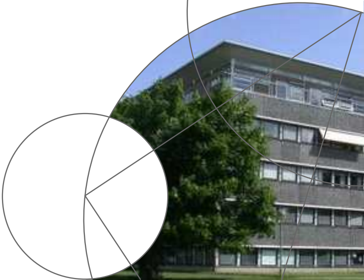
Knowledge modelling after Shannon

Flemming Topsøe, topsoe@math.ku.dk
Department of Mathematical Sciences, University of Copenhagen



Knowledge modelling after Shannon

List of Content

I: Introduction, Information Theoretical Inference

The start: **Shannon**[1], a myriad of followers; relevant here: **Kullback, Čencov, Csiszár, Jaynes, Rissanen, Barron,** later **Grunwald, Dawid, Lauritsen, Matús ...**

> **Ingarden** & **Urbanik**, 1962: *"... information seems intuitively a much simpler and more elementary notion than that of probability ... [it] represents a more primary step of knowledge than that of cognition of probability ..."*
> **Kolmogorov**, $\approx$ 1970: *"Information theory must preceed probability theory and not be based on it"*

... so the need arose to develop a Theory of *Information without probability.*

[1] born 1916, so this year we celebrate the Shannon centenary!

I': Abstract Quantitative Theories of Information

Possible approaches can be based on

- on geometry (**Amari**[2], **Nagaoka**),
- on convexity (**Csiszár, Matús**),
- on complexity (**Solominov, Kolmogorov**),
- or on games (**Pfaffelhuber, FT**).

We shall focus on the approach via games. Convexity will creep in ...

My original motivation: To understand better Tsallis entropy, a purely probabilistic notion, for which the physicists had no natural interpretation. I discovered that my approach (solution!?) to that problem was to a large extent abstract, based on non-probabilistic thinking.

[2] 80 years, thanks and congratulations!

II: Overall Philosophical Basis for Approach

> Mans encounters with the outside world are viewed as situations of conflict between two sides with widely different characteristica and capabilities: Observer and Nature.

Philosophical and also psychological considerations and guiding principles will play a role.

II': Nature and Observer, Roles and Capabilities

- **Nature** holds the truth ($x \in X$, the state space);
- **Observer** seeks the truth but is relegated to belief ($y \in Y$, the belief reservoir.)
  In general $Y \supseteq X$; we assume $Y = X$;
- Nature has no mind!
- Observer has – and can use it constructively, designing experiments or making measurements with the goal to extract knowledge with as little effort as possible;
- Observer can prepare a situation from the world which the players are placed in (a preparation: $\mathcal{P} \subseteq X$).

[If you like, take Nature as female, Observer as male!]

III: 1st guiding Principle, Properness

Properness - or the Perfect Matching Principle:
Minimizing effort should have a training effect.

- An effort function is a function $\Phi : X \times Y \to]-\infty, \infty]$ such that, for all (x, y), $\Phi(x, y) \geq \Phi(x, x)$;
- Φ is proper if, further, equality only holds if $y = x$ (unless $\Phi_x \equiv \infty$);
- $x \mapsto \Phi(x, x)$ is necessity or entropy. Notation: $\mathrm{H}(x)$;
- The excess is divergence: $\mathrm{D}(x, y)$. Thus the important linking identity holds:

$$\Phi(x, y) = \mathrm{H}(x) + \mathrm{D}(x, y).$$

Effort given by Φ you may often think of as description effort.

IV: Three Examples, first one probabilistic:

Shannon Theory. Take $X = Y =$ a probability simplex, say over a finite alphabet $\mathbb{A}$. With

$$\Phi(x, y) = \sum_{i \in \mathbb{A}} x_i \log \frac{1}{y_i} \text{ (Kerridge inaccuracy)}$$

we find the the well known formulas

$$\mathrm{H}(x) = \sum_{i \in \mathbb{A}} x_i \log \frac{1}{x_i} \text{ and } \mathrm{D}(x, y) = \sum_{i \in \mathbb{A}} x_i \log \frac{x_i}{y_i}.$$

(Shannon entropy and Kullback-Leibler divergence.)

