Potential Discrimination Identification Using Pseudo-Boolean Functions

Jinwook Lee, Lanqing Du, Jade Zhang, Matthew J. Schneider

Decision Sciences and MIS, Drexel University

April 2, 2022

NRC

List of topics

1. Section 1. Introduction
2. Section 2. Multi-linear polynomial representation of the metrics for datasets with potential bias issues
   - Some basic notations and definitions of pseudo-Boolean functions
   - Quadratic pseudo-Boolean functions and the metrics for datasets with potential bias issues
3. Section 3. Probability of the union of partially ordered sets
   - Quadratic pseudo-Boolean function for the bivariate case
   - A desirable partial order relation
   - General form of metrics for discrimination and bias
4. Section 4. Sequential time series data aggregation using a suitable matching algorithm

- Section 1. Introduction

Introduction

This paper contributes mainly to three literatures: the growing body of work on data with potential discrimination and bias issues; the study of data aggregation; and their counterparts on the graph. There are a wide range of metrics to measure unfairness and discrimination among different groups.

Why is discrimination detection important? How to detect?

- Social good: one should not single out any particular communities based on characteristics such as race, religion, gender, education, income level, etc.
- There are various metrics for discrimination detection, classification fairness for both group and individual levels.

Introduction

- We use pseudo-Boolean functions for the generalization of such metrics, that are used for effective detection and aggregation of data.
- Discrimination scenarios could be related to, for example, education, healthcare, recommendation systems, etc.

Asian	African American	Hispanic	White	Non-Resident Aliens	Other Races
0	0	0	2	19	0

Table 1: "Creative Aggregation": Ethnic Diversity of Statistics Master's Degree Recipients at University of California - Davis [1]

- Pseudo-Boolean functions and their applications have been used in most areas including statistics and probability theory, computer science, operations research, finance, manufacturing, etc. We refer the readers to the literature (see, e.g., [2], [3], [4], [5]).

Some notable discrimination and bias metrics

- Mean Difference: this metric measures the difference of the predicted outcome between the protected group and the general group.

$$d = P(\hat{y} = 1|g_a) - P(\hat{y} = 1|g_b) \tag{1}$$

- Equal opportunity: Considering the subtle differences across the groups, the Confusion matrix is introduced to assist the fairness measurement.

$$d = P(\hat{y} = 1|y = 1\&g_a) - P(\hat{y} = 1|y = 1\&g_b) \tag{2}$$

- Test fairness: Instead of comparing the outcomes across groups, test fairness requires the same likelihood of prediction for individuals regardless the group membership.

$$d = P(\hat{y} = 1|s_a\&g_a) - P(\hat{y} = 1|s_b\&g_b), \text{where } s_a = s_b \tag{3}$$

Some notable discrimination and bias metrics

- Situation testing: This metrics investigates if there is any significant difference of treatment between an individual and a similar person in a different group. Based on k-nearest neighbors, situation testing estimates the probability for an individual to have positive outcome by calculating the proportions of its neighbors who received positive outcomes in both protected group and general group.
- Bounded group loss: A bounded group loss measures the worst loss on any protected group. The prediction is considered satisfying the bounded group loss at level ϵ if the expected worst loss of a group is below ϵ.

$$E(l(\hat{y},y)|g_i) \leq \epsilon \tag{4}$$

- Regression slope test, difference of mean test, difference of proportions of multiple groups, etc.

- Section 2. Multi-linear polynomial representation of the metrics for datasets with potential bias issues

Some basic notions of Boolean functions

Our probability space is $(\Omega, \mathcal{F}, P) = (\mathbb{R}^n, \mathcal{B}, P)$, where $\mathcal{F} = \mathcal{B}$ is the σ-field of Borel subsets of Ω. Let $A_1, \ldots, A_n$ be events in a probability space Ω. If we subdivide Ω into 2^n disjoint subsets in such a way that we take all intersections of the original events and their complementary events, i.e., the collection of events:

$$\begin{aligned} &A_{i_1} \ldots A_{i_k} \bar{A}_{j_1} \ldots \bar{A}_{i_{n-k}}, \\ &1 \leq i_1 < \cdots < i_k \leq n, \{j_1, \ldots, j_{n-k}\} = \{1, \ldots, n\} \backslash \{i_1, \ldots, i_k\} \\ &k = 1, \ldots, n. \end{aligned} \tag{5}$$

For example, if $n = 3$, then we have the $2^3 = 8$ intersections:

$$\begin{array}{ll} A_1 A_2 A_3 & \bar{A}_1 \bar{A}_2 A_3 \\ \bar{A}_1 A_2 A_3 & \bar{A}_1 A_2 \bar{A}_3 \\ A_1 \bar{A}_2 A_3 & A_1 \bar{A}_2 \bar{A}_3 \\ A_1 A_2 \bar{A}_3 & \bar{A}_1 \bar{A}_2 \bar{A}_3 \end{array}$$

Some basic notions of pseudo-Boolean functions

In a more general formulation of the Boolean probability problem we can write probabilities as follows.

$$P(A_{i_1} \dots A_{i_k}),\ 1 \leq i_1 < \cdots < i_k \leq m$$
$$k = 1, \dots, m.$$

Let us consider indices of the sets A_i's. Then we let $V = [n] = \{1, 2, \dots, n\}$, a set of integers (or indices), together with $\mathbb{B} = \{0, 1\}$. For any subset $S \subseteq V$, we define the following characteristic vector as:

$$\mathbb{1}_j^S = \begin{cases} 1 & \text{if } j \in S \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Some basic notions of pseudo-Boolean functions

