Noise Sensitivity on Equivalence Relations

Omar Elshinawy [‡]

A Thesis submitted in partial fulfillment of the requirements for the degree of

Bachelor of Science in Mathematics

May 20, 2025, Bremen

Abstract

Noise sensitivity was first introduced by Benjamini, Kalai and Schramm in their seminal work on Boolean functions. We propose a construction of a similar taste on binary relations. By *flipping* every relation with a small probability p, a natural question on recoverability arises, to which give a positive answer in the case of an equivalence relation $(X, \sim)$. We prove that equivalence relations are *noise-stable* under the prescribed model. In particular, we propose a simple reconstruction algorithm, and show that it achieves an asymptotically zero misclassification error.

[‡] Constructor University | School of Computer Science and Engineering
Supervised by Prof. Dr. Keivan Mallahi Karai

Statutory Declaration

Family Name, First Name	Elshinawy, Omar Mohamed Hasan Mostafa
Matriculation number	30006525
Kind of thesis submitted	Bachelor Thesis

Declaration of Authorship

I hereby declare that the thesis submitted was created and written solely by myself without any external support. Any sources, direct or indirect, are marked as such. I am aware of the fact that the contents of the thesis in digital form may be revised with regard to usage of unauthorized aid as well as whether the whole or parts of it may be identified as plagiarism. I do agree my work to be entered into a database for it to be compared with existing sources, where it will remain in order to enable further comparisons with future theses. This does not grant any rights of reproduction and usage, however.

This document was neither presented to any other examination board nor has it been published.

Erklärung der Autorenschaft (Urheberschaft)

Ich erkläre hiermit, dass die vorliegende Arbeit ohne fremde Hilfe ausschließlich von mir erstellt und geschrieben worden ist. Jedwede verwendeten Quellen, direkter oder indirekter Art, sind als solche kenntlich gemacht worden. Mir ist die Tatsache bewusst, dass der Inhalt der Thesis in digitaler Form geprüft werden kann im Hinblick darauf, ob es sich ganz oder in Teilen um ein Plagiat handelt. Ich bin damit einverstanden, dass meine Arbeit in einer Datenbank eingegeben werden kann, um mit bereits bestehenden Quellen verglichen zu werden und dort auch verbleibt, um mit zukünftigen Arbeiten verglichen werden zu können. Dies berechtigt jedoch nicht zur Verwendung oder Vervielfältigung.

Diese Arbeit wurde noch keiner anderen Prüfungsbehörde vorgelegt noch wurde sie bisher veröffentlicht.

20.05.2025, Omar Elshinawy

Date, Signature

Contents

1 Noise Sensitivity

1.1 On Boolean Functions

The notion of Noise sensitivity was first introduced by Benjamini, Kalai and Schramm [1] in their seminal work on Boolean functions. Given a string x of $n-$many bits, set $\tilde{x}$ to be the string that disagrees with x on the i-th bit with a small probability p. Here, $\tilde{x}$ is seen as a random perturbation of x. The *noise sensitivity*[†] of f is then measured by

$$\mathrm{NS}_p(f) := \mathbb{P}[f(\tilde{x}) \neq f(x)]$$

which captures how much noise is needed to perturb the output. We then say that f is *noise sensitive* if there exists a $p \in (0,1)$ for which

$$\lim_{n\to\infty} \mathrm{NS}_p(f) = \frac{1}{2}.$$

It is *noise insensitive* if the same limit converges to $(0, \frac{1}{2})$, and *noise stable* if for vanishingly small p, the limit converges to zero uniformly in n. For instance, it is not difficult to believe that the parity function is a great example of a *noise sensitive* boolean function, whereas the majority function is *noise stable* for vanishingly small perturbations. It is important to note that convergence of $\mathrm{NS}_p(f)$ to $\frac{1}{2}$ encodes the idea that the odds are no better than a random output, that is either zero or one with probability $\frac{1}{2}$. In that respect, f is sensitive.

1.2 On Binary Relations

We propose a construction of a similar taste on binary relations. For every pair $i, j \in X$, let X_{ij} be a Bernoulli random variable with parameter p. Next, require that

$$i \sim' j \iff \begin{cases} i \sim j \ \wedge \ X_{ij} = 0 \\ i \not\sim j \ \wedge \ X_{ij} = 1 \end{cases} .$$

In a new *noisy* relation $\sim'$, we eliminate pairs $i \sim j$ with $X_{ij} = 1$ and introduce $i \not\sim j$ with $X_{ij} = 1$. The challenge here is that, unlike on boolean functions, only one flip is required to yield a completely new relation, making it impractical to directly compare $\sim'$ to $\sim$. To address this, we define a recovery-based notion of stability,

$$\mathrm{NS}(\sim) := \inf_{\mathbb{A}} \ \mathbb{P}[\mathbb{A}(\sim') \neq (\sim)]$$

where $\mathbb{A}$ is understood to be a recovery algorithm. It is important to note that if $|X| = n$, an order $\Theta(1)$ of noise will asymptotically destroy all structure, as one expects $\Theta(n^2)$ flips on all pairs $(i,j) \in X^2$. Setting $p = \Theta(1)$ therefore not only trivialises the theory, it goes against the original idea of small perturbations. For these reasons we would like to relax the noise condition.

