Determining maximal entropy functions for objective Bayesian inductive logic

Juergen Landes, Soroush Rafiee Rad, Jon Williamson

Progic, Munich
1st September 2021

1 Objective Bayesian inductive logic

Inductive entailment relationship (Haenni et al., 2011):

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx\!\!\!\models \psi^Y.$$

where

$\varphi_1, \ldots, \varphi_k, \psi$ are sentences of a logical language $\mathcal{L}$

$X_1, \ldots, X_k, Y$ are probabilities or sets of probabilities.

Objective Bayesian inductive logic:

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \overset{\circ}{\approx\!\!\!\models} \psi^Y$$

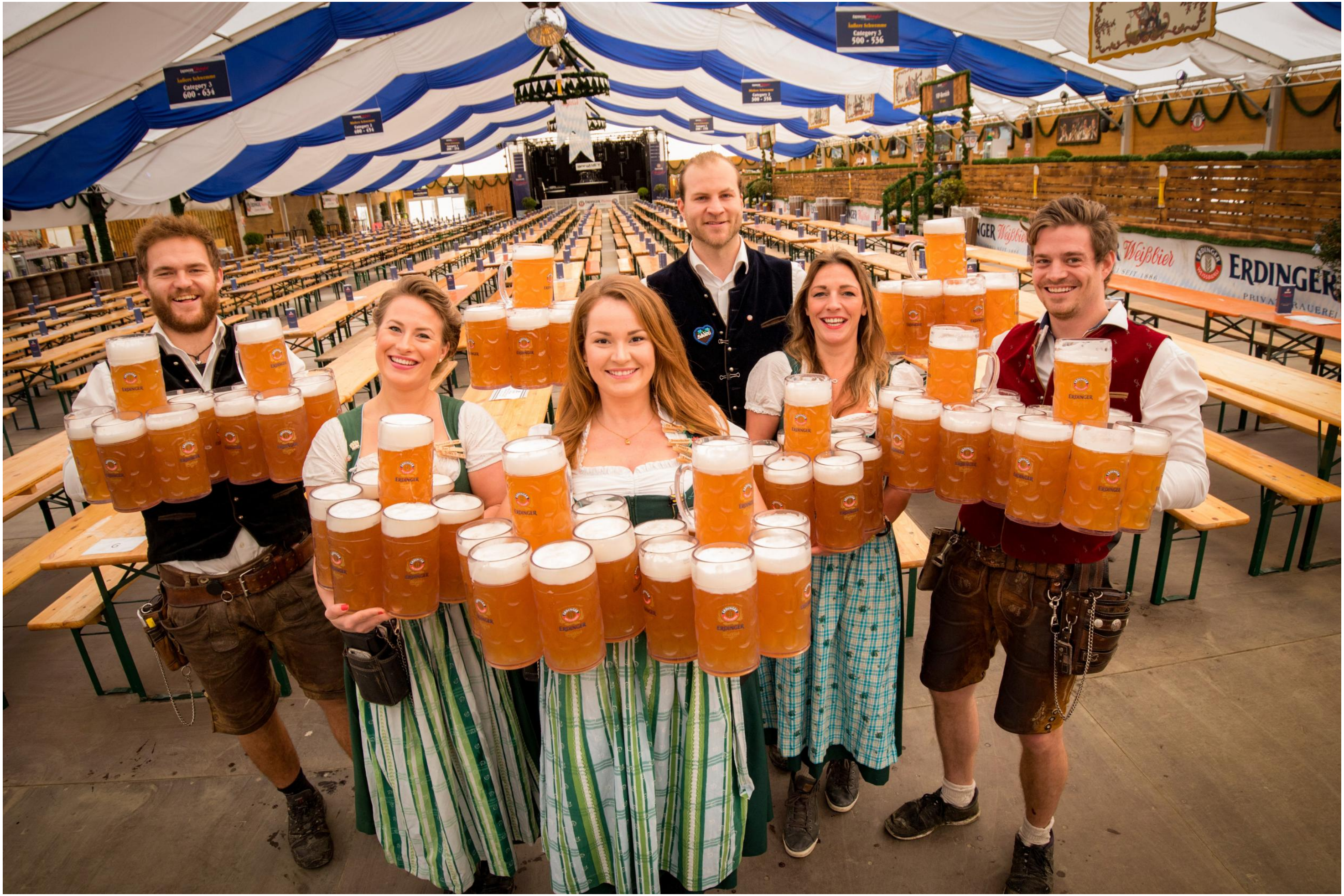
iff $P^\dagger(\psi) \in Y$ for every probability function $P^\dagger$, from all those that satisfy the premisses, that has maximal entropy.

