Reprinted from
Kybernetes 1982, Vol. 11 (3) Printed in Great Britain

The Probabilistic–Informational Concept of an Opacity Functional

M. Mugur-Schächter and N. Hadjisavvas

Kybernetes
1982, Vol. 11, pp. 189–193

THE PROBABILISTIC—INFORMATIONAL CONCEPT OF AN OPACITY FUNCTIONAL: NEW PROOF OF ITS MAIN PROPERTIES

M. MUGUR-SCHÄCHTER and N. HADJISAVVAS

Laboratoire de Mécanique Quantique de l'Université de Reims (France).

(Received March 31, 1981)

A new compact proof is given for the main properties of a previously defined and studied probabilistic—informational concept, the functional of opacity of a statistics with respect to the underlying probability law.

1. BRIEF RESTATEMENT OF PREVIOUS RESULTS

In previous papers[1,2] one of us has constructed a new probabilistic—informational concept, the opacity functional of a statistics:

Consider a finite universe $\mathscr{U} = \{\epsilon_i, i = 1, 2, \ldots, \lambda\}$ of λ elementary events ϵ_i, of which the probabilities are $p(\epsilon_i) = p_i$. The set $\{p_i, i = 1, 2, \ldots, \lambda\}$ of elementary probabilities is termed *the fundamental probability law*. Let $\mathscr{U}^N$ be the metauniverse consisting of all the possible sequences σ^N of N events $\epsilon_i \in \mathscr{U}$, different or *not* and taken in any order, $\sigma^N = (\epsilon_{i_1}, \ldots, \epsilon_{i_l}, \ldots, \epsilon_{i_N})$, $i_l = 1, 2, \ldots, \lambda$. Furthermore, consider any set $\mathscr{S}_j = \{f_{ij}, i = 1, 2, \ldots, \lambda\}$ of λ rational numbers f_{ij}, $0 \leqslant f_{ij} \leqslant 1$, $\sum_{i=1}^{\lambda} f_{ij} = 1$. The λ relative frequencies $n_{i\sigma}/N$ realized for the ϵ_i in any chosen sequence $\sigma^N \in \mathscr{U}^N$, $\forall N$ will form some such set $\mathscr{S}_j$ which therefore is termed *a statistics j*.

Obviously one given statistics j can be found realized in various sequences σ^N differing either by the values of N, or by the order of the ϵ_{i_l} in σ^N, or by both. We now fix N and consider the set $\{\sigma^N\}_j$ of all the sequences $\sigma^N \in \mathscr{U}^N$ possessing the same statistics j but differing from one another by the order of the ϵ_{i_l}. The set $\{\sigma^N\}_j$ belongs to the total tribe on $\mathscr{U}^N$, where it defines the event consisting of the realization of the statistics j via a sequence $\sigma^N \in \mathscr{U}^N$. The probability $p(\{\sigma^N\}_j) = p^N(j)$ of this event can be calculated as a function of the fundamental probability law $\{p_i, i = 1, 2, \ldots, \lambda\}$ and of the considered statistics j; then one can form the quantity $-\log p^N(j)/N$. If it is assumed that

$$p(\sigma^N) = \prod_{i=1}^{\lambda} p_i^{n_{i\sigma}}, \quad \forall \sigma^N \in \mathscr{U}^N, \ \forall N, \tag{1}$$

then the form found for $-\log p^N(j)/N$ is ([1] pg 40).

$$\frac{-\log p^N(j)}{N} = \sum_{i=1}^{\lambda} \left(\frac{n_{ij}}{N} \log \frac{n_{ij}}{N} - \frac{n_{ij}}{N} \log p_i \right) + \frac{T(\lambda, N, j)}{N} \tag{2}$$

where $n_{ij} = N f_{ij}$ and $T(\lambda, N, j)$ is a sum of terms depending upon λ, N, j:

$$T(\lambda, N, j) = -\frac{1}{2}\left[\log N - \sum_{i=1}^{\lambda} \log n_{ij} + (1 - \lambda) \log 2\pi\right] - \log(1 + f(N)) + \sum_{i=1}^{\lambda} \log(1 + f(n_{ij})) \tag{3}$$

Here, f is a function with $\lim_{x \to 0} f(x) = 0$, and we introduce the convention that $\log 0 = 0$.

