Jürgen Landes

presenting joint work with

Jon Williamson

Centre for Reasoning
University of Kent
Canterbury

Second Reasoning
Club Conference
17.06.2013—19.06.2013

J.landes@kent.ac.uk | http://www.kent.ac.uk/secl/philosophy/staff/j-landes.html

Preface

Dear young reader, to understand the following story let me briefly tell you that Objective Bayesianism is a normative approach to rational belief formation stipulating that

A. Beliefs should satisfy the axioms of probability.

B. Beliefs should satisfy constraints imposed by ones evidence.

C. Beliefs should maximize entropy among the probability functions satisfying the constraints imposed by the agent's evidence.

A Bedtime Story

Chapter 1

So spoke the all-knowing oracle: ``Your beliefs shall be coherent (probabilistic). If they are not the Dutch-Book will make *sure* that you loose money.''

Chapter 2

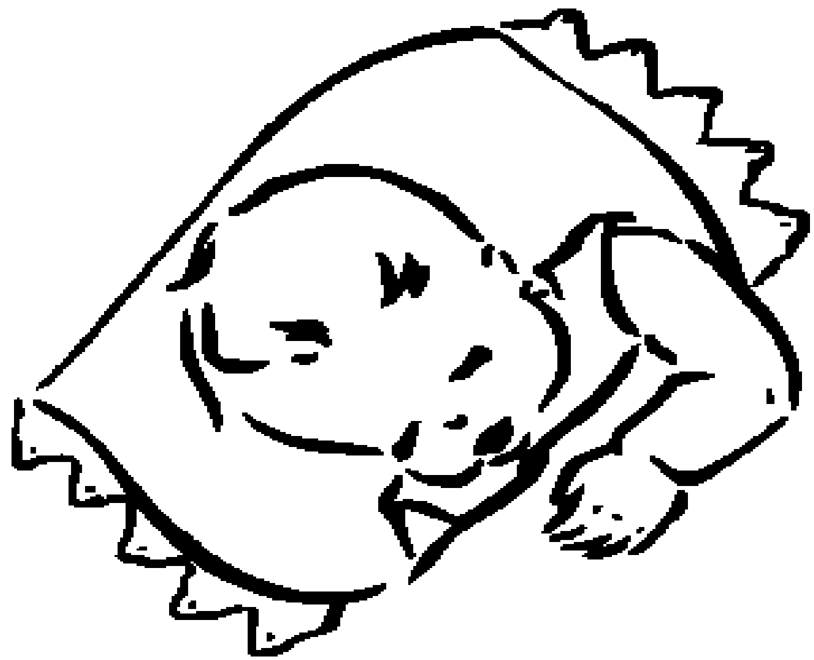
So spoke the all-knowing oracle: ``Your beliefs shall be calibrated. Otherwise, *repeated* betting will loose you money.”

Chapter 3

So spoke the all-knowing oracle: ``Your beliefs shall be maximally equivocal. Otherwise, your ***worst-case expectation*** betting returns are too low.'' *

Fin.

And since the boy was a good Objective Bayesian he slept well; every single night.



One night the son asks his dad: Why should I avoid three different types of loss (sure loss, expected loss, worst-case expected loss)?

His dad did not have an answer and our little hero had a really bad night.

Cooking up a story

Scoring Rules 101 - The usual story

- Idea: Ask DM for a forecast expressing her beliefs, i.e. $bel : S\mathcal{L} \to [0,1]$.
- Denote by Ω the set of states ($\omega = \bigwedge_{1 \le i \le n} \pm x_i$; elementary events).
- If $\omega \in \Omega$ obtains, then DM will suffer loss $L(\omega, bel)$.
- Expected loss then leads to the notion of a *scoring rule*

$$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) L(\omega, bel) \quad .$$

- P is the chance function (distribution of some random variable).
- *Low* score is good!

Scoring Rules 101 - The usual story

- Idea: Ask DM for a forecast expressing her beliefs, i.e. $bel : S\mathcal{L} \to [0,1]$.
- Denote by Ω the set of states ($\omega = \bigwedge_{1 \leq i \leq n} \pm x_i$; elementary events).
- If $\omega \in \Omega$ obtains, then DM will suffer loss $L(\omega, bel)$.
- Expected loss then leads to the notion of a *scoring rule*

$$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) L(\omega, bel) \quad .$$

- P is the chance function (distribution of some random variable).
- *Low* score is good!

