mathematics

MDPI

Article

On Generalizing Sarle's Bimodality Coefficient as a Path towards a Newly Composite Bimodality Coefficient

Nicolae Tarbă †, Mihai-Lucian Voncilă † and Costin-Anton Boiangiu *

Computer Science and Engineering Department, Faculty of Automatic Control and Computers, Politehnica University of Bucharest, Splaiul Independenței 313, 060042 Bucharest, Romania; nicolae.tarba@upb.ro (N.T.); mihai_lucian.voncila@stud.acs.upb.ro (M.-L.V.)

* Correspondence: costin.boiangiu@upb.ro

† These authors contributed equally to this work.

Abstract: Determining whether a distribution is bimodal is of great interest for many applications. Several tests have been developed, but the only ones that can be run extremely fast, in constant time on any variable-size signal window, are based on Sarle's bimodality coefficient. We propose in this paper a generalization of this coefficient, to prove its validity, and show how each coefficient can be computed in a fast manner, in constant time, for random regions pertaining to a large dataset. We present some of the caveats of these coefficients and potential ways to circumvent them. We also propose a composite bimodality coefficient obtained as a product of the weighted generalized coefficients. We determine the potential best set of weights to associate with our composite coefficient when using up to three generalized coefficients. Finally, we prove that the composite coefficient outperforms any individual generalized coefficient.

Keywords: bimodality coefficient; bimodality; distributions; image binarization; Sarle's coefficient

MSC: 62-08

Citation: Tarbă, N.; Voncilă, M.-L.; Boiangiu, C.-A. On Generalizing Sarle's Bimodality Coefficient as a Path towards a Newly Composite Bimodality Coefficient. *Mathematics* **2022**, *10*, 1042. https://doi.org/10.3390/math10071042

Academic Editor: Ion Mihai

Received: 22 February 2022
Accepted: 22 March 2022
Published: 24 March 2022

1. Introduction

Statistics represent an important part of dealing with everyday data, i.e., data that follow a certain distribution. Determining the number of modes in a distribution can be a crucial step in making further decisions based on said data. However, this can prove to be rather difficult, in most cases, and many approaches simply rely on just assigning the best number that fits, via trial and error.

Many have explored ways to test whether a distribution is bimodal and have found better methods than the ones available before them. This field has been growing for a long time and there seems to be no end in sight. An important application for these tests is the binarization of an image, for which worldwide competitions are often held to encourage finding an even better algorithm than the ones before it. Since these sorts of applications just require finding a threshold, we only need to determine whether a distribution is either bimodal or unimodal to find an optimal threshold. Thus, a metric that could ascertain this characteristic would be beneficial. Such metrics have been proposed before, but none are without flaws, so a more robust solution is needed.

The dip test of unimodality is a non-parametric test proposed by Hartigan and Hartigan [1] that measures multimodality in an empirical distribution by computing, over all points, the maximum difference between the empirical distribution and the unimodal distribution that minimizes said maximum difference. However, this test only determines if a distribution is multimodal or not and does not help with determining how many modes a multimodal distribution has.

Silverman [2] uses kernel density estimates to investigate multimodality. His test consists of finding the maximal window h for which the kernel density estimate has

k modes or less. A large value for this window indicates that the distribution has k modes or less and a small value indicates that it has more than k modes. This test is computationally demanding and must be run for both the number of modes you want to test for, and that number minus one.

Muller and Sawitzki [3] propose a test based on the excess mass functional, which measures excessive empirical mass in comparison with multiples of uniform distributions. Estimators for the excess mass functional are built iteratively and they converge uniformly towards it.

The MAP test for multimodality proposed by Rozal and Hartigan [4] uses minimal ascending path spanning trees to compute the MAPk statistic which indicates k-modality if it has a large value. Computing the MAPk statistic is straightforward, but the trees are created with Prim's algorithm which has a complexity of log-linear order. Creating the trees is time-consuming.

Ashman et al. [5] use the KMM algorithm to test for multimodality. They fit a user-specified number of Gaussian distributions to the empirical distribution and iteratively adjust the fit until the likelihood function converges to its maximum value. A good fit for a k-modal model indicates k-modality, but multiple values for k must be tested. Fitting a complex model is extremely cumbersome, especially for high values for k, and the test is not reliable for distributions drastically different from the model.

Zhang et al. [6] present three measures for bimodality after fitting a bimodal normal mixture to a distribution. The first one, bimodal amplitude, is calculated as $A_B = \frac{A_M - A_V}{A_M}$, with A_V being the amplitude of the minimum PDF between the two peaks and A_M the amplitude of the smaller peak of the two. The second one, bimodal separation, is based on the characteristics of the two normal distributions that form the mixture and has the following formula $S = \frac{\mu_2 - \mu_1}{2(\sigma_1 + \sigma_2)}$. The third one, bimodal ratio, is computed with $R = \frac{A_R}{A_L}$, where A_R and A_L are the amplitudes of the right and left peaks.

Wang et al. [7] also fit a bimodal normal mixture to a distribution, but with equal variances, and define their bimodality index as $\mathrm{BI} = \frac{|\mu_1 - \mu_2|}{\sigma}\sqrt{\pi(1-\pi)}$, where $\mu_{1,2}$ are the means of the normal distributions, σ is their common variance, and π and $1-\pi$ are the mixing weights of the normal distributions.

Bayramoglu [8] defines a bimodality degree $\delta = min\left\{\frac{f(x_1)}{f(x_2)}, \frac{f(x_1)}{f(x_3)}\right\}$, where f is the probabilty density function of the distribution, x_1 is a local minimum of f, and x_2, x_3 are local maxima of f. This bimodality degree takes values from 0 to 1, with 1 indicating unimodality and lower values indicating a more pronounced antimode. In practice, finding local minima and maxima from samples of a distribution requires fitting a model $f(x)$ over the observed samples and solving $f'(x) = 0$.

Jammalamadaka et al. [9] propose a Bayesian test for the number of modes in a Gaussian mixture, and compute a Bayes factor for a mixture of two Gaussians $B_{10} = \frac{P(H_1|x)P(H_0)}{P(H_0|x)P(H_1)}$, where H_0 is the hypothesis that the mixture is unimodal and H_1 is the hypothesis that the mixture is bimodal. Higher values for this Bayes factor indicate bimodality, but a suitable threshold must be selected.

Chaudhuri and Agrawal [10] measure the bimodality of a distribution after applying a threshold k with the following function $W(k) = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n\sigma^2}$, where n_1 is the number of samples with values $\leq k$, σ_1^2 is their variance, n_2, σ_2^2 are the analogs for the samples with values $> k$ and n, σ^2 for all the samples. Low values for this function indicate bimodality.

Van Der Eijk [11] presents a way to measure agreement in a discrete distribution with a finite range of possible values. The formula for this agreement is $A = U\left(1 - \frac{S-1}{K-1}\right)$, where U is a measure of unimodality, S is the number of distinct values present in the sample, and K is the total number of possible values for the distribution. This formula takes values between -1 and 1, with 1 indicating unimodality and -1 indicating multimodality.

Wilcock [12] defines a bimodality parameter based on the amplitudes and probabilities of the modes. The formula is $B = \sqrt{\frac{A_r}{A_l}}(P_l + P_r)$, where A_l, A_r are the amplitudes of the left and right mode and P_l, P_r are the probabilities. High values for B indicate bimodality.

Smith et al. [13] propose an alternative bimodality index with the formula $B^* = |\phi_2 - \phi_1|\frac{P_2}{P_1}$, where $\phi_{1,2}$ are the modal sizes in phi units, subscript 1 refers to the primary mode, and subscript 2 to the secondary mode. If the modes have equal amplitudes, then 1 refers to the right mode and 2 to the left one. The two indices are numerically similar in the range of $3 < A_r/A_l < 30$, which covers most of the bimodal distributions presented in [13], but the index presented in [13] has an asymptotic behavior of tending to zero as the separation of modes vanishes.

Contreras-Reyes [14] constructs an asymptotic test for bimodality for a bimodal skew-symmetric normal (BSSN) distribution from the Kullback–Leibler divergence. For the δ parameter of the BSSN distribution, he selects a threshold δ_0 and tests the hypothesis $H_0 : \delta \leq \delta_0$ versus the alternative $H_1 : \delta > \delta_0$. H_0 indicates bimodality and H_1 unimodality.

