# Symmetric priors *vs* asymmetric expectation

Hykel Hosni

CPNSS - London School of Economics

# Outline

# Reminder: Ellsberg's 1 urn problem

| | 1/3 | 2/3 | |
|---|---|---|---|
| | R | Y | B |
| $1_R$ | 1 | 0 | 0 |
| $1_Y$ | 0 | 1 | 0 |
| $1_{R\cup B}$ | 1 | 0 | 1 |
| $1_{Y\cup B}$ | 0 | 1 | 1 |

### Ellsberg's preference

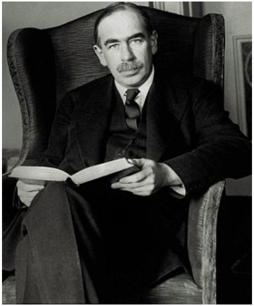
$1_R \succ 1_Y$ and $1_{Y\cup B} \succ 1_{R\cup B}$ which is *inconsistent* with:

1. the probabilistic representation of uncertainty
2. the maximisation of (subjective) expected utility as the norm of rational decision

See Richard's third Decision Theory Masterclass – Link

# THE PROBLEM ISN'T VERY NEW

*The sense in which I am using the term ['uncertain' knowledge] is that in which the prospect of a European war is uncertain [...] About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know.* **Nevertheless, the necessity for action and for decision compels us as practical men to do our best to overlook this awkward fact and to behave exactly as we should if we had behind us a good Benthamite calculation of a series of prospective advantages and disadvantages, each multiplied by its appropriate probability, waiting to be summed.**



J. M. Keynes, *The General Theory of Employment* Q.J.E. 51 (1937): 209-23.

# Goal

Illustrate the key idea behind some well-established models of uncertainty quantification and decision-making which can be taken as ways of *doing the Benthamite calculation*

## Relevance to MSU

- Keynes's challenge amounts to saying that you may well have an Ellsberg preference for playing the roulette over betting at the races, yet, how should you gamble at the races if you are forced to?
- less frivolously, how should you do your Benthamite calculation if you are sitting on the IPCC panel?

# Symmetric vs asymmetric probability averaging

The comparison of a selection of models provides useful conceptual, technical and methodological insights.

**MaxEnt** Symmetric average of the probabilities which are not being ruled out by the available data

**Minimax** Asymmetric average of the probabilities which are not being ruled out by the available data

**Choquet** Non-additive average of the probabilities which are not being ruled out by the available data

**Variational** Minimally Asymmetric average of the probabilities which are not being ruled out by the available data

# Outline

A reminder of material already covered by Richard in his Masterclasses

**Idea**

- agent $i \cong$ preference relation $\succsim_i$ over uncertain prospects
- rationality of $i \cong$ consistency of $\succsim_i \Leftrightarrow$ maximisation of the utility function $U_i$ which represents $\succsim_i$

- Pascal (1669) *Pensées*, part III, §233
- Bernoulli, D. (1738). "Exposition of a New Theory on the Measurement of Risk" . *Econometrica*, 22(1), 23–36.
- [von Neumann and Morgenstern(1953)]
- [Savage(1972)] (combining ideas of vnM, Ramsey, de Finetti)
- [Anscombe and Aumann(1963)] (combining vnM and Savage)

# A set up for the A-A representation

- $S$ is a possibly finite set of *states* endowed with an algebra $\Sigma$
- $\Delta(\Sigma)$ is the set of probability functions $P : \Sigma \to [0, 1]$
- $C$ is a set of *consequences*, which we assume is a *convex* subset in a vector space – typically lotteries
- $F$ is the set of *acts* of the form $f : S \to C$
- The convexity of $C$ guarantees that for all $f, g \in F$ and $\alpha \in (0, 1)$ *mixed acts* can always be defined as a convex combinations of acts $\alpha f + (1 - \alpha)g$, such that for all $s \in S$

$$(\alpha f + (1 - \alpha)g)(s) = \alpha f(s) + (1 - \alpha)g(s)$$

- $\succsim \subseteq F^2$ is a(n idealised agent's) preference relation

# Anscombe-Aumann Axioms

## Weak Ordering

$$\succsim \subseteq F^2 \text{ is total and transitive} \tag{A-A.1}$$

## Continuity (Archimedean)

For $f, g, h \in F$ such that $f \succ g \succ h$, $\exists \alpha\,\beta \in (0,1)$ such that

$$f \succ g \Rightarrow \alpha f + (1-\alpha)h \succ g \succ \beta f + (1-\beta)h. \tag{A-A.2}$$

## Independence

For $f, g, h \in F,$ and $\alpha \in (0,1)$,

$$f \succ g \Rightarrow \alpha f + (1-\alpha)h \succ \alpha g + (1-\alpha)h. \tag{A-A.3}$$

## Monotonicity

$$\text{If } f(s) \succsim g(s) \forall s \in S, \text{ then } f \succsim g \tag{A-A.4}$$