Definition 1.1 (Noise Sensitivity on Binary Relations)**.** Let $\sim$ be a binary relation on X, and set $p = \mathcal{O}(1/n)$. Next, define

$$L := \lim_{|X|\to\infty} \mathrm{NS}(\sim).$$

1. If $L = 0$, then $\sim$ is *noise stable.*[‡]
2. If $L \in (0,1)$, and $\sim$ is not *noise stable*, then $\sim$ is *noise insensitive.*
3. If $L = 1$, then $\sim$ is *noise sensitive.*

Remark. $\sim$ is noise sensitive, then any algorithm almost surely fails, and is no better than a random prediction. This is consistent with the boolean setting, with the subtle difference that the target space of binary relations is infinite. A random prediction is therefore incorrect with probability 1, and not $\frac{1}{2}$ as observed earlier.

The following discussion studies noise sensitivity on an equivalence relation, and motivates a recovery algorithm that utilises its natural properties.

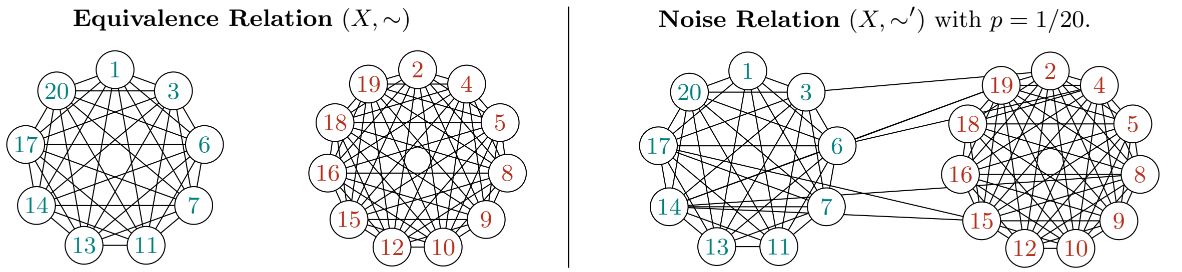
[1] *Noise Sensitivity of Boolean Functions and Applications to Percolation.*

[†]While this is not the original definition, one may show that it is equivalent to *asymptotic noise sensitivity* in [1].

[‡]The uniformity condition is satisfied, as an algorithm $\mathbb{A}$ that tolerates a noise rate p also tolerates $p' < p$.

1.3 On Equivalence Relations

One starts with an equivalence relation $(X, \sim)$ of size n. Here, we apply noise *symmetrically*[*] to all distinct pairs $i, j \in X$. To give an example,



N.B: *Reflexivity is not reflected in this graph.*

Central to our discussion is the question on recoverability. One may concede that it is quite tempting to simply conclude that X_1 and X_2 are two separate equivalence classes; however, this intuition needs be formalised.

The focus then turns to the transitive encoding of $\sim$, which is lost with probability one. All transitivity, however, is not lost! *Much* of the transitivity survives under noise. The first question this thesis answers is that of quantifying how *much* is lost.

Definition 1.2 (Score)**.** For $i, j \in X$, define

$$s(i,j) = \sum_{w \in X \setminus \{i,j\}} \mathbb{1}[i \sim w] \cdot \mathbb{1}[w \sim j]$$

to be the count of neighbours of w that witness the relation of i, j.

Remark. This natural measure pushes the transitive property to its full length. It is not difficult to see that for members $i \sim j$ there are $n_\alpha - 2$ other witnesses $w \in X_\alpha$, with n_α the size of X_α. If $i \not\sim j$, then the score is simply null.

Definition 1.3 (Score Matrix)**.** Define

$$S_\sim := \begin{pmatrix} s(1,1) & s(1,2) & \dots & s(1,n) \\ s(2,1) & s(2,2) & \dots & s(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ s(n,1) & \dots & s(n,n-1) & s(n,n) \end{pmatrix}$$

to store the scores $s(i,j)$ for every $i, j \in X$.

