The n-Distribution Bhattacharyya Coefficient

Soo Min Kang and Richard P. Wildes

Technical Report EECS-2015-02

March 9 2015

Department of Electrical Engineering and Computer Science
4700 Keele Street, Toronto, Ontario M3J 1P3 Canada

The n-distribution Bhattacharyya Coefficient

Soo Min Kang and Richard P. Wildes

Department of Electrical Engineering and Computer Science
York University
Toronto, Ontario
Canada

Abstract

The Bhattacharyya coefficient is a widely used statistical measure in various application areas of mathematics, such as computer science, which measures the similarity of two normalized distributions. In this report, we extend the measure by defining an n-distribution Bhattacharyya coefficient, which measures the overlap of n normalized distributions instead of just two. To affirm appropriateness of the measure, we provide useful properties, such as boundedness and properties along its boundaries with proofs, and supply illustrative examples in 1-dimension to demonstrate clearly the effect of the measure in various instances.

Table of Contents

Chapter 1

Introduction

The Bhattacharyya distance was first introduced by Anil Kumar Bhattacharyya as a measure to calculate the similarity between two distributions [3]. It has appeared in various fields from classical statistics [12] to numerous application areas in artificial intelligence, such as speech recognition [13], texture segmentation [10], colour and texture matching [11, 6], feature extraction, image segmentation [1, 9], and action recognition [5].

Many statisticians use the Bhattacharyya measure over others for its simplicity, since they are easier to evaluate than the exact average probability of error [7]. In addition, this measure has several useful mathematical properties that not all distribution comparison methods have. For example, the Bhattacharyya coefficient makes use of the summed unity structure of distributions, unlike L_p-based match measures, it is bounded below by zero and above by one [4], where zero indicates no overlap and one indicates a perfect match between the two distributions, and the bounded nature of the Bhattacharyya coefficient makes it robust to small outliers.

Kailath compared the Bhattacharyya distance with (Kullback-Leibler) divergence in [8], and observed that the Bhattacharyya yields either better or equivalent results in all its tests. In [2], various similarity measures: Bhattacharyya, Euclidean, Kullback-Leibler, and Fisher are studied, analyzed, and compared to examine the effect of each measure when used in image discrimination. The authors concluded that the Bhattacharyya distance is the most effective among the ones studied. In [6], a study compared how different approaches measure the similarity of image textures that are represented using distributions, the Bhattacharyya coefficient generally outperformed the approaches that were considered.

In this report, we extend the usefulness of the Bhattacharyya coefficient for computing the similarity between two distributions to n-distributions, where n is a positive integer, by defining the *n-distribution Bhattacharyya coefficient*, examining its properties, and visualizing its effects when applied to distributions of various types. We will begin by reviewing some mathematical concepts in Chapter 2 that are needed to thoroughly understand the formulation and proof of the n-distribution Bhattacharyya coefficient. In Chapter 3, we will define the n-distribution Bhattacharyya coefficient and examine its properties. In Chapter 4, we will observe the effect of the Bhattacharyya coefficient when applied to various combinations of distributions by looking at concrete examples. Finally, we will close the report with some concluding remarks of the n-distribution Bhattacharyya coefficient in Chapter 5.

Chapter 2

Mathematical Background

In this chapter, we review some mathematical terminology, theorems, and claims that will be used in the next chapter to prove the boundedness of an n-distribution Bhattacharyya coefficient.

Definition 1. *A real-valued function f is said to be **concave** over (a,b) if, for every $x, y \in (a,b)$ and $0 \leq \lambda \leq 1$, the following inequality holds:*

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y).$$

The definition of a convex function is used more frequently, which states that $-f$ is concave if and only if f is convex. However, in this paper, the definition of a concave function has a more practical use. Thus, the term *concave* is formally defined and will be used in subsequent chapters.

Claim 1. *Suppose f and g are non-decreasing real-valued positive concave functions. Then its product $f \cdot g$ is also concave.*

Proof. f and g are positive concave functions on (a,b) for $x, y \in (a,b)$ and $0 \leq \lambda \leq 1$ implies

$$f(\lambda x + (1-\lambda)y) g(\lambda x + (1-\lambda)y) \geq [\lambda f(x) + (1-\lambda) f(y)] [\lambda g(x) + (1-\lambda) g(y)] .$$

