# APPENDIX FORMULAS

# **A1: Concentration Index**

Consider a resident population of P individuals and let X= {} be a specific characteristic of the residential population grouped in categories (e.g. age groups, IRSD quantiles, etc.). Let P(X)= {} be the probability mass function of X (i.e. ). The increasing order of the observations is denoted:

Then the Lorenz Curve is the piecewise linear connection of points:

,

in the unit square. Accordingly, the *Lorenz Function* is:

It follows that the area under the Lorenz Curve is:

Therefore, the *Lorenz Zonoid* is

Finally, the CI is given by the following formula:

The Upper Bound of the Concentration Index is (it is attained when all the observation except one are equal to zero). As a result, the Lorenz Zonoid is divided by (See **Theorem A1.1**). It follows that the CI is a function bounded between zero and one.

**Theorem A1.1**: Let P(X) be a probability density function of the random variable

X= {}. Then .

*Proof*:

→

= =

It follows that:

***Q.E.D***

**Theorem A1.2**: Let LF(X) be the Lorenz Function of a random variable X= {}. Then, the area under the Lorenz Curve is

*Proof*:

Reminding the definition of LF(X):

Considering the following identities:

We can compute the Lorenz Curves’ Area:

***Q.E.D***

# **A2: Value validity CI**

To determine the condition for a CI to have Value validity, Kvålseth recently introduced the lambda distribution **(λ-distribution**) [6]. The term lambda distribution has been chosen for the distribution by Kvålseth because of its L-shape (Greek lambda: **λ**) from. For nominal categorical variable the **λ-distribution** is a single parameter distribution where the parameter **λ** reflects the uniformity or evenness of the distribution and the number of categories:

Basically, it is a weighted mean of the one-point distribution , sometimes called singleton or degenerate distribution, and the uniform distribution . Note that a concentration measure for nominal variables is a (permutation) symmetric function and therefore any choice of the one-point distribution could be made. However, we shall consider to be the first element of and 1 to be the first element of . Then, λ is a rectangularity (uniformity) parameter with increasing λ indicating a decreasing approach of from the singleton located in the first category to the rectangular (uniform) distribution . Then, a concentration measure has **Value validity** if it regularly decreases from the one-point distribution to the uniform distribution. This means that adding or subtracting an amount δ to or from λ should cause the same absolute change in the value of CI. This is equivalent to define the CI as a function of a single variable λ, that is . It follows that all the CI values lies on the line segment between and .

# **A3: Divergence Index**

In the statistical literature, for ordinal types of data, are known lots of indicators to measure the degree of the polarization phenomenon. Typically, many of the widely used measures of distributional variability are defined as a function of a reference point, which in some “sense” could be considered representative for the entire population. This function indicates how much all the values differ from the point that is considered “typical”. Of all measures of variability, the variance is a well-known example that use the mean as a reference point. However, mean-based measures depend to the scale applied to the categories and are highly sensitive to outliers.

Then, is there another way to compare the dispersion of a distribution that does not depend on its location?

To address this challenge, we propose a new measure of polarization, the Divergence Index (DI). The definition of the DI is based on a new representation of probability measures, the Bilateral Cumulative Distribution function (**BCDF**), which derives from a generalization of the Cumulative Distribution Function (CDF). Basically, it is an extended CDF that can be easily obtained by folding its upper part, commonly known as Survival function or complementary CDF. Unlike the CDF, this functional has a finite constant area independently of the probability distribution and, therefore, more convenient for any distribution comparison. For the definition of the **BCDF** see (**A3.1: Bilateral Cumulative Distribution Function**). For the area invariance property see theorem A3.1.2.

Accordingly, the DI is a non-parametric measure that compares the BCDF of two data sets. More specifically, the idea is to compare the distribution shape with that of minimum dispersion, that is the singleton distribution. On this basis, we completely defined the shape of a probability distribution by its BCDF autocorrelation function (**BCDFA**). The main advantage of this representation is that it is invariant to the location of a distribution and therefore is only sensitive to the distribution shape. For example, any singleton distribution will have the same BCDFA curve. Similarly, distribution with same shape but different means and medians can be represented with a unique curve. In this way, to quantify the distance between a pdf and the singleton distribution, it is only a matter of choosing an appropriate metric, which would assign large values to distribution with more dispersed BCDFA than the singleton. For the definition of the **BCDFA** see (**A3.2: Bilateral Cumulative Distribution Function Autocorrelation**).