Let us introduce n binary variables $x_1, x_2, \ldots, x_n$. Then we can write $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{B}^n$, a binary vector. For the sake of completeness, we present some basic notions. The variables x_i and their complements $\bar{x}_i = 1 - x_i, \forall i \in V$ consist of the set of literals: $\mathbf{L} = \{x_1, \bar{x}_1, \ldots, x_n, \bar{x}_n\}$. Mappings $f : \mathbb{B}^n \to \mathbb{R}$ are called *pseudo-Boolean functions.* All pseudo-Boolean functions are uniquely represented as *multi-linear polynomials* of the following form:

$$f(x_1, \ldots, x_n) = \sum_{S \subseteq V} c_S \prod_{j \in S} x_j, \tag{7}$$

and by definition, we have $\prod_{j \in \emptyset} x_j = 1$. The size of the largest subset $S \subseteq V$ for which $c_S \neq 0$ is called the degree of f. The first order derivative of a pseudo-Boolean function with respect to the ith variable can be written up as:

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n). \tag{8}$$

Characteristic set function

Let $X \in \mathbb{R}^n$ be a random variable. Suppose that our given dataset is based on a random vector (or multivariate random variable) X. Then we can write the following:

Definition (Characteristic set function)

In case of a finite support, there exist $N \geq 1$ and $X^{(1)}, \ldots, X^{(N)}$ such that $\{X^{(1)}, \ldots, X^{(N)}\}$, where $X^{(i)} \in \mathbb{R}^n$. Let A denote an arbitrary event. Then we can define a set function $h : \mathbb{R}^n \to \mathbb{B}^n$ as follows.

$$h(X^{(i)}) = \mathbf{x}, \tag{9}$$

where $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{B}^n$. Then the support of $h(X) \in \mathbb{B}^n$ can be written up as a binary matrix $\mathbb{B}^{N \times n}$.

Note that if X has a discrete distribution on $\mathbb{Z}^n$, then its support is finite or countably infinite.

Quadratic pseudo-Boolean functions and the metrics for datasets with potential bias issues

Quadratic pseudo-Boolean functions in our case can be represented by

$$f(x_1, \ldots, x_n) = c_0 \pm \sum_{j \in N^{(1)}} c_j x_j \pm \sum_{(i,j) \in N^{(2)}} c_{ij} x_i x_j, \tag{10}$$

where $N^{(1)}$ and $N^{(2)}$ are given sets of singles and pairs in which one might be interested.

Some of the prominent metrics are conditional expectations and probabilities. In order to connect the pseudo-Boolean functions with them, we let $c_S = E(\psi(\mathbf{x}) \mid j \in S), S \subseteq V = [n]$. This includes the following.

- $c_i = P(x_i = 1),\ c_{ij} = P(x_i = 1, x_j = 1)$
- $c_i = P(y = 1 \mid x_i = 1),\ c_{ij} = P(y = 1 \mid x_i = 1, x_j = 1)$
- $c_i = E(\psi(\mathbf{x}) \mid x_i = 1),\ c_{ij} = E(\psi(\mathbf{x}) \mid x_i = 1, x_j = 1).$

Simple numerical example 1/2

Public disaggregated enrollment data from University of California ([6]).

Group of interest:

- (African American, CA residence, Female) $= (x_{1(1)} = 1, x_{2(1)} = 1, x_3 = 1)$,
- (African American, CA residence, Male) $= (x_{1(1)} = 1, x_{2(1)} = 1, x_3 = 0)$,
- (White Caucasian, CA residence, Female) $= (x_{1(2)} = 1, x_{2(1)} = 1, x_3 = 1)$,
- (White Caucasian, CA residence, Male) $= (x_{1(2)} = 1, x_{2(1)} = 1, x_3 = 0)$,

Simple numerical example 2/2

$P(AA) = 0.0240$	$P(CA) = 0.1408$	$P(AA,CA) = 0.0027$	$f(AA,CA) = 0.1621$
$P(WC) = 0.3109$	$P(CA) = 0.1408$	$P(WC,CA) = 0.0360$	$f(WC,CA) = 0.4157$
$P(AA) = 0.0240$	$P(F) = 0.5152$	$P(AA,F) = 0.0136$	$f(AA,F) = 0.5256$
$P(WC) = 0.3109$	$P(F) = 0.5152$	$P(WC,F) = 0.1625$	$f(WC,F) = 0.6636$
$P(AA) = 0.0240$	$P(M) = 0.2105$	$P(AA,M) = 0.0045$	$f(AA,M) = 0.2300$
$P(WC) = 0.3109$	$P(M) = 0.2105$	$P(WC,M) = 0.0633$	$f(WC,M) = 0.4581$

Table 2: pseudo-Boolean function for the probability of union

$P(F\|AA) = 0.5658$	$P(F\|CA) = 0.0855$	$P(F\|AA,CA) = 0.0516$
$P(F\|WC) = 0.5227$	$P(F\|CA) = 0.0855$	$P(F\|WC,CA) = 0.0676$
$P(AA\|F) = 0.0263$	$P(AA\|CA) = 0.0189$	$P(AA\|CA,F) = 0.0027$
$P(WC\|F) = 0.3155$	$P(WC\|CA) = 0.2554$	$P(WC\|CA,F) = 0.0034$
$P(M\|AA) = 0.1864$	$P(M\|CA) = 0.2264$	$P(M\|AA,CA) = 0.2433$
$P(M\|WC) = 0.2036$	$P(M\|CA) = 0.2264$	$P(M\|WC,CA) = 0.2493$