By expanding the R.H.S. and adding $-[\lambda (f \cdot g)(x) + (1-\lambda)(f \cdot g)(y)]$ to both sides of the inequality, we get

$$\begin{aligned} \Leftrightarrow (f \cdot g)(\lambda x + (1-\lambda)y) - [\lambda (f \cdot g)(x) + (1-\lambda)(f \cdot g)(y)] &\geq \lambda^2 (f \cdot g)(x) + \lambda(1-\lambda) [f(x) \cdot g(y)] \\ &+ \lambda(1-\lambda) [f(y) \cdot g(x)] + (1-\lambda)^2 (f \cdot g)(y) - [\lambda (f \cdot g)(x) + (1-\lambda)(f \cdot g)(y)]. \end{aligned}$$

By collecting like-terms on the R.H.S. of the inequality, we further derive

$$\Leftrightarrow (f \cdot g)(\lambda x + (1-\lambda)y) - [\lambda(f \cdot g)(x) + (1-\lambda)(f \cdot g)(y)]$$
$$\geq [\lambda(\lambda-1)][(f \cdot g)(x) - f(x) \cdot g(y) - f(y) \cdot g(x) + (f \cdot g)(y)].$$

Next, we can factor the R.H.S. of the inequality to achieve a more concise expression

$$\Leftrightarrow (f \cdot g)(\lambda x + (1-\lambda)y) - [\lambda(f \cdot g)(x) + (1-\lambda)(f \cdot g)(y)] \geq [\lambda(\lambda-1)][f(x) - f(y)][g(x) - g(y)].$$

Since f and g are non-decreasing functions (i.e. $f(x) - f(y), g(x) - g(y) \leq 0 \; \forall \; x < y$), the product of their differences is non-negative. Together with the fact that $0 \leq \lambda \leq 1$, we can further abridge the R.H.S. of the inequality

$$\Rightarrow (f \cdot g)(\lambda x + (1-\lambda)y) - [\lambda(f \cdot g)(x) + (1-\lambda)(f \cdot g)(y)] \geq 0.$$

By rearranging the inequality, we obtain an inequality that satisfies the definition of a concave function

$$\Leftrightarrow (f \cdot g)(\lambda x + (1-\lambda)y) \geq \lambda(f \cdot g)(x) + (1-\lambda)(f \cdot g)(y).$$

□

The concavity of the product of two concave functions will be a necessity to prove the boundedness of the Bhattacharyya coefficient for n-distributions in Chapter 3.

A Jensen's inequality, named after a Danish mathematician Johan Jensen, relates the convex function of an integral to the integral of the convex function. Loosely speaking, the Jensen's inequality states that the average of a convex function is greater than or equal to the function of the average. As mentioned earlier, since concave functions will be more convenient to use in this paper, we state the Jensen's inequality formally with the reversed inequality along with concave functions below.

Theorem 1. *[Jensen's Inequality] If f is a concave function, and $\sum_{i=1}^{I} p_i = 1 \; \forall \; p_i \geq 0$, then*

$$\sum_{i=1}^{I} p_i f(x_i) \leq f\left(\sum_{i=1}^{I} p_i x_i\right).$$

Proof. We begin the proof of the Jensen's inequality by extracting the first term of the series inside the function

$$f\left(\sum_{i=1}^{I} p_i x_i\right) = f\left(p_1 x_1 + \sum_{i=2}^{I} p_i x_i\right).$$

By multiplying the second term by $\sum_{i=2}^{I} p_i / \sum_{i=2}^{I} p_i$, we get

$$f\left(\sum_{i=1}^{I} p_i x_i\right) = f\left(p_1 x_1 + \left[\sum_{i=2}^{I} p_i\right]\left[\frac{\sum_{i=2}^{I} p_i x_i}{\sum_{i=2}^{I} p_i}\right]\right).$$

By setting $\lambda = p_1 = \frac{p_1}{\sum_{i=1}^{I} p_i}$, we have $1 - \lambda = 1 - \frac{p_1}{\sum_{i=1}^{I} p_i} = \frac{\sum_{i=1}^{I} p_i - p_1}{\sum_{i=1}^{I} p_i} = \frac{\sum_{i=2}^{I} p_i}{\sum_{i=1}^{I} p_i} = \sum_{i=2}^{I} p_i$. We can apply the definition of a concave function using these substitutions such that

$$f\left(\sum_{i=1}^{I} p_i x_i\right) \geq p_1 f(x_1) + \sum_{i=2}^{I} p_i f\left(\frac{\sum_{i=2}^{I} p_i x_i}{\sum_{i=2}^{I} p_i}\right).$$

Likewise, we can extract $p_2 x_2$ from the summation inside the latter function to get

$$f\left(\sum_{i=1}^{I} p_i x_i\right) \geq p_1 f(x_1) + \sum_{i=2}^{I} p_i f\left(\frac{p_2 x_2}{\sum_{i=2}^{I} p_i} + \frac{\sum_{i=3}^{I} p_i x_i}{\sum_{i=2}^{I} p_i}\right).$$

