Fundamentally finitary foundations for probability and bounded probability

Matthias C. M. Troffaes

Department of Mathematical Sciences
Durham University, United Kingdom

31 January 2024
SIPTA Seminar

Aim

- remove non-constructive elements (e.g. axiom of choice)
  in the foundations of probability and bounded probability
- bring together three seemingly unrelated theories: Williams, Lindley, and Nelson

Outline

Outline

De Finetti

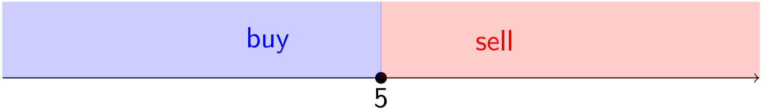
Probability Through Betting

Coin Bet

- Consider the outcome of tossing a fair coin.

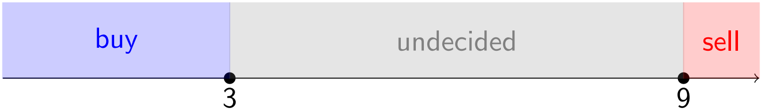
  Would you pay me $\$5 - \epsilon$ *now*
  if I promise to pay out \$10 to you
  if the coin lands heads?

  If I payed you $\$5 + \epsilon$ *now*
  would you promise to pay out \$10 to me
  if the coin lands heads?



Snow Bet

- Consider whether or not it snows in Moscow tomorrow.
  Would you pay me $\$5 - \epsilon$ *now*
  if I promise to pay out $10 to you
  if it snows in Moscow tomorrow?
  If I payed you $\$5 + \epsilon$ *now*
  would you promise to pay out $10 to me
  if it snows in Moscow tomorrow?



Williams: Recap of theory

Events

- Ω = possibility space — set of exhaustive and mutually exclusive outcomes of an experiment
- $A, B, \ldots \subseteq \Omega$ = event
- $\wp$ = power set of Ω — set of all events

Random Quantities

- $f, g, \ldots : \Omega \to \mathbb{R}$ bounded = random quantity
- $\mathcal{L}$ = set of all random quantities (on Ω)
- I_A = indicator function

$$I_A(\omega) = 1 \text{ if } \omega \in A \text{ else } 0 \tag{1}$$

Williams: Recap of theory

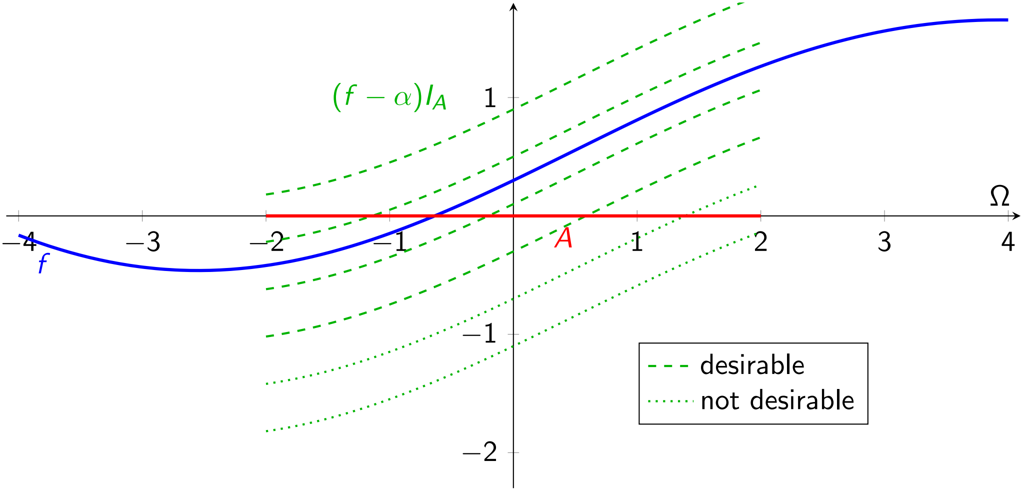
Lower Previsions

- $\underline{P}\colon \mathcal{L} \times (\wp \setminus \{\emptyset\}) \rightharpoonup \mathbb{R}$ (partial real function on $\mathcal{L} \times (\wp \setminus \{\emptyset\})$) = lower prevision
- $\operatorname{dom} \underline{P}$ = domain of $\underline{P}$
- $\overline{P}(f \mid A) := -\underline{P}(-f \mid A)$ = upper prevision

subject's $\underline{P}$ is understood as them stating that for each $(f, A) \in \operatorname{dom} \underline{P}$
$\underline{P}(f \mid A)$ is supremum called-off buying price for f if A obtains

$$(f - \alpha)I_A \text{ is desirable for all } \alpha < \underline{P}(f \mid A)$$