Now, it is possible to qualify the evolution of (2) when N increases, *while* (j) *remains fixed*:

Consider a sequence $(N_{1j} < N_{2j} < \cdots N_{kj} < \cdots)$ of integers N_{kj} such that

$$\forall N_{kj}, \left(\exists n_{ik}, \forall i : \frac{n_{ik}}{N_{kj}} = f_{ij}, \sum_{i=1}^{\lambda} n_{ik} = N_{kj} \right).$$

If in (2) N is increased towards infinity exclusively via values N_{kj} belonging to the sequence defined above, then the statistics $j = \{f_{ij}, i = 1, 2, \ldots, \lambda\}$ can be maintained invariant. Let us denote this operation by $\lim N \to \infty / j$ fixed. When applied to the quantity $\dfrac{-\log p^N(j)}{N}$ from (2), this operation yields (ref. 1, p. 42)

$$\lim N \to \infty / j \text{ fixed} \left(\frac{-\log p^N(j)}{N} \right)$$

$$= \sum_{i=1}^{\lambda} f_{ij} \log f_{ij} - \sum_{i=1}^{\lambda} f_{ij} \log p_i$$

$$= \sum_{i=1}^{\lambda} f_{ij} \log \frac{f_{ij}}{p_i} = \Omega(j/\{p_i\}) \tag{4}$$

where $\Omega(j/\{p_i\})$ is a real number depending on the fundamental probability law $\{p_i, i = 1, 2, \ldots, \lambda\}$ and on the statistics j considered. The quantity $\Omega(j/\{p_i\})$ has been termed the functional of opacity of the statistics j with respect to the underlying probability law $\{p_i, i = 1, 2, \ldots, \lambda\}$, in short, *the opacity of* j *with respect to the* p_i.

The first term $\sum_1 f_{ij} \log f_{ij}$ in the opacity functional has the structure of an entropy function for the statistics j, characterizing this statistics *in itself, independently* of the probability law $\{p_i, i = 1, 2, \ldots, \lambda\}$ which acts while j emerges. The second term $\sum_i f_{ij} \log p_i$—termed the modulation function of the probability law $\{p_i, i = 1, 2, \ldots, \lambda\}$ by the statistics j—connects the considered statistics j with the underlying probability law:

$$\sum_{i=1}^{\lambda} f_{ij} \log p_i = \sum_i \frac{n_{ij}}{N} \log p_i = \frac{1}{N} \log \prod_i p_i^{\frac{n_{i\sigma}}{N}}$$

$$= \frac{1}{N} \log p(\sigma^N), \quad \forall \sigma^N \in \mathcal{U}^N, \forall N$$

The properties of the opacity functional (4) bring forth new and profound relations between the entropy $-\sum_i f_{ij} \log f_{ij}$ of the statistics j, the informational entropy $-\sum_1 p_i \log p_i$ of the fundamental probability law which acts while j emerges, the metaprobability $p^N(j)$ of j, and the (weak) law of large numbers (ref. 1, pp. 43–65). These relations have been established before via a succession of rather complicated proofs. The specific aim of this work is to yield a more synthetic view on these relations by help of a vector representation for the fundamental probability law and for a statistics.

2. NEW COMPACT PROOF OF THE MAIN PROPERTIES OF THE OPACITY

Let us represent the fundamental probability law $p_i, i = 1, 2, \ldots, \lambda$ by a vector $\mathbf{p} = (p_1, \ldots, p_\lambda) \in R^\lambda$. Let us furthermore represent a statistics $\mathcal{S}_j$ by a vector $f_j = (f_{1j}, f_{2j}, \ldots, f_{\lambda j})$ of the same space.