Table 3: Conditional probabilities as coefficients for pseudo-Boolean functions

(Un)Favorable events in $\mathbb{R}^n$

Suppose that a pseudo-Boolean function f is related to losses from unwanted events with non-decreasing structure (i.e., $f(t_1) \geq f(t_2)$ if $t_1 \geq t_2, t_1 \neq t_2$, where $t_1, t_2 \in \mathbb{B}^n$) as we assume less social discrimination comes with more benefit. Using the general form of pseudo-Boolean function f and a set function h in (9) we can write, for $z \in \mathbb{Z}^n$,

$$H_s(X) = \{z \mid f_X(h(z)) \geq s\}, \tag{11}$$

where and $s \in \mathbb{R}$ denotes some boundary value. If we let s be a maximum allowable loss level, a favorable event for X can be defined as the following

$$X \in W_s = \bigcup_{z \in H_s(X)} (z + \mathbb{R}^n_-), \tag{12}$$

together with its complementary event (i.e., unfavorable event):

$$X \in W_s^c = \bigcap_{z \in H_s(X)} (z + \mathbb{R}^n_-)^c. \tag{13}$$

Description of (un)favorable events

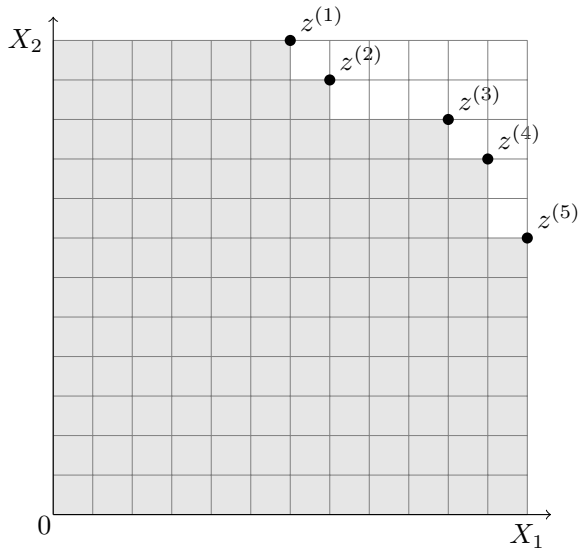

Figure 1: Description of favorable event (shaded region)

- The shaded region in the figure can be considered a union of orthants.

Description of (un)favorable events

- If we let the $X_i, i = 1, \ldots, n$ designate the proportion of itself to total, then we will have an n-dimensional hypercube.
- Any given requirements and/or regulations can be represented as a region in the hypercube.
- Since $u \in \mathbb{U}^n$ includes every possible values in the unit hypercube, we let $A_i = \{u \in \mathbb{U}^n \mid u \leq s^{(i)}\}$ for $i = 1, \ldots, N$ for our favorable events.
- Example:

$$\begin{aligned} A_1 &= \{u \in \mathbb{U}^n \mid u_1 \leq 0.05, u_2 \leq 0.05, u_3 \leq 1, \ldots, u_n \leq 1\} \\ &= \{u \in \mathbb{U}^n \mid u \leq (0.05, 0.05, 1, \ldots, 1)^T\} \end{aligned}$$

- The volume of A_i can be thought of as a threshold probability level.
- Therefore, we're interested in knowing if there exist cases in at least one of such events (or not).
- It would also be useful to know the following probabilities: union of such events, at least k events occur, exactly k events occur, etc.

- Section 3. Probability of the union of partially ordered sets

The formula for the bivariate case

Theorem (Modified formula for the bivariate case by [7])

In any given probability space, on a finite poset in $\mathbb{R}^2$ *with the maximal elements* $A_i = A(s^{(i)}) = \{z \in \mathbb{R}^2 \mid z \leq s^{(i)}\}, i = 1, \ldots, N$, *sorted on their 1st component such that* $z_1 < \cdots < z_n$, *we have the formula:*

$$P\left(\bigcup_{i=1}^{N} A_i\right) = \sum_{i=1}^{N} P(A_i) - \sum_{i=1}^{N-1} P(A_i A_{i+1}) = S_1 - S_2', \tag{14}$$

where S_1 *is the first binomial moment of the events* $A_1, \ldots A_N$, *and* S_2' *is the sum of the probabilities of the "incomparable" pairwise intersections.*

The formula for the bivariate case

Proof.

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{N} A_i\right) &= P(A_1) + P(A_2\bar{A}_1) + P(A_3\bar{A}_1\bar{A}_2) + \cdots \\
&\quad + \cdots + P(A_k\bar{A}_1\bar{A}_2\ldots\overline{A_{k-1}}) + \cdots + P(A_N\bar{A}_1\bar{A}_2\ldots\overline{A_{N-1}}) \\
&= P(A_1) + P(A_2\bar{A}_1) + P(A_3\bar{A}_2) + \cdots + P(A_k\overline{A_{k-1}}) + \cdots + P(A_N\overline{A_{N-1}}) \\
&= P(A_1) + P(A_2) - P(A_1A_2) + P(A_3) - P(A_2A_3) + \cdots \\
&\quad + \cdots + P(A_k) - P(A_{k-1}A_k) + \cdots + P(A_N) - P(A_{N-1}A_N) \\
&= \sum_{i=1}^{N} P(A_i) - \sum_{i=1}^{N-1} P(A_iA_{i+1}) = S_1 - S_2'.
\end{aligned}
\tag{15}
$$

□

The formula for the bivariate case with c.d.f.

Corollary (Modified inclusion-exclusion formula for the bivariate case with c.d.f.)

Let F be the given c.d.f. of a random vector $Z \in \mathbb{R}^2$. Then the formula (14) can be written as:

$$P\left(\bigcup_{i=1}^{N} A_i\right) = S_1 - S_2' = \sum_{i=1}^{N} F(s^{(i)}) - \sum_{j=1}^{N-1} F(t^{(j)}), \tag{16}$$

where $t^{(j)}, j = 1, \ldots, N-1$ are the vertices of the incomparable pairs.

Quadratic pseudo-Boolean function representation

Corollary (The formula for the bivariate case)

On a finite poset in $\mathbb{R}^2$ with a random vector $X \in \mathbb{R}^2$, let $x_i = 1$ if $X \in A_i$ (otherwise, 0).

$$P\left(\bigcup_{i=1}^{N} A_i\right) = \sum_{i=1}^{N} c_i x_i - \sum_{(i,j)\in M^*} c_{ij} x_i x_j, \tag{17}$$

where M^* denotes the index set of the incomparable pairs (i.e., the neighboring maximal pairs); $c_i = p(x_i), c_{ij} = p(x_i, x_j)$.

A Desirable Partial Order Relation

We have seen so far that, for any incomparable orthants, as long as they have the same shape, we can use an efficient system for the union of such events in the bivariate case. However, such a desirable relation (i.e., $x_1 < \cdots < x_n$ and $y_1 > \cdots > y_n$) does not hold in $\mathbb{R}^n, n \geq 3$ in general.

