Elicitation and Machine Learning

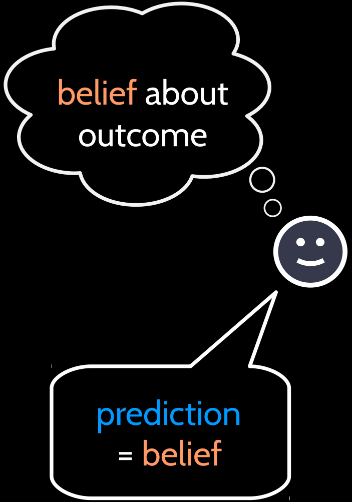
a tutorial at EC 2016

Part II

Rafael Frongillo and Bo Waggoner

25 July 2016

Scoring Rules



$$\text{Score}\,(\,\text{prediction}\,,\,\text{outcome}\,)$$

$$\max \; \mathbb{E}_{\text{outcome} \sim \text{belief}} \left[\; \text{Score} \;\right]$$

$$\min \sum_{\text{d in data}} \text{Loss}\,(\,\text{pred}\,,\,\text{d}\,)$$

Loss Function

$$L(\, r, y \,)$$

Objective
Distance
Penalty
Error
Loss
...

Parameter
Prediction
Estimate
Report
...

Observation
Data point
Sample
Truth
...

The Many Faces of Elicitation

Applications to algorithmic economics, machine learning, statistics, finance, engineering, ...

Formalism of elicitation used for model selection, estimation, empirical risk minimization (ERM), generalized regression, forecast evaluation / comparison / ranking, outlier detection, ...

Outline of Part II

Goal: survey *property elicitation* (asking for statistics rather than full distributions), show how it applies to machine learning in particular

1. Fundamentals of property elicitation

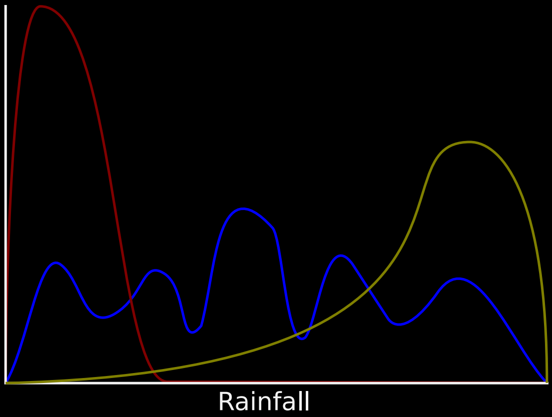
   break

2. "Elicitation complexity" and indirect elicitation
3. Machine learning applications and open problems

II.1. Property Elicitation

Information Overload

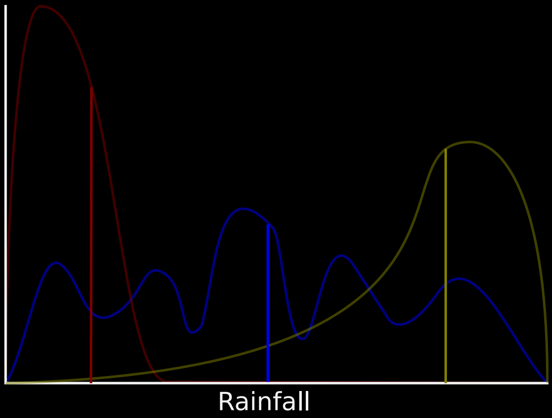
How much rain do you believe will fall today?



A lot of bits to communicate. . .

Information Overload

How much rain do you *expect* will fall today?



. . . if we just need a single number.

Example properties

- mean, variance, median, mode, moments of the distribution
- modal mass: what is the probability of the most likely outcome?
- confidence interval: an a, b such that w.prob 0.9, $a \le X \le b$.
- p-norm of the distribution
- $\ldots$

Research program

Loss $L(\hat{y}, y)$			Statistic Γ
Squared	$(\hat{y} - y)^2$	$\longrightarrow$	mean
Absolute	$\lvert\hat{y} - y\rvert$	$\longrightarrow$	median
Pinball	$(\hat{y} - y)(\mathbb{1}_{\hat{y} \geq y} - \alpha)$	$\longrightarrow$	α-quantile
	$\lvert 1_{\hat{y} \geq y} - \tau\rvert(\hat{y} - y)^2$	$\longrightarrow$	τ-expectile

- Which statistics (properties) can we compute by minimizing a loss (maximizing a score) over data?
- What are **all** losses minimized by the same statistic?
- How to **construct** losses for a statistic with good properties?

Outline for II.1

1. Definitions and recap of proper scoring rule result
2. Basic geometry and tools for impossibility
3. Survey of known characterizations

Definitions

A *property* is a function $\Gamma : \Delta_{\mathcal{Y}} \to \mathcal{R}$.
A *scoring rule* $S : \mathcal{R} \times \mathcal{Y} \to \mathbb{R}$ *elicits* Γ if

$$\Gamma(p) = \arg\max_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{p} S(r, Y).$$

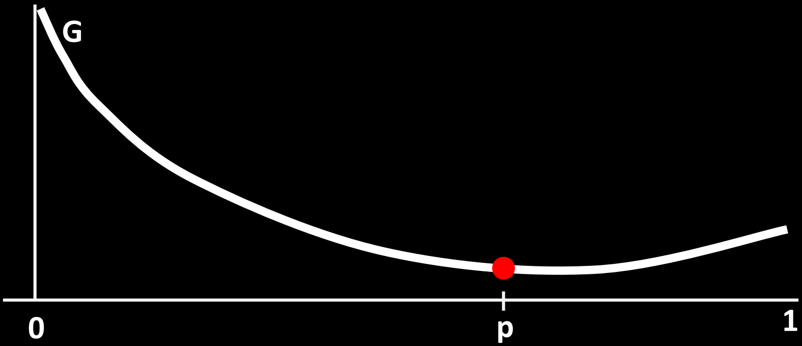
"an agent with belief p maximizes expected score by reporting $r = \Gamma(p)$."

