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

(if interested, ask FT or Robert Niven).
Overview

Entropy \textit{without} games:

- Overall setting
- Listing some properties
- More specifics on structure
- Results for "\(f\)-entropies", especially on \textbf{composability}

Entropy \textit{with} games:

- Complexity
- Defining entropy \textbf{two types of entropy}!
- Defining divergence
- MaxEnt via robustness \ldots
- Not primary focus on entropy - \textbf{complexity is what matters!}

Conclusions:

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

\textit{But}: does the question make sense?
Entropic without games

Overall setting: probabilities on discrete spaces!

Properties to consider include:
• minimal (0) on $\delta_i$’s (deterministic), max on uniform
• continuous (lower semi-cont. in infinite case)
• concave: $H(\text{mixture}) \geq \text{mixture of } H$
• data-reduction inequality: $H(\text{coarse}) \leq H(\text{fine})$

• 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: $H(P \otimes Q) = g(H(P), H(Q))$
• – or even additive): $H(P \otimes Q) = H(P) + H(Q)$
• acceptable, physically significant interpretation!
Key results on $f$-entropies

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

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

$$H_f(P) = - \sum f(p_i) \text{ or } H_f(P) = - \sum p_i \tilde{f}\left(\frac{1}{p_i}\right)$$

– with $\tilde{f}$ the Csiszár dual of $f$: $\tilde{f}(x) = xf\left(\frac{1}{x}\right)$.

BGS (classical): $f(x) = x \ln(x), \quad \tilde{f}(x) = \ln\left(\frac{1}{x}\right)$.

Tsallis family: (via "deformed logarithms"):

$$H_T^q(P) = \frac{1}{1 - q} \left(\sum p_i^q - 1\right).$$

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

---

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

$$H_T^q(P \otimes Q) = H_T^q(P) + H_T^q(Q)$$

$$+ (1 - q) H_T^q(P) \cdot H_T^q(Q).$$
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:**

\[
\Phi^{BGS} = \sum p_i \ln \frac{1}{q_i} : \text{BGS}
\]

\[
\Phi^{R} = \frac{1}{1-q} \ln \frac{\sum p_i^q}{\sum p_i^q q_i^{1-q}} : \text{Rényi}
\]

\[
\Phi^{T} = \frac{1}{1-q} \left( \frac{\sum p_i^q}{\sum p_i^q q_i^{1-q}} - 1 \right) : \text{Tsallis}
\]
... continued: Entropy, Divergence, MaxEnt

Entropy = minimal complexity: \( H(P) = \min_Q \Phi(P, Q) \).

Divergence = actual − minimal complexity:
\[
D(P, Q) = \Phi(P, Q) - H(P) \quad (= \Phi(P, Q) - \Phi(P, P)).
\]

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

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

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 \( H_{\text{max}}(\mathcal{P}) = h \).

Proof. Firstly: \( H(Q) = \Phi(Q, Q) = h \).
Secondly: if \( P \neq Q \) and \( P \in \mathcal{P} \), then
\[
H(P) < H(P) + D(P, Q) = \Phi(P, Q) = h. \quad \Box
\]
The examples (only $\Phi$ and $\mathcal{H}$)

<table>
<thead>
<tr>
<th>name</th>
<th>complexity</th>
<th>function of</th>
</tr>
</thead>
<tbody>
<tr>
<td>BGS</td>
<td>$\sum p_i \ln \frac{1}{q_i}$</td>
<td>$\langle \ln \frac{1}{Q}, P \rangle$</td>
</tr>
<tr>
<td>$q$-Rényi</td>
<td>$\frac{1}{1-q} \ln \frac{1}{q_i} \frac{\sum p_i}{\sum p_i q_i}$</td>
<td>$\langle Q^{1-q}, P(q) \rangle$</td>
</tr>
<tr>
<td>$q$-Tsallis$^1$</td>
<td>$\frac{1}{1-q} \left( \frac{\sum p_i}{\sum p_i q_i} - 1 \right)$</td>
<td>$\langle Q^{1-q}, P(q) \rangle$</td>
</tr>
<tr>
<td>$q$-Tsallis$^2$</td>
<td>$\frac{1}{1-q} \sum p_i q_i (1 - q_i^{1-q})$</td>
<td>$1 - Q^{1-q}, P^q$</td>
</tr>
<tr>
<td>$q$-Tsallis$^3$</td>
<td>$\sum \left( q_i^q - \frac{p_i (1 - qq_i^{q-1})}{1-q} \right)$ $\sum q_i^q, \langle Q^{q-1}, P \rangle$</td>
<td></td>
</tr>
</tbody>
</table>

$P^q$: the $q$-escort distribution: $i \sim p_i^q/\sum p_i^q$.

$P^q$: the (non-normalized) measure $i \sim p_i^q$. 
<table>
<thead>
<tr>
<th>name</th>
<th>entropy</th>
<th>dual entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>BGS</td>
<td>$H^{BGS}$</td>
<td>$H^{BGS}$</td>
</tr>
<tr>
<td>$q$-Rényi</td>
<td>$H_q^R$</td>
<td>$H^{BGS}$</td>
</tr>
<tr>
<td>$q$-Tsallis$^1$</td>
<td>$H_q^T$</td>
<td>$H_q^T$</td>
</tr>
<tr>
<td>$q$-Tsallis$^2$</td>
<td>$H_q^T$</td>
<td>$H_{2-q}^T$</td>
</tr>
<tr>
<td>$q$-Tsallis$^3$</td>
<td>$H_q^T$</td>
<td>$H_q^T$</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Property</th>
<th>Rényi</th>
<th>Tsallis</th>
</tr>
</thead>
<tbody>
<tr>
<td>consistent inf.</td>
<td>$q &lt; 1$ only</td>
<td>$q &lt; 1$ only</td>
</tr>
<tr>
<td>concave</td>
<td>$q &lt; 1$, few other</td>
<td>all $q$</td>
</tr>
<tr>
<td>composable</td>
<td>all $q$</td>
<td>all $q$</td>
</tr>
<tr>
<td>additive</td>
<td>all $q$</td>
<td>no $q$</td>
</tr>
<tr>
<td>interpretation</td>
<td>hmmm</td>
<td>hmmm</td>
</tr>
<tr>
<td>experimental evidence</td>
<td>hmmm</td>
<td>hmmm</td>
</tr>
</tbody>
</table>
Φ-exponential families etc.

To simplify, assume structure as in Tsallis$^3$ (the Bregman case):

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

$\tilde{Q}$: a certain transform of $Q$.

(Problem: Interpretation?)

For functions $f = (f_1, \cdots, f_k)$, define the $\Phi$-exponential family $E$ as the set of distributions $Q$ for which there exist constants $\lambda_0$ and $\lambda_1, \cdots, \lambda_k$ such that:

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

The natural preparations are those of the form

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

From robustness criterion we find immediately:

**Theorem** If $Q \in E \cap \mathcal{P}_a$, then $Q$ is the $\Phi$-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!