# The Principle of Maximum Entropy

Silviu Guiasu and Abe Shenitzer

## Mathematical Modelling and Variational Principles

There is no need to stress the importance of variational problems in mathematics and its applications. The list of variational problems, of different degrees of difficulty, is very long, and it stretches from famous minimum and maximum problems of antiquity, through the variational problems of analytical mechanics and theoretical physics, all the way to the variational problems of modern operations research. While maximizing or minimizing a function or a functional is a routine procedure, some special variational problems give solutions which either unify previously unconnected results or match surprisingly well the results of our experiments. Such variational problems are called variational principles. Whether or not the architecture of our world is based on variational principles is a philosophical problem. But it is a sound strategy to discover and apply variational principles in order to acquire a better understanding of a part of this architecture. In applied mathematics we get a model by taking into account some connections and, inevitably, ignoring others. One way of making a model convincing and useful is to obtain it as the solution of a variational problem.

The aim of the present paper is to bring some arguments in favour of the promotion of the variational problem of entropy maximization to the rank of a variational principle.

## Entropy as a Measure of Uncertainty

Sometimes a variational principle deals with the maximization or minimization of a function or a functional without special significance. In such cases the acceptance of the variational principle is justified by the properties of its solution. A relevant example is the principle of minimum action in analytical mechanics. Here the so called "action" has no direct and natural physical interpretation but the solution (the Hamilton canonical equations) gives just the law of motion. In case of the principle of maximum entropy, the function which is maximized, namely the entropy, does have remarkable properties entitling it to be considered a good measure of the amount of uncertainty contained in a probability distribution.

Let $\bar{p} = (p_1, \ldots, p_m)$ be a finite probability distribution, i.e., $m$ real numbers satisfying

$$p_k \geqslant 0, \ (k = 1, \ldots, m); \ \sum_{k=1}^{m} p_k = 1. \tag{1}$$

The number $p_k$ may represent the probability of the $k$-th outcome of a probabilistic experiment or the probability of the $k$-th possible value taken on by a finite discrete random variable.

The entropy attached to the probability distribution (1) is the number

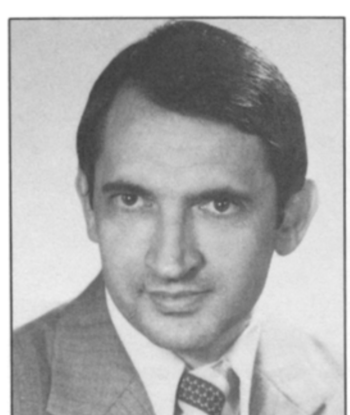
$$H_m(\bar{p}) = H_m(p_1, \ldots, p_m) = -\sum_{k=1}^{m} p_k \ln p_k \tag{2}$$

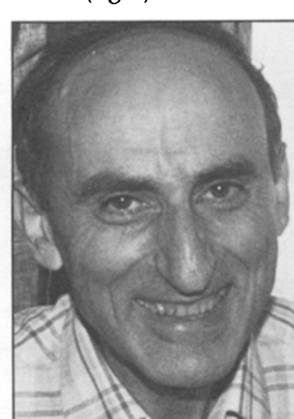
where we put $0 \cdot \ln 0 = 0$ to insure the continuity of the function $-x \ln x$ at the origin. For each positive integer $m \geqslant 2$, $H_m$ is a function defined on the set of probability distributions satisfying (1).

Entropy has several properties with interesting interpretations. We mention some of them.

**1.** $H_m(\bar{p}) \geqslant 0$, continuous, and invariant under any permutation of the indices.

**Silviu Guiasu** (left) and **Abe Shenitzer** (right)





2. If $\overline{p}$ has only one component which is different from zero (i.e., equal to 1) then $H_m(\overline{p}) = 0$.

3. $H_m(p_1, \ldots, p_m) = H_{m+1}(p_1, \ldots, p_m, 0)$.