Γ is *directly elicitable* if there exists S eliciting it.

Part I: The Simplest Property

Recall/reinterpret: a *proper scoring rule* elicits the property $\Gamma(p) = p$.

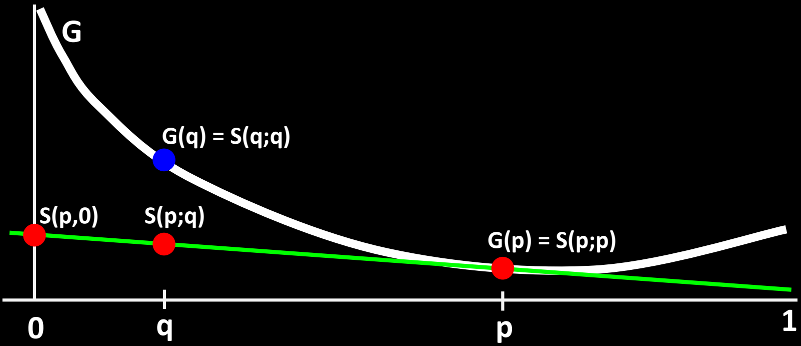
We showed: any proper scoring rule can be constructed from a convex G: **How?**



Part I: The Simplest Property

Recall/reinterpret: a *proper scoring rule* elicits the property $\Gamma(p) = p$.

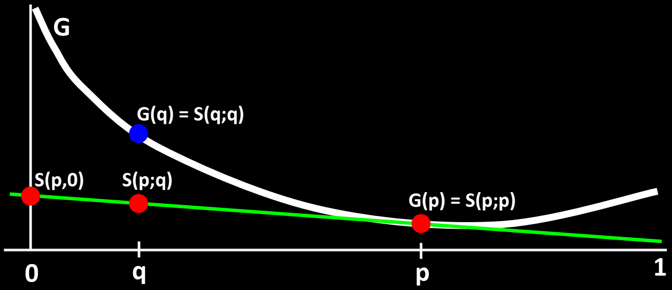
We showed: any proper scoring rule can be constructed from a convex G:



Theorem (Scoring Rule Characterization)

A scoring rule S is (strictly) proper ***if and only if*** *there exists a (strictly) convex G with*

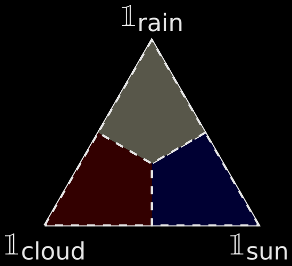
$$S(p,y) = G(p) + dG_p \cdot (\mathbb{1}_y - p).$$



Recall: level sets

The **level set** of r is $\{p : \Gamma(p) = r\}$.

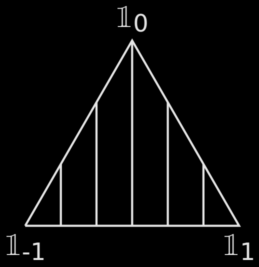
"the set of distributions all mapping to r"



Here: We drew the simplex $\Delta_{\{\text{clouds},\text{sun},\text{rain}\}}$
$\Gamma(p) =$ "most likely outcome" (mode).

A three-outcome example

Level set of the **mean**: all p with equal expectation
Here: $Y \in \{-1, 0, 1\}$.



Each line is a level set (*e.g.* distributions with mean 0).

Necessary geometry for elicitability

Theorem

If Γ is elicitable, then its level sets are convex.

Proof: Suppose $\Gamma(p) = \Gamma(p') = r$. Let $q = \lambda p + (1-\lambda)p'$.

Then $\forall r'$,

$$\mathop{\mathbb{E}}_{p} S(r,Y) > \mathop{\mathbb{E}}_{p} S(r',Y) \qquad \text{and}$$

$$\mathop{\mathbb{E}}_{p'} S(r,Y) > \mathop{\mathbb{E}}_{p'} S(r',Y)$$

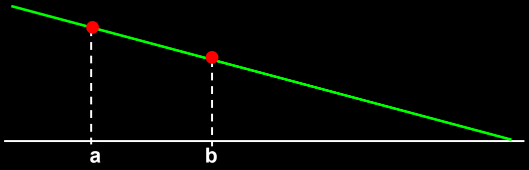
$$\implies \mathop{\mathbb{E}}_{q} S(r,Y) > \mathop{\mathbb{E}}_{q} S(r',Y).$$

Necessary geometry for elicitability

Theorem

If Γ is elicitable, then its level sets are convex.

Proof by picture: Consider $G(p) =$ expected utility.
If $\Gamma(a) = \Gamma(b)$, they must lie on the same hyperplane.

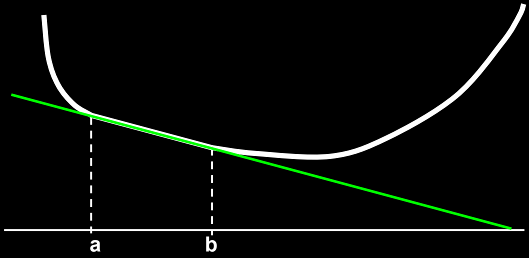


Necessary geometry for elicitability

Theorem

If Γ is elicitable, then its level sets are convex.