Multiplying the second term inside the latter function by $\sum_{i=3}^{I} p_i / \sum_{i=3}^{I} p_i$ and rearranging the order of the terms gives us

$$f\left(\sum_{i=1}^{I} p_i x_i\right) \geq p_1 f(x_1) + \sum_{i=2}^{I} p_i f\left(\frac{p_2}{\sum_{i=2}^{I} p_i} x_2 + \frac{\sum_{i=3}^{I} p_i}{\sum_{i=2}^{I} p_i} \frac{\sum_{i=3}^{I} p_i x_i}{\sum_{i=3}^{I} p_i}\right).$$

By setting $\lambda = \frac{p_2}{\sum_{i=2}^{I} p_i}$, we obtain $1 - \lambda = 1 - \frac{p_2}{\sum_{i=2}^{I} p_i} = \frac{\sum_{i=2}^{I} p_i - p_2}{\sum_{i=2}^{I} p_i} = \frac{\sum_{i=3}^{I} p_i}{\sum_{i=2}^{I} p_i}$. Again, using the definition of concave functions,

$$f\left(\sum_{i=1}^{I} p_i x_i\right) \geq p_1 f(x_1) + \sum_{i=2}^{I} p_i \left[\frac{p_2}{\sum_{i=2}^{I} p_i} f(x_2) + \frac{\sum_{i=3}^{I} p_i}{\sum_{i=2}^{I} p_i} f\left(\frac{\sum_{i=3}^{I} p_i x_i}{\sum_{i=3}^{I} p_i}\right)\right].$$

By cancelling the summation that appears in front of the square parenthesis, we obtain

$$f\left(\sum_{i=1}^{I} p_i x_i\right) \geq p_1 f(x_1) + p_2 f(x_2) + \sum_{i=3}^{I} p_i f\left(\frac{\sum_{i=3}^{I} p_i x_i}{\sum_{i=3}^{I} p_i}\right).$$

We can continue extracting a term inside the series, set it to λ, then use the definition of concave functions to obtain a linear combination of functions $f(x_i)$ with a coefficient of p_i for every $i = 1, \dots, I$. $\square$

The Jensen's inequality of concave functions will be used in the next chapter to prove the boundedness of the n-distribution Bhattacharyya coefficient.

Chapter 3

n-distribution Bhattacharyya Coefficient

In this chapter, we define the n-distribution Bhattacharyya coefficient for n normalized distributions, and state useful properties followed by a proof for each one.

3.1 Definition of the n-distribution Bhattacharyya Coefficient

The Bhattacharyya coefficient introduced by Bhattacharyya is a statistical measure that evaluates the overlap between two normalized probability distributions [3]. Formally speaking, for two normalized discrete distributions f and g (i.e. $\sum_k f(x_k) = 1$ and $\sum_k g(x_k) = 1$), the Bhattacharyya coefficient is defined as:

$$B(f, g) = \sum_k \sqrt{f(x_k)g(x_k)}.^{1}$$

The Bhattacharyya coefficient is not to be confused with the *Bhattacharyya distance*, which is defined as:

$$d_B(f, g) = -\ln B(f, g),$$

where $B(f, g)$ denotes the Bhattacharyya coefficient. The Bhattacharyya coefficient is bounded above and below by 0 and 1, respectively, while its distance is not bounded above. That is, $0 \leq B(f, g) \leq 1$ and $0 \leq d_B(f, g) \leq \infty$ [8]. In this chapter, we extend the commonly

[1] In this paper, we use discrete distributions $f(x_k)$ and $g(x_k)$ where $\sum_k f(x_k) = 1, \sum_k g(x_k) = 1$. However, the Bhattacharyya coefficient definition holds for two continuous functions f and g, which is defined as $B(f, g) = \int \sqrt{f(x)g(x)}dx$ such that $\int f(x)dx = 1$ and $\int g(x)dx = 1$.

used Bhattacharyya coefficient for a pair of distributions to a Bhattacharyya coefficient for n-distributions, and demonstrate the boundedness through an analogous proof of the base case.

