ST790 — Fall 2022

Imprecise-Probabilistic Foundations of Statistics

Ryan Martin
North Carolina State University
www4.stat.ncsu.edu/~rmartin

Week 13b

This lecture

- Formal decision theory
- Precise-probabilistic version
  - von Neumann–Morganstern
  - maximize expected utility
- Imprecise-probabilistic version
  - Choquet integrals define lower/upper expected utility
  - how to optimize?
- Applications along the way

Setup

- Ingredients:
  - uncertain variable $X \in \mathbb{X}$
  - collection $f \in \mathcal{F}$ of possible actions
  - reward/value/utility[1] function $u : \mathcal{F} \times \mathbb{X} \to \mathbb{R}$
- Then the "game" goes as follows:
  - if I choose action f
  - and it happens that $X = x$
  - then I get utility $u_f(x)$
- Decision theory: which f should I pick???
- *Fundamentally important problem!*
  - of course in STAT, ML, AI, etc.
  - also in real, everyday life

[1] We want to maximize "utility." An alternative formulation uses "loss," which we want to minimize. The two are equivalent: loss $= -$utility

Setup, cont.

- Why is it challenging?
- In general, no f uniformly maximizes utility
- Toy example taken from Denoeux (*IJAR* 2019, Example 1)
- Not much more[2] can be said without more structure
  - relaxing "uniformly maximizes" and/or
  - non-vacuous UQ about X

Utility	x_1	x_2	x_3
f_1	37	25	23
f_2	49	70	2
f_3	4	96	1
f_4	22	76	25
f_5	35	20	23

[2] Note that f_5 is always *strictly worse* than f_1

Setup, cont.

- Need to formalize what it means to compare actions
- Requires a *preference relation*, a sort of "ordering" on $\mathcal{F}$
- For $f, f^\star \in \mathcal{F}$, say
  - $f \preceq f^\star$ iff "$f^\star$ is at least as desirable as f"
  - $f \prec f^\star$ iff "$f^\star$ is strictly more desirable than f"
- A preference relation is a *preorder* iff
  - *reflexive:* $f \preceq f$ for all f
  - *transitive:* $f \preceq f', f' \preceq f^\star \implies f \preceq f^\star$
- Preorder is *complete* if in every pair of actions, one is at least as desirable as the other, i.e., $f \preceq f'$ or $f' \preceq f$
- A complete preorder has at least one *greatest element*, i.e., a $f^\star$ such that $f^\star \succeq f$ for all f

Vacuous UQ

- Suppose we're ignorant[3] about X
- What are some reasonable preference relations?
- Sort numerical summaries of $x \mapsto u_f(x)$
- Two extreme possibilities, *optimistic* and *pessimistic*[4]
  - maximax: $f \preceq f^\star \iff \max_x u_f(x) \leq \max_x u_{f^\star}(x)$
  - maximin: $f \preceq f^\star \iff \min_x u_f(x) \leq \min_x u_{f^\star}(x)$

Utility	x_1	x_2	x_3	min	max
f_1	37	25	23	23	37
f_2	49	70	2	2	70
f_3	4	96	1	1	96
f_4	22	76	25	22	76

[3] Like in non-Bayesian statistical decision theory

[4] This is the familiar pessimistic notion of "minimax" in statistics

Probabilistic UQ

- Suppose there's a (precise) prob distribution P on $\mathbb{X}$
- Totally reasonable strategy:
  - $X \mapsto u_f(X)$ is a random variable for each f
  - seek f that maximizes expected utility $f \mapsto \mathsf{P}u_f$
  - i.e., $f \preceq f^\star \iff \mathsf{P}u_f \leq \mathsf{P}u_{f^\star}$
- "MEU" is Bayesian decision theory in statistics
  - prior distribution for Θ
  - data y + prior + Bayes's rule = posterior Π_y
  - Bayes estimator minimizes $f \mapsto \Pi_y \ell_f$

Probabilistic UQ, cont.

- Toy example, revisited:
  - maximax
  - maximin
  - $\max \mathrm{P} u_f$ for $\mathrm{P} = (0.30, 0.50, 0.20)$

Utility	x_1	x_2	x_3	min	max	$\mathrm{P} u_f$
f_1	37	25	23	23	37	28.1
f_2	49	70	2	2	70	50.1
f_3	4	96	1	1	96	49.4
f_4	22	76	25	22	76	49.6

Probabilistic UQ, cont.