Williams: Recap of theory



Williams: Recap of theory

Consistency (Williams [Wil75])

$\underline{P}$ is called consistent if
for all $\epsilon > 0$, non-empty $\mathcal{K} \Subset \operatorname{dom} \underline{P}$, and $\lambda \colon \mathcal{K} \to \mathbb{R}$, there is no $\lambda \succ 0$ such that

$$\sum_{(f,A)\in\mathcal{K}} \lambda_{f,A}(f - \underline{P}(f \mid A) + \epsilon) I_A \leq 0 \tag{2}$$

Inference (Williams [Wil75])

Assume $\underline{P}$ consistent.
The natural extension $\underline{E}$ of $\underline{P}$ is, for each $g \in \mathcal{L}$ and $B \in \wp \setminus \{\emptyset\}$, defined as

$$\underline{E}(g \mid B) := \lim_{\substack{\epsilon>0 \\ \mathcal{K}\Subset\operatorname{dom}\underline{P}}} \sup_{\substack{\alpha\in\mathbb{R} \\ \lambda\colon \mathcal{K}\to\mathbb{R} \\ \lambda\geq 0}} \left\{ \alpha \colon (g - \alpha) I_B \geq \sum_{(f,A)\in\mathcal{K}} \lambda_{f,A}(f - \underline{P}(f \mid A) + \epsilon) I_A \right\} \tag{3}$$

Williams: Recap of theory

(also see lower envelope theorem by Walley [Wal91, §3.3.3, §3.4.1]])

Theorem (Duality Theorem by Troffaes and Cooman [TC14])

$\underline{P}$ avoids sure loss if and only if there is a finitely additive probability measure μ on $\wp$ such that

$$\forall f \in \operatorname{dom} \underline{P},\ \int f \, d\mu \geq \underline{P}(f) \tag{4}$$

If $\underline{P}$ avoids sure loss then for all $g \in \mathcal{L}$

$$\underline{E}(g) = \min_{\mu} \left\{ \int g \, d\mu \colon \forall f \in \operatorname{dom} \underline{P},\ \int f \, d\mu \geq \underline{P}(f) \right\} \tag{5}$$

- non-constructive proof: needs ultrafilter principle (related to axiom of choice)
- f.a. probability measures on $\wp$ are intangible (see Schechter [Sch97])
- conditional version needs full conditional f.a. probability measures which are notoriously complex (see Krauss [Kra68])

Williams: Contributions from this talk

Later in this talk I will show...

- fully constructive duality theorem for Williams's theory
- technique: use nets of quite simple 'finite' objects, namely Lindley's urns
  no full conditional f.a. probability measures needed!

Note...

- net can be replaced with single nonstandard 'finite' object
  e.g. using Nelson's internal set theory
  (but I will not discuss this further in this talk)

Outline

Lindley: High risk events

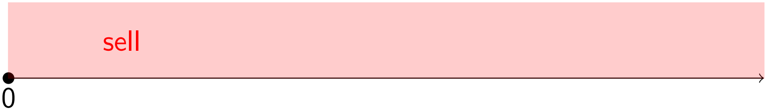
e.g. nuclear engineering, oncology, asteroid impact events, ...

Doomsday Bet

- suppose we both believe that
  the world will come to an end in the next week with probability 1/2
- for simplicity we ignore non-linear utility for monetary values

  Would you pay me $\$5 - \epsilon$ *now*
  if I promise to pay out $10 to you in one week
  if the world comes to an end during that period?

  If I payed you $\$5 + \epsilon$ *now*
  would you promise to pay out $10 to me in one week
  if the world comes to an end during that period?

