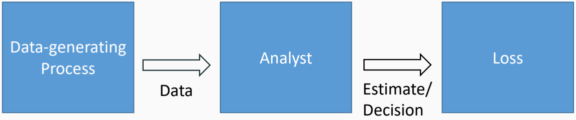
A Model of Scientific Communication

Isaiah Andrews (Harvard)
Jesse Shapiro (Brown)

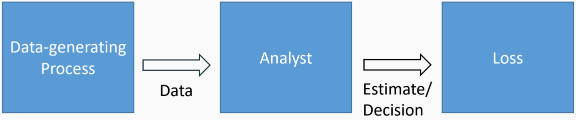
Classical Model of Statistics (Wald 1950)

- Analyst observes data $X \in \mathcal{X}$
- Uses X to form estimate of unknown parameter $\theta \in \Theta$
- Estimate is "good" if close to true value of parameter
- Formalized by imagining a decision problem in which
  - estimate is a decision $d \in \mathcal{D}$
  - want to minimize loss $L(d, \theta)$
- Dominant paradigm for point estimation
  - e.g., $L(d, \theta) = (d - \theta)^2$ gives MSE criterion
  - Foundation of most optimality claims

Classical Model of Statistics (Wald 1950)

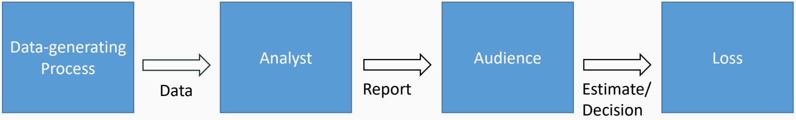


Classical Model of Statistics (Wald 1950)

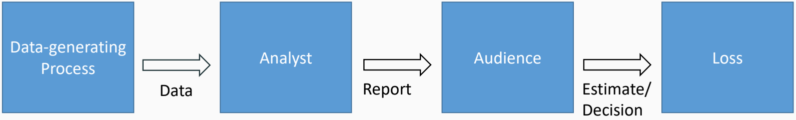


- Example: Analyst works for a firm that must make a pricing decision

Alternative Model of Statistics in Science



Alternative Model of Statistics in Science



- Example: Analyst reports to scientists with diverse opinions, policymakers with diverse objectives

Today

- Argue that these two models represent fundamentally different, and at times conflicting, views of the analyst's goal
- Can lead to very different recommended procedures
- (Time permitting) Discuss possible implications for empirical research

Setting

Timing

- Analyst publicly commits to a rule $c : \mathcal{X} \to \mathcal{D}$
- Analyst observes data $X \in \mathcal{X}$, where $X \sim F_\theta$
- Analyst makes report $c(X)$ to an audience $\mathcal{A}$
- Each agent $a \in \mathcal{A}$ selects decision d and realizes loss $L(d, \theta)$

Audience

- Agents $a \in \mathcal{A}$ have different priors on θ
  - Write $\mathsf{E}_a[\cdot]$ for expectation under a's prior
  - Identify each agent with their prior, so $\mathcal{A} \subseteq \Delta(\Theta)$
- All disagreement expressed via priors
- Paper shows that nests cases with disagreement over
  - Loss function L
  - Likelihood F_θ

Analyst's Goal

- Analyst tries to minimize expected loss (i.e. *risk*) for the agents
  - Benevolent analyst: no conflict of interest between analyst and agents
- Consider two possible definitions for the risk of rule c for agent a
  - *Decision risk (classical model)*

$$\mathrm{E}_a \left[L \left(c \left(X \right), \theta \right) \right],$$

    as if analyst makes decision on agent's behalf
  - *Communication risk (alternative model)*

$$\mathrm{E}_a \left[\min_{d \in \mathcal{D}} \mathrm{E}_a \left[L \left(d, \theta \right) | c \left(X \right) \right] \right],$$

    as if agent makes optimal decision given report

Analyst's Goal

- In special case of squared-error loss $L(d,\theta) = (d-\theta)^2$
  - *Decision risk (classical model)*

$$\mathrm{E}_a\left[L(c(X),\theta)\right] = \mathrm{E}_a\left[(c(X)-\theta)^2\right],$$

    is mean squared error
  - *Communication risk (alternative model)*

$$\mathrm{E}_a\left[\min_{d\in\mathcal{D}} \mathrm{E}_a\left[L(d,\theta)\,|c(X)\right]\right] = \mathrm{E}_a\left[\mathrm{Var}_a(\theta|c(X))\right],$$

    is expected posterior variance
- Decision/communication distinction irrelevant when $|\mathcal{A}| = 1$
  - Benevolent analyst will pick $c(X) = \mathrm{E}_a[\theta|X]$, so coincide
- Distinction can matter when $|\mathcal{A}| > 1$

Example

Example

- Analyst conducts a randomized trial with a binary outcome
- Goal is to learn the success probability $\theta = (\theta_1, ..., \theta_J)$ at each of a finite set of ordered treatments $\{1, ..., J\}$
  - e.g., Probability of purchase at a set of prices
  - e.g., Probability of callback at a set of unemployment spell lengths
- Success probabilities known to be decreasing, $\theta_1 \geq \theta_2 \geq ... \geq \theta_J$
  - e.g., Demand slopes down
  - e.g., Longer unemployment spells deter employers
- Quadratic loss $L(d, \theta) = \sum_j \left(d_j - \theta_j\right)^2$

