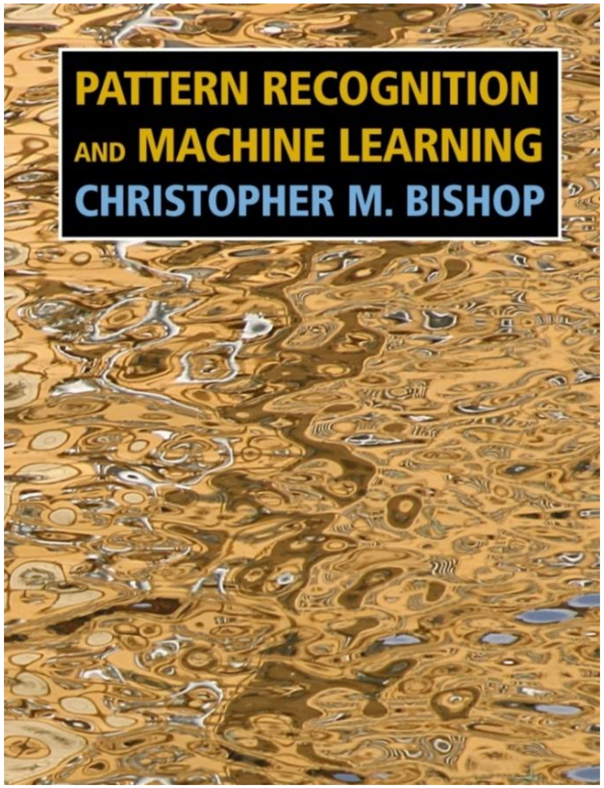
INTRODUCTION TO BAYESIAN INFERENCE – PART 1

CHRIS BISHOP

Please ask questions!



http://research.microsoft.com/~cmbishop

First Generation

"Artificial Intelligence" (GOFAI)

Within a generation ... the problem of creating 'artificial intelligence' will largely be solved

Marvin Minsky (1967)

Expert Systems (1980s)

knowledge-based AI

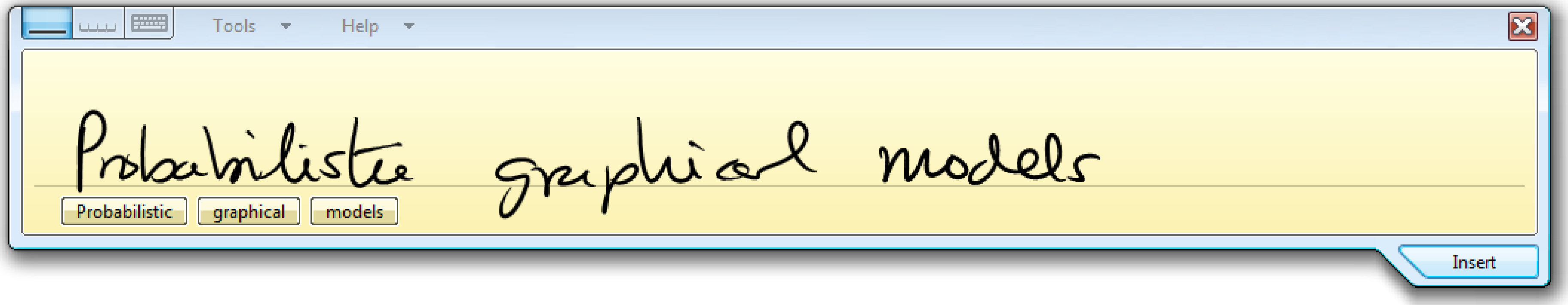
rules elicited from humans

Combinatorial explosion

General theme: hand-crafted rules

Second Generation

Neural networks, support vector machines



Difficult to incorporate complex domain knowledge

General theme: black-box statistical models

Third Generation

General theme: deep integration of domain knowledge and statistical learning

Bayesian framework

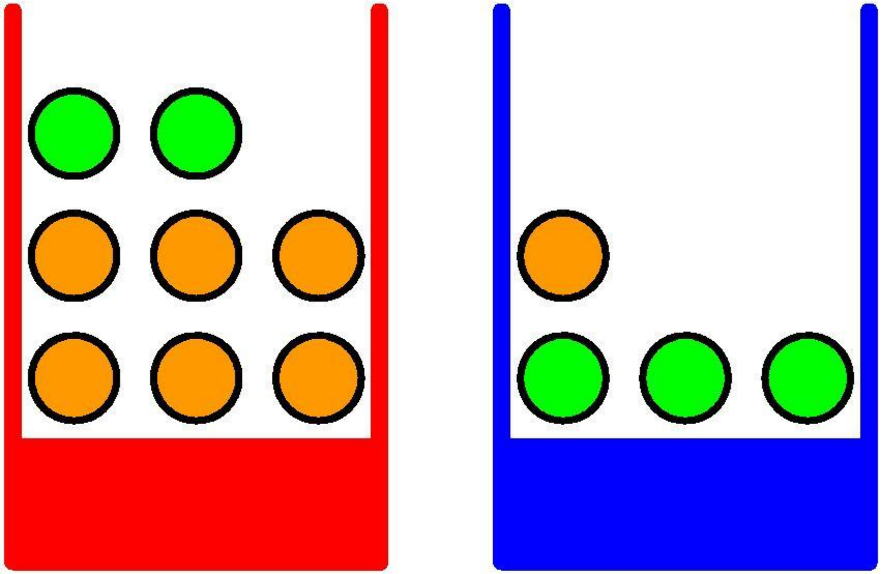
Probabilistic graphical models

Fast inference using local message-passing

Origins: Bayesian networks, decision theory, HMMs, Kalman filters, MRFs, mean field theory, ...

Probability Theory

Apples and Oranges



Fruit is orange, what is probability that box was blue?