Scoring Rules 101 - The usual story

- Idea: Ask DM for a forecast expressing her beliefs, i.e. $bel : S\mathcal{L} \to [0, 1]$.
- Denote by Ω the set of states ($\omega = \bigwedge_{1 \leq i \leq n} \pm x_i$; elementary events).
- If $\omega \in \Omega$ obtains, then DM will suffer loss $L(\omega, bel)$.
- Expected loss then leads to the notion of a *scoring rule*

$$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) L(\omega, bel) \quad .$$

- P is the chance function (distribution of some random variable).
- *Low* score is good!

University of Kent CENTRE FOR REASONING
www.kent.ac.uk/reasoning

Scoring Rules 101 - The usual story

- Idea: Ask DM for a forecast expressing her beliefs, i.e. $bel : S\mathcal{L} \to [0, 1]$.
- Denote by Ω the set of states ($\omega = \bigwedge_{1 \le i \le n} \pm x_i$; elementary events).
- If $\omega \in \Omega$ obtains, then DM will suffer loss $L(\omega, bel)$.
- Expected loss then leads to the notion of a *scoring rule*

$$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) L(\omega, bel) \ .$$

- P is the chance function (distribution of some random variable).
- *Low* score is good!

University of Kent CENTRE FOR REASONING
www.kent.ac.uk/reasoning

Scoring Rules 101 - The usual story

- Idea: Ask DM for a forecast expressing her beliefs, i.e. $bel : S\mathcal{L} \to [0, 1]$.
- Denote by Ω the set of states ($\omega = \bigwedge_{1 \leq i \leq n} \pm x_i$; elementary events).
- If $\omega \in \Omega$ obtains, then DM will suffer loss $L(\omega, bel)$.
- Expected loss then leads to the notion of a *scoring rule*

$$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) L(\omega, bel) \ .$$

- P is the chance function (distribution of some random variable).
- *Low* score is good!

Scoring Rules 101 - The usual story

- Idea: Ask DM for a forecast expressing her beliefs, i.e. $bel : S\mathcal{L} \to [0,1]$.
- Denote by Ω the set of states ($\omega = \bigwedge_{1\leq i\leq n} \pm x_i$; elementary events).
- If $\omega \in \Omega$ obtains, then DM will suffer loss $L(\omega, bel)$.
- Expected loss then leads to the notion of a *scoring rule*

$$S(P, bel) := \sum_{\omega\in\Omega} P(\omega)L(\omega, bel) \ .$$

- P is the chance function (distribution of some random variable).
- *Low* score is good!

University of Kent CENTRE FOR REASONING
www.kent.ac.uk/reasoning

Scoring Rules - Reloaded

- Normally, there are other good reasons (Dutch Book, Cox's Theorem) to adopt a probability function.
- We want to give one account, which makes DM adopt a probability function, i.e. get rid of nightmares.
- Thus, a scoring rule $S(P, bel)$ which only depends on the $bel(\omega)$ for $\omega \in \Omega$ is not going to cut it. – We would have no way to constrain $bel(\omega_1 \vee \omega_2)$.
- Instead, we will consider *extended score*

$$S(P, bel) = \sum_{F \subseteq \Omega} P(F) \cdot L(F, bel)$$

$$\text{compare with } \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, bel) \ .$$

University of Kent CENTRE FOR REASONING
www.kent.ac.uk/reasoning

Scoring Rules - Reloaded

- Normally, there are other good reasons (Dutch Book, Cox's Theorem) to adopt a probability function.
- We want to give one account, which makes DM adopt a probability function, i.e. get rid of nightmares.
- Thus, a scoring rule $S(P, bel)$ which only depends on the $bel(\omega)$ for $\omega \in \Omega$ is not going to cut it. – We would have no way to constrain $bel(\omega_1 \vee \omega_2)$.
- Instead, we will consider *extended score*

$$S(P, bel) = \sum_{F \subseteq \Omega} P(F) \cdot L(F, bel)$$

$$\text{compare with } \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, bel) \quad .$$

