\prec} Q \underset{*}{\prec} N \underset{*}{\prec} O$.

Now, some conclusive remarks are needed:

1. If the categories may be thought as ''equidistant'', the (''arithmetic'') bipolar mean of **Definition 2** may be a good – probably the best - candidate to synthesize a distribution.
2. In most cases, on the other hand, the categories may be thought as non equidistant. So, a different mean induced by (3) seems to be a suitable choice.
3. According to the previous points, a researcher may choose a bipolar mean in (6) according to:
   - the evaluation of distances between categories, from a subjective point of view, by considering the ordering of some distributions in β;
   - the estimation of distances between categories, as in a customer satisfaction analysis. For this case, new proposals are needed.

ACKNOWLEDGEMENT

The author wishes to thank the anonymous reviewers for their useful comments which improved the manuscript.

RIASSUNTO

Il lavoro presenta e discute una generalizzazione della media bipolare, recentemente introdotta in letteratura da Maffenini e Zenga (2005). Essa appare interessante nei contesti nei quali le categorie di una variabile ordinale possono pensarsi ''non equidistanti'', come accade in certi problemi di analisi della customer satisfaction. Si sottolinea che la particolare media bipolare scelta per un problema può modificare l'ordinamento fra le distribuzioni di frequenze a somma fissata N.

REFERENCES

Feller W. (1968). *An introduction to probability theory and its applications*. Volume 1. Third Edition. John Wiley & Sons, New York. London, p. 38.

Maffenini W., Zenga Mi. (2005). Bipolar mean for ordinal variables, *Statistica & Applicazioni*, Vol. III (1), pp. 3-18.

Zenga Mi. (2007). *Lezioni di statistica descrittiva*. Giappichelli, Torino.