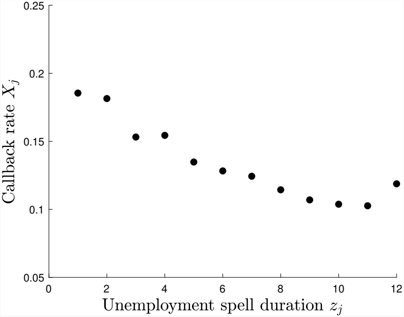
Example

- n independent observations for each treatment
- Data $X = (X_1, ..., X_J)$ are fraction of successes for each
- Decision space $\mathcal{D} = \mathcal{X}$ rich enough to communicate full data
- Audience $\mathcal{A} = \Delta(\Theta)$ includes all possible priors
  - Everyone agrees that $\theta_j \geq \theta_{j+1}$ for all j
  - ...but may disagree about everything else

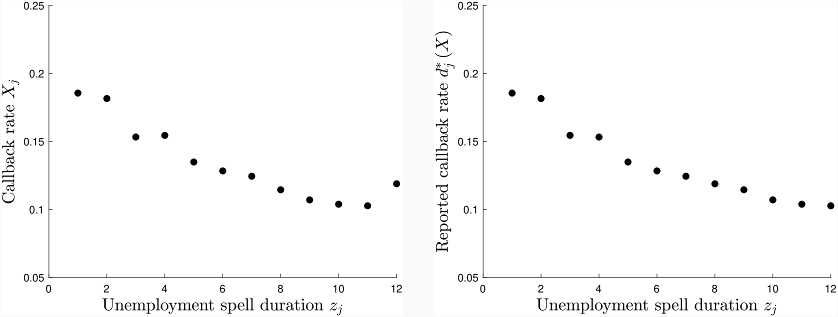
Two Rules

- Consider two possible rules
  - Full data: $c_j(X) = X_j$
    - Reports success fraction for each treatment j
  - Rearranged data: $c_j^*(X) =$ jth highest element of $\{X_1, ..., X_J\}$
    - Sorts success fractions in descending order

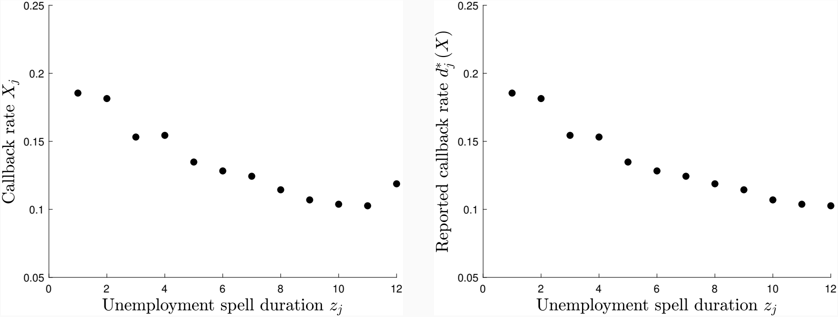
Illustration



Illustration

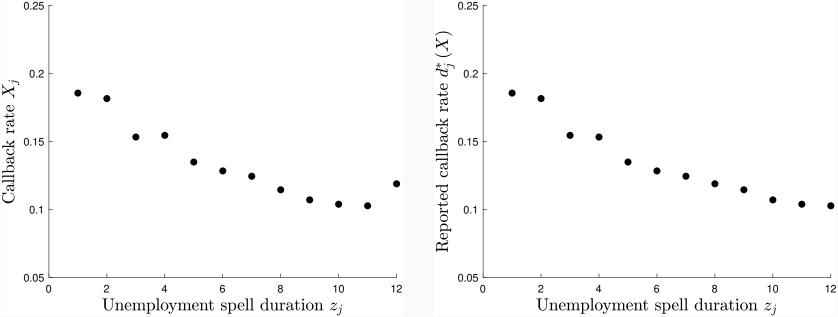


Decision Risk Perspective

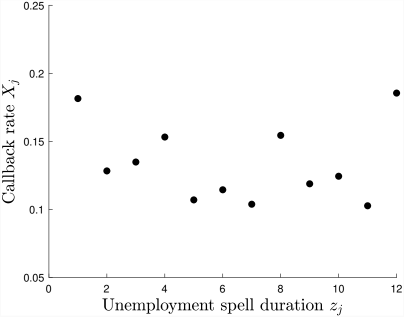


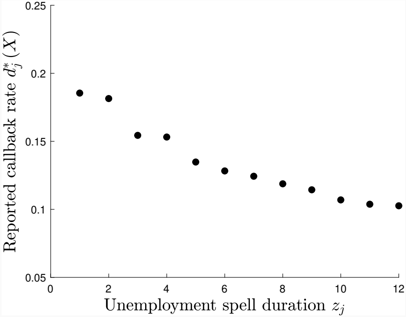
- Rearranged data c^* dominates full data c in decision risk
  - Achieves weakly lower risk for all agents, strictly lower for some
  - Intuitively, gets closer to true parameter
  - cf. Chernozhukov et al. (2009)
- Classical model would recommend c^* over c