IV': Second example, projection in Hilbert Space:

Take $X = Y =$ a Hilbert space, let $y_0 \in Y$, a prior, and take

$$\Phi(x,y) = \|x-y\|^2 - \|x-y_0\|^2. \text{ Then:}$$
$$\mathrm{H}(x) = -\|x-y_0\|^2 \text{ and } \mathrm{D}(x,y) = \|x-y\|^2 .$$

With x restricted to a preparation $\mathcal{P}$, maximizing entropy (Jaynes Principle) corresponds to seeking a (the) projection of y_0 on $\mathcal{P}$.

More natural to work with $-\Phi$, best thought of as a utility function , in fact $\mathrm{U}(x,y) = -\Phi(x,y)$ is a natural measure of the updating gain when replacing the prior y_0 by posterior y.

Results on effort give at the same time results about utility!

IV": Third example, also geometric, but queer:

$X = Y =$ Hilbert space. Now take

$$\Phi(x, y) = \|x - y\|^2 \, .$$

Perfectly acceptable proper effort function, but queer:
Entropy vanishes identically: $\boxed{\mathrm{H} \equiv 0!}$ and $\boxed{\mathrm{D} = \Phi}$, thus the linking identity becomes something very tame in this case.

We will later see how to "un-tame" it and obtain an example related to a classical problem within location theory:

Sylvester's Problem: *To determine the point in the plane with the least maximal distance to a given finite set of points.*

V: Visibility

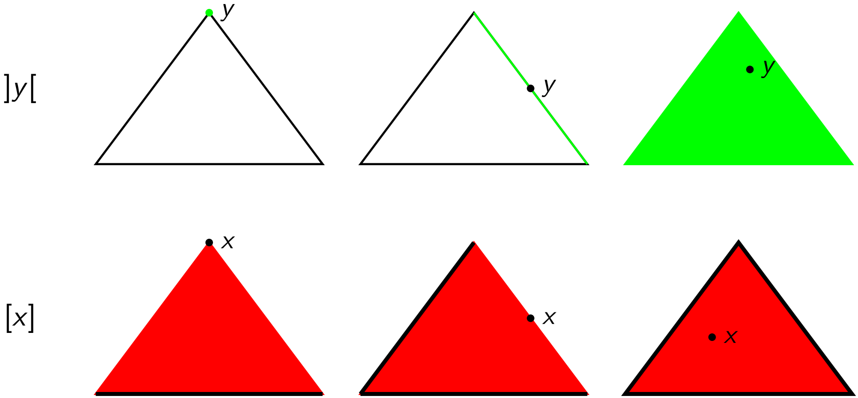
This us an innocent refinement, which you may at first choose to ignore. What we do is to replace $X \times Y$ by a

relation $X \otimes Y$, called visibility. A pair $(x, y) \in X \otimes Y$ is an atomic situation and we write $y \succ x$ and say that x is visible from y. We assume that $x \succ x$ for all states x. Notation: $]y[= \{x|y \succ x\}$ and $[x] = \{y|y \succ x\}$. Example: next slide!

An effort function is now defined only on $X \otimes Y$. Likewise for divergence. Entropy is defined on all of X.

Other possible refinements include the introduction of a subset $Y_{\text{det}} \subseteq Y$ of certain beliefs.

V': Visibility in a Probability Simplex



VI: 2nd Guide: From belief to Action and Control

Good's mantra:
Belief is a tendency to act!

Introduce a map $y \mapsto \hat{y}$, called response, which maps Y into an action space W. Response need not be injective. We write $W = \hat{Y}$. Elements in W are actions, or controls. W may contain $w_\emptyset$, the empty action or empty control. We assume that $\hat{y} = w_\emptyset$ if $y \in Y_{\text{det}}$.

