Demonstration: Equitable and Non-equitable Distribution among Categories

Youssoupha Gaye[1[0009−0003−6101−9995]], Maissa Mbaye[1[0000−0002−5748−446X]], Dame Diongue[1[0000−0002−3452−4174]], and Ousmane Dieng[2[0000−0003−3320−7887]]
Emmanuel Adetiba[3,4] Joke A. Badejo[3[0000−0002−5646−7385]]

[1] Laboratoire d'Analyse Numerique et Informatique/LANI, CEA-MITIC, Gaston Berger University, Saint-Louis PB 234, Senegal
{gaye.youssoupha,maissa.mbaye,dame.diongue}@ugb.edu.sn
[2] Power-management and Real-Time Systems Lab, University of Pittsburgh, Pittsburgh, PA 15260, USA
oud5@pitt.edu
[3] Covenant Applied Informatics and Communication African Center of Excellence, Covenant University, Ota, Ogun State, Nigeria.
{emmanuel.adetiba,joke.badejo}@covenantuniversity.edu.ng
[4] HRA, Institute for Systems Science, Durban University of Technology, Durban 1334, South Africa

1 The stabilizer factor

The stabilized factor adjusts the value of H in order to distinguish identifier and evenly distributed categories for a given attribute.

Definition 1. *Let p be the relative frequency that represents the probability that a value is in category $C(i)$. The stabilizer factor P is defined as follows :*

$$P = \prod_{i=1}^{s} p(x = C(i)),\ \ where\ p(x = C(i)) = \frac{n_i}{N}\ \ and\ \sum_{n=1}^{S} n_i = N. \tag{1}$$

1.1 Case 1: Equitable distribution among categories

– if $S = N$ then $n_1 = n_2 = n_3 = ... = n_S = 1$

$$P = \frac{n_1}{N} \times \frac{n_2}{N} \times \frac{n_3}{N} \times ... \times \frac{n_S}{N} = \frac{1}{N^N}$$

– if $S \in]1, N[$ then $n_1 = n_2 = n_3 = ... = n_S = a > 1$

$$P = \frac{n_1}{N} \times \frac{n_2}{N} \times \frac{n_3}{N} \times ... \times \frac{n_S}{N} = (\frac{a}{N})^S$$

Lemma 1. *Let D_1 and D_2 be two equitable distributions of the same attribute $C(i)$ with, respectively, S_1 and S_2 as a number of categories so that $S_1 < S_2$. $S_1, S_2, N \in \mathbb{N}^*$. $n_1 = n_2 = ... = n_{S_1} = a_1;$ $m_1 = m_2 = ... = m_{S_2} = a_2$*

$$\sum_{n=1}^{S_1} n_i = \sum_{m=1}^{S_2} m_i = N \ ; \ P_1 = \prod_{i=1}^{S_1} \frac{n_i}{N} = (\frac{a_1}{N})^{S_1} \ ; \ P_2 = \prod_{i=1}^{S_2} \frac{m_i}{N} = (\frac{a_2}{N})^{S_2}$$

$$S_1 < S_2 \Rightarrow P_1 > P_2 \tag{2}$$

Proof. Let $S_1, S_2, N \in \mathbb{N}^*$; $S_1 < S_2$

$$\frac{P_1}{P_2} = \frac{(\frac{a_1}{N})^{S_1}}{(\frac{a_2}{N})^{S_2}}. \text{ However } N = a_1 S_1 = a_2 S_2, \text{then}$$

$$= \frac{(\frac{a_1}{a_1 S_1})^{S_1}}{(\frac{a_2}{a_2 S_2})^{S_2}} = \frac{(\frac{1}{S_1})^{S_1}}{(\frac{1}{S_2})^{S_2}} = \frac{{S_2}^{S_2}}{{S_1}^{S_1}} > 1$$

$$\frac{P_1}{P_2} > 1, \text{ then } P_1 > P_2$$

Conclusion: This mathematical proof demonstrates that when it comes to equitable distribution, the more categories there are, the lower the stabilizing factor's value becomes.

1.2 Case 2: Non-equitable distribution among categories

This is the case when there are at least two categories that do not have the same frequency.

