Games, Entropy and Composability

Henrik Densing Petersen & Flemming Topsøe
University of Copenhagen
Department of Mathematical Sciences
Presentation for CTnext07, Catania, July 2007

Goal

Operational definitions of entropy and related quantities covering the classical as well as non-extensive settings, thereby understanding which entropy measures are relevant for physics

Announcement: Workshop: "Facets of Entropy", Copenhagen, October 24-26, 2007 (if interested, ask FT or Robert Niven).

Overview

Entropy *without* **games:**

- **Overall setting**
- **Listing some properties**
- **More specifics on structure**
- **Results for "f-entropies", especially on** **composability**

Entropy *with* **games:**

- **Complexity**
- **Defining entropy** **two types of entropy!**
- **Defining divergence**
- **MaxEnt via robustness** $\cdots$
- **Not primary focus on entropy -** **complexity is what matters!**

Conclusions:

- **Which entropy ? – e.g. Tsallis or Rényi?**

But: **does the question make sense?**

Entropy *without* **games**

Overall setting: probabilities on discrete spaces!

Properties to consider include:

- minimal (0) on δ_i's (deterministic), max on uniform

- continuous (lower semi-cont. in infinite case)

- concave: H(mixture) $\geq$ mixture of H

- datareduction inequality: H(coarse)$\leq$ H(fine)

- MaxEnt-principle should make sense for "natural" preparations (models) – and the nature of the entropy function should facilitate MaxEnt-calculations

- consistency: no feasible state is ignored under inference when you use the MaxEnt principle

- composable: $\mathrm{H}(P \otimes Q) = g(\mathrm{H}(P), \mathrm{H}(Q))$

- – or even additive): $\mathrm{H}(P \otimes Q) = \mathrm{H}(P) + \mathrm{H}(Q)$

- acceptable, physically significant interpretation!

Key results on f-entropies

"f-entropy": Based on generator f which is assumed to be nice convex and satisfy

$$f(0) = f(1) = 0,\ f'(1) = 1\,.$$

$$\mathsf{H}_f(P) = -\sum f(p_i) \text{ or } \mathsf{H}_f(P) = -\sum p_i \widetilde{f}(\frac{1}{p_i})$$

– with $\widetilde{f}$ the Csiszár dual of f: $\widetilde{f}(x) = x f(\frac{1}{x})$.
BGS (classical): $f(x) = x\ln(x)$, $\widetilde{f}(x) = \ln(\frac{1}{x})$,
Tsallis family: (via "deformed logarithms"):

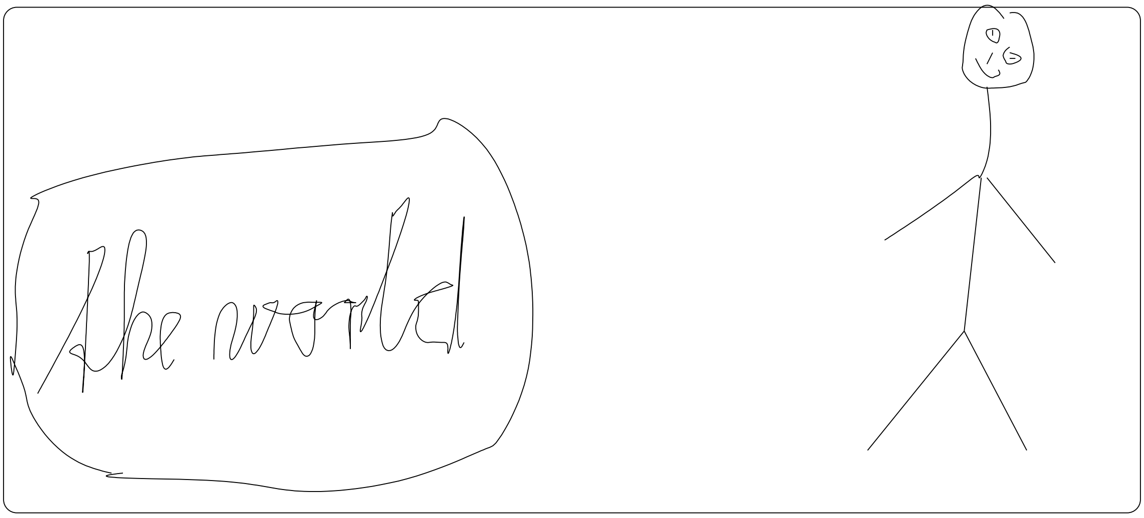
$$\mathsf{H}_q^T(P) = \frac{1}{1-q}\Big(\sum p_i^q - 1\Big)\,.$$

Well-known: H_f continuous, concave, satisfies datareduction principle – and MaxEnt? Wait!

Key result: Among f-entropies, only Tsallis entropies are composable. For these:

$$\mathsf{H}_q^T(P\otimes Q) = \mathsf{H}_q^T(P) + \mathsf{H}_q^T(Q) + (1-q)\,\mathsf{H}_q^T(P)\cdot\mathsf{H}_q^T(Q)\,.$$

Entropy *with* games



Natures side: P Observers side (you!): Q
connected by complexity function $\Phi = \Phi(P,Q)$.