Further, we assume given a relation $X \otimes \hat{Y}$ from X to $\hat{Y}$, controlability. Pairs $(x, w) \in X \otimes \hat{Y}$ are atomic situations (in the $\hat{Y}$-domain); we write $w \succ x$ and say that w controls x. If $w = \hat{x}$, w is adapted to x. We assume that $\hat{x} \succ x$ for all x. Often there will exist universal controls: $(w \succ x \; \forall x \in X)$.

Now focus on functions for $\hat{Y}$-domain in place of $(\Phi, \mathrm{H}, \mathrm{D})$:

VI': New definitions ($\hat{Y}$-domain)

- An effort function ($\hat{Y}$-domain) is a function $\hat{\Phi} : X \otimes \hat{Y} \to]-\infty, \infty]$ such that, for all atomic situations, $\hat{\Phi}(x, w) \geq \hat{\Phi}(x, \hat{x})$;
- $\hat{\Phi}$ is proper if, further, equality only holds if $w = \hat{x}$ (unless $\hat{\Phi}_x \equiv \infty$); more general definition later
- $x \mapsto \hat{\Phi}(x, \hat{x})$ is entropy. Notation unchanged: $\mathrm{H}(x)$;
- The excess is redundancy: $\hat{\mathrm{D}}(x, w)$. Thus the important linking identity holds:

$$\hat{\Phi}(x, w) = \mathrm{H}(x) + \hat{\mathrm{D}}(x, w)$$

If need be, introduce derived visibility, derived effort and derived divergence:

$X \otimes Y = \{(x, y) | (x, \hat{y}) \in X \otimes \hat{Y}\}$;
$\Phi(x, y) = \hat{\Phi}(x, \hat{y})$, $\mathrm{D}(x, y) = \hat{\mathrm{D}}(x, \hat{y})$ for $(x, y) \in X \otimes Y$.

VI'': Some merits

Merits of working in $\hat{Y}$-domain:

- formally,more general (as response need not be injective);
- useful;
- natural;
- a simple extension to work with.

In many examples we do not need to care much about Y.
But caution: Φ derived from a proper $\hat{\Phi}$ need not be proper as you can then only conclude $\hat{y} = \hat{x}$ from $\Phi(x, y) = \mathrm{H}(x)$.

In the further development we shall focus not only on effort, but on all three functions appearing in the linking identity.

VII: Information Triples

Given X, W $(= \hat{Y})$, response $(x \in X \mapsto w = \hat{x} \in W)$ and controllability $X \otimes \hat{Y}$, consider the following properties of a triple $(\hat{\Phi}, \mathrm{H}, \hat{\mathrm{D}})$:

- L (linking): $\hat{\Phi}(x,w) = \mathrm{H}(x) + \hat{\mathrm{D}}(x,w)$;
- F (fundamental inequality): $\hat{\mathrm{D}}(x,w) \geq 0$;
- S (soundness): $\hat{\mathrm{D}}(x,\hat{x}) = 0$;
- P (properness): $w \neq \hat{x} \Rightarrow \hat{\mathrm{D}}(x,w) > 0$. **Definitions:**

- $(\hat{\Phi}, \mathrm{H}, \hat{\mathrm{D}})$ is an (effort based) information triple if L,F and S hold. $\hat{\Phi}$ is effort, H is entropy and $\hat{\mathrm{D}}$ redundancy.
- $(\hat{\Phi}, \mathrm{H}, \hat{\mathrm{D}})$ is an (effort based) proper information triple if L,F,S and P hold (in that case, $\hat{\Phi}$ is a proper effort function as defined before);
- Given only $\hat{\mathrm{D}}$, $\hat{\mathrm{D}}$ is a proper redundancy function if F,S and P hold.

