Maximum Entropy and Inductive Logic II

Jürgen Landes

Spring School on Inductive Logic

Canterbury, 20.04.2015 - 21.04.2015

Outline

...

- Rational subjective Beliefs
- Finite propositional language L
- Variables $v_1, \ldots, v_n$
- Sentences of L, SL
- No funny business, self-reference, truth predicates, self-fulfilling

...

- Rational subjective Beliefs
- Finite propositional language L
- Variables $v_1, \ldots, v_n$
- Sentences of L, SL
- No funny business, self-reference, truth predicates, self-fulfilling

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

...

- Rational subjective Beliefs
- Finite propositional language L
- Variables $v_1, \ldots, v_n$
- Sentences of L, SL
- No funny business, self-reference, truth predicates, self-fulfilling

...

- Rational subjective Beliefs
- Finite propositional language L
- Variables $v_1, \ldots, v_n$
- Sentences of L, SL
- No funny business, self-reference, truth predicates, self-fulfilling

...

- Rational subjective Beliefs
- Finite propositional language L
- Variables $v_1, \ldots, v_n$
- Sentences of L, SL
- No funny business, self-reference, truth predicates, self-fulfilling

Probabilities

- $P : SL \to [0, 1]$
- Set of probability functions $\mathbb{P}$
- Possible world, states

$$\omega = v_1 \wedge v_2 \wedge \neg v_3 \wedge \ldots \wedge v_{n_1} \wedge \neg v_n$$

- $$P(\varphi) = \sum_{\substack{\omega \in \Omega \\ \omega \models \varphi}} P(\omega).$$

- A probability function $P \in \mathbb{P}$ is uniquely determined by its values on possible worlds, $\langle P(\omega) \; : \; \omega \in \Omega \rangle$.

Probabilities

- $P : SL \to [0, 1]$
- Set of probability functions $\mathbb{P}$
- Possible world, states

$$\omega = v_1 \wedge v_2 \wedge \neg v_3 \wedge \ldots \wedge v_{n_1} \wedge \neg v_n$$

-

$$P(\varphi) = \sum_{\substack{\omega \in \Omega \\ \omega \models \varphi}} P(\omega).$$

- A probability function $P \in \mathbb{P}$ is uniquely determined by its values on possible worlds, $\langle P(\omega) \ : \ \omega \in \Omega \rangle$.

Probabilities

- $P : SL \to [0,1]$
- Set of probability functions $\mathbb{P}$
- Possible world, states

$$\omega = v_1 \wedge v_2 \wedge \neg v_3 \wedge \ldots \wedge v_{n_1} \wedge \neg v_n$$

-

$$P(\varphi) = \sum_{\substack{\omega \in \Omega \\ \omega \models \varphi}} P(\omega).$$

- A probability function $P \in \mathbb{P}$ is uniquely determined by its values on possible worlds, $\langle P(\omega) \ : \ \omega \in \Omega \rangle$.

Probabilities

- $P : SL \to [0,1]$
- Set of probability functions $\mathbb{P}$
- Possible world, states

$$\omega = v_1 \wedge v_2 \wedge \neg v_3 \wedge \ldots \wedge v_{n_1} \wedge \neg v_n$$

- $$P(\varphi) = \sum_{\substack{\omega \in \Omega \\ \omega \models \varphi}} P(\omega).$$
- A probability function $P \in \mathbb{P}$ is uniquely determined by its values on possible worlds, $\langle P(\omega) \ : \ \omega \in \Omega \rangle$.

Probabilities

- $P : SL \rightarrow [0,1]$
- Set of probability functions $\mathbb{P}$
- Possible world, states

$$\omega = v_1 \wedge v_2 \wedge \neg v_3 \wedge \ldots \wedge v_{n_1} \wedge \neg v_n$$

-

$$P(\varphi) = \sum_{\substack{\omega \in \Omega \\ \omega \models \varphi}} P(\omega).$$

- A probability function $P \in \mathbb{P}$ is uniquely determined by its values on possible worlds, $\langle P(\omega) \ : \ \omega \in \Omega \rangle$.

Inference Processes

- Knowledge K leads to $\mathbb{E} \subseteq \mathbb{P}$.
- Formally, an *inference process* is a map from a set of probability functions (here $\mathbb{E}$) to the set of probability functions.
- An inference process is a map (or function)
- Input: $\mathbb{E}$
- Output: P^{+}.

Inference Processes

- Knowledge K leads to $\mathbb{E} \subseteq \mathbb{P}$.
- Formally, an *inference process* is a map from a set of probability functions (here $\mathbb{E}$) to the set of probability functions.
- An inference process is a map (or function)
- Input: $\mathbb{E}$
- Output: P^+.

Inference Processes

- Knowledge K leads to $\mathbb{E} \subseteq \mathbb{P}$.
- Formally, an *inference process* is a map from a set of probability functions (here $\mathbb{E}$) to the set of probability functions.
- An inference process is a map (or function)
- Input: $\mathbb{E}$
- Output: P^+.

Inference Processes

- Knowledge K leads to $\mathbb{E} \subseteq \mathbb{P}$.
- Formally, an *inference process* is a map from a set of probability functions (here $\mathbb{E}$) to the set of probability functions.
- An inference process is a map (or function)
- Input: $\mathbb{E}$
- Output: P^+.

Inference Processes

- Knowledge K leads to $\mathbb{E} \subseteq \mathbb{P}$.
- Formally, an *inference process* is a map from a set of probability functions (here $\mathbb{E}$) to the set of probability functions.
- An inference process is a map (or function)
- Input: $\mathbb{E}$
- Output: P^+.

Obviously right

- Adopt the function, $P^\dagger$, which solve this optimisation problem

$$\text{maximise: } -\sum_{\omega\in\Omega} P(\omega)\log(P(\omega))$$
$$\text{subject to: } P \in \mathbb{E} \ .$$

- Shannon Entropy: $H(P) = -\sum_{\omega\in\Omega} P(\omega)\log(P(\omega))$.
- Maximum Entropy Inference Process (MaxEnt):

$$\{P^{+}\} = \arg\sup_{P\in\mathbb{E}} H(P)$$

Obviously right

- Adopt the function, $P^\dagger$, which solve this optimisation problem

$$\text{maximise: } -\sum_{\omega\in\Omega} P(\omega)\log(P(\omega))$$
$$\text{subject to: } P \in \mathbb{E} \ .$$

- Shannon Entropy: $H(P) = -\sum_{\omega\in\Omega} P(\omega)\log(P(\omega))$.
- Maximum Entropy Inference Process (MaxEnt):

$$\{P^\dagger\} = \arg\sup_{P\in\mathbb{E}} H(P)$$

Obviously right

- Adopt the function, $P^{\dagger}$, which solve this optimisation problem

$$\text{maximise: } -\sum_{\omega\in\Omega} P(\omega)\log(P(\omega))$$
$$\text{subject to: } P \in \mathbb{E}\ .$$

- Shannon Entropy: $H(P) = -\sum_{\omega\in\Omega} P(\omega)\log(P(\omega))$.
- Maximum Entropy Inference Process (MaxEnt):

$$\{P^{+}\} = \arg\sup_{P\in\mathbb{E}} H(P)$$

Desideratum: Internal

- $P^{+} \in \mathbb{E}$
- if $\mathbb{E}$ is non-empty, convex and closed.