Definition 2. *Suppose $f_1, \ldots, f_n$ are real-valued functions such that $\sum_k f_i(x_k) = 1$ and $0 \leq f_i(x_k) \leq 1 \; \forall \; i \in \mathbb{Z}^+$. The n-distribution Bhattacharyya coefficient is defined as*

$$B(f_1, \ldots, f_n) = \sum_k \sqrt[n]{\prod_{i=1}^{n} f_i(x_k)}.$$

3.2 Properties of the n-distribution Bhattacharyya Distribution

In this section, properties of the n-distribution Bhattacharyya coefficient is established. Specifically, the bounds of the coefficient will be examined. A proof of each property will be provided, which will show that each property is a direct extension of the Bhattacharyya coefficient in the case of two distributions.

Claim 2. $0 \leq B(f_1, \ldots, f_n) \leq 1$ *for* $\forall \; f_i$ *such that* $\sum_k f_i(x_k) = 1$ *and* $0 \leq f_i(x_k) \leq 1$.

Proof. **Lower-bound:** $0 \leq B(f_1, \ldots, f_n)$
Since $f_i \geq 0 \; \forall \; x_k$, the product of f_i is non-negative, and its n-th root also remains non-negative:

$$\sqrt[n]{f_1(x_k) \cdot \ldots \cdot f_n(x_k)} \geq 0 \; \forall \; x_k, f_i.$$

Furthermore, the sum of non-negative values yield a non-negative value:

$$B(f_1, \ldots, f_n) = \sum_k \sqrt[n]{f_1(x_k) \cdot \ldots \cdot f_n(x_k)} \geq 0.$$

Thus, the n-distribution Bhattacharyya coefficient is non-negative.

Upper-bound: $B(f_1, \ldots, f_n) \leq 1$

We will prove the claim via proof by induction. We begin the proof for the base case ($n = 2$):

$$B(f_1, f_2) = \sum_k \sqrt{f_1(x_k) f_2(x_k)}.$$

By multiplying the product under the square root by ${}^{f_2(x_k)}/_{f_2(x_k)}$ and simplifying, we get

$$B(f_1, f_2) = \sum_k f_2(x_k) \sqrt{\frac{f_1(x_k)}{f_2(x_k)}}.$$

Since $f(x) = \sqrt{x}$ is a concave function, $\sum_k f_2(x_k) = 1$ and $f_2(x_k) > 0 \ \forall \ x_k$, we can apply the Jensen's inequality (Theorem 1) to obtain

$$B(f_1, f_2) \leq \sqrt{\sum_k f_2(x_k) \frac{f_1(x_k)}{f_2(x_k)}}.$$

We can simplify the equation by cancelling the common term $f_2(x_k)$ under the root to obtain the inequality

$$B(f_1, f_2) \leq \sqrt{\sum_k f_1(x_k)}.$$

Since $\sum_k f_1(x_k) = 1$ by construction,

$$B(f_1, f_2) \leq 1,$$

which proves the base case of the Bhattacharyya coefficient.

In the inductive case, we suppose the claim is true for $n - 1$. That is,

$$B(f_1, \ldots, f_{n-1}) = \sum_k \sqrt[n-1]{f_1(x_k) \cdot \ldots \cdot f_{n-1}(x_k)} \leq 1.$$

We implement a similar series of steps to prove the upper bound of the Bhattacharyya coefficient for n-distributions:

$$B(f_1, \ldots, f_n) = \sum_k \sqrt[n]{f_1(x_k) \cdot \ldots \cdot f_{n-1}(x_k) f_n(x_k)}.$$

By multiplying the product under the n^{th}-root by $\left[\frac{f_n(x_k)}{f_n(x_k)}\right]^{n-1}$ then simplifying, we have

$$B(f_1, \ldots, f_n) = \sum_k f_n(x_k) \sqrt[n]{\frac{f_1(x_k) \cdot \ldots \cdot f_{n-1}(x_k)}{[f_n(x_k)]^{n-1}}}.$$

The equation can be rewritten in the following form

$$B(f_1, \ldots, f_n) = \sum_k f_n(x_k) \left\{ \frac{[f_1(x_k) \cdot \ldots \cdot f_{n-1}(x_k)]^{\frac{1}{n-1}}}{f_n(x_k)} \right\}^{\frac{n-1}{n}}.$$

Since $f(x) = x^{n-1}$ and $g(x) = x^{1/n}$ $\forall\ n \in \mathbb{Z}^+$ are non-decreasing positive concave functions, by Claim 1, its product $(f \cdot g)(x) = x^{(n-1)/n}$ is also concave. Together with the given condition, $\sum_k f_n(x_k) = 1$ and $f_n(x_k) \geq 0$ $\forall\ x_k$, we apply the Jensen's inequality to obtain

$$B(f_1, \ldots, f_n) \leq \left\{ \sum_k f_n(x_k) \frac{[f_1(x_k) \cdot \ldots \cdot f_{n-1}(x_k)]^{\frac{1}{n-1}}}{f_n(x_k)} \right\}^{\frac{n-1}{n}}.$$