VII': Utility-based Information Triples

Given X, W $(= \hat{Y})$, response $(x \in X \mapsto w = \hat{x} \in W)$ and controllability $X \otimes \hat{Y}$, consider the following properties of a triple $(\hat{\mathrm{U}}, \mathrm{M}, \hat{\mathrm{D}})$:

- L (linking): $\hat{\mathrm{U}}(x, w) = \mathrm{M}(x) - \hat{\mathrm{D}}(x, w)$;
- F (fundamental inequality): $\hat{\mathrm{D}}(x, w) \geq 0$;
- (soundness): $\hat{\mathrm{D}}(x, \hat{x}) = 0$;
- P (properness): $w \neq \hat{x} \Rightarrow \hat{\mathrm{D}}(x, w) > 0$. **Definitions:**

- $(\hat{\mathrm{U}}, \mathrm{M}, \hat{\mathrm{D}})$ is a (utility-based) information triple if L,F and S hold. $\hat{\mathrm{U}}$ is utility, M is max-utility and $\hat{\mathrm{D}}$ is redundancy.
- $(\hat{\mathrm{U}}, \mathrm{M}, \hat{\mathrm{D}})$ is an (utility-based) proper information triple if L,F,S and P hold.
- Given only $\hat{\mathrm{D}}$, $\hat{\mathrm{D}}$ is a proper redundancy function if F,S and P hold.

Thus, $(\hat{\mathrm{U}}, \mathrm{M}, \hat{\mathrm{D}})$ has nice property as a utility-based triple if and only if $(-\hat{\mathrm{U}}, -\mathrm{M}, \hat{\mathrm{D}})$ has so as an effort-based triple.

VII'': Repeating definitions, adding some facts...

L: $\hat{\Phi} = \mathrm{H} + \hat{\mathrm{D}}$; **F:** $\hat{\mathrm{D}} \geq 0$; **S:** $\hat{\mathrm{D}}(x, \hat{x}) = 0$; **P:** $\hat{\mathrm{D}}(x, w) > 0$ for $w \neq \hat{x}$.

- $(\hat{\Phi}, \mathrm{H}, \hat{\mathrm{D}})$ is an information triple (I-Trip) $.\!\cdot.$ **L,F,S** hold;
- An I-Trip is degenerate $.\!\cdot.$ $\hat{\mathrm{D}} \equiv 0$; then I-Trip is $(\mathrm{H}, \mathrm{H}, 0)$;
- An I-Trip is proper $.\!\cdot.$ also **P** holds;
- Given only a function $\hat{\mathrm{D}}$, $\hat{\mathrm{D}}$ is proper $.\!\cdot.$ **F,S,P** hold; then $(\hat{\mathrm{D}}, 0, \hat{\mathrm{D}})$ is a proper I-Trip;
- I-Trips are equivalent $.\!\cdot.$ they have the same redundancy;
- Initial I-Trip of I-Trip $(\hat{\Phi}, \mathrm{H}, \hat{\mathrm{D}})$ $.\!\cdot.$ the I-Trip $(\hat{\mathrm{D}}, 0, \hat{\mathrm{D}})$;
- Adding (or integrating) I-Trips leads to I-Trips; if one is proper, so is the resulting one;
- Given an I-Trip, any equivalent one is obtained from the initial I-Trip by adding any degenerate I-Trip.

VII''': More on structure of information triples

- If two I-Trips differ by a positive factor, they are scalarly equivalent – choice among scarlarly equivalent I-Trips amounts to a choice of unit;
- Relativization involves prior and choice by Observer of posterior. Already indicated in an example; classically leads to information projections;
- Randomization requires affine structure on X, illustrated on a following slide for Sylvester example; classically leads to capacity determination for information channels.

Natural Problem: Representation via "primitive" triples.
Leads to Bregman set-up...

VIII: Game Theory applied to I-Triples

Given proper I-Trip $(\hat{\Phi}, \mathrm{H}, \hat{\mathrm{D}})$ and a preparation $\mathcal{P} \subseteq X$.

For the effort game $\hat{\gamma}(\mathcal{P})$:
- strategies for Nature are $x \in \mathcal{P}$, for Observer $w \succ \mathcal{P}$,
- object function $\hat{\Phi}$, Nature maximizer, Observer minimizer.