The Rules of Probability

Sum rule

$$p(X) = \sum_{Y} p(X, Y)$$

Product rule

$$p(X, Y) = p(Y|X)p(X)$$

$$p(X) \geqslant 0 \qquad \sum_{X} p(X) = 1$$

Bayes' Theorem

$$p(Y|X) = \frac{p(X,Y)}{p(X)}$$

$$= \frac{p(X|Y)p(Y)}{p(X)}$$

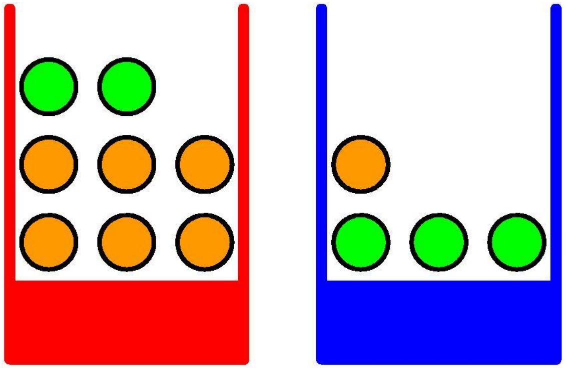
$$p(X) = \sum_{Y} p(X,Y)$$

$$= \sum_{Y} p(X|Y)p(Y)$$

Oranges and Apples

Suppose $p(B=r)=2/5$

Suppose we select an orange



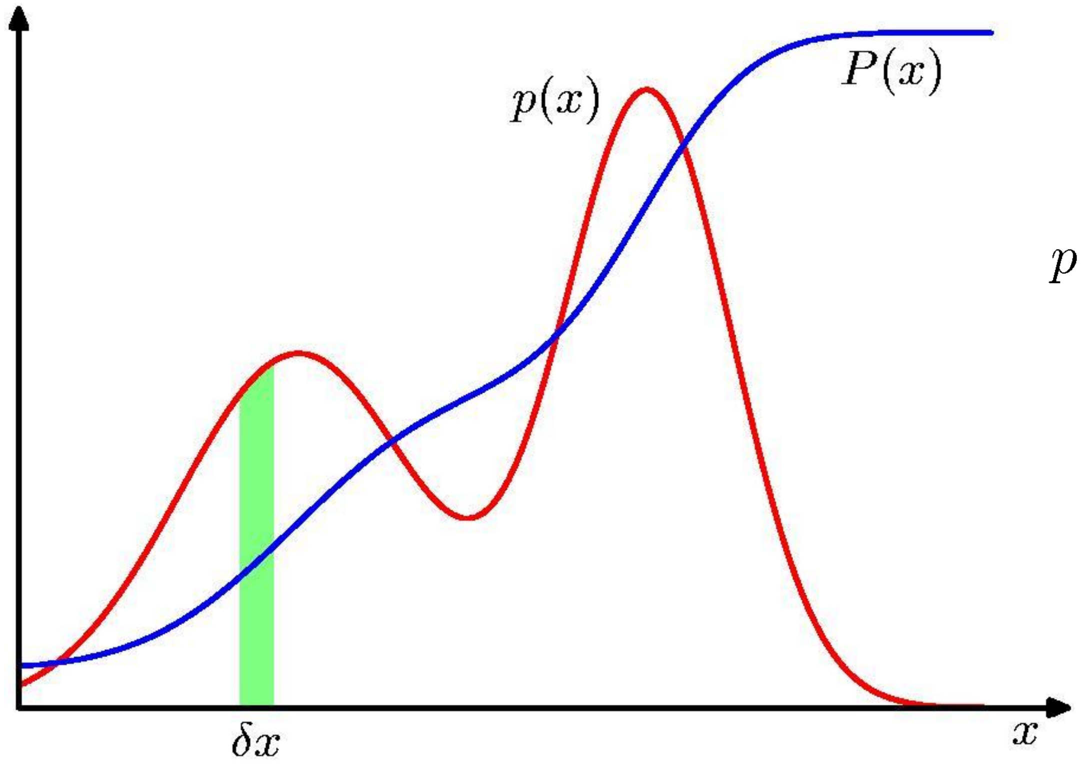
Then

$$\begin{aligned} p(F=o) &= p(F=o|B=r)p(B=r)+p(F=o|B=b)p(B=b) \\ &= 9/20 \end{aligned}$$

and hence

$$\begin{aligned} p(B=r|F=o) &= \frac{p(F=o|B=r)p(B=r)}{p(F=o)} \\ &= 2/3 \end{aligned}$$

Probability Densities



$$p(x \in (a,b)) = \int_a^b p(x)\,\mathrm{d}x$$

$$P(z) = \int_{-\infty}^{z} p(x)\,\mathrm{d}x$$

$$p(x) \geqslant 0 \qquad \int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1$$

Bayesian Inference

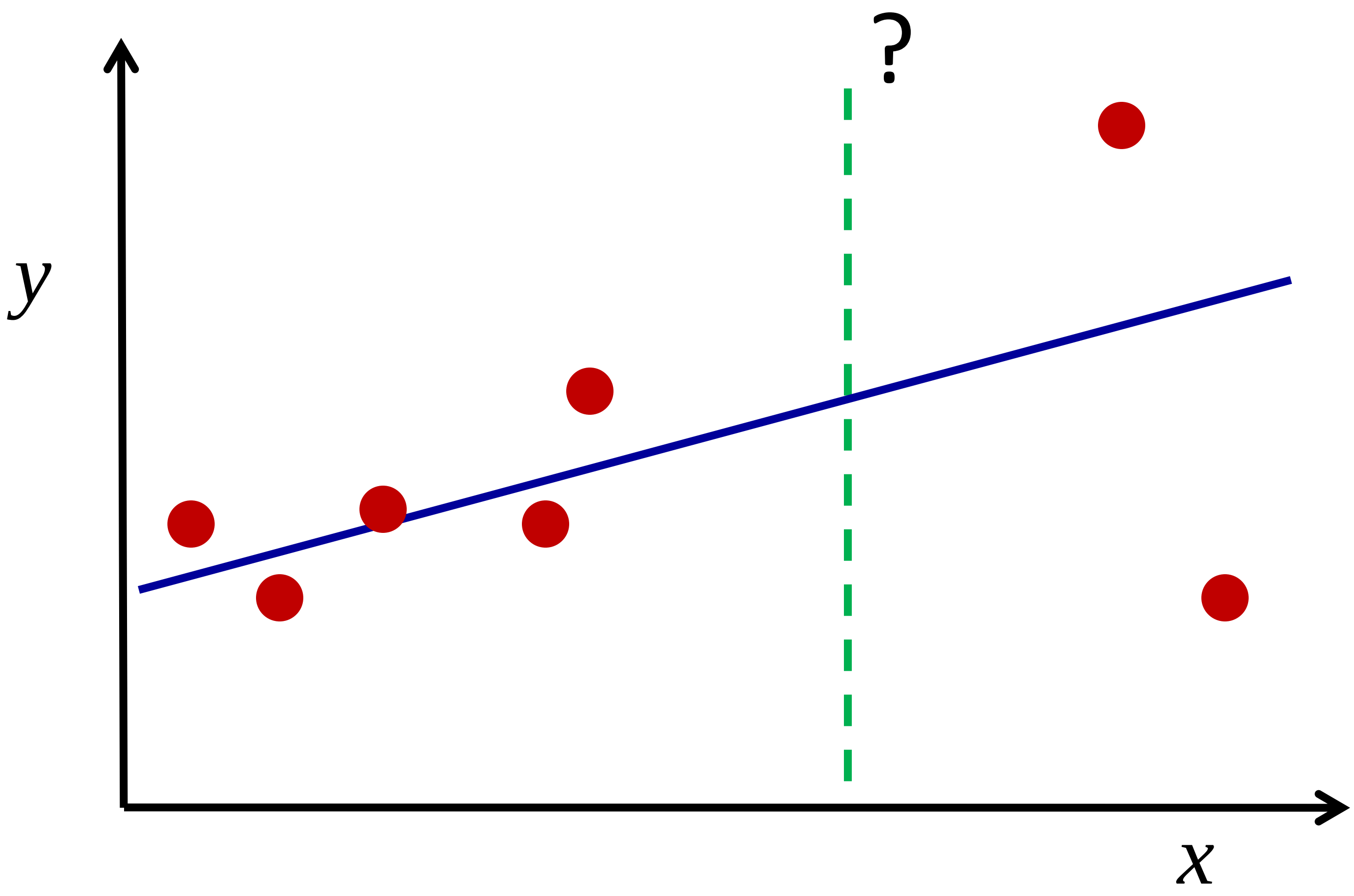
Consistent use of probability to quantify uncertainty

