Regular sets

and conditional density:

an extension of Benford's law

by

Rita Giuliano Antonini - Georges Grekos

The classical Benford's law (first digit problem)

Let $q \in \mathbb{N}^*$, $q \in \{1, 2, \ldots, 9\}$.

A_q = integers beginning by the digit q:

$$A_q = \cup_n \big([q10^n, (q+1)10^n[\cap \mathbb{N}^* \big).$$

Denote by $\partial(A_q)$ the logarithmic density of A_q.

Classical Benford's law

It states that

$$\partial(A_q) = \log\left(1 + \frac{1}{q}\right).$$

Historical discussion: see [R], [K-N-R-S].

General definition of density

Measure on $\mathbb{N}^*$

Given by

$$\mu = \sum_{n\in\mathbb{N}^*} \mu(n)\epsilon_n$$

where
(i) $(\mu(n)_n)$ is a sequence of non negative real numbers with $\sum_n \mu(n) = \infty$;
(ii) ϵ_n = measure of mass 1 concentrated in the integer n.

Put

$$F_\mu(n) = \sum_{k=1}^{n} \mu(k)$$

Upper μ-density of $A \subseteq \mathbb{N}^*$

$$\overline{\delta}_\mu(A) = \limsup_{n\to\infty} \frac{\mu(A \cap [1,n])}{F_\mu(n)}.$$

Lower μ-density of $A \subseteq \mathbb{N}^*$

Analogous definition, with $\liminf$ in place of $\limsup$. Denoted by $\underline{\delta}_\mu(A)$.

μ-density of $A \subseteq \mathbb{N}^*$

Analogous definition, when the limit exists. Denoted by

$$\delta_\mu(A)$$

Let $\mathbb{H} \subseteq \mathbb{N}^*$. If μ is concentrated on $\mathbb{H}$ (i. e. $\mu(n) = 0$ if $n \notin \mathbb{H}$) the μ-density is called **conditional with respect to $\mathbb{H}$**.

Logarithmic density of $A \subseteq \mathbb{N}^*$

Determined by the measure

$$\mu = \sum_{k \in \mathbb{N}} \frac{1}{k} \epsilon_k,$$

Denoted by $\partial(A)$ in place of $\delta_\mu(A)$

For stating extensions, we need the definition of

Logarithmic conditional density with respect to H

Take $\mathbb{H} \subseteq \mathbb{N}^*$.

The logarithmic density of $A \subseteq \mathbb{N}^*$ **conditional to $\mathbb{H}$** is determined by the measure

$$\nu = \sum_{k \in \mathbb{H}} \frac{1}{k} \epsilon_k,$$

(the mass is again $1/k$ but only for the points in $\mathbb{H}$).

Denoted by $\partial_{\mathbb{H}}(A)$.

Extensions of Benford's law

First extension

For $\mathbb{H} = \mathbb{P}$

The extension says that :

$$\partial_{\mathbb{P}}(A_q) = \partial(A_q)$$

Reference [K-N-R-S]

Second extension

For $\mathbb{H} = \mathbb{P}$

The extension says that :

$$\partial_{\mathbb{P}}(A) = \partial(A)$$

for a large class of sets A, to which of course A_q belongs.

Reference [F-L]

Natural questions

Question 1. For which other sets $\mathbb{H}$ (different from $\mathbb{P}$) we have

$$\partial_{\mathbb{H}}(A_q) = \partial(A_q)$$

Answer. The answer is for *regular sets*:

DEFINITION. (by [P-S]): Let $\mathbb{H} \subseteq \mathbb{N}^*$; put

$$H(x) = \mathrm{card}(\mathbb{H} \cap [1, x])$$

$\mathbb{H}$ is *regular* with exponent $\lambda \in (0, 1]$ if the function

$$L(x) = \frac{H(x)}{x^\lambda}$$

is slowly varying as $x \to \infty$.

DEFINITION. L is *slowly varying* as $x \to \infty$. if, for every $t > 0$

$$\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1.$$

Examples of regular sets

1. Let $r \geq 1$ be a fixed integer and

$$\mathbb{H} = \{n^r,\, n \in \mathbb{N}^*\}.$$

It is easy to see that $H(x) = \lfloor x^{1/r} \rfloor$, hence

$$L(x) = \frac{H(x)}{x^{1/r}} \sim 1.$$

Hence $\mathbb{H}$ is regular with exponent $\lambda = 1/r$.

2. Let $\mathbb{H}$ be the set of all powers. $\mathbb{H}$ is regular with exponent $1/2$ since

Lemma (G. Grekos)

$$H(x) \sim \sqrt{x}.$$

3. $\mathbb{P}$ is regular since its counting function π verifies

$$\pi(x) \sim \frac{x}{\log x}$$

($L(x) = \frac{1}{\log x}$ is slowly varying).