Now, we can simplify the equation by cancelling the common term $f_n(x_k)$ in the numerator and in the denominator and combining the exponents such that

$$B(f_1, \ldots, f_n) \leq \left[\sum_k \sqrt[n-1]{f_1(x_k) \cdot \ldots \cdot f_{n-1}(x_k)} \right]^{\frac{n-1}{n}}.$$

Next, by inductive hypothesis, we know that the term inside the parenthesis is less than or equal to 1. Thus,

$$B(f_1, \ldots, f_n) \leq [1]^{\frac{n-1}{n}} = 1.$$

Therefore, we can conclude that the upper-bound of the Bhattacharrya coefficient for calculating the overlap of n-distributions is 1. □

Next, we examine the meaning of these values when applied to measure the overlap of n distributions. The simplest way to determine what each bound refers to would be to consider the two most extreme cases: n distributions with no overlap and n distributions with perfect overlap. We will explore these two cases in the next two remarks.

Remark 1. *For any pair of real-valued functions $f_i, f_j \in \{f_1, \ldots, f_n\}$ that are orthogonal, the n-distribution Bhattacharyya coefficient containing the orthogonal functions is* *equal to 0. That is, if $f_i \cap f_j = \emptyset$ then $B(f_1, \dots, f_n) = 0$.*

Proof. By the definition of an n-distribution Bhattacharyya coefficient, we have

$$B(f_1, \dots, f_n) = \sum_k \left\{ \sqrt[n]{\prod_{i=1}^{n} f_i(x_k)} \right\}.$$

Expanding the product yields

$$B(f_1, \dots, f_n) = \sqrt[n]{f_1(x_1) \cdot \ldots \cdot f_n(x_1)} + \cdots + \sqrt[n]{f_1(x_k) \cdot \ldots \cdot f_n(x_k)}.$$

Since f_i and f_j do not overlap, we have $f_i(x_k) \cdot f_j(x_k) = 0 \ \forall \ x_k$ (i.e. the product of any two non-overlapping functions is 0). Then the product under each nth root is zero:

$$B(f_1, \dots, f_n) = \sqrt[n]{0} + \cdots + \sqrt[n]{0}.$$

The zero under each nth root remains zero, and the sum of these terms yield zero:

$$B(f_1, \dots, f_n) = 0.$$

Therefore, the n-distribution Bhattacharyya coefficient is 0 if $f_i \cap f_j = \emptyset$ for any $i, j \in \{1, \dots, n\}$. □

Remark 2. *If all functions are the same, then the n-distribution Bhattacharyya coefficient is equal to 1.*

Proof. By the definition of an n-distribution Bhattacharyya coefficient, we have

$$B(f_1, \dots, f_n) = \sum_k \left\{ \sqrt[n]{f_1(x_k) \cdot \ldots \cdot f_n(x_k)} \right\}.$$

Since f_i are the same $\forall \ 1 \leq i \leq n$ (i.e. $f_1 = \ldots = f_n$), the product under the root can be rewritten into a single term with an exponent:

$$B(f_1, \dots, f_n) = \sum_k \left\{ \sqrt[n]{[f_1(x_k)]^n} \right\}.$$

Using the properties of exponents, we can cancel the exponent n and the n-th root to obtain

$$B(f_1, \dots, f_n) = \sum_k f_1(x_k).$$

By construction, the sum of all k values of a function f_1 is 1:

$$B(f_1, \ldots, f_n) = 1.$$

Thus, the n-distribution Bhattacharyya coefficient of n distributions that are the same is 1. $\square$

Chapter 4

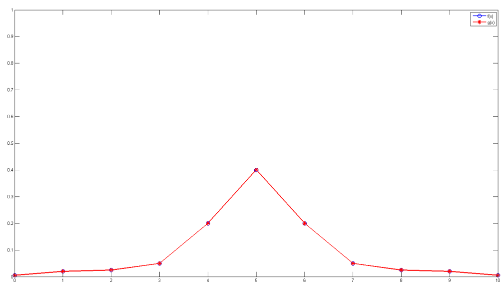
Illustrative Examples

In this chapter, we provide illustrative examples of the n-dimensional Bhattacharyya coefficient. We begin with examples of two distributions with perfect overlap, no overlap, and some overlap. To observe how the value of the n-distribution Bhattacharyya coefficient is affected with n-distributions, examples containing more than two distributions will follow.

