Algorithmic Bayesian Epistemology

Eric Neyman

January 24, 2024

Supervised by Tim Roughgarden

Outline

- What is algorithmic Bayesian epistemology?
- Technical content
    1. Incentivizing precise forecasts
    2. Arbitrage-free contract functions
    3. Quasi-arithmetic pooling
    4. Learning weights for logarithmic pooling
    5. Robust aggregation of substitutable signals
    6. When does agreement imply accuracy?
    7. Deductive circuit estimation

Outline

- What is algorithmic Bayesian epistemology?
- Technical content
    1. Incentivizing precise forecasts
    2. Arbitrage-free contract functions
    3. Quasi-arithmetic pooling
    4. Learning weights for logarithmic pooling
    5. Robust aggregation of substitutable signals
    6. When does agreement imply accuracy?
    7. Deductive circuit estimation
- Conclusion: the most exciting questions in ABE

The blue chapters are my favorite contributions! They will be accompanied by "future directions" slides.

Outline

- **What is algorithmic Bayesian epistemology?**
- Technical content
    1. Incentivizing precise forecasts
    2. Arbitrage-free contract functions
    3. Quasi-arithmetic pooling
    4. Learning weights for logarithmic pooling
    5. Robust aggregation of substitutable signals
    6. When does agreement imply accuracy?
    7. Deductive circuit estimation
- Conclusion: the most exciting questions in ABE

Bayesian epistemology

- **Epistemology:** the study of knowledge and uncertainty
- **Bayesian** epistemology: a formal approach to epistemology that interprets beliefs as subjective probabilities over outcomes
    - Observer assigns probabilities to uncertain events and updates those probabilities in light of new evidence

The algorithmic lens

- Many theoretical questions are optimization questions: *what is the optimal solution to problem X?*
- Example: welfare-maximizing auctions
  - Economist's solution: VCG (find optimal allocation, charge bidders their externalities)
  - Computer scientist's complaint: this can't be done efficiently!
    - Algorithmic lens: how can you achieve **approximately optimal welfare** in **polynomial computation + communication?**
- Instead of finding the *optimal* solution, looking for **satisfactory solutions** that **adhere to real-world constraints**
  - Coined at Berkeley by members of Theory of Computing research group *c.* 2000

Algorithmic Bayesian epistemology (ABE)

- The application of the algorithmic lens to Bayesian epistemology

*A question belongs to the field of **algorithmic Bayesian epistemology (ABE)** if it involves reasoning about uncertainty from a Bayesian perspective, but under constraints that prevent complete assimilation of all existing information.*

What kinds of constraints?

- Computational constraints
  - Approximating Bayesian inference
- Informational constraints
  - Forecast aggregation under incomplete information
- Communication constraints
  - Agreement protocols
- Strategic constraints
  - Prediction markets

Outline

- What is algorithmic Bayesian epistemology?
- **Technical content**
    1. **Incentivizing precise forecasts**
    2. Arbitrage-free contract functions
    3. Quasi-arithmetic pooling
    4. Learning weights for logarithmic pooling
    5. Robust aggregation of substitutable signals
    6. When does agreement imply accuracy?
    7. Deductive circuit estimation
- Conclusion: the most exciting questions in ABE

Background on proper scoring rules

- You want to elicit the probability that it will rain tomorrow from an expert
  - A **scoring rule** is a way of paying the expert depending on their forecast and whether or not it rains
  - A scoring rule s is **proper** if the expert is incentivized to report their true belief
- **Example 1:** quadratic (Brier) scoring rule
  - Penalty based on expert's squared error
  - If expert says 70% chance of rain, penalty = 0.3^2 if it rains, 0.7^2 if it doesn't
- **Example 2:** logarithmic scoring rule
  - Score = log of probability assigned to outcome
  - If expert says 70% chance of rain, score = $\ln(0.7)$ if it rains, $\ln(0.3)$ if it doesn't