For the previous example, $S_\sim$ is just

$$\begin{array}{cccccccccccccccccccc} \underline{1} & \underline{2} & \underline{3} & \underline{4} & \underline{5} & \underline{6} & \underline{7} & \underline{8} & \underline{9} & \underline{10} & \underline{11} & \underline{12} & \underline{13} & \underline{14} & \underline{15} & \underline{16} & \underline{17} & \underline{18} & \underline{19} & \underline{20} \end{array}$$
$$\begin{pmatrix} 8 & 0 & 7 & 0 & 0 & 7 & 7 & 0 & 0 & 0 & 7 & 0 & 7 & 7 & 0 & 0 & 7 & 0 & 0 & 7 \\ 0 & 10 & 0 & 9 & 9 & 0 & 0 & 9 & 9 & 9 & 0 & 9 & 0 & 0 & 9 & 9 & 0 & 9 & 9 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}_{20 \times 20} \begin{array}{l} |1 \\ |2 \\ \\ \end{array}$$

whereas a noisy $S_{\sim'}$ is given by

$$\begin{array}{cccccccccccccccccccc} \underline{1} & \underline{2} & \underline{3} & \underline{4} & \underline{5} & \underline{6} & \underline{7} & \underline{8} & \underline{9} & \underline{10} & \underline{11} & \underline{12} & \underline{13} & \underline{14} & \underline{15} & \underline{16} & \underline{17} & \underline{18} & \underline{19} & \underline{20} \end{array}$$
$$\begin{pmatrix} 7 & 1 & 6 & 1 & 0 & 6 & 6 & 0 & 0 & 0 & 6 & 0 & 6 & 7 & 1 & 0 & 6 & 0 & 1 & 6 \\ 1 & 11 & 0 & 9 & 9 & 3 & 1 & 9 & 9 & 9 & 1 & 8 & 1 & 4 & 9 & 9 & 2 & 8 & 9 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}_{20 \times 20} \begin{array}{l} |1 \\ |2 \\ \\ \end{array}$$

Fixing $x \in X$, observe the significant *gap* between scores of members $i \sim x$ and those of non-members $j \not\sim x$.

[*] in the sense that $X_{ij} = X_{ji}$.

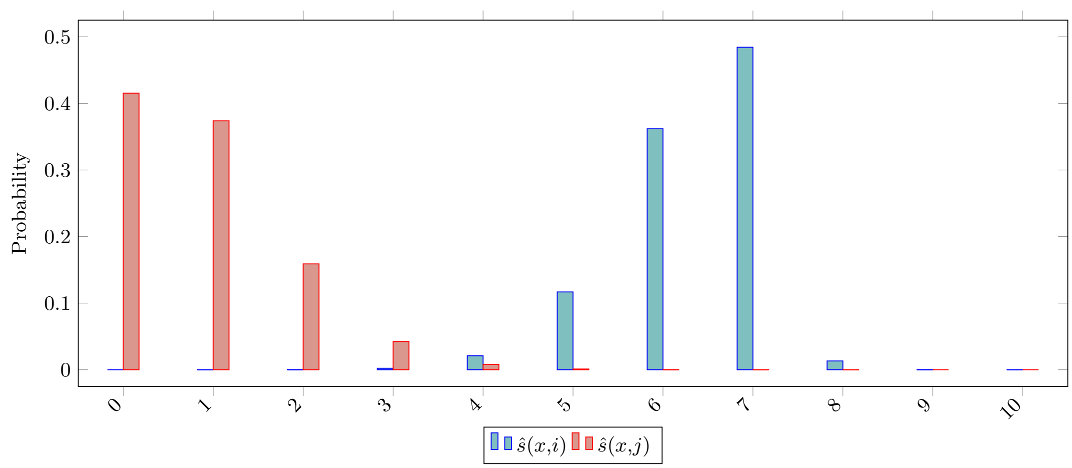
For instance, sorting the first row recovers members $i \in X_1$ completely. Not only that, but it naturally recommends 2 to sieve out its members from the set of possible $j \in X_2$.

$$\begin{matrix} \underline{5} & \underline{8} & \underline{9} & \underline{10} & \underline{12} & \underline{16} & \underline{18} & \underline{2} & \underline{4} & \underline{15} & \underline{19} & \underline{3} & \underline{6} & \underline{7} & \underline{11} & \underline{13} & \underline{17} & \underline{20} & \underline{1} & \underline{14} \end{matrix}$$
$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 7 & 7 \\ 9 & 9 & 9 & 9 & 8 & 9 & 8 & 11 & 9 & 9 & 9 & 0 & 3 & 1 & 1 & 1 & 2 & 1 & 1 & 4 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}_{20\times 20} \begin{matrix} |\,1 \\ |\,2 \\ \\ \end{matrix}$$