The two values for the game are

$$\sup_{x \in \mathcal{P}} \inf_{w \succ x} \hat{\Phi}(x, w) = \sup_{x \in \mathcal{P}} \mathrm{H}(x) = \mathrm{H}_{\max}(\mathcal{P})\,;$$

$$\inf_{w \succ \mathcal{P}} \sup_{x \in \mathcal{P}} \hat{\Phi}(x, w) = \inf_{w \succ \mathcal{P}} \hat{\mathrm{Ri}}(w|\,\mathcal{P}) = \hat{\mathrm{Ri}}_{\min}(\mathcal{P}).$$

("Ri" for "risk"). The minimax-inequality $\mathrm{H}_{\max}(\mathcal{P}) \leq \hat{\mathrm{Ri}}_{\min}(\mathcal{P})$ always holds. Notions of equilibrium (à la Nash) and optimal strategies are introduced as usual.

VIII':Typical results for $\hat{\gamma}(\mathcal{P})$

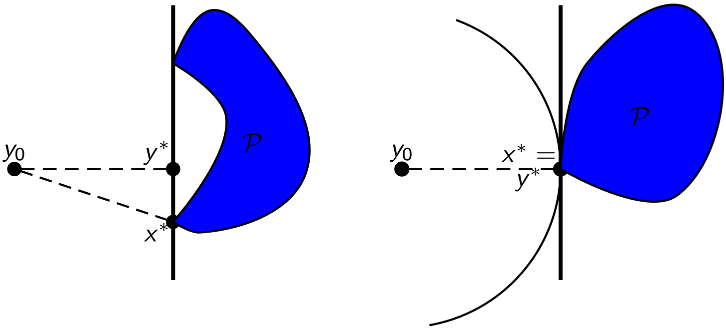
Thesis: "normally" $\hat{\gamma}(\mathcal{P})$ is in Nash-equilibrium and there exists a bi-optimal pair of strategies, (x^*, w^*) such that $w^* = \hat{x^*}$. w^* is the unique optimal strategy for Observer and all optimal strategies for Nature are equivalent under response (hence unique if response is injective). Further, the direct as well as the indirect Pythagorean inequalities hold:

$$\mathrm{H}(x) + \hat{\mathrm{D}}(x, w^*) \leq \mathrm{H}(x^*) \text{ for } x \in \mathcal{P}$$

$$\hat{\mathrm{Ri}}(w^*|\mathcal{P}) + \hat{\mathrm{D}}(x^*, w) \leq \hat{\mathrm{Ri}}(w|\mathcal{P}) \text{ for } w \succ \mathcal{P}\,.$$

Important special cases where this can be checked: If w^* is robust i.e., for some constant h, effort is independent of Nature's strategy: $\hat{\Phi}(x, w^*) = h$ for all $x \in \mathcal{P}$. Then Pythagoras inequality holds with equality.

(related to exponential families)


y_0
y^*
$\mathcal{P}$
x^*
y_0
$x^* =$
y^*
$\mathcal{P}$

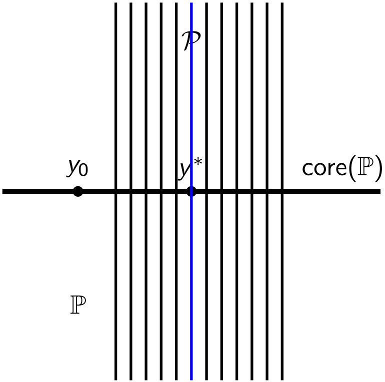


Figure: Preparation family and its core

IX: Randomization, Sylvester, Capacity

Find location $\eta \in \mathbb{R}^2$ *with* $\max_{\xi \in A} \|\xi - \eta\|$ *minimal (A finite)*