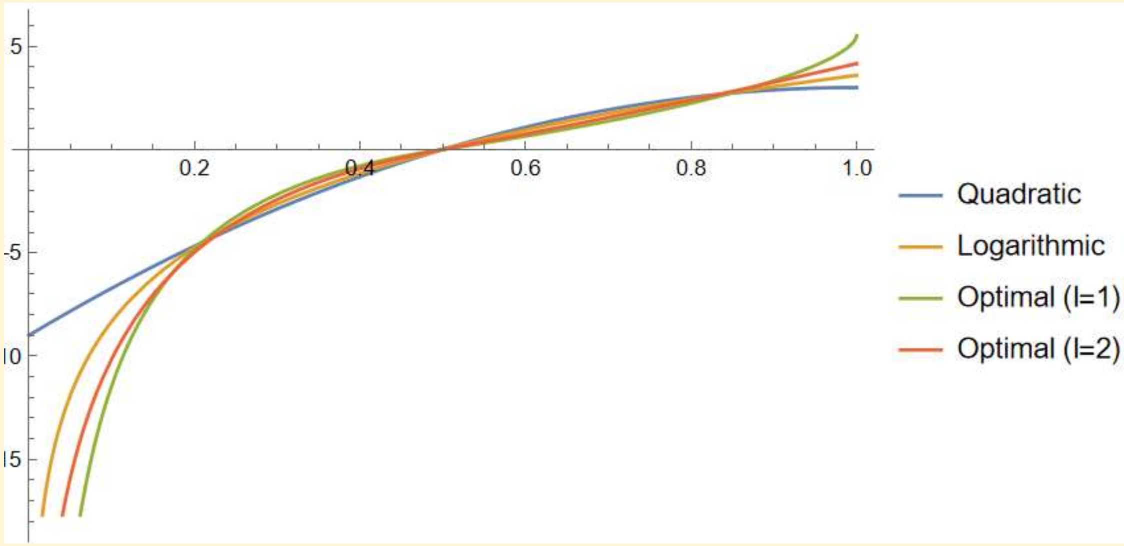
Incentivizing precision (1/2)

(Joint work with George Noarov and Matt Weinberg)

- All proper scoring rules incentivize *accuracy* in forecasts, but what about *precision?*
  - Which proper scoring rule most incentivizes experts to do research before reporting a forecast?
- Coin with bias uniformly chosen from $[0,1]$
  - Expert can flip coin at small cost c per flip
  - Expert will report a forecast and be scored according to a scoring rule
- Which proper scoring rule incentivizes expert to flip the coin a lot?
  - As $c \to 0$, which proper scoring rule minimizes expert’s expected error?
- We define an **incentivization index** to measure this
  - And then optimize the index

Incentivizing precision (2/2)

(Joint work with George Noarov and Matt Weinberg)



Quadratic and logarithmic scoring rules, together with optimal scoring rules for minimizing expected absolute error and squared error

Outline

- What is algorithmic Bayesian epistemology?
- **Technical content**
  1. Incentivizing precise forecasts
  2. **Arbitrage-free contract functions**
  3. Quasi-arithmetic pooling
  4. Learning weights for logarithmic pooling
  5. Robust aggregation of substitutable signals
  6. When does agreement imply accuracy?
  7. Deductive circuit estimation
- Conclusion: the most exciting questions in ABE

Arbitrage-free contract functions

(Joint work with Tim Roughgarden)

- Now suppose we have multiple experts. Experts can collude!
  - E.g. if s is the quadratic scoring rule, and 3 experts believe 40%, 50%, 90%, they can all say 60% and profit, no matter the outcome ("arbitrage")
- What if experts' scores are allowed to depend on *other experts'* reports?
  - This is called a *contract function*
- Do any contract functions prevent all arbitrage opportunities?
  - Yes! For expert $i \in [m]$, if $\overline{\boldsymbol{p}}_{-i}$ is average of other experts' reports, i's reward is

$$\Pi_i(\boldsymbol{P}; j) = \underbrace{s_{\text{quad}}(\boldsymbol{p}_i; j) - (m-1)^2 s_{\text{quad}}(\overline{\boldsymbol{p}}_{-i}; j)}_{\text{Total reward depends only on average of all reports}} + \underbrace{\alpha \overline{\boldsymbol{p}}_{-i,j}}_{\text{Total score is lower for some outcome}}$$