Consider now the probability space $(\mathcal{U}^N, \tau, p^N)$ where τ is the total tribe on $\mathcal{U}^N$ and p^N is any probability measure on τ (restriction to the law (1) characterizing independent trials will be considered later). We can define in this space a vectorial random variable $\mathbf{Y}_N$ by

$$\mathbf{Y}_N(\sigma^N) = \left(\frac{n_{i\sigma}}{N} \right)_{1 \leq i \leq \lambda}.$$

It is well known that, under certain circumstances, the vectorial random variable $\mathbf{Y}_N$ tends in probability towards $\mathbf{p}$ as N increases towards infinity (the weak law of large numbers). This fact can be tightly connected with the opacity functional and thereby with the concepts of statistical entropy and informational entropy. We shall now show how this connection can be expressed in vector terms.

Consider the first term in the second member of (2) as a random variable:

$$\Omega^N = \sum_{i=1}^{\lambda} Y_{Ni} \log \frac{Y_{Ni}}{p_i}$$

defined on the space $(\mathcal{U}^N, \tau, p^N)$, $\forall N$. Let us call this variable the opacity variable. We begin by showing that

Proposition 1. The opacity variable Ω^N is non-negative. Furthermore, one has $\Omega^N(\sigma^N) = 0$ for a $\sigma^N \in \mathcal{U}^N$ iff $\mathbf{Y}_N(\sigma^N) = \mathbf{p}$ for this σ^N.

Proof. We shall make use of the following inequality of Cziszar,[3] which will be also of use for the proof of Theorem 1: Let $(\alpha_i)_{1 \leqslant i \leqslant \lambda}$, $(\beta_i)_{1 \leqslant i \leqslant \lambda}$ be two families of numbers. Then:

$$\left(\alpha_i \geqslant 0, \beta_i > 0, \sum_i \alpha_i = \sum_i \beta_i = 1\right)$$

$$\Rightarrow \sum_i \alpha_i \log \frac{\alpha_i}{\beta_i} \geqslant \frac{1}{2} \sum_i |\alpha_i - \beta_i|^2. \quad (5)$$

Taking $\alpha_i = Y_{N_i}$, $\beta_i = p_i$ we see immediately from (5) that $\Omega^N \geqslant 0$. It is now obvious that if $\mathbf{p}$ is a possible value for $\mathbf{Y}_N$ and if $\mathbf{Y}_N(\sigma^N) = \mathbf{p}$, then $\Omega^N(\sigma^N) = 0$. On the other hand, if $\Omega^N(\sigma^N) = 0$ then (5) shows that $Y_{N_i} = p_i$, i.e. $\mathbf{Y}_N(\sigma^N) = \mathbf{p}$.

We now shall show that

Theorem 1. The following two conditions are equivalent:

(a) the vectorial random variable $\mathbf{Y}_N$ converges in probability towards $\mathbf{p}$ when $N \to \infty$ (weak law of large numbers)

(b) the opacity variable Ω^N converges in probability towards zero when $N \to \infty$.

Proof.

(a) $\Rightarrow$ (b):

Condition (a) can be written as: $\forall \eta > 0$, $\forall \delta > 0$, $\exists N_0(\eta, \delta)$:

$$N \geqslant N_0(\eta, \delta) \Rightarrow p^N\left(\bigcap_{i=1}^{\lambda} [|Y_{N_i} - p_i| \leqslant \eta]\right) \geqslant 1 - \delta. \quad (6)$$

By the continuity of the logarithm function one has:

$$\forall \epsilon > 0, \exists \eta_i(\epsilon) > 0: |y - p_i| < \eta_i(\epsilon) \Rightarrow |\log y - \log p_i| < \epsilon$$

so that the following inclusion between events does hold:

$$\bigcap_{i=1}^{\lambda} [|Y_{N_i} - p_i| < \eta_i(\epsilon)] \subset \bigcap_{i=1}^{\lambda} \left[\left|\log \frac{Y_{N_i}}{p_i}\right| < \epsilon\right] \quad (7)$$

Taking $\eta(\epsilon) = \inf_{1 \leqslant i \leqslant \lambda} \eta_i(\epsilon)$ we have also

$$\bigcap_{i=1}^{\lambda} [|Y_{N_i} - p_i| < \eta(\epsilon)] \subset \bigcap_{i=1}^{\lambda} [|Y_{N_i} - p_i| < \eta_i(\epsilon)]. \quad (8)$$

