A Characterization of Scoring Rules for Linear Properties

Rafael Frongillo

Department of Computer Science
University of California at Berkeley

June 26, 2012

Joint work with Jake Abernethy

A Characterization of Proper Losses for Linear Properties

Rafael Frongillo

Department of Computer Science
University of California at Berkeley

June 26, 2012

Joint work with Jake Abernethy

The unstoppable Jake Abernethy

Now a postdoc at UPenn with Michael Kearns

The unstoppable Jake Abernethy

Now a postdoc at UPenn with Michael Kearns

Proper Losses

Typical setting: classification

- Labels $y \in [n] = \{1, \ldots, n\}$
- Prediction $p \in \Delta_n$
- Loss $\ell : \Delta_n \to \mathbb{R}^n$ ⟵ *a **vector**: loss of p and y is $\ell[p]_y$*
- ℓ is *proper* if $p = \arg\min_q \{ \ell[q]\,p \}$

$\mathbb{E}_{y \sim p}[\ell[q]_y]$

Example: log loss

- Take $\ell[p]_y = -\log p_y$
- Now $\ell[q]\,p = -\sum p_y \log q_y = \mathrm{KL}(p\|q) + H(p)$

  Minimized at $q = p$

Proper Losses

Typical setting: classification

- Labels $y \in [n] = \{1, \ldots, n\}$
- Prediction $p \in \Delta_n$
- Loss $\ell : \Delta_n \to \mathbb{R}^n$ ⟵ *a **vector**: loss of p and y is $\ell[p]_y$*
- ℓ is *proper* if $p = \arg\min_q \{ \ \ell[q]\,p \ \}$

$\mathbb{E}_{y \sim p}[\ell[q]_y]$

Example: log loss

- Take $\ell[p]_y = -\log p_y$
- Now $\ell[q]\,p = -\sum p_y \log q_y = \mathrm{KL}(p \| q) + H(p)$

  Minimized at $q = p$

Proper Losses

Typical setting: classification

- Labels $y \in [n] = \{1, \dots, n\}$
- Prediction $p \in \Delta_n$
- Loss $\ell : \Delta_n \to \mathbb{R}^n$ $\longleftarrow$ *a **vector**: loss of p and y is* $\ell[p]_y$
- ℓ is *proper* if $p = \operatorname{argmin}_q \{ \ell[q]\, p \}$

$\ell[q]\, p$ ← $\mathbb{E}_{y \sim p}[\ell[q]_y]$

Example: log loss

- Take $\ell[p]_y = -\log p_y$
- Now $\ell[q]\, p = -\sum p_y \log q_y = \mathrm{KL}(p \| q) + H(p)$

  Minimized at $q = p$

Proper Losses

Typical setting: classification

- Labels $y \in [n] = \{1, \ldots, n\}$
- Prediction $p \in \Delta_n$
- Loss $\ell : \Delta_n \to \mathbb{R}^n$ $\longleftarrow$ *a **vector**: loss of p and y is $\ell[p]_y$*
- ℓ is *proper* if $p = \arg\min_q \{ \ell[q]\, p \}$

$\ell[q]\,p$ ← $\mathbb{E}_{y \sim p}[\ell[q]_y]$

Example: log loss

- Take $\ell[p]_y = -\log p_y$
- Now $\ell[q]\,p = -\sum p_y \log q_y = \mathrm{KL}(p \| q) + H(p)$

  Minimized at $q = p$

Proper Losses... for Properties

Our setting: properties of distributions

- Outcomes $\omega \in \Omega$
- Distributional *property* $\Gamma : \Delta_\Omega \to \mathcal{V} \subseteq \mathbb{R}^k$ *summary information*
- Prediction $v \in \mathcal{V}$
- Loss $\ell : \mathcal{V} \to \mathbb{R}^\Omega$ *loss of v and ω is $\ell[v]_\omega$*
- ℓ is *Γ-proper* if $\Gamma(p) = \operatorname{argmin}_v \{ \ell[v]\, p \}$