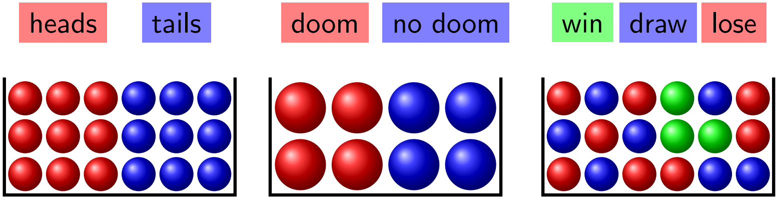


Lindley: Urn model

See Lindley [Lin06] and Aven and Reniers [AR13]:

- think of each $\omega \in \Omega$ as a colour — Ω need not be finite!
- an urn is a finite collection of coloured balls with colours from Ω
- a subject's urn is a standard for their uncertainty about $\omega \in \Omega$:
  by specifying an urn, the subject declares that
  their uncertainty about the colour of one ball uniformly drawn from their urn
  is equivalent to their uncertainty about the state of the world
- since one ball is drawn uniformly, only proportions matter

Lindley: Urn model



Lindley: Urn Model

Definition

An urn is a function $q\colon \Omega \to \mathbb{Q}_{\geq 0}$ with

- finite support $\{\omega\colon q(\omega) > 0\}$ is a finite set
- $\sum_{\omega\in\Omega} q(\omega) = 1$

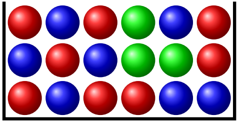
Some extra notation:

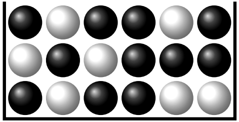
- $q(A) := q(I_A) = \sum_{\omega\in A} q(\omega)$ σ-additive probability measure
  represents the proportion of balls with colours in A
- $q(B \mid A) := \frac{q(B\cap A)}{q(A)}$ (provided $q(A) > 0$) ~~full~~ conditional prob. measure
  represents the conditional proportion of colours in B restricted to colours in A
- $q(f) := \sum_{\omega\in\Omega} q(\omega) f(\omega)$
  represents the average when each ball of colour ω is labelled $f(\omega)$
- $q(f \mid A) := \frac{\sum_{\omega\in A} q(\omega) f(\omega)}{\sum_{\omega\in A} q(\omega)} = \frac{q(fI_A)}{q(A)}$ (provided $q(A) > 0$)
  represents the conditional average of f restricted to balls with colours in A

Lindley: Urn Model

win draw lose

$A = \{\text{win}, \text{lose}\}$

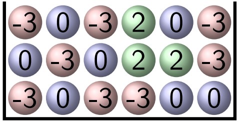




$q(A) = 11/18$

$$f(W) = 2,\ f(D) = 0,\ f(L) = -3$$

$q(\text{win}) = 3/18$
$q(\text{draw}) = 8/18$
$q(\text{lose}) = 7/18$



$$q(f) = \frac{8\times(-3)+7\times0+3\times2}{8+7+3}$$

Lindley: Urn model

- conceptually simple, easy to understand for people untrained in probability
- based on introspection: not experimentally verifiable
- consistency immediate
- rules of inference immediate from properties of proportions and averages
- provides sensible foundation for decision making
- practically limited to finite Ω since finite number of balls
- probabilities are proportions so only rational valued probabilities
- no full conditional probabilities
- hard to specify the urn under severe uncertainty
  can we use bounding?

contributions from this talk: solve the last four problems

Outline

Nelson: Disclaimer

Disclaimer

- will try explain the intuition behind Nelson's radically elementary probability theory using nets (or, sequences)
- beware this re-interpretation is 'incorrect' in various ways
- this will not do full justice to Nelson's internal set theory approach
  it is only my personal subjective re-interpretation of Nelson's nonstandard analysis using only objects from standard analysis

Why?

- for the sake of this audience being mostly unfamiliar with nonstandard analysis
- nets of probability mass functions will play essential role later

Nelson: Central limit theorem

An example...

- Consider $X_n \sim \mathrm{Bin}(n, 1/2)$
- Let $Y_n := \frac{X_{n^2} - n^2/2}{n/2}$

$$\mathrm{E}(Y_n) = \frac{\mathrm{E}(X_{n^2}) - n^2/2}{n/2} = 0 \tag{6}$$

$$\mathrm{Var}(Y_n) = \frac{\mathrm{Var}(X_{n^2})}{n^2/4} = 1 \tag{7}$$

$$\mathrm{P}(Y_n \in \mathbb{Q}) = 1 \tag{8}$$

What does probability theory say about $\lim_{n\to\infty} Y_n$?