$$\underbrace{p(\boldsymbol{\theta}|\widehat{\mathbf{x}}, \mathbf{X})}_{\text{posterior}} \propto \underbrace{p(\widehat{\mathbf{x}}|\boldsymbol{\theta})}_{\text{likelihood function}} \, \underbrace{p(\boldsymbol{\theta}|\mathbf{X})}_{\text{prior}}$$

Predictions involve marginalisation, e.g.

$$p(\mathbf{y}|\mathbf{X}) = \int p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{X}) \, \mathrm{d}\boldsymbol{\theta}$$

Why is prior knowledge important?



Probabilistic Graphical Models

Combine probability theory with graphs

- ✓ new insights into existing models
- ✓ framework for designing new models
- ✓ Graph-based algorithms for calculation and computation (c.f. Feynman diagrams in physics)
- ✓ efficient software implementation

Directed graphs to specify the model

Factor graphs for inference and learning

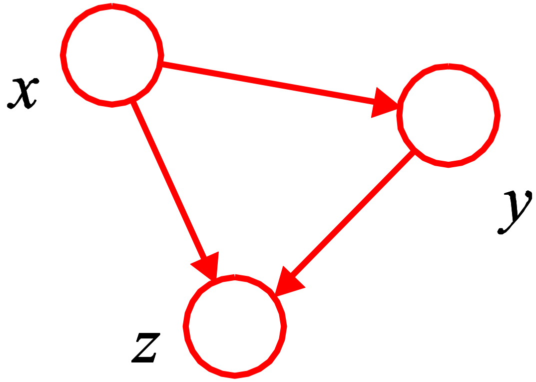
Decomposition

Consider an arbitrary joint distribution

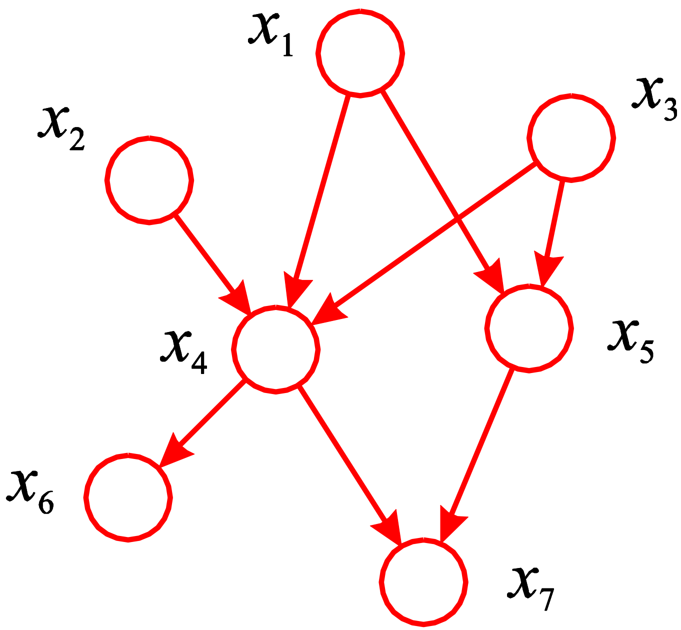
$$p(x, y, z)$$

By successive application of the product rule:

$$\begin{aligned} p(x, y, z) &= p(x)p(y, z|x) \\ &= p(x)p(y|x)p(z|x, y) \end{aligned}$$



Directed Graphs



$$p(x_1, \ldots, x_7) = p(x_1)p(x_2)p(x_3)$$
$$p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_3)$$
$$p(x_6|x_4)p(x_7|x_4, x_5)$$

