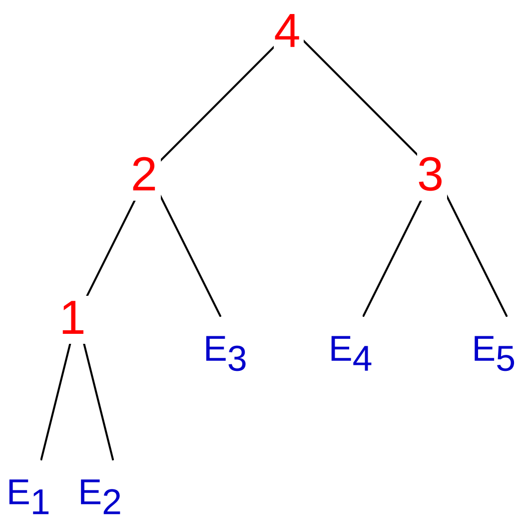
Penalty Functions for Inference Algorithms

Robert McEliece and Jeremy Thorpe

California Institute of Technology
Pasadena, California USA



Hattori Information Processing Laboratory
S & S Architecture Center
Sony Corporation, Tokyo
Thursday, Sept. 19, 2002

A Puzzle

- $\boldsymbol{X} = (X_1, X_2, X_3, X_4, X_5, X_6)$ is an unknown binary vector with components ± 1, selected randomly from the 64 possibilities. Thus initially we have

$$\Pr\{X_i = +1\} = \Pr\{X_i = -1\} = 1/2, \quad \text{for } i = 1, \dots, 6.$$

?

A Puzzle

- $\boldsymbol{X} = (X_1, X_2, X_3, X_4, X_5, X_6)$ is an unknown binary vector with components ± 1, selected randomly from the 64 possibilities. Thus initially we have

$$\Pr\{X_i = +1\} = \Pr\{X_i = -1\} = 1/2, \quad \text{for } i = 1, \ldots, 6.$$

- Suppose we gather the following "evidence" about $\boldsymbol{X}$:

$$\begin{aligned} Y_1 &= X_1 + X_2 - X_3 &&= -1 \\ Y_2 &= X_3 - X_4 - X_5 &&= +1 \\ Y_3 &= X_1 + X_5 + X_6 &&= +1 \end{aligned}$$

A Puzzle

- $\boldsymbol{X} = (X_1, X_2, X_3, X_4, X_5, X_6)$ is an unknown binary vector with components ± 1, selected randomly from the 64 possibilities. Thus initially we have

$$\Pr\{X_i = +1\} = \Pr\{X_i = -1\} = 1/2, \quad \text{for } i = 1, \ldots, 6.$$

- Suppose we gather the following "evidence" about $\boldsymbol{X}$:

$$\begin{aligned} Y_1 &= X_1 + X_2 - X_3 &&= -1 \\ Y_2 &= X_3 - X_4 - X_5 &&= +1 \\ Y_3 &= X_1 + X_5 + X_6 &&= +1 \end{aligned}$$

- What can be deduced about $\boldsymbol{X}$, given $\boldsymbol{Y} = (Y_1, Y_2, Y_3)$?

The Puzzle

$$
\begin{aligned}
X_1 + X_2 - X_3 \qquad\qquad\qquad\qquad &= -1 \\
X_3 - X_4 - X_5 \qquad &= +1 \\
X_1 \qquad\qquad\qquad\qquad + X_5 + X_6 &= +1
\end{aligned}
$$

The Puzzle

$$
\begin{aligned}
X_1 + X_2 - X_3 \qquad\qquad\qquad\qquad &= -1 \\
X_3 - X_4 - X_5 \qquad &= +1 \\
X_1 \qquad\qquad\qquad\qquad + X_5 + X_6 &= +1
\end{aligned}
$$

Anybody ? ...

Answer to the Puzzle

- Answer: Only three values of $\boldsymbol{X}$ are consistent with the evidence $\boldsymbol{Y}$:

$$(+1, -1, +1, +1, -1, +1)$$
$$(+1, -1, +1, -1, +1, -1)$$
$$(-1, +1, +1, -1, +1, +1),$$

Answer to the Puzzle

- Answer: Only three values of $\boldsymbol{X}$ are consistent with the evidence $\boldsymbol{Y}$:

$$(+1, -1, +1, +1, -1, +1)$$
$$(+1, -1, +1, -1, +1, -1)$$
$$(-1, +1, +1, -1, +1, +1),$$

so that

$$\Pr\{X_1 = +1 | \boldsymbol{Y}\} = 2/3$$
$$\Pr\{X_2 = +1 | \boldsymbol{Y}\} = 1/3$$
$$\Pr\{X_3 = +1 | \boldsymbol{Y}\} = 1$$
$$\Pr\{X_4 = +1 | \boldsymbol{Y}\} = 1/3$$
$$\Pr\{X_5 = +1 | \boldsymbol{Y}\} = 2/3$$
$$\Pr\{X_6 = +1 | \boldsymbol{Y}\} = 2/3.$$

The General Puzzle – A Simplified Sensor Network

- $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$, a list of N uniform i.i.d. ± 1 random variables, represents an unknown enviroment.

The General Puzzle – A Simplified Sensor Network

- $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$, a list of N uniform i.i.d. ± 1 random variables, represents an unknown enviroment.
- There are M remote sensors $S_1, \ldots, S_M$, each characterized by a sparse vector $\boldsymbol{S}_j = (s_{j,1}, \ldots, s_{j,n})$ whose nonzero components are ± 1.

