# A Triple Uniqueness of the Maximum Entropy Approach

**Jürgen Landes**

Panhellenic Logic Symposium 2022

06.07.2022

# Outline

# General Picture

- Problem: How strongly do we believe things to be true?
- 1st Answer: Depends on the available information.
- 2nd Answer: In interesting cases, beliefs are modelled by probabilities.
- 3rd Answer: In case there's more than one probability function consistent with the evidence further constraining is necessary.

# General Picture

- Problem: How strongly do we believe things to be true?
- 1st Answer: Depends on the available information.
- 2nd Answer: In interesting cases, beliefs are modelled by probabilities.
- 3rd Answer: In case there's more than one probability function consistent with the evidence further constraining is necessary.

# General Picture

- Problem: How strongly do we believe things to be true?
- 1st Answer: Depends on the available information.
- 2nd Answer: In interesting cases, beliefs are modelled by probabilities.
- 3rd Answer: In case there's more than one probability function consistent with the evidence further constraining is necessary.

# General Picture

- Problem: How strongly do we believe things to be true?
- 1st Answer: Depends on the available information.
- 2nd Answer: In interesting cases, beliefs are modelled by probabilities.
- 3rd Answer: In case there's more than one probability function consistent with the evidence further constraining is necessary.

# Inductive Logic

- Given is a fixed first order languages $L$ with constants $t_1, \ldots,$ relation symbols, connectives $(\vee, \wedge, \neg)$ and quantifiers $(\forall, \exists)$ but without function symbols and without $\equiv$.
- Define probability functions over the sentences of $L$.
- Given premisses (evidence): $P(\varphi_1) \in X_1, \ldots, P(\varphi_k) \in X_k$ $(X_i \subset [0, 1])$.
- Question: Which probabilities to attach to a sentence $\psi$ of the language?
- Formally:

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx \psi^? \; . \qquad (1)$$

# Inductive Logic

- Given is a fixed first order languages $L$ with constants $t_1, \ldots,$ relation symbols, connectives $(\vee, \wedge, \neg)$ and quantifiers $(\forall, \exists)$ but without function symbols and without $\equiv$.
- Define probability functions over the sentences of $L$.
- Given premisses (evidence): $P(\varphi_1) \in X_1, \ldots, P(\varphi_k) \in X_k$ $(X_i \subset [0,1])$.
- Question: Which probabilities to attach to a sentence $\psi$ of the language?
- Formally:

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx \psi^? \; . \quad (1)$$

# Inductive Logic

- Given is a fixed first order languages $L$ with constants $t_1, \ldots,$ relation symbols, connectives $(\vee, \wedge, \neg)$ and quantifiers $(\forall, \exists)$ but without function symbols and without $\equiv$.
- Define probability functions over the sentences of $L$.
- Given premisses (evidence): $P(\varphi_1) \in X_1, \ldots, P(\varphi_k) \in X_k$ $(X_i \subset [0, 1])$.
- Question: Which probabilities to attach to a sentence $\psi$ of the language?
- Formally:

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx \psi^{?} \ . \qquad (1)$$

# Inductive Logic

- Given is a fixed first order languages $L$ with constants $t_1, \ldots,$ relation symbols, connectives $(\vee, \wedge, \neg)$ and quantifiers $(\forall, \exists)$ but without function symbols and without $\equiv$.
- Define probability functions over the sentences of $L$.
- Given premisses (evidence): $P(\varphi_1) \in X_1, \ldots, P(\varphi_k) \in X_k$ $(X_i \subset [0, 1])$.
- Question: Which probabilities to attach to a sentence $\psi$ of the language?
- Formally:

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx \psi^? \quad . \tag{1}$$

# Inductive Logic

- Given is a fixed first order languages $L$ with constants $t_1, \ldots,$ relation symbols, connectives $(\vee, \wedge, \neg)$ and quantifiers $(\forall, \exists)$ but without function symbols and without $\equiv$.
- Define probability functions over the sentences of $L$.
- Given premisses (evidence): $P(\varphi_1) \in X_1, \ldots, P(\varphi_k) \in X_k$ $(X_i \subset [0,1])$.
- Question: Which probabilities to attach to a sentence $\psi$ of the language?
- Formally:

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx\!\!\!\mid\; \psi^{?} \quad . \tag{1}$$

# Outline

# Outline

# General Picture: Maximum Entropy

- 3rd Answer: In case there's more than one probability function consistent with the evidence further constraining is necessary.
- **MaxEnt**: Pick the probability function that maximises Shannon Entropy

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

- MaxEnt is well-understood on finite domains.
- On infinite domains, such as $L$, MaxEnt cannot directly be applied.

