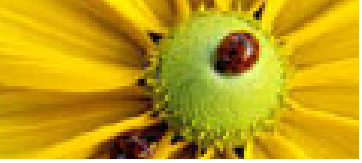


# Database Embedded Likelihood Models

Brian Dolan  
Director of Research Analytics  
Fox Interactive Media

21 January, 2009



# HAPPY OBAMA DAY!

What Is FIM?  
HAPPY OBAMA  
DAY!

What is FIM?

Motivation

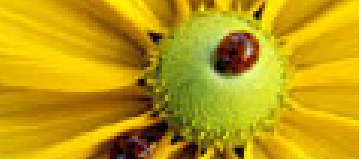
Likelihood Models

Example

The Guys

# HAPPY OBAMA DAY!

## Obamanos!



# What is FIM?

What Is FIM?  
HAPPY OBAMA  
DAY!

What is FIM?

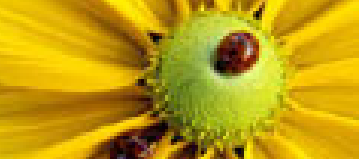
Motivation

Likelihood Models

Example

The Guys

- Fox Interactive Media is the parent company to MySpace, IGN, Rotten Tomatoes, American Idol.com and others.
- We are a major internet arm of News Corp.
- FIM provided me with a nice slide I didn't use.
- My group acts as the Subject Matter Experts on Statistics, Experiment Design, Machine Learning and Information Retrieval.



# Why Embed Your Analysis?

What Is FIM?

Motivation

Why Embed Your  
Analysis?

Likelihood Models

Example

The Guys

- Structured data lives in the database
- At large volumes, it is impractical to move the data out of the database
- Parallelization is free
- RDBMS have great triggering and automation support
- Other people manage the ETL process for you



# What is a Likelihood Model?

What Is FIM?

Motivation

Likelihood Models

What is a Likelihood Model?

Generalize This

Example

The Guys

Suppose you are trying to decide if a coin is fair. You flip it  $n = 12$  times to get

$$X = (T, H, H, H, H, T, T, H, T, H, H, T)$$

This leads you to two natural probabilities:

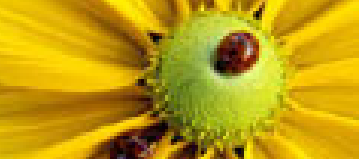
Null Model  $H_0$ : Probability of  $X$  if the coin is fair?

$$P(H) = 6/12$$

Alternate Model  $H_A$ : Probability of  $X$  if the coin is not fair?

$$P(H) = 7/12$$

Since we are dealing with a change in the parameters (e.g. where  $P(H)$  varies) we use the term “Likelihood” instead of “Probability”. There are deep reasons for this...



## Generalize This

What Is FIM?

Motivation

Likelihood Models  
What is a Likelihood  
Model?

Generalize This

Example

The Guys

A “Likelihood Ratio” is the ratio of two models of the same data set.

$$LLR \equiv \frac{P(X|H_0)}{P(X|H_A)}$$

The two models can be drastically different, such as “Multinomial” vs “Exponential”. Of course, they both should make sense with the data.



# Multinomial Distribution

What Is FIM?

Motivation

Likelihood Models

Example

Multinomial  
Distribution

Multinomial  
Likelihood Ratio

LLR Calculation

Enter The Data

The SQL For

$P(X|H_A)$

Get The LLR

The Guys

- The multinomial generalizes the Binomial distribution. The toy example is rolling a die.

- Roll a six-sided die 8 times with outcomes

$$X = (4, 5, 1, 2, 5, 4, 2, 6)$$

- This would lead you to formulate

$$H_0 : P(x = i) = 1/6, \quad i \in \{1, 2, 3, 4, 5, 6\}$$

$$H_A : P(x = 1) = 1/8, \quad P(x = 2) = 2/8,$$

$$P(x = 3) = 0/8, \quad P(x = 4) = 2/8,$$

$$P(x = 5) = 2/8, \quad P(x = 6) = 1/8$$



# Multinomial Likelihood Ratio

What Is FIM?