Lemma 2. *Let D_1 and D_2 be two non-equitable distributions of the same attribute $C(i)$ with, respectively, S_1 and S_2 as the number of categories so that $S_1 < S_2$.*
$S_1, S_2, N \in \mathbb{N}^*$. $(n_{1,i})_{i=1}^{S_1}$ and $(n_{2,i})_{i=1}^{S_2}$ with $\begin{cases} n_{j,i} \in N^*; \ j \in \{1,2\}; \ i \in [1, S_j] \\ \sum_{i=1}^{S_j} n_{j,i} = N \end{cases}$

$$P_1 = \prod_{i=1}^{S_1} \frac{n_{1,i}}{N} \ ; \ P_2 = \prod_{i=1}^{S_2} \frac{n_{2,i}}{N}$$

$$S_1 < S_2 \Rightarrow P_1 > P_2 \tag{3}$$

Proof. It is established by the recurrence

- **Assumption:** $S_1 < S_2 \Rightarrow P_1 > P_2$
  To prove this, we use the following lemma:

- **Lemma (To be proved):** Let $N \in \mathbb{N}^*$ and $k < N \in \mathbb{N}^*$.

$$If\ (n_{k,i})_{i=1}^{k} \text{ and } (n_{k+1,i})_{i=1}^{k+1} : \ \begin{cases} n_{k,i},\ n_{k+1,i} \in N^* \\ \sum_{i=1}^{k} n_{k,i} = \sum_{i=1}^{k+1} n_{k+1,i} = N \end{cases}$$

$$\text{then } P_k > P_{k+1} \text{ with } P_k = \prod_{i=1}^{k} \frac{n_{k,i}}{N}\ ;\ P_{k+1} = \prod_{i=1}^{k+1} \frac{n_{k+1,i}}{N}$$

- **Proof of the lemma (By recurrence on k)**
  Initialization: $k = 1$

$$k = 1\ ;\ n_{1,1} = N \Rightarrow P_1 = \frac{n_{1,1}}{N} = \frac{N}{N} = 1$$

$$\text{and } k+1 = 2\ ;\ n_{2,1} + n_{2,2} = N \Rightarrow P_2 = \frac{n_{2,1}}{N} \times \frac{n_{2,2}}{N} < 1$$

$$\Rightarrow P_1 > P_2$$

Heredity: $k < N$
It is assumed that for all sequences:

$$(n'_{k,i})_{i=1}^{k} \text{ and } (n'_{k+1,i})_{i=1}^{k+1} : \ \sum_{i=1}^{k} n'_{k,i} = \sum_{i=1}^{k+1} n'_{k+1,i} = N$$

$$\Rightarrow P'_k > P'_{k+1}$$

Let us show that for

$$(n_{k+1,i})_{i=1}^{k+1} \text{ and } (n_{k+2,i})_{i=1}^{k+2} : \ \sum_{i=1}^{k+1} n_{k+1,i} = \sum_{i=1}^{k+2} n_{k+2,i} = N$$

$$\Rightarrow P_{k+1} > P_{k+2}\ (for\ k < N-1)$$

$$\text{Let } (n_{k+1,i})_{i=1}^{k+1} \text{ and } (n_{k+2,i})_{i=1}^{k+2} : \ \sum_{i=1}^{k+1} n_{k+1,i} = \sum_{i=1}^{k+2} n_{k+2,i} = N$$

Since we are comparing the product of all the terms in the two sequences, we can arbitrarily assume that $(n_{k+1,i})_{i=1}^{k+1}$ is increasing and $(n_{k+2,i})_{i=1}^{k+2}$ is decreasing. (Because the product of integers is commutative)

(i) Let's show that:

$$\exists (l,p) \in [1, k+1] \times [1, k+1] :\ n_{k+1,l} > n_{k+2,p}$$

By absurdity, we assume that:

$$\forall (l,p) \in [1, k+1] \times [1, k+1],\ n_{k+1,l} \leq n_{k+2,p}$$

$$\text{then for } l = p = i \in [1, k+1],\ n_{k+1,i} \leq n_{k+2,i}$$

$$\Rightarrow \sum_{i=1}^{k+1} n_{k+1,i} \leq \sum_{i=1}^{k+1} n_{k+2,i}$$

$$\begin{aligned}\text{However } \sum_{i=1}^{k+1} n_{k+1,i} = N \text{ and } \sum_{i=1}^{k+1} n_{k+2,i} &= \sum_{i=1}^{k+2} n_{k+2,i} - n_{k+2,k+2} \\ &= N - n_{k+2,k+2}\end{aligned}$$

$$\text{then } N \leq N - n_{k+2,k+2} \text{ (Absurd because } n_{k+2,k+2} \in \mathbb{N}^*)$$

$$\underline{So\ \exists (l,p) \in [1, k+1] \times [1, k+1] :\ n_{k+1,l} > n_{k+2,p}}$$