Finally, since

$$\Omega^N = |\Omega^N| \leqslant \sum_i Y_{N_i} \left|\log \frac{Y_{N_i}}{p_i}\right|$$

one has

$$\bigcap_{i=1}^{\lambda} \left[\left|\log \frac{Y_{N_i}}{p_i}\right| < \epsilon\right] \subset [\Omega^N < \epsilon]. \quad (9)$$

Combining (7), (8), (9), we see that condition (a), i.e. relation (6), implies

$$\forall N \geqslant N_0(\eta(\epsilon), \delta): p^N([\Omega^N < \epsilon])$$

$$\geqslant p^N\left(\bigcap_{i=1}^{\lambda} [|Y_{N_i} - p_i| < \eta]\right) \geqslant 1 - \delta$$

i.e. condition (b).

(b) $\Rightarrow$ (a):

Condition (b) reads:

$\forall \delta > 0$, $\forall \eta > 0$, $\exists N_0(\delta, \eta)$ such that

$$N \geqslant N_0(\delta, \eta) \Rightarrow p^N([\Omega^N \leqslant \eta]) \geqslant 1 - \delta.$$

Applying inequality (5) for $\alpha_i = Y_{N_i}$ and $\beta_i = p_i$ we get

$$\Omega^N = \sum_i Y_{N_i} \log \frac{Y_{N_i}}{p_i} \geqslant \tfrac{1}{2} \sum_i (Y_{N_i} - p_i)^2.$$

It follows that

$$\bigcap_{i=1}^{\lambda} \left[|Y_{N_i} - p_i|^2 > \frac{2\eta}{\lambda}\right] \subset \left[\sum_{i=1}^{\lambda} |Y_{N_i} - p_i|^2 > 2\eta\right] \subset [\Omega^N > \eta]$$

and for $N \geqslant N_0(\delta, \eta)$,

$$p^N\left(\bigcap_{i=1}^{\lambda} \left[|Y_{N_i} - p_i| > \sqrt{\frac{2\eta}{\lambda}}\right]\right) \leqslant p^N[\Omega^N > \eta] \leqslant \delta$$

which shows (a).

Since the proof of Theorem 1 is valid not only for independent trials but also for Markov chains, it yields a necessary and sufficient condition for the ergodicity of Markov chains expressed *in terms of entropy* (by definition a Markov chain is called "ergodic" if the weak law of large numbers holds.[4])

But let us now restrict the examination to the special case of independent trials. Theorem 1 establishes that for this case (in particular) the weak law of large numbers entails that the opacity variable Ω^N tends in probability towards zero when N tends towards infinity. Now, this fact can be established also without any reference to the law of large numbers. The proof can be achieved in vector terms as follows:

Theorem 2: If the N trials are independent, then the opacity variable Ω^N tends towards zero in probability when $N \to \infty$.

Proof We wish to show that

$$\forall \epsilon > 0 : p^N([\Omega^N > \epsilon]) \xrightarrow[N\to\infty]{} 0. \qquad (10)$$

The events $\left[\mathbf{Y}_N = \frac{\mathbf{x}}{N}\right]$ for different values of $\mathbf{x} = (x_1, \ldots, x_\lambda)$ such that $\sum_i x_i = N$, $x_i \in N$ are mutually disjoint and cover $\mathscr{U}^N$. Thus

$$[\Omega^N > \epsilon] = \bigcup_{\mathbf{x}} \left([\Omega^N > \epsilon] \cap \left[\mathbf{Y}_N = \frac{\mathbf{x}}{N}\right]\right)$$

which implies

$$p^N([\Omega^N > \epsilon]) = \sum_{\mathbf{x}} p^N\left([\Omega^N > \epsilon] \cap \left[\mathbf{Y}_N = \frac{\mathbf{x}}{N}\right]\right) \qquad (11)$$

So we have to estimate the terms in the right member of (11), using relations (2) and (3). From our convention $\log 0 = 0$ we deduce

$$\forall \mathbf{x} : \sum_i \log x_i \geqslant 0. \qquad (12)$$

Since $f(n) \xrightarrow[n\to\infty]{} 0$, f is bounded, and so is also $\log(1 + f(n))$. Thus there exists a constant $B \in R$ such that

$$\forall N, \forall \mathbf{x} : \frac{\lambda - 1}{2} \log 2\pi - \log(1 + f(N)) + \sum_{i=1}^{\lambda} \log(1 + f(x_i)) > B. \qquad (13)$$

