Between Truth and Description

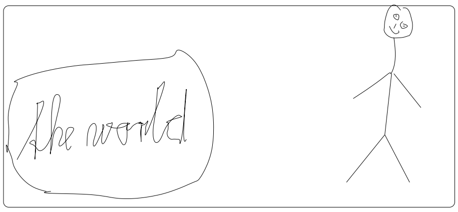
Flemming Topsøe

Abstract: The philosophy behind game-theoretical modelling of conflict situations between “man” , attempting to describe the world, and “nature” trying, so it seems, to conceal the “truth” is discussed. An analysis of instances of such games related to the *Maximum Entropy Principle* is discussed. The possibility of *entropy loss*, a phenomenon related to distributions with ultra heavy tails is discussed.

MSC 2000: 91A05, 91B50, 94A15, 94A17
Key words: equilibrium, maximum entropy, loss of entropy, hyperbolic distribution

1 Philosophy

A substantial portion of science deals with the situation indicated in the figure. The real world – a suitable part of it – is looked upon by the observer. We view the situation as one of *conflict* between two *“players”*: The real world, personified as *NATURE, GOD, THE SYSTEM, THE DEVIL* or some other choice according to individual taste, and then a more easily understandable person: *THE OBSERVER, “YOU”*. Neutrally, we refer to the players as *“Player I”* and *“Player II”*.



We assume that the players have opposing goals, hence the conflict. It is easy to understand what Player II is striving for, typically efficiency of observation, enabling extraction of as much information about the system as possible. It is more questionable to attach a deliberate intent to Player I, to NATURE. And even if we do, why should NATURE interfere with the activities of Player II and bother to make things difficult for man? Has not NATURE once and for all fixed the “laws of NATURE” , so that nothing can be changed, no influence on the outcome of the game exercised as opposed to the situation in a normal conflict between real persons?

In an attempt to answer these questions, consider Player II. As the observer, Player II should take all available information into account but otherwise be neutral

The research is supported by the Danish Natural Science Research Council.

to the possibilities and be prepared for any eventuality. Thinking about it, you realize that though you may well believe in some absolute truth, it is *you* who model what is going on – well, *could* be going on – at Player I's end and, therefore, modelling of the behaviour of Player I is not so much modelling of an absolute but really something that reflects *your* knowledge about the system under study. Thus, Player I is often best thought of as an expression for another side of yourself and you have the schizophrenic situation that you, Player II, are in conflict with another side of yourself, the side expressing knowledge, or perhaps rather absence of knowledge.

To sum-up: Player II is *you* as the observer who can choose any available tool you can think of in fighting your own ignorance, personified as Player I.

As a final philosophical remark, we note that there is a pronounced *asymmetry* between the players. Typically, this manifests itself by the fact that Player II has an optimal strategy to follow whereas this need not be the case for Player I.

2 History

Some hints to the development which reflects the philosophy outlined in Section 1 are in place.

The ideas of Section 1 point to *game theory*, especially the theory of *two-person zero-sum (non-cooperative) games*, a theory initiated in a famous paper by John von Neumann from 1928, [27]. More recent general references include [24] and [16].

The further study of the type of games we have in mind has statistical physics, information theory and mathematical statistics as major sources of inspiration. In particular, the concepts of information theory with their convincing interpretations and the by now well developed general results enable a rather complete study of games of great significance for theory as well as for applications.

The influence from information theory starts, not surprisingly, with Shannons celebrated contribution from 1948, [23], where a new way of thinking was born together with associated key concepts, notably that of *Shannon entropy* (below just *entropy*). In 1951, Kullback and Leibler introduced one further basic concept, *Kullback-Leibler divergence* (below just *divergence*), cf. [21]. Kullback also demonstrated the potential of information theoretical thinking for statistical inference in his monograph [20] from 1959. Of special significance here is Kullback's *minimum divergence principle*.

Regarding influences from physics, our views are consistent with views brought forward in 1957 by Jaynes, [14], see also [15]. As with other sources mentioned, Jaynes' writings basically focus on what goes on at Player I's end. In the foreground here is Jaynes *Maximum Entropy Principle*. The significance of this principle for a large range of theoretical and applied fields is widely recognized, cf. Kapur, [17], for example. Another important contribution, [13] by Ingarden and Urbanik from 1962, is the first example of a phenomenon of *collapse of the entropy function* or *loss of entropy* which we shall deal with in Section 4.