Desideratum: Internal

- $P^{+} \in \mathbb{E}$
- if $\mathbb{E}$ is non-empty, convex and closed.

Desideratum: Open-mindedness

- $P^{+}(\omega) > 0$, if there exists a $P \in \mathbb{E}$ such that $P(\omega) > 0$.

Desideratum: Language Invariance

- L' generated by $v_1, v_2, \ldots, v_n, v_{n+1}$
- and the same knowledge and the same patient? For all $\varphi \in SL$

$$P'(\varphi) = P^+(\varphi) \ .$$

- Repeat argument for even larger languages.

Desideratum: Language Invariance

- L' generated by $v_1, v_2, \ldots, v_n, v_{n+1}$
- and the same knowledge and the same patient? For all $\varphi \in SL$

$$P'(\varphi) = P^+(\varphi) \ .$$

- Repeat argument for even larger languages.

Desideratum: Language Invariance

- L' generated by $v_1, v_2, \ldots, v_n, v_{n+1}$
- and the same knowledge and the same patient? For all $\varphi \in SL$

$$P'(\varphi) = P^+(\varphi) \ .$$

- Repeat argument for even larger languages.

Outline

Outline

1. Recap
2. The Classical Derivation
   - Desiderata
   - The Classical Result
3. A Modern Approach
   - Decision Making
   - Justification
4. Universes
5. Objections, your Honor!
   - Bug or Feature?
   - Issues of Language
   - Risqué

Renaming

- Names should not matter.

Irrelevance

- Knowledge entirely irrelevant to the problem in hand can he ignored.
- Languages: L_1 with variables $v_1, \ldots, v_s$, L_2 with variables $v_{s+1}, \ldots, v_n$.
- 2 Bodies of Knowledge: K_1 formulated within L_1, K_2 formulated with in L_2
- For all $\varphi \in L_1$ (problem at hand)

$$IP(\{v_1, \ldots, v_n\}, K_1)(\varphi) = IP(\{v_1, \ldots, v_n\}, K_1 \cup K_2)(\varphi) \ .$$

Irrelevance

- Knowledge entirely irrelevant to the problem in hand can he ignored.
- Languages: L_1 with variables $v_1, \ldots, v_s$, L_2 with variables $v_{s+1}, \ldots, v_n$.
- 2 Bodies of Knowledge: K_1 formulated within L_1, K_2 formulated with in L_2
- For all $\varphi \in L_1$ (problem at hand)

$$IP(\{v_1, \ldots, v_n\}, K_1)(\varphi) = IP(\{v_1, \ldots, v_n\}, K_1 \cup K_2)(\varphi) \quad .$$

Irrelevance

- Knowledge entirely irrelevant to the problem in hand can he ignored.
- Languages: L_1 with variables $v_1, \ldots, v_s$, L_2 with variables $v_{s+1}, \ldots, v_n$.
- 2 Bodies of Knowledge: K_1 formulated within L_1, K_2 formulated with in L_2
- For all $\varphi \in L_1$ (problem at hand)

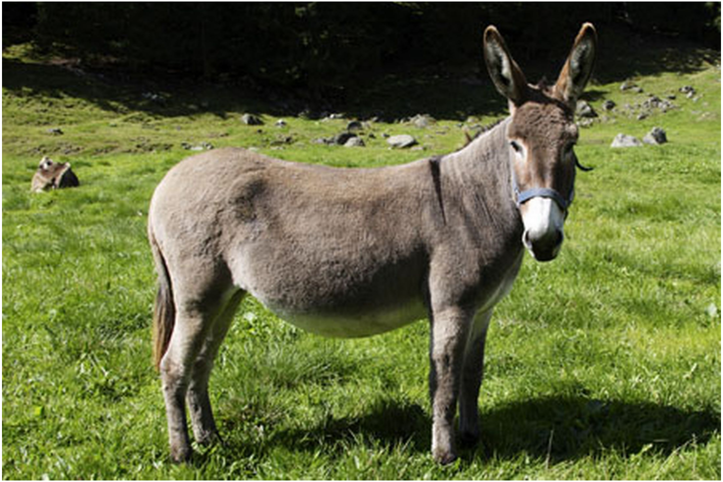
$$IP(\{v_1, \ldots, v_n\}, K_1)(\varphi) = IP(\{v_1, \ldots, v_n\}, K_1 \cup K_2)(\varphi) \quad .$$

Irrelevance

- Knowledge entirely irrelevant to the problem in hand can he ignored.
- Languages: L_1 with variables $v_1, \dots, v_s$, L_2 with variables $v_{s+1}, \dots, v_n$.
- 2 Bodies of Knowledge: K_1 formulated within L_1, K_2 formulated with in L_2
- For all $\varphi \in L_1$ (problem at hand)

$$IP(\{v_1, \dots, v_n\}, K_1)(\varphi) = IP(\{v_1, \dots, v_n\}, K_1 \cup K_2)(\varphi) \ .$$

I-AH



Obstinacy

- Learning something one already beliefs should not make any difference.
- Given two consistent bodies of knowledge, K_1, K_2 (on the same language)
- If $IP(K_1)$ (which is a probability function), is consistent with K_2, then
-
-
- $IP(K_1) = IP(K_1 \cup K_2)$.

Obstinacy

- Learning something one already beliefs should not make any difference.
- Given two consistent bodies of knowledge, K_1, K_2 (on the same language)
- If $IP(K_1)$ (which is a probability function), is consistent with K_2, then
-
-
- $IP(K_1) = IP(K_1 \cup K_2)$.

Obstinacy

- Learning something one already beliefs should not make any difference.
- Given two consistent bodies of knowledge, K_1, K_2 (on the same language)
- If $IP(K_1)$ (which is a probability function), is consistent with K_2, then
- $IP(K_1) = IP(K_1 \cup K_2)$.

