Interaction between Truth, Belief and Knowledge as the key to entropy measures of statistical physics

Flemming Topsøe
University of Copenhagen
Department of Mathematical Sciences
Presentation at MECO34, Leipzig 29.03-02.04, 2009
What is entropy?

Entropy operates in the interface
God ↔ man or Nature ↔ man or
system ↔ physicist studying or actually observing.

Entropy assigns to the true state \( x \) of a system the complexity of the system in that state as seen by man. More precisely, the entropy of \( x \) is given by

\[
S(x) = \min_y \Phi(x, y)
\]

where

\[
\Phi(x, y) = \text{description cost (or effort) needed to describe the system when the truth is } x \text{ and your belief is } y.
\]

So: **Entropy is minimal description effort**

Note: Focus on \( \Phi \) with \( S \) as a derived concept. Another important derived concept is redundancy (divergence) \( D \) defined by \( D(x, y) = \Phi(x, y) - S(x) \).
**Problem:** $\Phi = ?$

Answer requires knowledge about nature, the world we operate in, and input from man who should design efficient description strategies via *descriptors* (see later).

Regarding $z$: outcome from extended observations, synthesis of extended experience, knowledge.

$x$ and $y$ will be distributions over some *alphabet*, $\mathbb{A}$: $x = (x_i)_{i \in \mathbb{A}}, y = (y_i)_{i \in \mathbb{A}}$ and similarly for $y$. 
Possible worlds

Key postulate: There is an interaction between $x, y, z$ of the form $z = \Pi(x, y)$. $\Pi$ characterizes the world.

Some possibilities:

$\Pi(x, y) = x$: classical world,
$\Pi(x, y) = y$: black hole,
$\Pi(x, y) = qx + (1 - q)y$: mixtures.

Assume that interaction acts locally via interactor $\pi$: $z = \Pi(x, y)$ with $z_i = \pi(x_i, y_i)$. Always assume that $\pi(s, t) = s$ if $t = s$ (soundness).

Examples: $\pi_q(s, t) = qs + (1 - q)t$ corresponding to classical world ($q = 1$), black hole ($q = 0$) or mixtures.

There are many other possible interactors, but:

under consistency the $\pi_q$’s are the only ones.

Consistency:
$\sum z_i = 1$, for $x, y$ probability distributions, $z = \Pi(x, y)$. 
What can the physicist do?

We assume that the world is known to the physicist through the interactor $\pi$.

Every observation entails a cost (or effort). The cost depends on the event being observed. An event with high probability has little cost. Thus define a descriptor $\kappa$ to be a decreasing function on $[0, 1]$ with $\kappa(1) = 0$. Also insist that $\kappa$ satisfies the normalization condition $\kappa'(1) = -1$, corresponding to a choice of unit.

Interpretation: $\kappa(t)$ is the effort associated with observations from an event which you believe occurs with probability $t$ (equal to some $y_i$, say).

Total description cost, $\Phi(x, y)$, when truth is $x$ and belief is $y$ is given by

$$
\Phi(x, y) = \sum_{i \in \mathbb{A}} z_i \kappa(y_i) = \sum_{i \in \mathbb{A}} \pi(x_i, y_i) \kappa(y_i).
$$
The perfect match principle (PMP)

To design $\kappa$, hence $\Phi$, apply perfect match principle:

| PMP: $\Phi$ is the smallest when belief matches truth: $\Phi(x, y) \geq \Phi(x, x)$, i.e. $\Phi(x, y) \geq S(x)$. |

**Theorem** Given $\pi$, the only possible descriptor satisfying PMP is the solution to

$$\frac{\partial \pi}{\partial t}(t, t)\kappa(t) + t\kappa'(t) = -1; \kappa(1) = 0.$$

If $\pi = \pi_q$, this becomes

$$(1 - q)\kappa(t) + t\kappa'(t) = -1; \kappa(1) = 0$$

with solution $\kappa = \kappa_q$ given by

$$\kappa_q(t) = \ln_q \frac{1}{t}$$

with $q$-logarithm given by

$$\ln_q x = \begin{cases} \ln x & \text{if } q = 1, \\ \frac{1}{1-q}(x^{1-q} - 1) & \text{if } q \neq 1. \end{cases}$$

However, this only satisfies PMP if $q \geq 0$. 
Comments on the proof

The differential equation comes up via standard variational principles (introduce Lagrange multipliers!).

Re $\pi_q, \kappa_q$: Failure of PMP for $q < 0$: simple direct counter examples with a 3-element alphabet.