Outline

- What is algorithmic Bayesian epistemology?
- **Technical content**
  1. Incentivizing precise forecasts
  2. Arbitrage-free contract functions
  3. **Quasi-arithmetic pooling**
  4. Learning weights for logarithmic pooling
  5. Robust aggregation of substitutable signals
  6. When does agreement imply accuracy?
  7. Deductive circuit estimation
- Conclusion: the most exciting questions in ABE

Quasi-arithmetic pooling (1/2)

(Joint work with Tim Roughgarden)

- After eliciting forecasts from multiple experts, how should the aggregator combine them?
  - Intuition: should depend on scoring rule. Different scoring rules incentivize precision in different ways, e.g. log score incentivizes precision near extreme probabilities (compared to quadratic score).
- We define an aggregation method called *quasi-arithmetic (QA) pooling* (with respect to a given proper scoring rule) that takes this into account
  - QA pooling averages forecasts based on the experts' preferences over outcomes (induced by the scoring rule)

Quasi-arithmetic pooling (2/2)

(Joint work with Tim Roughgarden)

- Nice properties of QA pooling
  - Maps two most popular scoring rules (quadratic and logarithmic) to two most well-studied pooling methods (linear and logarithmic)
  - Max-min optimality: maximizes the worst-case improvement over a random expert
  - Learning expert weights: QA pooling allows for experts to have weights. The score of a QA pool of experts is concave in the experts' weights. So if s is bounded, weights for QA pooling can be no-regret learned efficiently.
  - Ties together two notions of overconfidence
  - Axiomatization: the space of QA pools (one per scoring rule) corresponds precisely to the space of pooling methods obeying a natural list of axioms

QA pooling: Future directions

- What about QA pooling with weights adding to > 1?
  - Makes sense if experts have pretty different information sources
  - Do our results generalize to arbitrary weights?
- **Bayesian justifications** of "generalized" QA pooling
  - I.e. an information structure in which generalized QA pooling is exactly correct
  - Satopää et al. (2017) give a Bayesian justification for generalized *linear* pooling
  - I give a Bayesian justification for generalized *logarithmic* pooling
  - Is there a Bayesian justification for generalized QA pooling ***for all s***?

Outline

- What is algorithmic Bayesian epistemology?
- **Technical content**
    1. Incentivizing precise forecasts
    2. Arbitrage-free contract functions
    3. Quasi-arithmetic pooling
    4. **Learning weights for logarithmic pooling**
    5. Robust aggregation of substitutable signals
    6. When does agreement imply accuracy?
    7. Deductive circuit estimation
- Conclusion: the most exciting questions in ABE

Learning weights for log pooling

(Joint work with Tim Roughgarden)

- Recall: "**If *s* is bounded**, weights for QA pooling can be no-regret learned efficiently"
- But what about the log scoring rule?
  - We show how to do no-regret learning of experts' weights even in this setting, provided that experts are *calibrated*
  - We use a modification of the online mirror descent (OMD) algorithm, with the Tsallis entropy regularizer

Outline

- What is algorithmic Bayesian epistemology?
- **Technical content**
  1. Incentivizing precise forecasts
  2. Arbitrage-free contract functions
  3. Quasi-arithmetic pooling
  4. Learning weights for logarithmic pooling
  5. **Robust aggregation of substitutable signals**
  6. When does agreement imply accuracy?
  7. Deductive circuit estimation
- Conclusion: the most exciting questions in ABE

Robust aggregation of substitutable signals (1/3)