Suppose $\mathcal{L}$ is a propositional language with atomic sentences $a_1, \ldots, a_n$:

$\Omega_n \stackrel{\text{df}}{=} \{\pm a_1 \wedge \cdots \wedge \pm a_n\}.$

$H_n(P) \stackrel{\text{df}}{=} -\sum_{\omega \in \Omega} P(\omega) \log P(\omega).$

Find the function $P^\dagger$, from those that satisfy $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$, with maximum entropy.


Category 3
600 - 634
Category 3
500 - 536
Weißbier
ERDINGER

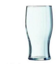
Answer the following question:

$$g \rightarrow h, h \wedge l \wedge \neg g^{1/16} \mid\!\approx g^{?}$$

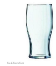
Think of the total probability as contained in a jug and to be distributed evenly amongst people at a table, who represent possible outcomes, given constraints:

hlg	$hl\neg g$	$h\neg lg$	$h\neg l\neg g$	$\neg hlg$	$\neg hl\neg g$	$\neg h\neg lg$	$\neg h\neg l\neg g$

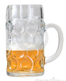
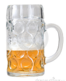
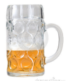
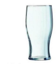
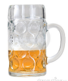
First constraint: $g \rightarrow h$

				 0		 0	
hlg	$hl\neg g$	$h\neg lg$	$h\neg l\neg g$	$\neg hlg$	$\neg hl\neg g$	$\neg h\neg lg$	$\neg h\neg l\neg g$

Second constraint: $P(hl\neg g) = 1/16$

							
	1/16			0		0	
hlg	$hl\neg g$	$h\neg lg$	$h\neg l\neg g$	$\neg hlg$	$\neg hl\neg g$	$\neg h\neg lg$	$\neg h\neg l\neg g$

Then share evenly:

							
3/16	1/16	3/16	3/16	0	3/16	0	3/16
hlg	$hl\neg g$	$h\neg lg$	$h\neg l\neg g$	$\neg hlg$	$\neg hl\neg g$	$\neg h\neg lg$	$\neg h\neg l\neg g$

A truth table can be used instead:

P	h	l	g	$g \to h$	$h \wedge l \wedge \neg g$	g
3/16	T	T	T	T	F	T
1/16	T	T	F	T	T	F
3/16	T	F	T	T	F	T
3/16	T	F	F	T	F	F
0	F	T	T	F	F	T
3/16	F	T	F	T	F	F
0	F	F	T	F	F	T
3/16	F	F	F	T	F	F

$$\begin{aligned} P(g) &= P(hlg) + P(h\neg lg) + P(\neg hlg) + P(\neg h\neg lg) \\ &= 3/16 + 3/16 + 0 + 0 \\ &= 6/16 \\ &= 3/8 \end{aligned}$$

So

$$g \to h, h \wedge l \wedge \neg g^{1/16} \mathrel{|\!\!\approx} g^{3/8}.$$

What if $\mathcal{L}$ is a first-order predicate language?

Finitely many relation symbols $U_1, \ldots, U_l$,

Countably many constant symbols $t_1, t_2, \ldots$

Atomic sentences $a_1, a_2, \ldots$ ordered such that those involving only $t_1, \ldots, t_n$ occur before those involving t_{n+1}, for each n,

Finite sublanguages $\mathcal{L}_n$ involve only $t_1, \ldots, t_n$.

