From Truth, Belief and Knowledge to Tsallis Entropy

Flemming Topsøe, topsoe@math.ku.dk
Department of Mathematical Sciences, University of Copenhagen
indications: key notions, key questions

truth
belief, tendency to act
knowledge, perception, extended experience
interaction
experiment, preparation, control
description, effort, information

what is “information”? – a potential saving of effort!
what then is “entropy”? – minimal effort, given the truth!
and, what c a n we know? – well, depends on your belief …
information: cost and associated effort

What is the cost of information or, how much are you willing to pay – or have to pay – in order to know that an event has happened?

Or, what is the effort you are willing to/have to allocate?

Depends on the probability $t$, you believe the event has: $\kappa(t)$. $\kappa$ tells us the individual effort. It is the effort-function or the descriptor.

effort $\longleftrightarrow$ description?

Requirements: $\kappa(1) = 0$, $\kappa$ is smooth (and decreasing). Further, natural with normalization via the differential cost $\iota = -\kappa'(1)$. If $\iota = 1$, we obtain natural units, nats; if $\iota = \ln 2$, we measure in binary units, bits.
the power hierarchy, the exponential hierarchy

Which kind of descriptors would you expect?
Note that any “reasonable” monotone function \( f \) defines a descriptor via linearization. Simply take

\[
\kappa(t) = \frac{f(1) - f(t)}{f'(1)}.
\]

Suggestions: The power hierarchy is defined from the functions \( t \mapsto t^a \) \((-\infty < a < \infty)\) and gives the descriptors:

\[
t \mapsto \frac{1 - t^a}{a}.
\]

... and the exponential hierarchy is defined from the functions \( t \mapsto b^t \) \((b > 0)\) and gives the descriptors

\[
t \mapsto \frac{1 - b^{t-1}}{\ln b}.
\]

But are any of these “sensible”? – and what does that mean?
accumulated effort in a probabilistic context

Consider distributions over a discrete alphabet $\mathbb{A}$:

$x = (x_i)_{i \in \mathbb{A}}$ represents truth, $y = (y_i)_{i \in \mathbb{A}}$ represents belief.

Accumulated effort (expected per observation) is

$$\Phi(x, y) = \sum_{i \in \mathbb{A}} x_i \kappa(y_i).$$

$\Phi$ satisfies the perfect match principle (for short is proper) if

$$\Phi(x, y) \geq \Phi(x, x)$$

with equality only for $y = x$ (or $\Phi(x, x) = \infty$).

**Theorem** There is only one descriptor $\kappa$, the classical descriptor, for which $\Phi$ above is proper, viz. (nats)

$$\kappa(t) = \ln \frac{1}{t}.$$
questioning the basic definition

Surely, $\Phi(x, y) = \sum x_i \kappa(y_i)$ is the right expression for accumulated effort as seen by someone, who knows the truth as well as the belief ...  

... but is this how you perceive accumulated effort?

What if the $x_i$’s above are not what you perceive as truth?

... perhaps this also depends on what you believe – and $\Phi$ should rather be something like $\sum \pi(x_i, y_i) \kappa(y_i)$.

Let’s leave the probabilistic setting for a while and go philosophical:
the beginnings of a philosophy of information

The whole is the world, $\mathcal{V}$

Situations from the world involve Nature and you, Observer.

Nature has no mind but holds the truth ($x$),

Observer has a creative mind,

- seeks the truth ($x$)
- is confined to belief ($y$)
- aims at knowledge ($z$).

But what is “knowledge”? Knowledge is

- the synthesis of extensive experience
- an expression of how Observer perceives situations from $\mathcal{V}$
- a manifestation of truth for Observer, for you.
interaction and effort

**Proposal**: Knowledge \((z)\) depends on truth and belief via a characteristic interactor \(\Pi: [z = \Pi(x, y)] \quad \mathcal{V} = \mathcal{V}_\Pi\).

\(\Pi_1 : (x, y) \mapsto x\) defines the classical world \(\mathcal{V}_1\)
\(\Pi_0 : (x, y) \mapsto y\) defines a black hole \(\mathcal{V}_0\)
\(\Pi_q : (x, y) \mapsto qx + (1 - q)y\) defines mixtures, Tsallis’ \(\mathcal{V}_q\)’s

Associated with \(\mathcal{V}_\Pi\) are, possibly many, effort functions, \(\Phi\)’s. Extending the previous definition, an effort function is proper if it satisfies the perfect match principle (PMP):
\(\Phi(x, y) \geq \Phi(x, x)\) with equality iff there is a perfect match, i.e. \(y = x\) (or \(\Phi(x, x) = \infty\)).

**Thesis** Given \(\mathcal{V}_\Pi\), there is at most one proper \(\Phi\)-function
entropy, divergence, the fundamental inequality

Abstract modelling involves effort ($\Phi$), entropy ($H$), and divergence ($D$). $\Phi$ is assumed proper.