Projection onto lower dimension

Suppose there are m distinct z_1 values (any other component can be chosen) from the orthants $A((z_{1_i}, z_{2_i}, z_{3_i})), i = 1, \ldots, N$. (i.e., $z_1^{(1)} < \cdots < z_1^{(m)}$). If n_j denotes the number of (z_2, z_3) pairs with $z_1^{(j)}, j = 1, \ldots, m$, then $N = n_1 + \cdots + n_m$.

Functional representation of the set $\{z^{(1)}, \ldots, z^{(N)}\}$

Suppose there are N sets, $A_i = A(z^{(i)}), i = 1, \ldots, N$. Let $A(z) = \{v \mid v \leq z\} \subset \mathbb{R}^n$. Then $A(z) \supset \Pi_{i=1}^{n} A(z^{(i)})$ if and only if $z \geq (\min_i(z_1^{(i)}), \ldots, \min_i(z_n^{(i)})), i = 1, \ldots, N$.

Theorem ([8])

If we have $A_1, \ldots, A_N \subset \mathbb{R}^n$ with $N > n$, then at least one of them contains the intersection of the others, i.e., for at least one i,

$$A_i \supset \Pi_{j \in \{1,\ldots,N\} \setminus \{i\}} A_j.$$

Proof.

The smallest one out of N orthants in $\mathbb{R}^n$ can be represented by at most n orthants. But we have $N > n$. □

System of distinct representatives (SDR)

Theorem ([8])

If we have $A_1, \ldots, A_N \subset \mathbb{R}^n$ with $N > n$, then the inclusion exclusion formula for the union of them will have the following terms only: singles, pairs, triples, ..., n-tuples, but not the entire members of their family.

Proof.

Any intersection of $A_1, \ldots, A_N \subset \mathbb{R}^n$ with $N > n$ can be expressed by n-tuples at most, i.e., there exists j such that $\prod_{j \in \{1,\ldots,n\}} A_j = \prod_{i \in \{1,\ldots,N\}} A_i$. □

Definition (System of distinct representatives (SDR))

Suppose that $A_1, A_2, \ldots, A_N$ are sets. The family of sets $A_1, A_2, \ldots, A_N$ has a system of distinct representatives (SDR) if and only if there exist distinct elements $z^{(1)}, z^{(2)}, \ldots, z^{(N)}$ such that $z^{(i)} \in A_i$ for each $i = 1, \ldots, N$.

First Duality for SDR

Theorem (Duality by [8])

The minimum number of non-redundant events in $A_{i_1}, \ldots, A_{i_r}$ is equal to the maximum number of distinct representatives in $z^{(i_1)}, \ldots, z^{(i_r)}$.

$\{A_1, A_2, \ldots, A_N\}$ is called a discrete convex set if none of the points $z^{(1)}, \ldots, z^{(N)}$ is in the relative interior of the convex hull of $\{z^{(1)}, \ldots, z^{(N)}\}$. Note that we can write, for the ith component of a vector $z \in \mathbb{R}^n$, $z_i = z_i(z_1, \ldots z_{i-1}, z_{i+1}, \ldots, z_n)$, $i = 1, \ldots, n$. Also note that the sets $(z_1, \ldots, z_n)$ and $(z_1, \ldots z_{i-1}, z_i(z_1, \ldots z_{i-1}, z_{i+1}, \ldots, z_n), z_{i+1}, \ldots, z_n)$ are the same.

Second Duality for SDR

Theorem (Second duality theorem for the case of a convex set $\{z^{(1)}, \ldots, z^{(N)}\}$ by [8])

Consider $z_n = z_n(z_1, \ldots, z_{n-1})$, or any other subscript instead of n. Consider the projection of the set $\{z^{(1)}, \ldots, z^{(N)}\}$ onto the space of $z_1, \ldots, z_{n-1}$. Then in $A_{i_1}, \ldots, A_{i_r}$ the event i_j is redundant if and only if $z^{(i_j)} \leq \sum_{k=1}^{r} \lambda_k z^{(i_k)}$, without the component $z_n^{(i_k)} = z_n^{(i_k)}(z_1^{(i_k)}, \ldots, z_{n-1}^{(i_k)})$, where $\lambda_k \geq 0, k = 1, \ldots, r$ and $\sum_{k=1}^{r} \lambda_k = 1$.

We can fully eliminate the redundant events by the use of the first duality theorem. The second theorem helps because we can look at sets in the space of $z_1, \ldots, z_{n-1}$ and if a z is inside the convex hull of such a set in the $n-1$-space, then we can eliminate it.

The general case in $\mathbb{R}^n, n \geq 3$

Let us introduce the following notations:

$$Z_i = \left\{ z_{(i)}^{(1)}, \ldots, z_{(i)}^{(m_{j_i})} \right\}, i = 1, \ldots, n, \tag{18}$$

$$Z_i^{(j_i)} = \left\{ z_{(i)} \mid z_{(i)} \leq z_{(i)}^{(j_i)} \right\}, j_i = 1, \ldots, m_{j_i}, \tag{19}$$

where the subscript (i) means the ith selected – one of the components of a vector $\mathbf{z} = (z_1, \ldots, z_n)$, excluding previously selected ones; the superscript (m_{j_i}) is the number of distinct values of the $z_{(i)}$. For example, $z_{(1)}$ designates the first selected component and there are m_{j_1} distinct values of it. If $z_{(1)} = z_n$ then $z_{(2)}$ can be selected from $(z_1, z_2, \ldots, z_{n-1}) \in \mathbb{R}^{n-1}$.
For the unions of projections, we need to find the SDRs for our desirable partial order relation.