Obstinacy

- Learning something one already beliefs should not make any difference.
- Given two consistent bodies of knowledge, K_1, K_2 (on the same language)
- If $IP(K_1)$ (which is a probability function), is consistent with K_2, then
-
-
- $IP(K_1) = IP(K_1 \cup K_2)$.

Obstinacy

- Learning something one already beliefs should not make any difference.
- Given two consistent bodies of knowledge, K_1, K_2 (on the same language)
- If $IP(K_1)$ (which is a probability function), is consistent with K_2, then
-
-
- $IP(K_1) = IP(K_1 \cup K_2)$.

Obstinacy

- Learning something one already beliefs should not make any difference.
- Given two consistent bodies of knowledge, K_1, K_2 (on the same language)
- If $IP(K_1)$ (which is a probability function), is consistent with K_2, then
-
-
- $IP(K_1) = IP(K_1 \cup K_2)$.

Relativisation

- Suppose you know the probability of φ.
- That is, for all $P \in \mathbb{E}$ there exists some $c \in [0, 1]$ such that $P(\varphi) = c$.
- If $\omega^- \models \varphi$, then $IP(\omega^-)$ should not depend on your knowledge about the $\neg\varphi$-worlds.

Relativisation

- Suppose you know the probability of φ.
- That is, for all $P \in \mathbb{E}$ there exists some $c \in [0, 1]$ such that $P(\varphi) = c$.
- If $\omega^- \models \varphi$, then $IP(\omega^-)$ should not depend on your knowledge about the $\neg\varphi$-worlds.

Relativisation

- Suppose you know the probability of φ.
- That is, for all $P \in \mathbb{E}$ there exists some $c \in [0, 1]$ such that $P(\varphi) = c$.
- If $\omega^- \models \varphi$, then $IP(\omega^-)$ should not depend on your knowledge about the $\neg\varphi$-worlds.

Independence

- If K does not contain any information which makes v_1, v_2 conditionally dependent on v_3, then v_1, v_2 should be conditionally independent given v_3.
- If $K = \{P(v_3) = \gamma, P(v_1|v_3) = \frac{\alpha}{\gamma}, P(v_2|v_3) = \frac{\beta}{\gamma}\}$ $(P(\gamma) > 0)$, then

$$IP(K)(v_1 \wedge v_2|v_3) = \frac{\alpha}{\gamma}\frac{\beta}{\gamma} \ .$$

Independence

- If K does not contain any information which makes v_1, v_2 conditionally dependent on v_3, then v_1, v_2 should be conditionally independent given v_3.
- If $K = \{P(v_3) = \gamma, P(v_1|v_3) = \frac{\alpha}{\gamma}, P(v_2|v_3) = \frac{\beta}{\gamma}\}$ $(P(\gamma) > 0)$, then

$$IP(K)(v_1 \wedge v_2|v_3) = \frac{\alpha}{\gamma}\frac{\beta}{\gamma} \quad .$$

Continuity

- If $\mathbb{E}'$ can be obtained from $\mathbb{E}$ by moving or deforming $\mathbb{E}$ a bit, then
- $IP(\mathbb{E}') \approx IP(\mathbb{E})$.
- For all sentences $\varphi \in SL$: $IP(\mathbb{E}')(\varphi) \approx IP(\mathbb{E})(\varphi)$.

Continuity

- If $\mathbb{E}'$ can be obtained from $\mathbb{E}$ by moving or deforming $\mathbb{E}$ a bit, then
- $IP(\mathbb{E}') \approx IP(\mathbb{E})$.
- For all sentences $\varphi \in SL$: $IP(\mathbb{E}')(\varphi) \approx IP(\mathbb{E})(\varphi)$.

Continuity

- If $\mathbb{E}'$ can be obtained from $\mathbb{E}$ by moving or deforming $\mathbb{E}$ a bit, then
- $IP(\mathbb{E}') \approx IP(\mathbb{E})$.
- For all sentences $\varphi \in SL$: $IP(\mathbb{E}')(\varphi) \approx IP(\mathbb{E})(\varphi)$.

Outline

Alena & Jeff

Theorem

If $\mathbb{E}$ is closed, convex and non-empty, IP satisfies Renaming, Irrelevance, Obstinacy, Relativisation, Independence and Continuity, then IP is MaxEnt.
MaxEnt satisfies language invariance and open-mindedness and is internal.

Outline

Outline

Continuity - Bristolean

- One considers the decision problem of setting degrees of belief
- and wonders which beliefs are *best*.
- A *utility function u* is used to measure the goodness / badness / utility of a belief function.
- Determine which beliefs have the "best utility".
- The usual caveats for decision making apply: Non-causal, act-state independence, etc.
- Decision theoretic norm still undetermined.

Continuity - Bristolean

- One considers the decision problem of setting degrees of belief
- and wonders which beliefs are *best*.
- A *utility function u* is used to measure the goodness / badness / utility of a belief function.
- Determine which beliefs have the "best utility".
- The usual caveats for decision making apply: Non-causal, act-state independence, etc.
- Decision theoretic norm still undetermined.

Continuity - Bristolean

- One considers the decision problem of setting degrees of belief
- and wonders which beliefs are *best*.
- A *utility function u* is used to measure the goodness / badness / utility of a belief function.
- Determine which beliefs have the "best utility".
- The usual caveats for decision making apply: Non-causal, act-state independence, etc.
- Decision theoretic norm still undetermined.

Continuity - Bristolean

- One considers the decision problem of setting degrees of belief
- and wonders which beliefs are *best*.
- A *utility function u* is used to measure the goodness / badness / utility of a belief function.
- Determine which beliefs have the "best utility".
- The usual caveats for decision making apply: Non-causal, act-state independence, etc.
- Decision theoretic norm still undetermined.

Continuity - Bristolean

- One considers the decision problem of setting degrees of belief
- and wonders which beliefs are *best*.
- A *utility function u* is used to measure the goodness / badness / utility of a belief function.
- Determine which beliefs have the "best utility".
- The usual caveats for decision making apply: Non-causal, act-state independence, etc.
- Decision theoretic norm still undetermined.

Continuity - Bristolean

- One considers the decision problem of setting degrees of belief
- and wonders which beliefs are *best*.
- A *utility function* u is used to measure the goodness / badness / utility of a belief function.
- Determine which beliefs have the "best utility".
- The usual caveats for decision making apply: Non-causal, act-state independence, etc.
- Decision theoretic norm still undetermined.