Highly specialized tests are much more accurate than the more general ones; however, they should only be used for the specific distributions they are specialized for. For example, Vichitlimaporn et al. [15] present highly specialized bimodality criterions for the molecular weight distributions in copolymers. Hassanian-Moghaddam et al. [16] present a similar criterion for the bivariate distribution of chain length and chemical composition of copolymers. Vořišek [17] uses Cardan's discriminant to determine whether or not a cusp distribution is bimodal.

Sarle's bimodality coefficient [18] uses a formula based on skewness and kurtosis that takes values between 0 and 1, with the value of 1 corresponding only to the perfect bimodal distribution. Values closer to 1 usually indicate bimodality, but not necessarily. Hildebrand [19] shows that kurtosis can be used as a bimodality indicator for some distributions but says that using it uncritically is hazardous and proves that for some distributions it cannot be used as such. Even though this test is not reliable enough to prove that a certain distribution is bimodal, it is quite easy to compute, especially when using summed-area tables to quickly compute the values of skewness and kurtosis.

In this paper, we discuss the strengths and weaknesses of Sarle's bimodality coefficient and improve upon it by proposing generalized bimodality coefficients and combining several of them to create and fine-tune a composite bimodality coefficient that offers better results on the datasets we tested, while still being able to run in constant time on variable-sized datasets.

In Section 2 we present Sarle's bimodality coefficient in greater detail, the summed-area table technique and how we use it to compute the coefficient in constant time, a generalization of Sarle's bimodality coefficient, and how we combine multiple generalized bimodality coefficients to create a composite bimodality coefficient. In Section 3, we describe the testing methodology, and in Section 4 we present the numerical results. In Section 5, we discuss the benefits and risks of using the generalized bimodality coefficients and the composite bimodality coefficient. In Section 6, we conclude that the composite bimodality coefficient can be a step-up from Sarle's bimodality coefficient with the correct ratios.

2. Materials and Methods

2.1. Sarle's Bimodality Coefficient (BC)

Sarle's bimodality coefficient is a straightforward method to test for bimodality. It is calculated according to the formula $\text{BC} = \frac{skewness^2+1}{kurtosis}$, where skewness is the third standardized moment of the distribution and kurtosis is the fourth one. Pearson, in an editorial note to Shohat [20], proved that skewness^2—kurtosis is greater than or equal to 1, with equality only for the two-point Bernoulli distribution [21]. Because the two-point Bernoulli distribution is the "most" bimodal distribution, the BC tends to take values closer to 1 for "more" bimodal distributions and closer to 0 for "less" bimodal ones.

However, this test is not perfect. High values do not always indicate bimodality and low values do not always indicate non-bimodality. Figure 1 showcases this caveat with the

unimodal beta distribution with probability density function (PDF) $f(x|\alpha = 25, \beta = 1) = \frac{\Gamma(26)}{\Gamma(25)\Gamma(1)} x^{24}, x \in [0,1]$, and the bimodal binormal distribution with PDF $g(x|\mu_1 = 0, \mu_2 = 1, \sigma_1 = \sigma_2 = 0.25, \omega = 0.5) = \frac{2}{\sqrt{2\pi}}\left(e^{-8x^2} + e^{-8(x-1)^2}\right), x \in [0,1]$, where ω and $1-\omega$ are the mixing weights. In practice, a common threshold for this coefficient is 0.555. Distributions with values greater than this are considered bimodal, and the ones with values lower than this are not. Knapp [18] compares the performance of this BC with the performance of other tests and shows a unimodal distribution for which the BC takes a very high value (0.78) compared to the actual bimodal distributions (0.56–0.66).

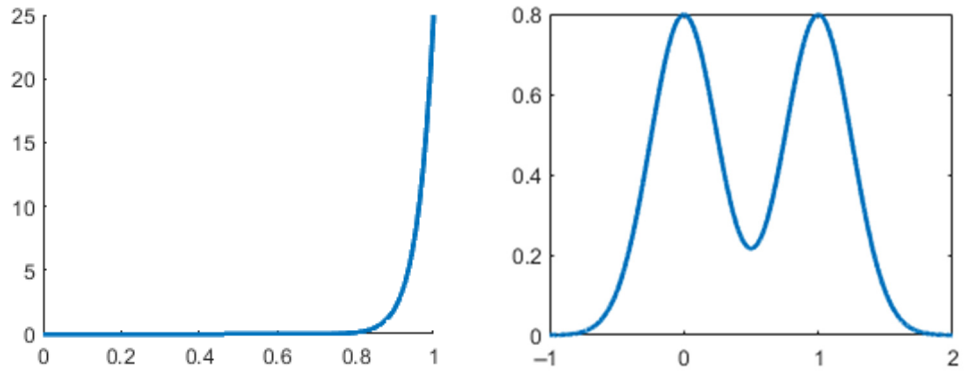

Figure 1. Unimodal distribution (**left**) with high BC (0.8), and bimodal distribution (**right**) with low BC (0.58).

2.2. Fast Generation of Moments

The reason we are interested in bimodality tests that can be run in constant time is the development of applications that are dealing with large amounts of data and must partition or classify portions of data as fast as possible. Such an application may be one that binarizes a sample from an unknown n-dimensional signal by identifying, for each sample point, the best window for which to apply a specific threshold-based binarization algorithm (e.g., Otsu's [22] method).

Sarle's bimodality coefficient requires the computation of standardized moments 3 and 4 (skewness and kurtosis), which in turn require the computation of raw moment 1 and central moment 2 (mean and variance). However, the brute-force approach for these computations is of a linear order ($\Theta(n)$), and the number of windows in the sample space is of a squared order ($\Theta(n^2)$), resulting in the complexity of a cubed order ($\Theta(n^3)$). Luckily, a constant order ($\Theta(1)$) approach for the in-window computations exists and is presented in the following subsections.

2.2.1. Summed-Area Table Technique

The mean of a window can be computed in constant time using a technique called the summed-area table [23]. The idea behind this is that by pre-computing the sums of the samples, the sum of a certain window can be computed as a function of the sums of the extremity points of the window. For a one-dimensional signal, the sum of a window would simply be the sum up to the point at the end of the window, minus the sum up to the point before the beginning of the window. Figure 2 shows how to compute the sum of a window for a two-dimensional signal. The sum of the window of interest (red) is equal to the sum of the orange window, whose value is stored in the summed-area table at the coordinates of the bottom right corner, plus the sum of the purple window (top left corner), minus the sum of the blue window (top right corner), minus the sum of the green window (bottom left corner). This technique can be applied to an n-dimensional signal for any natural number n.

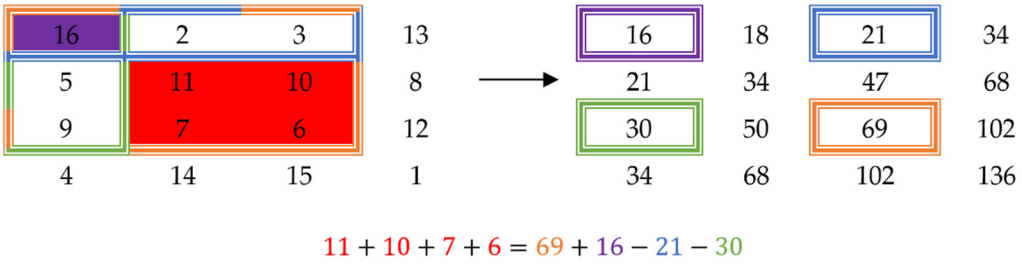

Figure 2. Example of a summed-area table (**right**) obtained from a 2D matrix (**left**).