The General Puzzle – A Simplified Sensor Network

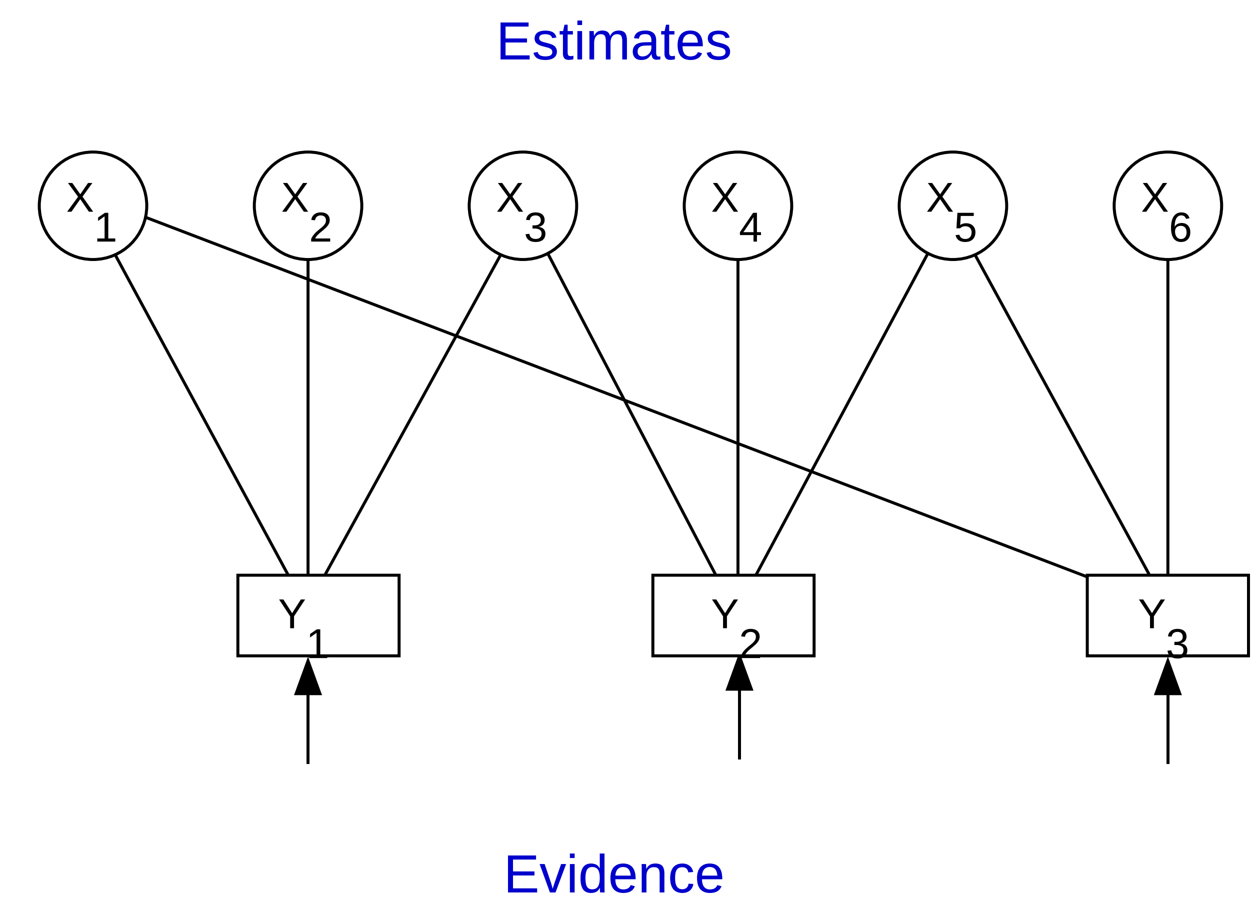
- $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$, a list of N uniform i.i.d. ± 1 random variables, represents an unknown enviroment.
- There are M remote sensors $S_1, \ldots, S_M$, each characterized by a sparse vector $\boldsymbol{S}_j = (s_{j,1}, \ldots, s_{j,n})$ whose nonzero components are ± 1.
- The measurement taken by the jth sensor $\boldsymbol{S}_j$ is $Y_j = \boldsymbol{X} \cdot \boldsymbol{S}_j$. Thus the aggregate evidence about the environment provided by the sensor network is the vector $Y = (Y_1, \ldots, Y_M)$.

The General Puzzle – A Simplified Sensor Network

- $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$, a list of N uniform i.i.d. ± 1 random variables, represents an unknown enviroment.
- There are M remote sensors $S_1, \ldots, S_M$, each characterized by a sparse vector $\boldsymbol{S}_j = (s_{j,1}, \ldots, s_{j,n})$ whose nonzero components are ± 1.
- The measurement taken by the jth sensor $\boldsymbol{S}_j$ is $Y_j = \boldsymbol{X} \cdot \boldsymbol{S}_j$. Thus the aggregate evidence about the environment provided by the sensor network is the vector $Y = (Y_1, \ldots, Y_M)$.
- Example. $N = 6$ and $M = 3$:

$$\begin{aligned}
\boldsymbol{S}_1 &= (+1, +1, -1, \ \ 0, \ \ 0, \ \ 0) \\
\boldsymbol{S}_2 &= (\ \ 0, \ \ 0, +1, -1, -1, \ \ 0) \\
\boldsymbol{S}_3 &= (+1, \ \ 0, \ \ 0, \ \ 0, +1, +1)
\end{aligned}$$

A Factor Graph for the Problem



An Experiment

- An "environment" $\boldsymbol{X} = (X_1, \ldots, X_N)$ is chosen at random from $\{+1, -1\}^N$.

An Experiment

- An "environment" $\boldsymbol{X} = (X_1, \ldots, X_N)$ is chosen at random from $\{+1, -1\}^N$.
- A "sensor network" $(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_M)$ is selected at random, where each $\boldsymbol{S}_j$ has exactly s nonzero (± 1) components.