Outline

Logarithmic Utility

- If ω is the true world, then the utility is $u(\omega, P^+) = \log(P^+(\omega))$.
- Expected utility for $P \in \mathbb{E}$ is: $\sum_{\omega\in\Omega} P(\omega)\log(P^+(\omega))$.
- If $\mathbb{E}$ is closed, convex and non-empty, maximise worst case expected utility:

$$\arg\sup_{P^+\in\mathbb{P}} \inf_{P\in\mathbb{E}} \sum_{\omega\in\Omega} P(\omega)\log(P^+(\omega)) = \{P^\dagger\} \quad .$$

- The latest rage is to understand u as an accuracy measure.
- Measure closeness to the truth.

Logarithmic Utility

- If ω is the true world, then the utility is $u(\omega, P^+) = \log(P^+(\omega))$.
- Expected utility for $P \in \mathbb{E}$ is: $\sum_{\omega\in\Omega} P(\omega)\log(P^+(\omega))$.
- If $\mathbb{E}$ is closed, convex and non-empty, maximise worst case expected utility:

$$\arg\sup_{P^+\in\mathbb{P}} \inf_{P\in\mathbb{E}} \sum_{\omega\in\Omega} P(\omega)\log(P^+(\omega)) = \{P^\dagger\} \ .$$

- The latest rage is to understand u as an accuracy measure.
- Measure closeness to the truth.

Logarithmic Utility

- If ω is the true world, then the utility is $u(\omega, P^+) = \log(P^+(\omega))$.
- Expected utility for $P \in \mathbb{E}$ is: $\sum_{\omega \in \Omega} P(\omega) \log(P^+(\omega))$.
- If $\mathbb{E}$ is closed, convex and non-empty, maximise worst case expected utility:

$$\arg \sup_{P^+ \in \mathbb{P}} \inf_{P \in \mathbb{E}} \sum_{\omega \in \Omega} P(\omega) \log(P^+(\omega)) = \{P^\dagger\} \ .$$

- The latest rage is to understand u as an accuracy measure.
- Measure closeness to the truth.

Logarithmic Utility

- If ω is the true world, then the utility is $u(\omega, P^+) = \log(P^+(\omega))$.
- Expected utility for $P \in \mathbb{E}$ is: $\sum_{\omega\in\Omega} P(\omega)\log(P^+(\omega))$.
- If $\mathbb{E}$ is closed, convex and non-empty, maximise worst case expected utility:

$$\arg\sup_{P^+\in\mathbb{P}} \inf_{P\in\mathbb{E}} \sum_{\omega\in\Omega} P(\omega)\log(P^+(\omega)) = \{P^\dagger\} \ .$$

- The latest rage is to understand u as an accuracy measure.
- Measure closeness to the truth.

Logarithmic Utility

- If ω is the true world, then the utility is $u(\omega, P^+) = \log(P^+(\omega))$.
- Expected utility for $P \in \mathbb{E}$ is: $\sum_{\omega\in\Omega} P(\omega)\log(P^+(\omega))$.
- If $\mathbb{E}$ is closed, convex and non-empty, maximise worst case expected utility:

$$\arg\sup_{P^+\in\mathbb{P}} \inf_{P\in\mathbb{E}} \sum_{\omega\in\Omega} P(\omega)\log(P^+(\omega)) = \{P^\dagger\} \ .$$

- The latest rage is to understand *u* as an accuracy measure.
- Measure closeness to the truth.

Utility Theory

- If you do have a utility function and a Decision Theoretic Norm,
- then you can show that a particular inference process is optimal with respect to the above.
- Justifications in terms of *common-sense* principles hinge on the common-sensicality of the principles.
- Justifications in terms of utility functions *appear* much more objective.
- However, one has to give a story explaining where the utility function and the DTN come from.

Utility Theory

- If you do have a utility function and a Decision Theoretic Norm,
- then you can show that a particular inference process is optimal with respect to the above.
- Justifications in terms of *common-sense* principles hinge on the common-sensicality of the principles.
- Justifications in terms of utility functions *appear* much more objective.
- However, one has to give a story explaining where the utility function and the DTN come from.

Utility Theory

- If you do have a utility function and a Decision Theoretic Norm,
- then you can show that a particular inference process is optimal with respect to the above.
- Justifications in terms of *common-sense* principles hinge on the common-sensicality of the principles.
- Justifications in terms of utility functions *appear* much more objective.
- However, one has to give a story explaining where the utility function and the DTN come from.

Utility Theory

- If you do have a utility function and a Decision Theoretic Norm,
- then you can show that a particular inference process is optimal with respect to the above.
- Justifications in terms of *common-sense* principles hinge on the common-sensicality of the principles.
- Justifications in terms of utility functions *appear* much more objective.
- However, one has to give a story explaining where the utility function and the DTN come from.

Utility Theory

- If you do have a utility function and a Decision Theoretic Norm,
- then you can show that a particular inference process is optimal with respect to the above.
- Justifications in terms of *common-sense* principles hinge on the common-sensicality of the principles.
- Justifications in terms of utility functions *appear* much more objective.
- However, one has to give a story explaining where the utility function and the DTN come from.

Outline

Ensembles, populations

- Consider an entire population M with $|M|$ members.
- K consists of statements of the form
- 32.2% of patients complain of symptom φ
- $$|\{x \in M \mid x \text{ complains about } \varphi\}| \approx \frac{32.2}{100}|M|$$
- Consider all populations M of fixed large size $|M|$ which satisfy the *above*.
- Then almost all *such* populations M satisfy

$$\frac{|\{x \in M \mid x \text{ complains about } \varphi\}|}{|M|} \approx P^{\dagger}(\varphi) \ .$$

Ensembles, populations

- Consider an entire population M with $|M|$ members.
- K consists of statements of the form
- 32.2% of patients complain of symptom φ
- $$|\{x \in M \mid x \text{ complains about } \varphi\}| \approx \frac{32.2}{100}|M|$$
- Consider all populations M of fixed large size $|M|$ which satisfy the *above*.
- Then almost all *such* populations M satisfy

$$\frac{|\{x \in M \mid x \text{ complains about } \varphi\}|}{|M|} \approx P^{\dagger}(\varphi) .$$

Ensembles, populations

- Consider an entire population M with $|M|$ members.
- K consists of statements of the form
- 32.2% of patients complain of symptom φ
- $$|\{x \in M \mid x \text{ complains about } \varphi\}| \approx \frac{32.2}{100}|M|$$
- Consider all populations M of fixed large size $|M|$ which satisfy the *above*.
- Then almost all *such* populations M satisfy

$$\frac{|\{x \in M \mid x \text{ complains about } \varphi\}|}{|M|} \approx P^{\dagger}(\varphi) \ .$$