Illustration

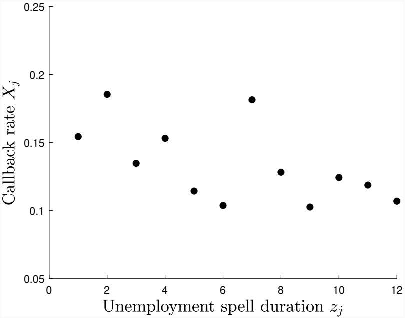


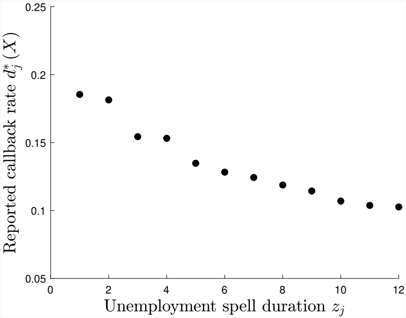
Illustration



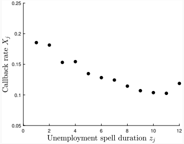


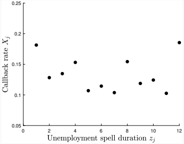
Illustration

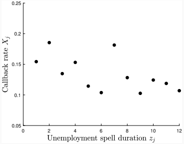


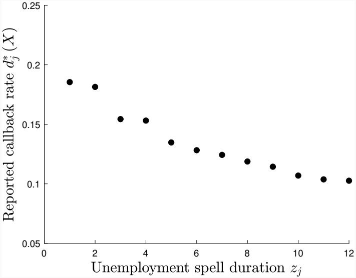


Illustration

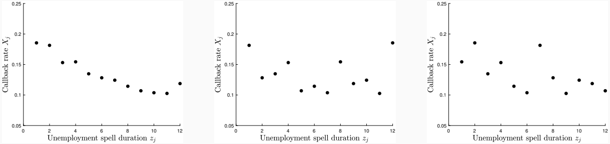


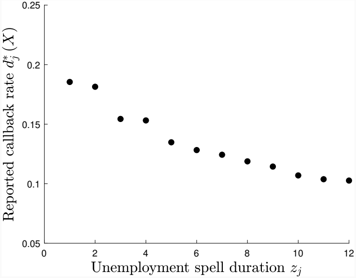






Communication Risk Perspective





- Full data c dominates rearranged data c^* in communication risk
  - Intuitively, preserves decision-relevant information

Conflict in Admissibility

- So far, we've shown that different models made different selections from the pair of rules $\{c, c^*\}$
- A stronger statement is true
- **Definition**: A rule is admissible (in a given notion of risk) if it is not dominated by another rule
- In this example, any rule that is admissible in decision risk is inadmissible in communication risk, and vice versa
    - No choice of rule resolves conflict between two notions of risk

Takeaways

- Shows conflict between goals of decision and communication
- *Recommendations of classical model may not achieve goals of scientific analyst who cares about communication*
- In this example, communication-optimal rules seem more in line with empirical practice
    - e.g. we're not aware of any unemployment audit studies that report only the sorted data, though many report unsorted results
    - Kroft, Lange, and Notowidigdo (2013) report both unsorted and sorted versions

Generalizations, and Implications

Generalizations

- Paper considers more general settings
  - Focus on discrete $\mathcal{X}$, Θ, $\mathcal{D}$
  - Results for some continuous cases in supplement
- Provide sufficient conditions for admissibility conflict
- Intuition is the same: good decision rules discard useful information

Generalizations

- We also provide results for other optimality criteria
    - Weighted average of risk over the audience
    - Worst-case risk over the audience

Generalizations

- We also provide results for other optimality criteria
  - Weighted average of risk over the audience
  - Worst-case risk over the audience
- Negative results extend to weighted average case
- For worst case risk, get a positive result

Implications for Practice

- In example, analyst concerned with communication can solve problem by reporting X
- Doesn't seem fully satisfactory in general
  - Otherwise, why does anyone write papers?
- Suggests communication or information processing constraints
- Raises question of optimal constrained communication
  - Optimal rules will depend on details of how model constraints
- Less ambitious: short of optimal rules, can we find simple, practical ways for analyst to reduce communication risk?
  - Andrews, Gentzkow, Shapiro (2020), "Transparency in Structural Research" discusses a range of practices
  - e.g. showing sensitivity to misspecification in the spirit of Conley, Hansen, and Rossi (2012), Andrews, Gentzkow, and Shapiro (2017)

Summary

- Focusing on communication rather than decision-making changes understanding of the goals of empirical scientist
- Leads to very different recommendations than classical decision-theoretic model in some cases
- Hope that change in perspective may help suggest good procedures for communicating scientific results

Thank you!