An Experiment

- An "environment" $\boldsymbol{X} = (X_1, \ldots, X_N)$ is chosen at random from $\{+1, -1\}^N$.
- A "sensor network" $(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_M)$ is selected at random, where each $\boldsymbol{S}_j$ has exactly s nonzero (± 1) components.
- "Evidence" $Y_j = \boldsymbol{X} \cdot \boldsymbol{S}_j$, for $j = 1, \ldots, M$, is generated.

An Experiment

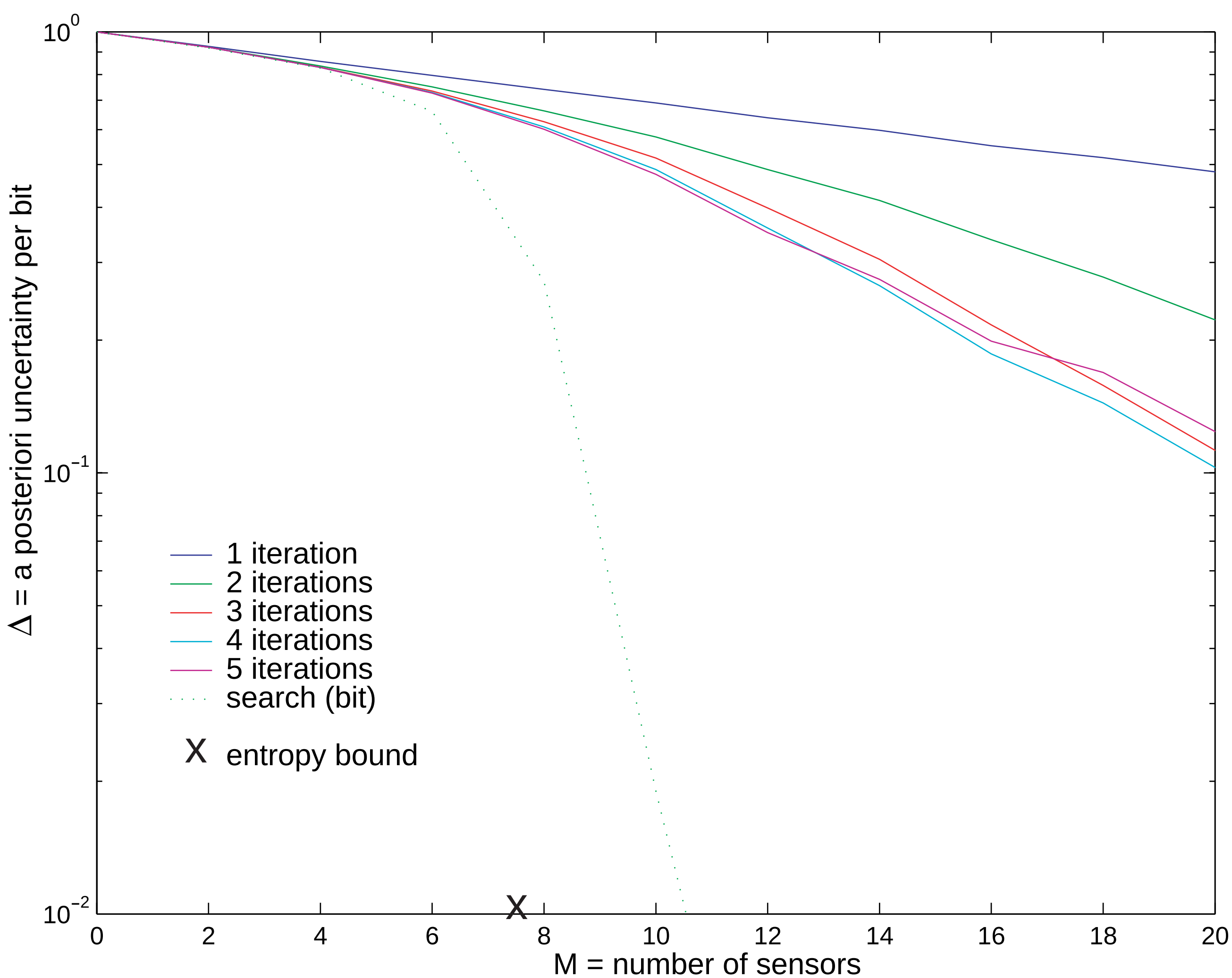
- An "environment" $\boldsymbol{X} = (X_1, \ldots, X_N)$ is chosen at random from $\{+1, -1\}^N$.
- A "sensor network" $(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_M)$ is selected at random, where each $\boldsymbol{S}_j$ has exactly s nonzero (± 1) components.
- "Evidence" $Y_j = \boldsymbol{X} \cdot \boldsymbol{S}_j$, for $j = 1, \ldots, M$, is generated.
- Belief propagation is run, using the evidence $(y_1, \ldots, y_M)$, and returns a set of beliefs (approximate a posteriori probabilities) $(b_i(+1), b_i(-1))$, for $i = 1, \ldots, N$.