Ensembles, populations

- Consider an entire population M with $|M|$ members.
- K consists of statements of the form
- 32.2% of patients complain of symptom φ
- $$|\{x \in M \mid x \text{ complains about } \varphi\}| \approx \frac{32.2}{100}|M|$$
- Consider all populations M of fixed large size $|M|$ which satisfy the *above*.
- Then almost all *such* populations M satisfy

$$\frac{|\{x \in M \mid x \text{ complains about } \varphi\}|}{|M|} \approx P^{\dagger}(\varphi) \ .$$

Ensembles, populations

- Consider an entire population M with $|M|$ members.
- K consists of statements of the form
- 32.2% of patients complain of symptom φ
- $$|\{x \in M \mid x \text{ complains about } \varphi\}| \approx \frac{32.2}{100}|M|$$
- Consider all populations M of fixed large size $|M|$ which satisfy the *above*.
- Then almost all *such* populations M satisfy

$$\frac{|\{x \in M \mid x \text{ complains about } \varphi\}|}{|M|} \approx P^{\dagger}(\varphi) \quad .$$

Ensembles, populations

- Consider an entire population M with $|M|$ members.
- K consists of statements of the form
- 32.2% of patients complain of symptom φ
- $$|\{x \in M \mid x \text{ complains about } \varphi\}| \approx \frac{32.2}{100}|M|$$
- Consider all populations M of fixed large size $|M|$ which satisfy the *above*.
- Then almost all *such* populations M satisfy

$$\frac{|\{x \in M \mid x \text{ complains about } \varphi\}|}{|M|} \approx P^{\dagger}(\varphi) \quad .$$

Alena & Jeff 2 – one last justification

Theorem

For all $\delta > 0$ there exists some natural number M_0 such that for all sentences $\varphi \in SL$ and all fixed sizes of populations $|M| \geq M_0$ the proportion of populations M of fixed size $|M| \geq M_0$ which satisfy

$$\left| \frac{|\{x \in M \mid x \text{ complains about } \varphi\}|}{|M|} - P^{\dagger}(\varphi) \right| < \delta$$

is greater or equal than $1 - \delta$.

Outline

Outline

1. Recap
2. The Classical Derivation
   - Desiderata
   - The Classical Result
3. A Modern Approach
   - Decision Making
   - Justification
4. Universes
5. Objections, your Honor!
   - Bug or Feature?
   - Issues of Language
   - Risqué

Equivocation, come what may

- The least opinionated function is
- $P_{=}(\omega) = \frac{1}{|\Omega|}$, $\{P_{=}\} = \arg\sup_{P\in\mathbb{E}} H(P)$.
- If $P_{=} \in \mathbb{E}$, then $P^{\dagger} = P_{=}$.
- Entropy strictly decreases along the rays originating from $P_{=}$.
- Entropy maximiser different from $P_{=}$ face $P_{=}$.
- If $\mathbb{E} = \{P \in \mathbb{P} \mid 0 \leq P(v_1) \leq 0.5\}$, then $P^{\dagger}(v_1) = 0.5$.
- Can this be right?

Equivocation, come what may

- The least opinionated function is
- $P_=(\omega) = \frac{1}{|\Omega|}$, $\{P_=\} = \arg\sup_{P \in \mathbb{E}} H(P)$.
- If $P_= \in \mathbb{E}$, then $P^\dagger = P_=$.
- Entropy strictly decreases along the rays originating from $P_=$.
- Entropy maximiser different from $P_=$ face $P_=$.
- If $\mathbb{E} = \{P \in \mathbb{P} \mid 0 \leq P(v_1) \leq 0.5\}$, then $P^\dagger(v_1) = 0.5$.
- Can this be right?

Equivocation, come what may

- The least opinionated function is
- $P_{=}(\omega) = \frac{1}{|\Omega|}$, $\{P_{=}\} = \arg\sup_{P \in \mathbb{E}} H(P)$.
- If $P_{=} \in \mathbb{E}$, then $P^{\dagger} = P_{=}$.
- Entropy strictly decreases along the rays originating from $P_{=}$.
- Entropy maximiser different from $P_{=}$ face $P_{=}$.
- If $\mathbb{E} = \{P \in \mathbb{P} \mid 0 \leq P(v_1) \leq 0.5\}$, then $P^{\dagger}(v_1) = 0.5$.
- Can this be right?

Equivocation, come what may

- The least opinionated function is
- $P_=(\omega) = \frac{1}{|\Omega|}$, $\{P_=\} = \arg\sup_{P \in \mathbb{E}} H(P)$.
- If $P_= \in \mathbb{E}$, then $P^\dagger = P_=$.
- Entropy strictly decreases along the rays originating from $P_=$.
- Entropy maximiser different from $P_=$ face $P_=$.
- If $\mathbb{E} = \{P \in \mathbb{P} \mid 0 \leq P(v_1) \leq 0.5\}$, then $P^\dagger(v_1) = 0.5$.
- Can this be right?

Equivocation, come what may

- The least opinionated function is
- $P_=(\omega) = \frac{1}{|\Omega|}$, $\{P_=\} = \arg\sup_{P \in \mathbb{E}} H(P)$.
- If $P_= \in \mathbb{E}$, then $P^\dagger = P_=$.
- Entropy strictly decreases along the rays originating from $P_=$.
- Entropy maximiser different from $P_=$ face $P_=$.
- If $\mathbb{E} = \{P \in \mathbb{P} \mid 0 \leq P(v_1) \leq 0.5\}$, then $P^\dagger(v_1) = 0.5$.
- Can this be right?

Equivocation, come what may

- The least opinionated function is
- $P_=(\omega) = \frac{1}{|\Omega|}$, $\{P_=\} = \arg\sup_{P \in \mathbb{E}} H(P)$.
- If $P_= \in \mathbb{E}$, then $P^\dagger = P_=$.
- Entropy strictly decreases along the rays originating from $P_=$.
- Entropy maximiser different from $P_=$ face $P_=$.
- If $\mathbb{E} = \{P \in \mathbb{P} \mid 0 \leq P(v_1) \leq 0.5\}$, then $P^\dagger(v_1) = 0.5$.
- Can this be right?

Equivocation, come what may

- The least opinionated function is
- $P_=(\omega) = \frac{1}{|\Omega|}$, $\{P_=\} = \arg\sup_{P\in\mathbb{E}} H(P)$.
- If $P_= \in \mathbb{E}$, then $P^\dagger = P_=$.
- Entropy strictly decreases along the rays originating from $P_=$.
- Entropy maximiser different from $P_=$ face $P_=$.
- If $\mathbb{E} = \{P \in \mathbb{P} \mid 0 \le P(v_1) \le 0.5\}$, then $P^\dagger(v_1) = 0.5$.
- Can this be right?

