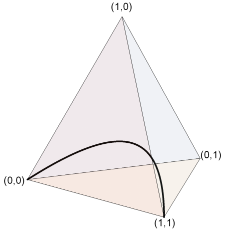
Multi-Observation Elicitation



Sebastian Casalaina-Martin Colorado
Rafael Frongillo Colorado
Tom Morgan Harvard
Bo Waggoner **UPenn**

July 2017

Background: Properties of distributions

Property or *statistic* of a probability distribution:

$$\Gamma : \Delta_{\mathcal{Y}} \to \mathcal{R}$$

Examples:

- $\Gamma(p) = \mathbb{E}_{Y\sim p} Y$ *mean*
- $\Gamma(p) = \sum_y p(y) \log \frac{1}{p(y)}$ *entropy*
- $\Gamma(p) = \mathrm{argmax}_y\, p(y)$ *mode*
- $\Gamma(p) = \mathbb{E}_{Y\sim p}(Y - \mathbb{E}\, Y)^2$ *variance*

Background: Elicitation (1)

If we minimize **expected loss**, what do we get?

Background: Elicitation (1)

If we minimize **expected loss** under a distribution p,
what **property** of p do we get?

$$r^* = \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{Y \sim p} \ell(r, Y)$$

minimize loss

Background: Elicitation (1)

If we minimize **expected loss** under a distribution p,
what **property** of p do we get?

$$r^* = \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{Y \sim p} \ell(r, Y) \qquad \text{minimize loss}$$

$$\Gamma(p) = \psi(r^*) \qquad \text{link}$$

Background: Elicitation (1)

If we minimize **expected loss** under a distribution p,
what **property** of p do we get?

$$r^* = \operatorname*{argmin}_{r \in \mathcal{R}} \; \mathop{\mathbb{E}}_{Y \sim p} \ell(r, Y) \qquad \text{minimize loss}$$

$$\Gamma(p) = \psi(r^*) \qquad \text{link}$$

Motivation: **statistically consistent** losses.

- Finite property space: classification, ranking, ...
- $\Gamma(p) \in \mathbb{R}^d$: regression, ...

Background: Elicitation (2)

If we minimize **expected loss** under a distribution p,
what **property** of p do we get?

$$r^* = \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{Y \sim p} \ell(r, Y) \qquad \text{minimize loss}$$

$$\Gamma(p) = \psi(r^*) \qquad \text{link}$$

Examples:

- **The mean** is elicited by **squared loss**.
- **Variance**: elicit mean and second moment, then link.
- **Any property** is a link from the *whole distribution* . . .
  but **dimension** of prediction r is unbounded. . .

This paper

What if the loss takes **multiple** i.i.d. observations?

$$r^* = \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{Y_1,\ldots,Y_m \sim p} \ell(r, Y_1, \ldots, Y_m)$$

This paper

What if the loss takes **multiple** i.i.d. observations?

$$r^* = \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{Y_1,\ldots,Y_m \sim p} \ell(r, Y_1, \ldots, Y_m)$$

Examples:

- $\mathrm{Var}(p) = \operatorname{argmin}_r \; \mathbb{E}\left(r - \frac{1}{2}(Y_1 - Y_2)^2\right)^2$.
- 2-norm: unbounded dimension $\rightarrow$ 1 dimension, 2 observations!

This paper

What if the loss takes **multiple** i.i.d. observations?

$$r^* = \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{Y_1,\ldots,Y_m \sim p} \ell(r, Y_1, \ldots, Y_m)$$

Examples:

- $\mathrm{Var}(p) = \mathrm{argmin}_r \; \mathbb{E}\left(r - \frac{1}{2}(Y_1 - Y_2)^2\right)^2$.
- 2-norm: unbounded dimension $\rightarrow$ 1 dimension, 2 observations!

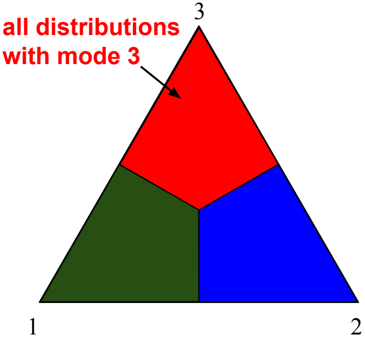
Motivating applications:

- Crowd labeling
- Numerical simulations *climate science, engineering, . . .*
- Regression?

Key concepts from prior research

Elicitable properties have convex **level sets**, linear structures.

Simplex on $\mathcal{Y} = \{1, 2, 3\}$:

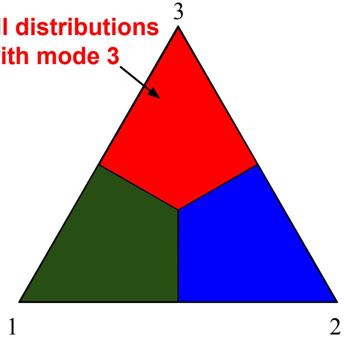


Mode

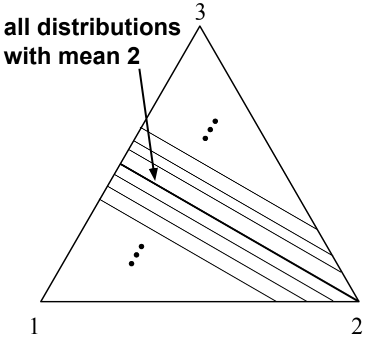
Key concepts from prior research

Elicitable properties have convex **level sets**, linear structures.