The selection of a measure to compare probability distribution is not a trivial matter and usually depends on the objectives. In this work we propose the use of the Jensen Shannon Divergence (JSD) [7]. Since the DI is based on the JSD we dub this new measure of polarization **DI-JSD**. For the definition see (**A3.3: DI-JSD**). This measure does not need to be normalized since is a bounded value in the unit interval see **theorem A3.3.1.**

# **A3.1: Bilateral Cumulative Distribution Function**

Let be the probability mass function of the discrete random variable

X = such that is a categorical ordinal variable of x (e.g. Age, IRSD score, etc.). It follows that:

Let Y = be a discrete random variable such that:

Then the *Bilateral Cumulative Distribution Function* **BCDF** of X is given by the following formula:

It follows that the recursive definition is:

It is worth noting that: for → and . Basically, the BCDF Complementary CDF.

Considering, therefore, the recursive definition and the integral relation we can compute the BCDF in terms of **the convolution operator** (see **Theorem A3.1.1**).

**Theorem A3.1.1**: **Convolution operator**

Let be the BCDF of the discrete random variable X [see A3.1] and P(X) = the corresponding probability mass function vector. Let be a n-dimensional unitary vector, then .

*Proof*:

Since is simply the folded sequence shifted to the right by units (if k is positive) or to the left (if k is negative), it follows:

*Q.E.D*

**Theorem A3.1.2: Area invariance property**

Let be the BCDF of the discrete random variable , and a N-dimensional unitary vector. Then

*Proof*:

Let A denote the area of BCDF:

Then

This can be proved by:

Since the area (do not change with a shift in τ). It follows that:

*Q.E.D*

# **A3.2: Bilateral Cumulative Distribution Function Autocorrelation**

Let denote the folding sequence of :

Then the autocorrelation function of is given by the following formula:

Since the maximum value of the autocorrelation function occurs when the sequence correlates perfectly with itself with zero delay and the symmetry property, we can use the following relations:

It follows that the maximum occurs for

Let denote the singleton distribution of the random variable :

Then the autocorrelation function of , is given by the following formula:

The BCDFA is a symmetric function that preserves the variance of the pdf. See **theorem A3.2.1**

For the parametric formula of the minimum and maximum value of the BCDFA see **theorem A3.2.2**

**Theorem A3.2.1: variance BCDFA**

Let be the BCDFA of the random variable and a sequence of unit impulses spaced at unit interval. Let’s denote the variance of the BCDFA, the impulses sequence and the probability distribution respectively. Then:

*Proof*:

It is known that the variance , which is a measure of spread of the probability distribution , is increased under convolution according to the result: .

It follows that the BCDF of x (i.e. F(Y)) has variance: .

Similarly, the variance of the BCDFA is given by the formula:

It follows that:

*Q.E.D*

**Theorem A3.2.2:**

Let be the autocorrelation function of the random variable . Then:

*Proof*:

Based on theorem A2.2 and the definition of the BCDF (A2):

For sake of simplicity, let’s denote . It follows that:

Substituting the following relations:

It follows that:

The autocorrelation function of the BCDF is therefore a conic. In order to find the maximum and minimum value of the autocorrelation function, it is convenient to reduce the conic in a canonical form:

In order to eliminate the linear terms and operate the translation to the origin, we consider the following conditions:

Since :

These two distributions correspond to the minimum and maximum variance of a bounded probability distribution:

It’s worth noting that in the case N=2 .

***Q.E.D***

# **A3.3: DI-JSD**

Let and be the autocorrelation functions of the random variable and the impulse distribution [see A3].

Based on the area invariance property theorem A3.1.2

It follows that the normalized sequences are:

and

For sake of simplicity let denote and . Then, based on the Jensen-Shannon Divergence (JSD), we propose the following divergence measure:

Where:

)

Since , we assume = 0.

**Theorem A3.3.1: Divergence index bound property**

Let DI(X) be the divergence index of the random variable . Then the DI(X) is a bounded metric such that:

{DI(X)} =

*Proof:*

Let and be the autocorrelation function of the BCDF and its central coordinate (i.e. ). Then we introduce the autocorrelation width as:

That is the area of the autocorrelation function divided by its central ordinate. Based on theorem A3.2.2, it follows that:

{ {

Then:

Since then:

We therefore define as a measure of compactness for certain shapes of distribution:

The minimum value can be easily computed, considering the autocorrelation of the BCDF

It follows that:

The Divergence index is therefore a bounded metric. ***Q.E.D***

# **A3.4: Value validity HI**

To determine the condition for an HI to have Value validity, we shall use the two parameters distribution, where the parameter reflects the polarization or dispersion of the distribution for a given value of . Thus, instead of we shall use:

which reduces to equation for and to:

for For example, the two-point extreme distribution is the most dispersed ( and the least heterogeneous configuration. On the other hand, the is the least dispersed and heterogeneous distribution . Lastly, the uniform distribution is the only distribution where and therefore the most heterogeneous . It is assumed of course, that is strictly decreasing in μ for any given λ. This result is a natural consequence of the simple aversion to median-preserving spreads property.