Donkey

- If $\mathbb{E} = \mathbb{P}$, then $P^{\dagger} = P_{=}$.
- But what if your learn this?
- Can this be right?

Donkey

- If $\mathbb{E} = \mathbb{P}$, then $P^{\dagger} = P_{=}$.
- But what if your learn this?
- Can this be right?

Donkey

- If $\mathbb{E} = \mathbb{P}$, then $P^\dagger = P_=$.
- But what if your learn this?
- Can this be right?

Intuition?!?

- Maximising $-\sum_{\omega \in \Omega} P(\omega) \log(P(\omega))$ is sooo counter-intuitive.
- How would you(!) respond to this objection?
- If the aim is to reconstruct "rational" human thinking, then MaxEnt fails; I claim.

Intuition?!?

- Maximising $-\sum_{\omega\in\Omega} P(\omega)\log(P(\omega))$ is sooo counter-intuitive.
- How would you(!) respond to this objection?
- If the aim is to reconstruct "rational" human thinking, then MaxEnt fails; I claim.

Intuition?!?

- Maximising $-\sum_{\omega\in\Omega} P(\omega)\log(P(\omega))$ is sooo counter-intuitive.
- How would you(!) respond to this objection?

- If the aim is to reconstruct “rational” human thinking, then MaxEnt fails; I claim.

Intuition?!?

- Maximising $-\sum_{\omega \in \Omega} P(\omega) \log(P(\omega))$ is sooo counter-intuitive.
- How would you(!) respond to this objection?

- If the aim is to reconstruct "rational" human thinking, then MaxEnt fails; I claim.

Intuition?!?

- Maximising $-\sum_{\omega\in\Omega} P(\omega)\log(P(\omega))$ is sooo counter-intuitive.
- How would you(!) respond to this objection?

- If the aim is to reconstruct “rational” human thinking, then MaxEnt fails; I claim.

Intuition?!?

- Maximising $-\sum_{\omega\in\Omega} P(\omega)\log(P(\omega))$ is sooo counter-intuitive.
- How would you(!) respond to this objection?

- If the aim is to reconstruct “rational” human thinking, then MaxEnt fails; I claim.

Outline

1. Recap
2. The Classical Derivation
   - Desiderata
   - The Classical Result
3. A Modern Approach
   - Decision Making
   - Justification
4. Universes
5. Objections, your Honor!
   - Bug or Feature?
   - Issues of Language
   - Risqué

Dependence

- No knowledge, $\mathbb{E} = \mathbb{P}$.
- Possible worlds: red and blue.
- $P^{\dagger}(red) = \frac{1}{2}$. Okay.
- Possible worlds: red and light blue and dark blue.
- $P^{\dagger}(red) = \frac{1}{3}$. Ohho!
- This phenomenon is called language dependence.
- It is quite common.

Dependence

- No knowledge, $\mathbb{E} = \mathbb{P}$.
- Possible worlds: red and blue.
- $P^{\dagger}(red) = \frac{1}{2}$. Okay.
- Possible worlds: red and light blue and dark blue.
- $P^{\dagger}(red) = \frac{1}{3}$. Ohho!
- This phenomenon is called language dependence.
- It is quite common.

Dependence

- No knowledge, $\mathbb{E} = \mathbb{P}$.
- Possible worlds: red and blue.
- $P^{\dagger}(red) = \frac{1}{2}$. Okay.
- Possible worlds: red and light blue and dark blue.
- $P^{\dagger}(red) = \frac{1}{3}$. Ohho!
- This phenomenon is called language dependence.
- It is quite common.

Dependence

- No knowledge, $\mathbb{E} = \mathbb{P}$.
- Possible worlds: red and blue.
- $P^{\dagger}(red) = \frac{1}{2}$. Okay.
- Possible worlds: red and light blue and dark blue.
- $P^{\dagger}(red) = \frac{1}{3}$. Ohho!
- This phenomenon is called language dependence.
- It is quite common.

Dependence

- No knowledge, $\mathbb{E} = \mathbb{P}$.
- Possible worlds: red and blue.
- $P^{\dagger}(red) = \frac{1}{2}$. Okay.
- Possible worlds: red and light blue and dark blue.
- $P^{\dagger}(red) = \frac{1}{3}$. Ohho!
- This phenomenon is called language dependence.
- It is quite common.

Dependence

- No knowledge, $\mathbb{E} = \mathbb{P}$.
- Possible worlds: red and blue.
- $P^{\dagger}(red) = \frac{1}{2}$. Okay.
- Possible worlds: red and light blue and dark blue.
- $P^{\dagger}(red) = \frac{1}{3}$. Ohho!
- This phenomenon is called language dependence.
- It is quite common.

Dependence

- No knowledge, $\mathbb{E} = \mathbb{P}$.
- Possible worlds: red and blue.
- $P^{\dagger}(red) = \frac{1}{2}$. Okay.
- Possible worlds: red and light blue and dark blue.
- $P^{\dagger}(red) = \frac{1}{3}$. Ohho!
- This phenomenon is called language dependence.
- It is quite common.

Invariance

- $L = \langle v_1, v_2, \ldots, v_n \rangle$, $L' = \langle v_1, v_2, \ldots, v_n, v_{n+1} \rangle$.
- Knowledge only concerns $v_1, \ldots, v_n$.
- $\varphi \in SL$.

$$IP(L)(\varphi) = IP(L')(\varphi)$$

- Centre of Mass is not language invariant! MaxEnt is language invariant.
- Ay, caramba!

Invariance

- $L = \langle v_1, v_2, \ldots, v_n \rangle$, $L' = \langle v_1, v_2, \ldots, v_n, v_{n+1} \rangle$.
- Knowledge only concerns $v_1, \ldots, v_n$.
- $\varphi \in SL$.

$$IP(L)(\varphi) = IP(L')(\varphi)$$

- Centre of Mass is not language invariant! MaxEnt is language invariant.
- Ay, caramba!

Invariance

- $L = \langle v_1, v_2, \ldots, v_n \rangle$, $L' = \langle v_1, v_2, \ldots, v_n, v_{n+1} \rangle$.
- Knowledge only concerns $v_1, \ldots, v_n$.
- $\varphi \in SL$.

$$IP(L)(\varphi) = IP(L')(\varphi)$$

- Centre of Mass is not language invariant! MaxEnt is language invariant.
- Ay, caramba!

