ZIPF'S LAW AND MANDELBROT'S IDEA

P. GIORDANO

We have a system whose state is described by a vector $y \in M \subseteq \mathbb{R}^n$. The systems has to be thought as a complex adaptive system that changes its state so as to decrease a suitable cost function $C : M \longrightarrow \mathbb{R}_{>0}$ and, at the same time, to increase a corresponding information entropy

$$H(y) = -\sum_{j=1}^{d} p_j(y) \cdot \log_2 p_j(y) > 0 \qquad \forall y \in M.$$

More precisely, the system adapts (e.g. it evolves) so as to minimize the ratio $\frac{C}{H}$ at the interior point $x \in M$:

$$\forall y \in M : \ 0 < \frac{C(x)}{H(x)} \le \frac{C(y)}{H(y)}.$$

By assumption, the probability always enter into the state of the system

$$p_j(y) = y_j \qquad \forall j = 1, \dots, d,$$

where $0 < d \le n$. This implies that by changing these probability, we are describing a different system having a different state.
The cost function C must satisfies the inequality

$$\partial_k C(x) \le \alpha(x) \cdot \log_2 k \qquad \forall k = 2, \dots, d \tag{0.1}$$

for some $\alpha : M \longrightarrow \mathbb{R}_{>0}$. Note that (0.1) is not required to hold for $k = 1$. Examples of cost functions that satisfy (0.1) are given by the average value $C(y) = \sum_{j=1}^{d} c_k(y) \cdot p_k(y)$, where the k-th component of the cost c_k can be any one of the following examples:

(i) $c_k(y) = \frac{a}{p_k(y)^s} \log_2 k = \frac{a}{y_k^s} \log_2 k$, where $a \in \mathbb{R}_{>0}$ and $s \in \mathbb{R}_{\ge 1}$. This example represents costs that are decreasing with an increasing of the probabilities y_k and they are increasing with an increasing of the number of bits $\log_2 k$ which are necessary to transmit the rank k. In this case we have $\partial_k C(x) = \partial_k c_k(x) \cdot x_k + c_k(x) \le \frac{a}{x_k^s}(1 - s) \le 0$. Therefore, it suffices to take $\alpha(x) \equiv 1$ (or lower, depending on the next assumption (0.3), see below).

(ii) $c_k(y) \le a \cdot \log_b(k + k_0) + j_0$, where a, $j_0 \in \mathbb{R}_{>0}$, $b > 2$ and $k_0 \le b - 2$. This example (with the equality sign) is essentially considered in [1]. Usually, b is the number of letters in an alphabet we are considering. In this case, the costs c_k do not depend on the probabilities y_k and hence $\partial_k C(x) = c_k(x)$. We have $\log_b(k + k_0) \le \log_2 k$ if and only if

$$k + k_0 \le b^{\log_2 k} = \left(b^{\log_b k}\right)^{\frac{1}{\log_b 2}} = k^{\log_2 b}. \tag{0.2}$$

Since $b > 2$, we have $\log_2 b > 1$ and the function $k^{\log_2 b} - k$ is increasing in k. Therefore if $k_0 \le 2^{\log_2 b} - 2 = b - 2$ the inequality (0.2) always holds, and hence $c_k(x) \le a \log_2 k + j_0$. If we take $\alpha \equiv a + j_0$, we get $\alpha \log_2 k =$

$a \log_2 k + j_0 \log_2 k \geq a \log_2 k + j_0 \geq a \cdot \log_b(k + k_0) + j_0 \geq c_k(x) = \partial_k C(x)$ for $k \geq 2$.

(iii) We can choose a better estimate of α if $c_k(y) \leq j_0 + a \cdot \log_b k$, i.e. if $k_0 = 0$ in the previous example. This case is considered in [2]. In fact, for $k \geq 2$ we have

$$\begin{aligned}\partial_k C(x) = c_k(x) \leq j_0 + a \log_b k = j_0 \log_2 2 + a \log_b 2 \cdot \log_2 k \leq \\ \leq j_0 \log_2 k + a \log_b 2 \cdot \log_2 k = (a \log_b 2 + j_0) \cdot \log_2 k.\end{aligned}$$

We can hence set $\alpha \equiv a \log_b 2 + j_0$.

(iv) $c_k(y) = \gamma_k > 0$. This is the case of constant costs depending only on the rank k (e.g. we can have that γ_k is proportional to the length of the words of rank k). We therefore have to take α so that $\alpha(x) \geq \frac{\gamma_k}{\log_2 k}$.

Note that in the first example (i) it is not reasonable to assume that this formula holds also for $k = 1$ because this would yield a null cost $c_1(y) = 0$. This, and the calculations we realized in the second example, motivate that (0.1) holds only for $k \geq 2$.

