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 $\ell[p]_y$
- $\ell$ is proper if $p = \arg\min_q \{ \ell[q] p \}$

Example: log loss
- Take $\ell[p]_y = - \log p_y$
- Now $\ell[q] p = - \sum p_y \log q_y = \text{KL}(p\|q) + H(p)$
  Minimized at $q = p$
Proper Losses

Typical setting: classification

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

Example: log loss

- Take $\ell[p]_y = -\log p_y$
- Now $\ell[q]p = -\sum p_y \log q_y = \text{KL}(p\|q) + H(p)$
  
  Minimized at $q = p$
Proper Losses

Typical setting: classification

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

Example: log loss

- Take $\ell[p]_y = -\log p_y$
- Now $\ell[q]p = -\sum p_y \log q_y = \text{KL}(p\|q) + H(p)$
  Minimized at $q = p$
Proper Losses

Typical setting: classification

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

Example: log loss

- Take $l[p]_y = -\log p_y$
- Now $l[q]_p = -\sum p_y \log q_y = KL(p\|q) + H(p)$ 
  *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 $\nu \in \mathcal{V}$
- Loss $l : \mathcal{V} \to \mathbb{R}^\Omega$ loss of $\nu$ and $\omega$ is $l[\nu]_\omega$
- $l$ is $\Gamma$-proper if $\Gamma(p) = \arg\min_{\nu} \{ l[\nu] p \}$

We will consider linear $\Gamma$:
- $\Gamma(p) = \mathbb{E}_{\omega \sim p} [\phi(\omega)]$ for some $\phi : \Omega \to \mathcal{V}$ i.e. means
Our setting: properties of distributions

- Outcomes $\omega \in \Omega$
- Distributional property $\Gamma : \Delta_\Omega \rightarrow \mathcal{V} \subseteq \mathbb{R}^k$ summary information
- Prediction $\nu \in \mathcal{V}$
- Loss $\ell : \mathcal{V} \rightarrow \mathbb{R}^\Omega$ loss of $\nu$ and $\omega$ is $\ell[\nu]_\omega$
- $\ell$ is $\Gamma$-proper if $\Gamma(p) = \arg\min_{\nu} \{ \ell[\nu]_p \}$

We will consider linear $\Gamma$:

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

A Characterization of Proper Losses for Linear Properties

Our goal

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

... Why bother?
**Motivation: Proper**

Proper losses are *well-calibrated*

Example: learning a coin’s bias $p$
- Want $\ell$ to measure *performance*
- After $N \gg 1$ 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: Proper

Proper losses are *well-calibrated*

Example: learning a coin’s bias $p$

- Want $\ell$ to measure *performance*
- After $N \gg 1$ 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 $q$
Proper losses are **well-calibrated**

Example: learning a coin’s bias $p$

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

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

"expected" loss of predicting $q$
Motivation: Characterization

Loss should \textit{quantify} error

Two losses for eliciting a mean

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

Very different notions of error

Given a notion of error, when can I \textit{design} a proper loss to match?
Loss should *quantify* error

Two losses for eliciting a mean

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

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:** \( \ell[v]_\omega = KL(\omega || v) \)

Very different notions of error

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

Two losses for eliciting a mean

- **Squared:** \( \ell[\nu]_\omega = (\nu - \omega)^2 \)
- **Log:** \( \ell[\nu]_\omega = \text{KL}(\omega \| \nu) \)

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:** $\ell[\nu]_\omega = (\nu - \omega)^2$
- **Log:** $\ell[\nu]_\omega = \text{KL}(\omega \| \nu)$

Very different notions of error

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

Two losses for eliciting a mean

- **Squared**: \( \ell [\nu]_\omega = (\nu - \omega)^2 \)
- **Log**: \( \ell [\nu]_\omega = \text{KL}(\omega \| \nu) \)

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 ($\infty$) number of classes?