Scoring Rules - Reloaded

- Normally, there are other good reasons (Dutch Book, Cox's Theorem) to adopt a probability function.
- We want to give one account, which makes DM adopt a probability function, i.e. get rid of nightmares.
- Thus, a scoring rule $S(P, bel)$ which only depends on the $bel(\omega)$ for $\omega \in \Omega$ is not going to cut it. – We would have no way to constrain $bel(\omega_1 \vee \omega_2)$.
- Instead, we will consider *extended score*

$$S(P, bel) = \sum_{F \subseteq \Omega} P(F) \cdot L(F, bel)$$

$$\text{compare with } \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, bel) \quad .$$

Scoring Rules - Reloaded

- Normally, there are other good reasons (Dutch Book, Cox's Theorem) to adopt a probability function.
- We want to give one account, which makes DM adopt a probability function, i.e. get rid of nightmares.
- Thus, a scoring rule $S(P, bel)$ which only depends on the $bel(\omega)$ for $\omega \in \Omega$ is not going to cut it. – We would have no way to constrain $bel(\omega_1 \vee \omega_2)$.
- Instead, we will consider *extended score*

$$S(P, bel) = \sum_{F \subseteq \Omega} P(F) \cdot L(F, bel)$$

$$\text{compare with } \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, bel) \ .$$

Worst-Case Loss

- However, DM does not know P^*, all she knows is $P^* \in \mathbb{E} \subseteq \mathbb{P}$. Minimizing worst case loss makes sense.
-

$$\sup_{P \in \mathbb{E}} S(P, bel) = \sup_{P \in \mathbb{E}} \sum_{F \subseteq \Omega} P(F) \cdot L(F, bel) \quad .$$

Worst-Case Loss

- However, DM does not know P^*, all she knows is $P^* \in \mathbb{E} \subseteq \mathbb{P}$. Minimizing worst case loss makes sense.
- $$\sup_{P \in \mathbb{E}} S(P, bel) = \sup_{P \in \mathbb{E}} \sum_{F \subseteq \Omega} P(F) \cdot L(F, bel) \ .$$

Proposition Score and Proposition Entropy

- We aim to justify adopting the $P^\dagger$ which maximizes

$$H_\Omega(P) = \sum_{\omega \in \Omega} -P(\omega) \cdot \log(P(\omega)) \quad .$$

- So our loss function will have to be logarithmic.
- Axioms L1 – L4 imply that $L(F, bel) = -\log(bel(F))$.
- $L(F, bel) = \log(bel(F))$ is interpreted as the loss distinct to F, if F obtains.
-

$$S_{\mathcal{P}\Omega}(P, B) := -\sum_{F \subseteq \Omega} P(F) \cdot \log(bel(F))$$

$$H_{\mathcal{P}\Omega}(P) := S_{\mathcal{P}\Omega}(P, P) \quad .$$

Proposition Score and Proposition Entropy

- We aim to justify adopting the $P^\dagger$ which maximizes

$$H_\Omega(P) = \sum_{\omega\in\Omega} -P(\omega)\cdot\log(P(\omega)) \quad .$$

- So our loss function will have to be logarithmic.
- Axioms L1 – L4 imply that $L(F, bel) = -\log(bel(F))$.
- $L(F, bel) = \log(bel(F))$ is interpreted as the loss distinct to F, if F obtains.
-

$$S_{\mathcal{P}\Omega}(P, B) := -\sum_{F\subseteq\Omega} P(F)\cdot\log(bel(F))$$

$$H_{\mathcal{P}\Omega}(P) := S_{\mathcal{P}\Omega}(P, P) \quad .$$

University of Kent CENTRE FOR REASONING www.kent.ac.uk/reasoning

Proposition Score and Proposition Entropy

- We aim to justify adopting the $P^\dagger$ which maximizes

$$H_\Omega(P) = \sum_{\omega\in\Omega} -P(\omega)\cdot\log(P(\omega)) \quad .$$

- So our loss function will have to be logarithmic.
- Axioms L1 – L4 imply that $L(F, bel) = -\log(bel(F))$.
- $L(F, bel) = \log(bel(F))$ is interpreted as the loss distinct to F, if F obtains.
-

$$S_{\mathcal{P}\Omega}(P, B) := -\sum_{F\subseteq\Omega} P(F)\cdot\log(bel(F))$$