Invariance

- $L = \langle v_1, v_2, \ldots, v_n \rangle$, $L' = \langle v_1, v_2, \ldots, v_n, v_{n+1} \rangle$.
- Knowledge only concerns $v_1, \ldots, v_n$.
- $\varphi \in SL$.

$$IP(L)(\varphi) = IP(L')(\varphi)$$

- Centre of Mass is not language invariant! MaxEnt is language invariant.
- Ay, caramba!

Invariance

- $L = \langle v_1, v_2, \ldots, v_n \rangle$, $L' = \langle v_1, v_2, \ldots, v_n, v_{n+1} \rangle$.
- Knowledge only concerns $v_1, \ldots, v_n$.
- $\varphi \in SL$.

$$IP(L)(\varphi) = IP(L')(\varphi)$$

- Centre of Mass is not language invariant! MaxEnt is language invariant.
- Ay, caramba!

Outline

MaxEnt_W

- MaxEnt_W: An agent ought to equivocate (sufficiently) between the basic possibilities that she can express.
- Language now contains the chance functions $\mathbb{P}$.
- Density functions:

$$\mathbb{C}^1_{\mathbb{E}} := \{f_1 : \mathbb{E} \to [0,1] \ : \ \int_{P \in \mathbb{E}} f_1(P) P \, dp \in \mathbb{P}\}.$$

- "Likelihood of a chance function"
- MaxEnt_W: Pick density with greatest entropy

MaxEnt_W

- MaxEnt_W: An agent ought to equivocate (sufficiently) between the basic possibilities that she can express.
- Language now contains the chance functions $\mathbb{P}$.
- Density functions:

$$\mathbb{C}^1_{\mathbb{E}} := \{f_1 : \mathbb{E} \to [0,1] \ : \ \int_{P \in \mathbb{E}} f_1(P)P \, dp \in \mathbb{P}\}.$$

- "Likelihood of a chance function"
- MaxEnt_W: Pick density with greatest entropy

MaxEnt_W

- MaxEnt_W: An agent ought to equivocate (sufficiently) between the basic possibilities that she can express.
- Language now contains the chance functions $\mathbb{P}$.
- Density functions:

$$\mathbb{C}^1_{\mathbb{E}} := \{f_1 : \mathbb{E} \to [0,1] \ : \ \int_{P \in \mathbb{E}} f_1(P) P \, dp \in \mathbb{P}\}.$$

- "Likelihood of a chance function"
- MaxEnt_W: Pick density with greatest entropy

MaxEnt$_W$

- MaxEnt$_W$: An agent ought to equivocate (sufficiently) between the basic possibilities that she can express.
- Language now contains the chance functions $\mathbb{P}$.
- Density functions:

$$\mathbb{C}^1_{\mathbb{E}} := \{f_1 : \mathbb{E} \to [0,1] \ : \ \int_{P \in \mathbb{E}} f_1(P) P \, dp \in \mathbb{P}\}.$$

- "Likelihood of a chance function"
- MaxEnt$_W$: Pick density with greatest entropy

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

MaxEnt$_W$

- MaxEnt$_W$: An agent ought to equivocate (sufficiently) between the basic possibilities that she can express.
- Language now contains the chance functions $\mathbb{P}$.
- Density functions:

$$\mathbb{C}^1_{\mathbb{E}} := \{f_1 : \mathbb{E} \to [0,1] \ : \ \int_{P \in \mathbb{E}} f_1(P) P \, dp \in \mathbb{P}\}.$$

- "Likelihood of a chance function"
- MaxEnt$_W$: Pick density with greatest entropy

Level One

- Entropy: $H(f_1) := -\int_{P\in\mathbb{E}} f_1(P)\cdot\log(f_1(P))dp$.
- Let $f_1^\dagger$ be the density in $\mathbb{C}_{\mathbb{E}}^1$ with maximal entropy.
- Pick a probability function $P^+ : \int_{P\in\mathbb{E}} f_1^\dagger(P)PdP$.
- $P^+ = P_{CoM}$.

Level One

- Entropy: $H(f_1) := -\int_{P\in\mathbb{E}} f_1(P)\cdot\log(f_1(P))dp$.
- Let $f_1^\dagger$ be the density in $\mathbb{C}_{\mathbb{E}}^1$ with maximal entropy.
- Pick a probability function $P^+ : \int_{P\in\mathbb{E}} f_1^\dagger(P)PdP$.
- $P^+ = P_{CoM}$.

Level One

- Entropy: $H(f_1) := -\int_{P\in\mathbb{E}} f_1(P)\cdot\log(f_1(P))dp$.
- Let $f_1^\dagger$ be the density in $\mathbb{C}_{\mathbb{E}}^1$ with maximal entropy.
- Pick a probability function $P^+ : \int_{P\in\mathbb{E}} f_1^\dagger(P)PdP$.
- $P^+ = P_{CoM}$.

Level One

- Entropy: $H(f_1) := -\int_{P\in\mathbb{E}} f_1(P)\cdot\log(f_1(P))dp$.
- Let $f_1^\dagger$ be the density in $\mathbb{C}_{\mathbb{E}}^1$ with maximal entropy.
- Pick a probability function $P^+ : \int_{P\in\mathbb{E}} f_1^\dagger(P)PdP$.
- $P^+ = P_{CoM}$.

Levels

- Density functions for agents with richer languages

$$\mathbb{C}_{\mathbb{E}}^{n+1} := \{f_{n+1} : \mathbb{E} \to [0,1] \ : \int_{f_n \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}(f_n) f_n \, df_n \in \mathbb{C}_{\mathbb{E}}^n\}.$$

- Entropy: $H(f_{n+1}) := -\int_{f_n \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}(f_n) \cdot \log(f_{n+1}(f_n)) df_n$.
- Let $f_{n+1}^{\dagger}$ be the density in $\mathbb{C}_{\mathbb{E}}^{n+1}$ with maximal entropy (it is flat).
- Pick a n density $f_n^{\dagger} : \int_{f \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}^{\dagger}(f_n) f_n df_n$.
- Eventually this determines some $P^{++} \in \mathbb{E}$.
- $P^{++} = P_{CoM}$.

Levels

- Density functions for agents with richer languages

$$\mathbb{C}_{\mathbb{E}}^{n+1} := \{f_{n+1} : \mathbb{E} \to [0,1] \ : \ \int_{f_n \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}(f_n) f_n \, df_n \in \mathbb{C}_{\mathbb{E}}^n\}.$$

- Entropy: $H(f_{n+1}) := -\int_{f_n \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}(f_n) \cdot \log(f_{n+1}(f_n)) df_n$.
- Let $f_{n+1}^{\dagger}$ be the density in $\mathbb{C}_{\mathbb{E}}^{n+1}$ with maximal entropy (it is flat).
- Pick a n density $f_n^{+} : \int_{f \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}^{\dagger}(f_n) f_n df_n$.
- Eventually this determines some $P^{++} \in \mathbb{E}$.
- $P^{++} = P_{CoM}$.