2.2.2. Generation of Moments Using Sums of Power

However, these sums divided by the number of samples in the window only yield the raw moments. Luckily, central, and standardized moments can be expressed as a function of raw moments of rank lower or equal to theirs as follows [24]:

$$\mathrm{M}_n = \sum_{j=0}^{n} C_n^j (-1)^{n-j} \mu'_j \mu^{n-j}, \tag{1}$$

where μ_n is the central moment of order n, and μ'_j is the raw moment of order j and is the mean value of the distribution. Denoting $\widetilde{\mu}_n = \frac{\mu_n}{\sigma^n}$ as the standardized moment of order n, with σ representing the standard deviation of the distribution, we have the formula for Sarle's BC as:

$$\mathrm{BC} = \frac{\widetilde{\mu}_3^2 + 1}{\widetilde{\mu}_4} = \frac{\sigma^4 + \sigma(2\mu^3 - 3\mu'_2\mu + \mu'_3)}{-3\mu^4 + 6\mu'_2\mu^2 - 4\mu'_3\mu + \mu'_4} \tag{2}$$

where $\sigma = \sqrt{-\mu^2 + \mu'_2}$. Applying the technique described in the previous section on the matrix in Figure 2, we can obtain the summed-area tables for higher powers, as illustrated in Figure 3.

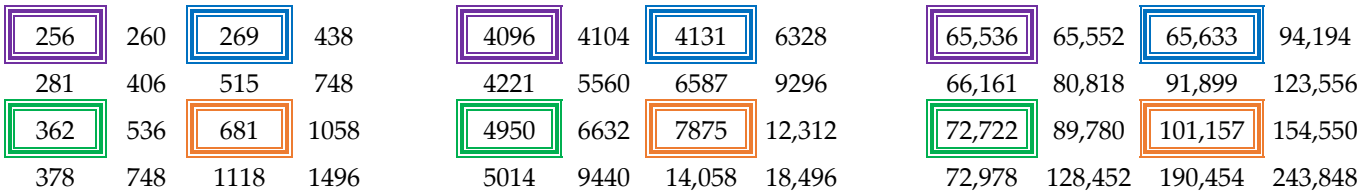

Figure 3. Summed-area tables for higher powers: (**left**) 2nd, (**middle**) 3rd, (**right**) 4th.

2.3. Generalized Bimodality Coefficient (GBC)

Theorem 1. *Let X be a distribution, μ its central moments, and $\widetilde{\mu}$ its standardized moments. Let*

$$GBC_k = \begin{cases} 0, \text{ for distributions concentrated on one point } \left(\text{to avoid } \frac{0}{0}\right) \\ \frac{\mu_{2k+1}^2 + \mu_2^{2k+1}}{\mu_{2k+2}\mu_{2k}} = \frac{\widetilde{\mu}_{2k+1}^2 + 1}{\widetilde{\mu}_{2k+2}\widetilde{\mu}_{2k}}, \text{ otherwise} \end{cases} \tag{3}$$

Then, the inequality $GBC_k \leq 1$ holds for any distribution, with equality only for distributions concentrated on two points.

Proof. Let

$$z_k = \widetilde{\mu}_{2k+2}\widetilde{\mu}_{2k} - \widetilde{\mu}_{2k+1}^2 - 1.$$
$$z_k \geq 0 \Leftrightarrow GBC_k \leq 1$$

Since z_k is concave and all finite distributions with zero mean are (convex) mixtures of distributions concentrated on at most two points [25], the theorem will be proven if we can show that $z_k = 0$ for all two-point distributions.

A two-point distribution takes the form $X = \begin{pmatrix} x & y \\ p(x) & p(y) \end{pmatrix}$. Then $\mu_k = \frac{a^k b - a b^k}{b-a} = \frac{-ab}{b-a}\left(b^{k-1} - a^{k-1}\right)$, where $a = x - E(X), b = y - E(X)$. We can substitute this value in the equation we want to prove and find that:

$$\begin{aligned} z_k &= \frac{(-ab)^{-k}}{b-a}\left(b^{2k+1} - a^{2k+1}\right)\frac{(-ab)^{-k}}{b-a}\left(b^{2k-1} - a^{2k-1}\right) - \left(\frac{(-ab)^{1-\frac{2k+1}{2}}}{b-a}\left(b^{2k} - a^{2k}\right)\right)^2 - 1 \\ &= \frac{(-ab)^{1-2k}}{(b-a)^2}\left(\left(b^{2k+1} - a^{2k+1}\right)\left(b^{2k-1} - a^{2k-1}\right) - \left(b^{2k} - a^{2k}\right)^2\right) - 1 \\ &= \frac{(-ab)^{1-2k}}{(b-a)^2}\left(b^{4k} - a^{2k+1}b^{2k-1} - a^{2k-1}b^{2k+1} + a^{4k} - b^{4k} + 2a^{2k}b^{2k} - a^{4k}\right) - 1 \\ &= \frac{(-ab)^{1-2k}}{(b-a)^2}\left(-a^{2k+1}b^{2k+1} - a^{2k-1}b^{2k-1} + 2a^{2k}b^{2k} - (b-a)^2(-ab)^{2k-1}\right) \\ &= \frac{(-ab)^{1-2k}}{(b-a)^2}\left(-a^{2k+1}b^{2k+1} - a^{2k-1}b^{2k-1} + 2a^{2k}b^{2k} + a^{2k-1}b^{2k+1} - 2a^{2k}b^{2k} + a^{2k+1}b^{2k-1}\right) \\ &= 0, \end{aligned} \tag{4}$$

which is equivalent to $GBC_k = 1$ for all distributions concentrated on at most two points. □

2.4. Composite Bimodality Coefficient (CBC)

As previously shown, comparing a bimodality coefficient's value to a set threshold is a weak test on its own. To have a strong test, we consider different formulas to mix our GBCs and obtain a CBC (see Appendix A). After running different tests with said formulas, we define our *CBC* according to a mixed power law, in the form seen in Equation (5):

$$GBC_n = \prod_{k=1}^{n} GBC_k^{p_k}. \tag{5}$$

This formula ensures that each *GBC* contributes to the test, and their respective powers represent the weight with which they do.

Since the problem at hand deals with a variety of possible distributions, potential mixtures of them, and imperfect data pertaining to said distributions, finding a set of powers that can be considered optimal in terms of general approximation is a difficult challenge. With that in mind, we found no way to analytically determine the best set of powers and thus resorted to the empirical method. We decided that a good way to establish the veracity of the CBC is by testing its performance in an image binarization application.

There are multiple ways to go about finding a set of powers that can fit a generic CBC. One that we tried initially was doing a simple grid-search using an image dataset, by starting around a series of points representing the powers of each GBC and shifting around them to increase the accuracy of our results. However, this method proved to be both lengthy and inefficient, as determining the accuracy of a set of powers in the image dataset took a while to compute (~10–20 min); given the thousands of possible combinations, the time to test them would increase to days. Subsequently, another factor that was an issue was that each new image that was added would change the order of the sets around, with the same one almost never being at the top again, whilst the accuracy of all sets was very similar to one another and within the margin of error. The third issue with this approach was that we do not know ahead of time whether a distribution is bimodal or not.

Thus, we needed a different approach that we could run on a dataset of known distributions in a fast time, obtain a set of powers on said dataset, and then run the CBC obtained by combining those powers against the original image dataset. For this purpose, we chose to train and use a very simple neural network, as the process for finding the desired values using one is very similar to that of a grid-search.

Our approach consisted of training a neural network to predict a potential set of powers that can be used in our composite coefficient. Since we were interested in obtaining a simple formula that can be applied in a fast, straightforward manner to compute values in real-time, we aimed to build a very rudimentary network configuration, more specifically

that of a perceptron [26]. To determine a proper configuration for our network, we also considered the way the value of a node is computed, and the one-to-one relationship between the set of values that we wanted to find and our bimodality coefficients. Thus, we could build a straightforward network where our only output was represented by our CBC, and the inputs, associated with our GBCs, were directly connected to it.