We will consider *linear* Γ:

- $\Gamma(p) = \mathbb{E}_{\omega \sim p}[\phi(\omega)]$ for some $\phi : \Omega \to \mathcal{V}$ *i.e. means*

Proper Losses... for Properties

Our setting: properties of distributions

- Outcomes $\omega \in \Omega$
- Distributional *property* $\Gamma : \Delta_\Omega \to \mathcal{V} \subseteq \mathbb{R}^k$ *summary information*
- Prediction $v \in \mathcal{V}$
- Loss $\ell : \mathcal{V} \to \mathbb{R}^\Omega$ *loss of v and ω is $\ell[v]_\omega$*
- ℓ is *Γ-proper* if $\Gamma(p) = \operatorname{argmin}_v \{ \ell[v]\, p \}$

We will consider *linear* Γ:

- $\Gamma(p) = \mathbb{E}_{\omega \sim p}[\phi(\omega)]$ for some $\phi : \Omega \to \mathcal{V}$ *i.e. means*

Motivation

This talk:

A Characterization of Proper Losses for Linear Properties

Our goal

Given some linear property $\Gamma : \Delta_\Omega \to \mathcal{V}$, determine exactly the losses $\ell : \mathcal{V} \to \mathbb{R}^\Omega$ which are Γ-proper

... Why bother?

Motivation: Proper

Proper losses are *well-calibrated*

Example: learning a coin's bias p

- Want ℓ to measure *performance*
- After $N >> 1$ flips, we want

$$p \approx \underset{q}{\operatorname{argmin}} \left\{ \frac{\#\text{heads}}{N} \ell[q]_{\text{heads}} + \frac{\#\text{tails}}{N} \ell[q]_{\text{tails}} \right\}$$

"expected" loss of predicting q

Motivation: Proper

Proper losses are *well-calibrated*

Example: learning a coin's bias p

- Want ℓ to measure *performance*
- After $N >> 1$ flips, we want

$$p \approx \operatorname*{argmin}_q \left\{ \frac{\#\text{heads}}{N} \ell[q]_{\text{heads}} + \frac{\#\text{tails}}{N} \ell[q]_{\text{tails}} \right\}$$

"expected" loss of predicting q

Motivation: Proper

Proper losses are *well-calibrated*

Example: learning a coin's bias p

- Want ℓ to measure *performance*
- After $N >> 1$ flips, we want

$$p \approx \arg\min_q \left\{ \frac{\#\text{heads}}{N} \ell[q]_{\text{heads}} + \frac{\#\text{tails}}{N} \ell[q]_{\text{tails}} \right\}$$

"expected" loss of predicting q

Motivation: Characterization

Loss should *quantify* error

Two losses for eliciting a mean

- Squared: $\ell[v]_\omega = (v - \omega)^2$
- Log: $\ell[v]_\omega = \mathrm{KL}(\omega \| v)$

Very different notions of error

Given a notion of error, when can I *design* a proper loss to match?

Motivation: Characterization

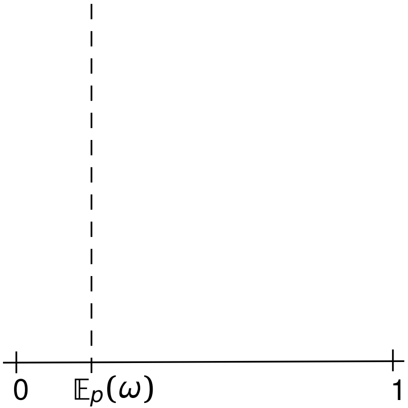
Loss should *quantify* error

Two losses for eliciting a mean

- Squared: $\ell[v]_\omega = (v - \omega)^2$
- Log: $\ell[v]_\omega = \mathrm{KL}(\omega \| v)$

Very different notions of error

Given a notion of error, when can I *design* a proper loss to match?