(ii) Let's show that

$$n_{k+1,k} > n_{k+2,k+1}$$

$$\text{Since } \exists (l,p) \in [1, k+1]^2 : n_{k+1,l} > n_{k+2,p}$$

$$\text{however } n_{k+1,k} \geq n_{k+1,l} \text{ because } (n_{k+1,i})_{i=1}^{k+1} \text{ is increasing}$$

$$n_{k+2,k} \leq n_{k+2,p} \text{ because } (n_{k+2,i})_{i=1}^{k+2} \text{ is decreasing}$$

$$\Rightarrow n_{k+1,k} \geq n_{k+1,l} > n_{k+2,p} \geq n_{k+2,k+1}$$

$$\underline{\Rightarrow n_{k+1,k} > n_{k+2,k+1}}$$

(iii) Let's build

$$(n'_{k,i})_{i=1}^{k} \text{ and } (n'_{k+1,i})_{i=1}^{k+1} : \sum_{i=1}^{k} n'_{k,i} = \sum_{i=1}^{k+1} n'_{k+1,i} = N$$

$$\text{Let's set } n'_{k,i} : \begin{cases} n'_{k,i} = n_{k+1,i}\ for\ i \in [1, k-1] \\ n'_{k,k} = n_{k+1,k} + n_{k+1,k+1} \end{cases}$$

$$\begin{aligned}\text{then } \sum_{i=1}^{k} n'_{k,i} &= \sum_{i=1}^{k-1} n'_{k,i} + n'_{k,k} \\ &= \sum_{i=1}^{k-1} n_{k+1,i} + (n_{k+1,k} + n_{k+1,k+1}) \\ &= \sum_{i=1}^{k+1} n_{k+1,i} \\ &= N\end{aligned}$$

$$\text{and } n'_{k+1,i} : \begin{cases} n'_{k+1,i} = n_{k+2,i}\ for\ i \in [1, k] \\ n'_{k+1,k+1} = n_{k+2,k+1} + n_{k+2,k+2} \end{cases}$$

$$\text{then } \sum_{i=1}^{k+1} n'_{k+1,i} = \sum_{i=1}^{k+2} n_{k+2,i} = N$$

$$(n'_{k,i})_{i=1}^{k} \text{ and } (n'_{k+1,i})_{i=1}^{k+1} : \sum_{i=1}^{k} n'_{k,i} = \sum_{i=1}^{k+1} n'_{k+1,i} = N$$

$$\text{then } P'_k > P'_{k+1} \text{ (by heredity hypothesis)}$$

$$\text{however } P'_k = \prod_{i=1}^{k} \frac{n'_{k,i}}{N} = \underbrace{\prod_{i=1}^{k-1} \frac{n'_{k,i}}{N}}_{A} \times \overbrace{\frac{n'_{k,k}}{N}}^{B}$$

$$\begin{aligned}
\text{moreover } A &= \prod_{i=1}^{k-1} \frac{n'_{k,i}}{N} = \prod_{i=1}^{k-1} \frac{n_{k+1,i}}{N} \\
&= \prod_{i=1}^{k-1} \frac{n_{k+1,i}}{N} \times \frac{\frac{n_{k+1,k}}{N} \times \frac{n_{k+1,k+1}}{N}}{\frac{n_{k+1,k}}{N} \times \frac{n_{k+1,k+1}}{N}} \\
&= \prod_{i=1}^{k+1} \frac{n_{k+1,i}}{N} \times \frac{1}{\frac{n_{k+1,k} \times n_{k+1,k+1}}{N^2}} \text{ however } P_{k+1} = \prod_{i=1}^{k+1} \frac{n_{k+1,i}}{N}
\end{aligned}$$

$$A = P_{k+1} \times \frac{N^2}{n_{k+1,k} \times n_{k+1,k+1}}$$

$$B = \frac{n'_{k,k}}{N} = \frac{n_{k+1,k} + n_{k+1,k+1}}{N}$$

$$\begin{aligned}
\text{then } P'_k &= A \times B \\
&= P_{k+1} \times \frac{N^2}{n_{k+1,k} \times n_{k+1,k+1}} \times \frac{n_{k+1,k} + n_{k+1,k+1}}{N} \\
&= P_{k+1} \times N \times \frac{n_{k+1,k} + n_{k+1,k+1}}{n_{k+1,k} \times n_{k+1,k+1}}
\end{aligned}$$

$$P'_{k+1} = \prod_{i=1}^{k+1} \frac{n'_{k+1,i}}{N} = \underbrace{\prod_{i=1}^{k} \frac{n'_{k+1,i}}{N}}_{A'} \times \overbrace{\frac{n'_{k+1,k+1}}{N}}^{B'}$$

$$A' = \prod_{i=1}^{k} \frac{n'_{k+1,i}}{N} = P_{k+2} \times \frac{N^2}{n_{k+2,k+1} \times n_{k+2,k+2}}$$

$$B' = \frac{n'_{k+1,k+1}}{N} = \frac{n_{k+2,k+1} + n_{k+2,k+2}}{N}$$

$$\begin{aligned} P'_{k+1} &= A' \times B' \\ &= P_{k+2} \times N \times \frac{n_{k+2,k+1} + n_{k+2,k+2}}{n_{k+2,k+1} \times n_{k+2,k+2}} \end{aligned}$$