4. $H_m(p_1, \ldots, p_m) \leq H_m(1/m, \ldots, 1/m)$, with equality if and only if $p_k = 1/m$, $(k = 1, \ldots, m)$.

5. If $\overline{\pi} = (\pi_{1,1}, \ldots, \pi_{m,n})$ is a joint probability distribution whose marginal probability distributions are $\overline{p} = (p_1, \ldots, p_m)$ and $\overline{q} = (q_1, \ldots, q_n)$, respectively, then

$$H_{mn}(\pi_{1,1}, \ldots, \pi_{m,n}) = H_m(p_1, \ldots, p_m) + \sum_{k=1}^{m} p_k H_n(\pi_{k,1}/p_k, \ldots, \pi_{k,n}/p_k) \tag{3}$$

where the conditional entropy $H_n(\pi_{k,1}/p_k, \ldots, \pi_{k,n}/p_k)$ is computed only for those values of $k$ for which $p_k \neq 0$.

6. With the notations given above

$$\sum_{k=1}^{m} p_k H_n(\pi_{k,1}/p_k, \ldots, \pi_{k,n}/p_k) \leq H_n(q_1, \ldots, q_n) \tag{4}$$

with equality if and only if

$$\pi_{k,\ell} = p_k q_\ell \; (k = 1, \ldots, m; \ell = 1, \ldots, n),$$

in which case (3) becomes

$$H_{mn}(\overline{\pi}) = H_m(\overline{p}) + H_n(\overline{q}).$$

All these properties can be proved in an elementary manner. Without entering into the technical details, we note that properties 1–3 are obvious while property 5 can be obtained by a straightforward computation taking into account only the definition of entropy. Finally, from Jensen's inequality

$$\sum_{k=1}^{m} a_k f(b_k) \leq f\left(\sum_{k=1}^{m} a_k b_k\right)$$

applied to the concave function $f(x) = -x \ln x$, we obtain property 4 by putting $a_k = 1/m$, $b_k = p_k$, $k = 1, \ldots, m$, and the inequality (4) by putting $a_k = p_k$, $b_k = \pi_{k,\ell}/p_k$, $k = 1, \ldots, m$, for any $\ell = 1, \ldots, n$, and, in the last case, summing the resulting $n$ inequalities.

Interpretation of the above properties agrees with common sense, intuition, and the reasonable requirements that can be asked of a measure of uncertainty. Indeed, a probabilistic experiment which has only one possible outcome (that is, a strictly deterministic experiment) contains no uncertainty at all; we know what will happen before performing the experiment. This is just property 2. If to the possible outcomes of a probabilistic experiment we add another outcome having the probability zero, the amount of uncertainty with respect to what will happen in the experiment remains unchanged (property 3). Property 4 tells us that in the class of all probabilistic experiments having $m$ possible outcomes, the maximum uncertainty is contained in the special probabilistic experiment whose outcomes are equally likely. Before interpreting the last two properties let us consider two discrete random variables $X$ and $Y$, whose ranges contain $m$ and $n$ numerical values, respectively. Using the same notations as in property 5, suppose that $\overline{\pi}$ is the joint probability distribution of the pair $(X, Y)$, and $\overline{p}$ and $\overline{q}$ are the marginal probability distributions of $X$ and $Y$, respectively. In this case equality (3) may be written more compactly

$$H(X, Y) = H(X) + H(Y|X) \tag{5}$$

where

$$H(X, Y) = H_{mn}(\pi_{1,1}, \ldots, \pi_{m,n})$$
$$H(X) = H_m(p_1, \ldots, p_m)$$

and where

$$H(Y|X) = \sum_{k=1}^{m} p_k H_n(\pi_{k,1}/p_k, \ldots, \pi_{k,n}/p_k)$$

is the conditional entropy of $Y$ given $X$. According to (5), the amount of uncertainty contained in a pair of random variables (or, equivalently, in a compound—or product—probabilistic experiment) is obtained by summing the amount of uncertainty contained in one component (say $X$) and the uncertainty contained in the other component ($Y$) conditioned by the first one ($X$). Similarly, we get for $H(X, Y)$ the decomposition

$$H(X, Y) = H(Y) + H(X|Y) \tag{6}$$

where,

$$H(Y) = H_n(q_1, \ldots, q_n)$$

and

$$H(X|Y) = \sum_{=1}^{n} q_\ell H_m\left(\frac{\pi_{1,\ell}}{q_\ell}, \ldots, \frac{\pi_{m,\ell}}{q_\ell}\right).$$