$\Omega_n \stackrel{\text{df}}{=} \{\pm a_1 \wedge \cdots \wedge \pm a_{r_n}\}$, the states of $\mathcal{L}_n$.

$\mathbb{E} \stackrel{\text{df}}{=} \{P \in \mathbb{P} : P(\varphi_1) \in X_1, \ldots, P(\varphi_k) \in X_k\}$.

$\text{maxent}\,\mathbb{E}$ is the set of functions in $\mathbb{E}$ that are not dominated in entropy:

$$\text{maxent}\,\mathbb{E} \stackrel{\text{df}}{=} \{P \in \mathbb{E} : \text{there is no } Q \in \mathbb{E}, N \in \mathbb{N} \text{ such that } H_n(Q) > H_n(P) \text{ for } n \geq N\}.$$

But how do we find these maximal entropy probability functions?

2 Entropy limit points

In certain circumstances $\text{maxent}\,\mathbb{E}$ is a limit of n-entropy maximisers.

Suppose $X_1, \ldots, X_k$ are convex and the premisses are satisfiable.

Consider:

$$\mathbb{H}_n = \{P \in \mathbb{E} : H_n(P) \text{ is maximised }\}.$$

$P \in \mathbb{P}$ is an **entropy limit point** of $\mathbb{E}$ if for each n there is some $Q_n \in \mathbb{H}_n$ such that $|H_n(Q_n) - H_n(P)| \longrightarrow 0$ as $n \longrightarrow \infty$.

This property enables us to characterise $\text{maxent}\,\mathbb{E}$ more constructively:

Theorem 1. *If $\mathbb{E}$ contains an entropy limit point P then*

$$\text{maxent}\,\mathbb{E} = \{P\}.$$

Note that there can be at most one entropy limit point P.

Example 2. Suppose we have a single premiss $\forall x U x^c$ where $\mathcal{L}$ has a single unary predicate U. The following probability function is an entropy limit point:

$$P(\omega_n) = \begin{cases} c + \frac{1-c}{2^n} & : \quad \omega_n = Ut_1 \wedge \cdots \wedge Ut_n \\ \frac{1-c}{2^n} & : \quad \omega_n \models \neg(Ut_1 \wedge \cdots \wedge Ut_n) \end{cases} .$$

$P \in \mathbb{E}$ because $P(\forall x U x) = \lim_{n\to\infty} P(\theta_n) = c$. Hence by Theorem 1, $\mathrm{maxent}\,\mathbb{E} = \{P\}$.

Example 3. Consider a single categorical premiss $U_1 t_1 \vee \exists x \forall y U_2 xy$. In this case $\mathbb{H}_n = \{P_=\}$ for all n. Thus the equivocator function is the unique entropy limit point of $\mathbb{E}$. However, the equivocator function is not in $\mathbb{E}$, so it cannot be the maximal entropy function.

3 Conditionalisation and entropy limit points

Special case:

> The premisses are categorical sentences $\varphi_1, \ldots \varphi_k$ of $\mathcal{L}$.

Let φ abbreviate $\varphi_1 \wedge \cdots \wedge \varphi_k$.

We will consider $\mathbb{E} = \mathbb{E}_\varphi \stackrel{\text{df}}{=} \{P \in \mathbb{P} : P(\varphi) = 1\}$.