Theorem 2. *Let $I_n = \{x_1, \ldots, x_n\}$ represent the inputs of our network, $W_n = \{w_1, \ldots, w_n\}$ the associated weights, and y the output of our network to which all the inputs are connected directly, with f(x) as its activation function and b as its activation bias. If*

$$y = f\left(\sum_{k=1}^{n} w_k x_k\right) + b, \exists g(x), I_n = \{g(GBC_1), \ldots, g(GBC_n)\} \Rightarrow y = \prod_{k=1}^{n} GBC_k^{w_k}. \quad (6)$$

Proof. Let

$$g(x) = ln(x), f(x) = e^x, b = 0 \Rightarrow y = \prod_{j=1}^{n} GBC_j^{w_j}$$

Something to be noted is that we chose to use only the first 3 GBCs in our approach, due to two reasons. First, as previously discussed, higher-order GBCs tend to be more error-prone to noise; thus, using them with a variety of distributions can potentially lead to poorer results. Secondly, computing higher-order GBCs requires a lot of resources to guarantee the precision necessary to store the values which can also take a lot of time, and is not the intended use for our coefficient that is meant to be fast and reliable. □

3. Preparing the Dataset for Training and Testing

For our training dataset, we chose to generate it from a variety of distributions, for two simple reasons. First, since we know the distributions involved in generating the data, we know ahead of time if a sample should follow a one or two-point focused distribution. Second, due to controlling the distributions involved in generating our samples, we can limit some of the caveats that were previously mentioned that negatively impact our GBC. For this purpose, we chose to generate random data points using twelve different distributions (see Appendix B), respectively a mixture of them, aimed at a variety of potential unimodal and bimodal results, generating a total of approximately 25,000 distributions. In a similar fashion, to vary the potential values of the GBCs pertaining to each distribution, we generated sets with a varying number of points, between 500 and 10,000. Because we wanted to test the veracity of the CBC in a binarization application, all values were generated in the interval $[0, 1]$, for an easier overall approach; however, this does not alter the result, as any distribution could be shifted and scaled to any other interval of choice, and the standardized central moments would not change.

After we generated the data, we computed the following for each set of points pertaining to a distribution: the moments, respectively the GBC from them, as well as a value that we would like our CBC to tend to for that set of points. We denoted these values as CBC′ and they will be used as the output of the network during the training step. Since we know the original distributions that generated our sets of points, we can compute CBC′ in two ways, depending on whether or not the original distribution was unimodal or bimodal:

$$CBC'(f(x)) = \begin{cases} CBC'_u(f(x)), if\ f(x)\ is\ unimodal \\ CBC'_b(f(x)), if\ f(x)\ is\ bimodal \end{cases} \quad (7)$$

$$CBC'_u(f(x)) = \frac{min(f(x))}{max(f(x))} |argmax(f(x)) - argmin(f(x))| \quad (8)$$

For a bimodal distribution $f(x)$, we denote with $x_{m1}, x_{m2}, x_{m1} \leq x_{m2}$ the position of the two modes of the distribution.

$$d_m = (x_{m2} - x_{m1})^{\left(\frac{1}{n_f(x_{m2}-x_{m1})}\right)} \tag{9}$$

$$CBC'_b(f(x)) = \begin{cases} d_m \left(\frac{max(f(x))-min(f(x))}{max(f(x))}\right)^{d_m}, if\ f(x_{m1}) = f(x_{m2}) \\ d_m \left(\frac{f(x_{m1})}{f(x_{m2})}\right)^{1-d_m}, if\ f(x_{m1}) < f(x_{m2}) \\ d_m \left(\frac{f(x_{m2})}{f(x_{m1})}\right)^{1-d_m},\ otherwise \end{cases} \tag{10}$$

Theorem 3. *Let* $Codomain(CBC') = \{CBC'(f(x)) | \forall x \in [0,1], \forall f(x) : [0,1] \to [0,+\infty)\}$. *Then* $Codomain(CBC') \subseteq [0,1]$.

Proof.

$$\begin{gathered} \frac{min(f(x))}{max(f(x))} \in [0,1], |argmax(f(x)) - argmin(f(x))| \in [0,1] \\ \Rightarrow Codomain(CBC'_u) \subseteq [0,1] \end{gathered} \tag{11}$$

$$(x_{m2} - x_{m1}) \in (0,1] \Rightarrow (x_{m2} - x_{m1})^u \in [0,1], \forall u \in R \Rightarrow d_m \in (0,1] \tag{12}$$

$$\text{If } f(x_{m1}) = f(x_{m2}),\ \frac{max(f(x)) - min(f(x))}{max(f(x))} \in (0,1] \Rightarrow Codomain\left(CBC'_b\right) \subseteq (0,1] \tag{13}$$

$$\frac{f(x_{m1})}{f(x_{m2})} \in [0,1), \text{if } f(x_{m1}) < f(x_{m2}), \Rightarrow Codomain\left(CBC'_b\right) \subseteq [0,1) \tag{14}$$

The last case for CBC'_b is similar to (14). From (11)–(14) we can see that $Codomain(CBC') \subseteq [0,1]$. □

In the above equations, $d_m \in (0,1]$ represents a correction factor based on the distance between the two modes of the distribution, such that two close modes do not give a high value for the CBC′. Similarly, n_f is a normalization factor, typically chosen in the interval $[100,\ 300]$ to ensure we always raise to subunit powers, and that d_m can raise exponentially closer to 1 when the modes start spreading apart. The reason for using a correction factor is because a set of points generated by a bimodal distribution with two close modes would be almost indistinguishable from one generated by a unimodal distribution with a similar shape and with its mode lying somewhere between the two modes of the bimodal one. In a similar fashion, we used $1 - d_m$ when the value of the PDF differs between the two nodes so that modes that are very spread vertically, but not horizontally, do not result in high CBC′ values.

3.1. Training the Network

Based on the previously computed values, we could train our network using the natural logarithms of GBCs as inputs and CBC′ as the output. However, before training our network, we also considered a second potential remap of our inputs, before applying our logarithmic function on the GBCs, one based on polynomial fitting (see Appendix C). This approach did lead to increased overall accuracy; thus, it was kept both for later training and testing purposes.

Since the training of the network was based on gradient descent, and we used a very simple network configuration, we had no certainty that our search would hit a global optimum, and, even if it did, we could not be sure of the general applicability of such an optimum. Consequently, since we did not know which set of powers might behave best in practice, we needed to obtain multiple sets and compare them. Thus, we trained different networks, with different constraints, such as using a different dataset when training, or constraining the weights, either to a set interval or simply making them strictly positive, to obtain our possible CBC parameters.

3.2. Testing the CBC

Tests were conducted against an image dataset generated by combining the following datasets: the DIBCO datasets [27–35], the PHIBD 2012 [36] dataset, the Nabuco dataset [37], and the Rahul Sharma dataset [38], the combination of which will be further denoted simply as the Images dataset. On this dataset, we ran a binarization algorithm starting from multiple possible choices, with various window sizes. We decided to settle on one based on Otsu's method with a fixed 8×8 window size centered on each pixel. There were also separate tests conducted against a synthetic dataset obtained from random data pertaining to multiple distributions and mixtures of them.

4. Numerical Results

4.1. Accuracy of GBC

To quantify the accuracy, we used the F-measure (*FM*) defined as follows:

$$FM = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}, \tag{15}$$

where *TP* stands for true positives, *FP* for false positives, and *FN* for false negatives. In Table 1, we can see the different F-measures associated with each GBC on both datasets. One thing to note is that higher-order GBCs tend to decrease in accuracy when trying to identify whether a distribution is bimodal or not, a point which will be discussed later as well as why higher-order GBCs might still be valuable, given certain scenarios.

Table 1. F-measures of the first three GBCs.

GBC	F-Measure (%)	Dataset
GBC_1	76.829	Images
	64.193	Synthetic
GBC_2	76.714	Images
	47.688	Synthetic
GBC_3	74.934	Images
	46.546	Synthetic

Figure 4 shows the class separation corresponding to each GBC, showing how higher-order GBCs tend to have their values more spread out, thus making it harder to pick a threshold. The peaks close to 1, for the unimodal cases, in all GBCs, are indicative of distributions with high skewness and low kurtosis, for the synthetic dataset.

Table 2 shows the behaviors of the GBCs on a binormal distribution with varying means and variances. All of them decrease in values as the means get closer to each other and increase in value as the variances decrease, which indicates that they encourage bimodal distributions with more distinct peaks, with the higher-order GBCs being increasingly more sensitive. We compensated for the lower values of the higher-order GBCs by raising the powers of the lower-order GBCs.

Table 3 shows the behaviors of the GBCs on a binormal distribution with fixed means and varying variances. The table is symmetrical, which indicates that they are insensitive to flipping transformations on the dataset.