Validity of PMP for $q \geq 0$: via PFI, the pointwise fundamental inequality, which states that $d(s, t) \geq 0$ where

$$d(s, t) = (\pi(s, t)\kappa(t) + t) - (s\kappa(s) + s).$$

Indeed,

$$\mathcal{D}(x, y) = \sum d(x_i, y_i)$$

and PMP really says that $\mathcal{D}(x, y) \geq 0$ (with “$=$” iff $y = x$). Writing up the formulas with $\pi_q$ and $\kappa_q$ in place of $\pi$ and $\kappa$ one finds that a simple application of the geometric/arithmetic mean inequality leads to the desired inequality $d(s, t) \geq 0$. 
Comments on result:

- Strong argument in favour of the view that only possible entropy measures of statistical physics are those in the Tsallis family: $S_q(x) = \sum x_i \kappa_q(x_i)$

\[
S_q(x) = \left( \sum x_i^q - 1 \right)/(1 - q)
\]

for $q \geq 0$. Classical value for $q = 1$, and for black hole, $S(x) = n - 1$, number of degrees of freedom ($n = \text{size of alphabet}$).

- $\Phi$ more important than $S$, gives more, e.g. shows which are the feasible preparations, viz. those of the form $\mathcal{P} = \{x | \Phi(x, y_0) = c\}$, and $\Phi$ also assists greatly in finding equilibrium distributions (via MaxEnt, and even without introducing Lagrange multipliers – instead a natural, therefore better, intrinsic approach is used). (more on next slides)

- The roles of God and man (nature and man, system and man) clearly separated!

- Formulas are mathematically attractive, e.g. lines up with popular Bregman divergencies ... there “must be some truth in them”!

- main outstanding issues are: interaction, how? description, how – via coding as in classical case (à la Shannon)?
Equilibrium calculations

Background theorem  \( \mathcal{P} \) any preparation (set of \( x \)'s). Given \( x^*, y^* \) such that: \( x^* \in \mathcal{P}, y^* \) robust, i.e. \( \exists h \forall x \in \mathcal{P} : \Phi(x, y^*) = h \), and \( y^* = x^* \). Then \( x^* \) is the MaxEnt distribution of \( \mathcal{P} \) (and \( y^* \) the MinRisk strategy for the physicist, i.e. \( \arg\min \left( \max_{x \in \mathcal{P}} \Phi(x, y) \right) = y^* \)).

Proof: Assume \( \Phi(x, y^*) = h \) for all \( x \in \mathcal{P} \). Then \( S(x^*) = \Phi(x^*, x^*) = \Phi(x^*, y^*) = h \), and, for \( x \in \mathcal{P} \) and \( x \neq x^* \) we have \( S(x) < S(x) + D(x, y^*) = h \). The min-risk part is proved just as easily. \( \square \)

Define: \( y^* \in \mathcal{E}(y) \), the exponential family of \( y \), if \( \Phi(x, y^*) \) only depends on \( \Phi(x, y) \). In short:
\[
\mathcal{E}(y) = \{ y^* | \exists \xi : \Phi(x, y^*) = \xi(\Phi(x, y)) \}.
\]

Corollary Put \( L_y(h) = \{ x | \Phi(x, y) = h \} \). If \( \mathcal{P} = L_y(h) \) and \( x^*, y^* \) are given such that: \( x^* \in \mathcal{P}, y^* \in \mathcal{E}(y) \) and \( y^* = x^* \), then \( x^* \) is the MaxEnt distribution of \( \mathcal{P} \).
Let $L^y$ be the class of non-empty models of the form $L^y(h)$ with $h \in \mathbb{R}$. The associated exponential family or equilibrium generating family, is the family

$$
\mathcal{E}(y) = \{x | \forall L \in L^y \exists c \in \mathbb{R} : L \subseteq L^x(c)\}.
$$

$x \in \mathcal{E}(y), \ S(x) = h \Rightarrow x$ is MaxEnt dist. of $L^x(h)$

Thus one should try and determine the elements in $\mathcal{E}(y)$. Looking at it, you find that for the worlds determined by one of the interactions $\pi_q$, every $x$ for which there exist constants $\alpha$ and $\beta$ such that

$$
\forall i \in A : \kappa(x_i) = \alpha + \beta \kappa(y_i)
$$

are in $\mathcal{E}(y)$. For $q = 1$ this leads to classical analysis (with partition function etc.) and even without the use of Lagrange multipliers. For the general case, you have to adjust constants so that

$$
\sum_{i \in A} \kappa^{-1}(\alpha + \beta \kappa(y_i)) = 1.
$$
Figure shows family \((\kappa_q)_{q \geq 0}\) of descriptors

So, given a probability \(t\), you can see what effort is needed, measured in nats (natural units), in order to describe events with probability \(t\) when you use the chosen descriptor.
Figure shows inverses of descriptors.

“probability checkers”

You can use these functions (**q-deformed exponentials**) to check, for a chosen descriptor, how complicated events you can describe with a given number of nat’s available, i.e. how low a probability an event can have and still be describable with the available number of nat’s. This kind of consideration is important in order to carry out MaxEnt calculations indicated previously.