$$H_{\mathcal{P}\Omega}(P) := S_{\mathcal{P}\Omega}(P, P) \quad .$$

Proposition Score and Proposition Entropy

- We aim to justify adopting the $P^\dagger$ which maximizes

$$H_\Omega(P) = \sum_{\omega\in\Omega} -P(\omega)\cdot\log(P(\omega)) \ .$$

- So our loss function will have to be logarithmic.
- Axioms L1 – L4 imply that $L(F, bel) = -\log(bel(F))$.
- $L(F, bel) = \log(bel(F))$ is interpreted as the loss distinct to F, if F obtains.
-

$$S_{\mathcal{P}\Omega}(P, B) := -\sum_{F\subseteq\Omega} P(F)\cdot\log(bel(F))$$

$$H_{\mathcal{P}\Omega}(P) := S_{\mathcal{P}\Omega}(P, P) \ .$$

University of Kent CENTRE FOR REASONING www.kent.ac.uk/reasoning

Proposition Score and Proposition Entropy

- We aim to justify adopting the $P^\dagger$ which maximizes

$$H_\Omega(P) = \sum_{\omega \in \Omega} -P(\omega) \cdot \log(P(\omega)) \quad .$$

- So our loss function will have to be logarithmic.
- Axioms L1 – L4 imply that $L(F, bel) = -\log(bel(F))$.
- $L(F, bel) = \log(bel(F))$ is interpreted as the loss distinct to F, if F obtains.
-

$$S_{\mathcal{P}\Omega}(P, B) := -\sum_{F \subseteq \Omega} P(F) \cdot \log(bel(F))$$

$$H_{\mathcal{P}\Omega}(P) := S_{\mathcal{P}\Omega}(P, P) \quad .$$

University of Kent CENTRE FOR REASONING www.kent.ac.uk/reasoning

The loss function L for general beliefs

- Our story is along the lines: Minimize (...) logarithmic loss!
- If $bel(F) = 1$ for all $F \subseteq \Omega$, then $L(F, bel) = -\log(1) = 0$.
- Thus, $S_{\mathcal{P}\Omega}(P, bel) = \sum_{F \subseteq \Omega} P(F) \cdot 0 = 0$.
- So, $bel = 1$ minimizes loss! This is BAD.
- Houston, we have a problem!
- Fact: This same problem raises its ugly head for every local extended strictly proper scoring rule.

The loss function L for general beliefs

- Our story is along the lines: Minimize (...) logarithmic loss!
- If $bel(F) = 1$ for all $F \subseteq \Omega$, then $L(F, bel) = -\log(1) = 0$.
- Thus, $S_{\mathcal{P}\Omega}(P, bel) = \sum_{F \subseteq \Omega} P(F) \cdot 0 = 0$.
- So, $bel \equiv 1$ minimizes loss! This is BAD.
- Houston, we have a problem!
- Fact: This same problem raises its ugly head for every local extended strictly proper scoring rule.

The loss function L for general beliefs

- Our story is along the lines: Minimize (...) logarithmic loss!
- If $bel(F) = 1$ for all $F \subseteq \Omega$, then $L(F, bel) = -\log(1) = 0$.
- Thus, $S_{\mathcal{P}\Omega}(P, bel) = \sum_{F \subseteq \Omega} P(F) \cdot 0 = 0$.
- So, $bel \equiv 1$ minimizes loss! This is BAD.
- Houston, we have a problem!
- Fact: This same problem raises its ugly head for every local extended strictly proper scoring rule.

The loss function L for general beliefs

- Our story is along the lines: Minimize (...) logarithmic loss!
- If $bel(F) = 1$ for all $F \subseteq \Omega$, then $L(F, bel) = -\log(1) = 0$.
- Thus, $S_{\mathcal{P}\Omega}(P, bel) = \sum_{F \subseteq \Omega} P(F) \cdot 0 = 0$.
- So, $bel \equiv 1$ minimizes loss! This is BAD.
- Houston, we have a problem!
- Fact: This same problem raises its ugly head for every local extended strictly proper scoring rule.