# General Picture: Maximum Entropy

- 3rd Answer: In case there's more than one probability function consistent with the evidence further constraining is necessary.
- **MaxEnt**: Pick the probability function that maximises Shannon Entropy

$$H(P) := -\sum_{\omega \in \Omega} P(\omega) \log(P(\omega)) \ .$$

- MaxEnt is well-understood on finite domains.
- On infinite domains, such as $L$, MaxEnt cannot directly be applied.

# General Picture: Maximum Entropy

- 3rd Answer: In case there's more than one probability function consistent with the evidence further constraining is necessary.
- **MaxEnt**: Pick the probability function that maximises Shannon Entropy

$$H(P) := -\sum_{\omega\in\Omega} P(\omega)\log(P(\omega)) \ .$$

- MaxEnt is well-understood on finite domains.
- On infinite domains, such as $L$, MaxEnt cannot directly be applied.

# General Picture: Maximum Entropy

- 3rd Answer: In case there's more than one probability function consistent with the evidence further constraining is necessary.
- **MaxEnt**: Pick the probability function that maximises Shannon Entropy

$$H(P) := -\sum_{\omega \in \Omega} P(\omega) \log(P(\omega)) \ .$$

- MaxEnt is well-understood on finite domains.
- On infinite domains, such as $L$, MaxEnt cannot directly be applied.

# Definition of the Maximum Entropy Approach

- $n$-entropy: $H_n(P) = -\sum_{\omega_n \in \Omega_n} P(\omega_n) \cdot \log(P(\omega_n))$, "Entropy of the states expressible in $L_n$".
- $P$ has greater entropy than $Q$, if and only if there is some natural number $N$ such that for all $n \geq N$ it holds that $H_n(P) > H_n(Q)$, denoted by $P \succ Q$. "Eventually, $P$ has greater $n$-entropy than $Q$.
- Set of probability functions consistent with the evidence, $\mathbb{E}$.
- Set of maximum entropy functions

$$\text{maxent}_{\succ}\mathbb{E} := \{P \in \mathbb{E} : \text{ there is no } Q \in \mathbb{E} \setminus \{P\} \text{ with } Q \succ P\}$$

Probability functions in $\mathbb{E}$ which are "not dominated in entropy".

# Definition of the Maximum Entropy Approach

- $n$-entropy: $H_n(P) = -\sum_{\omega_n \in \Omega_n} P(\omega_n) \cdot \log(P(\omega_n))$, "Entropy of the states expressible in $L_n$".
- $P$ has greater entropy than $Q$, if and only if there is some natural number $N$ such that for all $n \geq N$ it holds that $H_n(P) > H_n(Q)$, denoted by $P \succ Q$. "Eventually, $P$ has greater $n$-entropy than $Q$.
- Set of probability functions consistent with the evidence, $\mathbb{E}$.
- Set of maximum entropy functions

$$\text{maxent}_{\succ}\,\mathbb{E} := \{P \in \mathbb{E} : \text{ there is no } Q \in \mathbb{E} \setminus \{P\} \text{ with } Q \succ P\}$$

Probability functions in $\mathbb{E}$ which are "not dominated in entropy".

# Definition of the Maximum Entropy Approach

- $n$-entropy: $H_n(P) = -\sum_{\omega_n \in \Omega_n} P(\omega_n) \cdot \log(P(\omega_n))$, "Entropy of the states expressible in $L_n$".
- $P$ has greater entropy than $Q$, if and only if there is some natural number $N$ such that for all $n \geq N$ it holds that $H_n(P) > H_n(Q)$, denoted by $P \succ Q$. "Eventually, $P$ has greater $n$-entropy than $Q$.
- Set of probability functions consistent with the evidence, $\mathbb{E}$.
- Set of maximum entropy functions

$$\text{maxent}_{\succ}\,\mathbb{E} := \{P \in \mathbb{E} : \text{ there is no } Q \in \mathbb{E} \setminus \{P\} \text{ with } Q \succ P\}$$

Probability functions in $\mathbb{E}$ which are "not dominated in entropy".

# Definition of the Maximum Entropy Approach

- *n*-entropy: $H_n(P) = -\sum_{\omega_n \in \Omega_n} P(\omega_n) \cdot \log(P(\omega_n))$, "Entropy of the states expressible in $L_n$".
- $P$ has greater entropy than $Q$, if and only if there is some natural number $N$ such that for all $n \geq N$ it holds that $H_n(P) > H_n(Q)$, denoted by $P \succ Q$. "Eventually, $P$ has greater *n*-entropy than $Q$.
- Set of probability functions consistent with the evidence, $\mathbb{E}$.
- Set of maximum entropy functions

  $$\text{maxent}_{\succ} \mathbb{E} := \{P \in \mathbb{E} : \text{ there is no } Q \in \mathbb{E} \setminus \{P\} \text{ with } Q \succ P\}$$

  Probability functions in $\mathbb{E}$ which are "not dominated in entropy".