In Čencov [2] from 1972, and Csiszár [3] from 1975, these authors, independently of each other, focus on key elements of information theoretical optimization techniques. A special role is played by the *Pythagorean inequality* and Csiszár's notion of *I-projection.* Again, the focus is on Player I. Later developments, notably initiated by Csiszár discuss the significance for statistical inference. The reader may consult the recent paper [4] by Csiszár and Matus, and references there.

The game theoretical thinking, focusing also on Player II, originated in independent studies by Pfaffelhuber, [22] from 1972, and by the author, [25] from 1979. The latter builds on and expands Csiszárs results. Later studies, partly re-discoveries, are Kazakos [18] and Haussler [12]. A thorough study from 2004 with emphasis on statistical inference based on the game theoretical approach is due to Grünwald and Dawid, [9].

Since 1998, the author and Peter Harremoës have researched the area further. Suffice it here to point to [10], [11] and to [26] and references to be found there. These papers demonstrate the wide applicability of the game theoretical approach. For instance, it follows from Lemma 1 of [26] that the Pythagorean inequality – as well as a *reverse Pythagorean inequality* (relevant at Player II's side of the game) – are general game-theoretical principles which are not necessarily tied to special information theoretical concepts. The generality of the game theoretical approach is also emphasized in [9].

We should also mention research on the *minimum description length principle* initiated by Rissanen. The overall philosophy is independent of, but related to that of the game theoretical approach. For details, see [1].

3 Maximum entropy, the basics

Taking the example of the arguably most important information theoretical game, the *code length game* which is the game behind the maximum entropy principle, we show here and in Section 4 a quick route to the key results.

By $\mathbb{A}$ we denote a countable set, the *alphabet*, by $M_+^1(\mathbb{A})$ the set of probability measures over $\mathbb{A}$ with the topology of pointwise convergence and by $K(\mathbb{A})$ the set of *(idealized) codes* over $\mathbb{A}$, i.e. the set of mappings $\kappa : \mathbb{A} \to [0,\infty]$ such that $\sum_{i\in\mathbb{A}} \exp(-\kappa_i) = 1$. We say that $\kappa \in K(\mathbb{A})$ *is adapted to* $P \in M_+^1(\mathbb{A})$ or that P *is ideal for* κ if, for all $i \in \mathbb{A}$, $\kappa_i = -\ln p_i$, equivalently, $p_i = \exp(-\kappa_i)$. For $P = (p_i)_{i\in\mathbb{A}}$ and $Q = (q_i)_{i\in\mathbb{A}}$ in $M_+^1(\mathbb{A})$, *divergence between* P *and* Q is defined as usual, i.e., observing standard conventions, as

$$D(P\|Q) = \sum_{i\in\mathbb{A}} p_i \ln \frac{p_i}{q_i}\,. \tag{1}$$

From the identity

$$D(P\|Q) = \sum_{i\in\mathbb{A}} p_i\Big(\ln\frac{p_i}{q_i} - (1 - \frac{q_i}{p_i})\Big) \tag{2}$$

we realize that divergence is well-defined, non-negative on all of $M_+^1(\mathbb{A}) \times M_+^1(\mathbb{A})$ and only vanishes on the diagonal $P = Q$. Furthermore, (2) implies that divergence – as a sum of non-negative lower semi-continuous functions – is itself lower semi-continuous, even jointly. Divergence between $P \in M_+^1(\mathbb{A})$ and $\kappa \in K(\mathbb{A})$ is defined by

$$D(P\|\kappa) = D(P\|Q) \text{ with } Q \text{ ideal for } \kappa\,. \tag{3}$$

A *preparation* is a non-empty subset $\mathcal{P}$ of $M_+^1(\mathbb{A})$. If $P \in \mathcal{P}$ we call P a *consistent distribution.* By $\gamma(\mathcal{P})$ we denote the two-person zero-sum game with consistent distributions as strategies for Player I and codes over $\mathbb{A}$ as strategies for Player II. As objective function, taken as a cost to Player II, we take *average code length*

$$\langle \kappa, P \rangle = \sum_{i \in \mathbb{A}} \kappa_i p_i\,. \tag{4}$$

By the general *minimax-inequality*,

$$\sup_{P \in \mathcal{P}} \inf_{\kappa \in K(\mathbb{A})} \langle \kappa, P \rangle \le \inf_{\kappa \in K(\mathbb{A})} \sup_{P \in \mathcal{P}} \langle \kappa, P \rangle\,. \tag{5}$$

The game $\gamma(\mathcal{P})$ is in *equilibrium* if equality holds in (5) and if the common value, the *value of the game*, is finite. *Optimal strategies* are defined in the usual way, e.g. P^* is an optimal strategy for Player I if $P^* \in \mathcal{P}$ and if $\inf_{\kappa \in K(\mathbb{A})} \langle \kappa, P^* \rangle$ coincides with the left hand side of (5).