Motivation: Characterization

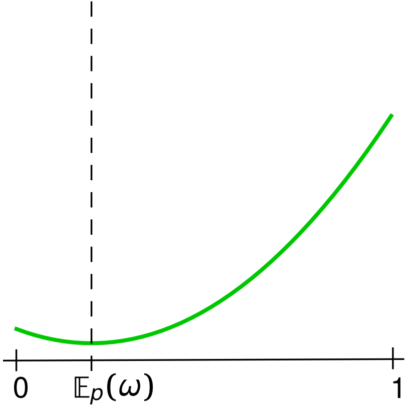
Loss should *quantify* error

Two losses for eliciting a mean

- Squared: $\ell[v]_\omega = (v - \omega)^2$
- Log: $\ell[v]_\omega = \mathrm{KL}(\omega \| v)$

Very different notions of error

Given a notion of error, when can I *design* a proper loss to match?



Motivation: Characterization

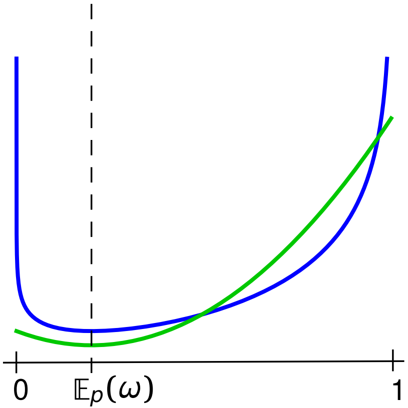
Loss should *quantify* error

Two losses for eliciting a mean

- Squared: $\ell[v]_\omega = (v - \omega)^2$
- Log: $\ell[v]_\omega = \mathrm{KL}(\omega \| v)$

Very different notions of error

Given a notion of error, when can I *design* a proper loss to match?



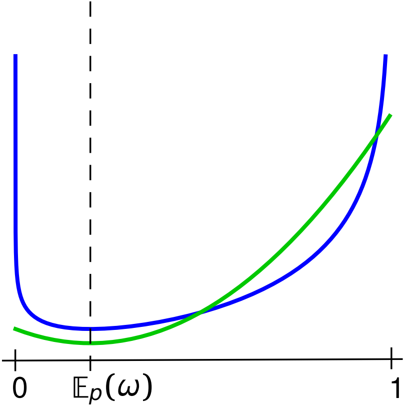
Motivation: Characterization

Loss should *quantify* error

Two losses for eliciting a mean

- Squared: $\ell[v]_\omega = (v - \omega)^2$
- Log: $\ell[v]_\omega = \mathrm{KL}(\omega \| v)$

Very different notions of error



Given a notion of error, when can I *design* a proper loss to match?

Motivation: Characterization

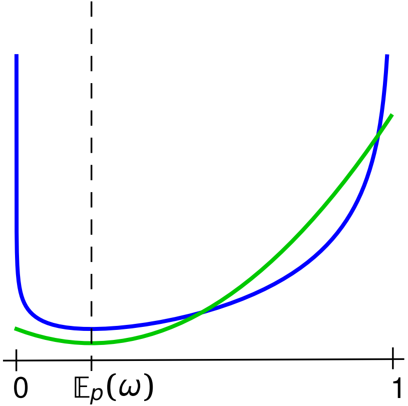
Loss should *quantify* error

Two losses for eliciting a mean

- Squared: $\ell[v]_\omega = (v - \omega)^2$
- Log: $\ell[v]_\omega = \mathrm{KL}(\omega \| v)$

Very different notions of error

Given a notion of error, when can I *design* a proper loss to match?



Motivation: Properties

Problem: What if your "classification" problem has a huge (∞) number of classes?

E.g. Price of gas next month?