The general case in $\mathbb{R}^n, n \geq 3$

First, we let $K_{n-i}^{(j)} \in \mathbb{R}^{n-i}$ denote the SDR of projection onto the space of one less dimension together with the value $z_i^{(j)}$, for $i = 1, \ldots, n$. (i.e., projection onto the space of $n - i$ dimension of the remaining components except $z_{(i)}$, after all previous projection operations). Then we make $K_{n-i}^{\prime(j)} \in \mathbb{R}^{n-i}$ by the sequential subtraction operation and find $H_{n-i}^{(j_i)}$ for all $j \geq j_i$ by the following:

Algorithm for $H_{n-i}^{(j_i)}$ by projection onto $\mathbb{R}^{n-i}$ from $\mathbb{R}^{n-i+1}$ with $z_i^{(j)}$

Step 1. Make sets $K_{n-i}^{(j)}$ = the SDR of projection with $z_i^{(j)}$

Step 2. By the sequential subtraction operation, find sets

$$K_{n-i}^{\prime(j_i)} = \text{ the SDR of } \bigcup_{j \geq j_i} K_{n-i}^{(j)} \tag{20}$$

Note that if there are m_j distinct $z_i^{(j)}$'s then we have $H_{n-i}^{(1)}, H_{n-i}^{(2)}, \ldots, H_{n-i}^{(m_j)}$.

The general case in $\mathbb{R}^n, n \geq 3$

The process (20) can be efficiently done by backwards from the last one to the first one.

$$\begin{aligned}
K'_m &= K_m \\
K'_{m-1} &= \text{the SDR of } K_m \cup K_{m-1} \\
K'_{m-2} &= \text{the SDR of } K_m \cup K_{m-1} \cup K_{m-2} \\
&\vdots \\
K'_2 &= \text{the SDR of } K_m \cup K_{m-1} \cup K_{m-2} \cup \cdots \cup K_3 \cup K_2 \\
K'_1 &= \text{the SDR of } K_m \cup K_{m-1} \cup K_{m-2} \cup \cdots \cup K_3 \cup K_2 \cup K_1,
\end{aligned} \tag{21}$$

The general case in $\mathbb{R}^n, n \geq 3$

which can be written equivalently as the following recursion

$$\begin{array}{lcl} K'_m & = & K_m \\ K'_{m-1} & = & \text{the SDR of } K'_m \cup K_{m-1} \\ K'_{m-2} & = & \text{the SDR of } K'_{m-1} \cup K_{m-2} \\ & \vdots & \\ K'_2 & = & \text{the SDR of } K'_3 \cup K_2 \\ K'_1 & = & \text{the SDR of } K'_2 \cup K_1. \end{array} \tag{22}$$

We can use (22) for (20) of the algorithm for our desirable partial order construction.

The general case in $\mathbb{R}^n, n \geq 3$

Let $H_{n-i}^{(k)} \subseteq \mathbb{R}^{n-i}$ denote the union of projections onto the space of one less dimension together with the value $z_{(i)}^{(k)}$ (after all previous projection operations). Then we can write the following:

$$\bigcup_{i=1}^{N} A_i = \bigcup_{j_1=1}^{m_{j_1}} H_{n-1}^{(j_1)} Z_1^{(j_1)}, \tag{23}$$

and

$$H_{n-(i-1)}^{(j_{i-1})} = \bigcup_{j_i=1}^{m_{j_i}} H_{n-i}^{(j_i)} Z_i^{(j_i)}, i = 2, \ldots, n-2, \tag{24}$$

where, in the RHS, m'_{j_i} is the number of distinct elements of Z_i as defined in (18). The above recursive equation goes all the way down to $H_2 \subseteq \mathbb{R}^2$. Thus, for the general case we have the following partial order relation, for $1 \leq k < l \leq m_{j_i}$,

$$z_{(i)}^{(k)} < z_{(i)}^{(l)} (\text{or, } Z_i^{(k)} \subset Z_i^{(l)}) \Rightarrow H_{n-i}^{(k)} \supseteq H_{n-i}^{(l)}. \tag{25}$$

The general case in $\mathbb{R}^n, n \geq 3$

Theorem (The formula for the union of orthants in $\mathbb{R}^n, n \geq 3$ by [8])

In any given probability space, on a finite poset in $\mathbb{R}^n$, $n \geq 3$ with the maximal elements $A_i = A(z^{(i)}) = \{z \in \mathbb{R}^n \mid z \leq z^{(i)}\}, i = 1, \ldots, N$, we have the formula:

$$P\left(\bigcup_{i=1}^{N} A_i\right) = P\left(Q_n\right), \tag{26}$$

where

$$\begin{array}{rcl} Q_n & = & Q(Z_1, Q'_{n-1}) \\ Q_{n-1} & = & Q(Z_2, Q'_{n-2}) \\ & \vdots & \\ Q_4 & = & Q(Z_{n-3}, Q'_3), \\ Q_3 & = & Q(Z_{n-2}, Q'_2), \\ Q_2 & = & Q(Z_{n-1}, Z_n), \end{array} \tag{27}$$

where we have the sequential subtraction operation ($'$) for the next recursion, and the function Q for the recursion is:

$$Q(A, B) = \sum_{k=1}^{r} A^{(k)} B^{(k)} - \sum_{k=1}^{r-1} A^{(k)} B^{(k+1)}, \tag{28}$$

where the sets A and B has our desirable partial order relation:

$$A^{(1)} \subset A^{(2)} \subset \cdots \subset A^{(r)} \text{ and } B^{(1)} \supseteq B^{(2)} \supseteq \cdots \supseteq B^{(r)}.$$

The nested representation for the union of events

We can also write Theorem equivalently, without a recursion function Q as the following:

Corollary (The nested representation for the union of events in $\mathbb{R}^n$)

In any given probability space, on a finite poset in $\mathbb{R}^n, n \geq 3$ with the maximal elements $A_i = A(z^{(i)}) = \{z \in R^n \mid z \leq z^{(i)}\}, i = 1, \ldots, N$, we have:

$$\bigcup_{i=1}^{N} A_i = \bigcup_{j_1=1}^{m_{j_1}} \left(\bigcup_{j_2=1}^{m_{j_2}} \left(\bigcup_{j_3=1}^{m_{j_3}} \left(\cdots \bigcup_{j_{n-3}=1}^{m_{j_{n-3}}} \left(\bigcup_{j_{n-2}=1}^{m_{j_{n-2}}} H_{(j_{n-2})} z_{n-2}^{(j_{n-2})} \right)'_{(j_{n-3})} z_{n-3}^{(j_{n-3})} \cdots \right)'_{(j_3)} z_3^{(j_3)} \right)'_{(j_2)} z_2^{(j_2)} \right)'_{(j_1)} z_1^{(j_1)}, \tag{29}$$

where $(')$ designates the sequential subtraction operation, and H is the unions as defined in (24).