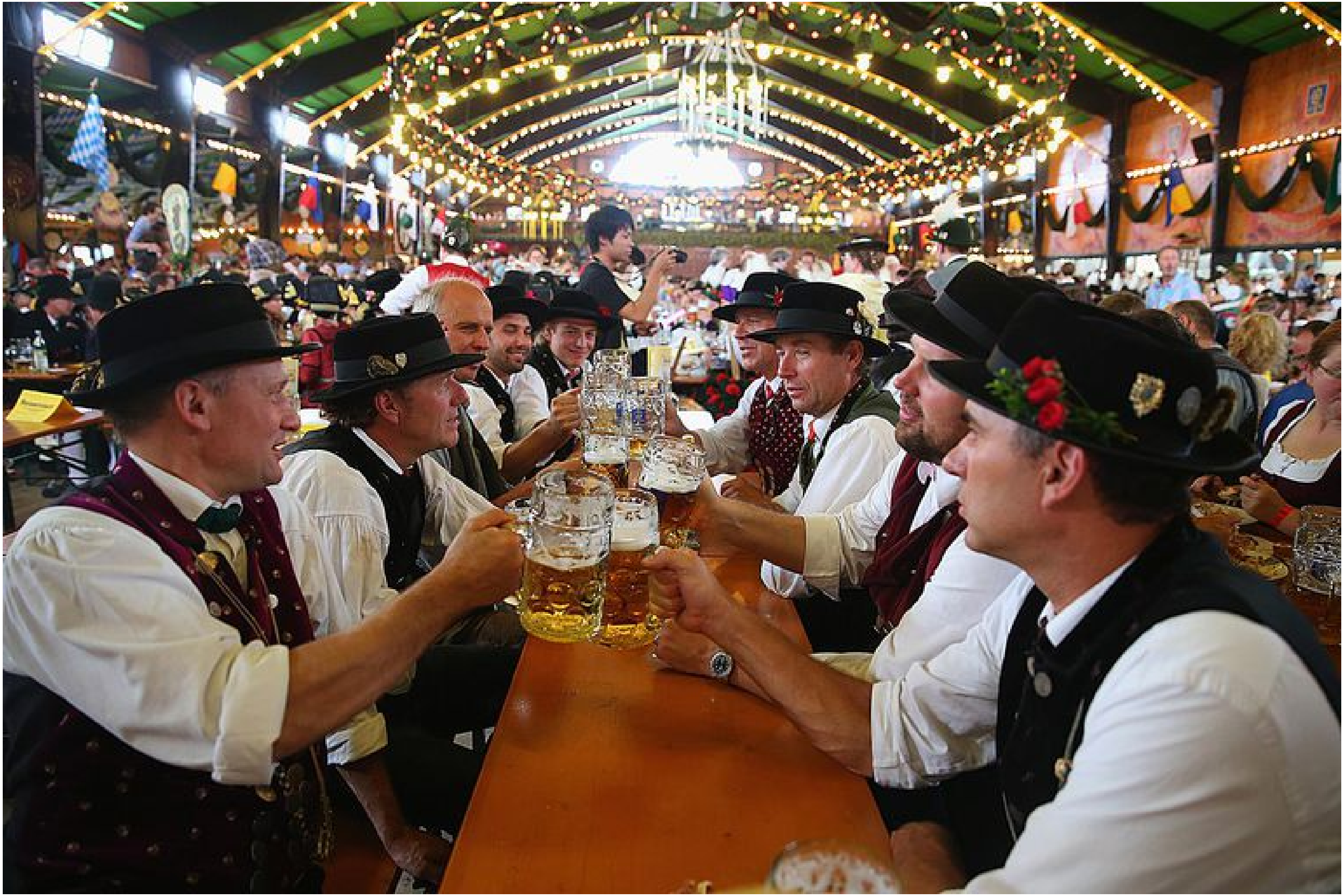
Corollary 4. *If $P_=(\cdot|\varphi)$ is an entropy limit point of $\mathbb{E}_\varphi$ then*

$$\text{maxent}\,\mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\}.$$

Corollary 5. *If $\mathbb{H}_n$ contains $P_=(\cdot|\varphi)$ for sufficiently large n then*

$$\text{maxent}\,\mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\}.$$

Example 6. Suppose we have a single categorical premiss $\exists x Ux$, where $\mathcal{L}$ has a single unary predicate symbol U. $P_=(\exists x Ux) = 1$, so $P_=(\cdot|\exists x Ux) = P_=(\cdot)$. $P_= \in \mathbb{H}_1, \mathbb{H}_2, \ldots,$ so Corollary 5 applies and $\text{maxent}\,\mathbb{E}_\varphi = \{P_=\}$.