Solution: Use a $\Gamma : \Delta_\Omega \to \mathcal{V} \subseteq \mathbb{R}^k$

Only extract the "relevant information" from your data

Means are quite expressive:

- First k moments of a distribution: $\phi(\omega) = (\omega, \omega^2, \ldots, \omega^k)$
- Covariance matrix: $\phi(\omega)_{(i,j)} = \omega_i \omega_j$

Motivation: Properties

Problem: What if your “classification” problem has a huge (∞) number of classes?

E.g. Price of gas next month?

Solution: Use a $\Gamma : \Delta_\Omega \to \mathcal{V} \subseteq \mathbb{R}^k$

Only extract the “relevant information” from your data

Means are quite expressive:

- First k moments of a distribution: $\phi(\omega) = (\omega, \omega^2, \ldots, \omega^k)$
- Covariance matrix: $\phi(\omega)_{(i,j)} = \omega_i \omega_j$

Motivation: **Linear** Properties

Problem: What if your "classification" problem has a huge (∞) number of classes?

E.g. Price of gas next month?

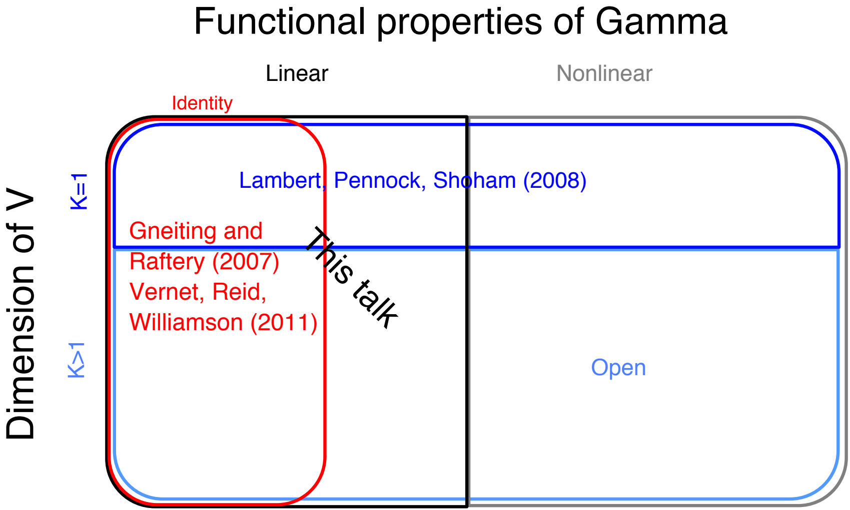
Solution: Use a $\Gamma : \Delta_\Omega \to \mathcal{V} \subseteq \mathbb{R}^k$

Only extract the "relevant information" from your data

Means are quite expressive:

- First k moments of a distribution: $\phi(\omega) = (\omega, \omega^2, \ldots, \omega^k)$
- Covariance matrix: $\phi(\omega)_{(i,j)} = \omega_i \omega_j$

Known characterizations of proper losses



Bregman divergences

Given convex $f : \mathcal{V} \to \mathbb{R}$, the Bregman *divergence* w.r.t. f:

$$D_f(x, y) := f(x) - f(y) - \nabla f(y) \cdot (x - y)$$

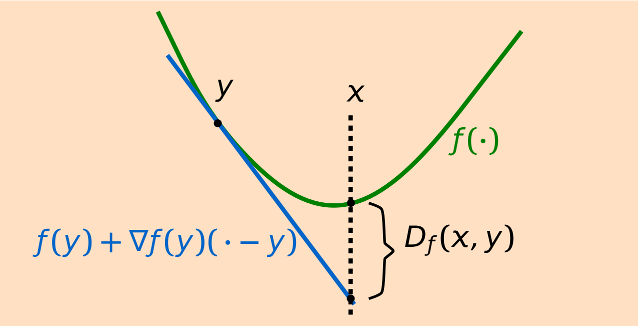
f is called: Bayes risk, regularizer, generalized entropy