This seems to be no isolated case, but rather a global phenomena. Sorting each row ascendingly, one may observe the sharp jumps in scores.

$$\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \mathbf{6} & 6 & 6 & 6 & 6 & 6 & 6 & 7 & 7 \\
0 & 1 & 1 & 1 & 1 & 1 & 2 & 3 & 4 & \mathbf{8} & 8 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 11 \\
0 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 3 & 3 & \mathbf{6} & 6 & 7 & 7 & 7 & 7 & 7 & 7 & 9 \\
1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 4 & \mathbf{8} & 8 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 10 & 11 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 2 & 3 & \mathbf{8} & 8 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 10 \\
1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 3 & 3 & 4 & \mathbf{6} & 7 & 7 & 7 & 7 & 7 & 7 & 7 & 10 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & \mathbf{6} & 6 & 7 & 7 & 7 & 7 & 7 & 7 & 8 \\
0 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 3 & \mathbf{8} & 8 & 9 & 9 & 9 & 9 & 9 & 9 & 10 & 10 & 11 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 2 & 3 & \mathbf{8} & 8 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 10 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 2 & 3 & \mathbf{8} & 8 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 10 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & \mathbf{6} & 6 & 7 & 7 & 7 & 7 & 7 & 7 & 8 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 2 & 3 & \mathbf{8} & 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 & 9 & 9 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & \mathbf{6} & 6 & 7 & 7 & 7 & 7 & 7 & 7 & 8 \\
2 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 4 & 4 & \mathbf{6} & 6 & 6 & 6 & 6 & 7 & 7 & 7 & 10 \\
1 & 1 & 2 & 2 & 2 & 2 & 3 & 3 & 4 & 8 & 8 & 9 & 9 & 9 & 9 & 9 & 9 & 10 & 10 & 12 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 2 & 3 & \mathbf{8} & 8 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 9 & 10 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 3 & \mathbf{6} & 7 & 7 & 7 & 7 & 7 & 7 & 7 & 9 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 2 & 3 & \mathbf{8} & 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 & 9 & 9 \\
1 & 2 & 2 & 2 & 2 & 2 & 3 & 3 & 3 & \mathbf{8} & 8 & 9 & 9 & 9 & 9 & 9 & 10 & 10 & 10 & 12 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 2 & 2 & \mathbf{6} & 6 & 7 & 7 & 7 & 7 & 7 & 7 & 8
\end{pmatrix}
\begin{matrix} |\,1 \\ |\,2 \\ |\,3 \\ |\,4 \\ |\,5 \\ |\,6 \\ |\,7 \\ |\,8 \\ |\,9 \\ |\,10 \\ |\,11 \\ |\,12 \\ |\,13 \\ |\,14 \\ |\,15 \\ |\,16 \\ |\,17 \\ |\,18 \\ |\,19 \\ |\,20 \end{matrix}$$

Fixing $x \in X_1$, one finds the score distribution† for *noisy* scores $\hat{s}(x, i)$ and $\hat{s}(x, j)$ below. Here, $x \sim i$ but $x \not\sim j$. The key observation is that the distributions of $\hat{s}(x, i)$ and $\hat{s}(x, j)$ are *almost* distinguishable.



If we may possibly distinguish between the two scores via some *score gap*, then one only needs to compute the scores to determine membership. This is a rather exciting prospect, whose feasibility we shall explore in the following discussion.

2 Formalism

For this discussion, fix $x, i \in X_i$ to be members of the same class, and $j \in X_j$ to be a member of another class. If $\hat{\xi}_{\text{in}} := \hat{s}(x, i)$ is and $\hat{\xi}_{\text{out}} = \hat{s}(x, j)$ are the perturbed scores of i, j with respect to x, set

$$\hat{\xi} := \hat{\xi}_{\text{in}} - \hat{\xi}_{\text{out}}$$

to be the random variable that captures the score gap. Let us prove a few interesting results.

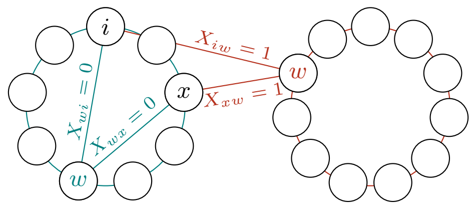
†Both random variables follow a Poisson Binomial distribution, see (**2.1**), (**2.2**).