Entropy is defined as minimal effort, given the truth, divergence as excess (or redundant) effort:

$$H(x) = \Phi(x, x); \quad D(x, y) = \Phi(x, y) - H(x).$$

The properness of $\Phi$ may be expressed in terms of $D$ by the fundamental inequality of information theory (FI):

$$D(x, y) \geq 0 \quad \text{with equality iff } y = x.$$

Further notions and properties are best discussed for probabilistic modelling.
probabilistic modelling (discrete)

Truth-, belief- and knowledge instances are $x = (x_i)$, $y = (y_i)$ and $z = (z_i)$ ($i$ ranging over an alphabet $\mathbb{A}$). $x$ and $y$ are probability distributions, $z$ just a function on $\mathbb{A}$.

Interaction, $\Pi$, acts via the local interactor $\pi$:

$$(\Pi(x, y))_{i \in \mathbb{A}} = (\pi(x_i, y_i))_{i \in \mathbb{A}}.$$ $\pi$ is always assumed sound, i.e. $\pi(s, t) = s$ if $t = s$ (perfect match).

$\pi$ is weakly consistent if $\forall x \forall y : \sum z_i = 1$. Strong consistency requires that $z$ is always a probability distribution.

**Proposition:** Only the $\pi_q$’s given by $\pi_q(s, t) = qs + (1 - q)t$ are weakly consistent; strong consistency requires $0 \leq q \leq 1$. 
accumulated effort, the one and only
Accumulated effort always chosen among $\Phi_{\pi, \kappa}$ where $\kappa$ is a descriptor and

$$\Phi_{\pi, \kappa}(x, y) = \sum_{i \in A} \pi(x_i, y_i) \kappa(y_i).$$

**Theorem** (modulo regularity conditions). Given $\pi = \pi(s, t)$, let $\pi'_2 = \frac{\partial \pi}{\partial t}$ and put $\chi(t) = \pi'_2(t, t)$. Only one among the $\Phi_{\pi, \kappa}$'s can be proper, viz. the solution to

$$t\kappa'(t) + \chi(t)\kappa(t) = -1, \quad \kappa(1) = 0. \quad (*)$$

If $\pi$ is consistent, hence one of the $\pi_q$'s, then a proper $\Phi_{\pi, \kappa}$ exists iff $q > 0$ ($q = 0$ OK as a singular case, though).
If so, the unique descriptor concerned is the one depending linearly on $t^{q-1}$, i.e. $\boxed{\kappa_q(t) = \ln_q \frac{1}{t}}$ (recall: $\ln_q u = \frac{1}{1-q}(u^{1-q} - 1)$).
gross effort, pointwise fundamental inequality

Introduce gross (accumulated) effort and gross entropy by adding a term representing overhead cost (or effort):

\[
\tilde{\Phi}(x, y) = \sum_{i \in \mathcal{A}} \left( \pi(x_i, y_i) \kappa(y_i) + y_i \right) = \Phi(x, y) + 1,
\]

\[
\tilde{H}(x) = \sum_{i \in \mathcal{A}} \left( x_i \kappa(x_i) + x_i \right) = H(x) + 1.
\]

Clearly, “gross divergencee” = divergence and, defining the divergence generator by

\[
\delta(s, t) = (\pi(s, t) \kappa(t) + t) - (s \kappa(s) + s),
\]

one has

\[
D(x, y) = \sum \delta(x_i, y_i).
\]

We refer to the inequality \( \delta \geq 0 \) as the pointwise fundamental inequality (PFI). Clearly

**Conjecture** Converse also true

In practice, PMP and Fl are always proved via PFI!
given $\kappa$, which world are you in?

Given $\pi$, we insist, when possible, to choose $\kappa$ such that the resulting function $\Phi$ is proper. This gives a unique choice, the ideal descriptor. You determine $\kappa$ from $\pi$, but

**Warning:** you cannot determine $\pi$ from $\kappa$

Thus knowing the entropy function does not reveal the world.

Examples: Let $\pi = \pi_q (q > 0)$ and consider $\pi^\xi$ of the form

$$\pi^\xi(s, t) = \xi^{-1}(\pi(\xi(s), \xi(t))).$$

Then the differential equation (*) is unchanged, hence you find the same descriptor $\kappa_q$. E.g. for $\xi(u) = \ln u$, $\pi^\xi(s, t) = s^q t^{1-q}$; by PFI, the associated effort is proper.

**Problem** which $\kappa$’s are associated with (meaningful) $\pi$’s?

e.g. $\kappa(t) = \frac{1}{2}(t^{-2} - 1)$; or $\kappa(t) = 1 - \exp(t - 1)$?
what can we know? (abstract modelling)

Setting: World $\mathcal{Y}_\pi$ with ideal descriptor and effort fct. $\Phi$.
I.J. Good (1952): **Belief is a tendency to act!**

To us, this is expressed via controls, $w$’s. There is a bijection $y \leftrightarrow w$ ( $w = \hat{y}; y = \hat{w}$). In our probabilistic modelling this is given by $w_i = \kappa(y_i); i \in \mathbb{A}$.
Expressed via controls, the effort function is denoted $\Psi$:
$\Psi(x, w) = \Phi(x, y)$ with $y \leftrightarrow w$.