Bregman divergences

Given convex $f : \mathcal{V} \to \mathbb{R}$, the Bregman *divergence* w.r.t. f:

$$D_f(x, y) := f(x) - f(y) - \nabla f(y) \cdot (x - y)$$

f is called: Bayes risk, regularizer, generalized entropy



Divergences and means

Definition: ℓ is *divergence-based* if $\exists f, \phi$ s.t.

$$\ell[v]_\omega = D_f(\phi(\omega), v)$$

Fact: this ℓ is proper for linear property $\Gamma(p) = \mathbb{E}_{\omega \sim p}[\phi(\omega)]$

$$\begin{aligned}
&\operatorname*{argmin}_v \{\ell[v]p\} \\
&= \operatorname*{argmin}_v \left\{ \mathbb{E}_{\omega \sim p} \left[f(\phi(\omega)) - f(v) - \nabla f(v) \cdot (\phi(\omega) - v) \right] \right\} \\
&= \operatorname*{argmin}_v \{ -f(v) - \nabla f(v) \cdot (\Gamma(p) - v) \} \\
&= \operatorname*{argmin}_v \{ D_f(\Gamma(p), v) - f(\Gamma(p)) \} = \Gamma(p)
\end{aligned}$$

Divergences and means

Definition: ℓ is *divergence-based* if $\exists f, \phi$ s.t.

$$\ell[v]_\omega = D_f(\phi(\omega), v)$$

Fact: this ℓ is proper for linear property $\Gamma(p) = \mathbb{E}_{\omega \sim p}[\phi(\omega)]$

$$\begin{aligned}
&\operatorname*{argmin}_v \{\ell[v]p\} \\
&= \operatorname*{argmin}_v \left\{ \mathbb{E}_{\omega \sim p} \left[f(\phi(\omega)) - f(v) - \nabla f(v) \cdot (\phi(\omega) - v) \right] \right\} \\
&= \operatorname*{argmin}_v \{ -f(v) - \nabla f(v) \cdot (\Gamma(p) - v) \} \\
&= \operatorname*{argmin}_v \{ D_f(\Gamma(p), v) - f(\Gamma(p)) \} = \Gamma(p)
\end{aligned}$$

Divergences and means

Definition: ℓ is *divergence-based* if $\exists f, \phi$ s.t.

$$\ell[v]_\omega = D_f(\phi(\omega), v)$$

Fact: this ℓ is proper for linear property $\Gamma(p) = \mathbb{E}_{\omega\sim p}[\phi(\omega)]$

$$\operatorname*{argmin}_v \{\ell[v]p\}$$

$$= \operatorname*{argmin}_v \left\{ \mathbb{E}_{\omega\sim p}\left[f(\phi(\omega)) - f(v) - \nabla f(v)\cdot(\phi(\omega) - v)\right]\right\}$$

$$= \operatorname*{argmin}_v \{-f(v) - \nabla f(v)\cdot(\Gamma(p) - v)\}$$

$$= \operatorname*{argmin}_v \left\{D_f(\Gamma(p), v) - f(\Gamma(p))\right\} = \Gamma(p)$$

Divergences and means

Definition: ℓ is *divergence-based* if $\exists f, \phi$ s.t.

$$\ell[v]_\omega = D_f(\phi(\omega), v)$$

Fact: this ℓ is proper for linear property $\Gamma(p) = \mathbb{E}_{\omega\sim p}[\phi(\omega)]$

$$\operatorname*{argmin}_v \{\ell[v]p\}$$
$$= \operatorname*{argmin}_v \left\{ \mathop{\mathbb{E}}_{\omega\sim p} \left[f(\phi(\omega)) - f(v) - \nabla f(v) \cdot (\phi(\omega) - v) \right] \right\}$$
$$= \operatorname*{argmin}_v \{ -f(v) - \nabla f(v) \cdot (\Gamma(p) - v) \}$$
$$= \operatorname*{argmin}_v \left\{ D_f(\Gamma(p), v) - f(\Gamma(p)) \right\} = \Gamma(p)$$

Divergences and means

Definition: ℓ is *divergence-based* if $\exists f, \phi$ s.t.