An Experiment

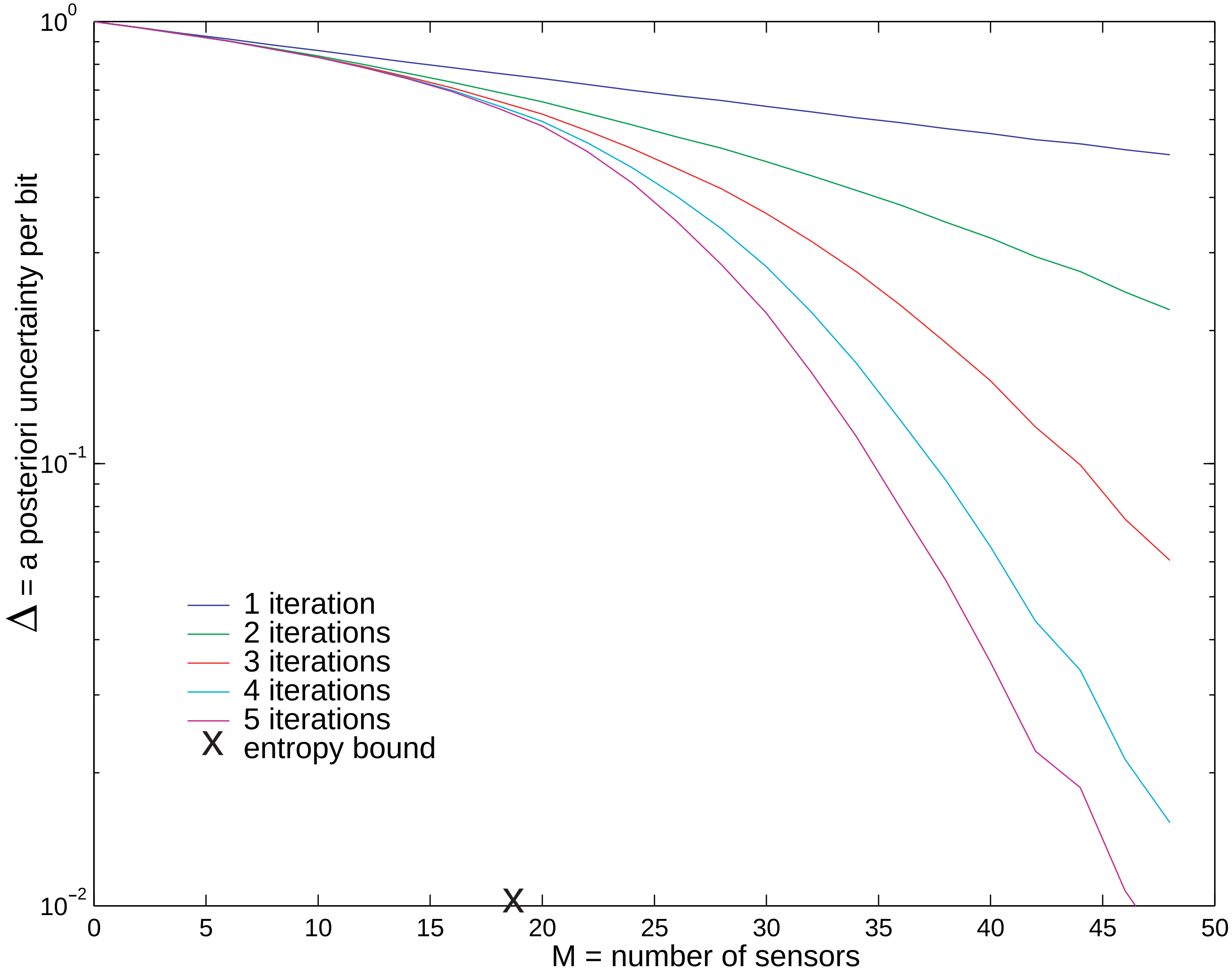
- An "environment" $\boldsymbol{X} = (X_1, \ldots, X_N)$ is chosen at random from $\{+1, -1\}^N$.
- A "sensor network" $(\boldsymbol{S}_1, \ldots, \boldsymbol{S}_M)$ is selected at random, where each $\boldsymbol{S}_j$ has exactly s nonzero (± 1) components.
- "Evidence" $Y_j = \boldsymbol{X} \cdot \boldsymbol{S}_j$, for $j = 1, \ldots, M$, is generated.
- Belief propagation is run, using the evidence $(y_1, \ldots, y_M)$, and returns a set of beliefs (approximate a posteriori probabilities) $(b_i(+1), b_i(-1))$, for $i = 1, \ldots, N$.
- The "score" of the algorithm's output is defined as

$$\Delta = -\frac{1}{N} \sum_{i=1}^{N} \log b_i(X_i).$$

Experimental Results, $N = 20$, $s = 10$



Experimental Results, $N = 50$, $s = 10$



Another Puzzle

Why did we use $-\log b(X)$ as the penalty measure?

Another Puzzle

Why did we use $-\log b(X)$ as the penalty measure?

Anybody?

The Oracle Judging Problem

- An event E with n possible outcomes $\{E_1, \ldots, E_n\}$ is about to occur.

The Oracle Judging Problem

- An event E with n possible outcomes $\{E_1, \ldots, E_n\}$ is about to occur.
- An oracle makes a "soft" prediction about the outcome:

$$\boldsymbol{b} = (b_1, \ldots, b_n)$$

(b_i = oracle's "belief" in outcome E_i, $\sum_i b_i = 1$.)

The Oracle Judging Problem

• An event E with n possible outcomes $\{E_1, \ldots, E_n\}$ is about to occur.

• An oracle makes a "soft" prediction about the outcome:

$$\boldsymbol{b} = (b_1, \ldots, b_n)$$

(b_i = oracle's "belief" in outcome E_i, $\sum_i b_i = 1$.)

• Problem: How shall we measure the accuracy (or inaccuracy) of $\boldsymbol{b}$ as a predictor of E?

Examples

- A bookmaker quoting odds on a horserace.

Examples

- A bookmaker quoting odds on a horserace.
- A TV weatherman predicting tomorrow's weather (Sunny, Cloudy, Rain, Typhoon, ...)

Examples

- A bookmaker quoting odds on a horserace.
- A TV weatherman predicting tomorrow's weather (Sunny, Cloudy, Rain, Typhoon, ...)
- A number-theorist predicting the number of distinct prime factors of an integer selected at random between 1 and 10^{20}.

Examples

- A bookmaker quoting odds on a horserace.
- A TV weatherman predicting tomorrow's weather (Sunny, Cloudy, Rain, Typhoon, ...)
- A number-theorist predicting the number of distinct prime factors of an integer selected at random between 1 and 10^{20}.
- A Go-playing algorithm predicting the next move of a 9-dan player.