Simplex on $\mathcal{Y} = \{1, 2, 3\}$:



Mode

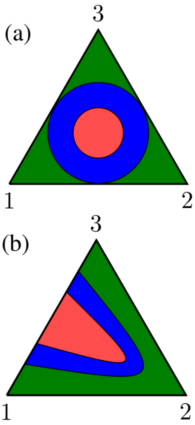


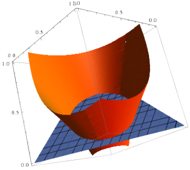
Mean

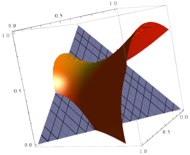
Results (1)

Geometric approach

Summary: k-observation level sets $\leftrightarrow$ zeros of degree-k polynomials







Results (2)

Upper and lower bounds.

Key example: (integer) k-**norm**$(p) = \left(\sum_y p(y)^k\right)^{1/k}$.

Idea: $\mathbf{1}[Y_1 = \cdots = Y_k]$ is an **unbiased estimator** for $\|p\|_k^k$.

Results (2)

Upper and lower bounds.

Key example: (integer) k-**norm**$(p) = \left(\sum_y p(y)^k\right)^{1/k}$.

Idea: $\mathbf{1}[Y_1 = \cdots = Y_k]$ is an **unbiased estimator** for $\|p\|_k^k$.

Loss$(r, Y_1, \ldots, Y_k) = \left(r - \mathbf{1}[Y_1 - \cdots = Y_k]\right)^2$.
Link$(r) = r^{1/k}$.

Results (2)

Upper and lower bounds.

Key example: (integer) k-**norm**$(p) = \left(\sum_y p(y)^k\right)^{1/k}$.

Idea: $\mathbf{1}[Y_1 = \cdots = Y_k]$ is an **unbiased estimator** for $\|p\|_k^k$.

Loss$(r, Y_1, \ldots, Y_k) = \left(r - \mathbf{1}[Y_1 - \cdots = Y_k]\right)^2$.
Link$(r) = r^{1/k}$.

- Similar approach for products of expectations.
- Lower bound: k-norm requires k observations.
- Lower bound approach is general (algebraic geometry).

Why could this be useful?

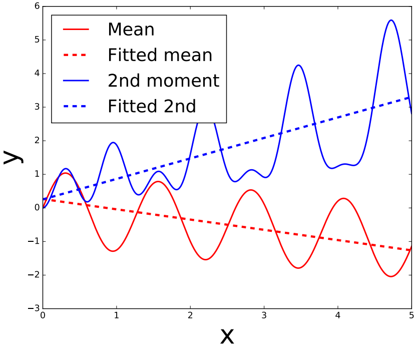
Problem: Regress x vs $\mathrm{Var}(y|x)$.

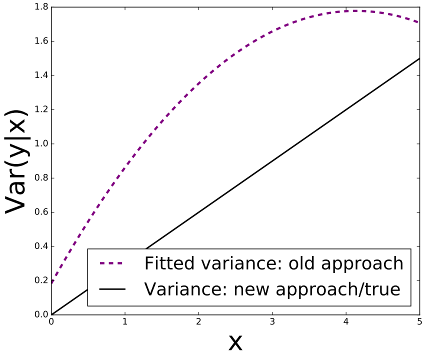
Why could this be useful?

Problem: Regress x vs $\mathrm{Var}(y|x)$.
Old approach: Regress on mean and second moment, then link.

Why could this be useful?

Problem: Regress x vs $\mathrm{Var}(y|x)$.
Old approach: Regress on mean and second moment, then link.





$\implies$ Requires good modeling and sufficient data for these (unimportant) proxies!

Future directions

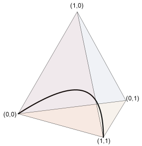
- **Elicitation frontiers** and (d, m)-elicitability

  In paper: central moments
- Regression

  In paper: preliminary results
- Additional useful examples

  e.g. expected max of k draws; risk measures
- Lots of COLT questions for multi-observation losses!

Thanks!



Aside - comparison to property testing

Property Testing

- **Algorithmic problem**
- Distribution p is initially unknown
- Algorithm draws samples to **estimate** property or **test** hypothesis

Aside - comparison to property testing

Property Testing

- **Algorithmic problem**
- Distribution p is initially unknown
- Algorithm draws samples to **estimate** property or **test** hypothesis

Property Elicitation

- **Existential questions**, e.g. . . .
- . . . does there exist a one-dim. loss function eliciting variance? *no*
- . . . two-dimensional? *yes*
- . . . describe all losses directly eliciting the mean *divergences*