Time complexity of the modified inclusion-exclusion formulas

If the number of events is N, it takes $O(2^N)$ to calculate the probability of union by the original inclusion-exclusion formula. For our events in $\mathbb{R}^n, n \geq 3$, in general, the worst case time complexity of solving the formula is $O(n^2N^2)$ for a sorted data set and $O(n^2N^2 + n \cdot N \log N)$ for unsorted data set. Note that it takes $O(N \log N)$ for the bivariate case for unsorted data, a linear time otherwise.

General form of metrics for discrimination and bias

Definition (General form of metrics for discrimination and bias)

Generally, metrics for discrimination and bias compare the outcomes across groups, e.g., the difference of some outcome between the protected and general groups. The difference is typically represented by expected value or probability. Therefore, we can write the following:

$$E(\psi(X) \mid X \in G_a) - E(\psi(X) \mid X \in G_b), \tag{30}$$

which can also measure the difference between probabilities when ψ is a function of binary vector. Note that G_a can be a favorable (or unfavorable) event as a protected group and G_b is a group to compare with any given protected groups.

General form of metrics for discrimination and bias

Let us introduce a fractional vector $\mathbf{u} \in \mathbb{U}^n$, i.e., $\mathbf{u} = (u_1, \ldots, u_n)$, where $u_i \in [0, 1], \forall i$. Using the following

$$x_i = \begin{cases} 1 & \text{if } X_i \in A_i \\ 0 & \text{otherwise,} \end{cases} \tag{31}$$

we write $P(x_i = 1) = P(X_i \in A_i) = u_i$, $P(x_i = 0) = P(X_i \in A_i^c) = 1 - u_i$.

Theorem (Pseudo-Boolean function for probability difference)

Let $G_a \subseteq \mathbb{U}^n$. Let $X \in \mathbb{R}^n$ be a random vector. If we have a mapping $h : \mathbb{R}^n \to \mathbb{B}^n$ with $\mathbf{x} = h(X)$, we have:

$$E\left(\frac{\partial f(\mathbf{x})}{\partial x_i}\right) = P(X_i \in G_{a_i}) - P(X_i \notin G_{a_i}), \tag{32}$$

where

$$f(\mathbf{x}) = \mathbf{1}^T \mathbf{x}. \tag{33}$$

- Section 4. Sequential time series data aggregation using a suitable matching algorithm

Sequential pairwise aggregation based on the correlation, distance, pseudo-Boolean function

For our new aggregation scheme for time series data, we quantify similarity between the time series by the following: i) correlation; ii) Euclidean distance; iii) fairness metrics based on a pseudo-Boolean function.

Sequential pairwise aggregation based on the correlation, distance, pseudo-Boolean function

Suppose that we are given N time series data X_i for $i = 1, \ldots, N$. Let $A = \{X_1, \ldots, X_N\}$ be the set of such given time series data. W.l.o.g, we let each time series vector X_i be m-dimensional column vector (i.e., $X_i \in \mathbb{R}^{m \times 1}$). Then A is a set of m-vectors and we typically have $m \ll N$. The set A can be represented, using m observations from each of the N time series, as an $m \times N$ data matrix X (i.e., N columns of m-vectors). We write the following:

$$X = [X_1, X_2, \ldots, X_N] = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mN} \end{pmatrix}, \tag{34}$$

where a column vector $X_i \in \mathbb{R}^m$ for $i = 1, \ldots N$.

Sequential pairwise aggregation based on the correlation, distance, pseudo-Boolean function

First, for efficient correlation calculation, let Y be an $m \times N$ mean-centered data matrix, where its (i,j) entries are $y_{ij} = x_{ij} - e_i$, where e_i denotes the sample mean of time series i, $i = 1, \ldots, N$ (i.e., $e_i = \frac{1}{N} \sum_{j=1}^{m} x_{ij}$). Then we can write the following for an $N \times N$ sample covariance matrix:

$$C = \frac{1}{N-1} Y^T Y, \tag{35}$$

rescaled to a correlation matrix (with zero diagonals) by

$$R = BCB - I, \tag{36}$$

where B is a diagonal matrix $B = \text{diag}[1/\sigma_1, \ldots, 1/\sigma_N]$. The entries $r_{ij} = \sigma_{ij}/(\sigma_i \sigma_j)$ and $r_{ii} = 0$. Note that we subtract an $N \times N$ identity matrix from the original correlation matrix calculation BCB.

Sequential pairwise aggregation based on the correlation, distance, pseudo-Boolean function

Second, for more efficient calculation for Euclidean distance between time series, let us introduce an Euclidean distance matrix D using the squared distances. Euclidean distance matrix would be useful in many aspects, e.g., more compact representation, efficient computation by matrix transformation and factorization, etc. We refer the readers to [9], [10], etc. The squared distance between the vectors X_i and X_j is given by $d_{i,j} = \|X_i - X_j\|^2$, and $d_{i,j}$ is the (i,j)th entry of the matrix D. Hence D is symmetric. (Also note that $\text{rank}(D) \leq m + 2$.) For our applications, the vector $X_j \in \mathbb{R}^m$, $j = 1, \ldots, N$ where $m \ll N$, typically. Suppose that we are given a data matrix X as in (34) (i.e., $X \in \mathbb{R}^{N \times m}$).