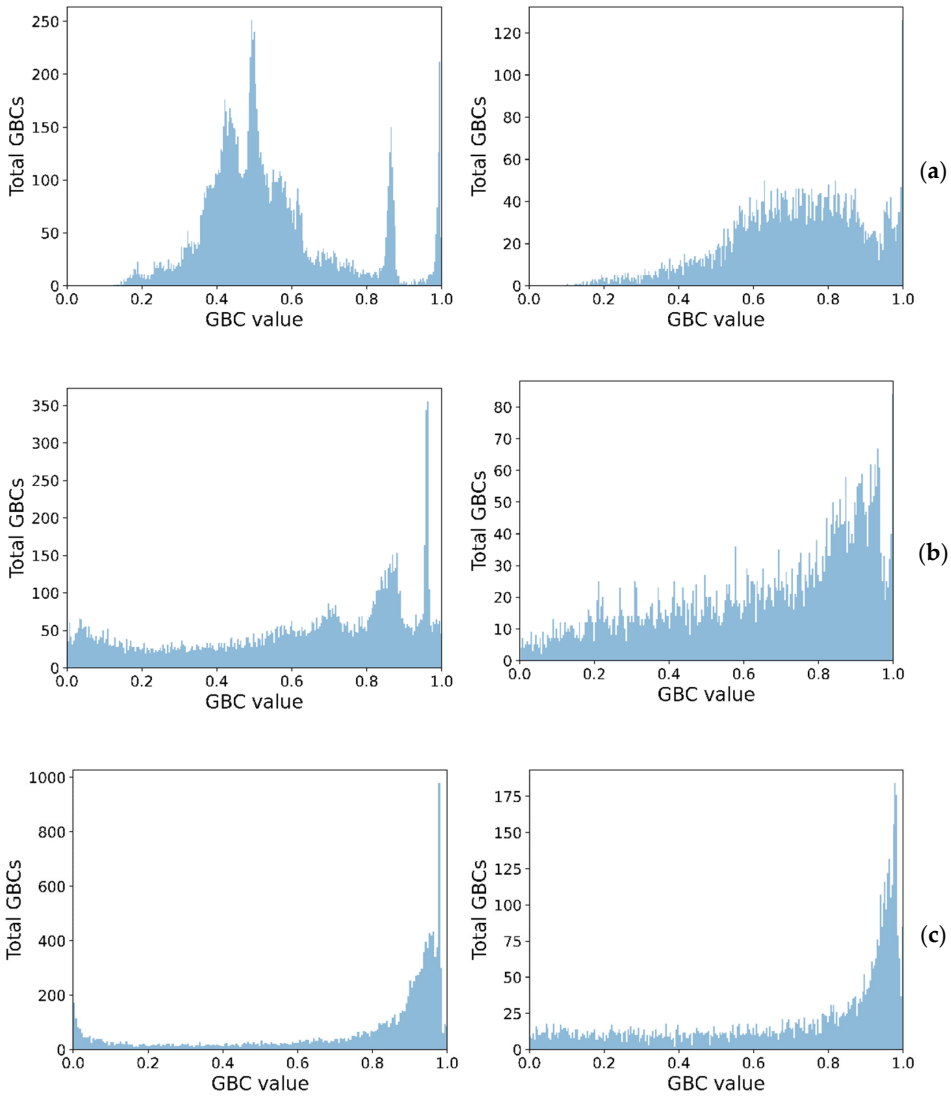

Figure 4. Spread of GBC values for the two classes (unimodal **left**, bimodal **right**) on the synthetic dataset (**a**) GBC_1, (**b**) GBC_1, and (**c**) GBC_3.

Table 2. Values for GBC_1, GBC_2, and GBC_3 for a binormal distribution with means μ_1, μ_2 and variances σ_1^2, σ_2^2.

μ_1, μ_2 \ σ_1, σ_2	**5, 5**	**2.5, 2.5**	**1, 1**
25, 75	0.8689, 0.5842, 0.3168	0.9621, 0.8586, 0.7136	0.9937, 0.9749, 0.9446
33, 66	0.7563, 0.3556, 0.1177	0.9185, 0.7186, 0.4872	0.9856, 0.9439, 0.8790
40, 60	0.5814, 0.1456, 0.0218	0.8141, 0.4613, 0.1965	0.9621, 0.8586, 0.7136

Table 3. Values for GBC_1, GBC_2, and GBC_3 for a binormal distribution with means $\mu_1 = 33$, $\mu_2 = 66$ and variances σ_1^2, σ_2^2.

σ_1 \ σ_2	**5**	**2.5**	**1**
5	0.7563, 0.3556, 0.1177	0.8309, 0.5287, 0.3728	0.8578, 0.6105, 0.5198
2.5	0.8309, 0.5287, 0.3728	0.9185, 0.7186, 0.4872	0.9509, 0.8234, 0.6625
1	0.8578, 0.6105, 0.5198	0.9509, 0.8234, 0.6625	0.9621, 0.8586, 0.7136

4.2. Accuracy of CBC

Table 4 shows some of the coefficients we ended up testing and their corresponding F-measures for both datasets. As we can see, just because a set of powers gives a good result for one dataset, does not necessarily mean it would do the same for the other, mostly due to the specific nature of the distributions that exist in each dataset. However, based on the overall performance, we would recommend the set of powers of either 3:2:1 or 3:0:1.

Table 4. Mean values for F-measures for different power ratios.

p_1	p_2	p_3	F-Measure (%)	Dataset
1	1	1	76.976	Images
			47.138	Synthetic
3.051	0	1	76.855	Images
			66.294	Synthetic
14.92	−1	1.806	76.841	Images
			66.227	Synthetic
16.5	−1	1.414	76.83	Images
			64.389	Synthetic
2.98	1.92	1	76.901	Images
			65.996	Synthetic
1.535	−1	1.064	74.687	Images
			65.251	Synthetic

5. Discussion

5.1. GBC Comparison with BC

As we could see from our results, higher order GBCs tend to be less accurate and less reliable as indices when identifying whether a distribution is bimodal or not. This is in part because of two reasons, the first being that higher order GBCs are more error-prone to noise, specifically because they require values that use higher powers for computation. The second thing to note is that it is harder to set a specific threshold at which to decide whether a distribution is bimodal or not, which could also be seen in our results.

However, this does not necessarily mean that using them in a composite coefficient is a bad thing. As we could see from the results pertaining to our CBC, it just means that using them individually can lead to worse results. However, for "well-behaved" distributions (low skewness, high kurtosis, and their respective higher-order moments), GBCs tend to have some benefits, when considering bimodal distributions:

- the higher the value of the two modes in a distribution, the higher the value of the GBC, with higher-order GBCs being more sensitive to this property;
- the more spread apart the two modes are, on the x-axis, the higher the value of the GBC, again, with higher-order GBCs being more sensitive;
- the closer the value of the two modes, on the y-axis, the higher the value of the GBC, where GBCs of higher-order behave better once again.

Something else to be noted is the fact that there is a huge jump when computing the F-measure for GBCs one and two on the Synthetic dataset, one of about 15; meanwhile, these values on the Images dataset tend to be very close together. This might be because the Synthetic dataset contains a lot of distributions with high skewness and low kurtosis, which, while they might make the dataset more generic, also have a negative influence on the results. Similarly, there is no guarantee that the set of synthetic distributions, or something close in nature to it, might exist in the real world. Thus, for future work, using sets that pertain to a more specific subset of distributions and mixtures of them, rather than such a large one, might be more beneficial.

5.2. CBC as a Bimodality Coefficient

As we could see from our results, the formula we used to compute our CBC gives better results, on both datasets we tested on, when compared to [18]. This does not mean that our coefficient is without issues, since distributions with high skewness and low kurtosis can still lead to misidentifying a unimodal distribution as a bimodal one.

A multitude of indices that were proposed [5–9] needs to make assumptions that the data they analyze fit a mixture of two normal distributions, which, often, is not necessarily the case. In this regard, from a generic point of view, our proposed coefficient can be more robust in some situations. A more significant advantage of our coefficient is its computational efficiency, especially when compared with the indices previously mentioned. Since we can compute the moments in a fast manner and we do not require to make assumptions or try to fit certain distributions to our data, we can determine in a fast and accurate manner whether some data fit a bimodal distribution or not. Table 5 shows the performance of our proposed coefficient, when using the powers 3:0:1, which compared with other indices, ran against a subsample of 4000 distributions of the Synthetic dataset. All tests were performed on an AMD Ryzen 9 5900HS processor. The times were averaged across 15 different runs of the program to address potential biases, whilst also looking for the best potential threshold for all indices, in order to maximize their F-measure, with the latter being a potential reason why a multitude of the indices shares the same F-measure.