# Outline

# Alternative Comparisons

- $P$ has greater entropy than $Q$, if and only if there is some natural number $N$ such that for all $n \geq N$ it holds that
  1. $P \succ Q$: $H_n(P) > H_n(Q)$,
  2. $P)Q$: $H_n(P) \geq H_n(Q)$,
  3. $P]Q$: $H_n(P) \geq H_n(Q)$ and there are infinitely many $n$ such that $H_n(P) > H_n(Q)$.
- All three look reasonable.

# Alternative Comparisons

- $P$ has greater entropy than $Q$, if and only if there is some natural number $N$ such that for all $n \geq N$ it holds that
  1. $P \succ Q$: $H_n(P) > H_n(Q)$,
  2. $P)Q$: $H_n(P) \geq H_n(Q)$,
  3. $P]Q$: $H_n(P) \geq H_n(Q)$ and there are infinitely many $n$ such that $H_n(P) > H_n(Q)$.
- All three look reasonable.

# Alternative Comparisons

- $P$ has greater entropy than $Q$, if and only if there is some natural number $N$ such that for all $n \geq N$ it holds that
  1. $P \succ Q$: $H_n(P) > H_n(Q)$,
  2. $P)Q$: $H_n(P) \geq H_n(Q)$,
  3. $P]Q$: $H_n(P) \geq H_n(Q)$ and there are infinitely many $n$ such that $H_n(P) > H_n(Q)$.
- All three look reasonable.

# Alternative Comparisons

- $P$ has greater entropy than $Q$, if and only if there is some natural number $N$ such that for all $n \geq N$ it holds that
  1. $P \succ Q$: $H_n(P) > H_n(Q)$,
  2. $P ) Q$: $H_n(P) \geq H_n(Q)$,
  3. $P ] Q$: $H_n(P) \geq H_n(Q)$ and there are infinitely many $n$ such that $H_n(P) > H_n(Q)$.
- All three look reasonable.

# Alternative Comparisons

- $P$ has greater entropy than $Q$, if and only if there is some natural number $N$ such that for all $n \geq N$ it holds that
  1. $P \succ Q$: $H_n(P) > H_n(Q)$,
  2. $P)Q$: $H_n(P) \geq H_n(Q)$,
  3. $P]Q$: $H_n(P) \geq H_n(Q)$ and there are infinitely many $n$ such that $H_n(P) > H_n(Q)$.
- All three look reasonable.

# Alternative Maximum Entropies

- Probability functions in $\mathbb{E}$ which are "not dominated in entropy":

$$\text{maxent}_{\succ}\mathbb{E} :=\{P\in\mathbb{E}: \text{ there is no } Q\in\mathbb{E}\setminus\{P\} \text{ with } Q\succ P\} \ ,$$
$$\text{maxent}_{)}\mathbb{E} :=\{P\in\mathbb{E}: \text{ there is no } Q\in\mathbb{E}\setminus\{P\} \text{ with } Q)P\} \ ,$$
$$\text{maxent}_{]}\mathbb{E} :=\{P\in\mathbb{E}: \text{ there is no } Q\in\mathbb{E}\setminus\{P\} \text{ with } Q]P\} \ .$$

- All three look reasonable.

# Alternative Maximum Entropies

- Probability functions in $\mathbb{E}$ which are "not dominated in entropy":

$$\text{maxent}_{\succ}\,\mathbb{E} :=\{P \in \mathbb{E} :\ \text{there is no } Q \in \mathbb{E} \setminus \{P\} \text{ with } Q \succ P\}\ ,$$
$$\text{maxent}_{)}\,\mathbb{E} :=\{P \in \mathbb{E} :\ \text{there is no } Q \in \mathbb{E} \setminus \{P\} \text{ with } Q ) P\}\ ,$$
$$\text{maxent}_{]}\,\mathbb{E} :=\{P \in \mathbb{E} :\ \text{there is no } Q \in \mathbb{E} \setminus \{P\} \text{ with } Q ] P\}\ .$$

- All three look reasonable.