$$\ell[v]_\omega = D_f(\phi(\omega), v)$$

Fact: this ℓ is proper for linear property $\Gamma(p) = \mathbb{E}_{\omega\sim p}[\phi(\omega)]$

$$\begin{aligned}
&\operatorname*{argmin}_v \{\ell[v]p\} \\
&= \operatorname*{argmin}_v \left\{ \mathop{\mathbb{E}}_{\omega\sim p}\left[f(\phi(\omega)) - f(v) - \nabla f(v)\cdot(\phi(\omega) - v)\right]\right\} \\
&= \operatorname*{argmin}_v \{-f(v) - \nabla f(v)\cdot(\Gamma(p) - v)\} \\
&= \operatorname*{argmin}_v \left\{D_f(\Gamma(p), v) - f(\Gamma(p))\right\} = \Gamma(p)
\end{aligned}$$

Divergences and means

Definition: ℓ is *divergence-based* if $\exists f, \phi$ s.t.

$$\ell[v]_\omega = D_f(\phi(\omega), v)$$

Fact: this ℓ is proper for linear property $\Gamma(p) = \mathbb{E}_{\omega\sim p}[\phi(\omega)]$

$$\begin{aligned}
&\arg\min_v \{\ell[v]p\} \\
&= \arg\min_v \left\{ \mathbb{E}_{\omega\sim p}\left[f(\phi(\omega)) - f(v) - \nabla f(v)\cdot(\phi(\omega) - v) \right] \right\} \\
&= \arg\min_v \{ -f(v) - \nabla f(v)\cdot(\Gamma(p) - v) \} \\
&= \arg\min_v \left\{ D_f(\Gamma(p), v) - f(\Gamma(p)) \right\} = \Gamma(p)
\end{aligned}$$

Characterization for linear properties

This shows divergence-based $\implies$ Γ-proper for some linear Γ

Q: Is every Γ-proper loss ℓ divergence-based?

A: Yes[1]!

Theorem (Abernethy, F.)

ℓ is Γ-proper for linear Γ $\iff$ ℓ is divergence-based

[1] with extremely weak differentiability assumptions

Characterization for linear properties

This shows divergence-based $\implies$ Γ-proper for some linear Γ

Q: Is every Γ-proper loss ℓ divergence-based?

A: Yes[1]!

Theorem (Abernethy, F.)

ℓ is Γ-proper for linear Γ $\iff$ ℓ is divergence-based

[1] with extremely weak differentiability assumptions

Proof Intuition

We draw intuition from the identity case *i.e.* $\Gamma(p) = p$

Theorem (Gneiting and Raftery, 2010)

$\ell : \Delta_\Omega \times \Omega \to \mathbb{R}$ *proper* $\implies$ ℓ *is divergence-based*

Their proof:

- Extract $f(p) = \ell[p]\, p$ *Bayes risk, concave*
- Observe $\ell[p]\, p + \ell[p]\,(q - p) \geq \ell[q]\, q$ *from propriety*
- Hence $\ell[p]$ is a gradient of f! $\implies$ divergence

Proof Intuition

We draw intuition from the identity case *i.e.* $\Gamma(p) = p$

Theorem (Gneiting and Raftery, 2010)

$\ell : \Delta_\Omega \times \Omega \to \mathbb{R}$ *proper* $\implies$ ℓ *is divergence-based*

Their proof:

- Extract $f(p) = \ell[p]\, p$ *Bayes risk, concave*
- Observe $\ell[p]\, p + \ell[p]\,(q-p) \geq \ell[q]\, q$ *from propriety*
- Hence $\ell[p]$ is a gradient of f! $\implies$ divergence

Proof Intuition

We draw intuition from the identity case *i.e.* $\Gamma(p) = p$

Theorem (Gneiting and Raftery, 2010)

$\ell : \Delta_\Omega \times \Omega \to \mathbb{R}$ *proper* $\implies$ ℓ *is divergence-based*

??