Try effort$=\|\xi - \eta\|^2$. Modify by randomization: State space: probability distributions over A, control space: $\mathrm{co}(A)$ (or $\mathbb{R}^2$) and response: barycentric map. Notation: State $P = (p_\xi)_{\xi \in A}$, control η, response: $\hat{P} = \sum_{\xi \in A} p_\xi \cdot \xi$. Now define effort by

$$\hat{\Phi}(P, \eta) = \sum_{\xi \in A} p_\xi \|\xi - \eta\|^2. \text{ Then}$$

$$\hat{\Phi}(P, \eta) = \sum_{\xi \in A} p_\xi \|(\xi - \hat{P}) + (\hat{P} - \eta)\|^2 = \sum_{\xi \in A} p_\xi \|\xi - \hat{P}\|^2 + \|\hat{P} - \eta\|^2,$$

hence $\mathrm{H}(P) = \sum_{\xi \in A} p_\xi \|\xi - \hat{P}\|^2$ and $\hat{\mathrm{D}}(P, \eta) = \|\hat{P} - \eta\|^2$.

$(\hat{\Phi}, \mathrm{H}, \hat{\mathrm{D}})$ is a proper I-Trip. For associated game and any control η: $\hat{\mathrm{Ri}}(\eta) = \max_P \hat{\Phi}(P, \eta) = \max_{\xi \in A} \|\xi - \eta\|^2$.

IX': Kuhn-Tucker type results

So $\hat{\mathrm{R}}\mathrm{i}_{\min} = \min_\eta \max_\xi \|\xi - \eta\|^2$, just what Sylvester looked for - except for the square. But who cares! By robustness, if a point η has the same distance to all points in A, this is the location Sylvester sought. With a simple extension of robustness which applies to randomized models (and with some extra work on necessity), one can prove:

Necessary and sufficient that $\eta \in \mathrm{co}(A)$ is a solution, necessarily unique, to Sylvester's problem is that, for some constant R, $\|\xi - \eta\| \le R$ for all $\xi \in A$ and that η can be written in the form $\eta = \hat{P}$ with $\|\xi - \eta\| = R$ for all $\xi \in A$ with $p_\xi > 0$.

Obs: Resemblance with well known Kuhn-Tucker type results of information theory on channel capacity. Proof is "the same" and can be based on any proper abstract divergence function which satisfies the so-called compensation identity, $\sum p_i \hat{\mathrm{D}}(x_i, w) = \hat{\mathrm{D}}\big(\sum p_i x_i, w\big) + \sum p_i \hat{\mathrm{D}}(x_i, \hat{x})$ with $x = \sum p_i x_i$.

X: Primitive I-Triples, Bregman Construction

A primitive I-Trip $(\phi, \mathrm{h}, \mathrm{d})$ is one for which $X = Y = I$ is an interval in $\mathbb{R}$. Important are those with affine marginals ϕ^u ($\phi = \phi(s, u)$. Normally we may take $I \otimes I = I \times I$, though variants may be convenient to handle endpoint behaviour.
Bregman construction: Let h be smooth strictly concave function on I. With

$$\phi(s,u) = \mathrm{h}(u) + (s-u)\,\mathrm{h}'(u)\,,$$
$$\mathrm{d}(s,u) = \mathrm{h}(u) - \mathrm{h}(s) + (s-u)\,\mathrm{h}'(u)\,,$$

$(\phi, \mathrm{h}, \mathrm{d})$ is a proper primitive I-Trip.

X': Bregman Construction

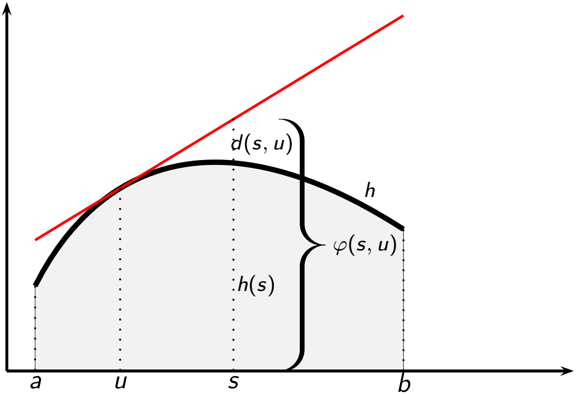


Figure: Bregman generator and primitive effort-based information triple

X": Standard concrete Examples

Example 1. Standard algebraic triple is given by

$$\phi(s,u) = u^2 - 2su \text{ (affine in } s!),$$
$$\mathrm{h}(s) = -s^2\,,$$
$$\mathrm{d}(s,u) = (s-u)^2$$