Levels

- Density functions for agents with richer languages

$$\mathbb{C}_{\mathbb{E}}^{n+1} := \{f_{n+1} : \mathbb{E} \to [0,1] \ : \ \int_{f_n \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}(f_n) f_n \, df_n \in \mathbb{C}_{\mathbb{E}}^n\}.$$

- Entropy: $H(f_{n+1}) := -\int_{f_n \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}(f_n) \cdot \log(f_{n+1}(f_n)) df_n$.
- Let $f_{n+1}^{\dagger}$ be the density in $\mathbb{C}_{\mathbb{E}}^{n+1}$ with maximal entropy (it is flat).
- Pick a n density $f_n^{+} : \int_{f \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}^{\dagger}(f_n) f_n df_n$.
- Eventually this determines some $P^{++} \in \mathbb{E}$.
- $P^{++} = P_{CoM}$.

Levels

- Density functions for agents with richer languages

$$\mathbb{C}_{\mathbb{E}}^{n+1} := \{f_{n+1} : \mathbb{E} \to [0,1] \ : \ \int_{f_n \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}(f_n) f_n \, df_n \in \mathbb{C}_{\mathbb{E}}^n\}.$$

- Entropy: $H(f_{n+1}) := -\int_{f_n \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}(f_n) \cdot \log(f_{n+1}(f_n)) df_n$.
- Let $f_{n+1}^{\dagger}$ be the density in $\mathbb{C}_{\mathbb{E}}^{n+1}$ with maximal entropy (it is flat).
- Pick a n density $f_n^{+} : \int_{f \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}^{\dagger}(f_n) f_n df_n$.
- Eventually this determines some $P^{++} \in \mathbb{E}$.
- $P^{++} = P_{CoM}$.

Levels

- Density functions for agents with richer languages

$$\mathbb{C}_{\mathbb{E}}^{n+1} := \{f_{n+1} : \mathbb{E} \to [0,1] \ : \ \int_{f_n \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}(f_n) f_n \, df_n \in \mathbb{C}_{\mathbb{E}}^n\}.$$

- Entropy: $H(f_{n+1}) := -\int_{f_n \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}(f_n) \cdot \log(f_{n+1}(f_n)) df_n$.
- Let $f_{n+1}^{\dagger}$ be the density in $\mathbb{C}_{\mathbb{E}}^{n+1}$ with maximal entropy (it is flat).
- Pick a n density $f_n^{+} : \int_{f \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}^{\dagger}(f_n) f_n df_n$.
- Eventually this determines some $P^{++} \in \mathbb{E}$.
- $P^{++} = P_{CoM}$.

Levels

- Density functions for agents with richer languages

$$\mathbb{C}_{\mathbb{E}}^{n+1} := \{f_{n+1} : \mathbb{E} \to [0,1] \ : \ \int_{f_n \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}(f_n) f_n \, df_n \in \mathbb{C}_{\mathbb{E}}^n\}.$$

- Entropy: $H(f_{n+1}) := -\int_{f_n \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}(f_n) \cdot \log(f_{n+1}(f_n)) df_n$.
- Let $f_{n+1}^{\dagger}$ be the density in $\mathbb{C}_{\mathbb{E}}^{n+1}$ with maximal entropy (it is flat).
- Pick a n density $f_n^{+} : \int_{f \in \mathbb{C}_{\mathbb{E}}^n} f_{n+1}^{\dagger}(f_n) f_n df_n$.
- Eventually this determines some $P^{++} \in \mathbb{E}$.
- $P^{++} = P_{CoM}$.

Density Invariance

- Equivocating over a density level ($\geq$ 1) leads to centre of mass, regardless of the level.
- Density Invariance
- Equivocating over $\mathbb{P}$ gives MaxEnt!
- The level matters.
- Ay, ay, caramba!

Density Invariance

- Equivocating over a density level ($\geq$ 1) leads to centre of mass, regardless of the level.
- Density Invariance
- Equivocating over $\mathbb{P}$ gives MaxEnt!
- The level matters.
- Ay, ay, caramba!

Density Invariance

- Equivocating over a density level ($\geq$ 1) leads to centre of mass, regardless of the level.
- Density Invariance
- Equivocating over $\mathbb{P}$ gives MaxEnt!
- The level matters.
- Ay, ay, caramba!

Density Invariance

- Equivocating over a density level ($\geq$ 1) leads to centre of mass, regardless of the level.
- Density Invariance
- Equivocating over $\mathbb{P}$ gives MaxEnt!
- The level matters.
- Ay, ay, caramba!

Density Invariance

- Equivocating over a density level ($\geq$ 1) leads to centre of mass, regardless of the level.
- Density Invariance
- Equivocating over $\mathbb{P}$ gives MaxEnt!
- The level matters.
- Ay, ay, caramba!

References I

Csiszár, I. (1991).
Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems.
The Annals of Statistics, 19(4):2032–2066.

Grünwald, P. D. and Dawid, A. (2004).
Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory.
Annals of Statistics, 32(4):1367–1433.

Jaynes, E. (1957).
Information Theory and Statistical Mechanics.
Physical Review, 106(4):620–630.

References II

Jaynes, E. T. (2003).
Probability Theory: The Logic of Science.
Cambridge University Press.

Paris, J. B. (1998).
Common Sense and Maximum Entropy.
Synthese, 117:75–93.

Paris, J. B. (2006).
The Uncertain Reasoner's Companion: A Mathematical Perspective, volume 39 of *Cambridge Tracts in Theoretical Computer Science*.
Cambridge University Press, 2 edition.

References III

Paris, J. B. and Vencovská, A. (1989).
On the applicability of maximum entropy to inexact reasoning.
International Journal of Approximate Reasoning, 3(1):1–34.

Paris, J. B. and Vencovská, A. (1990).
A note on the inevitability of maximum entropy.
International Journal of Approximate Reasoning, 4(3):183–223.

References IV

Paris, J. B. and Vencovská, A. (1997).
In Defense of the Maximum Entropy Inference Process.
International Journal of Approximate Reasoning, 17(1):77–103.

Williamson, J. (2010).
In Defence of Objective Bayesianism.
Oxford University Press.

That's it. Thank you! Questions? – Progic Tomorrow

Logarithmic Utility

- Intuitive human sensations tend to be logarithmic functions of the stimulus. – Jaynes
- Savage
- Jon's L1 – L4