# Outline

# Triple Uniqueness

## Theorem (Triple Uniqueness)

*If* $\mathbb{E} \neq \emptyset$ *is convex and at least one of* $\text{maxent}_{)}\,\mathbb{E}$, $\text{maxent}_{]}\,\mathbb{E}$ *or* $\text{maxent}_{\succ}\,\mathbb{E}$ *is a singleton, then*

$$\text{maxent}\,\mathbb{E}_{)} = \text{maxent}_{]}\,\mathbb{E} = \text{maxent}_{\succ}\,\mathbb{E} \quad .$$

$\mathbb{E} \neq \emptyset$ *is convex here (, if the* $X_i$ *are non-empty intervals).*

# Triple Uniqueness – Single Premiss

## Theorem (With Soroush Rafiee Rad and Jon Williamson, [9])

*For* $\varphi^c$ *with* $c \in (0, 1]\}$ *and* $P_=(\varphi) \in (0, 1)$ *it holds that*

$$\text{maxent}\,\mathbb{E} = \{c \cdot P_=(\cdot|\varphi) + (1 - c) \cdot P_=(\cdot|\neg\varphi)\} \ .$$

*MaxEnt updating is Jeffrey updating.*
*Corollary: For a single certain premiss with non-zero prior probability, MaxEnt updating is Bayesian updating.*

*In a large range of cases, there's just one MaxEnt updating to* **rule them all**.

# Triple Uniqueness – Multiple Premisses

**Theorem (With Soroush Rafiee Rad and Jon Williamson, [10])**

*If the premisses $\varphi_1^{c_1}, \ldots, \varphi_k^{c_k}$ are finitely satisfiable, then $\mathcal{H}(P^\dagger) > \mathcal{H}(P)$ for all other probability functions in $P \in \mathbb{E} \setminus \{P^\dagger\}$, where $P^\dagger$ is obtained by equivocating beyond the "final solution".*

*In words: $P^\dagger$ is the unique maximal entropy function.*

*In a large range of cases, there's just one MaxEnt updating to* **rule them all**.

# Outline

# $\Sigma_2$

- For the premiss $\varphi = \exists x \forall y Rxy$ all the maxent $\mathbb{E}$ are empty.
- Idea: modify the notion of comparative entropy.
- Let's try $P \} Q$, if and only if $H_n(P) > H_n(Q)$ for all $n \in \mathbb{N}$.
- Maximum entropy functions: $\text{maxent}_{\}} \mathbb{E}$.
- $\text{maxent}_{\}} \mathbb{E} = \mathbb{E}$ – Nonsense!

# $\Sigma_2$

- For the premiss $\varphi = \exists x \forall y Rxy$ all the maxent $\mathbb{E}$ are empty.
- Idea: modify the notion of comparative entropy.
- Let's try $P \succ Q$, if and only if $H_n(P) > H_n(Q)$ for all $n \in \mathbb{N}$.
- Maximum entropy functions: $\text{maxent}_{\succ} \mathbb{E}$.
- $\text{maxent}_{\succ} \mathbb{E} = \mathbb{E}$ – Nonsense!

# $\Sigma_2$

- For the premiss $\varphi = \exists x \forall y Rxy$ all the maxent $\mathbb{E}$ are empty.
- Idea: modify the notion of comparative entropy.
- Let's try $P \} Q$, if and only if $H_n(P) > H_n(Q)$ for all $n \in \mathbb{N}$.
- Maximum entropy functions: maxent$_{\}}$ $\mathbb{E}$.
- maxent$_{\}}$ $\mathbb{E} = \mathbb{E}$ – Nonsense!

# $\Sigma_2$

- For the premiss $\varphi = \exists x \forall y Rxy$ all the maxent $\mathbb{E}$ are empty.
- Idea: modify the notion of comparative entropy.
- Let's try $P \} Q$, if and only if $H_n(P) > H_n(Q)$ for all $n \in \mathbb{N}$.
- Maximum entropy functions: $\text{maxent}_{\}} \mathbb{E}$.
- $\text{maxent}_{\}} \mathbb{E} = \mathbb{E}$ – Nonsense!

# $\Sigma_2$

- For the premiss $\varphi = \exists x \forall y Rxy$ all the maxent $\mathbb{E}$ are empty.
- Idea: modify the notion of comparative entropy.
- Let's try $P \} Q$, if and only if $H_n(P) > H_n(Q)$ for all $n \in \mathbb{N}$.
- Maximum entropy functions: $\text{maxent}_{\}} \mathbb{E}$.
- $\text{maxent}_{\}} \mathbb{E} = \mathbb{E}$ – Nonsense!