In Figure 4.1a, it can be seen that taking the root of the product of two identical functions in its respective bins yield a distribution that is the same as the original function. Ergo, the bins sum to one by construction. On the contrary, the product of two non-overlapping distributions (as in Figure 4.1b) is zero at each bin, x, yielding a cumulative sum of zero. Hence, the Bhattacharyya coefficient of two non-overlapping distribution is 0 and two identical distributions is 1 as outlined in Remark 1 and Remark 2, respectively.

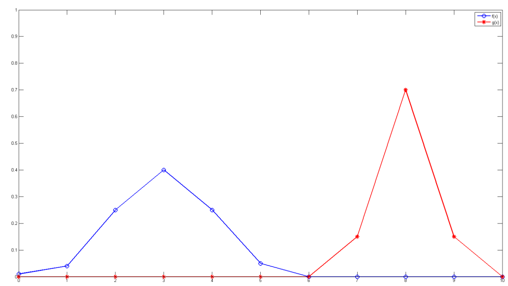
In Figure 4.2, we illustrate how the value of the Bhattacharyya coefficient is affected in areas with overlap for two distributions. In regions where there is overlap, namely bins $1 \leq x \leq 3$ and $7 \leq x \leq 9$, the square root of the products are greater than zero, which contributes to the cumulative sum of the Bhattacharyya coefficient. However, the product of the non-overlapping regions, $4 \leq x \leq 6$, equal zero. This prevents the sum from accumulating to its maximum value of 1.

In Figure 4.3a, Figure 4.3b, and Figure 4.4, we demonstrate the effect of comparing 3 distributions using the n-distribution Bhattacharyya coefficient. Figure 4.3a depicts the importance of overlap in all distributions to obtain a Bhattacharyya coefficient greater than zero. That is, although distribution f (blue) overlap distributions g (red) and h (black), since there is no overlap between g and h, the product of the three distributions is zero in all its bins x. This causes the n-distribution Bhattacharyya coefficient to remain at a value of zero as outlined in Remark 1. However, if there is some overlap between the distributions (as in Figure 4.3b), then the overlapping regions contribute to the cumulative sum yielding a Bhattacharyya coefficient value that is greater than zero. In Figure 4.4, we illustrate explicitly that the measure is not limited to unimodal distributions (a curve with one maximum), but can be applied to multi-modal distributions (distributions with multiple maxima). Again, it is only the regions where there is overlap, namely $2 \leq x \leq 3$ and $x = 7$ in Figure 4.4, that contribute to the sum in the n-distribution Bhattacharyya coefficient.



x	$f(x)$	$g(x)$	$\sqrt{f(x)g(x)}$
0	0.0050	0.0050	0.0050
1	0.0200	0.0200	0.0200
2	0.0250	0.0250	0.0250
3	0.0500	0.0500	0.0500
4	0.2	0.2	0.2
5	0.4	0.4	0.4
6	0.2	0.2	0.2
7	0.05	0.05	0.05
8	0.0250	0.0250	0.0250
9	0.02	0.02	0.02
10	0.005	0.005	0.005
$\sum_x$	1	1	1

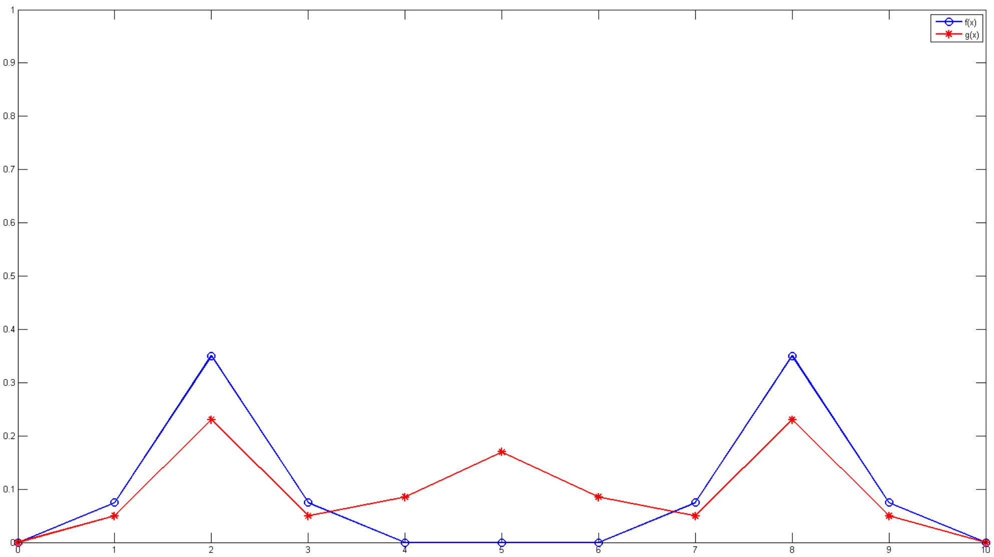
(a) Two distributions with perfect overlap.