Examples

- A bookmaker quoting odds on a horserace.
- A TV weatherman predicting tomorrow's weather (Sunny, Cloudy, Rain, Typhoon, ...)
- A number-theorist predicting the number of distinct prime factors of an integer selected at random between 1 and 10^{20}.
- A Go-playing algorithm predicting the next move of a 9-dan player.
- A soft-decision decoding algorithm predicting whether an information bit is a zero or a one.

$$\vdots$$

Examples

- A bookmaker quoting odds on a horserace.
- A TV weatherman predicting tomorrow's weather (Sunny, Cloudy, Rain, Typhoon, ...)
- A number-theorist predicting the number of distinct prime factors of an integer selected at random between 1 and 10^{20}.
- A Go-playing algorithm predicting the next move of a 9-dan player.
- A soft-decision decoding algorithm predicting whether an information bit is a zero or a one.

$$\vdots$$

Penalty Functions for Oracles

- Let $\Delta_i(\boldsymbol{b})$ denote the penalty charged to the belief vector $\boldsymbol{b} = (b_1, \ldots, b_n)$ if the actual outcome is E_i. For example,

$$\begin{aligned} \Delta_i(\boldsymbol{b}) &= -\log b_i \\ &= (1 - b_i)^2 \\ &= 1/b_i - 1 \\ &\vdots \end{aligned}$$

Penalty Functions for Oracles

- Let $\Delta_i(\boldsymbol{b})$ denote the penalty charged to the belief vector $\boldsymbol{b} = (b_1, \ldots, b_n)$ if the actual outcome is E_i. For example,

$$\begin{aligned}\Delta_i(\boldsymbol{b}) &= -\log b_i \\ &= (1 - b_i)^2 \\ &= 1/b_i - 1 \\ &\vdots\end{aligned}$$

What properties should $\Delta_i(\boldsymbol{b})$ have?

Some Innocuous Restrictions

Property 1.

$$\Delta_i(\boldsymbol{b}) \geq 0,$$

i.e., penalties are nonnegative.

Some Innocuous Restrictions

Property 1.

$$\Delta_i(\boldsymbol{b}) \geq 0,$$

i.e., penalties are nonnegative.

Property 2.

$$\Delta_i(\boldsymbol{b}) = 0 \quad \textit{if } b_i = 1,$$

i.e., there is no penalty if the oracle predicts the actual outcome with certainty.

The Sine Qua Non Axiom

Axiom 0. *If* $\boldsymbol{p} = (p_1, \ldots, p_n)$ *and* $\boldsymbol{b} = (b_1, \ldots, b_n)$ *are arbitrary probability vectors,*

$$\sum_{i=1}^{n} p_i \Delta_i(\boldsymbol{p}) \leq \sum_{i=1}^{n} p_i \Delta_i(\boldsymbol{b})$$

$$(\boldsymbol{p} \cdot \boldsymbol{\Delta}(\boldsymbol{p}) \leq \boldsymbol{p} \cdot \boldsymbol{\Delta}(\boldsymbol{b})),$$

with equality iff $\boldsymbol{b} = \boldsymbol{p}$.

The Sine Qua Non Axiom

Axiom 0. *If* $\boldsymbol{p} = (p_1, \ldots, p_n)$ *and* $\boldsymbol{b} = (b_1, \ldots, b_n)$ *are arbitrary probability vectors,*

$$\sum_{i=1}^{n} p_i \Delta_i(\boldsymbol{p}) \leq \sum_{i=1}^{n} p_i \Delta_i(\boldsymbol{b})$$
$$(\boldsymbol{p} \cdot \boldsymbol{\Delta}(\boldsymbol{p}) \leq \boldsymbol{p} \cdot \boldsymbol{\Delta}(\boldsymbol{b})),$$

with equality iff $\boldsymbol{b} = \boldsymbol{p}$.

This says that if the event E has *a priori* probability density

$$\Pr\{E = E_i\} = p_i,$$

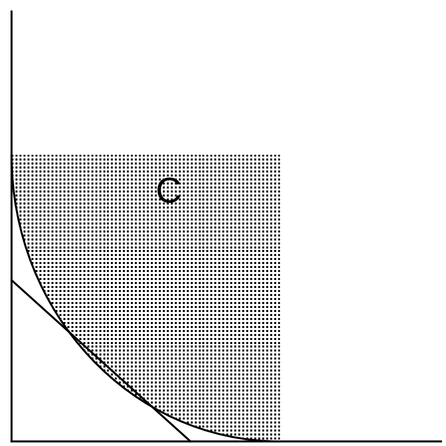
then the belief about E with the minimum average penalty is $\boldsymbol{p}$ itself.

Unfortunately, there are Infinitely Many Penalty Functions Which Satisfy the SQN Axiom:

Theorem. *If C is a convex set in R^n, like the one depicted,*

$$\boldsymbol{\Delta}(\boldsymbol{p}) = \underset{\boldsymbol{c} \in C}{\operatorname{argmin}}(\boldsymbol{p} \cdot \boldsymbol{c})$$

satisfies Axiom 0. (And Conversely.)



e.g. $C = \{x_i : \sum_i 2^{-x_i} \leq 1\}$

Another Axiom is needed!

A Possible Further Axiom:

Axiom 1. *(The No Partial Credit Axiom.)*

$$\Delta_i(\boldsymbol{b}) = f(b_i)$$

for some continuous function $f(x)$.

In words, the penalty assessed depends only on the belief assigned by the oracle to the actual outcome (and therefore not on the beliefs in the other outcomes).

This seems reasonable, at least when there is no notion of "closeness" of an incorrect prediction.

Now we're getting somewhere!

Theorem. *If Axioms 0 and 1 hold, and if $n \geq 3$, then the only possible penalty function is*

$$\Delta_i(\boldsymbol{b}) = -\log b_i\,,$$

where the base of the logarithm is arbitrary.