Sequential pairwise aggregation based on the correlation, distance, pseudo-Boolean function

Note that we can find the distance matrix D using the fact that $\|X_i - X_j\|^2 = (X_i - X_j)^T(X_i - X_j) = X_i^T X_i - X_i^T X_j - X_j^T X_i + X_j^T X_j$, which can be written up as the following:

$$D = \mathbf{1} diag(X^T X)^T - 2X^T X + diag(X^T X)\mathbf{1}^T, \tag{37}$$

where the symbol $\mathbf{1}$ denotes a column vector of N ones. It is easy to see that the column vector $diag(X^T X) = (\|x_1\|^2, \ldots, \|x_J\|^2)^T$. Let d_j denote the jth column of D. Then we can write the $N \times N$ distance matrix $D = [d_1, \ldots, d_N]$, where $d_j \in \mathbb{R}^N$.

Sequential pairwise aggregation based on the correlation, distance, pseudo-Boolean function

Third, for the fairness metric calculation, we use the following pseudo-Boolean function. For (i, j) pair, we let $h_{ij} : (\mathbb{R}^n, \mathbb{R}^n) \to \mathbb{B}^n$ with $\mathbf{y} = h_{ij}(X_i, X_j)$. For each component k of $\mathbf{y}$, with a small positive constant ϵ, we write the following:

$$y_k = \begin{cases} 1 & \text{if } \left|P(A_k \mid X_i) - P(A_k \mid X_j)\right| \leq \epsilon \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

For the pair of (i, j) time series, we have the following pseudo-Boolean function:

$$f_{ij}(\mathbf{y}) = c_0 + \sum_{k=1}^{n} c_k y_k + \sum_{1 \leq k < l \leq n} c_{kl} y_k y_l. \tag{39}$$

Sequential pairwise aggregation based on the correlation, distance, pseudo-Boolean function

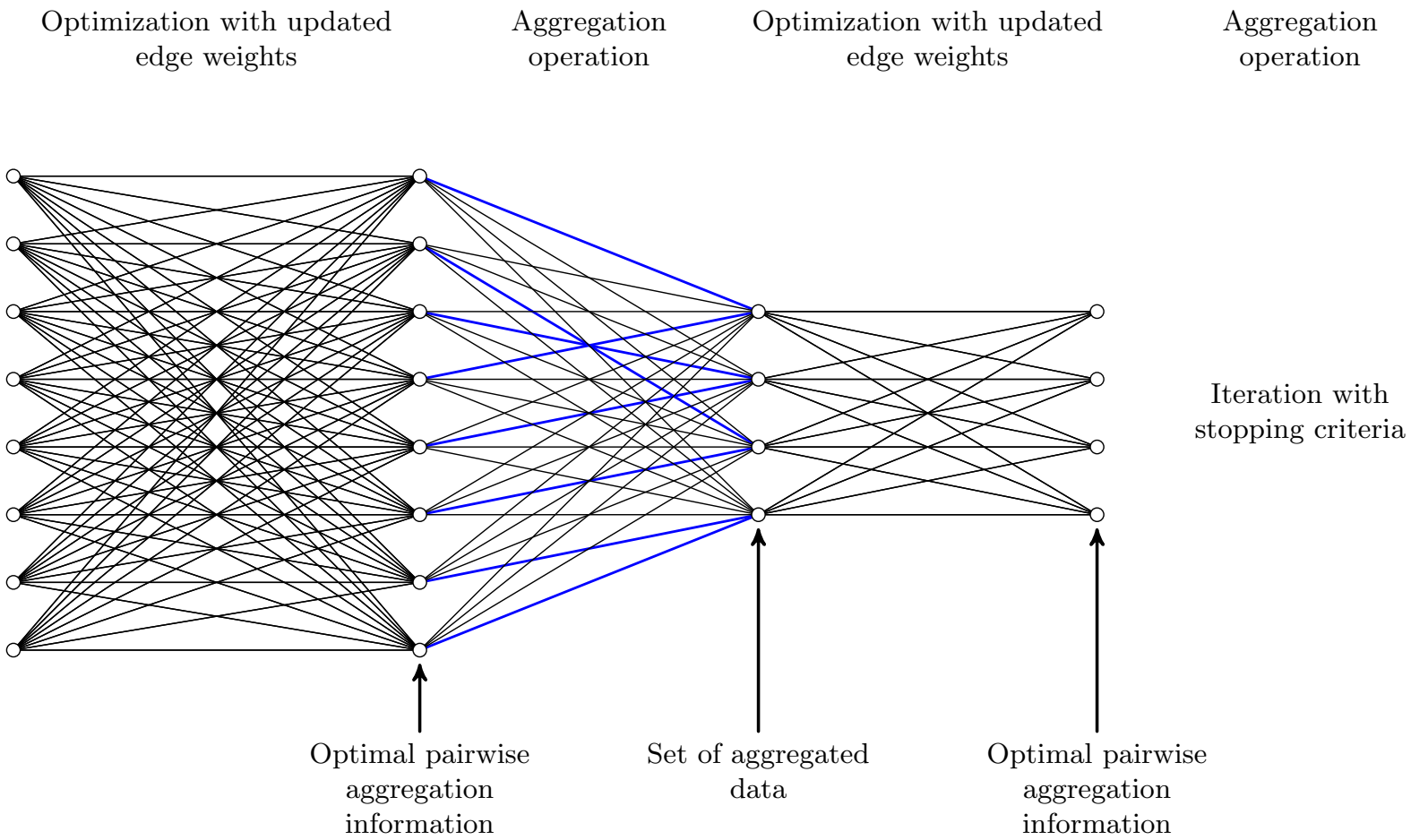

Figure 2: Description of the sequential aggregation process

Sequential pairwise aggregation based on the correlation, distance, pseudo-Boolean function