Assumptions Minimal on diagonal: $\Phi(P,Q) \geq \Phi(P,P)$.
Vanishes on deterministic dist.: $\Phi(\delta_i,\delta_i) = 0$.
Examples:

$\Phi^{BGS} = \sum p_i \ln \frac{1}{q_i}$: BGS

$\Phi_q^R = \frac{1}{1-q} \ln \frac{\sum p_i^q}{\sum p_i^q q_i^{1-q}}$: Rényi

$\Phi_q^T = \frac{1}{1-q}\left(\frac{\sum p_i^q}{\sum p_i^q q_i^{1-q}} - 1\right)$: Tsallis

... continued: Entropy, Divergence, MaxEnt

Entropy = minimal complexity: $\mathsf{H}(P) = \min_Q \Phi(P, Q)$.
Divergence $=$ actual $-$ minimal complexity:
$\mathsf{D}(P, Q) = \Phi(P, Q) - \mathsf{H}(P)$ $(= \Phi(P, Q) - \Phi(P, P))$.

Dual entropy anticipates unknown but deterministic distribution: $\widehat{\mathsf{H}}(Q) = \sum q_i \Phi(\delta_i, Q) = \sum q_i \, \mathsf{D}(\delta_i, Q)$.

MaxEnt-problem: given a preparation $\mathcal{P}$, to determine the MaxEnt-distribution and the corresponding MaxEnt-value: $\mathsf{H}_{\max} = \mathsf{H}_{\max}(\mathcal{P}) = \max_{P \in \mathcal{P}} \mathsf{H}(P)$.

A *highly useful, trivial,* but *neglected criterion*:

If $Q \in \mathcal{P}$ is robust: $\Phi(P, Q)$ independent of $P \in \mathcal{P}$, say $\forall P \in \mathcal{P} : \Phi(P, Q) = h$, then Q is the MaxEnt-distribution and $\mathsf{H}_{\max}(\mathcal{P}) = h$.

Proof. Firstly: $\mathsf{H}(Q) = \Phi(Q, Q) = h$.
Secondly: if $P \neq Q$ and $P \in \mathcal{P}$, then
$H(P) < H(P) + \mathsf{D}(P, Q) = \Phi(P, Q) = h$. □

The examples (only Φ and H)

name	complexity	function of
BGS	$\sum p_i \ln \frac{1}{q_i}$	$\langle \ln \frac{1}{Q}, P \rangle$
q-Rényi	$\frac{1}{1-q} \ln \frac{\sum p_i^q}{\sum p_i^q q_i^{1-q}}$	$\langle Q^{1-q}, P^{(q)} \rangle$
q-Tsallis^1	$\frac{1}{1-q} \left(\frac{\sum p_i^q}{\sum p_i^q q_i^{1-q}} - 1 \right)$	$\langle Q^{1-q}, P^{(q)} \rangle$
q-Tsallis^2	$\frac{1}{1-q} \sum p_i^q (1 - q_i^{1-q})$	$\langle 1 - Q^{1-q}, P^q \rangle$
q-Tsallis^3	$\sum \left(q_i^q - \frac{p_i(1-q q_i^{q-1})}{1-q} \right)$	$\sum q_i^q, \langle Q^{q-1}, P \rangle$

$P^{(q)}$: the q-escort distribution: $i \curvearrowright p_i^q / \sum p_i^q$.

P^q: the (non-normalized) measure $i \curvearrowright p_i^q$.

name	entropy	dual entropy
BGS	H^{BGS}	H^{BGS}
q-Rényi	H_q^R	H^{BGS}
q-Tsallis[1]	H_q^T	H_q^T
q-Tsallis[2]	H_q^T	H_{2-q}^T
q-Tsallis[3]	H_q^T	H_q^T

The entropies: BGS: $-\sum p_i \ln p_i$, Rényi: $\frac{1}{1-q} \ln \sum p_i^q$, Tsallis: $\frac{1}{1-q}(\sum p_i^q - 1)$.

Property	Rényi	Tsallis
consistent inf.	$q < 1$ only	$q < 1$ only
concave	$q < 1$, few other	all q
composable	all q	all q
additive	all q	no q
interpretation	hmmm	hmmm
experimental evidence	hmmm	hmmm

Φ-exponential families etc.

To simplify, assume structure as in Tsallis[3]
(the Bregman case):

$$\Phi(P,Q) = \text{fct. of } Q + \langle \widehat{Q}, P \rangle$$

$\widehat{Q}$: a certain transform of Q.
(Problem: Interpretation?)

For functions $\mathbf{f} = (f_1, \cdots, f_k)$, define the Φ-exponential family $\mathcal{E}$ as the set of distributions Q for which there exist constants λ_0 and $\lambda_1, \cdots, \lambda_k$ such that:

$$\widehat{Q} = \lambda_0 + (\lambda_1 f_1 + \cdots + \lambda_k f_k)$$

The natural preparations are those of the form

$$\mathcal{P}_{\mathbf{a}} = \{P | \langle f_1, P \rangle = a_1, \cdots, \langle f_k, P \rangle = a_k\} .$$

¿From robustness criterion we find immediately:

Theorem If $Q \in \mathcal{E} \cap \mathcal{P}_{\mathbf{a}}$, then Q is the Φ-MaxEnt distribution.

Conclusions

Recall the key technical result (due to HDP):

> Among f-entropies, only Tsallis entropies are composable.

This also covers entropies (like Rényi entropy) which are monotone functions of f-entropies.

Apart from the key result we conclude with some insights gained during the investigations:

(i) Never more use Lagrange multipliers!
– unless you deal with "ad hoc problem" or use these multipliers as a guide in preliminary investigations.

(ii) Be aware of the two types of entropies!

(iii) Never consider entropy measures alone! – you must supply with other considerations, at best:

> take as point of departure a suitable *complexity measure!*