The infimum on the left hand side of (5) is interpreted as *minimum description length* (description obtained via coding). This is what we understand by *entropy.* It is more conventional to define entropy by the familiar formula

$$H(P) = -\sum_{i \in \mathbb{A}} p_i \ln p_i\,. \tag{6}$$

Then, by the following simple, but crucial identity, the *linking identity*

$$\langle \kappa, P \rangle = H(P) + D(P\|\kappa)\,, \tag{7}$$

we realize that entropy is indeed minimum average code length. Accordingly, an optimal distribution for Player I is the same as a *maximum entropy distribution*, a consistent distribution of maximal entropy. Thus, we are led to Jaynes maximum entropy principle from our game theoretical considerations. The findings so far also justify the notation $H_{max}(\mathcal{P})$ for the left hand side of (5). Further, we put

$$R(\kappa|\mathcal{P}) = \sup_{P \in \mathcal{P}} \langle \kappa, P \rangle \tag{8}$$

and consider this quantity as a *risk* to Player II associated with the strategy κ. We denote the right hand side of (5) by $R_{min}(\mathcal{P})$. Then, an optimal strategy for player II, *an optimal code*, is a code κ for which $R(\kappa|\mathcal{P}) = R_{min}(\mathcal{P})$.

Theorem 3.1 (optimal codes)**.** *For every preparation $\mathcal{P}$, Player II has an optimal strategy. In case $R_{min}(\mathcal{P}) < \infty$, this strategy, the optimal code, is unique.*

Proof. We may assume that $R_{min}(\mathcal{P}) < \infty$. Let $\overline{K}(\mathbb{A})$ denote the set of functions κ on $\mathbb{A}$ with $\sum_{i\in\mathbb{A}} \exp(-\kappa_i) \le 1$ in the topology of pointwise convergence (equivalently, $\overline{K}(\mathbb{A})$ is the closure of $K(\mathbb{A})$). Extend the definitions (4) and (8) to be valid also for $\kappa \in \overline{K}(\mathbb{A})$. Note that the extended function $R(\cdot|\mathcal{P})$ is lower semi-continuous on the compact set $\overline{K}(\mathbb{A})$. It therefore assumes its minimal value. The minimum must be assumed for a $\kappa \in K(\mathbb{A})$ since otherwise we could subtract a positive constant from κ and decrease the risk. Finally, the uniqueness follows by applying the geometric-arithmetic mean inequality to a mixture of two postulated optimal codes (the mixture is in $\overline{K}(\mathbb{A})$ but only in $K(\mathbb{A})$ if the codes are identical). □

Natural questions to look into concern the existence of optimal strategies, also for Player I, and the problem of equilibrium. We start with the second question. First note that there are simple examples, say with preparations consisting of deterministic distributions or of small perturbations of such distributions, which show that these games (with $R_{min}(\mathcal{P}) < \infty$) are not always in equilibrium. Also, equilibrium may fail even when Player I too has a unique optimal strategy.

The discussion of what equilibrium really amounts to involves once more considerations of the compact and convex set $\overline{K}(\mathbb{A})$. Denoting "convex hull" by *co* we find:

Theorem 3.2 (equilibrium)**.** *Assume that $H_{max}(\mathcal{P}) < \infty$. Then a necessary and sufficient condition that $\gamma(\mathcal{P})$ is in equilibrium is that entropy cannot be increased by taking mixtures in the sense that*

$$H_{max}(co\mathcal{P}) = H_{max}(\mathcal{P})\,. \tag{9}$$

Proof. If $\gamma(\mathcal{P})$ is in equilibrium, then

$$H_{max}(co\mathcal{P}) \le R_{min}(co\mathcal{P}) = R_{min}(\mathcal{P}) = H_{max}(\mathcal{P}) \tag{10}$$

and (9) follows.