The loss function L for general beliefs

- Our story is along the lines: Minimize (...) logarithmic loss!
- If $bel(F) = 1$ for all $F \subseteq \Omega$, then $L(F, bel) = -\log(1) = 0$.
- Thus, $S_{\mathcal{P}\Omega}(P, bel) = \sum_{F \subseteq \Omega} P(F) \cdot 0 = 0$.
- So, $bel \equiv 1$ minimizes loss! This is BAD.
- Houston, we have a problem!
- Fact: This same problem raises its ugly head for every local extended strictly proper scoring rule.

The loss function L for general beliefs

- Our story is along the lines: Minimize (...) logarithmic loss!
- If $bel(F) = 1$ for all $F \subseteq \Omega$, then $L(F, bel) = -\log(1) = 0$.
- Thus, $S_{\mathcal{P}\Omega}(P, bel) = \sum_{F \subseteq \Omega} P(F) \cdot 0 = 0$.
- So, $bel \equiv 1$ minimizes loss! This is BAD.
- Houston, we have a problem!
- Fact: This same problem raises its ugly head for every local extended strictly proper scoring rule.

Normalize!

- For $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\pi = \langle(\omega_1, \omega_2, \omega_4), (\omega_3)\rangle$ is a partition of Ω.
- Let Π be the set of partitions of states of our language.
- Let $M := \max_{\pi\in\Pi} \sum_{F\in\pi} bel(F)$.
- Given a belief function $bel : \{F \subseteq \Omega\} \longrightarrow \mathbb{R}_{\geq 0}$ (bel not zero everywhere), its *normalisation* B is defined as $B(F) := bel(F)/M$.
- Set of normalized belief functions

$$\mathbb{B} := \{B : \{F \subseteq \Omega\} \longrightarrow [0,1] : \sum_{F\in\pi} B(F) = 1 \text{ for some } \pi$$

$$\text{and } \sum_{F\in\pi} B(F) \leq 1 \text{ for all } \pi \in \Pi\} \ .$$

Normalize!

- For $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\pi = \langle(\omega_1, \omega_2, \omega_4), (\omega_3)\rangle$ is a partition of Ω.
- Let Π be the set of partitions of states of our language.
- Let $M := \max_{\pi \in \Pi} \sum_{F \in \pi} bel(F)$.
- Given a belief function $bel : \{F \subseteq \Omega\} \longrightarrow \mathbb{R}_{\geq 0}$ (bel not zero everywhere), its *normalisation* B is defined as $B(F) := bel(F)/M$.
- Set of normalized belief functions

$$\mathbb{B} := \{B : \{F \subseteq \Omega\} \longrightarrow [0,1] : \sum_{F \in \pi} B(F) = 1 \text{ for some } \pi \text{ and } \sum_{F \in \pi} B(F) \leq 1 \text{ for all } \pi \in \Pi\} .$$

Normalize!

- For $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\pi = \langle(\omega_1, \omega_2, \omega_4), (\omega_3)\rangle$ is a partition of Ω.
- Let Π be the set of partitions of states of our language.
- Let $M := \max_{\pi\in\Pi} \sum_{F\in\pi} bel(F)$.
- Given a belief function $bel : \{F \subseteq \Omega\} \longrightarrow \mathbb{R}_{\geq 0}$ (bel not zero everywhere), its *normalisation* B is defined as $B(F) := bel(F)/M$.
- Set of normalized belief functions

$$\mathbb{B} := \{B : \{F \subseteq \Omega\} \longrightarrow [0,1] : \sum_{F\in\pi} B(F) = 1 \text{ for some } \pi$$

$$\text{and } \sum_{F\in\pi} B(F) \leq 1 \text{ for all } \pi \in \Pi\} \, .$$

Normalize!

- For $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\pi = \langle(\omega_1, \omega_2, \omega_4), (\omega_3)\rangle$ is a partition of Ω.
- Let Π be the set of partitions of states of our language.
- Let $M := \max_{\pi\in\Pi} \sum_{F\in\pi} bel(F)$.
- Given a belief function $bel : \{F \subseteq \Omega\} \longrightarrow \mathbb{R}_{\geq 0}$ (bel not zero everywhere), its *normalisation* B is defined as $B(F) := bel(F)/M$.
- Set of normalized belief functions

$$\mathbb{B} := \{B : \{F \subseteq \Omega\} \longrightarrow [0,1] : \sum_{F\in\pi} B(F) = 1 \text{ for some } \pi$$

$$\text{and } \sum_{F\in\pi} B(F) \leq 1 \text{ for all } \pi \in \Pi\} \; .$$

Normalize!