*E.g. Price of gas next month?*

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

*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 ($\infty$) 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 ($\infty$) 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

Functional properties of Gamma

- Linear
  - Identity
  - K > 1: Gneiting and Raftery (2007), Vernet, Reid, Williamson (2011)

- Nonlinear
  - Open

This talk
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
Given convex $f : \mathcal{V} \rightarrow \mathbb{R}$, 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 \textit{divergence-based} if \( \exists f, \phi \) s.t.

\[
l[\nu]_\omega = D_f(\phi(\omega), \nu)
\]

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

\[
\begin{align*}
\arg\min_{\nu} \{ l[\nu]p \} \\
= \arg\min_{\nu} \left\{ \mathbb{E}_{\omega \sim p} \left[ f(\phi(\omega)) - f(\nu) - \nabla f(\nu) \cdot (\phi(\omega) - \nu) \right] \right\} \\
= \arg\min_{\nu} \left\{ -f(\nu) - \nabla f(\nu) \cdot (\Gamma(p) - \nu) \right\} \\
= \arg\min_{\nu} \left\{ D_f(\Gamma(p), \nu) - f(\Gamma(p)) \right\} = \Gamma(p)
\end{align*}
\]
Divergences and means

Definition: \( \ell \) is **divergence-based** if \( \exists f, \phi \) s.t.

\[
\ell [\nu]_{\omega} = D_f(\phi(\omega), \nu)
\]

Fact: this \( \ell \) is proper for linear property \( \Gamma(p) = E_{\omega \sim p}[\phi(\omega)] \)

\[
\begin{align*}
\arg\min_{\nu} \{ \ell [\nu]_p \} \\
= \arg\min_{\nu} \left\{ \mathbb{E}_{\omega \sim p} \left[ f(\phi(\omega)) - f(\nu) - \nabla f(\nu) \cdot (\phi(\omega) - \nu) \right] \right\} \\
= \arg\min_{\nu} \left\{ -f(\nu) - \nabla f(\nu) \cdot (\Gamma(p) - \nu) \right\} \\
= \arg\min_{\nu} \left\{ D_f(\Gamma(p), \nu) - f(\Gamma(p)) \right\} = \Gamma(p)
\end{align*}
\]
Divergences and means

Definition: \( l \) is \textit{divergence-based} if \( \exists f, \phi \) s.t.

\[
l[\nu]_\omega = D_f(\phi(\omega), \nu)
\]

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

\[
\arg\min_{\nu} \{ l[\nu]p \} \\
= \arg\min_{\nu} \left\{ \mathbb{E}_{\omega \sim p} \left[ f(\phi(\omega)) - f(\nu) - \nabla f(\nu) \cdot (\phi(\omega) - \nu) \right] \right\} \\
= \arg\min_{\nu} \{ -f(\nu) - \nabla f(\nu) \cdot (\Gamma(p) - \nu) \} \\
= \arg\min_{\nu} \{ D_f(\Gamma(p), \nu) - f(\Gamma(p)) \} = \Gamma(p)
\]
Divergences and means

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

\[
l[\nu]_\omega = D_f(\phi(\omega), \nu)
\]

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

\[
\arg\min_{\nu} \{l[\nu]p\} = \arg\min_{\nu} \left\{ \mathbb{E}_{\omega \sim p} \left[ f(\phi(\omega)) - f(\nu) - \nabla f(\nu) \cdot (\phi(\omega) - \nu) \right] \right\} \\
= \arg\min_{\nu} \{ -f(\nu) - \nabla f(\nu) \cdot (\Gamma(p) - \nu) \} \\
= \arg\min_{\nu} \{ D_f(\Gamma(p), \nu) - f(\Gamma(p)) \} = \Gamma(p)
\]
Divergences and means

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

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

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

$$\arg\min_{\nu} \{ \ell[\nu]p \}$$

$$= \arg\min_{\nu} \left\{ \mathbb{E}_{\omega \sim p} \left[ f(\phi(\omega)) - f(\nu) - \nabla f(\nu) \cdot (\phi(\omega) - \nu) \right] \right\}$$

$$= \arg\min_{\nu} \left\{ -f(\nu) - \nabla f(\nu) \cdot (\Gamma(p) - \nu) \right\}$$

$$= \arg\min_{\nu} \left\{ D_f(\Gamma(p), \nu) - f(\Gamma(p)) \right\} = \Gamma(p)$$
Divergences and means