Arrows indicate causal relationships

MAAS

Manchester Asthma and Allergies Study

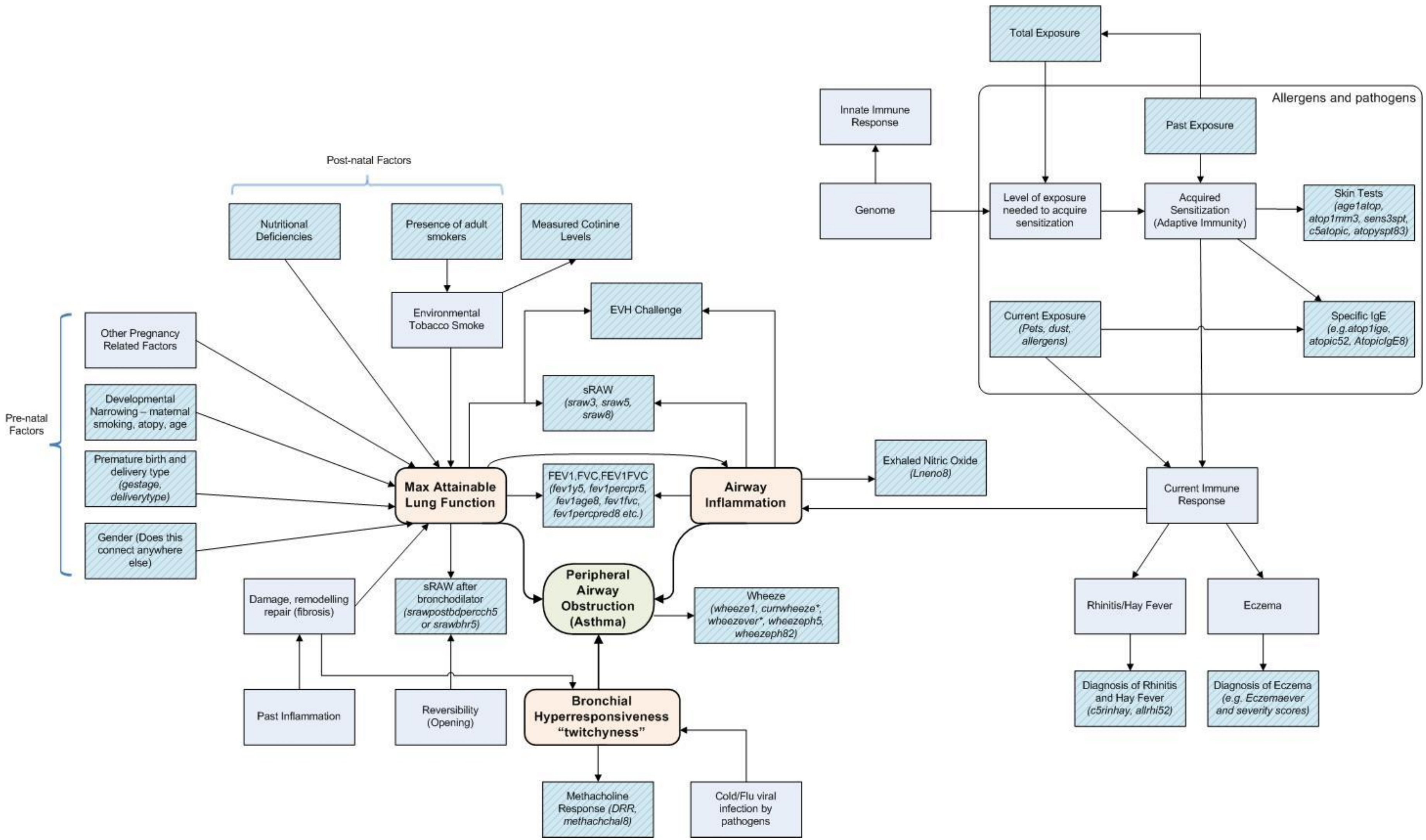
Goal: discover environmental and genetic causes of asthma