Corollary. *In this case we have*

$$\boldsymbol{p} \cdot \boldsymbol{\Delta}(\boldsymbol{b}) = H(\boldsymbol{p}) + D(\boldsymbol{p} \parallel \boldsymbol{b}),$$

i.e., if the a priori probability vector is $\boldsymbol{p}$, the minimum expected penalty is $H(\boldsymbol{p})$, with equality only for the oracle whose belief exactly matches the a priori *distribution.*

(The Case $n = 2$ is Ugly)

Theorem. *Let $n = 2$. Then Axioms 0 and 1 hold if and only if*

$$\Delta_i(\boldsymbol{b}) = f(b_i),$$

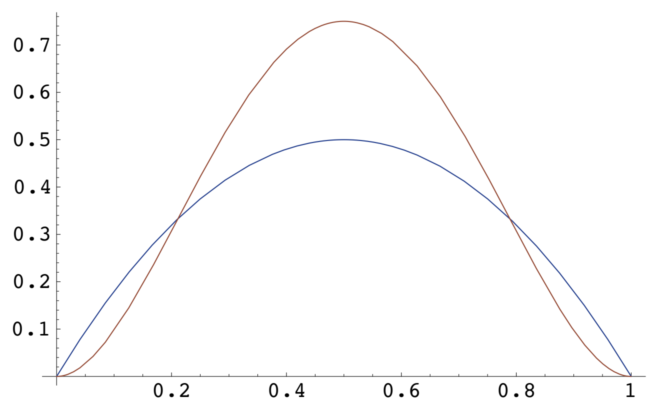
where $f(x)$ is of the form

$$f(x) = \int_x^1 \frac{g(t)}{t} dt,$$

for some function $g(t)$ which satisfies

$$g(t) \geq 0 \text{ and } g(t) = g(1-t) \text{ for all } t \in (0,1).$$

■



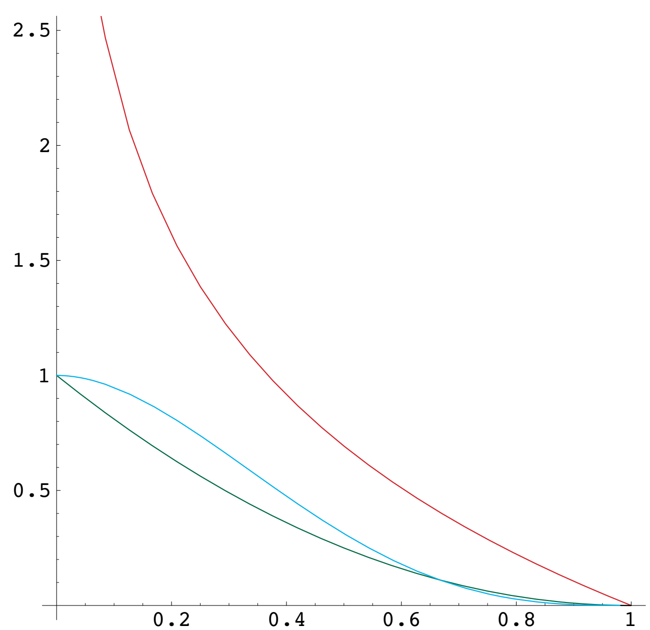
(Some $n = 2$ Examples)

$$g(t) = 1 \Rightarrow f(x) = -\log x$$

$$g(t) = 2t(1-t) \Rightarrow f(x) = (1-x)^2$$

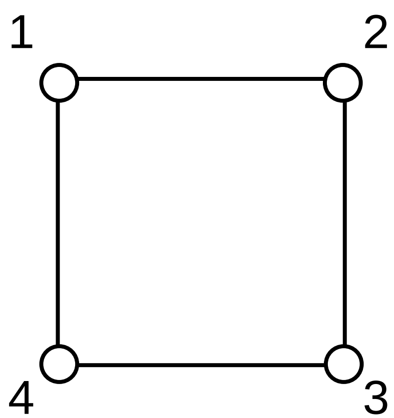
$$g(t) = 12t^2(1-t)^2 \Rightarrow f(x) = (1-x)^3(1+3x)$$

$$\vdots$$



An Extension — The Metric Case.

Suppose there is a distance measure $D = (d_{i,j})$ between the possible outcomes. (Some wrong guesses are better than others.)



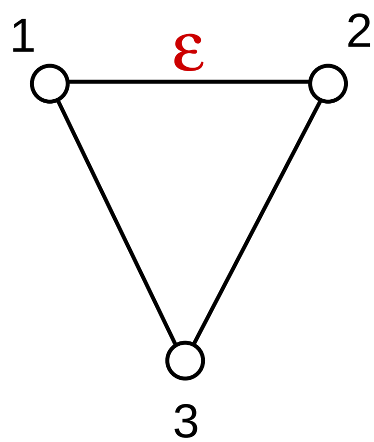
$$D = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \end{pmatrix} \end{array}$$

Call the corresponding penalty function $\boldsymbol{\Delta}(D, \boldsymbol{b})$.

What properties should $\boldsymbol{\Delta}(D, \boldsymbol{b})$ have?

The Metric Case —Continued.

- $\Delta_i(D, \boldsymbol{b}) = -\log b_i$ will no longer do:



$$D = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & \epsilon & 1 \\ \epsilon & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \end{array}$$

Arguably $\boldsymbol{b} = (b_1, b_2, b_3)$ and $\boldsymbol{b}' = (b_2, b_1, b_3)$ should be assessed identical penalties, as $\epsilon \to 0$.

The Metric Case —Continued.