Then assume that (9) holds. The essential thing to prove is that $co\mathcal{P}$ is in equilibrium. Assume for convenience that $\mathcal{P}$ itself is convex. Note that the map $(\kappa, P) \curvearrowright \langle\kappa, P\rangle$ of $\overline{K}(\mathbb{A}) \times M_+^1(\mathbb{A})$ into $[0,\infty]$ is affine in each variable and lower semi-continuous in the first variable. By *Kneser's minimax theorem*, a classical result, cf. [19], it follows that

$$\sup_{P\in\mathcal{P}} \min_{\kappa\in\overline{K}(\mathbb{A})} \langle\kappa, P\rangle = \min_{\kappa\in\overline{K}(\mathbb{A})} \sup_{P\in M_+^1(\mathbb{A})} \langle\kappa, P\rangle\,^* \tag{11}$$

*Originally Kneser's theorem assumes embedding in topological vector spaces and real-valued mappings. It is, however, easy to extend the result, assuming only embedding in topological cones (hence $\overline{K}(\mathcal{A})$ is embedded in $[0,\infty]^{\mathbb{A}}$) and to allow mappings which may assume the value ∞. Alternatively, one may use Ky Fan's result [6] (which only needs a minor adjustment to cover our set-up). For a discussion of these results, see [8].

By arguments as in the proof of Theorem 3.1, we realize that in (11), $\overline{K}(\mathcal{P})$ may be replaced by $K(\mathcal{P})$. Thus, $\gamma(\mathcal{P})$ is in equilibrium.† □

Games which are in equilibrium have nice properties. We need two concepts: A sequence $(P_n)_{n\geq 1} \subseteq \mathcal{P}$ is *asymptotically optimal for* $\mathcal{P}$ if $H(P_n) \to H_{max}(\mathcal{P})$. Further, $P^* \in M_+^1(\mathbb{A})$ is the *maximum entropy attractor of* $\mathcal{P}$ (or just the *attractor*) if $D(P_n \| P^*) \to 0$ for every asymptotically optimal sequence (P_n). The attractor need not exist, but when it does, it is unique and, by *Pinsker's inequality*, any asymptotically optimal sequence converges to it in the topology of total variation. Note that the attractor need not belong to the preparation under study.

Theorem 3.3 (properties under equilibrium)**.** *Assume that* $\gamma(\mathcal{P})$ *is in equilibrium and denote by* κ^* *the optimal code. Then* $\gamma(\mathcal{P})$ *has a maximum entropy attractor, viz. the ideal distribution* P^* *for* κ^**. Furthermore, for every* $P \in \mathcal{P}$*,*

$$H(P) + D(P\|P^*) \leq H_{max}(\mathcal{P}) \tag{12}$$

and, for every code κ*,*

$$R_{min} + D(P^*\|\kappa) \leq R(\kappa|\mathcal{P})\,. \tag{13}$$

Proof. Let $P \in \mathcal{P}$. Then $\langle \kappa^*, P\rangle \leq R_{min}(\mathcal{P}) = H_{max}(\mathcal{P})$, and (12) follows by the linking identity (7). And from (12) we see that P^* is the attractor.

Now, let $\kappa \in K(\mathbb{A})$ and consider any asymptotically optimal sequence (P_n). Then, $R(\kappa|\mathcal{P}) \geq \langle \kappa, P_n\rangle = H(P_n) + D(P_n\|\kappa)$. Going to the limit, (13) follows. □

Theorems 3.1 and 3.2 suffer from the usual defect of pure existence results. We turn to a simple, yet very useful criterion for equilibrium which allows, at the same time, to identify the optimal objects, assuming that suitable candidates are available. Some definitions: The code κ is *robust for* $\mathcal{P}$ if $\langle \kappa, P\rangle$ is finite and independent of $P \in \mathcal{P}$; and $P \in M_+^1(\mathbb{A})$ is *essentially consistent for* $\mathcal{P}$ if there exists $(P_n)_{n\geq 1} \subseteq \mathcal{P}$ such that $D(P_n\|P) \to 0$; finally, the attractor P^* for $\mathcal{P}$ is the *essential maximum entropy distribution* if $H(P^*) = H_{max}(\mathcal{P})$.