1,186 children monitored since birth

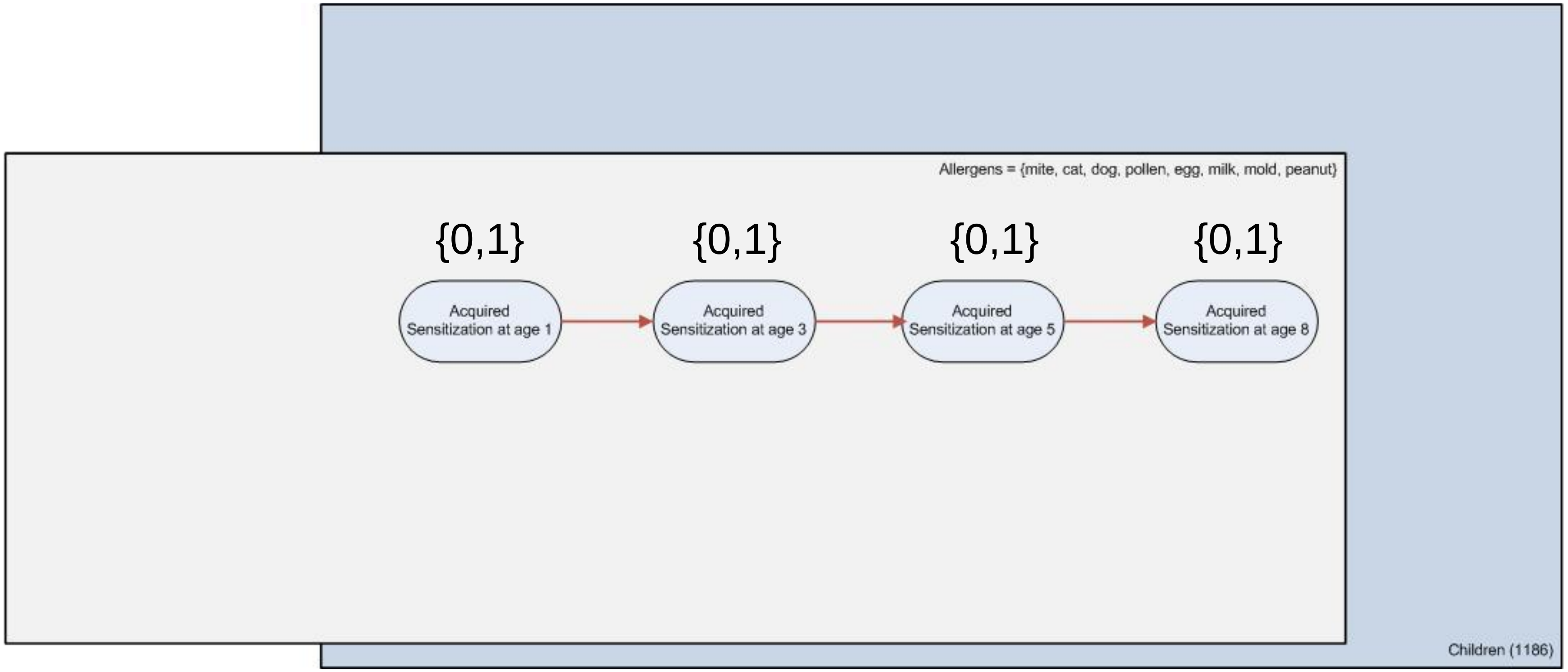
640k SNPs per child

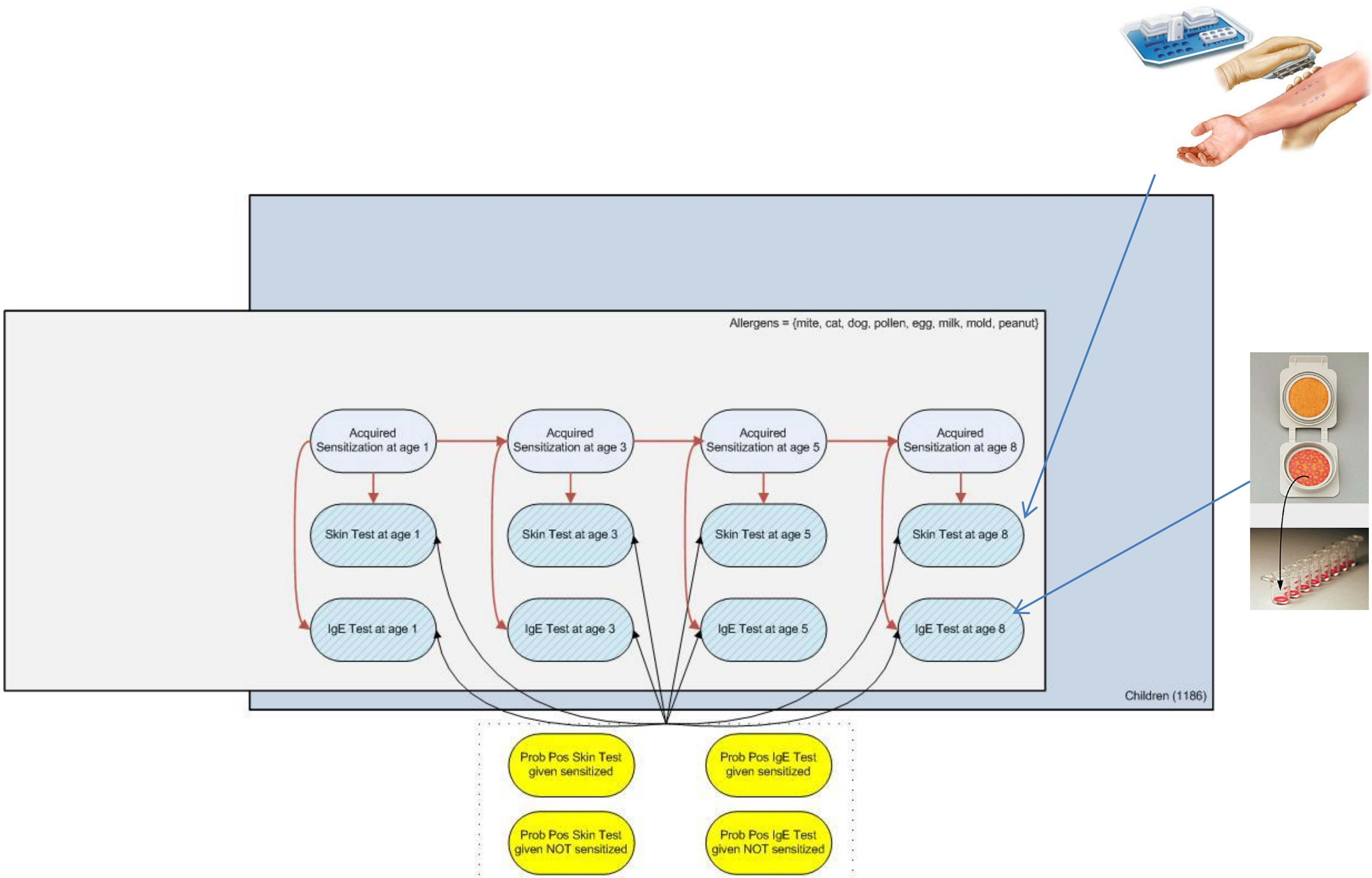
Many environment and physiological measurements:

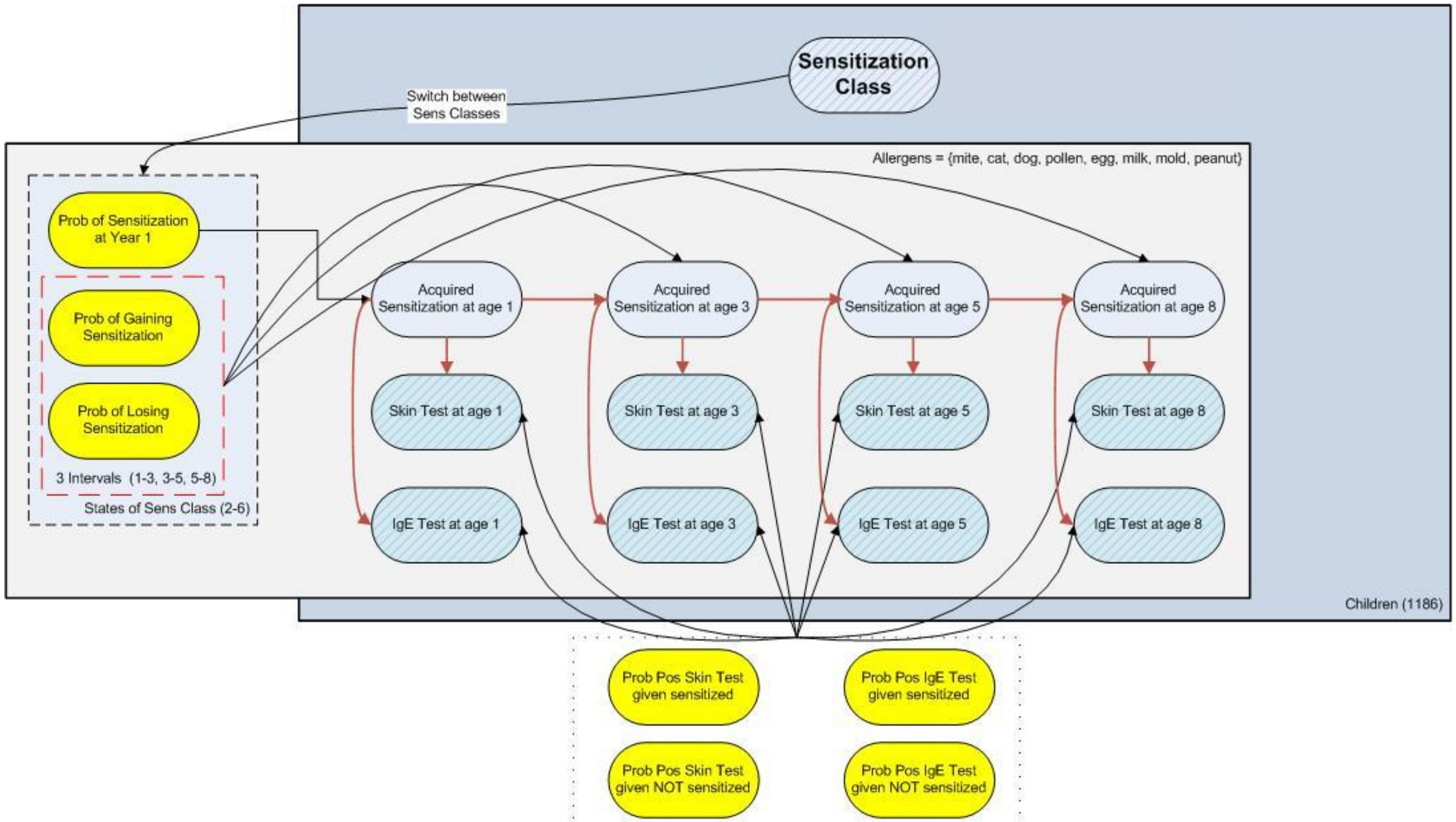
- skin and IgE blood tests at age 1, 3, 5, and 8
- wheezing, methacholine response,
- pets, parental smoking, day-care, breast feeding, ...