2.1 The Score Gap

Lemma 2.1. *The random variable* $\hat{\xi}_{in}$ *is given by*

$$\tilde{s}(x,i) = \sum_{w \in X_i \setminus \{x,i\}} (1 - X_{xw}) \cdot (1 - X_{wi}) + \sum_{w \notin X_i} X_{xw} \cdot X_{wi}.$$

Argument. A quick justification goes as follows. If $w \in X_i$, then $x \sim w \ \wedge \ w \sim i$. This relation survives in $\sim'$ if the relations do not *flip*, meaning $X_{xw} = X_{wi} = 0$. On the other hand, $w \notin X_i$ implies that w is not connected to either elements, and both relations must *flip* to get a contribution.

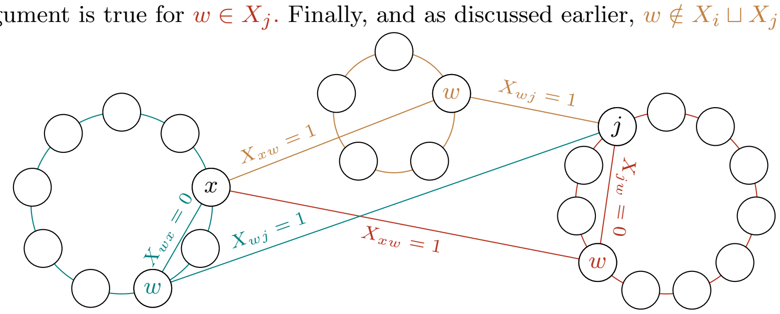


$\boxed{\xi}$

Lemma 2.2. *The random variable* $\hat{\xi}_{out}$ *is given by*

$$\tilde{s}(x,j) = \sum_{w \in X_i \setminus \{x\}} (1 - X_{xw}) \cdot X_{wj} + \sum_{w \in X_j \setminus \{j\}} X_{1w} \cdot (1 - X_{wj}) + \sum_{w \notin X_i \sqcup X_j} X_{xw} \cdot X_{wj}$$

Argument. $w \in X_i$ implies that the inner-connection X_{wx} must not flip, while the outer-connection X_{wj} must flip. The same argument is true for $w \in X_j$. Finally, and as discussed earlier, $w \notin X_i \sqcup X_j$ costs two flips.



$\boxed{\xi}$

Lemma 2.3. *The gap* $\hat{\xi} := \hat{\xi}_{in} - \hat{\xi}_{out}$ *is given by*

$$(X_{xi} + X_{xj} - 1) \cdot X_{ij} + \sum_{w \in X_i \setminus \{x,i\}} (1 - X_{xw}) \cdot (1 - X_{wi} - X_{wj}) + \sum_{w \in X_j \setminus \{j\}} X_{xw} \cdot (X_{wi} + X_{wj} - 1) + \sum_{w \notin X_i \sqcup X_j} X_{xw} \cdot (X_{iw} - X_{jw})$$

that is, a sum of independent random variables. Furthermore, if $\xi := \mathbb{E}[\hat{\xi}]$ *is the expected gap, then*

$$\xi = (n_i - 2) - (3n_i + n_j - 6) \cdot p + (2n_i + 2n_j - 4) \cdot p^2 = \mathcal{O}(n_i).$$

Computation. Using the previous two computations, we write

$$\hat{\xi} = \hat{\xi}_{\text{in}} - \hat{\xi}_{\text{out}} = \overbrace{\sum_{w \in X_i \setminus \{x,i\}} (1 - X_{xw}) \cdot (1 - X_{wi})}^{A} + \overbrace{\sum_{w \notin X_i} X_{xw} \cdot X_{wi}}^{B} \quad \text{(by 2.1)}$$

$$\underbrace{- \sum_{w \in X_i \setminus \{x\}} (1 - X_{xw}) \cdot X_{wj}}_{C} \ \underbrace{- \sum_{w \in X_j \setminus \{j\}} X_{xw} \cdot (1 - X_{wj})}_{D} \ \underbrace{- \sum_{w \notin X_i \sqcup X_j} X_{xw} \cdot X_{wj}}_{E}. \quad \text{(by 2.2)}$$

To obtain a sum of independent random variables, one would like to rewrite this expression over disjoint sets.

