Multi-Observation Elicitation

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

July 2017
Property or statistic of a probability distribution:

$$\Gamma : \Delta \mathcal{Y} \rightarrow \mathcal{R}$$

Examples:

- $$\Gamma(p) = \mathbb{E}_{Y \sim p} Y$$  \hspace{1cm} mean
- $$\Gamma(p) = \sum_y p(y) \log \frac{1}{p(y)}$$  \hspace{1cm} entropy
- $$\Gamma(p) = \operatorname{argmax}_y p(y)$$  \hspace{1cm} mode
- $$\Gamma(p) = \mathbb{E}_{Y \sim p} (Y - \mathbb{E} Y)^2$$  \hspace{1cm} variance
If we minimize expected loss, what do we get?
If we minimize \textbf{expected loss} under a distribution \( p \), what \textbf{property} of \( p \) do we get?

\[ r^* = \arg \min_{r \in \mathcal{R}} \mathbb{E}_{Y \sim p} \ell(r, Y) \]

\[ \Gamma(p) = \psi(r^*) \]

\textbf{Motivation:} statistically consistent losses.

Finite property space: classification, ranking, . . .

\[ \Gamma(p) \in \mathbb{R}^d: \text{regression, . . .} \]
If we minimize expected loss under a distribution $p$, what property of $p$ do we get?

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

minimize loss

$$\Gamma(p) = \psi(r^*)$$

link

Motivation: statistically consistent losses. Finite property space: classification, ranking, ... Regression, ...
If we minimize expected loss under a distribution $p$, what property of $p$ do we get?

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

minimize loss

$$\Gamma(p) = \psi(r^*)$$

Motivation: statistically consistent losses.

- Finite property space: classification, ranking, . . .
- $\Gamma(p) \in \mathbb{R}^d$: regression, . . .
If we minimize expected loss under a distribution $p$, what property of $p$ do we get?

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

minimize loss

$$\Gamma(p) = \psi(r^*)$$

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

\[ r^* = \arg\min_{r \in \mathbb{R}} \mathbb{E}_{Y_1, \ldots, Y_m \sim p} \ell(r, Y_1, \ldots, Y_m) \]
What if the loss takes multiple i.i.d. observations?

\[ r^* = \arg\min_{r \in \mathcal{R}} \mathbb{E}_{Y_1, \ldots, Y_m \sim p} \ell(r, Y_1, \ldots, Y_m) \]

Examples:

- \( \text{Var}(p) = \arg\min_r \mathbb{E} \left( r - \frac{1}{2}(Y_1 - Y_2)^2 \right)^2 \).
- 2-norm: unbounded dimension → 1 dimension, 2 observations!
This paper

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

\[ r^* = \arg\min_{r \in \mathcal{R}} \mathbb{E}_{Y_1, \ldots, Y_m \sim p} \ell(r, Y_1, \ldots, Y_m) \]

Examples:

- \( \text{Var}(p) = \arg\min_r \mathbb{E} \left( r - \frac{1}{2}(Y_1 - Y_2)^2 \right)^2 \).
- 2-norm: unbounded dimension \( \rightarrow \) 1 dimension, 2 observations!

Motivating applications:

- Crowd labeling
- Numerical simulations \( \text{climate science, engineering, \ldots} \)
- Regression?
Elicitable properties have convex **level sets**, linear structures.

Simplex on $\mathcal{Y} = \{1, 2, 3\}$:
Elicitable properties have convex **level sets**, linear structures.

Simplex on $\mathcal{Y} = \{1, 2, 3\}$:
**Results (1)**

**Geometric approach**

**Summary:** $k$-observation level sets $\leftrightarrow$ zeros of degree-$k$ polynomials
Results (2)
Upper and lower bounds.

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

Idea: \( 1[Y_1 = \cdots = Y_k] \) is an \textbf{unbiased estimator} for \( ||p||_k \).
Key example: (integer) \( k\)-norm \( p \) = \( \left( \sum_y p(y)^k \right)^{1/k} \).

Idea: \( 1[Y_1 = \cdots = Y_k] \) is an unbiased estimator for \( \|p\|_k \).

Loss \( (r, Y_1, \ldots, Y_k) = \left( r - 1[Y_1 = \cdots = Y_k] \right)^2 \).

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: $1[Y_1 = \cdots = Y_k]$ is an unbiased estimator for $\|p\|_k$.

Loss$(r, Y_1, \ldots, Y_k) = \left( r - 1[Y_1 - \cdots = Y_k] \right)^2$.

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 $\text{Var}(y|x)$.
Why could this be useful?

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

**Old approach:** Regress on mean and second moment, then link.
Why could this be useful?

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

**Old approach:** Regress on mean and second moment, then link.

⇒ 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**