Nelson: Finite Additivity

- Under weak-* (i.e. pointwise) convergence

$$P\left(\lim_{n\to\infty} Y_n \in A\right) := \lim_{n\to\infty} P(Y_n \in A) \tag{9}$$

it cannot have a σ-additive distribution on the Borel σ-field.

$$\lim_{n\to\infty} P(Y_n = y) = 0 \text{ for all } y \in \mathbb{Q}, \text{ yet } \lim_{n\to\infty} P(Y_n \in \mathbb{Q}) = 1 \tag{10}$$

- Solution from probability theory:

sequence Y_n converges in distribution

i.e. can only say $\lim_{n\to\infty} P(Y_n \leq y) = \Phi(y)$

- Solution from Nelson [Nel87] (re-interpreted):

Y_n has a weak-* converging subsequence to a f.a. probability measure on $\wp$

Nelson: Ambiguity

(Interpretation of) Nelson's theory

Represent all of probability though
weak-* converging nets of probability mass functions (e.g. urns)

Consequences

- normal distribution is not unique (many converging subsequences!)

  this ambiguity is essential part of theory
- normal distribution is allowed to be f.a. even on the Borel σ-field
- non-constructive (converging subsequence needs ultrafilter principle)
- his central limit theorem does not rely on σ-additivity
  (SIPTA community also rediscovered this: see Shafer and Vovk [SV01])

Thoughts

- $\liminf$ and $\limsup$ instead of $\lim$? makes theory fully constructive!
- rephrase central limit theorem as a bounding statement?

Outline

Urn Interpretation of Bounded Probability

- as with Williams, start with $\underline{P}\colon \mathcal{L} \times (\wp \setminus \{\emptyset\}) \rightharpoonup \mathbb{R}$
- but interpret $\underline{P}$ not through betting; instead...

Supremum Lower Average Interpretation

For every $\epsilon > 0$ and every $\mathcal{K} \Subset \operatorname{dom} \underline{P}$, your urn q is such that for every $(f, A) \in \mathcal{K}$, $q(A) > 0$ and $q(f \mid A) \geq \underline{P}(f \mid A) - \epsilon$.

- $\underline{P}(f \mid A)$ represents a
  lower bound on the subject's urn average (up to ϵ) for all $(f, A) \in \mathcal{K}$
  q is allowed to depend on choice of $\epsilon > 0$ and $\mathcal{K} \Subset \operatorname{dom} \underline{P}$
- q is never fully specified (unlike Lindley)
- can be used to check consistency and perform inference

Definition (Approximate Credal Set)

$$\mathcal{Q}_{\epsilon,\mathcal{K}}(\underline{P}) \coloneqq \{q\colon \forall (f, A) \in \mathcal{K},\ q(A) > 0 \text{ and } q(f \mid A) \geq \underline{P}(f \mid A) - \epsilon\} \tag{11}$$

Urn Interpretation of Bounded Probability

Theorem (Consistency)

The following are equivalent. If any is satisfied, we say $\underline{P}$ is consistent.

(A) *For all $\epsilon > 0$ and all $\mathcal{K} \Subset \operatorname{dom}\underline{P}$, we have $\mathcal{Q}_{\epsilon,\mathcal{K}}(\underline{P}) \neq \emptyset$, i.e. there is a q such that*

$$\forall (f,A) \in \mathcal{K}\colon \quad q(A) > 0 \text{ and } q(f \mid A) \geq \underline{P}(f \mid A) - \epsilon \qquad \textit{generalizes Lindley!}$$

(B) *There is a net of urns $(q_\alpha)_{\alpha \in \mathcal{A}}$ such that for all $\forall (f,A) \in \operatorname{dom}\underline{P}$,*

$$\exists \beta \in \mathcal{A}, \forall \alpha \geq \beta,\ q_\alpha(A) > 0 \text{ and } \liminf_{\alpha \geq \beta} q_\alpha(f \mid A) \geq \underline{P}(f \mid A) \qquad \textit{Nelson's nets!}$$

(C) *For all $\epsilon > 0$, $\emptyset \neq \mathcal{K} \Subset \operatorname{dom}\underline{P}$, and $\lambda\colon \mathcal{K} \to \mathbb{R}$, there is no $\lambda \succ 0$ such that*

$$\sum_{(f,A)\in\mathcal{K}} \lambda_{f,A}(f - \underline{P}(f \mid A) + \epsilon) I_A \leq 0 \qquad \textit{Williams avoiding sure loss!}$$

Urn Interpretation of Bounded Probability

Theorem (Inference)