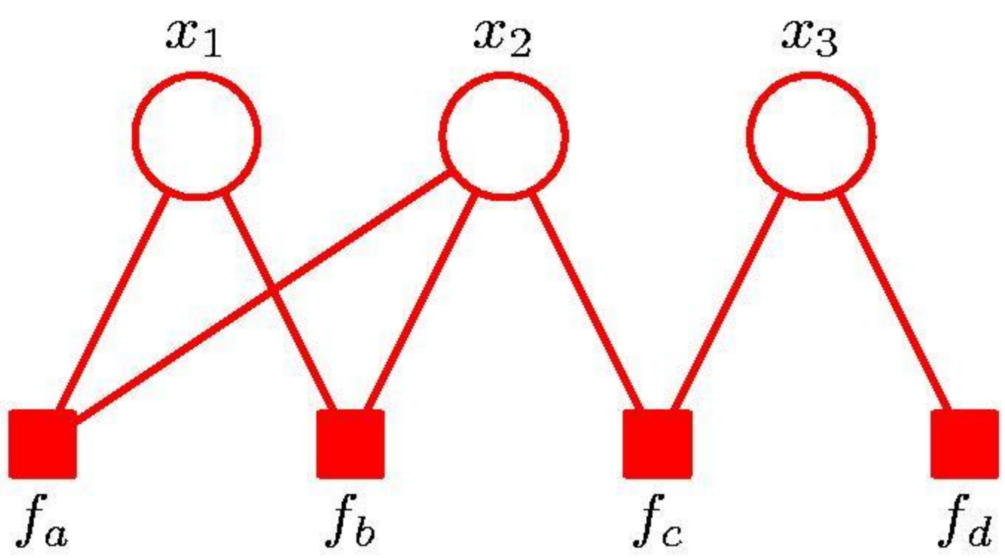
Total Exposure
Allergens and pathogens
Innate Immune Response
Past Exposure
Post-natal Factors
Genome
Level of exposure needed to acquire sensitization
Acquired Sensitization (Adaptive Immunity)
Skin Tests (age1atop, atop1mm3, sens3spt, c5atopic, atopyspt83)
Nutritional Deficiencies
Presence of adult smokers
Measured Cotinine Levels
Environmental Tobacco Smoke
EVH Challenge
Current Exposure (Pets, dust, allergens)
Specific IgE (e.g.atop1ige, atopic52, AtopicIgE8)
Other Pregnancy Related Factors
Pre-natal Factors
Developmental Narrowing – maternal smoking, atopy, age
sRAW (sraw3, sraw5, sraw8)
Premature birth and delivery type (gestage, deliverytype)
Max Attainable Lung Function
FEV1,FVC,FEV1FVC (fev1y5, fev1percpr5, fev1age8, fev1fvc, fev1percpred8 etc.)
Airway Inflammation
Exhaled Nitric Oxide (Lneno8)
Current Immune Response
Gender (Does this connect anywhere else)
Damage, remodelling repair (fibrosis)
sRAW after bronchodilator (srawpostbdpercch5 or srawbhr5)
Peripheral Airway Obstruction (Asthma)
Wheeze (wheeze1, currwheeze*, wheezever*, wheezeph5, wheezeph82)
Rhinitis/Hay Fever
Eczema
Past Inflammation
Reversibility (Opening)
Bronchial Hyperresponsiveness “twitchyness”
Diagnosis of Rhinitis and Hay Fever (c5rinhay, allrhi52)
Diagnosis of Eczema (e.g. Eczemaever and severity scores)
Methacholine Response (DRR, methachchal8)
Cold/Flu viral infection by pathogens


Allergens = {mite, cat, dog, pollen, egg, milk, mold, peanut}
{0,1}
{0,1}
{0,1}
{0,1}
Acquired Sensitization at age 1
Acquired Sensitization at age 3
Acquired Sensitization at age 5
Acquired Sensitization at age 8
Children (1186)


Allergens = {mite, cat, dog, pollen, egg, milk, mold, peanut}
Acquired Sensitization at age 1
Acquired Sensitization at age 3
Acquired Sensitization at age 5
Acquired Sensitization at age 8
Skin Test at age 1
Skin Test at age 3
Skin Test at age 5
Skin Test at age 8
IgE Test at age 1
IgE Test at age 3
IgE Test at age 5
IgE Test at age 8
Children (1186)
Prob Pos Skin Test given sensitized
Prob Pos IgE Test given sensitized
Prob Pos Skin Test given NOT sensitized
Prob Pos IgE Test given NOT sensitized


Sensitization Class
Switch between Sens Classes
Allergens = {mite, cat, dog, pollen, egg, milk, mold, peanut}
Prob of Sensitization at Year 1
Prob of Gaining Sensitization
Prob of Losing Sensitization
3 Intervals (1-3, 3-5, 5-8)
States of Sens Class (2-6)
Acquired Sensitization at age 1
Acquired Sensitization at age 3
Acquired Sensitization at age 5
Acquired Sensitization at age 8
Skin Test at age 1
Skin Test at age 3
Skin Test at age 5
Skin Test at age 8
IgE Test at age 1
IgE Test at age 3
IgE Test at age 5
IgE Test at age 8
Children (1186)
Prob Pos Skin Test given sensitized
Prob Pos IgE Test given sensitized
Prob Pos Skin Test given NOT sensitized
Prob Pos IgE Test given NOT sensitized