Here

$$H_m(\pi_{1,\ell}/q_\ell, \ldots, \pi_{m,\ell}/q_\ell)$$

is the conditional entropy of $X$ given the $\ell$-th value of $Y_\ell$. $H_m$ is defined only for those values of $\ell$ for which $q_\ell > 0$. From (5) and (6) we get

$$H(X) - H(X|Y) = H(Y) - H(Y|X)$$

which is the so-called "uncertainty balance", the only conservation law for entropy.

Finally, property 6 shows that some data on $X$ can only decrease the uncertainty on $Y$, namely

$$H(Y|X) \leq H(Y) \tag{7}$$

with equality if and only if $X$ and $Y$ are independent. From (5) and (7) we get

$$H(X, Y) \leq H(X) + H(Y)$$

with equality if and only if $X$ and $Y$ are independent. Fortunately this inequality holds for any number of components. More generally, for $s$ random variables with arbitrary finite range we can write

$$H(X_1, \ldots, X_s) \leq H(X_1) + \ldots + H(X_s)$$

with equality if and only if $X_1, \ldots, X_s$ are globally independent. Therefore

$$W(X_1, \ldots, X_s) = \sum_{i=1}^{s} H(X_i) - H(X_1, \ldots, X_s) \geq 0$$

measures the *global dependence* between the random variables $X_1, \ldots, X_s$, that is, the extent to which the system $(X_1, \ldots, X_s)$, due to interdependence, makes up "something more" than the mere juxtaposition of its components. In particular, $W = 0$ if and only if $X_1, \ldots, X_s$ are independent.

Note that the difference between the amount of uncertainty contained by the pair $(X, Y)$ and the amount of dependence between the components $X$ and $Y$, namely,

$$d(X, Y) = H(X, Y) - W(X, Y)$$

or, equivalently,

$$\begin{aligned} d(X, Y) &= 2H(X, Y) - H(X) - H(Y) \\ &= H(X|Y) + H(Y|X), \end{aligned}$$

is a *distance* between the random variables $X$ and $Y$, with the two random variables considered identical if either one completely determines the other, or if $H(X|Y) = 0$ and $H(Y|X) = 0$. Therefore, the "pure randomness" contained in the pair $(X, Y)$, i.e., the uncertainty of the whole, minus the dependence between the components, measured by $d(X, Y)$, is a distance. This geometrizes chaos!

Discrete entropy as a measure of uncertainty was introduced by C. E. Shannon [12] by analogy with Boltzmann's $H$ function [1] in statistical mechanics. It was also used by Shannon as a measure of information, considering information as removed uncertainty. *Before* a probabilistic experiment is performed the entropy measures the amount of uncertainty associated with the possible outcomes. *After* the experiment the entropy measures the amount of supplied information. We stress that this is the first time a mathematical function has aimed to measure the uncertainty contained in a probabilistic experiment—an entity so different from measurable characteristics of the real world such as length, area, volume, temperature, pressure, mass, charge, etc.

Is the Shannon entropy unique? The answer depends on what properties are taken as the axioms for the measure of uncertainty. Khintchine [9] proved that properties 1, 3, 4, and 5, taken as axioms (which is quite reasonable from an intuitive point of view), imply uniquely the expression (2) for the measure of uncertainty up to an arbitrary positive multiplicative constant. This allows us to choose arbitrarily a base greater than 1 for the logarithm without affecting the basic properties of the measure.