However, we have

$$\begin{aligned} P'_k > P'_{k+1} &\Rightarrow \frac{P'_k}{P'_{k+1}} > 1 \\ &\Rightarrow \frac{P_{k+1} \times N \times \dfrac{n_{k+1,k} + n_{k+1,k+1}}{n_{k+1,k} \times n_{k+1,k+1}}}{P_{k+2} \times N \times \dfrac{n_{k+2,k+1} + n_{k+2,k+2}}{n_{k+2,k+1} \times n_{k+2,k+2}}} > 1 \\ &\Rightarrow \frac{P_{k+1}}{P_{k+2}} > \frac{\dfrac{n_{k+2,k+1} + n_{k+2,k+2}}{n_{k+2,k+1} \times n_{k+2,k+2}}}{\dfrac{n_{k+1,k} + n_{k+1,k+1}}{n_{k+1,k} \times n_{k+1,k+1}}} \\ &\Rightarrow \frac{P_{k+1}}{P_{k+2}} > \frac{n_{k+2,k+1} + n_{k+2,k+2}}{n_{k+2,k+1} \times n_{k+2,k+2}} \times \frac{n_{k+1,k} \times n_{k+1,k+1}}{n_{k+1,k} + n_{k+1,k+1}} \end{aligned}$$

- $n_{k+2,k+1} + n_{k+2,k+2} \geq 2n_{k+2,k+2}$ because $(n_{k+2,i})_{i=1}^{k+2}$ is decreasing

$$\begin{aligned} \Rightarrow \frac{n_{k+2,k+1} + n_{k+2,k+2}}{n_{k+2,k+1} \times n_{k+2,k+2}} &\geq \frac{2n_{k+2,k+2}}{n_{k+2,k+1} \times n_{k+1,k+2}} \\ &\geq \frac{2}{n_{k+2,k+1}} \ (1) \end{aligned}$$

- $n_{k+1,k} + n_{k+1,k+1} \leq 2n_{k+1,k+1}$ because $(n_{k+1,i})_{i=1}^{k+1}$ is increasing

$$\Rightarrow \frac{n_{k+1,k} + n_{k+1,k+1}}{n_{k+1,k} \times n_{k+1,k+1}} \geq \frac{2n_{k+1,k+1}}{n_{k+1,k} \times n_{k+1,k+1}} = \frac{2}{n_{k+1,k}}$$

$$\Rightarrow \frac{n_{k+1,k} \times n_{k+1,k+1}}{n_{k+1,k} + n_{k+1,k+1}} \leq \frac{n_{k+2,k+1}}{2} \ (2)$$

$$\begin{aligned} (1) \text{ and } (2) \ \Rightarrow \frac{P_{k+1}}{P_{k+2}} &> \frac{2}{n_{k+2,k+1}} \times \frac{n_{k+1,k}}{2} \\ &> \frac{n_{k+1,k}}{n_{k+2,k+1}}. \text{ However } n_{k+1,k} \geq n_{k+2,k+1}, \text{ then} \\ &> 1 \end{aligned}$$

$$\Rightarrow P_{k+1} > P_{k+2}$$

Conclusion: $N \in \mathbb{N}^*$ and $k < N \in \mathbb{N}^*$

$$(n_{k,i})_{i=1}^{k} \text{ and } (n_{k+1,i})_{i=1}^{k+1} : \begin{cases} n_{k,i},\ n_{k+1,i} \in N^* \\ \sum_{i=1}^{k} n_{k,i} = \sum_{i=1}^{k+1} n_{k+1,i} = N \end{cases}$$

$$\text{then } P_k > P_{k+1} \text{ with } P_k = \prod_{i=1}^{k} \frac{n_{k,i}}{N} \text{ and } P_{k+1} = \prod_{i=1}^{k+1} \frac{n_{k+1,i}}{N}$$

We have thus proved the lemma.

– **Proof (Assumption)**

$$S_1 < S_2 \Rightarrow S_2 = S_1 + m \text{ (with } m = S_2 - S_1)$$

From the lemma

$$P_{S_1} > P_{S_1+1} > P_{S_1+2} > ... > P_{S_1+m}$$

$$\text{however } P_{S_1} = P_1 \text{ and } P_{S_1+m} = P_{S_2} = P_2$$

$$\text{then } P_1 > P_2$$