Proof by picture: Consider $G(p) =$ expected utility.
If $\Gamma(a) = \Gamma(b)$, they must lie on the same hyperplane.
But G is convex; must be flat between a and b.

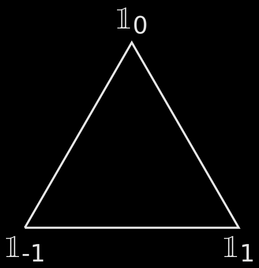


Obtaining Negative Results

Theorem

Variance is not directly elicitable.

Proof:



Each curve is a level set – not convex sets!

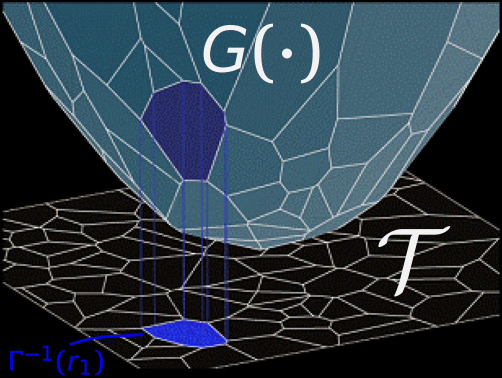
Survey of what we know

Cases that have been settled:

- $\Gamma(p) \in \mathcal{R}$, *finite "multiple-choice"*
- $\Gamma(p) = \mathbb{E}_p \, \phi(Y)$ *linear properties*
- $\Gamma(p) \in \mathbb{R}$ *scalar/one-dimensional*
- Others and general principles

Recall: finite properties

Finite properties are elicitable $\iff$ they are power diagrams; can construct scoring rule from diagram.



Linear properties

Theorem

Suppose $\Gamma(p) = \mathbb{E}_p \phi(Y)$. Then Γ is elicitable.

And: Γ is elicited by and only by S of the form

$$S(r, y) = G(y) + dG(y) \cdot (\phi(y) - r) + C_y$$

for some convex G.

Connections to:

- exponential families (ϕ is a sufficient statistic)
- prediction markets ($\phi \equiv$ the securities)

[Frongillo and Kash 2015]

One-dimensional properties

Identification function: $v : \mathcal{R} \to$ unit vectors in $\mathbb{R}^{\mathcal{Y}}$ such that, for all p, $\Gamma(p) = r \iff p \cdot v(r) = 0$.

Theorem

A continuous property $\Gamma : \Delta_{\mathcal{Y}} \to \mathbb{R}$ is elicitable ***if and only if*** *it has an identification function v.*
Furthermore, any scoring rule eliciting it has the form

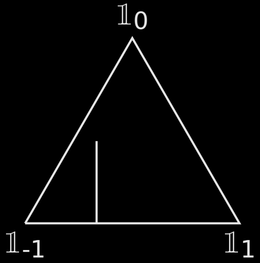
$$S(r, y) = C_y + \int_{r_0}^{r} \lambda(t) v(t)_y \, dt$$

for some positive $\lambda(t)$.

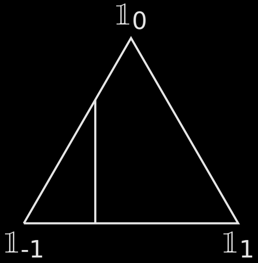
[Lambert et. al 2008]

Proof idea

1. Consider the constraint $\Gamma(p) = r$. Can show: this is a **linear constraint**, and since it's one constraint and $|\mathcal{Y}| - 1$ degrees of freedom, solutions lie on a $|\mathcal{Y}| - 2$ -dimensional subspace.



No!

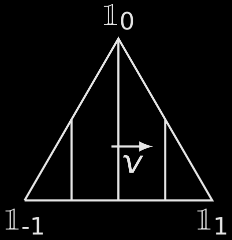


OK!

Proof idea

1. Consider the constraint $\Gamma(p) = r$. Can show: this is a **linear constraint**, *i.e.* solutions lie on an $|\mathcal{Y}| - 1$ -dimensional subspace.

2. These level sets are **ordered.**



$\Gamma(p)$ small $\longrightarrow$ $\Gamma(p)$ large
(direction given by v, the normal vector!)

Proof idea

1. Consider the constraint $\Gamma(p) = r$. Can show: this is a **linear constraint**, *i.e.* solutions lie on an $|\mathcal{Y}| - 1$ -dimensional subspace.

2. These level sets are **ordered.**

3. Can integrate along this direction with any given **weighting** λ. ("gold argument" of Savage 1971).

$$S(r, y) = C_y + \int_{r_0}^{r} \lambda(t) v(t)_y dt$$

The “gold argument”

Ask an agent to report her true value of gold in dollars/ounce, r.

Sell her a piece of gold at price 0/ounce. Another at price 1/ounce, . . . , up to r/ounce.

Truthful! She is happy with each transaction; reporting lower leaves money on the table and higher gives some undesirable transactions.

And: The pieces of gold could be any size! Could sell $\lambda(t)$ ounces of gold at each price t; still truthful.

Recap: state of knowledge

Known:

- $\Gamma(p) \in \mathcal{R}$, *finite "multiple-choice"*
- $\Gamma(p) = \mathbb{E}_p \phi(Y)$ *linear properties*
- $\Gamma(p) \in \mathbb{R}$ *scalar/one-dimensional*
- $\Gamma(p) = \mathbb{E}_p \phi(Y) / \mathbb{E} \psi(Y)$ *ratio of expectations*

Additionally:

- Tools for proving non-elicitability, *e.g.* convex level sets
- General principles (expected utility G must be convex, etc.)