$$\begin{aligned} A - C :=& \sum_{w \in X_i \setminus \{x,i\}} (1 - X_{xw}) \cdot (1 - X_{wi}) - \sum_{w \in X_i \setminus \{x\}} (1 - X_{xw}) \cdot X_{wj} \\ =& (X_{xi} - 1) \cdot X_{ij} + \sum_{w \in X_i \setminus \{x,i\}} (1 - X_{xw}) \cdot (1 - X_{wi} - X_{wj}) \quad \text{(a)} \\ B - E - D :=& \sum_{w \notin X_i} X_{xw} \cdot X_{wi} - \sum_{w \notin X_i \sqcup X_j} X_{xw} \cdot X_{wj} - \sum_{w \in X_j \setminus \{j\}} X_{xw} \cdot (1 - X_{wj}) \\ =& \sum_{w \in X_j} X_{xw} \cdot X_{wi} + \sum_{w \notin X_i \sqcup X_j} X_{xw} \cdot (X_{wi} - X_{wj}) + \sum_{w \in X_j \setminus \{j\}} X_{xw} \cdot (X_{wj} - 1) \\ =& X_{xj} \cdot X_{ji} + \sum_{w \notin X_i \sqcup X_j} X_{xw} \cdot (X_{iw} - X_{jw}) + \sum_{w \in X_j \setminus \{j\}} X_{xw} \cdot (X_{wi} + X_{wj} - 1) \quad \text{(b)} \end{aligned}$$

Adding (a) to (b) gives the result, where the singletons are factored together via $X_{ij} = X_{ji}$. By linearity of expectation, ξ is then the sum of expected values of each component of the sum.

$$\xi = (2p - 1) \cdot p + (n_i - 2) \cdot (1 - p) \cdot (1 - 2p) + (n_j - 1) \cdot p \cdot (2p - 1) + 0$$

Rewriting as a polynomial in p, one obtains the result. $\boxed{\xi}$

Lemma 2.4. *Var*$[\hat{\xi}]$ *is of order* $\mathcal{O}(n_i \cdot p)$. *More precisely,*

$$Var[\hat{\xi}] = (3n_i + n_j - 6) \cdot p + (2n - 11n_i - 5n_j + 18) \cdot p^2 + (12n_i + 8n_j - 2n - 20) \cdot p^3 + (-4n_i - 4n_j + 8) \cdot p^4$$

Proof. If $A, B \sim \text{Bern}(p)$, then

- $A^2 = A \implies \mathbb{E}[A^2] = p$
- $(1 - A)^2 = (1 - A) \implies \mathbb{E}[(1 - A)^2] = 1 - p$
- $(1 - A - B)^2 = (1 + 2AB - A - B) \implies \mathbb{E}[(1 - A - B)^2] = 1 - 2p + 2p^2$

Utilising $\text{Var}[Z] := \mathbb{E}[Z^2] - \mathbb{E}[Z]^2$, we get that

$$\begin{aligned} \text{Var}\left[(X_{xi} + X_{xj} - 1) \cdot X_{ij}\right] &= (1 - 2p + 2p^2) \cdot p - (1 - 2p)^2 \cdot p^2 = p - 3p^2 + 6p^3 - 4p^4 \\ \text{Var}\left[(1 - X_{xw}) \cdot (1 - X_{wi} - X_{wj})\right] &= (1 - p) \cdot (1 - 2p + 2p^2) - (1 - p)^2 \cdot (1 - 2p)^2 = 3p - 9p^2 + 10p^3 - 4p^4 \\ \text{Var}\left[X_{xw} \cdot (X_{wi} + X_{wj} - 1)\right] &= p \cdot (1 - 2p + 2p^2) - p^2 \cdot (1 - 2p)^2 = p - 3p^2 + 6p^3 - 4p^4 \\ \text{Var}\left[X_{xw} \cdot (X_{iw} - X_{jw})\right] &= p \cdot (2p - 2p^2) - 0 = 2p^2 - 2p^3 \end{aligned}$$

Then, by independence, $\text{Var}(\hat{\xi})$ is the sum of variances of individual terms. Computation (2.3) gives

$$\text{Var}[\hat{\xi}] = (n_i - 2) \cdot (3p - 9p^2 + 10p^3 - 4p^4) + n_j \cdot (p - 3p^2 + 6p^3 - 4p^4) + (n - n_i - n_j) \cdot (2p^2 - 2p^3).$$

This expression is indeed of order $\mathcal{O}(n_i \cdot p)$. Rewriting as a polynomial in p gives the result. $\boxed{\xi}$

2.2 Large Deviations

The implication of (2.3) is that the score gap grows linearly in n_i. This is great progress, but one still needs to gauge large deviations $\hat{\xi}$ below the expected gap ξ. We interest ourselves in lower-tail deviations, as such deviations close in the gap between the two distributions of $\hat{\xi}_{\text{in}}$ and $\hat{\xi}_{\text{out}}$, making them less distinguishable. For this specific purpose, the Bernstein bound comes into play, giving an exponentially decaying bound on such deviations.