- The average distance between the belief vector $\boldsymbol{b} = (b_1, \ldots, b_n)$. and the outcome i is

$$d(\boldsymbol{b}, i) = \sum_{j=1}^{n} d_{i,j} b_i,$$

- It is plausible to require that $\Delta_i(D, \boldsymbol{b})$ be an increasing function of $d(\boldsymbol{b}, i)$. (The farther $\boldsymbol{b}$ is from i, the larger the penalty.)
- So we might propose the

Strong Metric Axiom .

$$\Delta_i(\boldsymbol{b}) = f\left(d(\boldsymbol{b}, i)\right),$$

where $f(0) = 0$ and $f(x)$ is an increasing function of x.

The Strong Metric Axiom is Too Strong!

The Strong Metric Axiom plus the SQN axiom requires that

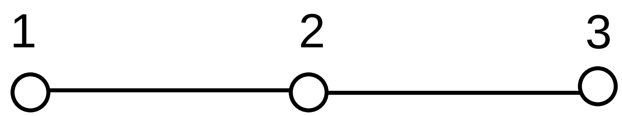
$$\sum_{i=1}^{n} p_i f\left(d(\boldsymbol{p}, i)\right) \leq \sum_{i=1}^{n} p_i f\left(d(\boldsymbol{b}, i)\right),$$

for all possible priors $\boldsymbol{p}$ and beliefs $\boldsymbol{b}$.

Unfortunately, this won't work.

Why it won't work.

- Example.



$$D = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix} \end{array}$$

Consider the two belief vectors $\boldsymbol{b}_1 = (1/3, 1/3, 1/3)$ and $\boldsymbol{b}_2 = (0, 1, 0)$, with average distances from $(1, 2, 3)$ given by

$$(d(\boldsymbol{b}_1, 1), d(\boldsymbol{b}_1, 2), d(\boldsymbol{b}_1, 3)) = (1, \frac{2}{3}, 1), \quad \text{and}$$
$$(d(\boldsymbol{b}_2, 1), d(\boldsymbol{b}_2, 2), d(\boldsymbol{b}_2, 3)) = (1, 0, 1).$$

If $\boldsymbol{b}_1$ is the prior, the SM axiom plus the SQN axiom therefore requires

$$\frac{1}{3}f(1) + \frac{1}{3}f(2/3) + \frac{1}{3}f(1) \leq \frac{1}{3}f(1) + \frac{1}{3}f(0) + \frac{1}{3}f(1),$$

i.e.,

$$f(2/3) \leq f(0) = 0,$$

which is false, since $f(x)$ is assumed increasing.

• In other words, if nature chooses $(1/3, 1/3, 1/3)$, the *metrically optimal* belief is $(0, 1, 0)$, and not $(1/3, 1/3, 1/3)$.

• (This is obvious to a game theorist.)

Some Weaker Metric Axioms

Again: There is a distance measure $D = (d_{i,j})$ between the possible outcomes. (Some wrong guesses are better than others.) What properties should $\boldsymbol{\Delta}(D, \boldsymbol{b})$ have?

- Continuous in D, i.e., if $D \to D'$, then $\boldsymbol{\Delta}(D, \boldsymbol{b}) \to \boldsymbol{\Delta}(D', \boldsymbol{b})$.
- Homogeneity: $\boldsymbol{\Delta}(\lambda D, \boldsymbol{b}) = \lambda \boldsymbol{\Delta}(D, \boldsymbol{b})$.
- Indistinguishable Outcomes: If $d_{i,j} = 0$ and $b_k = b'_k$ if $k \neq i, j$, then $\boldsymbol{\Delta}(D, \boldsymbol{b}) = \boldsymbol{\Delta}(D, \boldsymbol{b}')$.
- (No partial credit) If $D = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ then

$$\Delta_i(D, \boldsymbol{b}) = -\log b_i.$$

A Possibility ?

$$\Delta_i(\boldsymbol{b}) = -\log b_i$$

$$\Downarrow$$

$$\Delta_i(\boldsymbol{b}) = -\int_0^\infty \log b_i(r)dr,$$

where $b_i(r)$ denotes the probability that the oracle $\boldsymbol{b}$ is within distance r of the outcome i:

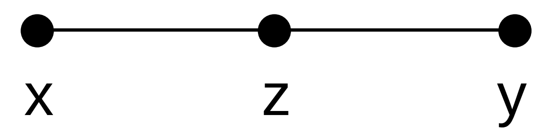
$$b_i(r) = \sum\{b_j : D_{i,j} \leq r\}.$$

This satisfies Properties (1)–(4) and satisfies SQN if D is an *ultrametric* !

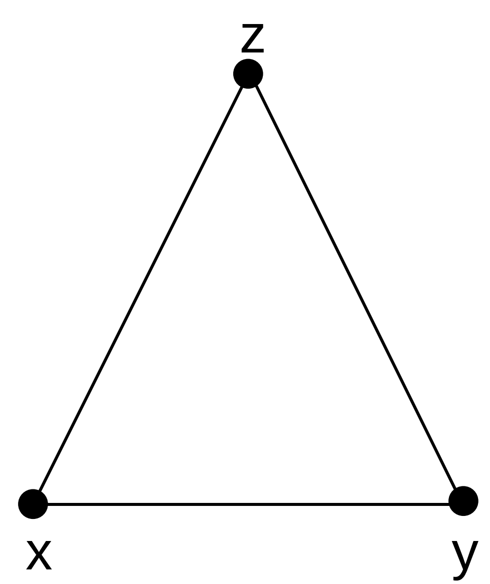
What is an Ultrametric?