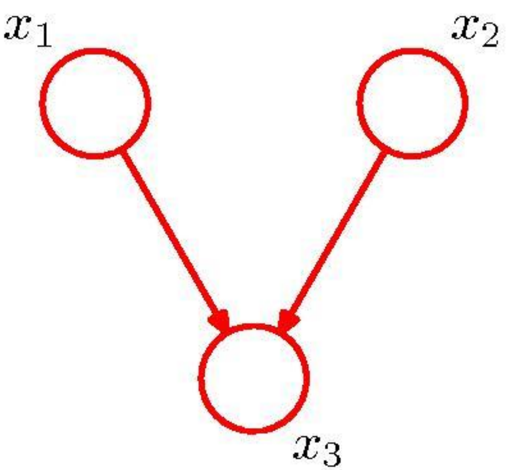
Factor Graphs

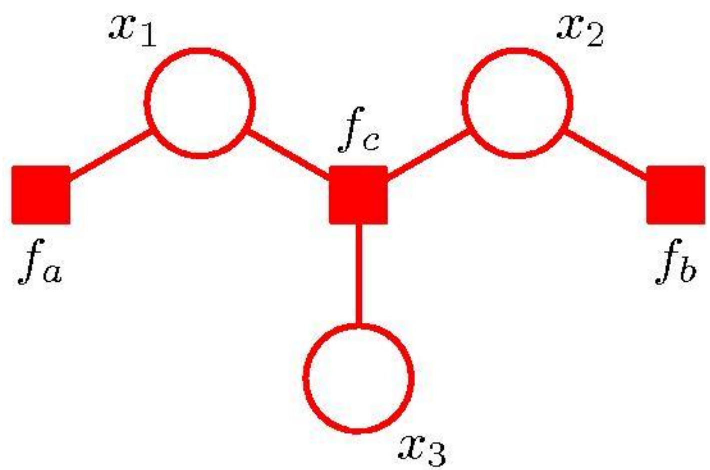


$$p(x_1, x_2, x_3) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3)$$

From Directed Graph to Factor Graph

$$p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3|x_1, x_2)$$



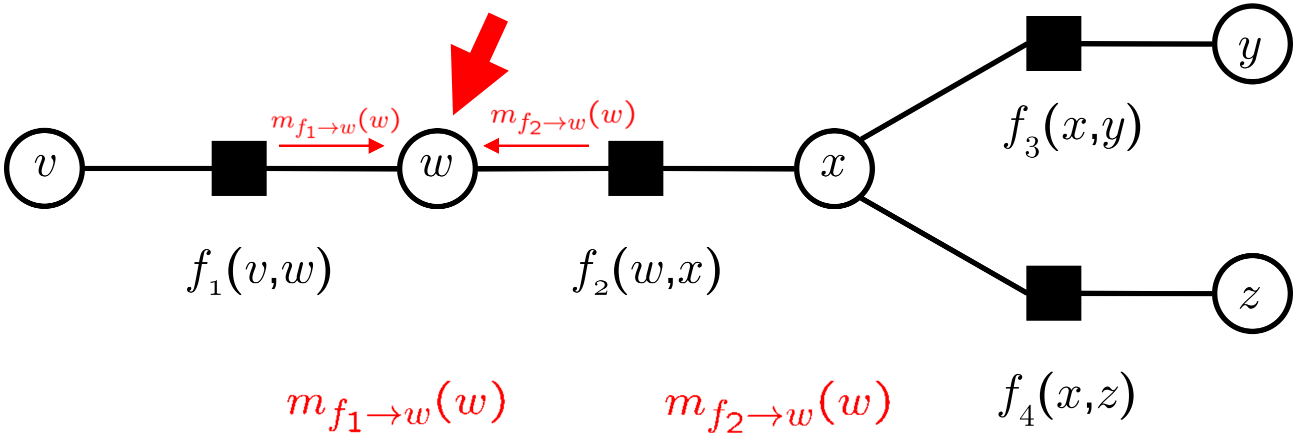


$$\begin{aligned} f_a(x_1) &= p(x_1) \\ f_b(x_2) &= p(x_2) \\ f_c(x_1, x_2, x_3) &= p(x_3|x_1, x_2) \end{aligned}$$

Inference on Graphs

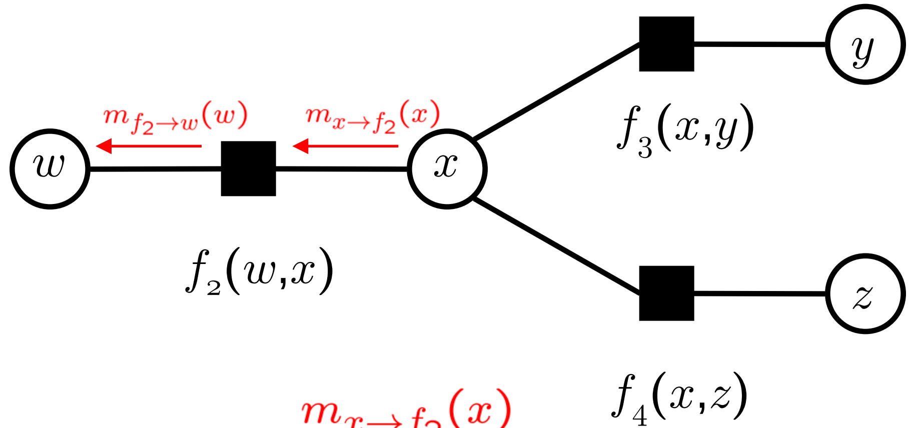
$$ab + ac = a(b + c)$$

Factor Trees: Separation



$$p(w) = \overbrace{\boxed{\left[\sum_v f_1(v,w)\right]}}^{m_{f_1 \to w}(w)} \overbrace{\boxed{\left[\sum_x \sum_y \sum_z f_2(w,x) f_3(x,y) f_4(x,z)\right]}}^{m_{f_2 \to w}(w)}$$

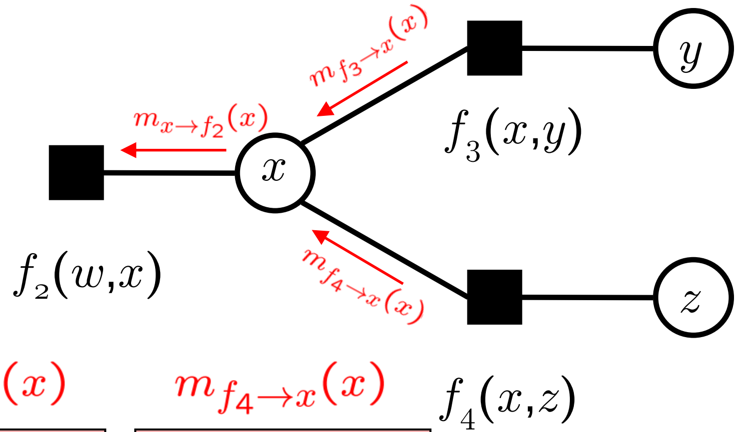
Messages: From Factors To Variables



$$m_{f_2\to w}(w) = \sum_x \sum_y \sum_z f_2(w,x) f_3(x,y) f_4(x,z)$$

$$m_{f_2\to w}(w) = \sum_x f_2(w,x) \overbrace{\boxed{\left[\sum_y \sum_z f_3(x,y) f_4(x,z)\right]}}^{m_{x\to f_2}(x)}$$