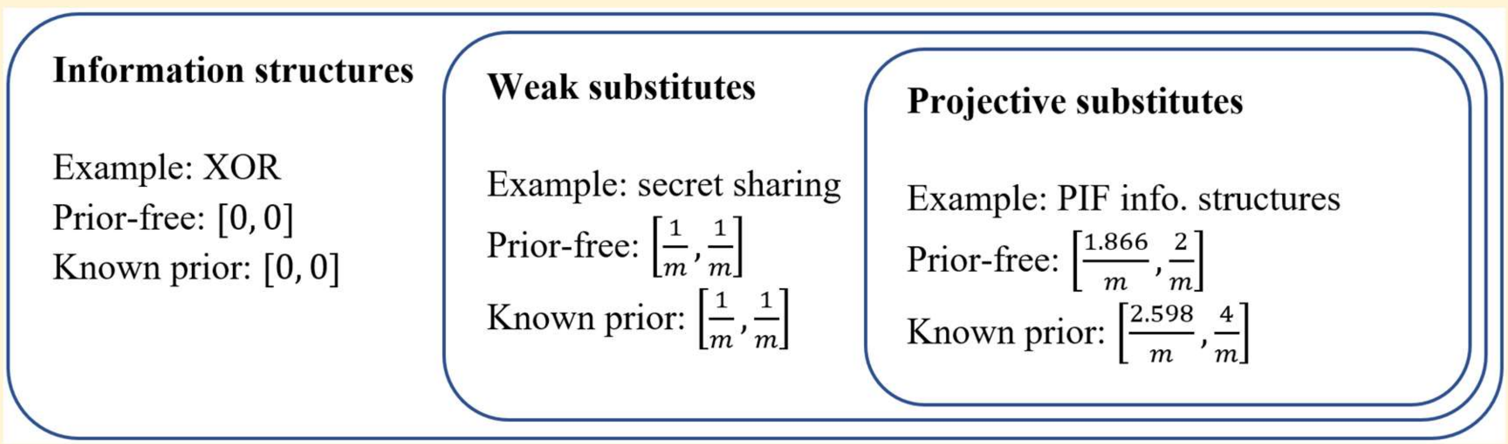
(Joint work with Tim Roughgarden)

- Alice says 60%, Bob says 75%, what should aggregator say?
  - If Alice is strictly more informed: 60%. If Bob is more informed: 75%.
  - If they are updating from 50/50 with conditionally independent evidence: 82%
  - The “right answer” could be anything – it depends!
  - “Robust” solution concept: **worst case over information structures**
- Goal: compete with “perfect” aggregator who knows all experts’ information
  - Problem: “XOR information structure” – Alice receives $a \in \{0,1\}$, Bob receives $b \in \{0,1\}$, answer is $a \oplus b$
  - Alice and Bob both say 0.5; aggregator can’t do any better
  - Restrict space of allowed information structures?

Robust aggregation of substitutable signals (2/3)

(Joint work with Tim Roughgarden)

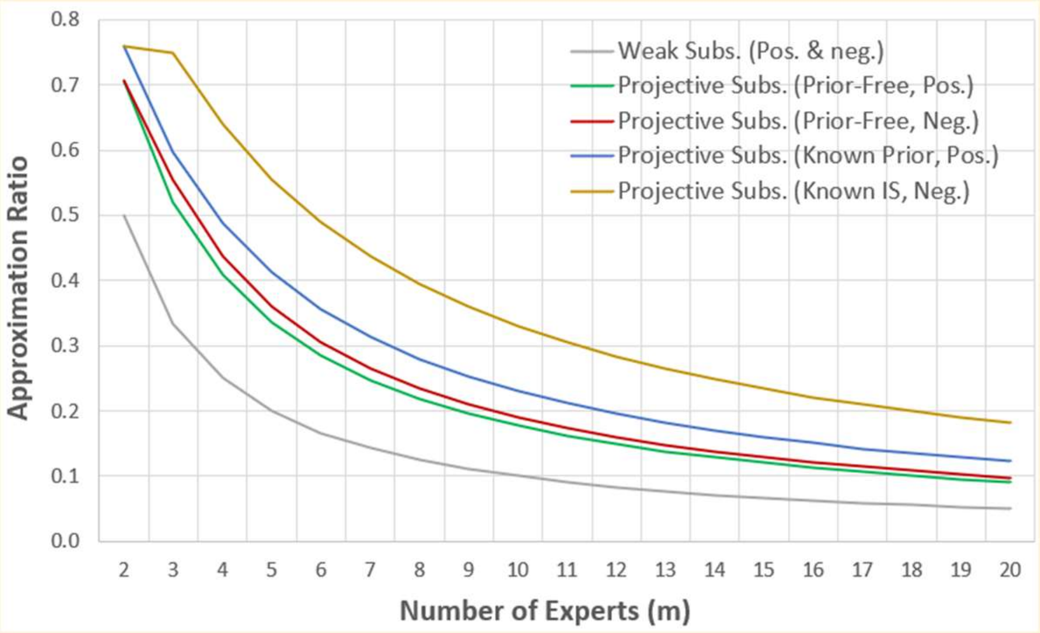
- We explore *informational substitutes* (roughly: experts' information is substitutable rather than complementary)
  - Standard notion: weak substitutes (Chen & Waggoner, 2016)
  - To get nontrivial results, we give a stronger notion: *projective substitutes*