What can Observer do? via control! preparations which are sets of $x$’s, typically denoted by $\mathcal{P}$.
A feasible preparation is one which Observer can realize.
more on preparations

Typical example (of genus 1): Fix a control $w$ and a level $h$. Set-up an experiment (!?) which constrains Nature's possibilities to the preparation

$$\mathcal{P}(w, h) = \{x | \Psi(x, w) = h\}$$

or to the variant $\mathcal{P}_{\leq}(w, h) = \{x | \Psi(x, w) \leq h\}.$

Finite non-empty intersections of such level sets (or sub-level sets) constitute the feasible preparations and show what Observer can know!
equilibrium, MaxEnt and all that!

A closer study of a fixed preparation $\mathcal{P}$ requires game theory and exploits thinking of John Nash. We shall only outline this. The two players are Nature with truth instances $x \in \mathcal{P}$ as strategies and Observer with controls $w$ as strategies. As objective function we take $\Psi = \Psi(x, w)$. The maxmin value is easily seen to be the MaxEnt value

$$H_{\text{max}}(\mathcal{P}) = \sup_{x \in \mathcal{P}} H(x).$$

If this equals the minmax value (required finite), the game is in equilibrium.

Another notion, often overlooked: A control $\varepsilon^*$ is robust if, for some finite $h$, $\Psi(x, \varepsilon^*) = h$ for all $x \in \mathcal{P}$. Then $h$ is the level of robustness.

By results of Nash:
robustness lemma, exponential families

**Robustness lemma** If $x^* \in \mathcal{P}$ and $\varepsilon^* = \hat{x}^*$ is robust with level $h$, then the game is in equilibrium with $x^*$ and $\varepsilon^*$ as optimal strategies, in particular, $x^*$ is the MaxEnt strategy. (Furthermore, the celebrated Pythagorean inequalities hold).

Let $w$ be a control, let $\mathcal{L}^w$ be the preparation family of non-empty sets of the form $\mathcal{P}(w, h)$. The associated exponential family, denoted $\hat{\mathcal{E}}^w$ is the set of controls $\varepsilon$ which are robust for all preparations in $\mathcal{L}^w$. From robustness lemma:

Consider a preparation family $\mathcal{L}^w$. Let $x^*$ be a truth instance, put $\varepsilon^* = \hat{x}^*$ and assume that $\varepsilon^* \in \hat{\mathcal{E}}^w$. Put $h = \Psi(x^*, w)$. Then the game corresponding to $\mathcal{P}(w, h)$ is in equilibrium and has $x^*$ and $\varepsilon^*$ as optimal strategies. In particular, $x^*$ is the MaxEnt distribution for $\mathcal{P}(w, h)$. 
sketch of MaxEnt calculations in $\mathcal{V}_q$

Return to probabilistic setting and consider a Tsallis world $\mathcal{V} = \mathcal{V}_q$, cor. to $\pi_q$ with $q > 0$.
Fix $y \longleftrightarrow w$. Then $\mathcal{L}^w$ consists of all preparations $\mathcal{P}$ for which $\Psi(x, w)$ is constant over $\mathcal{P}$.
But $\Psi(x, w) = \sum (q x_i + (1 - q) y_i) w_i$ so condition is equivalent to $\sum x_i w_i$ being constant over $\mathcal{P}$.
For fixed constants $\alpha$ and $\beta$ this implies that $\sum x_i (\alpha + \beta w_i)$ is constant over $\mathcal{P}$.
Now, if $\alpha + \beta w$ is a control, say $w^*$, $\sum x_i w_i^*$ is constant over $\mathcal{P}$, hence $\Psi(x, w^*)$ is constant over $\mathcal{P}$, i.e. $w^* \in \hat{\mathcal{E}}^w$ and the robustness lemma applies.
Then, given $\beta$, try to adjust $\alpha$ so that $\alpha + \beta w$ is a control.
Classically, $\alpha$ is the logarithm of the partition function.
Finally, adjust $\beta$ ($\approx$ inverse temperature) to desired level ...
what have we achieved?

- found a reasonably transparent interpretation of Tsallis entropy
- developed a basis for an abstract theory
- clarified role of FI via PMP; focus on PFI as the natural basis for establishing FI and hence PMP
- identified the unit of entropy as an overhead
- answered the question “what can we know”
- found good (the right ?) definition of an exponential family
- indicated dual role of preparations and exponential families
- exploited games and wisdom of Nash, enabled MaxEnt calculations without introducing Lagrange multipliers
- separated Nature from Observer in key expressions
what needs being done?

- interaction, how?
- description, how?
- control, how?
- expand, quantum setting ...
- link to information geometry
- ...

thank you!