Theorem 3.4 (identification)**.** *Let* $\mathcal{P}$ *be a preparation,* κ^* *a code and* P^* *the ideal distribution for* κ^**.*

(i). If $R(\kappa^*|\mathcal{P}) \leq H(P^*) < \infty$ *and if* P^* *is essentially consistent, then* $\gamma(\mathcal{P})$ *is in equilibrium,* κ^* *is the optimal code and* P^* *the essential maximum entropy distribution for* $\gamma(\mathcal{P})$*.*

(ii). If κ^* *is robust for* $\mathcal{P}$ *and if* P^* *is consistent, then* $\gamma(\mathcal{P})$ *is in equilibrium and both players have unique optimal strategies, viz.* P^* *for Player I (thus* P^* *is the unique maximum entropy distribution of* $\mathcal{P}$*) and* κ^* *for Player II.*

† in [25] a direct proof was given together with a hint that a deeper general minimax result could have been applied (this had Ky Fan's paper [7] in mind). However, all that is needed is Kneser's theorem (note the elegant and elementary proof in [19]) or the related result by Ky Fan, in [6]. This was pointed out recently to the author by Gabor Kassay, who also assisted with other advice regarding game theory and minimax theorems.

Proof. (i): Choose $(P_n)_{n\geq 1} \subseteq \mathcal{P}$ such that $D(P_n\|P^*) \to 0$. By assumption,

$$R_{min}(\mathcal{P}) \leq R(\kappa^*|\mathcal{P}) \leq \limsup_{n\to\infty}\langle \kappa^*, P_n\rangle = \limsup_{n\to\infty} H(P_n) + D(P_n\|P^*) = H_{max}(\mathcal{P})$$

which is finite and we realize that $\gamma(\mathcal{P})$ is in equilibrium with κ^* as optimal code. By Theorem 3.3, P^* is the maximum entropy attractor. Clearly, $H(P^*) = H_{max}(\mathcal{P})$.

(ii): This is an easy corollary to part (i). □

4 Collapse of the entropy function

In order to enable a more familiar formulation of the results to follow, and to simplify notation we now assume that $\mathbb{A} = \mathbb{N}$, the set of natural numbers.

If P^* is the maximum entropy attractor for a game $\gamma(\mathcal{P})$ in equilibrium (now with $\mathcal{P} \subseteq M_+^1(\mathbb{N})$), then $H(P^*) \leq H_{max}(\mathcal{P})$ by lower semi-continuity. The inequality may be strict, cf. [13] and results to follow. If so, we talk about *collapse of the entropy function* or *loss of entropy* (at P^*) , whereas, if equality holds, P^* is the essential maximum entropy distribution.

A distribution $P^* \in M_+^1(\mathbb{N})$ has *potential entropy loss* if it is the maximum entropy attractor for some model $\mathcal{P}$ in equilibrium with collapse of the entropy function: $H(P^*) < H_{max}(\mathcal{P})$. This phenomenon is only possible for attractors with ultra heavy tails. First define a distribution $P^* \in M_+^1(\mathbb{N})$ to be *power-dominated* if, for some $a > 1$, $p_n^* \leq n^{-a}$ for all n sufficiently large. If P^* is not power-dominated, we say that P^* is *hyperbolic*.

Theorem 4.1 (collapse of entropy)**.** *The necessary and sufficient condition that $P^* \in M_+^1(\mathbb{N})$ has potential entropy loss is that P^* is hyperbolic and $H(P^*)$ finite.*

Proof. Assume first that P^* is power-dominated. Clearly then, $H(P^*) < \infty$. Consider some game $\gamma(\mathcal{P})$ which is in equilibrium and has P^* as maximum entropy attractor. For the purpose of an indirect proof, assume that $H(P^*) < H_{max}(\mathcal{P})$.

Consider the code κ^* adapted to P^* and the Dirichlet series ‡

$$Z(x) = \sum_{n\in\mathbb{N}} \exp(-x\kappa_n^*)\,. \tag{14}$$

Denote by Q_x the distribution with point masses $\exp(-x\kappa_n^*)/Z(x)$; $n \in \mathbb{N}$ whenever $Z(x) < \infty$.