Question 2. For which other sets A (different from A_q) we have

$$\partial_{\mathbb{H}}(A) = \partial(A).$$

Answer. A first answer has been given by [F-L]: if

$$A = \cup_n \big([p_n, q_n[\cap \mathbb{N}^* \big),$$

they impose the conditions

$$q_{n-1} \sim \rho q_n, \quad p_n \sim \sigma q_n \quad \text{as } n \to \infty,$$

$(0 < \rho < \sigma < 1)$.

We shall see that the conditions by [F-L] are too restrictive, we need only

$$p_n \sim \sigma q_n \quad \text{as } n \to \infty.$$

Details

Our aim for both extensions is to use the following general Theorem (see [G]):

THEOREM

Let μ, ν be two measures on $\mathbb{N}^*$, both having infinite total mass. Let A be a subset of $\mathbb{N}^*$, given in the form

$$A = \mathbb{N}^* \cap (\cup_{n\geq 1}[p_n, q_n[)$$

where (p_n) and (q_n) are two sequences of integers, with $1 \leq p_n < q_n < p_{n+1}$.
Assume that the following conditions are satisfied:
(a) $F_\mu(p_n) \sim \alpha F_\nu(p_n)$;
(b) $F_\mu(q_n) - F_\mu(p_n) \sim \alpha(F_\nu(q_n) - F_\mu(p_n))$
as $n \to \infty$
for a suitable constant $\alpha > 0$. Then

$$\underline{\delta}_\mu(A) = \underline{\delta}_\nu(A), \qquad \overline{\delta}_\mu(A) = \overline{\delta}_\nu(A).$$

Let's explain why we need regularity, by discussing a particular case. Our aim is to apply the general Theorem, as we have said.

Take $\mathbb{H} \subseteq \mathbb{N}^*$; put

$$H(x) = \mathrm{card}(\mathbb{H} \cap [1, x])$$

The logarithmic density of $A \subseteq \mathbb{N}^*$ **conditional to** $\mathbb{H}$ is determined by the measure

$$\nu = \sum_{k \in \mathbb{H}} \frac{1}{k} \epsilon_k,$$

while the **non conditional** logarithmic density of A is determined by

$$\mu = \sum_{k \in \mathbb{N}} \frac{1}{k} \epsilon_k,$$

BUT, as a preliminary step, we content ourselves to find μ of the kind

$$\mu = \sum_{k \in \mathbb{N}} \frac{L(k)}{k} \epsilon_k ,$$

(interpret such μ as a sort of "spreading" of the mass of ν over **all** the integers by means of the function L; this interpretation is due to G. Grekos.)
For the case $\mathbb{H} = \mathbb{P}$, it is known [Rita] that a possible choice for L is

$$L(x) = \frac{1}{\log x}.$$

We observe that

$$\frac{1}{\log x} \sim \frac{\pi(x)}{x}$$

(π = counting function of $\mathbb{P}$.)

So, let's try

$$L(x) = \frac{H(x)}{x}$$

in the general case.

Integration by parts formula gives

$$\begin{aligned} F_\mu(n) &= \sum_{\substack{k \leq n \\ k \in \mathbb{H}}} \frac{1}{k} \\ &= L(n) + \int_1^n \frac{L(t)}{t} \\ &\sim L(n) + \sum_{k=1}^{n} \frac{L(k)}{k} \\ &= L(n) + F_\mu(n) \end{aligned}$$

i.e.

$$F_\mu(n) \sim L(n) + F_\mu(n)$$

hence, (a) of the general Theorem follows if we put a condition on L that assures

$$L(n) = o\left(\sum_{k=1}^{n} \frac{L(k)}{k}\right).$$

Of course this is true if $L = \text{const.}$!!

But it is not difficult to see that any slowly varying function L will produce the same result!! The reason is that slowly varying functions behave nearly as constants in many situations. Hence $\mathbb{H}$ is regular (in this particular case the exponent of regularity is 1).

The next step is to see that we can eliminate L in the definition of the measure μ, again because it is slowly varying, hence near to a constant!

Verifying condition (b) of the general Theorem requires the same technique as above but in the calculations we only need that A satisfies

$$p_n \sim \sigma q_n$$

$(\sigma < 1)$.
This answers the second question (which A besides A_q).

REFERENCES

[F-L] **A. Fuchs - G. Letta**, *Le problème du premier chiffre décimal pour les nombres premiers*, The Electronic Journal of Combinatorics 3 (2) (1996) R25.

[G] **R. Giuliano Antonini**, *Comparaison de densités sur certains ensembles d'entiers*, Rend. Acc. Naz. delle Scienze detta dei XL, Mem. di Matematica e Applicazioni, 115 (1997), Vol. XXI, 1, 147-155.

[K-N-R-S] **S. Kanemitsu - K. Nagasaka - G. Rauzy - J. S. Shiue**, *On Benford's law: the first digit problem*, L. N. 1299, Springer, 1988, 158-169.

[P-S] **G. Pólya, G. Szegö**, *Problems and Theorems in Analysis I*, Springer, 1978.

[R] **R. A. Raimi**, *The first digit problem*, Amer. Math. Monthly 83 (1976), 521-538.