Table 5. Performance of different indices and methods on the Synthetic dataset.

Index/Method	Runtime (s)	F-Measure (%)
Ashman et al. [5]	37.179	44.294
Zhang et al. [6]	37.286	44.77
Wang et al. [7]	36.986	44.77
Chaudhuri et al. [10]	14.338	44.77
Eijk [11]	2.881	46.976
Wilcock et al. [12]	1.772	44.113
CBC (3:0:1)	0.272	54.104

Similarly, the computational efficiency of our coefficient can be seen when comparing it with the multitude of bimodality/multimodality tests proposed by [5–10]. Since most of these work directly on the histogram pertaining to a data sample, they tend to be slow, whilst also working with an uncertain stop condition, due to trying to find the best fit for said data. The indices proposed by the authors of [11,12] also work on the histogram, since they need to make approximations regarding the modes of the distribution. This can be both slow and can lead to poor results for bimodal distributions which have modes that are close together. Whilst our coefficient is not impervious to the closeness of the two modes, the fact that it is easier and faster to compute makes it more desirable.

Due to the fact that the image binarization test scenario is extremely complex, and running it on the Images dataset will result in an astonishing number of calls to any benchmarked bimodality coefficient (we counted about 300 million calls per sample from the Images dataset), it is unfeasible to run on the aforementioned dataset other coefficients that may be employed in Table 5 besides CBC, simply because they cannot be computed in constant time (O(1) complexity).

6. Conclusions

Determining whether some data follow a unimodal or bimodal distribution is an important step in figuring out a proper approach, in order to deal with said data, in a variety of situations. Thus, a proper way to ascertain this property is desired. A bimodality coefficient is a good starting point if it is robust.

In our case, the CBC performs better than any GBC on its own, for most of the composition ratios we tested, on average obtaining F-measure increases of 1.5–2%. However, finding an optimal ratio is no easy task, and even then, we have no certainty on how a CBC

based on such a ratio might behave in more generic scenarios. For this reason, we also recommend two possible ratios to be considered: 3:2:1 and 3:0:1.

A certain downside of using the bimodality coefficient on its own is the fact that we do not know how data that follow a multimodal distribution with more than two modes behave. Thus, a multimodality coefficient could be considered for a better approach, and in future works, we could investigate ways to compute one.

Author Contributions: Conceptualization, N.T., M.-L.V. and C.-A.B.; Data curation, N.T., M.-L.V. and C.-A.B.; Formal analysis, N.T., M.-L.V. and C.-A.B.; Investigation, N.T., M.-L.V. and C.-A.B.; Methodology, N.T., M.-L.V. and C.-A.B.; Project administration, C.-A.B.; Resources, C.-A.B.; Software, N.T., M.-L.V. and C.-A.B.; Supervision, C.-A.B.; Validation, N.T., M.-L.V. and C.-A.B.; Visualization, N.T., M.-L.V. and C.-A.B.; Writing—original draft, N.T., M.-L.V. and C.-A.B.; Writing—review and editing, N.T., M.-L.V. and C.-A.B. All authors have read and agreed to the published version of the manuscript.

Funding: The APC has been supported by Politehnica University of Bucharest, through the PubArt program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

It is unreasonable to consider that a CBC comprised of a product of GBCs would yield the best results in practice. Thus, various other formulas were also considered as potential approaches for the problem at hand; however, none seemed to give many more beneficial results when considering both the results on the synthetic dataset of generated distributions or the image binarization dataset. Moreover, even when they did, it was very situational and only for a subset of the entire dataset. The F-measures for some of these formulas can be seen in Table A1, once again considering just the first three GBCs for practical reasons. These are not all the formulas we tested, but just some of those that we considered being more relevant in our case. The formulas are also presented without additional coefficients attached to our GBCs for simplicity. In practice, we tested both GBCs raised to different powers or multiplied with different coefficients.

Table A1. Results for different formulas we experimented with.

General Formula	Tested Formula	F-Measure (%)	Dataset
$\frac{1}{n}\sum_{k=1}^{n} p_k GBC_k$	$\frac{1}{3}(GBC_1 + GBC_2 + GBC_3)$	76.953	Images
		47.845	Synthetic
$\prod_{k=1}^{n} GBC_k^{p_k}$	$GBC_1 GBC_2 GBC_3$	76.976	Images
		47.138	Synthetic
$\frac{n}{\sum_{k=1}^{n} \frac{1}{GBC_k^{p_k}}}$	$\frac{3}{\frac{1}{GBC_1}+\frac{1}{GBC_2}+\frac{1}{GBC_3}}$	65.513	Images
		44.366	Synthetic
$\sqrt[n]{\prod_{k=1}^{n} GBC_k}$	$\sqrt[3]{GBC_1 GBC_2 GBC_3}$	58.98	Images
		64.69	Synthetic
$\sqrt{\sum_{k=1}^{n} GBC_k^2}$	$\sqrt{GBC_1^2 + GBC_2^2 + GBC_3^2}$	58.98	Images
		62.788	Synthetic

Appendix B

In order to generate a dataset both for training and testing our network, we chose to combine the following distributions: Beta [39,40], Burr [41,42], Exponential [43–45],

Frechet [46–48], Gumbel [49,50], Laplace [51,52], Logistic [53,54], Log-Logistic [55–57], Log-Normal [58–60], Normal [61,62], Pareto [63,64], and Weibull [65,66]. Obviously, there are a lot more distributions that we could have used, but we considered these to be representative enough for our purposes. As can be seen, we chose to use only continuous distributions, as discrete ones would lead to similar results, with no additional benefits when generating the data itself.

Each distribution has a different set of parameters associated with it, that determines its shape. For non-mixed distributions, we simply chose these parameters randomly and generated the distribution and the random set of points based on it. For mixtures of distribution, in a similar fashion, we chose the two sets of parameters randomly, whilst also picking a scaling coefficient for each distribution that determined how much it contributed to the result. The sum of these scaling coefficients was one. Table A2 shows the set of these distributions as well as the parameters associated with each of them, and the values we considered for them, with $x \in [0, 1]$. For some of these distributions, we skewed the selection of the parameters to certain intervals to obtain more representative distributions, with the probability of selecting the skewed interval for the parameters being denoted with p.

Table A2. Distributions used in creating the Synthetic dataset.

Distribution	PDF	Parameters
Beta	$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$	$\alpha, \beta \in (0,1), p = 0.3$ $\alpha, \beta \in (1,10], p = 0.7$
Burr	$f(x;c,k) = ck\frac{x^{c-1}}{(1+x^c)^{k+1}}$	$c \in (1,10]$ $k \in (0,10]$
Exponential	$f(x;\lambda) = \lambda e^{-\lambda x}$	$\lambda \in (0,0.2]$
Frechet	$f(x;\alpha,\mu,\sigma) = \frac{\alpha}{\sigma}\left(\frac{x-\mu}{\sigma}\right)^{-1-\alpha} e^{-(\frac{x-\mu}{\sigma})^{-\alpha}}$	$\alpha \in (0,10]$ $\mu \in [0,1]$ $\sigma \in (0,0.2]$
Gumbel	$f(x;\mu,\sigma) = \frac{1}{\sigma}e^{-\frac{x-\mu}{\sigma}+e^{-z}}$	$\mu \in [0,1]$ $\sigma \in (0,0.2]$
Laplace	$f(x;\mu,\sigma) = \frac{1}{2\sigma}e^{-\frac{\lvert x-\mu\rvert}{\sigma}}$	$\mu \in [0,1]$ $\sigma \in (0,0.2]$
Logistic	$f(x;\mu,\sigma) = \frac{e^{\frac{x-\mu}{\sigma}}}{\sigma\left(1+e^{-(\frac{x-\mu}{\sigma})}\right)^2}$	$\mu \in [0,1]$ $\sigma \in (0,0.2]$
Log-Logistic	$f(x;\alpha,\sigma) = \frac{\left(\frac{\sigma}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\sigma-1}}{\left(1+\left(\frac{x}{\alpha}\right)^{\sigma}\right)^2}$	$\alpha \in (0,10]$ $\sigma \in (0,0.2]$
Log-Normal	$f(x;\alpha,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}}e^{(\frac{(lnx-\alpha)^2}{2\sigma^2})}$	$\alpha \in (0,100]$ $\sigma \in (0,0.2]$
Normal	$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$	$\mu \in [0,1]$ $\sigma \in (0,0.2]$
Pareto	$f(x;\alpha,\mu) = \frac{\alpha\mu^\alpha}{x^{\alpha+1}}$	$\mu \in [0,1]$ $\alpha \in (0,100]$
Weibull	$f(x;\alpha,\sigma) = \frac{\sigma}{\alpha}\left(\frac{\pi}{\alpha}\right)^{\sigma-1}e^{-(\frac{x}{\alpha})^\sigma}$	$\alpha \in (0,100]$ $\sigma \in (0,0.2]$