Lemma 2.5. *Let $\mathcal{A}$ be the event that any score gap $\hat{\xi}$ deviates below the expected gap ξ by at least a linear multiple $\alpha \cdot \xi$, for some $\alpha \in \mathbb{R}$. Assume further that the equivalence classes n_i grow linearly in n. Then,*

$$\mathbb{P}[\mathcal{A}] \underset{n\to\infty}{\longrightarrow} 0.$$

Proof. First, if there are n^2 scores, then there is at most n^4 possible score gaps. We know from (2.3) that $\hat{\xi}$ is a sum of independent random variables, bounded in absolute value by one. Furthermore, (2.4) tells us that every term of the sum has bounded variance. Applying the Bernstein bound gives

$$\mathbb{P}[\mathcal{A}] \leq n^4 \exp\left(-\frac{\alpha^2 \cdot \xi^2}{2\sigma^2 + 2/3 \cdot \alpha \cdot \xi}\right) \underset{n\to\infty}{\longrightarrow} n^4 \exp\left(-\frac{\Theta(n^2)}{\Theta(n \cdot p) + \Theta(n)}\right) = 0. \quad \boxed{\xi}$$

Remark. The statement might fail when n_i does not grow linearly in n, for example, if $n_i = \Theta(\log(n))$.

3 Main Results

To prove that the equivalence relation $\sim$ is *noise stable*, the strategy is to supply an algorithm $\mathbb{A}$ with an asymptotically zero misclassification error.

3.1 The Algorithm

Let us recall the key protagonists of this construction. With $x, i \in X_i$ and $j \in X_j$, the goal was to separate members from non-members. In theory, we have explicit expressions for the expected gap $\xi(n_i, n_j, p)$ with variance $\text{Var}[\xi] =: \sigma^2(n, n_i, n_j, p)$. In practice, however, this requires knowledge of, n_i, n_j, namely the size of each equivalence class sizes.

The idea is to establish a lower bound for the gap, regardless of n_j. For this discussion, set $\epsilon n \leq n_\alpha \leq n$.

Lemma 3.1. *Let $\xi_+(n, n_i, p) = (n_i - 2) - (3n_i + \epsilon n - 6) \cdot p + (2n - 4) \cdot p^2$. Then, $\xi_+ \geq \xi$.*

Computation. The contribution of n_j is negative, therefore we replace it with ϵn. $\boxed{\xi}$

Lemma 3.2. *Let $\sigma_+^2(n, n_i, p)$ be given by*

$$(n_i - 2) \cdot (3p - 9p^2 + 10p^3 - 4p^4) + n \cdot (p - 3p^2 + 6p^3 - 4p^4) + (n - n_i - \epsilon n) \cdot (2p^2 - 2p^3).$$

Then, $\sigma_+^2 \geq \sigma^2$.

Computation. We replace n_j with ϵn when the contribution is negative, and n when it is positive. $\boxed{\xi}$

Remark. The reason why we still get good estimates is that n_j is linear in n, and always has a factor of $p = \Theta(1/n)$.

The crucial insight is as follows. If the gap $\hat{\xi}$ deviates below ξ_+ by t, and $\xi_+ - \xi$ is positive (3.1), then $\hat{\xi}$ has surely deviated from ξ by at least t. From there we obtain t and may apply the Bernstein inequality,

$$\mathbb{P}[\xi - \hat{\xi} \geq t] \leq \exp\left(-\frac{t^2}{2\sigma^2 + 2/3 \cdot \alpha \cdot t}\right) \leq \exp\left(-\frac{t^2}{2\sigma_+^2 + 2/3 \cdot t}\right) \qquad \text{(by 3.2)}$$

This is a proper sieve for non-members j. Next, computing the score matrix can be computationally expensive. Let us propose a more sustainable solution.

Lemma 3.3. *Let $(X, \sim')$ be a noisy equivalence relation, and set A to be its adjacency matrix. Then,*

$$S_{\sim'} = A^2 - 2 \cdot A + I.$$

Proof. It is a well-known fact that the square of an adjacency matrix gives all paths $i \to k \to j$. This is very close to $s(i,j)$, with the exception that the score purely counts common neighbours. It counts every x for which

$$i \sim x \ \wedge \ x \sim j. \qquad (x \neq i \neq j)$$

If $x = i$ or $x = j$, the condition simplifies to $i \sim j$. When $i \neq j$, we have two different x that are counted, so A^2 differs from $S_{\sim'}$ by two. Otherwise, $i = j$ and x is counted once, so the identity matrix accounts for the over-subtraction. $\boxed{\xi}$

Our algorithm takes an adjacency matrix of size $n \times n$, and a parameter p.

