Justifying Objective Bayesianism with Scoring Rules

Jürgen Landes
Centre for Reasoning

Workshop on Scoring Rules

LSE

10.06.2013

Outline

Aims of this talk

- What do I do for a living.
- Please stop me!
- This talk is about you; not about me.

Aims of this talk

- What do I do for a living.
- Please stop me!
- This talk is about you; not about me.

Aims of this talk

- What do I do for a living.
- Please stop me!
- This talk is about you; not about me.

Objective Bayesianism

- 3 Principles of Rationality (a subjective belief function of a rational agent ought to satisfy):
  1. Probabilism – Beliefs should satisfy the axioms of probability.
  2. Calibration – Beliefs should satisfy constraints imposed by the available evidence.
  3. Equivocation – "Choose probability function consistent with evidence which is most open-minded."
     (Equivalently: maximize Shannon Entropy among calibrated probability functions)
     With asterisk

Objective Bayesianism

- 3 Principles of Rationality (a subjective belief function of a rational agent ought to satisfy):
  1. Probabilism – Beliefs should satisfy the axioms of probability.
  2. Calibration – Beliefs should satisfy constraints imposed by the available evidence.
  3. Equivocation – "Choose probability function consistent with evidence which is most open-minded."
     (Equivalently: maximize Shannon Entropy among calibrated probability functions)
     With asterisk

Objective Bayesianism

- 3 Principles of Rationality (a subjective belief function of a rational agent ought to satisfy):
  1. Probabilism – Beliefs should satisfy the axioms of probability.
  2. Calibration – Beliefs should satisfy constraints imposed by the available evidence.
  3. Equivocation – "Choose probability function consistent with evidence which is most open-minded."
     (Equivalently: maximize Shannon Entropy among calibrated probability functions)
     With asterisk

Objective Bayesianism

- 3 Principles of Rationality (a subjective belief function of a rational agent ought to satisfy):
  1. Probabilism – Beliefs should satisfy the axioms of probability.
  2. Calibration – Beliefs should satisfy constraints imposed by the available evidence.
  3. Equivocation – "Choose probability function consistent with evidence which is most open-minded."
     (Equivalently: maximize Shannon Entropy among calibrated probability functions)
     With asterisk

Justifying Objective Bayesianism

- The usual story
  1. Probabilism - Dutch Book (one single bet)
     Avoidance of **sure** loss.
  2. Calibration - Repeated betting
     Avoidance of **expected** loss.
  3. Equivocation - Repeated betting
     Avoidance of **worst-case expected** loss.
- Our current goal: Give **one single** justification for OB.
- No need to appeal to **three different** types of loss avoidance.

Justifying Objective Bayesianism

- The usual story
  1. Probabilism - Dutch Book (one single bet)
     Avoidance of **sure** loss.
  2. Calibration - Repeated betting
     Avoidance of **expected** loss.
  3. Equivocation - Repeated betting
     Avoidance of **worst-case expected** loss.
- Our current goal: Give **one single** justification for OB.
- No need to appeal to **three different** types of loss avoidance.

Justifying Objective Bayesianism

- The usual story
  1. Probabilism - Dutch Book (one single bet)
     Avoidance of **sure** loss.
  2. Calibration - Repeated betting
     Avoidance of **expected** loss.
  3. Equivocation - Repeated betting
     Avoidance of **worst-case expected** loss.
- Our current goal: Give **one single** justification for OB.
- No need to appeal to **three different** types of loss avoidance.

Justifying Objective Bayesianism

- The usual story
  1. Probabilism - Dutch Book (one single bet)
     Avoidance of **sure** loss.
  2. Calibration - Repeated betting
     Avoidance of **expected** loss.
  3. Equivocation - Repeated betting
     Avoidance of **worst-case expected** loss.
- Our current goal: Give **one single** justification for OB.
- No need to appeal to **three different** types of loss avoidance.