Finally, we assume that

$$\sum_{k=1}^{d} k^{-\alpha(x)\cdot\frac{H(x)}{C(x)}} =: N(x) \geq \frac{1}{p_1(x)} \geq e. \tag{0.3}$$

Note that this implies $p_1(x) \leq e^{-1} \simeq 0.368$. This is another restriction on the value $\alpha(x)$: whereas condition (0.1) states that we can take $\alpha(x)$ as large as we want, the inequality (0.3) yields that the larger is $\alpha(x)$, the more difficult will be to arrive at a value $N(x) \geq e$.

We have the following

Theorem 1. *Let $x \in M \subseteq \mathbb{R}^n$ be an open set and let $p_j \in \mathcal{C}^1(M, \mathbb{R}_{\geq 0})$, for all $j = 1, \dots, d \leq n$, be such that*

$$\forall y \in M : \ \left(p_j(y)\right)_{j=1,\dots,d} \ \text{is a probability.}$$

Set

$$H(y) := -\sum_{j=1}^{d} p_j(y) \cdot \log_2 j \qquad \forall y \in M.$$

Let $C \in \mathcal{C}^1(M, \mathbb{R}_{>0})$ be such that

$$\forall y \in M : \ 0 < \frac{C(x)}{H(x)} \leq \frac{C(y)}{H(y)}. \tag{0.4}$$

Finally assume that

$$\begin{gathered} p_j(y) = y_j \qquad \forall j = 1, \dots, d \ \forall y \in M \\ \partial_k C(x) \leq \alpha(x) \cdot \log_2 k \qquad \forall k = 2, \dots, d \\ \sum_{k=1}^{d} k^{-\alpha(x)\cdot\frac{H(x)}{C(x)}} =: N(x) \geq \frac{1}{p_1(x)} \geq e, \end{gathered} \tag{0.5}$$

where $\alpha : M \longrightarrow \mathbb{R}_{>0}$. Then we have

(i) $p_k(x) = p_1(x) \cdot k^{-\alpha(x)\cdot\frac{H(x)}{C(x)}}$ *for all $k = 1, \dots, d$.*

(ii) $p_1(x) = \frac{1}{N(x)}$.

Proof. Since M is an open set, $x \in M$ and C, $H \in \mathcal{C}^1(M, \mathbb{R}_{>0})$, by (0.4) we get $\partial_k \left(\frac{C}{H}\right)(x) = 0$. For simplicity, all the functions that will appear in the following are evaluated at the point x. We have

$$\partial_k C \cdot H + C \cdot \sum_j \left(\partial_k p_j \cdot \log_2 p_j + p_j \frac{1}{p_j} \log_2 e \cdot \partial_k p_j \right) = 0$$

But $\partial_k p_j(x) = \partial_k x_j = \delta_{kj}$, so

$$H \cdot \partial_k C + C \left(\log_2 p_k + \log_2 e \right) = 0.$$

By (0.5), for $k \geq 2$ we obtain

$$\log_2 p_k = -\frac{H}{C} \partial_k C - \log_2 e \geq -\frac{H}{C} \alpha \log_2 k - \log_2 e,$$

and hence

$$p_k \geq 2^{\left(-\frac{H}{C}\alpha \log_2 k - \log_2 e\right)} = k^{-\alpha \frac{H}{C}} e^{-1}.$$

We assumed that $N \geq e$, so

$$p_k \geq \frac{1}{N} \cdot k^{-\alpha \frac{H}{C}} \qquad \forall k = 1, \dots, d.$$

Note that this inequality holds also for $k = 1$ because we assumed that $p_k(x) \geq \frac{1}{N(x)}$.

Finally, we note that we cannot have $p_h > \frac{1}{N} \cdot h^{-\alpha \frac{H}{C}}$ for some $h = 1, \dots, d$ because otherwise we would have

$$\sum_{j=1}^{d} p_j = 1 > \sum_{j=1}^{d} \frac{1}{N} k^{-\alpha \frac{H}{C}} = 1.$$

Therefore, we must have $p_k = \frac{1}{N} k^{-\alpha \frac{H}{C}}$ for all $k = 1, \dots, d$. Finally, for $k = 1$ we get $p_1 = \frac{1}{N}$ which proves both our conclusions. □

References

[1] Mandelbrot B., An informational theory of the statistical structure of language, in W. Jackson (eds.), Communication Theory, Butterworths, London, 1953.

[2] B. Conrad, M. Mitzenmacher, Power Laws for Monkeys Typing Randomly: The Case of Unequal Probabilities, IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 50, NO. 7, JULY 2004.

Wolfgang Pauli Institute, Vienna, AT.