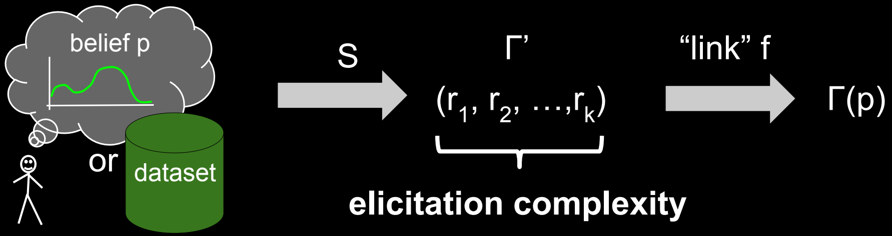
Not known: general multidimensional properties.

II.2. Elicitation Complexity

Back to Variance

- Var not elicitable with only one $\mathbb{R}$-valued report
- But what if you are allowed more?
- One idea: $\Gamma(p) = \left(\mathbb{E}_p[Y], \mathbb{E}_p[Y^2] \right) \in \mathbb{R}^2$
- Then $\mathrm{Var}(p) = \mathbb{E}_p[Y^2] - \mathbb{E}_p[Y]^2 = \Gamma(p)_2 - \left(\Gamma(p)_1\right)^2$
- Idea: $elic(\Gamma) :=$ min # of reports before you know Γ
  elicitation complexity *of* Γ
- Thus, $elic(\mathrm{Var}) = 2$

Indirect Elicitation



Competing Definitions

Γ is k-elicitable (i.e. $elic(\Gamma) \leq k$) if...

1. There exist k elicitable properties $\Gamma'_i : \Delta_{\mathcal{Y}} \to \mathbb{R}$ and link f such that $\Gamma = f \circ (\Gamma'_1, \ldots, \Gamma'_k)$. *[Lambert et al. 2008]*
2. There exists elicitable $\Gamma' : \Delta_{\mathcal{Y}} \to \mathbb{R}^k$ such that $\Gamma = \Gamma'_i$ *[Fissler & Ziegel 2015]*
3. There exists elicitable $\Gamma' : \Delta_{\mathcal{Y}} \to \mathbb{R}^k$ and link f such that $\Gamma = f \circ \Gamma'$ *[F & Kash 2015]*

Separating examples:

- $\Gamma(p) = \mathrm{Var}(p)$ — $elic =$ 1 2 2 2 3 2
- $\Gamma(p) = \mathbb{E}_p[Y]^2$ — $elic =$ 1 1 2 2 3 1
- $\Gamma(p) = \max_y p(y)$ — $elic =$ 1 $|\mathcal{Y}| - 1$ 2 2 3 2

The “Right” Definition

Problem: bijections from $\mathbb{R}^n$ to $\mathbb{R}$!
Solution: impose further structure.
$\mathcal{I} :=$ {identifiable props}. $\exists\ v\ s.t.\ \Gamma(p) = r \iff p \cdot v(r) = 0.$

$$elic_{\mathcal{I}}(\Gamma) = \min\{k : \text{exists elicitable } \Gamma' : \Delta_{\mathcal{Y}} \to \mathbb{R}^k \text{ in } \mathcal{I} \text{ and link } f \text{ such that } \Gamma = f \circ \Gamma' \}$$

“First elicit Γ', then apply f to get Γ”

Note: could choose any class $\mathcal{C}$ of “nice” properties.

$$elic_{\mathcal{C}}(\Gamma) = \min\{k : \text{exists elicitable } \Gamma' : \Delta_{\mathcal{Y}} \to \mathbb{R}^k \text{ in } \mathcal{C} \text{ and link } f \text{ such that } \Gamma = f \circ \Gamma' \}$$

Basics of Complexity

- Every continuous Γ has $elic_{\mathcal{I}}(\Gamma) \leq$ countable ∞
- "Full rank" linear $\Gamma : \Delta_{\mathcal{Y}} \to \mathbb{R}^k$ has $elic_{\mathcal{I}}(\Gamma) = k$
- $\Gamma = k$ distinct quantiles has $elic_{\mathcal{I}}(\Gamma) = k$
- $elic_{\mathcal{I}}(\{\Gamma_1, \Gamma_2\}) \leq elic_{\mathcal{I}}(\Gamma_1) + elic_{\mathcal{I}}(\Gamma_2)$

A Cool Trick: Modal Mass

$\Gamma(p) = \max_y p(y)$

Let $S(r, y) = 2r_1 \mathbb{1}\{r_2 = y\} - r_1^2$.

Then $\mathbb{E}_p S(r, Y) = 2r_1 p(r_2) - r_1^2$.

For any $r_1 > 0$, best r_2 is $\mathrm{argmax}_y\, p(y) =: \mathrm{mode}(p)$.

$\implies r_1 = p(r_2) = p(\mathrm{argmax}_y\, p(y)) = \max_y p(y) = \Gamma(p)$.

Hence, S elicits $(\mathrm{mode}(p), \Gamma(p)) \implies elic_I(\Gamma) = 2$.

An Upper Bound

- Let $W(p) = \max_{a \in \mathbb{R}^k} \mathbb{E}_p\, w(a, Y)$ where $w : \mathbb{R}^k \times \mathcal{Y} \to \mathbb{R}$
- In general, W is not elicitable...
- Let $a^*(p) = \operatorname{argmax}_{a \in \mathbb{R}^k} \mathbb{E}_p\, w(a, Y)$

  Note: a^* is a property elicited by w!