Let us consider given time series vectors as nodes and their pairwise relationships as edge weights. Then we can draw a graph $G = (V, E)$, which is a complete graph K_N, where we have N nodes and $N(N-1)/2$ edges (hence the total number of pairs of N nodes). We then let w_{ij} denote the edge weight between nodes i and j. The edge weight w_{ij} is the linear combination of (i,j) entry of R in (36) and that of D in (37) for integers $1 \leq i, j \leq N$, together with some suitable quadratic pseudo-Boolean function f as the following:

$$w_{ij} = \alpha_1 r_{ij} + \alpha_2 d_{ij} + \alpha_3 f_{ij}, \tag{40}$$

where $\sum_k \alpha_k = 1, \alpha_k \geq 0, k = 1,2,3.$

Sequential pairwise aggregation based on the correlation, distance, pseudo-Boolean function

In order to cover all the nodes and to find their best pairs (if there is any), we solve a perfect matching problem on the graph. Based on the graph K_N, we can construct a bipartite graph with two identical partitions $V_1 = V_2 = \{X_1, \ldots, X_N\}$, that is, $G = (V, E)$ and $V = V_1 \cup V_2$ and $E \subseteq \{v_1 v_2 : v_1 \in V_1, v_2 \in V_2\}$. Then, we can write the following maximum weight perfect matching problem:

$$\begin{aligned} &\max \textstyle\sum_{e \in E} w_e x_e \\ &\text{subject to} \\ &x(\delta(v_1)) = 1, \text{ for all } v_1 \in V_1 \\ &x_e \in \{0,1\}, \text{ for all } e \in E, \end{aligned} \tag{41}$$

where $\delta(v_1) = \{v_1 v_2 : v_1 v_2 \in E, v_1 \in V_1, v_2 \notin V_1\}$.

Sequential pairwise aggregation based on the correlation and distance

Equivalently, (41) can be written up as the following.

$$\begin{aligned} &\max \textstyle\sum_{v_1,v_2} w_{v_1v_2} x_{v_1v_2} \\ &\text{subject to} \\ &\sum_{v_1 \in V_1} x_{v_1v_2} = 1 \\ &\sum_{v_2 \in V_2} x_{v_1v_2} = 1 \\ &x_{v_1v_2} \in \{0,1\}. \end{aligned} \tag{42}$$

With optimal solution, in case of a perfect matching, time series aggregation operation can be done by a simple multiplication using an $N/2 \times N$ matrix regardless of the dimension. In general, we have a mapping from $\mathbb{R}^{N \times m}$ to $\mathbb{R}^{2/N \times m}$.

Sequential pairwise aggregation based on the correlation and distance

The aggregated data matrix can be found by the following multiplication.

$$X^{(1)} = \begin{pmatrix} X_1 + X_2 \\ X_3 + X_7 \\ X_4 + X_6 \\ X_5 + X_8 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_8 \end{pmatrix} = S^{(1)} X, \qquad (43)$$

where the matrix S can be obtained by the optimal solution from the model (42).

Zero-one matrix $S^{(k)}$ construction at the kth iteration

Require: Let i' denote the i'th row of the matrix $S^{(k)}$. Optimal solution x_{ij} of (41) based on the current original or aggregated data matrix and its correlation matrix.

Set $i = 1$ and $i' = 1$.

while $i \leq i'$ **do**

  if $x_{ij} = 1$ for any $i \leq j$ **then**

    $S_{i'i} = S_{i'j} = 1$ and set all the other columns to be zero.

  else

    Replace i by $i + 1$

  end if

  Replace i' by $i' + 1$.

end while

Sequential pairwise aggregation based on the correlation and distance

The aggregated data set $X^{(1)}$ can be written up as the following:

$$A^{(1)} = \{(X_1 + X_2), (X_3 + X_7), (X_4 + X_6), (X_5 + X_8)\}, \tag{44}$$

where the superscript (i) denotes the ith pairwise aggregation.
We assume that the optimal solution from the next iteration is $x_{12} = 1, x_{34} = 1$ with all others being zero. Then we get the following

$$X^{(2)} = \begin{pmatrix} (X_1 + X_2) + (X_3 + X_7) \\ (X_4 + X_6) + (X_5 + X_8) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} X^{(1)} = S^{(2)} X, \tag{45}$$

and equivalently, we have:

$$A^{(2)} = \{X_1 + X_2 + X_3 + X_7, X_4 + X_5 + X_6 + X_8\}. \tag{46}$$

Thank you. Questions or comments?

References I

College Factual, Statistics at University of California - Davis (2021). URL https://www.collegefactual.com/colleges/university-of-california-davis/academic-life/academic-majors/mathematics-and-statistics/mathematics-and-statistics/

P. L. Hammer, Some network flow problems solved with pseudo-boolean programming, Oper. Res 13 (1) (1965) 388–399.

P. L. Hammer, P. Hansen, B. Simeone, Roof duality, complementation and persistency inquadratic 0–1 optimization, Math. Programming 28 (1) (1984) 121–155.

E. Boros, Y. Crama, P. Hammer, Upper-bounds for quadratic 0–1 maximization, Operations Research Letters 9 (2) (1990) 73–79.

References II

- E. Boros, P. L. Hammer, Cut-polytopes, boolean quadric polytopes and nonnegative quadratic pseudo-boolean functions, Mathematics of Operations Research 18 (1) (1993) 245–253.
- UC System Infocenter, Disaggregated data (2021). URL `https://www.universityofcalifornia.edu/infocenter/disaggregated-data`
- J. Lee, A. Prékopa, On the probability of union in the n-space, Oper. Res. Lett 45 (1) (2017) 19–24.
- J. Lee, Computing the probability of union in the n-dimensional Euclidean space for application of the multivariate quantile: p-level efficient points, Oper. Res. Lett. 45 (3) (2017) 242–247.
- L. Liberti, C. Lavor, N. Maculan, A. Mucherino, Euclidean Distance Geometry and Applications, SIAM Review 56 (2014) 3–69.

References III

I. Dokmanić, R. Parhizkar, J. Ranieri, M. Vetterli, Euclidean Distance Matrices: Essential Theory, Algorithms and Applications, IEEE Signal Processing Magazine 32 (2015) 12–30.