It is easy to see that as P^* is power-dominated, the abscissa of convergence of the series (14) is less than 1. Note that $H(P^*) = \langle\kappa^*, Q_1\rangle$. Then, as, by assumption, $H(P^*) < H_{max}(\mathcal{P})$, $\langle\kappa^*, Q_x\rangle$ will be less than $H_{max}(\mathcal{P})$ for x's less than, but sufficiently close to 1. Therefore, we can choose $\beta < 1$ such that $H(P^*) < h < H_{max}(\mathcal{P})$ with $h = \langle\kappa^*, Q_\beta$.

‡the results we shall use about Dirichlet series are summarized in the appendix of [11].

Consider the preparation

$$\mathcal{P}_h = \{P \in M_+^1(\mathbb{N}) | \langle \kappa^*, P \rangle = h\} . \tag{15}$$

By construction, $\mathcal{P}_h \neq \emptyset$. We shall show that $\mathcal{P}_h$ is in equilibrium with P^* as attractor. To this end, consider an asymptotically optimal sequence $(P_k)_{k \geq 1}$ such that $H(P_k) > h$ for all $k \geq 1$. Then, as $\langle \kappa^*, P^* \rangle$ is less than and $\langle \kappa^*, P_k \rangle$ is larger than h, we can, for each $k \geq 1$, find a convex combination, say $Q_k = \alpha_k P^* + \beta_k P_k$, such that $Q_k \in \mathcal{P}_h$. By convexity of $D(\cdot \| P^*)$ and as $D(P_k \| P^*) \to 0$, we conclude from the linking identity (7) applied to $\langle \kappa^*, Q_k \rangle$ that $\lim_{k \to \infty} H(Q_k) = h$, hence $H_{max}(\mathcal{P}_h) \geq h$. On the other hand, by the definition of $\mathcal{P}_h$, $R(\kappa^* | \mathcal{P}_h) = h$. It follows that $\gamma(\mathcal{P}_h)$ is in equilibrium with κ^* as optimal code, hence with P^* as attractor. However, by robustness, part (ii) of Theorem 3.2, also Q_β is attractor for this preparation. As $Q_\beta \neq P^*$, we have arrived at a contradiction and conclude that there cannot be entropy loss for any preparation with P^* as attractor.

Now assume that P^* is hyperbolic and $H(P^*) < \infty$. Fix a finite $h > H(P^*)$ and consider $\mathcal{P}_h$ defined by (15). Then $\mathcal{P}_h \neq \emptyset$ (consider suitable distributions with finite support). By reasoning as in the first part of the proof, P^* is the attractor of $\mathcal{P}_h$ and $H_{max}(\mathcal{P}_h) = h$. Thus, the entropy function collapses at P^*. □

As examples of hyperbolic distributions with finite entropy, we mention those with point probabilities proportional to $n^{-1}(\ln n)^{-K}$ (for $n > 2$) with $K > 2$.

5 Discussion

Our technical results are, basically, contained in [11]. The interest here is the overall philosophy, cf. Section 1, the arrangement of the proofs, the new proof of Theorem 3.2 via standard minimax results and then the proof of Theorem 4.1 where some inaccuracies in [11] are corrected (in Section 8 of [11] the statements regarding $\mathcal{P}_h$ when $h = H_{max}(\mathcal{P})$ are erroneous). Technically and philosophically, there are good reasons for the focus on codes, somewhat at the expense of distributions. As an indication, note that for the setting of Section 3, the exponential family related to models $\mathcal{P} = \{P | \langle E, P \rangle = constant\}$ ("E" for *energy*), is better understood, not as a family of distrivbutions, but rather as the corresponding family of robust codes.

In [25] and [11] some further results are found, in particular of a topological nature.

The author expects that the mixed game- and information theoretical ideas will be integrated in central parts of probability theory and statistics, perhaps also in other areas. The references (and the homepages of Peter Harremoës and the author) serve to document the trend.

Key emphasis should be attached to Theorem 4.1. It demonstrates a potential for "generation" of entropy, almost contradicting the law of energy preservation. In fact, if a phenomenon is governed by a hyperbolic distribution P^*, this requires only finite "energy" , $H(P^*) = \langle \kappa^*, P^* \rangle$, but does lead to preparations (15) which operate at as high an "energy level" H_{max} we wish. In [11] this was applied to *Zipf's law* and used to explain that a stable, yet flexible language is possible with a potential for unlimited expressive power. Note that, typically, phenomena modelled by hyperbolic or other heavy-tailed distributions all seem to require high energies for their emergence (creation of a language, of the internet, of large economies, of the universe (!) etc.). Regarding the difficulties in handling these phenomena statistically, see [5]. One could even speculate that, in some precise sense, it may be impossible to handle statistically data generated by a hyperbolic distribution.