# Anscombe Aumann Representation

**Theorem ([Anscombe and Aumann(1963)])**

*The following are equivalent:*

1. $\succsim \subseteq F^2$ *satisfies WeO, Con, Ind, Mon*
2. $\exists$ *nonconstant* $U : C \to \mathbb{R}$ *and a unique* $P : \Sigma \to [0,1]$ *s.t.*

$$f \succsim g \Leftrightarrow E_P^{A-A}(f) \geq E_P^{A-A}(g),$$

*Moreover, for* $V : C \to \mathbb{R}$

$$f \succsim g \Leftrightarrow E_P^{A-A}(f) \geq E_P^{A-A}(g) \Leftrightarrow V(f) = aU(f) + b$$

*for* $a, b \in \mathbb{R}$, *with* $a > 0$, *i.e.* $U$ *is cardinally unique.*

# $E_P^{A-A}$

$$E_P^{A-A}(f) = \int_P (U(f(s)))\, dP(s).$$

The $E_P^{A-A}$ functional has an *inner part*

$$U(f(s)) = \sum_{x \in supp\, f(s)} u(x) f(s) \tag{1}$$

representing the objective expected utility of $f(s)$ (with $C = \Delta(X)$ and $u : X \to \mathbb{R}$) and an *outer* part

$$\int_S \left( \sum_{x \in Supp\, f(s)} u(x) f(s) \right) dP(s) \tag{2}$$

representing the agent's beliefs by averaging the expected utilities of (1) w.r.t. $P$

# Idea of the proof

1. The restriction of the A-A axioms to *certain acts* establishes the representation of the inner part (vNM risk preferences) i.e. for lotteries $p, q$, there exists $U : X \to \mathbb{R}$ such that

$$p \succsim q \Leftrightarrow \sum_{x \in supp\, p} U(x)p \geq \sum_{x \in supp\, q} U(x)q$$

2. The outer part is established by showing that the axioms are necessary and sufficient for the existence of a (normalised, positive and affine) functional which represents the subjective probability $P \in \Delta(\Sigma)$

# Outline

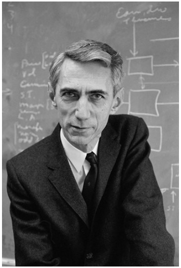
# THE SHANNON ENTROPY

### IDEA

Shannon's interest was in measuring the amount of information (that could be transmitted on a channel and its compression).

### AXIOMS/DESIDERATA

1. The information carried by an event depends only on its probability
2. the measure should be continuous
3. the measure should be additive



Fixing a *contextual meaning* for events,

$$H(P) = -\sum_{s \in S} P(s) log P(s)$$

is the *only* (up to affine transformations) measure satisfying the above

# Background

- Let $P_= \in \Delta(\Sigma)$ be the uniform distribution on $S$
- Let $P, Q \in \Delta(\Sigma)$. The *Kullback-Leibler divergence* is defined by

$$KL(P, Q) = \sum_{s \in S} P(s) log \frac{P(s)}{Q(s)}$$

- Minimising $K(P, P_=)$ is equivalent to maximising the Shannon entropy of $P$

$$H(P) = -\sum_{s \in S} P(s) \log P(s).$$

## Key references


Cover, T., & Thomas, J. (2009). *Elements of Information Theory* (2nd edition.). Wiley.

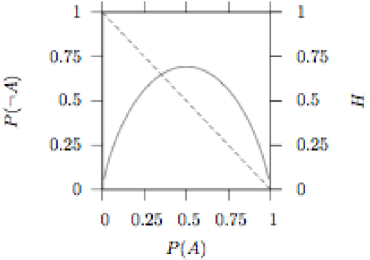
Paris, J. B. (1994). *The uncertain reasoner's companion: A mathematical perspective*. CUP

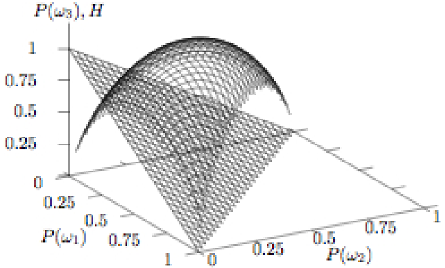
Williamson, J. (2010).*In defence of objective Bayesianism*. OUP.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. CUP.

## IDEA

When there is a lot at stake, (i.e. much to be lost as a consequence of misjudgment) prudence demands keeping as far as possible from *extreme degrees of belief*




Figures from Williamson (2010)

# Justification (Representation theorems)


- Shore, J. E. and Johnson, R. W. (1980) Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy, *IEEE Transaction on Information Theory*, IT-26, pp. 26–37.
- Paris, J. B., & Vencovská, A. (1997). In Defense of the Maximum Entropy Inference Process. *International Journal of Approximate Reasoning*, 17(97), 77–103.
- Grunwald, P. D., & Dawid, A. P. (2004). Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Annals of Statistics*, 32(4), 1367–1433.
- Landes, J., & Williamson, J. (2013). Objective Bayesianism and the Maximum Entropy Principle. *Entropy*, 15(9), 3528–3591.

# Relevance to MSU

MaxEnt *methods* meet Keynes's challenge by always feeding the classical Expected Utility norm with a single (unique) probability distribution on the variables of interest.

+'s

- confidence-stakes-uncertainty
- no risk of inaction
- several axiomatisations

-'s

- sensitive to $\mathcal{L}$
- worst case complexity
- knowledge and ignorance are treated symmetrically