4 An alternative route to conditionalisation

This section demonstrates agreement between the maximal entropy approach and conditionalisation without appeal to entropy limit points.

Again we consider categorical sentences $\varphi_1, \ldots, \varphi_k$ and abbreviate $\varphi_1 \wedge \cdots \wedge \varphi_k$ by φ.

Let sentence φ^n be the disjunction of those n-states ω such that $\varphi \wedge \omega$ has positive measure:

$$\varphi^n \stackrel{\text{df}}{=} \bigvee \{\omega \in \Omega_n : P_=(\omega \wedge \varphi) > 0\}.$$

Let N be the greatest index of constants appearing in φ.

Theorem 7 (Agreement with Conditionalisation)**.** *For all $\varphi \in S\mathcal{L}$ with $P_=(\varphi) \in (0,1]$ and all $n \geq N$ it holds that*

$$\text{maxent}\,\mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\} = \{P_=(\cdot|\varphi^n)\} = \{P_=(\cdot|\varphi^N)\} \ .$$

Example 8. For the premiss sentence $\varphi = (\exists x \forall y Uxy \wedge Ut_1t_1) \vee (\forall x \exists y \neg Uxy \wedge \neg Ut_1t_1)$ it holds that

$$\text{maxent}\,\mathbb{E} = \{P_=(\cdot|\neg Ut_1t_1)\} \ .$$

5 Jeffrey Conditionalisation

Consider the premiss φ^c for $P_=(\varphi), c \in (0,1)$.

Theorem 9 (Jeffrey Conditionalisation)**.** *For all* $c \in (0,1)$ *and all* $\varphi \in S\mathcal{L}$ *such that* $P_=(\varphi) \in (0,1)$*, the maximal entropy function is obtained by Jeffrey Conditionalisation with respect to the equivocator function:*

$$\text{maxent}\,\mathbb{E} = \{P^\dagger\}$$

where

$$P^\dagger(\cdot) = c \cdot P_=(\cdot|\varphi^N) + (1-c) \cdot P_=(\cdot|\neg\varphi^N) \;\; .$$