Theorem [F & Kash 2015]

If $a^* \in \mathcal{I}$, then $elic_{\mathcal{I}}(W) \leq k + 1$

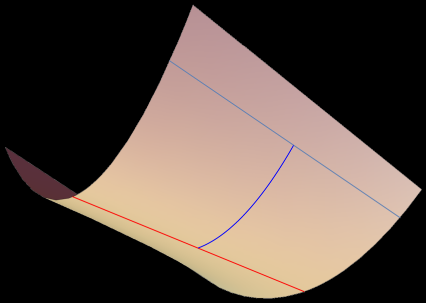
Proof:

$$S((r, a), y) = G(r) + dG_r \cdot (w(a, y) - r)$$

elicits (W, a^*) as long as $dG_r > 0$ everywhere.

So $G(r) = r^2$ works on $\mathbb{R}_+$, like prev slide.

A Lower Bound



Theorem [F & Kash 2015]

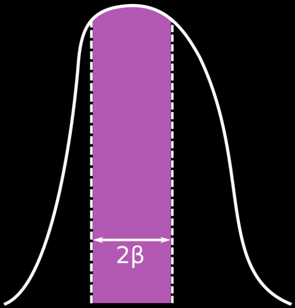
If $elic_{\mathcal{I}}(\alpha^*) = k$, then $elic_{\mathcal{I}}(W) \geq k + 1$

So $elic_{\mathcal{I}}(W) = k + 1$ for all such W!

Except when it's k...

Back to Modal Mass

- $\Gamma(p) = \max_a \mathbb{E}_p \, \mathbb{1}\{a = y\}$
- Take $w(a, y) = \mathbb{1}\{a = y\}$
  elicits the mode!
- $\Gamma(p) = W(p)$, so $elic_{\mathcal{I}}(\Gamma) = 2$
- More generally,
  $\Gamma_\beta(p) = \max_a \mathbb{E}_p \, \mathbb{1}_{|a - Y| < \beta}$



Aside: risk measures

Banks and Risk

Sometimes banks invest your money...

...and take on **risk**

What could possibly go wrong?

Quantifying & Regulating Risk

US law: banks can only take on so much risk

How to quantify? Financial **risk measures**.

Let p be distribution of believed financial losses Y
Risk measure is some $\rho : \mathcal{P} \to \mathbb{R}$

Introduced by [Artzner et al. 1998] *Cited by 5600+*
Various kinds: convex, coherent, distortion, spectral, ...

Which Risk Measure?

Most common: "value-at-risk" $\mathrm{V\alpha R}_\alpha$, the α-quantile of p
I.e. the amount y giving an α probability of losing $\geq y$

As of 2005: US banks required to calculate and report their $\mathrm{V\alpha R}_{0.01}$ estimates, over a 10 day horizon

New measure w/ better properties: "expected shortfall"

$$\mathrm{ES}_\alpha(p) = \min_{a \in \mathbb{R}} \left\{ \mathbb{E}_p \left[\frac{1}{\alpha}(a - Y)\mathbb{1}_{a \geq Y} - a \right] \right\}$$

Only problem: **not elicitable** [Gneiting 2011]
Needed for estimation, evaluation, "back-testing", ...

Rescuing ES

Cannot elicit ES, but it has low elicitation **complexity**

- *elic*(ES) ≤ 2 [Fissler & Ziegel 2015]
  more generally: spectral risk measures
  "Superquantile regression" of [Rockafellar et al. 2014]

- Special case of bounds we just gave:
  $\dim(\mathcal{A}) = \dim(\mathbb{R}) = 1 \implies \mathit{elic}_{\mathcal{I}}(\mathrm{ES}) = 2$

Punch line: elicitation complexity can save ~~lives~~ banks!

Other risk measures?

Recap, Open Questions

- Defined elicitation **complexity**: min # of reports/parameters until you have enough info to compute Γ
- Some tight bounds and examples

Many open questions. Complexity of:

- The mode when $\mathcal{Y} = \mathbb{R}$?? *We think* $elic_{\mathcal{I}}(mode) = \infty$
- Risk measures: distortion, spectral w/ cts support, ...
- Any non-elicitable statistic!
- $elic_{\mathcal{C}}$ for other $\mathcal{C}$ (stay tuned)

II.3. Machine Learning

ML Overview

Loss functions $L(r, y)$ used all over ML...

Unsurprisingly, property elicitation is a useful way to view some ML techniques/results.

1. Direct elicitation and **regression**
2. Indirect elicitation and **classification**

Note: many more intersections that we won't cover!

Empirical Risk Minimization

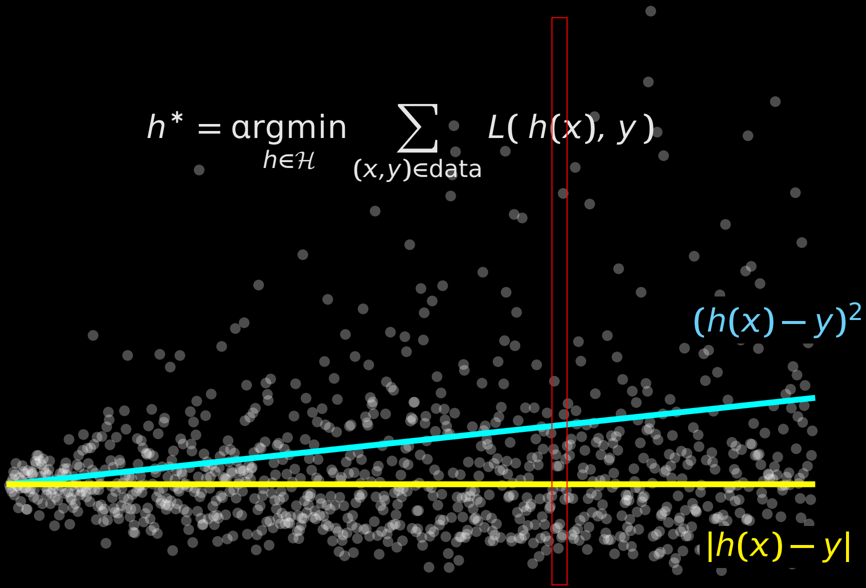
(for regression, or more generally)