Their proof:

- Extract $f(p) = \ell[p]\, p$ *Bayes risk, concave*
- Observe $\underbrace{\ell[p]\, p}_{f(p)} + \underbrace{\ell[p]}_{\partial f(p)}(q - p) \geq \underbrace{\ell[q]\, q}_{f(q)}$ *from propriety*
- Hence $\ell[p]$ is a gradient of f! $\implies$ divergence

Proof Intuition

We draw intuition from the identity case *i.e.* $\Gamma(p) = p$

Theorem (Gneiting and Raftery, 2010)

$\ell : \Delta_\Omega \times \Omega \to \mathbb{R}$ *proper* $\implies$ ℓ *is divergence-based*

Their proof:

- Extract $f(p) = \ell[p]\,p$ *Bayes risk, concave*
- Observe $\ell[p]\,p + \ell[p](q-p) \geq \ell[q]\,q$ *from propriety*

  $f(p)$ $\partial f(p)$ $f(q)$

- Hence $\ell[p]$ is a gradient of f! $\implies$ divergence

Proof Intuition

We draw intuition from the identity case *i.e.* $\Gamma(p) = p$

Theorem (Gneiting and Raftery, 2010)

$\ell : \Delta_\Omega \times \Omega \to \mathbb{R}$ *proper* $\implies$ ℓ *is divergence-based*

Their proof:

- Extract $f(p) = \ell[p]\, p$ *Bayes risk, concave*
- Observe $\underbrace{\ell[p]\, p}_{f(p)} + \underbrace{\ell[p]}_{\partial f(p)}(q-p) \geq \underbrace{\ell[q]\, q}_{f(q)}$ *from propriety*
- Hence $\ell[p]$ is a gradient of f! $\implies$ divergence

Proof Intuition

We draw intuition from the identity case *i.e.* $\Gamma(p) = p$

Theorem (Gneiting and Raftery, 2010)

$\ell : \Delta_\Omega \times \Omega \to \mathbb{R}$ *proper* $\implies$ ℓ *is divergence-based*

Their proof:

- Extract $f(p) = \ell[p]\, p$ *Bayes risk, concave*
- Observe $\underbrace{\ell[p]\, p}_{f(p)} + \underbrace{\ell[p]}_{\partial f(p)}(q-p) \geq \underbrace{\ell[q]\, q}_{f(q)}$ *from propriety*
- Hence $\ell[p]$ is a gradient of f! $\implies$ divergence

Proof Intuition

We draw intuition from the identity case *i.e.* $\Gamma(p) = p$

Theorem (Gneiting and Raftery, 2010)

$\ell : \Delta_\Omega \times \Omega \to \mathbb{R}$ *proper* $\implies$ ℓ *is divergence-based*

Their proof: ??

- Extract $f(p) = \ell[p]\, p$ *Bayes risk, concave*
- Observe $\ell[p]\, p + \ell[p]\,(q - p) \geq \ell[q]\, q$ *from propriety*

  $f(p)$ $\partial f(p)$ $f(q)$
- Hence $\ell[p]$ is a gradient of f! $\implies$ divergence

Proof Intiution

Their proof:

- Extract $f(p) = \ell[p]\, p$ **??**

Challenge: How to define f when $\mathcal{V} \neq \Delta_\Omega$?

- Let $\hat{p}$ such that $\Gamma \circ \hat{p} \equiv \mathrm{id}_{\mathcal{V}}$

  A "family" of distributions with "parameter space" $\mathcal{V}$

- Now $f(v) = \ell[v]\hat{p}[v]$

Proof Intiution

Their proof:

- Extract $f(p) = \ell[p]\, p$ **??**

Challenge: How to define f when $\mathcal{V} \neq \Delta_\Omega$?