Our bounds on what approximation guarantees are attainable (m = # experts)

Robust aggregation of substitutable signals (3/3)

(Joint work with Tim Roughgarden)



Robust aggregation: Future directions

- Generalizing robust aggregation results beyond squared error
  - E.g. KL divergence (more appropriate for probabilistic forecasts)
- “Exploring the playground” of robust aggregation
  - Which loss function?
  - Assumptions about information structure?
  - What does the aggregator know?
  - What does the aggregator learn from the experts?
  - Experts truthful or strategic?
  - Benchmark?

Outline

- What is algorithmic Bayesian epistemology?
- **Technical content**
  1. Incentivizing precise forecasts
  2. Arbitrage-free contract functions
  3. Quasi-arithmetic pooling
  4. Learning weights for logarithmic pooling
  5. Robust aggregation of substitutable signals
  6. **When does agreement imply accuracy?**
  7. Deductive circuit estimation
- Conclusion: the most exciting questions in ABE

Does agreement imply accuracy?

(Joint work with Raf Frongillo and Bo Waggoner)

- Alice and Bob have different information, leading to different beliefs
- Can efficiently exchange information in order to reach agreement?
  - Yes! (Aaronson 2004)
  - But the agreed-upon value might not be *accurate* (might be different from their belief if they exchanged *all* information)
- Are there natural sufficient conditions under which agreement implies accuracy?
  - Yes! A (different) notion of informational substitutes

Outline

- What is algorithmic Bayesian epistemology?
- **Technical content**
  1. Incentivizing precise forecasts
  2. Arbitrage-free contract functions
  3. Quasi-arithmetic pooling
  4. Learning weights for logarithmic pooling
  5. Robust aggregation of substitutable signals
  6. When does agreement imply accuracy?
  7. **Deductive circuit estimation**
- Conclusion: the most exciting questions in ABE

Deductive circuit estimation (1/3)

(Joint work with Paul Christiano, Jacob Hilton, Václav Rozhoň, and Mark Xu)

- How can you estimate the acceptance probability of a boolean circuit?
  - Obvious answer: sampling random inputs (or MCMC, etc.)
    - These are based on *inductive reasoning* about the circuit
  - Less obvious answer: *deductive* reasoning about the structure of the circuit
    - Ex. 1: $C(a, b, c) = 1$ if $\max(a, b) = \max(b, c)$
      - Reasoning: $b \geq a, c$ w.p. 1/3
    - Ex. 2: $C(x)$ computes SHA-256(x), returns 1 if first 128 bits > last 128 bits
      - Reasoning: can think of output of SHA-256 as independent random bits → ½
    - Ex. 3: C is a particular 3CNF with k clauses
      - Reasoning: on average, circuits with this structure have acceptance probability $\left(\frac{7}{8}\right)^k$
    - Ex. 4: C takes integer $k \in [e^{100}, e^{101}]$, outputs 1 if $k, k+2$ are both prime
      - Reasoning: density of primes $\approx 1\%$ → $(1\%)^2 = 0.01\%$
      - Better reasoning: k is prime → k is odd → $k + 2$ is odd → $k + 2$ more likely to be prime → 0.02%
      - Can refine this further, e.g. by considering that k is not divisible by 3