over $]-\infty,+\infty[$. By integration, this leads to basic concepts from Hilbert space theory.

Example 2. Standard logarithmic triple given by

$$\phi(s,u) = u - s + s\ln\frac{1}{u} \text{ (affine in } s!),$$
$$\mathrm{h}(s) = s\ln\frac{1}{s}\,,$$
$$\mathrm{d}(s,u) = u - s + s\ln\frac{s}{u}\,.$$

over $[0,\infty]$. By integration, this leads to basic concepts from Shannon theory.

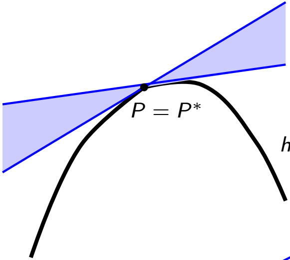
XI: Relaxed notion of properness

For Bregman's construction, it is natural to allow concave generators h which are not necessarily strictly concave. This can be achieved by two extended definitions:

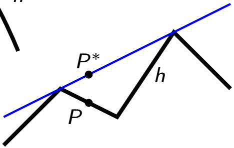
- a general extension of properness to weak properness (here corresponding to $\mathcal{P} = X$):
  P': If $w \neq \hat{x}$ and $\hat{\mathrm{D}}(x, w) = 0$, then $\hat{\mathrm{R}}\mathrm{i}(w) > \mathrm{H}(x)$;
  (a stronger form requires $\hat{\mathrm{R}}\mathrm{i}(w) > \hat{\mathrm{R}}\mathrm{i}(\hat{x})$ which in a way is a more natural condition);
- a notion of extended control for h, viz. that you, for $x \in I$, as $w = \hat{x}$ take that line through $(x, h(x))$ which controls h (lies on or above h) and is closest to a horizontal line.

Important to work in the $\hat{Y}$ domain (natural attempt for Y-domain will not allow h to have horizontal parts). You also have to extend the general theory (bi-optimality etc.) to the general case. Going further, h need not even be concave...

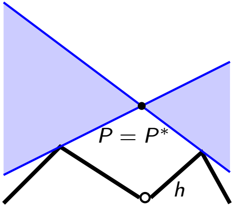
XI': Possible generalisation, Bregman Case



(a) Non-smooth generator



(b) Mixed concave/convex



(c) Upper semi-continuous generator with a discontinuity point

Figure: Possible types of generators.

XII: Uniqueness of Shannon and Tsallis entropy

Problem: How to choose the effort function?

We shall study models where truth and belief interact and result in perception for Observer. We only consider discrete probabilistic models with states $x = (x_i)_{i \in \mathbb{A}}$ and belief instances $y = (y_i)_{i \in \mathbb{A}}$ over an infinite alphabet $\mathbb{A}$ (possibly $\sum_i y_i \leq 1$). Write $y \succ x$ if $\text{supp}(x) \subseteq \text{supp}(y)$. Also put $Y_{det} = \{y | \exists i : y_i = 1\}$.

We assume that interaction acts locally. Formally: Let $\pi = \pi(s, u)$, the interactor, be defined on $[0, 1] \times [0, 1]$ and interpret $\pi(s, u)$ as the perceived intensity of an event with true intensity s and believed intensity u. Then, for an atomic situation (x, y), the perceived intensity of event with index i is $\pi(x_i, y_i)$.