$$h^* = \operatorname*{argmin}_{h \in \mathcal{H}} \sum_{(x,y) \in \text{data}} L(\, h(x), \, y\,) + \text{Reg}(h)$$

Note: regularization won't really matter...

The Loss Matters

$$h^* = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \sum_{(x,y) \in \text{data}} L(\, h(x), y \,)$$



Elicitation

Property $\Gamma : \Delta_{\mathcal{Y}} \to \mathcal{R}$ ("statistic")

L elicits Γ when

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

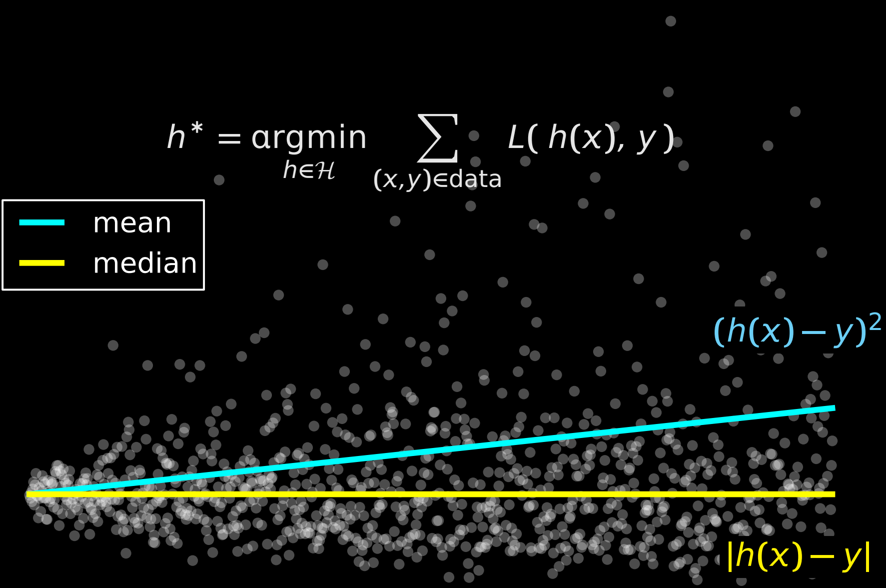
... for p = the empirical distribution $\hat{p}$ ...

$$= \operatorname*{argmin}_{r \in \mathcal{R}} \sum_{y \in \text{data set}} L(r, y)$$

- Mean: $\mathbb{E}_p[Y] = \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{p} (r - Y)^2$
- Median: $\text{med}(p) = \operatorname*{argmin}_{r \in \mathcal{R}} \mathop{\mathbb{E}}_{p} |r - Y|$

Elicitation is Key

$$h^* = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \sum_{(x,y) \in \text{data}} L(\, h(x), y\,)$$



Theorem

If the function $h^* : x \mapsto \Gamma(Y|X = x)$ *is in* $\mathcal{H}$*, then:*

$$L \textit{ elicits } \Gamma \implies \mathrm{ERM}_L(X, Y) = h^*$$

I.e., if your class $\mathcal{H}$ has a model h^* hitting the conditional statistic Γ (mean,median,etc) for every x, then ERM for *any* loss eliciting Γ will give h^*.

Takeaway:

"If $\mathcal{H}$ is expressive enough, elicitation tells all"

II.3.2. Indirect elicitation and classification

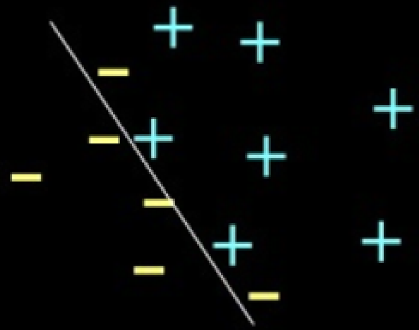
The Story

- Optimal classification is hard
- Many ML algorithms are like convex relaxations
- Still need asymptotic/statistical "consistency"
- Can view consistency as indirect elicitation

Classification

Input: Feature vectors $x \in \mathbb{R}^n$, labels $y \in \mathcal{Y}$
Here $\mathcal{Y}$ is a finite set, for now $\{+,-\}$.

Output: Classifier $h : \mathcal{X} \to \mathcal{Y}$ from some class $\mathcal{H}$
E.g. $\mathcal{H} =$ linear classifiers, $h(x) = \mathrm{sgn}(w \cdot x + b)$.



Direct Solution?