In this framework, we should expect that a valid HI will be closer to the CI curve for smaller value of the parameter μ and gives reasonably high values for the two-point extreme distribution. This insight leads to the problem of choosing a HI which should vary as little as possible from a selected target function and be not too far away from the CI curve. The task is therefore to find a Value validity condition which incorporates the idea that a well behaved or “smoother” HI is more appropriate than a “rougher” HI. This requirement implies that the HI is bounded above by the values assigned to a target function and below by a minimum curve. These functions set the boundaries of the feasible HI measures and identify a class of slowing decreasing functions or dual Schwartz functions [8] in the space. This means that small changes in the distribution vector should not lead to abrupt changes in the value of the index. This behavior does not overstate evenness and give acceptable values when the distributions have large variance. In this way, the upper and lower bound functions identify a feasible region for the HI. Then, the HI has Value validity if it satisfies the following two conditions:

(**C1**): this condition is based on a loss function which indicates how far away is the HI from the lower function. More precisely, the value of this function must be positive.

(**C2**): the value of the loss function must be less than or equal to the loss function for the target function .

# **A3.5: CI and True Diversity**

The true diversity gives the number of equally abundant categories needed to give a particular value of a concentration index.

The conversion of to any concentration measure, such as the CI, can be easily obtained for a countable number of categories: calculate the CI for s equally abundant categories (s bins with a frequency ) and assign that value to . But how is possible to link the CI to any value of ?

As discussed in section A3 the CI satisfies the Value validity property and therefore it follows a straight line that basically is an arithmetic progression. In a similar way, the CI value can be mapped to the true diversity number in a linear manner. Consequently, the direct and inverse formula are:

Where:

# **Appendix B: Location Index**

An important task of spatial data analysis consists in identifying a part of data which represents the typical features of the population living in a geographic area. Often the data points are assumed to vary around a centre that identifies a single value as representative of an entire distribution. As a result, we propose a new measure of central tendency, the “**Location Index**” (**LI**), that identifies the position of the bin in the distribution where the values are mostly concentrated (i.e. the values in the surrounding areas of the bin are noticeably higher than the others).

The basic idea is to map a nested family of sets around the location of the bin into a single number. If the distribution is univariate the set is a symmetric interval formed by a finite number of categories and the concentration value is given by summing up the likelihood of these regions. In this way, each location has a concentration value and the one with the maximum value corresponds to the point of the LI. For the definition of the LI see (**B1: Location Index**).

Moreover, the definition of the function used in the computation of the bin concentration value, the Bin Concentration Function (BCF), provides simple rules to determine the location of a distribution. For example, considering a decile distribution, the set of necessary and sufficient conditions to determine the LI value is given in Table 1.

|  |  |  |
| --- | --- | --- |
| Location  Index |  | |
| Decile |  |  |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |
| 7 |  |  |
| 8 |  |  |
| 9 |  |  |
| 10 |  |  |

**Table 1 Location Index conditions**

In this table, the first and remaining columns indicate the LI and the deciles intervals that contain at least of the data and those with less or equal to of the samples. For example, the first and last row show that a distribution with a LI equal to one or ten have at least of the data in the first or last decile. Similarly, the value of this index for the other bins indicates the decile position separating the greater and lesser halves of a distribution.

Therefore, the LI turns out to be an equivalent form of the median, having a breakdown point of : so long as no more than half the data are too much dispersed, the LI will not give an arbitrary large or small result. The basic difference of the LI in describing data compared to the median is that the outcome is always a member of the distribution and it is easier to calculate. In addition, since more than one value can be at the median level, this location parameter identifies a unique interval of values for which the conditions hold. Lastly, the LI definition can be generalized to a multivariate distribution.

Finally, Table 2 summarizes several of the properties of the central tendencies measures that have been discussed in the paper.

For the properties see Theorem B1.1 and lemma B1.1-5.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Properties of Measures of Center | | | |
|  | Mode | Median | Mean | LI |
| Level of measurement | Nominal or higher | Ordinal or higher | Metric usually | Ordinal or higher |
| Simple to understand | best | yes | yes | Yes |
| Easy to calculate | best | yes | yes | Yes |
| Algebraic | No | No | yes | Yes |
| Single valued | Not Always | yes | best | Not Always |
| Resistant to outliers | yes | yes | no | Yes |
| Generalize to  n-variables | no | no | yes | Yes |
| Easy interpretation for n variables | no | no | no | Yes |
| Equal to actual data values | yes | Not always | Not Always | Yes |
| Interpretation | Most typical value | Middle value | Average value  Center of mass | Set of middle values |
| Bad guess interpretation | Highest % accuracy | Closest to all scores | Minimized sum of squared deviations or signed deviation | Closest to all scores. Highest concentration |

Table 2 Central Tendency Measure Properties - Adaption of Central Tendency and variability [9]

# **B1: Location Index**

Let be a closed bounded interval of the random variable Then . It follows that . In the case of a symmetric closed bounded interval

Let’s denote the set of nested compact subset of with the following notation:

Then, let’s define the functions (**Likelihood Central Region function)** and (**Level Set Density function**).