6 Preservation

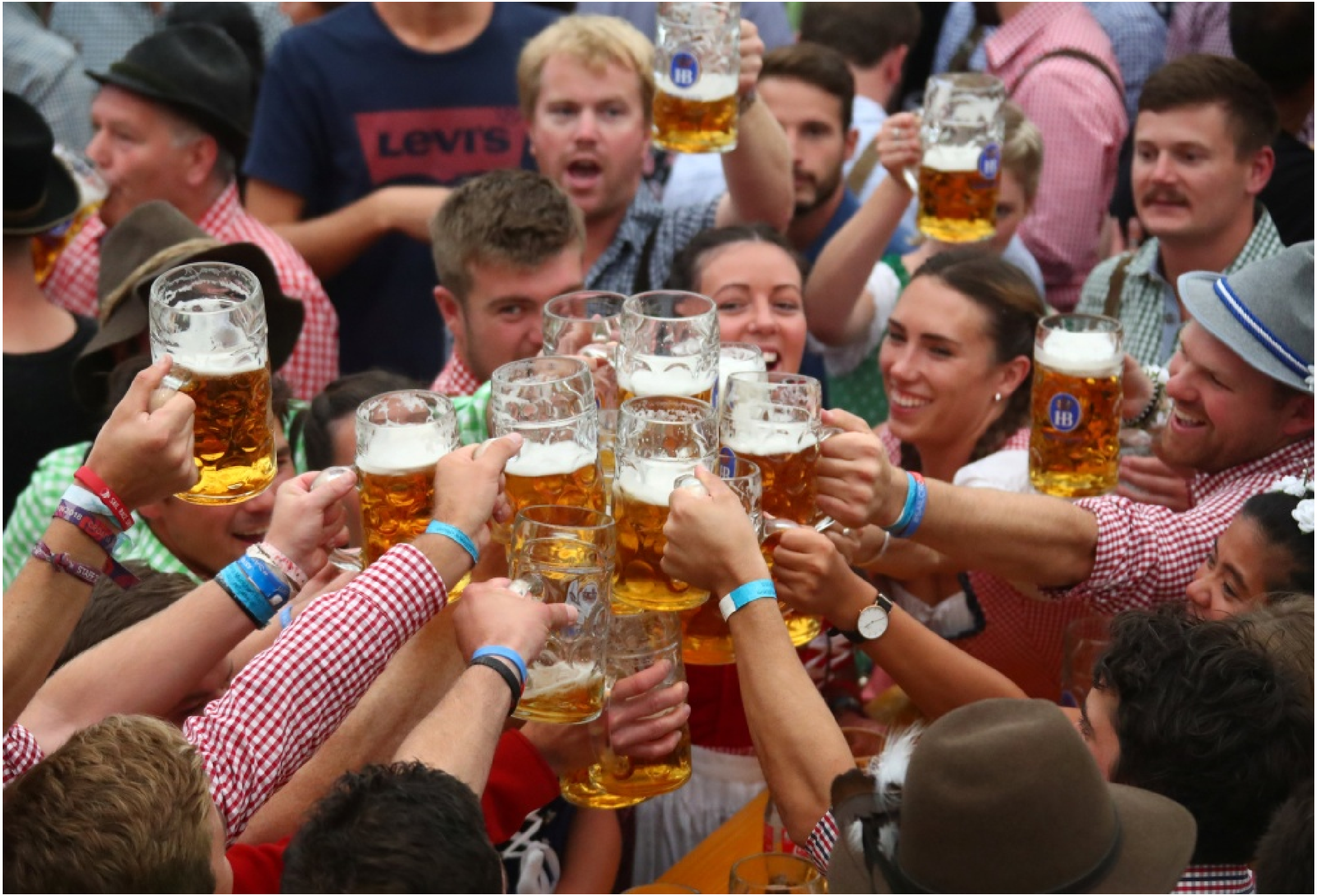
Theorem 10 (Preservation)**.** *If* $\mid\!\overset{\circ}{\approx} \theta$ *and* $\not\mid\!\overset{\circ}{\approx} \neg\varphi$*, then* $\varphi^c \mid\!\overset{\circ}{\approx} \theta$ *for any* $c \in (0, 1]$*.*

Related to:

Obstinacy (Paris, 1994, p. 99)

Rational Monotonicity (Lehmann and Magidor, 1992, §3.4)

Absolute Continuity


Levi's
HB

7 Conclusion

The concept of an entropy limit point can help with the general problem of determining maximal entropy functions.