- For $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\pi = \langle(\omega_1, \omega_2, \omega_4), (\omega_3)\rangle$ is a partition of Ω.
- Let Π be the set of partitions of states of our language.
- Let $M := \max_{\pi\in\Pi} \sum_{F\in\pi} bel(F)$.
- Given a belief function $bel : \{F \subseteq \Omega\} \longrightarrow \mathbb{R}_{\geq 0}$ (bel not zero everywhere), its *normalisation* B is defined as $B(F) := bel(F)/M$.
- Set of normalized belief functions

$$\mathbb{B} := \{B : \{F \subseteq \Omega\} \longrightarrow [0,1] : \sum_{F\in\pi} B(F) = 1 \text{ for some } \pi$$

$$\text{and } \sum_{F\in\pi} B(F) \leq 1 \text{ for all } \pi \in \Pi\} \; .$$

Good News Everyone!

Theorem – Norm 1, 2

For convex $\mathbb{E} \subseteq \mathbb{P}$

$$\arg \inf_{B\in\mathbb{B}} \sup_{P\in\mathbb{E}} S^{\log}_{\mathcal{P}\Omega}(P,B) = \arg \sup_{P\in\mathbb{E}} H_{\mathcal{P}\Omega}(P) = \{P^{\dagger}_{\mathcal{P}\Omega}\} \ .$$

Theorem – Norm 1, 2, 3

If $P_{=} \in \bar{\mathbb{E}}$, *then*

$$\arg \inf_{B\in\mathbb{B}} \sup_{P\in\mathbb{E}} S_{\mathcal{P}\Omega}(P,B) = \arg \sup_{P\in\mathbb{E}} H_{\mathcal{P}\Omega}(P) = \{P_{=}\} = \arg \sup_{P\in\mathbb{E}} H_{\Omega}(P)$$

Good News Everyone!

Theorem – Norm 1, 2

For convex $\mathbb{E} \subseteq \mathbb{P}$

$$\arg\inf_{B\in\mathbb{B}} \sup_{P\in\mathbb{E}} S^{\log}_{\mathcal{P}\Omega}(P,B) = \arg\sup_{P\in\mathbb{E}} H_{\mathcal{P}\Omega}(P) = \{P^{\dagger}_{\mathcal{P}\Omega}\} \ .$$

Theorem – Norm 1, 2, 3

If $P_{=} \in \bar{\mathbb{E}}$, *then*

$$\arg\inf_{B\in\mathbb{B}} \sup_{P\in\mathbb{E}} S_{\mathcal{P}\Omega}(P,B) = \arg\sup_{P\in\mathbb{E}} H_{\mathcal{P}\Omega}(P) = \{P_{=}\} = \arg\sup_{P\in\mathbb{E}} H_{\Omega}(P)$$

University of Kent CENTRE FOR REASONING
www.kent.ac.uk/reasoning

Not so good news

Theorem

There exists a convex $\mathbb{E}$ *such that*

$$\arg\inf_{B\in\mathbb{B}}\sup_{P\in\mathbb{E}} S_{\mathcal{P}\Omega}(P,B) = \{P^{\dagger}_{\mathcal{P}\Omega}\} \quad \neq \quad \arg\sup_{P\in\mathbb{E}} H_{\Omega}(P) \ .$$

Partition Entropy

- There is another plausible way to define extended score:

$$S_\Pi(P,B) := \sum_{\pi\in\Pi}\sum_{F\in\pi} -P(F)\cdot\log(B(F))$$
$$= \sum_{F\subseteq\Omega}\Big(\sum_{\substack{\pi\in\Pi\\ F\in\pi}} 1\Big) - P(F)\cdot\log(B(F))$$
$$H_\Pi(P) := S_\Pi(P,P) \quad .$$

Good News Everyone!