Then the **Bin Concentration function** (BCF), is given by the following formula:

Let be the BCF vector of the random varaible X.

Then, the **Location Index** of this random variable is given by the formula:

**Theorem B1.1- Location index convergence property:** Let be a sequence in , such that:

Then .

Proof:

Let and be two collections of non empty closed bounded intervals in ℤ such that

Since and , it follows that:

Then

It follows that

Therefore,

**Q.E.D**

**Lemma B1.1** Let be the *Likelihood Central Region Function* and be the BCF of the random variable . Then the BCF is a linear convex combination such that:

*Proof:*

Let , then the *Likelihood Central Region Function* is given by the following formula:

Since it follows that:

Then the *Level Set Density Function* is given by the following equation:

Then for

It follows that:

*Q.E.D*

Let’s prove for :

Since , it follows that:

Then the *Level Density Function* is given by the following equation:

Then for

It follows that:

*Q.E.D*

Let

The *Likelihood Central Region Function* is given by the following formula:

Since . It follows from the previous results that:

Then the *Level Set Density Function* is given by the following equation:

Let’s solve the first term:

Then for

Since

It follows that:

The second term:

Then for

Since

It follows that:

***Q.E.D***

**Corollary B1.1**: Let be the expected value of the uniform distribution and be the rounded floor value. Then:

*Proof:*

For

For

For

For

For

For

***Q.E.D***

**Lemma B1.2:** Let be the BCF vector of the random variable X

and . Then

*Proof:*

According to the *Corollary B1.1* :

Let’s prove

It follows that:

So

For sake of simplicity let’s denote , then:

It follows:

It follows that

Assume

Then

It follows that:

Therefore, for

Similarly, for

***Q.E.D***

**Lemma B1.3** Let be the LI of the random variable and its probability mass function. Then:

*Proof:*

Since the BCF is a linear convex combination:

It follows that the maximum value is attained when the coefficients of the polynomial are all set to zero execpt for In other words:

On the other hand, the minimum value is attained when all the bins have exactly the same concentration value. This situation occurs when .

Substituiting .

***Q.E.D***

**Lemma B1.4:** Let be the LI of the random variable and be the Bin concentration value. Then the normalized value is given by the following formula:

*Proof:*

According to Lemma B1.3:

It follows that:

It follows that

As for the Bin compactness deviation:

***Q.E.D***

**Lemma B1.5**: Let be the BCF of the random variable and be the average absolute deviation. Then:

*Proof:*

For sake of simplicity let’s dentoe:

It follows that:

Let

Let and be the mass probability of the random variable . Then, since , it follows that:

Let

It follows that:

Let

It follows that:

Where:

***Q.E.D***

# **BIBLIOGRAPHY**

|  |  |
| --- | --- |
| [1] | T. O. Kvalseth, "The lambda distribution and its applications to categorical summary measures," *Advances and Applications in Statistics,* vol. 24, no. 2, pp. 83-106, 2011. |
| [2] | J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Transactions on Information Theory,* 1991. |
| [3] | E. M. Stein and R. Shakarchi, Fourier Analysis an Introduction p.134, New Jersy: Princeton University Press, 2003. |
| [4] | Herbert.F.Weisberg, Central Tendency and Variability, p35, SAGE Publication, 1992. |
| [5] | T. O. Kvalseth, "Evenness indicies once again: critical analysis of properties," *Springerplus,* p. 4:232, 2015. |
| [6] | F. Gosselin, "Lorenz Partial Order: the Best Known Logical Framework to Define Evenness Indicies," *Community Ecology,* pp. 197-207, 2001. |
| [7] | G. Koshevoy and K. Mosler, "The Lorenz Zonoid of a Multivariate Distribution," *Journal of the American Statistical Association,* vol. 91, no. 434, p. 873, 1996. |
| [8] | R. Bathia and D. Chander, "A better Bound on the variance," *The American Mathematical Monthly,* pp. 353-357, 2000. |
| [9] | G. LV and Y. W.-Y. Xu, "On a new class of measures for health inequality based on ordinal data," *Journal of Economic Inequality,* pp. 465-477, 2015. |