Input: (A, p)

- $n \leftarrow A_{n\times n}$ ▷ Get n
- $S \leftarrow A^2 - 2 \cdot A + I_n$ ▷ (3.3)
- **for** each row $x \in \{1, \dots, n\}$ **do**
  - $\tilde{\mathbf{s}} = \{\tilde{s}(x,k)\}_{k=1}^{n}$ ▷ Score row
  - $\tilde{\mathbf{s}} \leftarrow \mathrm{sort}(\tilde{\mathbf{s}})$ ▷ Sort descendingly
  - **for** each element $k \in \{1, \dots, n\}$ **do** ▷ Assume $n_i = k$ and find best fit
    - $\sigma_+^2 \leftarrow \mathrm{Var}(n,k,p)$ ▷ (3.2)
    - $\xi_+ \leftarrow \xi_+(n,k,p)$ ▷ (3.1)
    - $\hat{\xi} \leftarrow \tilde{\mathbf{s}}(k) - \tilde{\mathbf{s}}(k+1)$ ▷ Measured gap
    - $t \leftarrow \xi_+ - \hat{\xi}$ ▷ Deviation Value
    - **if** $t < 0$ **then**
      - **continue** ▷ Skip upper deviations
    - **else**
      - $T(k) \leftarrow \mathrm{Bernstein}(\sigma_+^2, t)$ ▷ Compute deviation probabilities
    - **end if**
  - **end for**
  - $\tau(x) \leftarrow \tilde{\mathbf{s}}(\arg_k \max\ T(k))$ ▷ Least-unlikely approach
- **end for**
- **Return:** Thresholds τ

With the score thresholds in hand, one may easily recover $(X, \sim)$.[‡] This is what we prove next.

3.2 Noise Stability of Equivalence Relations

As with our earlier discussions, we set p to be of order $1/n$, and assume that the equivalence classes grow linearly in n.

Theorem 3.4. *Set $\mathbb{A}$ to be as described in Section 3.1. Then,*

$$\lim_{n\to\infty} \mathbb{P}[\mathbb{A}(X, \sim') \neq (X, \sim)] = 0.$$

Proof of Theorem. It is enough to show that the distributions of $\hat{\xi}_{\text{in}}$, $\hat{\xi}_{\text{out}}$ are almost surely distinguishable. This is not difficult to show, since

(2.3) the expected score gap $\xi = \Theta(n_i) = \Theta(n)$ grows linearly,

(2.4) variance $\sigma^2 = \Theta(n \cdot p) = \Theta(1)$ is finite,

(2.5) linear deviations from ξ are unlikely.

A misclassification is therefore not in question. $\boxed{\xi}$

[‡] For a thorough discussion of the algorithm's construction, please consult the Jupyter notebook https://elshenawyom.github.io/projects/noise-sensitivity associated to this thesis.

4 Closing Remarks

This thesis is not interesting for the classification result on equivalence classes, but rather for its introduction of a large class of open problems to study. For instance, one might study the same problem on partial orders, total orders, and perhaps even on unions of sets with different relations. Even in the case of an equivalence relation, one may choose to apply noise asymmetrically, so that the graph of $(X, \sim)$ is one that is directed.

The notion of *noise sensitivity* on binary relations may also be improved, as it is not entirely obvious how one might prove there is no recovery algorithm. It would be very interesting to find an example of a binary relation $\sim$ that is provably *noise-insensitive.* Whether the current definition allows for that is something to explore in the future.

It remains surprising how many ways exist to attack the same question we proposed. For instance, the question may be formulated in a purely *information-theoretic* language, or effectively solved via modularity and community detection, or portrayed in a graph-theoretic setting. Regardless of the approach, the pursuit of this question proves to be an intellectually rewarding experience.

4.1 Jupyter Notebook

A Jupyter notebook associated to this thesis may be found under

https://elshenawyom.github.io/projects/noise-sensitivity

Acknowledgments

I would like to thank and acknowledge Professor Keivan Mallahi−Karai for his mentorship and support. I am grateful to the Mathematics Society at Constructor University, to my professors and colleagues with whom I have shared the last three years. To my friends and family, without you I would not be standing till this day. Finally, this thesis is dedicated to Julius Liu, my fellow mathematician and dear friend who passed away in the summer of 2023.

None of this would have been possible without you. $\boxed{\xi}$

References

[1] Itai Benjamini, Gil Kalai, and Oded Schramm. *Noise Sensitivity of Boolean Functions and Applications to Percolation*. 2000. arXiv: math/9811157 [math.PR].

[2] Encyclopedia of Mathematics. *Bernstein inequality*. http://encyclopediaofmath.org/index.php?title=Bernstein_inequality&oldid=15217.