Theorem – Norm 1, 2

For convex $\mathbb{E} \subseteq \mathbb{P}$

$$\arg \inf_{B \in \mathbb{B}} \sup_{P \in \mathbb{E}} S_\Pi(P, B) = \arg \sup_{P \in \mathbb{E}} H_\Pi(P) = \{P_\Pi^\dagger\} \ .$$

Theorem – Norm 1, 2, 3

If $P_= \in \bar{\mathbb{E}}$, *then*

$$\arg \inf_{B \in \mathbb{B}} \sup_{P \in \mathbb{E}} S_\Pi^{\log}(P, B) = \arg \sup_{P \in \mathbb{E}} H_\Pi(P) = \{P_=\} = \arg \sup_{P \in \mathbb{E}} H_\Omega(P)$$

Good News Everyone!

Theorem – Norm 1, 2

For convex $\mathbb{E} \subseteq \mathbb{P}$

$$\arg \inf_{B \in \mathbb{B}} \sup_{P \in \mathbb{E}} S_\Pi(P, B) = \arg \sup_{P \in \mathbb{E}} H_\Pi(P) = \{P_\Pi^\dagger\} \ .$$

Theorem – Norm 1, 2, 3

If $P_= \in \bar{\mathbb{E}}$*, then*

$$\arg \inf_{B \in \mathbb{B}} \sup_{P \in \mathbb{E}} S_\Pi^{\log}(P, B) = \arg \sup_{P \in \mathbb{E}} H_\Pi(P) = \{P_=\} = \arg \sup_{P \in \mathbb{E}} H_\Omega(P)$$

Not so good news

Theorem

There exists a convex $\mathbb{E}$ *such that*

$$\arg\sup_{P\in\mathbb{E}} H_{\Pi}(P) \neq \arg\sup_{P\in\mathbb{E}} H_{\Omega}(P) \neq \arg\sup_{P\in\mathbb{E}} H_{\mathcal{P}\Omega}(P) \ .$$

g-Entropy

- There is a general way to define extended score:

$$S_g(P,B) := \sum_{\pi\in\Pi} g(\pi) \sum_{F\in\pi} -P(F)\cdot\log(B(F))$$
$$= \sum_{F\subseteq\Omega}\Big(\sum_{\substack{\pi\in\Pi\\ F\in\pi}} g(\pi)\Big) - P(F)\cdot\log(B(F))$$
$$H_g(P) := S_g(P,P) \ .$$

- $g:\Pi\to\mathbb{R}_{\geq 0}$ such that $\sum_{\substack{\pi\in\Pi\\ F\in\pi}} g(\pi) > 0$ for all $F\subseteq\Omega$.

g-Entropy

- There is a general way to define extended score:

$$S_g(P,B) := \sum_{\pi \in \Pi} g(\pi) \sum_{F \in \pi} -P(F) \cdot \log(B(F))$$
$$= \sum_{F \subseteq \Omega} \Big(\sum_{\substack{\pi \in \Pi \\ F \in \pi}} g(\pi) \Big) - P(F) \cdot \log(B(F))$$
$$H_g(P) := S_g(P,P) \quad .$$

- $g : \Pi \to \mathbb{R}_{\geq 0}$ such that $\sum_{\substack{\pi \in \Pi \\ F \in \pi}} g(\pi) > 0$ for all $F \subseteq \Omega$.

Good News Everyone!

Theorem – Norm 1, 2

For convex $\mathbb{E} \subseteq \mathbb{P}$

$$\arg \inf_{B \in \mathbb{B}} \sup_{P \in \mathbb{E}} S_g(P, B) = \arg \sup_{P \in \mathbb{E}} H_g(P) = \{P_g^\dagger\} \ .$$

Theorem – Norm 1, 2, 3

If $P_= \in \bar{\mathbb{E}}$ *and if* g *is symmetric, then*

$$\arg \inf_{B \in \mathbb{B}} \sup_{P \in \mathbb{E}} S_g^{\log}(P, B) = \arg \sup_{P \in \mathbb{E}} H_g(P) = \{P_=\} = \arg \sup_{P \in \mathbb{E}} H_\Omega(P)$$

Good News Everyone!

Theorem – Norm 1, 2

For convex $\mathbb{E} \subseteq \mathbb{P}$

$$\arg \inf_{B \in \mathbb{B}} \sup_{P \in \mathbb{E}} S_g(P, B) = \arg \sup_{P \in \mathbb{E}} H_g(P) = \{P_g^{\dagger}\} \quad .$$

Theorem – Norm 1, 2, 3

If $P_= \in \bar{\mathbb{E}}$ *and if g is symmetric, then*

$$\arg \inf_{B \in \mathbb{B}} \sup_{P \in \mathbb{E}} S_g^{\log}(P, B) = \arg \sup_{P \in \mathbb{E}} H_g(P) = \{P_=\} = \arg \sup_{P \in \mathbb{E}} H_\Omega(P)$$

Mixed News

Conjecture – Norm 3?