Natural objective: find the best model in $\mathcal{H}$

(fewest classification errors)

Corresponds to ERM with *0-1 loss* $L(r, y) = \mathbb{1}\{r \neq y\}$.

$$h^* = \operatorname*{argmin}_{h \in \mathcal{H}} \sum_{(x,y) \in \text{data}} \mathbb{1}\{h(x) \neq y\}$$

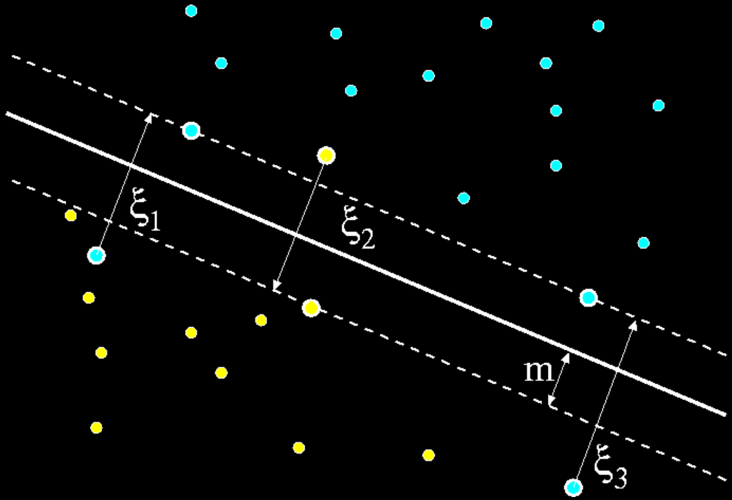
Problem: **NP-hard!** *[Arora et al. 1997]* (Also overfits...)

Solution: approximate 0-1 loss with a convex loss

logistic regression, SVMs, boosting, ...

Support Vector Machines (SVM)

Idea: find hyperplane with max *margin*, allowing errors



SVM Optimization

$$\min_{w,b,\xi} \quad \underbrace{\frac{1}{2}||w||^2}_{\text{large margin}} + \underbrace{C\sum_n \xi_n}_{\text{small slack}}$$

$$\text{subj. to} \quad y_n\left(w \cdot x_n + b\right) \geq 1 - \xi_n$$

$$\xi_n \geq 0$$

[HD]

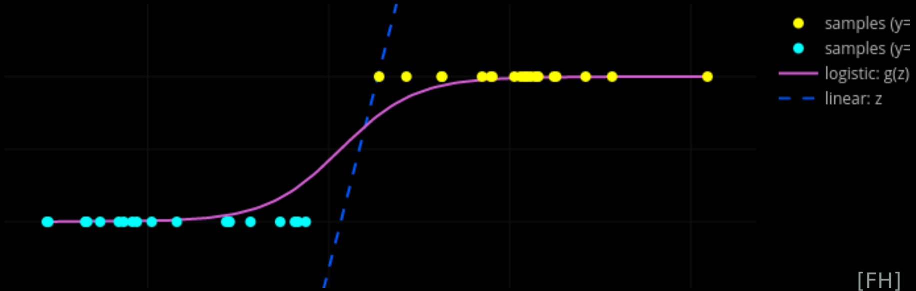
$$\min_{w,b} \quad \underbrace{\frac{1}{2}||w||^2}_{\text{large margin}} + \underbrace{C\sum_n \ell^{(\text{hin})}(y_n, w \cdot x_n + b)}_{\text{small slack}}$$

Can write as ERM! For *hinge loss* $L(r, y) = \max(0, 1 - ry)$:

$$(w^*, b^*) = \operatorname*{argmin}_{w \in \mathbb{R}^n, b \in \mathbb{R}} \sum_{(x,y) \in \text{data}} \max(0, 1 - y(w \cdot x + b)) + \frac{1}{2C}\|w\|^2$$

Logistic Regression

Idea: fit a model h to the log-odds ratio $\log \frac{\Pr[Y=+|X=x]}{\Pr[Y=-|X=x]}$.



[FH]

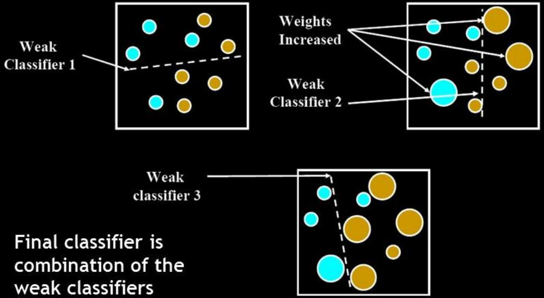
Then prediction $y = \operatorname{sgn}(h(x))$ is the most likely label.

ERM for *logistic loss* $L(r, y) = \log(1 + \exp(-ry))$.

$$(w^*, b^*) = \underset{w \in \mathbb{R}^n, b \in \mathbb{R}}{\operatorname{argmin}} \sum_{(x,y) \in \text{data}} \log\Big(1 + \exp(-y(w \cdot x + b))\Big)$$

AdaBoost

Idea: focus more on what you got wrong, and iterate



[KG]

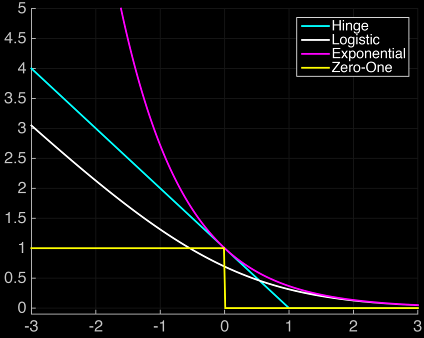
Each step, use exp weights to update data distribution
Then combine: $h(x) = \mathrm{sgn}(\alpha_1 h_1(x) + \alpha_2 h_2(x) + \alpha_3 h_3(x))$

Suprisingly, ERM for *exponential loss* $L(r, y) = \exp(-ry)$.

Each iteration is a coordinate descent step

Margin Losses, Calibrated

All these are *margin losses*: $L(r, y) = \phi(ry)$.