Assume $\underline{P}$ consistent. Then, for any $g \in \mathcal{L}$ and $B \in \wp \setminus \{\emptyset\}$:

$$\underline{E}(g \mid B) := \boldsymbol{\lim_{\substack{\epsilon > 0 \\ \mathcal{K} \Subset \operatorname{dom} \underline{P}}} \inf_{\substack{q \in \mathcal{Q}_{\epsilon, \mathcal{K}}(\underline{P}) \\ q(B) > 0}} q(g \mid B)}$$

$$= \lim_{\substack{\epsilon > 0 \\ \mathcal{K} \Subset \operatorname{dom} \underline{P}}} \sup_{\substack{a \in \mathbb{R} \\ \lambda \colon \mathcal{K} \to \mathbb{R} \\ \lambda \geq 0}} \left\{ a \colon (g - a) I_B \geq \sum_{(f, A) \in \mathcal{K}} \lambda_{f,A} (f - \underline{P}(f \mid A) + \epsilon) I_A \right\}$$

Urn Interpretation of Bounded Probability

A Few Corollaries

Corollary (Coherent Lower Prevision Representation Theorem)

Assume $\underline{P}$ consistent. Then for every $\epsilon > 0$ and every $\mathcal{J} \Subset \mathcal{L} \times \wp \setminus \{\emptyset\}$ there is a $\delta > 0$, $\mathcal{K} \Subset \operatorname{dom} \underline{P}$, and $Q \in \mathcal{Q}_{\delta,\mathcal{K}}(\underline{P})$ such that for all $(g, B) \in \mathcal{J}$ we have that

$$\underline{Q}(B) > 0 \text{ and } \left|\underline{E}(g \mid B) - \underline{Q}(g \mid B)\right| \leq \epsilon$$

proof does not hinge on Krein-Milman

Corollary (A Variation of Kraus's Representation Theorem)

Every full conditional f.a. probability measure can be represented by a single nonstandard probability mass function with finite support.

proof does not hinge on zero layers

Discussion

- new interpretation of lower previsions: supremum lower average
  - uses rational valued probability mass functions with finite support (i.e. 'urns')
  - works for high risk situations, unlike betting interpretation
  - possibly easier to understand than betting for some audiences
- new formulas for avoiding sure loss and natural extension
  - no need for finitely additive measures or full conditional measures
  - use nets of urns
  - new (simpler?) duality theorem for conditional lower previsions
- unifies Williams, Lindley, Nelson
  - fully generalizes Lindley's (somehwat limited) interpretation
  - shows Nelson's radically elementary probability can capture all of probability theory
- fully constructive: no reliance on ultrafilter principle or Hahn-Banach

Outline

Conclusions

Concerning Bounded Probability

- interpretation of lower previsions need not be betting
- Williams duality possible without full conditional f.a. probability measures

Concerning Probability

- finite additivity helps with limits
- bounding helps theory become more constructive
  e.g. replace lim with lim inf and lim sup
- every probabilist ought to take a serious look at Nelson's work
- is measure theory an 'unnecessary' part of foundations of probability?
- is σ-additivity 'hindering' probability theory?
- rethink scaling limits (such as the central limit theorem) through bounding?

Thank you!

References I

[AR13] Terje Aven and Genserik Reniers. "How to define and interpret a probability in a risk and safety setting". In: *Safety Science* 51 (Jan. 2013), pp. 223–231. DOI: 10.1016/j.ssci.2012.06.005.

[Kra68] Peter H. Krauss. "Representation of conditional probability measures on Boolean algebras". In: *Acta Mathematica Academiae Scientiarum Hungaricae* 19.3-4 (Sept. 1968), pp. 229–241. DOI: 10.1007/BF01894506.

[Lin06] Dennis V. Lindley. *Understanding Uncertainty*. Wiley, Apr. 2006. DOI: 10.1002/0470055480.

[Nel87] Edward Nelson. *Radically Elementary Probability Theory*. Annals of Mathematical Studies. New Jersey: Princeton University Press, 1987. URL: https://web.math.princeton.edu/~nelson/books/rept.pdf.

[Sch97] Eric Schechter. *Handbook of Analysis and Its Foundations*. San Diego: Academic Press, 1997.

References II

[SV01] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's Only a Game!* Wiley, 2001.

[TC14] Matthias C. M. Troffaes and Gert de Cooman. *Lower Previsions*. Wiley Series in Probability and Statistics. Wiley, 2014. ISBN: 978-0-470-72377-7.

[Wal91] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall, 1991.

[Wil75] Peter M. Williams. *Notes on conditional previsions*. Tech. rep. School of Math. and Phys. Sci., Univ. of Sussex, Feb. 1975.