For all (reasonable) g there exists a convex $\mathbb{E}$ such that

$$\arg \inf_{B\in\mathbb{B}} \sup_{P\in\mathbb{E}} H_g^{\log}(P) \neq \arg \sup_{P\in\mathbb{E}} H_\Omega(P) \quad .$$

Theorem – Norm 3 asterisk

For fixed $\mathbb{E}$ let $P_g^\dagger$ be the unique g-entropy maximizer, then

$$P_\Omega^\dagger \in \overline{\{P_g^\dagger \mid g \text{ senisble}\}} \quad .$$

Mixed News

Conjecture – Norm 3?

For all (reasonable) g there exists a convex $\mathbb{E}$ such that

$$\arg \inf_{B \in \mathbb{B}} \sup_{P \in \mathbb{E}} H_g^{\log}(P) \quad \neq \quad \arg \sup_{P \in \mathbb{E}} H_\Omega(P) \ .$$

Theorem – Norm 3 asterisk

For fixed $\mathbb{E}$ let $P_g^\dagger$ be the unique g-entropy maximizer, then

$$P_\Omega^\dagger \in \overline{\{P_g^\dagger \mid g \text{ senisble}\}} \ .$$

The Boy sleeps well indeed - he is still very young

The loss function L – Axiomatic Characterization

- L1 $L(F, bel) = 0$, if $bel(F) = 1$.
- L2 Loss strictly increases as $bel(F)$ decreases from 1 towards 0.
- L3 L is local. L is called *local*, if and only if $L(F, bel) = L(bel(F))$.
- L4 Losses are additive when the language is composed of independent sublanguages.
- L1 – L4 imply that $L(bel(F)) = -\log_b(bel(F))$ for some $b \in \mathbb{R}_{>0}$.

The loss function L – Axiomatic Characterization

- L1 $L(F, bel) = 0$, if $bel(F) = 1$.
- L2 Loss strictly increases as $bel(F)$ decreases from 1 towards 0.
- L3 L is local. L is called *local*, if and only if $L(F, bel) = L(bel(F))$.
- L4 Losses are additive when the language is composed of independent sublanguages.
- L1 – L4 imply that $L(bel(F)) = -\log_b(bel(F))$ for some $b \in \mathbb{R}_{>0}$.

University of Kent CENTRE FOR REASONING www.kent.ac.uk/reasoning

The loss function L – Axiomatic Characterization

- L1 $L(F, bel) = 0$, if $bel(F) = 1$.
- L2 Loss strictly increases as $bel(F)$ decreases from 1 towards 0.
- L3 L is local. L is called *local*, if and only if $L(F, bel) = L(bel(F))$.
- L4 Losses are additive when the language is composed of independent sublanguages.
- L1 – L4 imply that $L(bel(F)) = -\log_b(bel(F))$ for some $b \in \mathbb{R}_{>0}$.

The loss function L – Axiomatic Characterization

- L1 $L(F, bel) = 0$, if $bel(F) = 1$.
- L2 Loss strictly increases as $bel(F)$ decreases from 1 towards 0.
- L3 L is local. L is called *local*, if and only if $L(F, bel) = L(bel(F))$.
- L4 Losses are additive when the language is composed of independent sublanguages.
- L1 – L4 imply that $L(bel(F)) = -\log_b(bel(F))$ for some $b \in \mathbb{R}_{>0}$.

The loss function L – Axiomatic Characterization

- L1 $L(F, bel) = 0$, if $bel(F) = 1$.
- L2 Loss strictly increases as $bel(F)$ decreases from 1 towards 0.
- L3 L is local. L is called *local*, if and only if $L(F, bel) = L(bel(F))$.
- L4 Losses are additive when the language is composed of independent sublanguages.
- L1 – L4 imply that $L(bel(F)) = -\log_b(bel(F))$ for some $b \in \mathbb{R}_{>0}$.