## The Principle of Maximum Entropy

Let us go back to property 4: The uncertainty is maximum when the outcomes are equally likely. The uniform distribution maximizes the entropy; the uniform distribution contains the largest amount of uncertainty. But this is just Laplace's Principle of Insufficient Reason, according to which if there is no reason to discriminate between two or several events the best strategy is to consider them as equally likely. Of course, for Laplace this was a subjective point of view, based on prudence and on common sense. Indeed, without knowing anything about entropy we apply Laplace's Principle of Insufficient Reason in everyday life, even in analyzing the simplest experiments. Indeed, in tossing a coin we usually attach equal probabilities to the two possible outcomes not after a long series of repetitions of this simple experiment followed by a careful analysis of the stability of the relative frequencies of the possible outcomes but simply because we apply Laplace's Principle and realize that we have no good reasons for discriminating between the two outcomes. But, as we have already seen, if we accept the Shannon entropy as the measure of uncertainty, then property 4 is just the mathematical justification of the Principle of Maximum Entropy, which asserts that entropy is maximized by the uniform distribution when no constraint is imposed on the probability distribution. In such a case, our intuition, based on our past experience, gives us the right solution. But what happens when there are some constraints imposed on the probability distribution?

Before answering this question let us see what kinds of constraints may be imposed. Quite often in applications we have at our disposal one or several mean values of one or several random variables. Thus in statistical mechanics the state functions are random variables because the state space is a probability space and we can measure only some mean values of such state functions. For instance, to each microscopic state there corresponds a well-defined value of the energy of the system. But we cannot determine with certainty the real, unique, microscopic state of the system at some instant $t$, and so we construct instead a probability distribution on the possible states of the system. Then the energy becomes a random variable and what we can really measure, at the macroscopic level, is the mean value of this random variable, i.e., the macroscopic energy. The macroscopic level is the level of mean values and some of these mean values can be measured. But we need a probabilistic model of the microscopic level, i.e., a probability distribution on the possible microscopic states of the system. In general, there are many probability distributions (even an infinity!) compatible with the known mean values. Hence the question: What probability distribution is "best" and with respect to what criterion?

In 1957 E. T. Jaynes [8] gave a very natural criterion of choice by introducing the *Principle of Maximum Entropy:* From the set of all probability distributions compatible with one or several mean values of one or several random variables, choose the one that maximizes Shannon's entropy. Such a probability distribution is the "largest one"; it will ignore no possibility, being the most uniform one, subject to the given constraints. Introduced for solving a problem in statistical mechanics, the Principle of Maximum Entropy has become a widely applied tool for constructing the probability distribution in statistical inference, in decision theory, in pattern-recognition, in communication theory, and in time-series analysis, because in all these areas what we generally know is expressed by mean values of some random variables and what we need is a probability distribution which ignores no possibility subject to the relevant constraints.

To see how this principle works let us take the simplest possible case, the case in which we know the mean value $E(f)$ of a random variable $f$ whose possible values are $f_1, \ldots, f_m$. We need a probability distribution $\bar{p} = (p_1, \ldots, p_m)$,

$$p_k > 0, (k = 1, \ldots, m); \sum_{k=1}^{m} p_k = 1 \tag{8}$$

satisfying the constraint

$$E(f) = \sum_{k=1}^{m} f_k\, p_k. \tag{9}$$

In the trivial case $m = 2$, the mean value $E(f)$ uniquely defines the corresponding probability distribution from the linear equation

$$E(f) = f_1 p_1 + f_2(1 - p_1).$$

But for any $m \geqslant 3$ there is an infinity of probability distributions (8) satisfying (9). Applying the Principle of Maximum Entropy we choose the most uncertain probability distribution, i.e., the probability distribution that maximizes the entropy

$$H_m(\bar{p}) = -\sum_{k=1}^{m} p_k \ln p_k$$

subject to the constraints (8) and (9). Of course, $H_m$ is a concave and continuous function defined in the convex domain characterized by (8) and (9). There is only one global maximum point belonging to the open set

$$\{\bar{p} = (p_1, \ldots, p_m) | p_k > 0,\ k = 1, \ldots, m; \\ \sum_{k=1}^{m} p_k - 1 = 0,\ \sum_{k=1}^{m} f_k\, p_k - E(f) = 0\}.$$