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

\[
l[\nu]_\omega = D_f(\phi(\omega), \nu)
\]

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

\[
\arg\min_{\nu} \{ l[\nu] p \} \\
= \arg\min_{\nu} \left\{ \mathbb{E}_{\omega \sim p} \left[ f(\phi(\omega)) - f(\nu) - \nabla f(\nu) \cdot (\phi(\omega) - \nu) \right] \right\} \\
= \arg\min_{\nu} \{ -f(\nu) - \nabla f(\nu) \cdot (\Gamma(p) - \nu) \} \\
= \arg\min_{\nu} \{ D_f(\Gamma(p), \nu) - f(\Gamma(p)) \} = \Gamma(p)
\]
Characterization for linear properties

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

Q: Is every $\Gamma$-proper loss $\ell$ divergence-based?

A: Yes$^1$!

Theorem (Abernethy, F.)

$\ell$ is $\Gamma$-proper for linear $\Gamma$ $\iff$ $\ell$ is divergence-based

$^1$with extremely weak differentiability assumptions
This shows divergence-based $\implies \Gamma$-proper for some linear $\Gamma$

**Q:** Is every $\Gamma$-proper loss $\ell$ divergence-based?

**A:** Yes\(^1\)!

---

**Theorem (Abernethy, F.)**

$\ell$ is $\Gamma$-proper for linear $\Gamma$ $\iff \ell$ 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)**

\[ \ell : \Delta\Omega \times \Omega \rightarrow \mathbb{R} \text{ proper} \implies \ell \text{ is divergence-based} \]

Their proof:

- Extract \( f(p) = \ell[p] p \) \text{ Bayes risk, concave} \\
- Observe \( \ell[p] p + \ell[p](q-p) \geq \ell[q] q \) \text{ 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)

\[
\ell : \Delta_\Omega \times \Omega \to \mathbb{R} \text{ proper} \implies \ell \text{ is divergence-based}
\]

Their proof:

- Extract \( f(p) = \ell[p]p \) \textit{Bayes risk, concave}

- Observe \( \ell[p]p + \ell[p](q - p) \geq \ell[q]q \) \textit{from propriety}

- Hence \( \ell[p] \) is a gradient of \( f \) \implies \text{divergence}
Proof Intuition

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

Theorem (Gneiting and Raftery, 2010)

\[
\ell : \Delta_\Omega \times \Omega \rightarrow \mathbb{R} \text{ proper} \implies \ell \text{ is divergence-based}
\]

Their proof:

- Extract \( f(p) = \ell[p] p \) \( \text{Bayes risk, concave} \)

- Observe \( \ell[p] p + \ell[p] (q - p) \geq \ell[q] q \) \( \text{from propriety} \)

- Hence \( \ell[p] \) is a gradient of \( f \) \( \implies \text{divergence} \)
Proof Intuition

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

**Theorem (Gneiting and Raftery, 2010)**

\[
\ell : \Delta_\Omega \times \Omega \rightarrow \mathbb{R} \text{ proper} \implies \ell \text{ is divergence-based}
\]

Their proof:

- Extract \( f(p) = \ell[p]p \) \text{ Bayes risk, concave} 

- Observe \( \ell[p]p + \ell[p](q-p) \geq \ell[q]q \) \text{ from propriety}

- Hence \( \ell[p] \) is a gradient of \( f \) \implies \text{ divergence}
Proof Intuition

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

\[ \ell : \Delta_\Omega \times \Omega \rightarrow \mathbb{R} \text{ proper} \implies \ell \text{ is divergence-based} \]

Their proof:

- Extract \( f(p) = \ell[p] \, p \) \( \text{Bayes risk, concave} \)
- Observe \( \ell[p] \, p + \ell[p] (q-p) \geq \ell[q] \, q \) \( \text{from propriety} \)
- \( f(p) \partial f(p) \)
- Hence \( \ell[p] \) is a gradient of \( f \) \( \implies \text{divergence} \)
Proof Intuition

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

**Theorem (Gneiting and Raftery, 2010)**

\[\ell : \Delta_{\Omega} \times \Omega \rightarrow \mathbb{R} \text{ proper} \implies \ell \text{ is divergence-based}\]

Their proof:

- Extract \( f(p) = \ell[p] p \) \hspace{1cm} \text{Bayes risk, concave}

- Observe \( \ell[p] p + \ell[p](q-p) \geq \ell[q] q \) \hspace{1cm} \text{from propriety}

\[f(p) \quad \partial f(p) \quad f(q)\]

- Hence \( \ell[p] \) is a gradient of \( f \)! \hspace{1cm} \implies \text{divergence}
We draw intuition from the identity case \( i.e. \Gamma(p) = p \)

**Theorem (Gneiting and Raftery, 2010)**

\[ \ell : \Delta_{\Omega} \times \Omega \rightarrow \mathbb{R} \text{ proper} \implies \ell \text{ is divergence-based} \]

Their proof:

- Extract \( f(p) = \ell[p] \cdot p \) \( \text{Bayes risk, concave} \)

- Observe \( \ell[p] \cdot p + \ell[p](q-p) \geq \ell[q] \cdot q \) \( \text{from propriety} \)

- Hence \( \ell[p] \) is a gradient of \( f \! \equiv \! \text{divergence} \)
Proof Intuition

Their proof:

Extract \[ f(p) = \ell[p] p \]

**Challenge:** How to define \( f \) when \( \nu \neq \Delta_\Omega \)?

- Let \( \hat{\nu} \) such that \( \Gamma \circ \hat{\nu} \equiv \text{id}_\nu \)
  - A “family” of distributions with “parameter space” \( \nu \)
- Now \( f(\nu) = \ell[\nu] \hat{\nu}[\nu] \)
Proof Intuition

Their proof:

Extract

\[ f(p) = \ell[p] p \]

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

- Let \( \hat{\rho} \) such that \( \Gamma \circ \hat{\rho} \equiv \text{id}_{\mathcal{V}} \)

  *A “family” of distributions with “parameter space” \( \mathcal{V} \)*

- Now \( f(\mathcal{V}) = \ell[\mathcal{V}] \hat{\rho}[\mathcal{V}] \)
Switching Gears: Prediction Markets

<table>
<thead>
<tr>
<th>Candidate</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obama</td>
<td>$1</td>
</tr>
<tr>
<td>Romney</td>
<td>$1</td>
</tr>
<tr>
<td>Johnson</td>
<td>$1</td>
</tr>
</tbody>
</table>

$0.22 + $0.35 + $0.43 = 1

- Traders buy and sell these contracts
- Prices reflect the consensus prediction
Switching Gears: Prediction Markets

Traders buy and sell these contracts

Prices reflect the consensus prediction

Obama $1
Romney $1
Johnson $1

$0.22 + $0.35 + $0.43 = 1
Switching Gears: Prediction Markets

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

Obama: $1
Romney: $1
Johnson: $1

$0.22 + $0.35 + $0.43 = 1
Switching Gears: Prediction Markets

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

Obama $1
Romney $1
Johnson $1

$0.22 + $0.35 + $0.43 = 1
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 $p \rightarrow p'$ in the prices...

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

Theorem (Abernethy, F.)

Traders have profit $l[p]_\omega - l[p']_\omega \iff l$ 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 $p \rightarrow p'$ in the prices...

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

Theorem (Abernethy, F.)

Traders have profit $l[p]_\omega - l[p']_\omega \iff l$ 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 $p \rightarrow p'$ in the *prices*...

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

**Theorem (Abernethy, F.)**

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

*Aside: can use this framework for data mining competitions!*
Tying it all together

NIPS 2011:

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

COLT 2012:

$\ell$ divergence-based $\iff \ell$ proper loss for linear $\Gamma$

Hence, prediction markets $\iff$ proper losses for means!

i.e. Prediction Markets $\iff$ Market Scoring Rules
Tying it all together

NIPS 2011:

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

COLT 2012:

$\ell$ divergence-based $\iff \ell$ proper loss for linear $\Gamma$

Hence, prediction markets $\iff$ proper losses for means!

i.e. Prediction Markets $\iff$ Market Scoring Rules
NIPS 2011:

Traders have profit $l[p] \omega - l[p'] \omega \iff l$ divergence-based

COLT 2012:

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

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

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