Deductive circuit estimation (2/3)

(Joint work with Paul Christiano, Jacob Hilton, Václav Rozhoň, and Mark Xu)

- Why deductive reasoning?
  - Helps you understand **why** a circuit has a certain acceptance probability
  - Can notice **different reasons** for acceptance
    - Recall $C(a, b, c) = 1$ if $\max(a, b) = \max(b, c)$. Can notice $b \geq a, c$ vs. $a = c$
- Our goal: create a **deductive estimation algorithm:**
  - Input: boolean circuit C, set of deductive arguments about C
  - Output: estimate of C's acceptance probability based on those arguments
- Somewhat analogous to proof verification
  - Input to proof verifier: statement and alleged proof. Output: accept/reject.
  - It's not trying to find the proof, only assess the given proof!
  - Similarly, we aren't trying to find deductive arguments, just assess and incorporate the given ones.

Deductive circuit estimation (3/3)

(Joint work with Paul Christiano, Jacob Hilton, Václav Rozhoň, and Mark Xu)

- Our notation for deductive estimation algorithm: $G(\,C \mid \Pi\,)$
  - C is circuit; $\Pi = \{\pi_1, \dots, \pi_m\}$ is the set of arguments
- Desiderata for G:
  - **Linearity:** $G(\,C \mid \Pi\,) = \frac{1}{2}\big(G(\,C[x_i = 1] \mid \Pi\,) + G(\,C[x_i = 0] \mid \Pi\,)\big)$
  - **Respect for proofs:** a proof that $\lambda_1 \mathbb{E}[C_1] + \cdots + \lambda_k \mathbb{E}[C_k] \leq b$ can be turned into an argument π such that

$$\lambda_1 G(\,C_1 \mid \pi\,) + \cdots + \lambda_k G(\,C_k \mid \pi\,) \leq b$$

  - **0-1 boundedness:** $0 \leq G(\,C \mid \Pi\,) \leq 1$ for all C, Π
- We give an algorithm G that satisfies linearity + respect for proofs…
  - …but show that no polynomial-time G can satisfy all three (assuming $P \neq PP$)
- We discuss potential further desiderata for G

Deductive circuit estimation: Future directions

- Design a good deductive estimation algorithm!
  - Figure out what “good” means (state some formal desiderata)
  - Find an algorithm that satisfies those desiderata
- I find this problem compelling for two reasons
  - Seems like a fundamental theory problem (understanding and formalizing deductive argumentation)
  - Potentially useful for the AI alignment problem

Outline

- What is algorithmic Bayesian epistemology?
- Technical content
  1. Incentivizing precise forecasts
  2. Arbitrage-free contract functions
  3. Quasi-arithmetic pooling
  4. Learning weights for logarithmic pooling
  5. Robust aggregation of substitutable signals
  6. When does agreement imply accuracy?
  7. Deductive circuit estimation
- **Conclusion: the most exciting questions in ABE**

Conclusion: The most exciting questions in ABE

- I've given three highlights:
  - Finding Bayesian justifications for generalized QA pooling
  - Further investigating robust aggregation (e.g. w.r.t. KL divergence)
  - Finding a good deductive circuit estimation algorithm
- Some other exciting directions:
  - Sophisticated Bayesian models for forecast aggregation
  - Wagering mechanisms that produce good aggregate forecasts
- And many more!

Thank you!

- My advisor, Tim Roughgarden
- My undergraduate mentor, Matt Weinberg
- My research collaborators: Paul Christiano, Raf Frongillo, Jacob Hilton, George Noarov, Václav Rozhoň, Bo Waggoner, Mark Xu
- My friends, family, and communities