Taking the Lagrange function

$$L = H_m(p_1, \ldots, p_m) - \alpha\left(\sum_{k=1}^{m} p_k - 1\right) \\ - \beta\left(\sum_{k=1}^{m} f_k\, p_k - E(f)\right)$$

where $\alpha$ and $\beta$ are the Lagrange multipliers corresponding to the two constraints, and putting the first order partial derivatives equal to zero we get

$$\frac{\partial L}{\partial p_k} = -\ln p_k - 1 - \alpha - \beta f_k = 0, (k = 1, \ldots, m),$$

$$\frac{\partial L}{\partial \alpha} = 1 - \sum_{k=1}^{m} p_k = 0,$$

$$\frac{\partial L}{\partial \beta} = E(f) - \sum_{k=1}^{m} f_k\, p_k = 0.$$

Thus the solution is

$$p_k = \frac{e^{-\beta_0 f_k}}{\sum_{r=1}^{m} e^{-\beta_0 f_r}}, (k = 1, \ldots, m) \tag{10}$$

where $\beta_0$ is the solution of the exponential equation

$$\sum_{k=1}^{m} [f_k - E(f)]e^{-\beta(f_k - E(f))} = 0. \quad \textbf{(11)}$$

If the random variable $f$ is nondegenerate (i.e., if $f$ takes on at least two different values), such a solution exists and is unique because the function

$$G(\beta) = \sum_{k=1}^{m} [f_k - E(f)]e^{-\beta(f_k - E(f))} \quad \textbf{(12)}$$

is strictly decreasing with

$$\lim_{\beta \to -\infty} G(\beta) = +\infty, \quad \lim_{\beta \to +\infty} G(\beta) = -\infty.$$

We have already seen that when there is no constraint, the solution of the Principle of Maximum Entropy is the uniform probability distribution. When the mean value $E(f)$ of a random variable $f$ is given, then the solution of the Principle of Maximum Entropy is (10) or, equivalently,

$$p_k = \frac{1}{\Phi(\beta)} e^{-\beta_0 f_k}, \ (k = 1, \ldots, m)$$

where

$$\Phi(\beta) = \sum_{k=1}^{m} e^{-\beta f_k}$$

and $\beta_0$ is the unique solution of the equation

$$\frac{d \ ln \ \Phi(\beta)}{d\beta} = -E(f).$$

This is just the Gibbs, or canonical, distribution encountered in almost all books on statistical mechanics and, more recently, in some books on decision theory. Now we see why the canonical distribution is useful in applications: It is the most uncertain one, the most uniform one, it ignores no possibility subject to the constraint given by the mean value $E(f)$.

Since (11) is an exponential equation, its solution $\beta_0$ may be a transcendental number. However, the fact that the function $G$ given by (12) is strictly decreasing permits us to approximate the solution $\beta_0$ with great accuracy.

For instance, let $m = 3$, $f_1 = 12$, $f_2 = 15$, $f_3 = 20$ and the mean value $E(f) = 18.12$. Using a simple TI-57 pocket calculator, we can obtain in a few minutes the solution of equation (11) (correct to six decimals), namely, $\beta_0 \approx -0.2364201$. The corresponding solution of the Principle of Maximum Entropy is $p_1 = 0.1035103$, $p_2 = 0.2103835$, $p_3 = 0.6861062$. This is the most uniform probability distribution compatible with the given mean value. For some other constraints, the exact values of the Lagrange multipliers introduced for maximizing the entropy may be determined exactly. Without entering into technical details we mention some remarkable results relating to the Principle of Maximum Entropy:

a) If $f$ is a random variable whose range is countable, namely, if

$$\{ku | u > 0, \ k = 0,1,2, \ldots\}$$

(this is true of the energy in quantum mechanics—in which case $u$ is the quantum of energy—or of many discrete functions in operations research—in which case $u$ is the unit), and if the mean value $E(f)$ is given, then the probability distribution

$$p_k > 0, \ (k = 0,1, \ldots), \ \sum_{k=0}^{\infty} p_k = 1$$

maximizing the countable entropy:

$$H = -\sum_{k=0}^{\infty} p_k \ ln \ p_k$$

is

$$p_k = \frac{u(E(f))^k}{(u + E(f))^{k+1}}, \ k = 0,1,2, \ldots.$$

We see that the unit $u$ and the mean value $E(f)$ completely determine the solution of the Principle of Maximum Entropy. The importance of this probability distribution is stressed by M. Born [2].

Before discussing the continuous case, we note an unusual property of entropy that permits us to maximize it even when the solution is a sequence satisfying

$$p_k > 0, \ \sum_{k=0}^{\infty} p_k = 1.$$

In such a case, instead of computing the partial derivatives with respect to a countable set of variables ($p_0, p_1, \ldots$ and the Lagrange multipliers corresponding to the constraints), it is enough to take into account the simple equality

$$t \ ln \ t = (t - 1) + \frac{1}{2\tau}(t - 1)^2,$$

true for *any* $t > 0$, where $\tau$, depending on $t$, is a *positive* number located somewhere between 1 and $t$. (This equality is obtained from the Taylor expansion of $t \ln t$ about 1.) Applying this equality and considering the constraint

$$E(f) = \sum_{k=0}^{\infty} ku\, p_k < \infty$$

we have, for $\alpha > 0$, $\beta > 0$,

$$H - \alpha \cdot 1 - \beta E(f) = -\sum_{k=0}^{\infty} p_k \ln(p_k e^{\alpha+\beta ku}) =$$

$$-\sum_{k=0}^{\infty} e^{-\alpha-\beta ku}(p_k e^{\alpha+\beta ku}) \ln (p_k e^{\alpha+\beta ku}) \leqslant$$

$$-\sum_{k=0}^{\infty} e^{-\alpha-\beta ku}(p_k e^{\alpha+\beta ku} - 1) = -1 + \sum_{k=0}^{\infty} e^{-\alpha-\beta ku};$$

here the upper bound is independent of the probability distribution $\{p_k, k = 0,1, \ldots\}$, and we have equality if and only if

$$p_k = e^{-\alpha-\beta ku}, k = 0,1, \ldots$$

From the first constraint

$$\sum_{k=0}^{\infty} e^{-\alpha-\beta ku} = 1$$

we obtain

$$e^{-\alpha} = 1 - e^{-\beta u}$$

and from the second constraint

$$E(f) = \sum_{k=0}^{\infty} ku(1 - e^{-\beta u})\, e^{-\beta ku}$$

we obtain the solution

$$p_k = \frac{u(E(f))^k}{(u + E(f))^{k+1}}, k = 0,1, \ldots$$

b) In the continuous case, suppose that we know the mean value $\mu$ of a positive continuous random variable whose probability density function is square-integrable. In such a case, the continuous entropy

$$H(\delta) = -\int_{-\infty}^{+\infty} \delta(x) \ln \delta(x) dx \tag{13}$$

is maximized by

$$\delta(x) = \begin{cases} \frac{1}{\mu} e^{-\frac{1}{\mu}x} & \text{if } x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

which is just the well-known exponential probability density function. Now we have a justification for the usual assumption in queueing theory that the interarrival time is exponentially distributed. Such a probability distribution is the most uncertain one, the most prudent one, and it ignores no possibility subject to the mean interarrival time $\mu$.

c) Of course, it is possible to have many constraints. Suppose that, in the continuous case, we know both the mean $\mu$ and the variance $\sigma^2$ of a continuous random variable whose probability density function is square-integrable. The agreeable surprise is that, in such a case, the continuous entropy (13) is maximized just by

$$\delta(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, (-\infty < x < +\infty)$$

which is the probability density function of the normal distribution $N(\mu, \sigma^2)$. Now we see why this probability distribution has been frequently used in the applications of statistical inference and why it deserves the adjective "normal"; in the infinite set of square integrable probability density functions defined on the real line with mean $\mu$ and variance $\sigma^2$, the normal distribution (or de Moivre-Laplace-Gauss distribution) is the distribution that is most uncertain and that maximizes the entropy. Entropy would have had to be invented if only to demonstrate this variational property of the normal distribution!