Motivation

Likelihood Models

Example

Multinomial  
Distribution

Multinomial  
Likelihood Ratio

LLR Calculation

Enter The Data

The SQL For

$P(X|H_A)$

Get The LLR

The Guys

Now we can calculate

$$P(X|H_0) = \binom{8}{1,2,2,2,1} 1^{1/6} 2^{1/6} 2^{1/6} 2^{1/6} 1^{1/6}$$

$$P(X|H_A) = \binom{8}{1,2,2,2,1} 1^{1/8} 2^{1/8} 2^{2/8} 2^{2/8} 1^{1/8}$$

The Likelihood Ratio is then

$$LR = \frac{P(X|H_0)}{P(X|H_A)}.$$

But we want the Log-Likelihood Ratio

$$LLR = -2 \log \left( \frac{P(X|H_0)}{P(X|H_A)} \right).$$



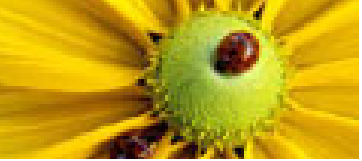
# LLR Calculation

Writing it out, you see how easy it is to implement in the database using standard SQL.

$$\begin{aligned} LLR &= -2 \log \left( \frac{P(X|H_0)}{P(X|H_A)} \right) \\ &= -2 \{ \log P(X|H_0) - \log P(X|H_A) \} \\ &= -2 \{ 1/6 \log(1) + 1/6 \log(2) \\ &\quad + 1/6 \log(2) + 1/6 \log(2) + 1/6 \log(1) \} \\ &\quad + 2 \{ 1/8 \log(1) + 2/8 \log(2) \\ &\quad + 2/8 \log(2) + 2/8 \log(2) \\ &\quad + 1/8 \log(1) \} \end{aligned}$$

if you skip some of the algebra.

- What Is FIM?
- Motivation
- Likelihood Models
- Example
- Multinomial Distribution
- Multinomial Likelihood Ratio
- LLR Calculation
- Enter The Data
- The SQL For  $P(X|H_A)$
- Get The LLR
- The Guys



# Enter The Data

What Is FIM?

Motivation

Likelihood Models

Example

Multinomial

Distribution

Multinomial

Likelihood Ratio

LLR Calculation

Enter The Data

The SQL For

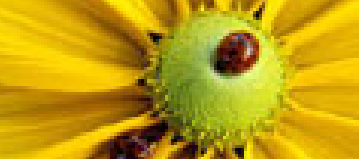
$P(X|H_A)$

Get The LLR

The Guys

	<i>value</i>	<i>model</i>	<i>p</i>
	1	<i>h0</i>	0.16666666666666667
	2	<i>h0</i>	0.16666666666666667
	3	<i>h0</i>	0.16666666666666667
	4	<i>h0</i>	0.16666666666666667
	5	<i>h0</i>	0.16666666666666667
	6	<i>h0</i>	0.16666666666666667
	1	<i>ha</i>	0.125
	2	<i>ha</i>	0.25
	3	<i>ha</i>	0
	4	<i>ha</i>	0.25
	5	<i>ha</i>	0.25
	6	<i>ha</i>	0.125

(12rows)



# The SQL For $P(X|H_A)$

What Is FIM?

Motivation

Likelihood Models

Example

Multinomial

Distribution

Multinomial

Likelihood Ratio

LLR Calculation

Enter The Data

The SQL For

$P(X|H_A)$

Get The LLR

The Guys

```
SELECT
model
,p.p * log(count(*)) AS ha_sum
,0::float as h0_sum
FROM
ra.lapug_data d
,ra.lapug_pdf p
WHERE
d.event = p.value
AND model = 'ha'
GROUP BY
model, p.p
```



# Get The LLR

What Is FIM?

Motivation

Likelihood Models

Example

Multinomial

Distribution

Multinomial

Likelihood Ratio

LLR Calculation

Enter The Data

The SQL For

$P(X|H_A)$

Get The LLR

The Guys

## ■ SELECT

```
sum(ha_sum) as ha_sum
```

```
,sum(ho_sum) as ho_sum
```

```
, 2 * (sum(ha_sum) - sum(ho_sum)) AS llr
```

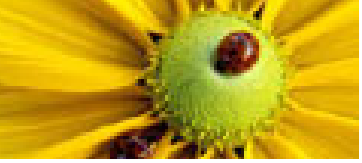
```
FROM (...) AS a
```

## ■ The result being

<i>ha_sum</i>	<i>ho_sum</i>	<i>llr</i>
0.23216	0.15051	0.16330

(1row)

And 0.16 is not statistically significant.



# The Guys

What Is FIM?

Motivation

Likelihood Models

Example

The Guys

The Guys

- Noelle Sio is a goddess. Her particular domain is Filipino Dance, so pray to her the next time you put on your poofy shirt.
- David Hubbard is the embodiment of Grace and Agility.