- WLOG, consider a slightly different setup:
  - action $f \in \mathcal{F}$ is a map $f : \mathbb{X} \to \mathbb{C}$
  - $\mathbb{C}$ is a space of *consequences*
  - taking action $f \in \mathcal{F}$ when $X = x$ leads to $f(x) \in \mathbb{C}$
  - then $u : \mathbb{C} \to \mathbb{R}$ measures utility of a consequence
- If $X \sim \mathrm{P}$, then $C = f(X)$ is a RV w/ distribution $\mathrm{P}f^{-1}$
- A probability distribution on $\mathbb{C}$ is called a *lottery*[5]
  - lottery returns to me a consequence chosen at random
  - I know which consequences are desirable and which ones aren't
  - so in principle there are some lotteries I prefer more others
- Properties of a preference order on lotteries...?[6]

[5] Lotteries will be denoted by letters $Q, Q^\star, \ldots$

[6] von Neumann–Morganstern (1947), *Theory Games & Economic Behavior*

Probabilistic UQ, cont.

Axioms for preferences $\preceq$ on lotteries.

- *Completeness:* $\preceq$ is a complete preorder
- *Continuity:* if $Q^- \prec Q \prec Q^+$, then there exists $\alpha, \beta \in (0,1)$ s.t.

$$\alpha Q^- + (1-\alpha)Q^+ \prec Q \prec \beta Q^- + (1-\beta)Q^+$$

- *Independence:* for any $(Q, Q', Q^\star)$ and any $\alpha \in (0,1)$,

$$Q \preceq Q^\star \iff \alpha Q + (1-\alpha)Q' \preceq \alpha Q^\star + (1-\alpha)Q'$$

Theorem (von Neumann–Morganstern).

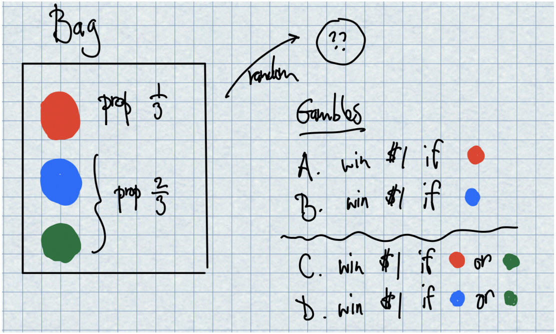
A preference $\preceq$ for lotteries on $\mathbb{C}$ satisfies the above axioms if and only if there exists a utility function $u : \mathbb{C} \to \mathbb{R}$ such that

$$Q \preceq Q^\star \iff Qu \leq Q^\star u$$

i.e., $\preceq$ satisfies axioms iff it corresponds to maximizing expected utility

Probabilistic UQ, cont.

- vNM's result is powerful, but if axioms don't match real-world preferences, then the theorem is useless
- Unfortunately,..... there's *Ellsberg paradox*[7]
  - vNM theory says $A \succeq B \implies C \succeq D$
  - Ellsberg: for some people, $A \succ B$ and $C \prec D$



[7] https://en.wikipedia.org/wiki/Ellsberg_paradox

Imprecise-probabilistic UQ

- For various reasons, e.g., Bayes with imprecise prior, we might not have a precise probability distribution on $\mathbb{X}$
- Let $(\underline{\mathrm{P}}, \overline{\mathrm{P}})$ be a lower/upper probability on $\mathbb{X}$
- Can calculate lower/upper expected utility using Choquet
- e.g., if $\underline{\mathrm{P}}$ is a belief function with mass m, then

$$\underline{\mathrm{P}} u_f = \sum_A \Big\{ \min_{x \in A} u_f(x) \Big\} \, m(A)$$

- Generalizations of vNM:[8]
  - *maximax:* maximize $f \mapsto \overline{\mathrm{P}} u_f$
  - *maximin:* maximize $f \mapsto \underline{\mathrm{P}} u_f$
- Ellsberg: strategy (A, D) is maximin optimal[9]

[8] e.g., Gilboa & Schmidler (*J Math Econ* 1989)

[9] see page 204 in *Intro to IP*

- Lots of other papers with different perspectives
- One that aligns with Walley's theory is *maximality*
- Action $f^\star$ is maximal iff $\underline{\mathsf{P}}(u_{f^\star} - u_f) \geq 0 \ \forall \ f \in \mathcal{F}$
- $f^\star$ maximal is equivalent to

$$\mathsf{P} u_{f^\star} \geq \mathsf{P} u_f \quad \text{for all } f \text{ and all } \mathsf{P} \in \mathscr{C}(\underline{\mathsf{P}})$$

- See Chapter 8 in *Intro to IP* and the references therein

Applications

- Robust/generalized Bayes has an imprecise prior distribution, characterized by a credal set of precise priors
- If there's just a single prior, then the Bayes estimator is the one that minimizes posterior expected loss
- Set of priors $\rightarrow$ set of posteriors
- Lower/upper posterior expected loss
- Common to use the minimax generalized Bayes rule[10]
- Computation can be a challenge, but Wasserman and others have provided formulas for certain prior classes

[10] Berger et al (*TEST* 1994), etc.