The fact that the Principle of Maximum Entropy can be used to obtain a unified variational treatment of some well-known probability distributions is just one reason for its importance. In fact, we can apply the same strategy (i.e., maximizing the entropy) subject to more numerous and more sophisticated constraints, such as a large number of mean values (moments of order greater than 2) of several random variables. The solution of the Principle of Maximum Entropy will give probability distributions never met before.

We conclude with some comments on what is subjective and what is objective in the use of the Principle of Maximum Entropy. As a variational problem (maximize the entropy subject to constraints expressed by mean values of some random variables) it is as objective as any other mathematical optimization problem. Accepting the probabilistic entropy as a measure of uncertainty and interpreting the solution of the Principle of Maximum Entropy from the viewpoint of the amount of uncertainty contained is, in spite of the "naturalness" of the properties 1–6 above, a subjective attitude. But the fact that some important probability distributions from statistical inference (exponential, canonical, uniform, and, above all, the most important one, the normal distribution) are solutions of it enables us to say that the use of the Principle of Maximum Entropy proves to be more than a simple convention.

The Principle of Maximum Entropy has implied both some other entropic variational problems (the minimization of the Kullback-Leibler divergence, the minimization of the interdependence) and many new applications (for example its recent applications in time-series analysis and the entropic algorithm for pattern-recognition which proves to have the smallest mean length); but this is another story.

## References


1. L. Boltzmann (1896) Vorlesungen über Gastheorie, Leipzig: J. A. Barth.
2. M. Born (1969) Atomic Physics. (8-th edition). London-Glasgow: Blakie & Son Ltd.
3. L. L. Campbell (1970) Equivalence of Gauss's principle and minimum discrimination information estimation of probabilities. Ann. Math. Statist., 41, 1011–1015.
4. J. W. Gibbs (1902) Elementary principles in statistical mechanics developed with especial reference to the rational foundation of thermodynamics. New Haven, Conn: Yale University Press.
5. S. Guiaşu (1977) Information theory with applications. New York-London-Düsseldorf: McGraw-Hill.
6. S. Guiaşu, T. Nguyen Ky (1982) On the mean length of the entropic algorithm of pattern-recognition. Journal of Combinatorics, Information & System Sciences, 7, 203–211.
7. S. Guiaşu, R. Leblanc, C. Reischer (1982) On the principle of minimum interdependence. Journal of Information & Optimization Sciences, 3, 149–172.
8. E. T. Jaynes (1957) Information theory and statistical mechanics. Phys. Rev. 106, 620–630; 108, 171–182.
9. A. I. Khinchin (1957) Mathematical foundations of information theory. New York: Dover Publications.
10. S. Kullback (1959) Information theory and statistics. New York: Wiley. London: Chapman & Hall.
11. E. Parzen (1982) Maximum entropy interpretation of autoregressive spectral densities. Statistics & Probability Letters 1, 7–11.
12. C. E. Shannon (1948) A mathematical theory of communication. Bell Syst. Techn. Journal, 27, 379–423, 623–656.
13. S. Watanabe (1969) Knowing and guessing. New York: Wiley.



*Silviu Guiasu*
*Department of Mathematics*
*York University*
*4700 Keele Street*
*Downsview, Ont. M3J 1P3*
*Canada*

*Abe Shenitzer*
*Department of Mathematics*
*York University*
*4700 Keele Street*
*Downsview, Ont. M3J 1P3*
*Canada*