Justifying Objective Bayesianism

- The usual story
  1. Probabilism - Dutch Book (one single bet)
     Avoidance of **sure** loss.
  2. Calibration - Repeated betting
     Avoidance of **expected** loss.
  3. Equivocation - Repeated betting
     Avoidance of **worst-case expected** loss.
- Our current goal: Give **one single** justification for OB.
- No need to appeal to **three different** types of loss avoidance.

Justifying Objective Bayesianism

- The usual story
  1. Probabilism - Dutch Book (one single bet)
     Avoidance of **sure** loss.
  2. Calibration - Repeated betting
     Avoidance of **expected** loss.
  3. Equivocation - Repeated betting
     Avoidance of **worst-case expected** loss.
- Our current goal: Give **one single** justification for OB.
- No need to appeal to **three different** types of loss avoidance.

Outline

Scoring Rules - Basic Notation

- Idea: Ask agent for a her beliefs, i.e. $bel : S\mathcal{L} \to [0, 1]$.
- Denote by Ω the set of worlds (elementary events, atoms).
- If $\omega \in \Omega$ obtains, then DM will suffer loss $L(\omega, bel)$.
- Expected loss then leads to the notion of a *scoring rule*

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

- *Low* score is good! – Avoid loss.

Scoring Rules - Basic Notation

- Idea: Ask agent for a her beliefs, i.e. *bel* : $S\mathcal{L} \to [0, 1]$.
- Denote by Ω the set of worlds (elementary events, atoms).
- If $\omega \in \Omega$ obtains, then DM will suffer loss $L(\omega, bel)$.
- Expected loss then leads to the notion of a *scoring rule*

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

- *Low* score is good! – Avoid loss.

Scoring Rules - Basic Notation

- Idea: Ask agent for a her beliefs, i.e. $bel : S\mathcal{L} \to [0, 1]$.
- Denote by Ω the set of worlds (elementary events, atoms).
- If $\omega \in \Omega$ obtains, then DM will suffer loss $L(\omega, bel)$.
- Expected loss then leads to the notion of a *scoring rule*

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

- *Low* score is good! – Avoid loss.

Scoring Rules - Basic Notation

- Idea: Ask agent for a her beliefs, i.e. $bel : S\mathcal{L} \to [0,1]$.
- Denote by Ω the set of worlds (elementary events, atoms).
- If $\omega \in \Omega$ obtains, then DM will suffer loss $L(\omega, bel)$.
- Expected loss then leads to the notion of a *scoring rule*

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

- *Low* score is good! – Avoid loss.

Scoring Rules - Basic Notation

- Idea: Ask agent for a her beliefs, i.e. $bel : S\mathcal{L} \to [0,1]$.
- Denote by Ω the set of worlds (elementary events, atoms).
- If $\omega \in \Omega$ obtains, then DM will suffer loss $L(\omega, bel)$.
- Expected loss then leads to the notion of a *scoring rule*

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

- *Low* score is good! – Avoid loss.

Interpreting the Functions

- $$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, bel) \quad .$$
- *bel* is the belief function DM announces.
- Suppose $P = bel^*$, private subjective beliefs.
- A DM minimizing $S(bel^*, bel)$ should announce a probability function, *because* her personal beliefs satisfy the axioms of probability.
- No justification of the probability norm nor the calibration norm!

Interpreting the Functions

- $$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, bel) \quad .$$
- *bel* is the belief function DM announces.
- Suppose $P = bel^*$, private subjective beliefs.
- A DM minimizing $S(bel^*, bel)$ should announce a probability function, *because* her personal beliefs satisfy the axioms of probability.
- No justification of the probability norm nor the calibration norm!

Interpreting the Functions

- $$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, bel) \quad .$$
- *bel* is the belief function DM announces.
- Suppose $P = bel^*$, private subjective beliefs.
- A DM minimizing $S(bel^*, bel)$ should announce a probability function, *because* her personal beliefs satisfy the axioms of probability.
- No justification of the probability norm nor the calibration norm!