Applications, cont.

- IM developments have focused on *inference*, not decisions
- Same is true for fiducial, confidence distributions, etc.[11]
- Nothing stopping us from using the above theory
- Suppose prior info about Θ is vacuous, and that $(\underline{\Pi}_y, \overline{\Pi}_y)$ is the valid IM based on data $Y = y$
- If it has the form of a possibility measure, then

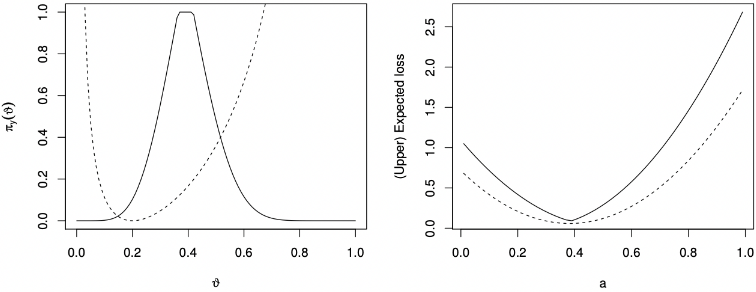
$$\overline{\Pi}_y h = \int_0^1 \Big\{ \sup_{\theta : \pi_y(\theta) > \alpha} h(\theta) \Big\} \, d\alpha, \quad h \geq 0$$

- For non-negative loss ℓ_f, define

$$\hat{f}(y) = \arg\min_f \overline{\Pi}_y \ell_f \qquad \leftarrow \text{IM's minimax action}$$

[11] Except for Taraldsen & Lindqvist (*Ann Stat* 2013)

Applications, cont.



(a) Possibility contour (and loss)

(b) (Upper) expected loss

Figure 4: Binomial example with $n = 18$ and $y = 7$. Panel (a) shows the possibility contour $\pi_y(\vartheta)$; dashed line shows the weighted squared error loss function $\ell_a(\vartheta)$, when $a = 0.2$. Panel (b) shows the (upper) expected loss functions, $a \mapsto \overline{\Pi}_y \ell_a$ (solid) and $a \mapsto Q_y^{\star} \ell_a$ (dashed), for the weighted squared error loss function.

Applications, cont.

- Action $\hat{f}(y)$ is, e.g., an estimate of some feature of Θ
- Can evaluate $\hat{f}(Y)$ based on sampling dist properties
  - unbiased
  - small mean square error
- But what role does the IM's validity property play?[12]
- Intuitively:
  - validity implies error control
  - errors in decision-making happen when there exists f such that $\overline{\Pi}_y \ell_f$ is much less than $\ell_f(\cdot)$ near the true θ
- Validity implies these "decision-making errors" are controlled in a certain sense that's hard to describe[13]

---

[12] M., `arXiv:2112.13247`

[13] I think better results are possible...

Applications, cont.

$$\widetilde{R}(y,\theta) \approx \inf_f \frac{\overline{\Pi}_y \ell_f}{\text{maximum of } \ell_f(\cdot) \text{ nearby true } \theta}$$

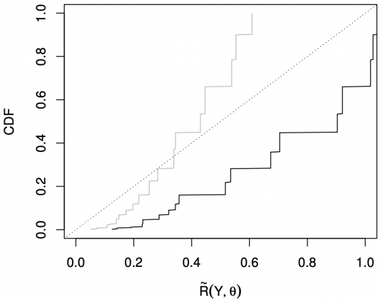


Figure 5: The black line shows the distribution function of $\widetilde{R}(Y,\theta)$, as defined in (14), for the binomial IM described in Section 4.3.2. The gray line shows the distribution function for the version of $\widetilde{R}(Y,\theta)$ for the fiducial IM. The diagonal line corresponds to the $\mathsf{Unif}(0,1)$ distribution function.

Conclusion

- Decision theory is a fundamentally important problem
- Pessimistic minimax/maximin is conservative
- With a prob distribution for X, more can be done
- This is *Bayesian decision theory*
- Having a precise prior might not be realistic in all applications, so extensions to imprecise probability are available
- This is robust/generalized Bayes decision theory
- Valid IMs and decisions...
  - very little is known, lots more work could be done
  - interested in case with partial priors

Next lecture

- Model uncertainty
- Imprecision in the model itself
- Missing/coarse data