References

[1] A. Barron, J. Rissanen, and B. Yu. The minimum description length principle in coding and modeling. *IEEE Trans. Inform. Theory*, 44:2743–2760, 1998.

[2] N. N. Čencov. *Statistical decision rules and optimal inference.* Nauka, Moscow, 1972. In russian, translation in "Translations of Mathematical Monographs", 53. American Mathematical Society, 1982.

[3] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.*, 3:146–158, 1975.

[4] I. Csiszár and F. Matús. Information projections revisited. *IEEE Trans. Inform. Theory*, 49:1474–1490, 2003.

[5] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events.* Springer, Berlin, Heidelberg, 1997.

[6] K. Fan. Minimax theorems. *Proc. Nat. Acad. Sci. U.S.A.*, 39:42–47, 1953.

[7] K. Fan. A Minimax Inequality and Applications. In *Inequalities III, Proc. 3rd. Symp., Los Angeles*, pages 103–113, 1972.

[8] J.B.G. Frenk, G. Kassay, and J. Kolumbán. On equivalent results in minimax theory. *European J. Oper. Res.*, 157:46–58, 2004.

[9] P. D. Grünwald and A. P. Dawid. Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Ann. Stat.*, 32:1367–1433, 2004.

[10] P. Harremoës and F. Topsøe. Unified approach to optimization techniques in Shannon theory. In *Proceedings, 2002 IEEE International Symposium on Information Theory*, page 238. IEEE, 2002.

[11] P. Harremoës and F. Topsøe. Maximum Entropy Fundamentals. *Entropy*, 3:191–226, Sept. 2001. http://www.unibas.ch/mdpi/entropy/ [ONLINE].

[12] D. Haussler. A general minimax result for relative entropy. *IEEE Trans. Inform. Theory*, 43:1276–1280, 1997.

[13] R. S. Ingarden and K. Urbanik. Quantum informational thermodynamics. *Acta Physica Polonica*, 21:281–304, 1962.

[14] E. T. Jaynes. Information theory and statistical mechanics, I and II. *Physical Reviews*, 106 and 108:620–630 and 171–190, 1957.

[15] E. T. Jaynes. *Probability Theory - The Logic of Science.* Cambridge University Press, Cambridge, 2003.

[16] A. J. Jones. *Game theory. Mathematical models of conflict.* Horwood Publishing, Chichester, 2000.

[17] J. N. Kapur. *Maximum Entropy Models in Science and Engineering.* Wiley, New York, 1993. first edition 1989.

[18] D. Kazakos. Robust noiceless source coding through a game theoretic approach. *IEEE Trans. Inform. Theory*, 29:577–583, 1983.

[19] H. Kneser. Sur un théorème fondamental de la théorie des jeux. *Comptes Rendues de l'Acadéemie des Sciences, Paris*, 234:2418–2420, 1952.

[20] S. Kullback. *Informaton Theory and Statistics.* Wiley, New York, 1959.

[21] S. Kullback and R. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22:79–86, 1951.

[22] E. Pfaffelhuber. Minimax information gain and minimum discrimination principle. In I. Csiszár and P. Elias, editors, *Topics in Information Theory*, volume 16 of *Colloquia Mathematica Societatis János Bolyai*, pages 493–519. János Bolyai Mathematical Society and North-Holland, 1977.

[23] C. E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423 and 623–656, 1948.

[24] J. Szép and F. Forgó. *Introduction to the Theory of Games.* Akadémiai Kiadó, Budapest, 1985.

[25] F. Topsøe. Information Theoretical Optimization Techniques. *Kybernetika*, 15:8–27, 1979.

[26] F. Topsøe. Two General Games of Information. In *Proceedings of the International Symposium on Information Theory and its Applications, Parma*, pages 1410–1415, 2004. [ONLINE: http://www.math.ku.dk/ topsoe].

[27] J. von Neumann. Zur Theorie der Gesellschaftsspiele. *Math. Ann.*, 100:295–320, 1928.

Flemming Topsøe: University of Copenhagen, Department of Mathematics, Universitetsparken 5, Copenhagen, 2100, Denmark, topsoe@math.ku.dk