Interpreting the Functions

- $$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, bel) \quad .$$
- *bel* is the belief function DM announces.
- Suppose $P = bel^*$, private subjective beliefs.
- A DM minimizing $S(bel^*, bel)$ should announce a probability function, *because* her personal beliefs satisfy the axioms of probability.
- No justification of the probability norm nor the calibration norm!

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

Interpreting the Functions

- $$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, bel) \quad .$$
- *bel* is the belief function DM announces.
- Suppose $P = bel^*$, private subjective beliefs.
- A DM minimizing $S(bel^*, bel)$ should announce a probability function, *because* her personal beliefs satisfy the axioms of probability.
- No justification of the probability norm nor the calibration norm!

Interpreting the Functions 2

- $$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, bel) \ .$$
- It makes much more sense to interpret P as the objective chance function P^* — if you believe in such a thing.
- Then, minimizing score can be interpreted as minimizing inaccuracy; with respect to L.
- However, DM does not know P^*, all she knows is $P^* \in \mathbb{E} \subseteq \mathbb{P}$. Minimizing worst case loss makes sense:

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

University of Kent Centre for Reasoning www.kent.ac.uk/reasoning

Interpreting the Functions 2

- $$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, bel) \quad .$$
- It makes much more sense to interpret P as the objective chance function P^* — if you believe in such a thing.
- Then, minimizing score can be interpreted as minimizing inaccuracy; with respect to L.
- However, DM does not know P^*, all she knows is $P^* \in \mathbb{E} \subseteq \mathbb{P}$. Minimizing worst case loss makes sense:

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

Interpreting the Functions 2

- $$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, bel) \quad .$$
- It makes much more sense to interpret P as the objective chance function P^* — if you believe in such a thing.
- Then, minimizing score can be interpreted as minimizing inaccuracy; with respect to L.
- However, DM does not know P^*, all she knows is $P^* \in \mathbb{E} \subseteq \mathbb{P}$. Minimizing worst case loss makes sense:

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

Interpreting the Functions 2

- $$S(P, bel) := \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, bel) \ .$$
- It makes much more sense to interpret P as the objective chance function P^* — if you believe in such a thing.
- Then, minimizing score can be interpreted as minimizing inaccuracy; with respect to L.
- However, DM does not know P^*, all she knows is $P^* \in \mathbb{E} \subseteq \mathbb{P}$. Minimizing worst case loss makes sense:

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

Local Scoring Rules

- Let us revisit the expression $P(\omega)L(\omega, bel)$.
- Imagine we want to implement our scoring rule by penalization (weather man).
- In case ω, it would be very strange, if forecaster's loss depended the forecast for $\omega' \neq \omega$.
- Thus, we desire that our scoring rules are *local*, i.e.

$$L(\omega, bel) = L(bel(\omega)).$$

Brier score is *not* local.

Local Scoring Rules

- Let us revisit the expression $P(\omega)L(\omega, bel)$.
- Imagine we want to implement our scoring rule by penalization (weather man).
- In case ω, it would be very strange, if forecaster's loss depended the forecast for $\omega' \neq \omega$.
- Thus, we desire that our scoring rules are *local*, i.e.

$$L(\omega, bel) = L(bel(\omega)).$$

Brier score is *not* local.

Local Scoring Rules

- Let us revisit the expression $P(\omega)L(\omega, bel)$.
- Imagine we want to implement our scoring rule by penalization (weather man).
- In case ω, it would be very strange, if forecaster's loss depended the forecast for $\omega' \neq \omega$.
- Thus, we desire that our scoring rules are *local*, i.e.

$$L(\omega, bel) = L(bel(\omega)).$$

Brier score is *not* local.

Local Scoring Rules

- Let us revisit the expression $P(\omega)L(\omega, bel)$.
- Imagine we want to implement our scoring rule by penalization (weather man).
- In case ω, it would be very strange, if forecaster's loss depended the forecast for $\omega' \neq \omega$.
- Thus, we desire that our scoring rules are *local*, i.e.

$$L(\omega, bel) = L(bel(\omega)).$$

Brier score is *not* local.