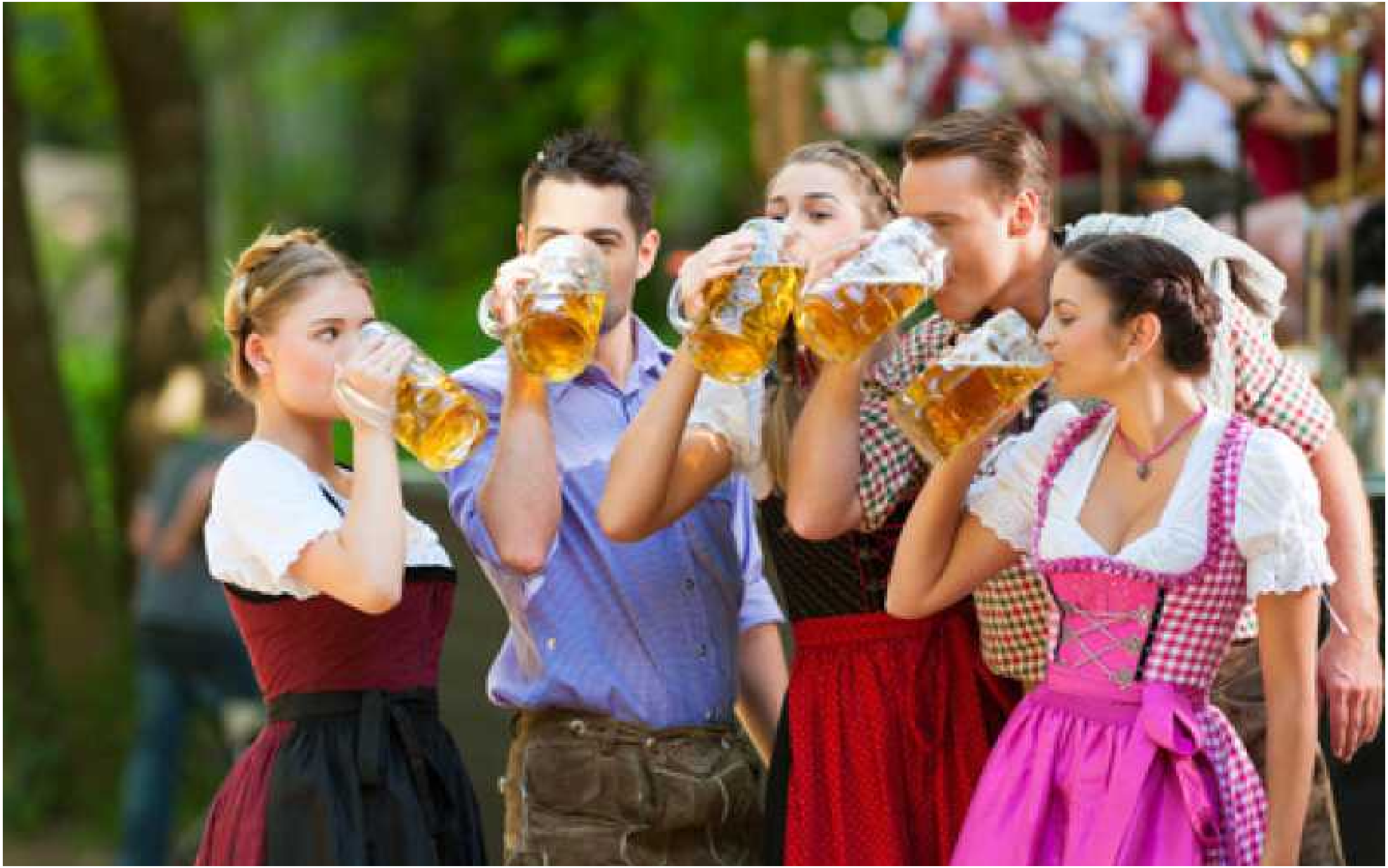
If premisses have positive measure, maxent agrees with Bayesian Conditionalisation.

This extends to Jeffrey Conditionalisation.

Open questions:

- How do we determine maxent when φ has zero measure?
  - In some cases we can appeal to entropy limit points.
  - EG $\forall x U x^c$
  - In some cases, there is no maximal entropy function.
  - EG $\exists x \forall Y U x y$, or $\forall x \exists y \forall z S x y z$
- Does maxent depend on the order of the constants?
  - (The *greater entropy* relation is relative to a specific ordering.)

Thanks to Alena Vencovska, Jeff Paris, Leverhulme Trust, DFG.



Bibliography

Haenni, R., Romeijn, J.-W., Wheeler, G., and Williamson, J. (2011). *Probabilistic logics and probabilistic networks*. Synthese Library. Springer, Dordrecht.

Lehmann, D. and Magidor, M. (1992). What does a conditional knowledge base entail? *Artificial Intelligence*, 55(1):1–60.

Paris, J. B. (1994). *The uncertain reasoner's companion*. Cambridge University Press, Cambridge.