x	$f(x)$	$g(x)$	$\sqrt{f(x)g(x)}$
0	0	0	0
1	0.01	0	0
2	0.04	0	0
3	0.25	0	0
4	0.4	0	0
5	0.25	0	0
6	0.05	0	0
7	0	0	0
8	0	0.15	0
9	0	0.7	0
10	0	0.15	0
$\sum_x$	1	1	0

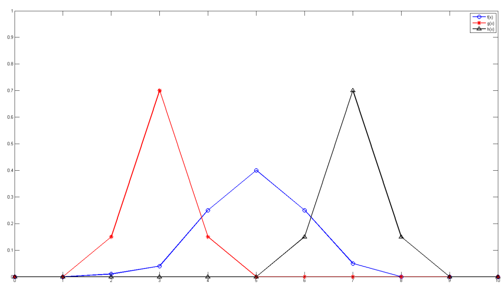
(b) Two distributions with no overlap.

Figure 4.1: Bhattacharyya coefficient illustration for two distributions.



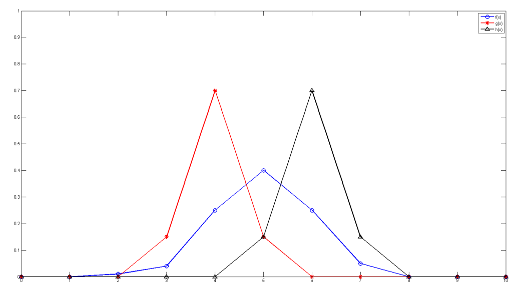
x	$f(x)$	$g(x)$	$\sqrt{f(x)g(x)}$
0	0	0	0
1	0.075	0.05	0.0612
2	0.35	0.23	0.2837
3	0.075	0.05	0.0612
4	0	0.085	0
5	0	0.17	0
6	0	0.085	0
7	0.075	0.05	0.0612
8	0.35	0.23	0.2837
9	0.075	0.05	0.0612
10	0	0.0	0
$\sum_x$	1	1	0.8124

Figure 4.2: Two non-unimodal distributions with some overlap.



x	$f(x)$	$g(x)$	$h(x)$	$\sqrt[3]{f(x)g(x)h(x)}$
0	0	0	0	0
1	0	0	0	0
2	0.01	0.15	0	0
3	0.04	0.7	0	0
4	0.25	0.15	0	0
5	0.4	0	0	0
6	0.25	0	0.15	0
7	0.05	0	0.7	0
8	0	0	0.15	0
9	0	0	0	0
10	0	0	0	0
$\sum_x$	1	1	1	0

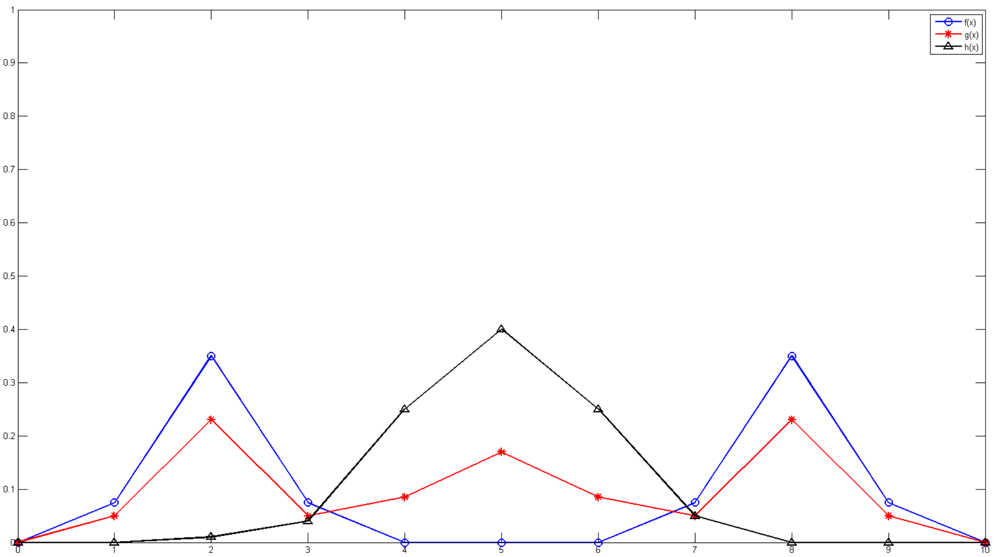
(a) Three distributions with no overlap.



x	$f(x)$	$g(x)$	$h(x)$	$\sqrt[3]{f(x)g(x)h(x)}$
0	0	0	0	0
1	0	0	0	0
2	0.01	0	0	0
3	0.04	0.15	0	0
4	0.25	0.7	0	0
5	0.4	0.15	0.15	0.2080
6	0.25	0	0.7	0
7	0.05	0	0.15	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0
$\sum_x$	1	1	1	0.2080

(b) Three distributions with some overlap.