Scoring Rules on Worlds

- Minimizing

$$\sup_{P\in\mathbb{E}} S(P, bel) = \sup_{P\in\mathbb{E}} \sum_{\omega\in\Omega} P(\omega)\cdot L(\omega, bel)$$
$$= \sup_{P\in\mathbb{E}} \sum_{\omega\in\Omega} P(\omega)\cdot L(bel(\omega))$$

- still falls well short for justification of probability norm!
- $bel(\omega_1 \cup \omega_2)$ does not appear in $S(P, bel)$!
- Instead, consider minimizing *extended* score

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

Scoring Rules on Worlds

- Minimizing

$$\sup_{P\in\mathbb{E}} S(P, bel) = \sup_{P\in\mathbb{E}} \sum_{\omega\in\Omega} P(\omega)\cdot L(\omega, bel)$$
$$= \sup_{P\in\mathbb{E}} \sum_{\omega\in\Omega} P(\omega)\cdot L(bel(\omega))$$

- still falls well short for justification of probability norm!
- $bel(\omega_1 \cup \omega_2)$ does not appear in $S(P, bel)$!
- Instead, consider minimizing *extended* score

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

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

Scoring Rules on Worlds

- Minimizing

$$\sup_{P\in\mathbb{E}} S(P, bel) = \sup_{P\in\mathbb{E}} \sum_{\omega\in\Omega} P(\omega)\cdot L(\omega, bel)$$
$$= \sup_{P\in\mathbb{E}} \sum_{\omega\in\Omega} P(\omega)\cdot L(bel(\omega))$$

- still falls well short for justification of probability norm!
- $bel(\omega_1 \cup \omega_2)$ does not appear in $S(P, bel)$!
- Instead, consider minimizing *extended* score

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

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

Scoring Rules on Worlds

- Minimizing

$$\begin{aligned}\sup_{P\in\mathbb{E}} S(P, bel) &= \sup_{P\in\mathbb{E}} \sum_{\omega\in\Omega} P(\omega)\cdot L(\omega, bel)\\ &= \sup_{P\in\mathbb{E}} \sum_{\omega\in\Omega} P(\omega)\cdot L(bel(\omega))\end{aligned}$$

- still falls well short for justification of probability norm!
- $bel(\omega_1 \cup \omega_2)$ does not appear in $S(P, bel)$!
- Instead, consider minimizing *extended* score

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

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

Constraining $L(F, bel)$

- 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 $L(F, bel) = -\log(bel(F))$.
- $L(F, bel) = L(bel(F))$ is interpreted as the loss distinct to F, if F obtains.

Constraining $L(F, bel)$

- 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 $L(F, bel) = -\log(bel(F))$.
- $L(F, bel) = L(bel(F))$ is interpreted as the loss distinct to F, if F obtains.

Constraining $L(F, bel)$

- 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 $L(F, bel) = -\log(bel(F))$.
- $L(F, bel) = L(bel(F))$ is interpreted as the loss distinct to F, if F obtains.

Constraining $L(F, bel)$

- 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 $L(F, bel) = -\log(bel(F))$.
- $L(F, bel) = L(bel(F))$ is interpreted as the loss distinct to F, if F obtains.

Outline

Local Scoring Rules for Probability function

A scoring rule S is called *strictly proper*, if and only if $S(P, X)$ is uniquely minimized by $X = P$.

Theorem – Savage 1971

$L(\omega, BEL) = -\lambda \cdot \log(BEL(\omega))$ *is the only strictly-*$BEL \in \mathbb{P}$*-proper scoring rule.* ($\lambda \in \mathbb{R}_{>0}$)

Theorem – Us 2012

There is no strictly-$BEL \in \mathbb{BEL}$*-proper local extended scoring rule.*

Local Scoring Rules for Probability function

A scoring rule S is called *strictly proper*, if and only if $S(P, X)$ is uniquely minimized by $X = P$.