Denote models of this type Ω_π. We assume that π is sound ($\pi(s, s) = s$) and sufficiently smooth.

Examples: Ω_q with $\pi_q(s, t) = qs + (1 - q)t$. For $q = 1$ you get the classical world, for $q = 0$ a black hole.

XII': Description effort

We base the analysis on the notion of a descriptor, again assumed to act locally. This is a function $\kappa : [0,1] \to [0,\infty]$. Interpretation: If Observer believes an event has probability u, then, with an effort $\kappa(u)$, the description effort, he can describe the event.

We require that κ is non-increasing, that $\kappa(1) = 0$ and that $\kappa'(1) = -1$ (a condition of normalization). By definition, this gives effort and information in natural units.

The pointwise effort function, respectively the full effort function are the functions

$$\phi(s,u) = \pi(s,u)\kappa(u)$$

$$\Phi(x,y) = \sum_{i \in I} \phi(x_i, y_i).$$

XII'': gross quantities rather than net quantities

Insight:

- ϕ is "never" proper;
- – but the gross pointwise effort function $\tilde{\phi}(s,u) = \pi(s,u)\kappa(u) + u$, hence also the integrated version $\tilde{\Phi}(x,y) = \sum_{i \in \mathbb{A}} \tilde{\phi}(x_i, y_i)$ stands a chance to be so. If that is the case, we say that κ is proper;
- Interpretation: extra term corresponds to overhead cost
- Working with overhead is technically simpler and helps us to interpret what the unit of information stands for.

XII''': Worlds Ω_π, especially Ω_q ($\pi = \pi_q$)

Theorem.

- For any interactor π, at most one descriptor κ is proper for the world Ω_π;
- No descriptor is proper for Ω_q if $q \leq 0$; however, $q = 0$ is a degenerate case with $\kappa_0(u) = \frac{1}{u} - 1$ and $\mathrm{H}_0(x) = |\operatorname{supp} x|$; Assume now that $q > 0$. Then:
- The proper descriptor κ_q exists and is given by $\kappa_q(u) = \ln_q \frac{1}{u}$, the q-logarithm of $\frac{1}{u}$: $\frac{1}{1-q}\left(u^{q-1} - 1\right)$). The associated entropy function is Havrda&Charvát-Lindhard&Nielsen-Tsallis$\cdots$ entropy;
- Other mean values than π_q, e.g. geometric and harmonic mean, also determine the same proper descriptor κ_q;
- The fundamental inequality even holds in the pointwise version: $\pi(s,u)\kappa(u) + u \geq s\kappa(s) + s$ and is most simply proved in this form.

(Special investigation required if $\mathbb{A}$ has only two elements.)

Conclusions (claims!)

- The theory developed is natural as it builds on sound philosophical considerations which are generally accepted (?!) as representing key features of mans encounters with situations from the world;
- the theory provides a common ground for diverse, at times seemingly unrelated applications
- a switch back and forth from effort to utility (or score) is trivial;
- technical handling is smooth (not much shown, though);
- I refuse to believe that apparent successes are coincidental and claim that the modelling genuinly reflects the "true nature" of basic elements of cognition.
- Main worry: Nothing is said about quantum modelling ... If my ambitious claims are justified, this should be possible!

- What can go wrong *does* go wrong - so better prepare for the worst [Murphy].
- Overhead cost is the natural unit of information.
- You can only know what you can describe.
- Belief is a tendency to act [Good].
- Information is that which induces a change of action or belief.[Caticha]
- Conflict and selfish behaviour can be modelled mathematically - not love and perhaps not even irrationality.
- Support learning by invoking sound training principles.
- Affinity appears to be a necessity behind successful quantitative modelling of information.
- When deciding, choose maximal necessary effort or, if you have something to compare with, minimal maximal deviation, always respecting available information. Any other decision would imply that you had known something more [Jaynes, Kullback].
- Search for natural structural explanations, and reserve the use of non-constructive methods to narrow down the search for solutions.
- Control is essential.