Appendix C

Something we also considered was using a function that could remap the value of our GBCs to better fit what we considered the "ideal" coefficient, CBC', as described in Section 3. For these, we considered using a polynomial approximation. We tried polynomials of different degrees; however, polynomials of degrees higher than 3 would give values outside of the interval $[0, 1]$, for some GBCs in $[0, 1]$. Thus, even though they were fitting the data better at some points, since they were not robust for our purpose they

were discarded and only degree 3 polynomials were considered. The formulas for the three polynomials used can be seen in Equations (A1)–(A3), where $f_i(x)$ corresponds to GBC_i.

$$f_1(x) = -2.81x^3 + 5.91x^2 - 2.77x + 0.42 \quad \text{(A1)}$$

$$f_2(x) = 2.97x^3 - 5.19x^2 + 2.56x \quad \text{(A2)}$$

$$f_3(x) = 1.3x^3 - 2.97x^2 + 1.68x + 0.17 \quad \text{(A3)}$$

Applying this conversion on the GBCs yielded better overall results, obtaining an F-measure of 65.025 on the 1:1:1 ratio, which was much higher than the initial 47.138, on the Synthetic dataset alone.

References

1. Hartigan, J.A.; Hartigan, P.M. The Dip Test of Unimodality. *Ann. Stat.* **1985**, *13*, 70–84. [CrossRef]
2. Silverman, B.W. Using Kernel Density Estimates to Investigate Multimodality. *J. R. Stat.* **1981**, *43*, 97–99. [CrossRef]
3. Muller, D.W.; Sawitzki, G. Excess Mass Estimates and Tests for Multimodality. *J. Am. Stat. Assoc.* **1991**, *86*, 738–746. [CrossRef]
4. Rozál, G.P.M.; Hartigan, J.A. The MAP Test for Multimodality. *J. Classif.* **1994**, *11*, 5–36. [CrossRef]
5. Ashman, K.M.; Bird, C.M.; Zepf, S.E. Detecting Bimodality in Astronomical Datasets. *Astron. J.* **1994**, *108*, 2348–2361. [CrossRef]
6. Zhang, C.; Mapes, B.; Soden, B. Bimodality in Tropical Water Vapor. *Q. J. R. Meteorol. Soc.* **2003**, *129*, 2847–2866. [CrossRef]
7. Wang, J.; Wen, S.; Symmans, W.F.; Pusztai, L.; Coombes, K.R. The Bimodality Index: A Criterion for Discovering and Ranking Bimodal Signatures from Cancer Gene Expression Profiling Data. *Cancer. Inform.* **2009**, *7*, 199–216. [CrossRef]
8. Bayramoglu, I. Bivariate and multivariate distributions with bimodal marginals. *Commun. Stat. Theory Methods* **2019**, *49*, 361–384. [CrossRef]
9. Jammalamadaka, S.; Jin, Q. A Bayesian Test for the Number of Modes in a Gaussian Mixture. *Asian, J. Stat. Sci.* **2021**, *1*, 9–22.
10. Chaudhuri, D.; Agrawal, A. Split-and-Merge Procedure for Image Segmentation Using Bimodality Detection Approach. *Def. Sci. J.* **2010**, *60*, 290–301. [CrossRef]
11. Van Der Eijk, C. Measuring Agreement in Ordered Rating Scales. *Qual. Quant.* **2001**, *35*, 325–341. [CrossRef]
12. Wilcock, P.R. Critical Shear Stress of Natural Sediments. *J. Hydraul. Eng.* **1993**, *119*, 491–505. [CrossRef]
13. Smith, G.H.S.; Nicholas, A.P.; Ferguson, R.I. Measuring and Defining Bimodal Sediments: Problems and Implications. *Water Resour. Res.* **1997**, *33*, 1179–1185. [CrossRef]
14. Contreras-Reyes, J.E. An asymptotic test for bimodality using the Kullback-Leibler divergence. *Symmetry* **2020**, *12*, 1013. [CrossRef]
15. Vichitlimaporn, C.; Anantawaraskul, S.; Soares, J.B. Molecular Weight Distribution of Ethylene/1-Olefin Copolymers: Generalized Bimodality Criterion. *Macromol. Theory Simul.* **2017**, *26*, 1600060. [CrossRef]
16. Hassanian-Moghaddam, D.; Moattari, M.M.; Rezania, A.; Sharif, F.; Ahmadi, M. Analytical representation of bimodality in bivariate distribution of chain length and chemical composition of copolymers. *Chem. Eng. J.* **2022**, *431*, 133229. [CrossRef]
17. Voříšek, J. Estimation and bimodality testing in the cusp model. *Kybernetika* **2018**, *54*, 798–814. [CrossRef]
18. Knapp, T. Bimodality Revisited. *JMASM* **2007**, *6*, 8–20. [CrossRef]
19. Hildebrand, D.K. Kurtosis measures bimodality? *Am. Stat.* **1971**, *25*, 42–43.
20. Shohat, J. Inequalities for moments of frequency functions and for various statistical constants. *Biometrika* **1929**, *21*, 361–375. [CrossRef]
21. Pearson, K. Editorial note on the Limitation of Frequency Constants. *Biometrika* **1929**, *21*, 370–375.
22. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern. Syst.* **1979**, *9*, 62–66. [CrossRef]
23. Crow, F.C. Summed-Area Tables for Texture Mapping. In Proceedings of the 11th annual conference on Computer graphics and interactive techniques, Minneapolis, MN, USA, 23–27 July 1984; SIGGRAPH '84. Association for Computing Machinery: New York, NY, USA, 1984; pp. 207–212. [CrossRef]
24. Small, C.G. *Expansions and Asymptotics for Statistics*; Chapman and Hall/CRC: New York, NY, USA, 2010.
25. Rohatgi, V.K.; Székely, G.J. Sharp Inequalities between Skewness and Kurtosis. *Stat. Probab. Lett.* **1989**, *8*, 297–299. [CrossRef]
26. Rosenblatt, F. *The Perceptron—A Perceiving and Recognizing Automaton*; Techreport 85-460–1; Cornell Aeronautical Laboratory: Ithaca, NY, USA, 1957.
27. Gatos, B.; Ntirogiannis, K.; Pratikakis, I. ICDAR 2009 Document Image Binarization Contest (DIBCO 2009). In Proceedings of the 2009 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 26–29 July 2009; pp. 1375–1382. [CrossRef]
28. Pratikakis, I.; Gatos, B.; Ntirogiannis, K. H-DIBCO 2010—Handwritten Document Image Binarization Competition. In Proceedings of the 2010 12th International Conference on Frontiers in Handwriting Recognition, Kolkata, India, 16–18 November 2010; pp. 727–732.
29. Pratikakis, I.; Gatos, B.; Ntirogiannis, K. ICDAR 2011 Document Image Binarization Contest (DIBCO 2011). In Proceedings of the 2011 International Conference on Document Analysis and Recognition, Beijing, China, 18–21 September 2011; pp. 1506–1510.