Theorem – Savage 1971

$L(\omega, BEL) = -\lambda \cdot \log(BEL(\omega))$ is the only strictly-$BEL \in \mathbb{P}$-proper scoring rule. ($\lambda \in \mathbb{R}_{>0}$)

Theorem – Us 2012

There is no strictly-$BEL \in \mathbb{BEL}$-proper local extended scoring rule.

Proof of no-locality for Belief Functions

- **Proof:** Assume that $S(P, BEL) = \sum_{F \subseteq \Omega} P(F)L(F, BEL)$ is a strictly proper extended scoring rule.
- Locality implies $S(P, BEL) = \sum_{F \subseteq \Omega} P(F)L(BEL(F))$.
- It is best to adopt $B(F) = x$ where $x \in [0, 1]$ minimizes $L(x)$ – regardless of P! $\blacksquare$

Proof of no-locality for Belief Functions

- **Proof:** Assume that $S(P, BEL) = \sum_{F \subseteq \Omega} P(F) L(F, BEL)$ is a strictly proper extended scoring rule.
- Locality implies $S(P, BEL) = \sum_{F \subseteq \Omega} P(F) L(BEL(F))$.
- It is best to adopt $B(F) = x$ where $x \in [0, 1]$ minimizes $L(x)$ – regardless of P! ■

Proof of no-locality for Belief Functions

- **Proof:** Assume that $S(P, BEL) = \sum_{F \subseteq \Omega} P(F)L(F, BEL)$ is a strictly proper extended scoring rule.
- Locality implies $S(P, BEL) = \sum_{F \subseteq \Omega} P(F)L(BEL(F))$.
- It is best to adopt $B(F) = x$ where $x \in [0, 1]$ minimizes $L(x)$ – regardless of P! ■

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, $S_g^{\log}(P, bel) = \sum_{F \subseteq \Omega} g(F)P(F) \cdot 0 = 0$.
- So, $bel \equiv 1$ minimizes loss! This is BAD.
- Houston, we have a problem!

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, $S_g^{\log}(P, bel) = \sum_{F \subseteq \Omega} g(F)P(F) \cdot 0 = 0$.
- So, $bel \equiv 1$ minimizes loss! This is BAD.
- Houston, we have a problem!

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, $S_g^{\log}(P, bel) = \sum_{F \subseteq \Omega} g(F)P(F) \cdot 0 = 0$.
- So, $bel \equiv 1$ minimizes loss! This is BAD.
- Houston, we have a problem!

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, $S_g^{\log}(P, bel) = \sum_{F \subseteq \Omega} g(F)P(F) \cdot 0 = 0$.
- So, $bel \equiv 1$ minimizes loss! This is BAD.
- Houston, we have a problem!

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, $S_g^{\log}(P, bel) = \sum_{F \subseteq \Omega} g(F)P(F) \cdot 0 = 0$.
- So, $bel \equiv 1$ minimizes loss! This is BAD.
- Houston, we have a problem!

Normalize!

- Let Π be the set of partitions of states of our language.
- For example for $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\pi = \langle(\omega_1, \omega_2, \omega_4), (\omega_3)\rangle$ is a partition.
- 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* $B : \{F \subseteq \Omega\} \longrightarrow [0,1]$ is defined as $B(F) := bel(F)/M$.
- Set of normalized belief functions

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

$$\text{and } \sum_{F \in \pi} B(F) = 1 \text{ for some } \pi\}.$$

Normalize!

- Let Π be the set of partitions of states of our language.
- For example for $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\pi = \langle(\omega_1, \omega_2, \omega_4), (\omega_3)\rangle$ is a partition.
- 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* $B : \{F \subseteq \Omega\} \longrightarrow [0, 1]$ is defined as $B(F) := bel(F)/M$.
- Set of normalized belief functions

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

Normalize!