Using (3), (12), (13) we find

$$\forall N \in N, \forall \mathbf{x} : \frac{T}{N} > -\frac{\log N}{N} + \frac{B}{N}. \qquad (14)$$

Since

$$-\frac{\log N}{N} + \frac{B}{N} \xrightarrow[N\to\infty]{} 0,$$

we deduce

$$\forall \epsilon > 0, \exists N_0(\epsilon) : N \geqslant N_0(\epsilon) \Rightarrow \frac{T}{N} > -\frac{\log N}{N} + \frac{B}{N} > -\frac{\epsilon}{2}. \qquad (15)$$

Now (15) together with (2) imply that if $N \geqslant N_0(\epsilon)$ and $p^N\left([\Omega^N > \epsilon] \cap \left[\mathbf{Y}_N = \frac{\mathbf{x}}{N}\right]\right) \neq 0$ then

$$\frac{-\log p^N\left([\Omega^N > \epsilon] \cap \left[\mathbf{Y}_N = \frac{\mathbf{x}}{N}\right]\right)}{N} > \epsilon - \frac{\epsilon}{2} = \frac{\epsilon}{2}$$

and thus

$$N \geqslant N_0(\epsilon) \Rightarrow p^N\left([\Omega^N > \epsilon] \cap \left[\mathbf{Y}_N = \frac{\mathbf{x}}{N}\right]\right) < \exp\left(-\frac{N\epsilon}{2}\right). \qquad (16)$$

For a given N, the number[5] of values of $\mathbf{x}$ for which

$$p\left([\Omega^N > \epsilon] \cap \left[\mathbf{Y}_N = \frac{\mathbf{x}}{N}\right]\right) \neq 0$$

is inferior to the number of *all* possible values of $\mathbf{x} = (x_1, \ldots, x_\lambda)$ with $x_i \in N$, $\sum_i x_i = N$, and this latter number equals $\frac{(N + \lambda - 1)!}{(\lambda - 1)!N!}$.

Combining this result with (16) and (11) we obtain $\forall N \geqslant N_0(\epsilon)$:

$$p^N([\Omega^N > \epsilon]) < \frac{(N+\lambda-1)!}{(\lambda-1)!N!} \exp\left(-\frac{N\epsilon}{2}\right)$$

$$= \frac{(N+1)(N+2)\dots(N+\lambda-1)}{(\lambda-1)!}$$

$$\times \exp\left(-\frac{N\epsilon}{2}\right)$$

$$\leqslant \frac{(N+\lambda-1)^\lambda}{(\lambda-1)!} \exp\left(-\frac{N\epsilon}{2}\right)$$

$$\xrightarrow[N\to\infty]{} 0.$$

Consequently $\forall \epsilon > 0$: $p^N([\Omega^N > \epsilon]) \xrightarrow[N\to\infty]{} 0$.

This result shows that the opacity variable is a concept self-consistent with respect to the law of large numbers, entailing this law as a *consequence* of the dynamics of the metaprobability $p^N(j)$, while N tends towards infinity.

3. CONCLUSION

Compared to the previous study of the opacity functional, the proofs given in this work offer an improved and more synthetic perception of the significant connections existing between the metaprobability $p^N(j)$ of a statistics j, the entropy of j, the informational entropy $-\sum_i p_i \log p_i$ of the fundamental probability law, which acts while the statistics j emerges, and the weak law of large numbers.

ACKNOWLEDGEMENT

This work has gained much from exchanges with Professor Raymond Payen.

REFERENCES

1. M. Mugur-Schächter, *Ann. Inst. Henri Poincaré* **32** (6), 33 (1980).
2. M. Mugur-Schächter, *C.R.A.S.* **288A,** 771 (1979).
3. I. Cziszar, *Stud. Scient. Math. Hung.* **2,** 299 (1967).
4. Khinchin, *Mathematical Foundations of Information Theory* (Dover, New York, 1957).
5. L. Comptet, *Analyse Combinatoire* (Presses Univ. de France, Paris, 1970).