[KW]

Theorem (Bartlett, Jordan, McAuliffe 2006)

Let ϕ be convex. Then L is *classification-calibrated* if and only if ϕ is differentiable at 0 and $\phi'(0) < 0$.

Calibrated → Indirect Elicitation

Def. L is *classification-calibrated* if

$$\begin{aligned} \operatorname{sgn}(\Gamma'(p)) = + \quad &\iff \quad \operatorname{mode}(p) = + \\ \min_{r>0} \mathbb{E}_p L(r,Y) > \min_{r<0} \mathbb{E}_p L(r,Y) \quad &\iff \quad p(+) > p(-) \end{aligned}$$

Indirect elicitation: $\Gamma = f \circ \Gamma'$

What are Γ, f here? $\Gamma = \operatorname{mode}$, $f = \operatorname{sgn}$

Alternate Def. L is *classification-calibrated* if it indirectly elicits the mode via link $f = \operatorname{sgn}$

Indirect Elicitation in ML

Recall: $elic_{\mathcal{C}}(\Gamma) = \min\{k : \text{exists elicitable } \Gamma' \text{ in } \mathcal{C}$
$\text{and link } f \text{ such that } \Gamma = f \circ \Gamma' \}$

General program: $\mathcal{C}$ = properties with "nice" losses

Approximate NP-hard objective with a nicer one

Elicitation keeps "calibration"

Here: $\mathcal{C}$ = properties elicited by convex margin losses.

Next: $\mathcal{C}$ = linear properties.

Another Application: Rankings

Given L : {possible rankings} × {relevant docs} → $\mathbb{R}$

Still hard to optimize for $\Gamma := \arg\min \mathbb{E}\, L$ directly...

Look for surrogate: want $\Gamma = f \circ \Gamma'$ for Γ' linear.
How big does Γ' need to be? I.e. what is $elic_{\text{linear}}(\Gamma)$?

Theorem (Agarwal, Agarwal 2015)

$elic_{linear}(\Gamma) = \text{affdim}(L)$. Think affdim = rank.

Proof sketch (upper bound):

Write $L = BA + c$ so that $L(r, y) = (BA)_{ry} + c$.

Let $\Gamma'(p) = \mathbb{E}_p A_{\cdot,Y} = Ap$, $f(a) = \arg\min_r (Ba)_r$.

Then $(f \circ \Gamma')(p) = \arg\min_r (BAp)_r = \arg\min_r \mathbb{E}_p L(r, Y) = \Gamma$.

Takeaway

1. In clasification, need to approximate a hard discrete problem, often with a continuous convex objective.

2. Elicitation keeps the limiting behavior the same.

3. Lots of open questions.

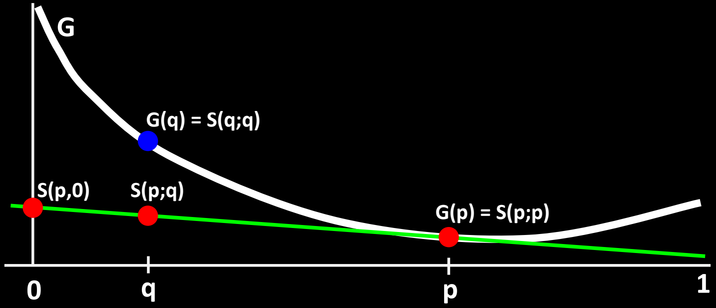
(I + II). Recap

Main questions:

- What properties can be elicited?
  or, how many reports does it take to elicit them?
- How to characterize **all** loss functions (scoring rules) eliciting a given property?
- How to construct loss functions in a principled way?

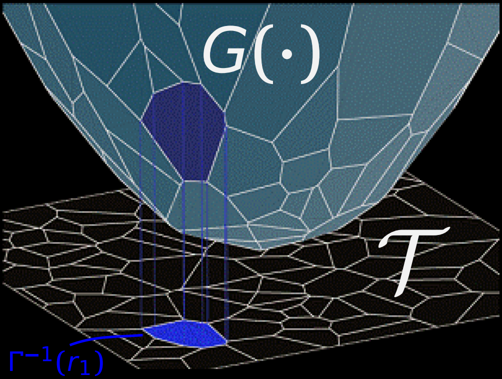
Known characterizations:

proper scoring rules



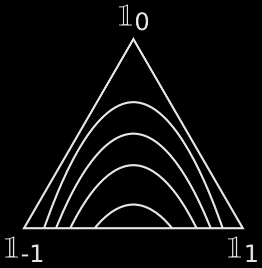
Known characterizations:

linear properties, finite properties, continuous 1-dimensional properties



Known principles:

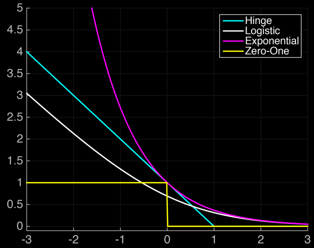
convexity, scoring rules as subgradients, . . .



Applications outside elicitation:

- Mechanism design: characterizing and constructing truthful mechanisms

- Machine learning: characterizing and constructing useful loss functions



[KW]

Open problems and research directions:

- Characterizations and constructions for more properties
- Mechanism-design applications with complex type spaces
- Elicitation complexity with efficiently-optimizable surrogate loss functions (ML motivation)
- More ML: general program to select (surrogate) losses in principled way using elicitation
- . . .

Uiteinde.

Thanks for coming!

References & Credits

[FH] Florian Hartl
[KW] Kilian Weinberger
[KG] Kristen Grauman
[HD] Hal Daumé