- Let Π be the set of partitions of states of our language.
- For example for $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\pi = \langle(\omega_1, \omega_2, \omega_4), (\omega_3)\rangle$ is a partition.
- 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* $B : \{F \subseteq \Omega\} \longrightarrow [0,1]$ is defined as $B(F) := bel(F)/M$.
- Set of normalized belief functions

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

$$\text{and } \sum_{F \in \pi} B(F) = 1 \text{ for some } \pi\}.$$

Normalize!

- Let Π be the set of partitions of states of our language.
- For example for $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\pi = \langle(\omega_1, \omega_2, \omega_4), (\omega_3)\rangle$ is a partition.
- 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* $B : \{F \subseteq \Omega\} \longrightarrow [0,1]$ is defined as $B(F) := bel(F)/M$.
- Set of normalized belief functions

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

$$\text{and } \sum_{F \in \pi} B(F) = 1 \text{ for some } \pi\}.$$

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

Normalize!

- Let Π be the set of partitions of states of our language.
- For example for $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\pi = \langle(\omega_1, \omega_2, \omega_4), (\omega_3)\rangle$ is a partition.
- 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* $B : \{F \subseteq \Omega\} \longrightarrow [0, 1]$ is defined as $B(F) := bel(F)/M$.
- Set of normalized belief functions

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

Outline

1. Introduction
   - Objective Bayesianism
2. Scoring Rules
   - Scoring Rules for Probability Functions
   - Scoring Rules for Belief Functions
   - The Probability Norm
   - Logarithmic Loss
3. Belief Functions
   - Locality
   - Normalization
4. Results

g-Score

- For a loss function L and a weighting function $g : \Pi \to \mathbb{R}_{>0}$ define expected g-loss

$$S_g^L(P,B) = \sum_{F \subseteq \Omega} \Big(\sum_{\substack{\pi \in \Pi \\ F \in \pi}} g(\pi) \Big) P(F) L(F,B) \ .$$

- With $L(F,B) = -\log(B(F))$ this becomes

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

- g-entropy is defined as

$$H_g(P) = S_g^{\log}(P,P) \ .$$

g-Score

- For a loss function L and a weighting function $g : \Pi \to \mathbb{R}_{>0}$ define expected g-loss

$$S_g^L(P,B) = \sum_{F \subseteq \Omega} \Big(\sum_{\substack{\pi \in \Pi \\ F \in \pi}} g(\pi) \Big) P(F) L(F,B) \ .$$

- With $L(F,B) = -\log(B(F))$ this becomes

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

- g-entropy is defined as

$$H_g(P) = S_g^{\log}(P,P) \ .$$

g-Score

- For a loss function L and a weighting function $g:\Pi\to\mathbb{R}_{>0}$ define expected g-loss

$$S_g^L(P,B)=\sum_{F\subseteq\Omega}\Big(\sum_{\substack{\pi\in\Pi\\ F\in\pi}}g(\pi)\Big)P(F)L(F,B)\ .$$

- With $L(F,B)=-\log(B(F))$ this becomes

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

- g-entropy is defined as

$$H_g(P)=S_g^{\log}(P,P)\ .$$

Good News Everyone!

Theorem – Norm 1, 2

$S_g^{\log}(P,\cdot)$ is strictly proper on $\mathbb{B}$. For convex $\mathbb{E}\subseteq\mathbb{P}$

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

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

$S_g^{\log}(P,\cdot)$ is strictly proper on $\mathbb{B}$. For convex $\mathbb{E} \subseteq \mathbb{P}$

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

Theorem – Norm 1, 2, 3

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

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

Mixed News

Conjecture – Norm 3?

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

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

Theorem – Norm 3 asterisk

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

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

Mixed News

Conjecture – Norm 3?

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

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

Theorem – Norm 3 asterisk

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

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

Thank You. Questions?

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 $b \in \mathbb{R}_{>0}$.

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 $b \in \mathbb{R}_{>0}$.