Statistical decision theory cannot justify randomization or pre-registration for experiments.

Maximilian Kasy

Department of Economics, University of Oxford

June 8, 2021

Introduction

1. Standard prescriptions for experiments in the social / life sciences:
   a) **Randomization**:
      To balance unobserved heterogeneity in treatment and control group.
   b) **Pre-registration**:
      To prevent "cherry picking" of results, which would distort inference.
2. Decision theory provides the foundation for statistics:
   - Experimental design / estimation / inference / policy choice as **decisions**
   - in the presence of an unknown **state** of the world.
   - The state of the world impacts the distribution of observed **data**,
   - as well as the **loss** function used to ultimately evaluate the decision.

My argument today

- Under-appreciated fact:
  - Statistical decision theory cannot justify
  - randomization or pre-registration for experiments!
- **Randomization**:
  - Expected loss of any randomized procedure
  - equals the **average** of the loss of the deterministic procedures averaged over.
  - This is dominated by the expected loss of the **best** deterministic procedure.
- **Pre-registration**:
  - Expected loss is **dynamically consistent**.
  - The decision maker's future self has no reason to cheat their past self.
  - Therefore there is no need for a **commitment device**.

Possible takeaways

1. The methodological conventions of experimentation need to be amended to bring them in line with decision theory.
2. Decision theory needs to be amended to make sense of the existing methodological conventions.
3. Both?

I will argue:

- We need new **social foundations** for statistics.
- The framework of **mechanism design** is useful for this purpose: Staying close to decision theory, while incorporating conflicts of interest, private information, etc.

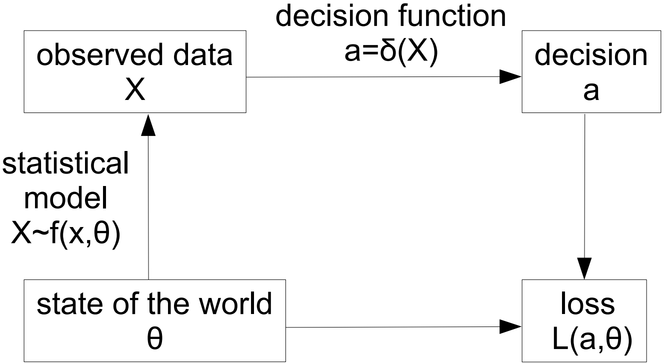
Introduction

Decision theory

- No randomization
- No commitment

Beyond decision theory

Decision theory – General setup



Notions of risk

- **Risk function:**
  Expected loss, averaging over sampling distribution, given the state of the world:
  $$R(\delta, \theta) = E_\theta[L(\delta(X), \theta)].$$
- **Bayes risk:**
  Average of risk function over some prior distribution (i.e., decision weights):
  $$R(\delta, \pi) = \int R(\delta, \theta)\pi(\theta) d\theta.$$
- **Worst case risk:**
  Maximum of risk function, over some set of θ, given $\delta(\cdot)$:
  $$\overline{R}(\delta) = \sup_{\theta \in \Theta} R(\delta, \theta).$$

No randomization

- We can allow δ to depend on some independent **randomization device** U: $a = \delta(X, U)$, where
$$P(U = u) = p_u$$
for $u = 1, \ldots, k$.
- Denote δ^u the deterministic decision rule $a = \delta(X, u)$.
- It follows from the definitions that
$$\begin{aligned} R(\delta,\theta) &= p_1 \cdot R(\delta^1,\theta) + \ldots + p_k \cdot R(\delta^k,\theta),\\ R(\delta,\pi) &= p_1 \cdot R(\delta^1,\pi) + \ldots + p_k \cdot R(\delta^k,\pi)\\ \overline{R}(\delta) &= p_1 \cdot \overline{R}(\delta^1) + \ldots + p_k \cdot \overline{R}(\delta^k).\end{aligned}$$
- **Averages** (over U) **are not as good as best cases**. Thus
$$R(\delta,\pi) \geq \min_u R(\delta^u,\pi),$$
$$\overline{R}(\delta) \geq \min_u \overline{R}(\delta^u).$$

No commitment

- Two alternatives:
  1. We can commit to (**pre-register**) a rule $\delta(\cdot)$ before observing X.
  2. We can pick $\delta(X)$ after observing X.
- By the **law of iterated expectations**

$$\begin{aligned} R(\delta, \pi) &= E[L(\delta(X), \theta)] \\ &= E[E[L(\delta(X), \theta)|X]] \\ &= \sum_x E[L(\delta(x), \theta)|X = x] \cdot P(X = x). \end{aligned}$$

- Therefore:
  - Picking the optimal $\delta(\cdot)$ (to minimize the sum) is the same
  - as picking the optimal $\delta(x)$ for every value of x (each term of the sum).
  - The decision-problem is **dynamically consistent**.

Some comments

- These arguments just rely on basic properties of probability.
- They don't depend on properties
    - of L (such as risk aversion),
    - or of the action space (such as finiteness or convexity).
- The arguments are
    - Immediate for Bayes risk,
    - but also true (if a little more subtle) for worst case ("frequentist") risk.

Introduction

Decision theory

- No randomization
- No commitment

Beyond decision theory

Beyond single-agent decision theory

- A key benefit of decision theory: It forces us to be explict about
  1. Objectives (loss function),
  2. possible actions that are considered,
  3. assumptions about data generation.
- But it cannot address important scientific and societal challenges:
  1. Replication crisis, publication bias, p-hacking, pre-registration, reforms of statistics teaching and the publication system.
  2. Artificial intelligence, questions of discrimination and inequality, value alignment.
- We need to recognize that
  1. Different agents have different (conflicting) objectives.
  2. The objectives of statistical inference or machine learning algorithms are socially determined.

A mechanism design perspective

- Statistical inference as a mechanism design problem.
- Take the perspective of a reader of empirical research who wants to implement a statistical decision rule (mapping from full data to a decision).
- Not all rules are implementable when researchers have divergent interests and private information about the data, and they can selectively report to readers.
- Agenda: Characterize optimal decision rules subject to implementability.

Applied to the question of pre-registration

- Work in progress with Jann Spiess:
  - Pre-analysis plans can be rationalized with multiple parties, conflicts of interest, and costly communication / asymmetric information.
  - We consider (optimal) statistical decision rules subject to the constraint of implementability.
- Our model:
  1. A journal commits to a publication / testing rule,
  2. then a researcher commits to a pre-analysis plan,
  3. then observes the data, reports selected statistics to the journal,
  4. which then applies the publication / testing rule.
- Pre-analysis plans are optimal when
  - there are many researcher degrees of freedom,
  - and/or communication costs are high.

References

- Kasy, M. (2016).
  Why experimenters might not always want to randomize, and what they could do instead.
  Political Analysis, 24(3):324–338.

- Kasy, M. and Spiess J. (2021).
  Pre-analysis plans and mechanism design.
  Work in progress.

More references

- Kasy, M. (2021)
  Of forking paths and tied hands:
  Selective publication of findings, and what economists should do about it.
  Journal of Economic Perspectives 35(2)

- Andrews, I. and Kasy, M. (2019).
  Identification of and correction for publication bias.
  American Economic Review, 109(8):2766–94.

- Frankel, A. and Kasy, M. (2021).
  Which findings should be published?
  American Economic Journal: Microeconomics, Forthcoming.

Thank you!