- Ordinary metric triangle inequality:

$$d(x, y) \leq d(x, z) + d(z, y).$$



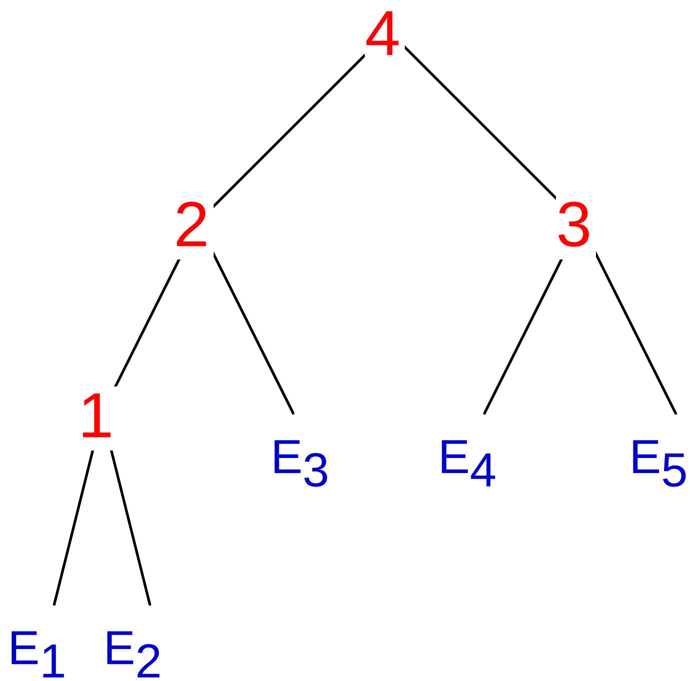
- Ultrametric triangle inequality:



$$d(x, y) \leq \max(d(x, z), d(z, y)).$$

(All triangles are isosceles.)

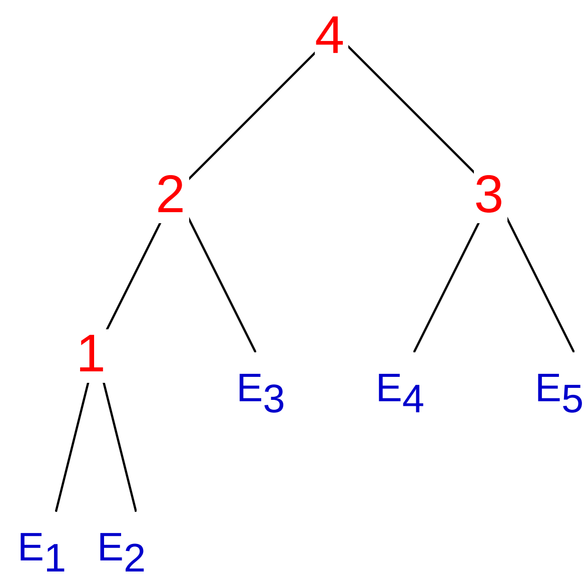
Ultrametrics Can be Represented by Trees



	E_1	E_2	E_3	E_4	E_5
E_1	0	1	2	4	4
E_2	1	0	2	4	4
E_3	2	2	0	4	4
E_4	4	4	4	0	3
E_5	4	4	4	3	0

Example of the Ultrametric Penalty Function

$$\Delta_i(\boldsymbol{b}) = -\int_0^\infty \log b_i(r)\, dr$$



$$\Delta_1(\boldsymbol{b}) = -\log b_1 - \log(b_1 + b_2) - 2\log(b_1 + b_2 + b_3)$$

$$\Delta_2(\boldsymbol{b}) = -\log b_2 - \log(b_1 + b_2) - 2\log(b_1 + b_2 + b_3)$$

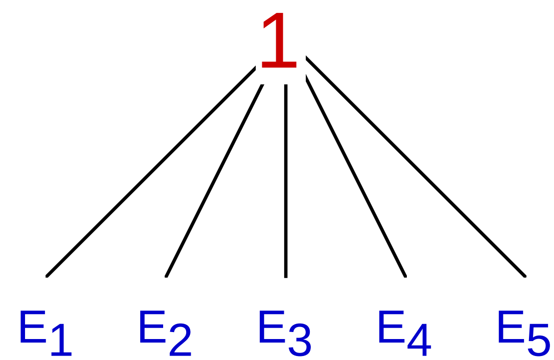
$$\Delta_3(\boldsymbol{b}) = -2\log b_3 - 2\log(b_1 + b_2 + b_3)$$

$$\Delta_4(\boldsymbol{b}) = -3\log b_4 - \log(b_4 + b_5)$$

$$\Delta_5(\boldsymbol{b}) = -3\log b_5 - \log(b_4 + b_5)$$

Another Example

$$\Delta_i(\boldsymbol{b}) = -\int_0^\infty \log b_i(r)\ dr$$



$$\Delta_1(\boldsymbol{b}) = -\log b_1$$
$$\Delta_2(\boldsymbol{b}) = -\log b_2$$
$$\Delta_3(\boldsymbol{b}) = -\log b_3$$
$$\Delta_4(\boldsymbol{b}) = -\log b_4$$
$$\Delta_5(\boldsymbol{b}) = -\log b_5$$

Conclusions and Open Questions

- We recommend using $\Delta_i(\boldsymbol{b}) = -\log b_i$, even with $n = 2$.

Conclusions and Open Questions

- We recommend using $\Delta_i(\boldsymbol{b}) = -\log b_i$, even with $n = 2$.

- Can the Weak Metric Axioms be satisfied for an ordianry metric?

Conclusions and Open Questions

- We recommend using $\Delta_i(\boldsymbol{b}) = -\log b_i$, even with $n = 2$.

- Can the Weak Metric Axioms be satisfied for an ordianry metric?
- Is the penalty function

$$\Delta_i(\boldsymbol{b}) = -\int_0^\infty \log b_i(r)\; dr$$

unique?

Conclusions and Open Questions

- We recommend using $\Delta_i(\boldsymbol{b}) = -\log b_i$, even with $n = 2$.

- Can the Weak Metric Axioms be satisfied for an ordianry metric?
- Is the penalty function

$$\Delta_i(\boldsymbol{b}) = -\int_0^\infty \log b_i(r)\; dr$$

unique?

- The End. Any Questions?