- Let $\hat{p}$ such that $\Gamma \circ \hat{p} \equiv \mathrm{id}_{\mathcal{V}}$

  A "family" of distributions with "parameter space" $\mathcal{V}$

- Now $f(v) = \ell[v]\hat{p}[v]$

Switching Gears: Prediction Markets

Obama $1		Romney $1		Johnson $1		
$0.22	+	$0.35	+	$0.43	=	1

- Traders buy and sell these contracts
- Prices reflect the consensus prediction

Switching Gears: Prediction Markets

Obama $1		Romney $1		Johnson $1		
$0.22	+	$0.35	+	$0.43	=	1

- Traders buy and sell these contracts
- Prices reflect the consensus prediction

Switching Gears: Prediction Markets

Obama $1		Romney $1		Johnson $1		
\$0.22	+	\$0.35	+	\$0.43	=	1

- Traders buy and sell these contracts
- Prices reflect the consensus prediction

Switching Gears: Prediction Markets

Obama $1		Romney $1		Johnson $1		
$0.22	+	$0.35	+	$0.43	=	1

- Traders buy and sell these contracts
- Prices reflect the consensus prediction

Quantifying the Wagers

In standard market maker model, prices *adapt* to trades

From NIPS 2011, we can describe the net *profit* of such a trade in terms of the change $\mathbf{p} \to \mathbf{p}'$ in the *prices*...

... as the drop in a divergence-based loss!

Theorem (Abernethy, F.)

Traders have profit $\ell[\mathbf{p}]_\omega - \ell[\mathbf{p}']_\omega \iff \ell$ *is divergence-based*

Aside: can use this framework for data mining competitions!

Quantifying the Wagers

In standard market maker model, prices *adapt* to trades

From NIPS 2011, we can describe the net *profit* of such a trade in terms of the change $\mathbf{p} \to \mathbf{p}'$ in the *prices*...

... as the drop in a divergence-based loss!

Theorem (Abernethy, F.)

Traders have profit $\ell[\mathbf{p}]_\omega - \ell[\mathbf{p}']_\omega \iff \ell$ *is divergence-based*

Aside: can use this framework for data mining competitions!

Quantifying the Wagers

In standard market maker model, prices *adapt* to trades

From NIPS 2011, we can describe the net *profit* of such a trade in terms of the change $\mathbf{p} \to \mathbf{p}'$ in the *prices*...

... as the drop in a divergence-based loss!

Theorem (Abernethy, F.)

Traders have profit $\ell[\mathbf{p}]_\omega - \ell[\mathbf{p}']_\omega \iff \ell$ *is divergence-based*

Aside: can use this framework for data mining competitions!

Tying it all together

NIPS 2011:

Traders have profit $\ell[\mathbf{p}]_\omega - \ell[\mathbf{p}']_\omega \iff \ell$ divergence-based

COLT 2012:

ℓ divergence-based $\iff$ ℓ proper loss for linear Γ

Hence, prediction markets $\overset{\text{1 to 1}}{\iff}$ proper losses for means!

i.e. Prediction Markets $\iff$ *Market Scoring Rules*

Tying it all together

NIPS 2011:

Traders have profit $\ell[\mathbf{p}]_\omega - \ell[\mathbf{p}']_\omega \iff \ell$ divergence-based

COLT 2012:

ℓ divergence-based $\iff$ ℓ proper loss for linear Γ

Hence, prediction markets $\overset{\text{1 to 1}}{\iff}$ proper losses for means!

i.e. Prediction Markets $\iff$ *Market Scoring Rules*

Tying it all together

NIPS 2011:

Traders have profit $\ell[\mathbf{p}]_\omega - \ell[\mathbf{p}']_\omega$ $\Longleftrightarrow$ ℓ divergence-based

COLT 2012:

ℓ divergence-based $\Longleftrightarrow$ ℓ proper loss for linear Γ

Hence, prediction markets $\overset{\text{1 to 1}}{\Longleftrightarrow}$ proper losses for means!

i.e. Prediction Markets $\Longleftrightarrow$ *Market Scoring Rules*

Thanks!