30. Pratikakis, I.; Gatos, B.; Ntirogiannis, K. ICFHR 2012 Competition on Handwritten Document Image Binarization (H-DIBCO 2012). In Proceedings of the 2012 International Conference on Frontiers in Handwriting Recognition, Bari, Italy, 18–20 September 2012; pp. 817–822.
31. Pratikakis, I.; Gatos, B.; Ntirogiannis, K. ICDAR 2013 Document Image Binarization Contest (DIBCO 2013). In Proceedings of the 2013 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, 25–28 August 2013; pp. 1471–1476.
32. Ntirogiannis, K.; Gatos, B.; Pratikakis, I. ICFHR2014 Competition on Handwritten Document Image Binarization (H-DIBCO 2014). In Proceedings of the 2014 14th International Conference on Frontiers in Handwriting Recognition, Hersonissos, Greece, 1–4 September 2014; pp. 809–813.
33. Pratikakis, I.; Zagoris, K.; Barlas, G.; Gatos, B. ICFHR2016 Handwritten Document Image Binarization Contest (H-DIBCO 2016). In Proceedings of the 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), Shenzhen, China, 23–26 October 2016; pp. 619–623.
34. Pratikakis, I.; Zagoris, K.; Barlas, G.; Gatos, B. ICDAR2017 Competition on Document Image Binarization (DIBCO 2017). In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 9–15 November 2017; pp. 1395–1403.
35. Pratikakis, I.; Zagoris, K.; Karagiannis, X.; Tsochatzidis, L.; Mondal, T.; Marthot-Santaniello, I. ICDAR 2019 Competition on Document Image Binarization (DIBCO 2019). In Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, Australia, 20–25 September 2019; pp. 1547–1556.
36. Nafchi, H.Z.; Ayatollahi, S.M.; Moghaddam, R.F.; Cheriet, M. Persian Heritage Image Binarization Dataset (PHIBD 2012), ID: PHIBD 2012_1. Available online: http://tc11.cvc.uab.es/datasets/PHIBD2012_1 (accessed on 20 March 2022).
37. Lins, R.D. Two Decades of Document Processing in Latin America. *J. Univers. Comput. Sci.* **2011**, *17*, 151–161.
38. Singh, B.M.; Sharma, R.; Ghosh, D.; Mittal, A. Adaptive binarization of severely degraded and non-uniformly illuminated documents. *Int. J. Doc. Anal. Recognit.* **2014**, *17*, 393–412. [CrossRef]
39. Ma, Z.; Leijon, A. Beta Mixture Models and the Application to Image Classification. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009; pp. 2045–2048. [CrossRef]
40. Gupta, A.K.; Nadarajah, S. Beta Function and the Incomplete Beta Function. In *Handbook of Beta Distribution and Its Applications*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2004; pp. 3–32.
41. Nadarajah, S.; Pogány, T.; Saxena, R. On the Characteristic Function for Burr Distributions. *Statistics* **2011**, *46*, 419–428. [CrossRef]
42. Ahmad, K.; Jaheen, Z.; Mohammed, H. Finite Mixture of Burr Type XII Distribution and Its Reciprocal: Properties and Applications. *Stat. Pap.* **2011**, *52*, 835–845. [CrossRef]
43. Gupta, A.K.; Zeng, W.-B.; Wu, Y. Exponential Distribution. In *Probability and Statistical Models: Foundations for Problems in Reliability and Financial Mathematics*; Gupta, A.K., Zeng, W.-B., Wu, Y., Eds.; Birkhäuser: Boston, MA, USA, 2010; pp. 23–43. [CrossRef]
44. Elfessi, A.; Reineke, D. A Bayesian Look at Classical Estimation: The Exponential Distribution. *J. Stat. Educ.* **2001**, *9*. [CrossRef]
45. Rather, N.; Rather, T. New Generalizations of Exponential Distribution with Applications. *J. Probab. Stat.* **2017**, *2017*, 2106748. [CrossRef]
46. Nadarajah, S.; Pogány, T. On the Characteristic Functions for Extreme Value Distributions. *Extremes* **2013**, *16*, 27–38. [CrossRef]
47. Gusmão, F.; Ortega, E.; Cordeiro, G. The Generalized Inverse Weibull Distribution. *Stat. Pap.* **2011**, *52*, 591–619. [CrossRef]
48. Khan, M.S.; Pasha, G.; Pasha, A. Theoretical Analysis of Inverse Weibull Distribution. *WSEAS Trans. Math.* **2008**, *7*, 30–38.
49. Gumbel, E.J. Les valeurs extrêmes des distributions statistiques. *Ann. L'institut Henri Poincaré* **1935**, *5*, 115–158.
50. Evin, G.; Merleau, J.; Perreault, L. Two-Component Mixtures of Normal, Gamma, and Gumbel Distributions for Hydrological Applications. *Water Resour. Res.* **2011**, *47*. [CrossRef]
51. Ding, P.; Blitzstein, J. On the Gaussian Mixture Representation of the Laplace Distribution. *Am. Stat.* **2017**, *72*, 172–174. [CrossRef]
52. Eltoft, T.; Kim, T.; Lee, T.-W. On the Multivariate Laplace Distribution. *IEEE Signal Process. Lett.* **2006**, *13*, 300–303. [CrossRef]
53. Decani, J.S.; Stine, R.A. A Note on Deriving the Information Matrix for a Logistic Distribution. *Am. Stat.* **1986**, *40*, 220–222. [CrossRef]
54. Kumar, C.S. A Note on Logistic Mixture Distributions. *Biom. Biostat. Int. J.* **2017**, *2*, 122–124. [CrossRef]
55. Ekawati, D.; Warsono, W.; Kurniasari, D. On the Moments, Cumulants, and Characteristic Function of the Log-Logistic Distribution. *IPTEK J. Sci.* **2015**, *25*. [CrossRef]
56. Makubate, B.; Oluyede, B.; Dingalo, N.; Fagbamigbe, A. The Beta Log-Logistic Weibull Distribution: Model, Properties and Application. *Int. J. Stat. Prob.* **2018**, *7*, 49. [CrossRef]
57. Oluyede, B.; Foya, S.; Warahena-Liyanage, G.; Huang, S. The Log-Logistic Weibull Distribution with Applications to Lifetime Data. *Austrian J. Stat.* **2016**, *45*, 43. [CrossRef]
58. Zhang, X.; Barnes, S.; Golden, B.; Myers, M.; Smith, P. Lognormal-Based Mixture Models for Robust Fitting of Hospital Length of Stay Distributions. *ORHC* **2019**, *22*, 100184. [CrossRef]
59. Sultan, K.; Al-Moisheer, A. Mixture of Inverse Weibull and Lognormal Distributions: Properties, Estimation, and Illustration. *Math. Probl.* **2015**, *2015*, 526786. [CrossRef]
60. Yang, M. Normal Log-Normal Mixture, Leptokurtosis and Skewness. *Appl. Econ.* **2008**, *15*, 737–742. [CrossRef]
61. Negarestani, H.; Jamalizadeh, A.; Shafiei, S.; Balakrishnan, N. Mean Mixtures of Normal Distributions: Properties. *Metrika* **2019**, *82*, 501–528. [CrossRef]

62. Wang, J.; Taaffe, M. Multivariate Mixtures of Normal Distributions: Properties, Random Vector Generation, Fitting, and as Models of Market Daily Changes. *INFORMS J. Comput.* **2015**, *27*, 193–203. [CrossRef]
63. Arnold, B.C. Pareto and Related Heavy-Tailed Distributions. In *Pareto Distributions*, 2nd ed.; Chapman and Hall/CRC: London, UK, 2015; pp. 41–115.
64. Paduthol, S.; Nair, M. On a Finite Mixture of Pareto Distributions. *CSAB* **2005**, *57*, 67–84. [CrossRef]
65. Mohammed, Y.; Yatim, B.; Ismail, S. A Parametric Mixture Model of Three Different Distributions: An Approach to Analyse Heterogeneous Survival Data. In Proceedings of the 21st National Symposium on Mathematical Sciences, Penang, Malaysia, 6–8 November 2013; Volume 1605. [CrossRef]
66. Jiang, R.; Murthy, D.N.P. A Study of Weibull Shape Parameter: Properties and Significance. *Reliab. Eng.* **2011**, *96*, 1619–1626. [CrossRef]