# Outline

# Some Lessons

- Uniqueness is noteworthy in light of a string of results suggesting and comparing different notions of entropy (maximisation), which lead to different maximum entropy functions [1, 2, 3, 4, 5, 6, 8, 11].
- Updating with evidence of zero prior probability remains hard. Nevertheless, what is to be done?
- We can't expect much from attempting to draw on finite settings where conditionalisation is not defined.
- Multiple uncertain premisses seem to pose really hard questions.

# Some Lessons

- Uniqueness is noteworthy in light of a string of results suggesting and comparing different notions of entropy (maximisation), which lead to different maximum entropy functions [1, 2, 3, 4, 5, 6, 8, 11].
- Updating with evidence of zero prior probability remains hard. Nevertheless, what is to be done?
- We can't expect much from attempting to draw on finite settings where conditionalisation is not defined.
- Multiple uncertain premisses seem to pose really hard questions.

# Some Lessons

- Uniqueness is noteworthy in light of a string of results suggesting and comparing different notions of entropy (maximisation), which lead to different maximum entropy functions [1, 2, 3, 4, 5, 6, 8, 11].
- Updating with evidence of zero prior probability remains hard. Nevertheless, what is to be done?
- We can't expect much from attempting to draw on finite settings where conditionalisation is not defined.
- Multiple uncertain premisses seem to pose really hard questions.

# Some Lessons

- Uniqueness is noteworthy in light of a string of results suggesting and comparing different notions of entropy (maximisation), which lead to different maximum entropy functions [1, 2, 3, 4, 5, 6, 8, 11].
- Updating with evidence of zero prior probability remains hard. Nevertheless, what is to be done?
- We can't expect much from attempting to draw on finite settings where conditionalisation is not defined.
- Multiple uncertain premisses seem to pose really hard questions.

# Join us!

- Computational complexity [12].
- Sound and/or complete rules of proof [7].
- Multiple Premisses [10].
- Inconsistent evidence <your name here >.
- ...

# Join us!

- Computational complexity [12].
- Sound and/or complete rules of proof [7].
- Multiple Premisses [10].
- Inconsistent evidence <your name here >.
- ...

# Join us!

- Computational complexity [12].
- Sound and/or complete rules of proof [7].
- Multiple Premisses [10].
- Inconsistent evidence <your name here >.
- ...

# Join us!

- Computational complexity [12].
- Sound and/or complete rules of proof [7].
- Multiple Premisses [10].
- Inconsistent evidence $<$your name here $>$.
- ...

# Join us!

- Computational complexity [12].
- Sound and/or complete rules of proof [7].
- Multiple Premisses [10].
- Inconsistent evidence $<$your name here $>$.
- ...

# Q & A

# References I

Vincenzo Crupi, Jonathan Nelson, Björn Meder, Gustavo Cevolani, and Katya Tentori.
Generalized information theory meets human cognition: Introducing a unified framework to model uncertainty and information search.
*Cognitive Science*, 42:1410–1456, 2018.

Imre Csiszár.
Axiomatic Characterizations of Information Measures.
*Entropy*, 10(3):261–273, 2008.

# References II

Huizi Cui, Qing Liu, Jianfeng Zhang, and Bingyi Kang.
An Improved Deng Entropy and Its Application in Pattern Recognition.
*IEEE Access*, 7:18284–18292, 2019.

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

# References III

R. Hanel, S. Thurner, and M. Gell-Mann.
Generalized entropies and the transformation group of superstatistics.
*Proceedings of the National Academy of Sciences*, 108(16):6390–6394, 2011.

Jürgen Landes.
Probabilism, Entropies and Strictly Proper Scoring Rules.
*International Journal of Approximate Reasoning*, 63:1–21, 2015.

# References IV

Jürgen Landes.
Rules of Proof for Maximal Entropy Inference.
*International Journal of Approximate Reasoning*, 2022.
38 pages, submitted.

Jürgen Landes and George Masterton.
Invariant Equivocation.
*Erkenntnis*, 82:141–167, 2017.

Jürgen Landes, Soroush Rafiee Rad, and Jon Williamson.
Determining maximal entropy functions for objective Bayesian inductive logic.
*Journal of Philosophical Logic*, 2022.
52 pages, accepted.

# References V

Jürgen Landes, Soroush Rafiee Rad, and Jon Williamson.
Multiple premisses in objective bayesian inductive logic.
*Artificial Intelligence*, 2023.
in preparation.

Jürgen Landes and Jon Williamson.
Objective Bayesianism and the maximum entropy principle.
*Entropy*, 15(9):3528–3591, 2013.

# References VI

Jürgen Landes and Jon Williamson.
Objective Bayesian Nets for Integrating Consistent Datasets.
*Journal of Artificial Intelligence Research*, 74:393–458, 2022.