Figure 4.3: Bhattacharyya coefficient illustration for three distributions.



x	$f(x)$	$g(x)$	$h(x)$	$\sqrt[3]{f(x)g(x)h(x)}$
0	0	0	0	0
1	0.075	0.05	0	0
2	0.35	0.23	0.01	0.0930
3	0.075	0.05	0.04	0.0531
4	0	0.085	0.25	0
5	0	0.17	0.4	0
6	0	0.085	0.25	0
7	0.075	0.05	0.05	0.0572
8	0.35	0.23	0	0
9	0.075	0.05	0	0
10	0	0	0	0
$\sum_x$	1	1	1	0.2034

Figure 4.4: Three non-unimodal distributions with some overlap.

Chapter 5

Conclusion

In this report, we extended the widely used Bhattacharyya coefficient, which compares two normalized distributions, to n normalized distributions. Similar to the Bhattacharyya coefficient of two distributions, the n-distribution Bhattacharyya coefficient is bounded by 0 and 1. The n-distribution Bhattacharyya coefficient is more sensitive to overlaps. That is, if there are any pairs of non-overlapping distributions, then the n-distribution Bhattacharyya coefficient will have a value of zero. Similar to the Bhattacharyya coefficient for two distributions, more overlap amongst the distributions yield a Bhattacharyya value closer to 1, and a perfect overlap between the distributions yield a value of exactly 1. The measure is not limited to unimodal distributions, but can be applied to multi-modal distributions, provided that the distributions are normalized.

Acknowledgments

The authors would like to thank Hang Gao for her valuable suggestion in the proof of the n-distribution Bhattacharyya coefficient boundedness.

Bibliography

[1] I.B. Ayed, H.M. Chen, K. Punithakumar, I. Ross, and S. Li. Graph cut segmentation with a global constraint: Recovering region distribution via a bound of the Bhattacharyya measure. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3288–3295, 2010.

[2] A.H. Bhalerao and N.M. Rajpoot. Discriminant feature selection for texture classification. In *Proceedings of the British Machine Vision Conferences*, pages 80.1–80.10. BMVA Press, 2003.

[3] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *The Indian Journal of Statistics*, 7(4):401–406, 1943.

[4] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):564–577, 2003.

[5] K.G. Derpanis, M. Sizintsev, K. Cannons, and R.P. Wildes. Action spotting and recognition based on a spatiotemporal orientation analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3):527–540, 2013.

[6] K.G. Derpanis and R.P. Wildes. Spacetime texture representation and recognition based on a spatiotemporal orientation analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(6):1193–1205, 2012.

[7] A. Djouadi, O. Snorrason, and F.D. Garber. The quality of training-sample estimates of the Bhattacharyya coefficient. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):92–97, 1990.

[8] T. Kailath. The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15(1):52–60, 1967.

[9] S.M. Kang and J.W.L. Wan. A multiscale graph cut approach to bright-field multiple cell image segmentation using a Bhattacharyya measure. In *Proceedings of SPIE Symposium on Medical Imaging: Image Processing*, volume 8669, pages 86693S1–86693S8, 2013.

[10] C.C. Reyes-Aldasoro and A. Bhalerao. The Bhattacharyya space for feature selection and its application to texture segmentation. *The Journal of the Pattern Recognition Society*, 39:812–826, 2006.

[11] Y. Rubner, J. Puzicha, C. Tomasi, and J. Buhmann. Empirical evaluation of dissimilarity measures for color and texture. *Computer Vision and Image Understanding*, 84(1):25–43, 2001.

[12] C. Stein. Approximation of Improper Prior Measures by Prior Probability Measures. Technical Report 12, Stanford University Department of Statistics, 1964.

[13] C.H. You, K.A. Lee, and H. Li. An SVM kernel with GMM-supervector based on the Bhattacharyya distance for speaker recognition. *IEEE Signal Processing Letters*, 16(1):49–52, 2009.