Messages: From Variables To Factors



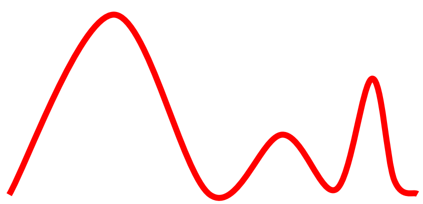
$$m_{x\to f_2}(x) = \overbrace{\left[\sum_y f_3(x,y)\right]}^{m_{f_3\to x}(x)} \cdot \overbrace{\left[\sum_z f_4(x,z)\right]}^{m_{f_4\to x}(x)}$$

What if the graph is not a tree?

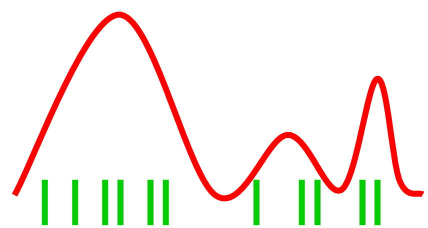
Keep iterating the messages:

loopy belief propagation

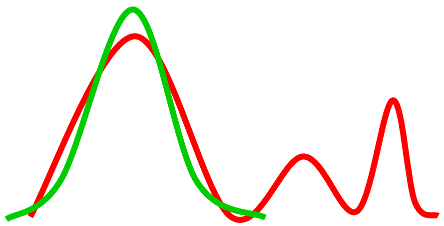
What if marginalisations are not tractable?



True distribution

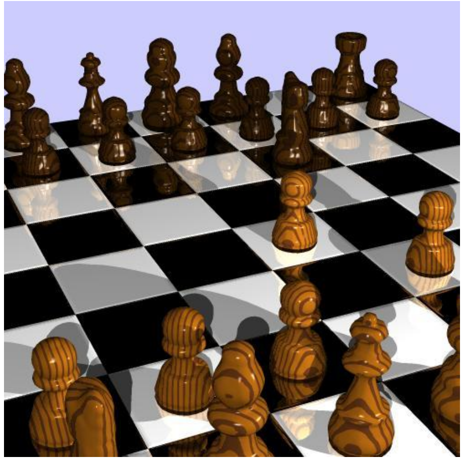


Monte Carlo



VMP / Loopy BP / EP

Illustration: Bayesian Ranking



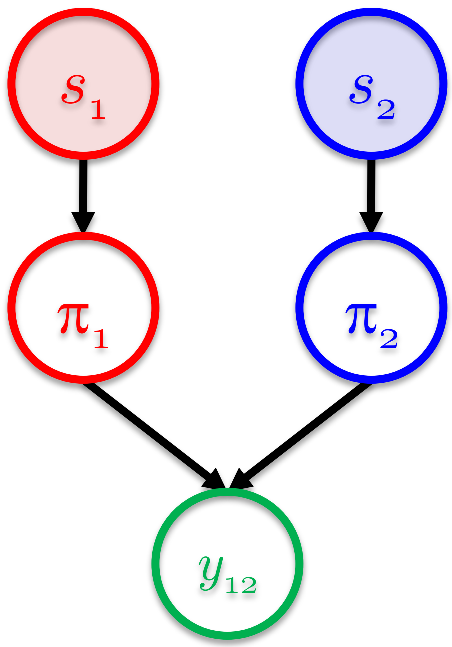
Ralf Herbrich
Tom Minka
Thore Graepel

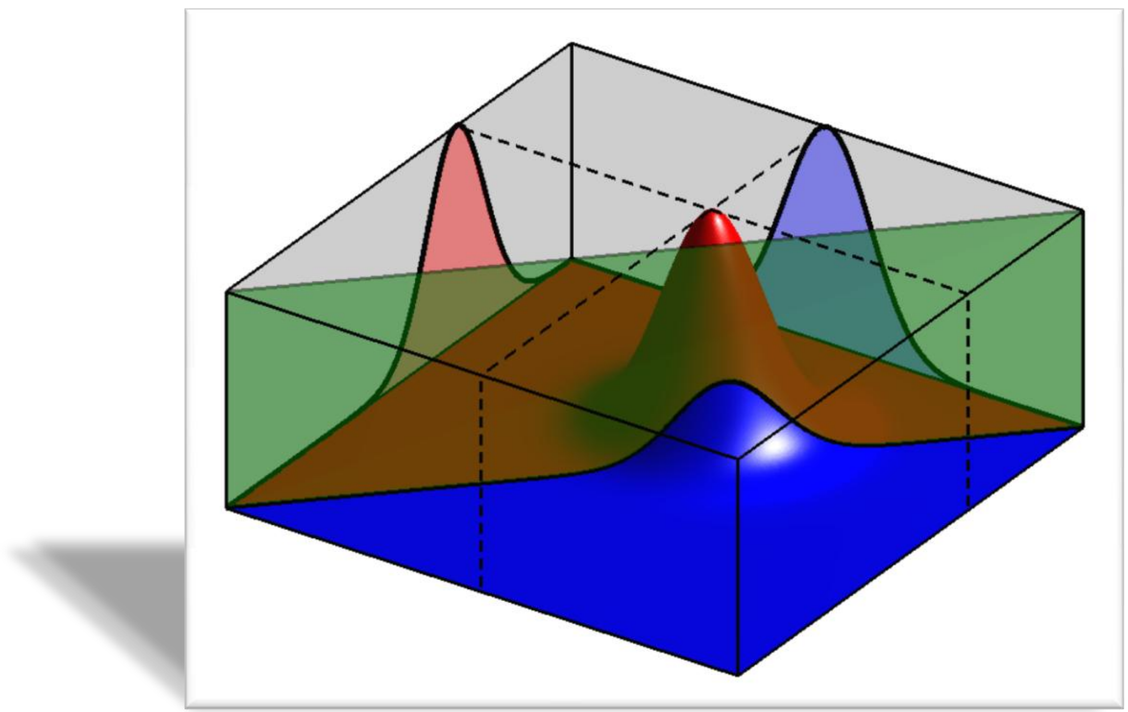
Goal: global ranking from noisy partial rankings

Conventional approach: Elo (used in chess)

- maintains a single strength value for each player
- cannot handle team games, or > 2 players

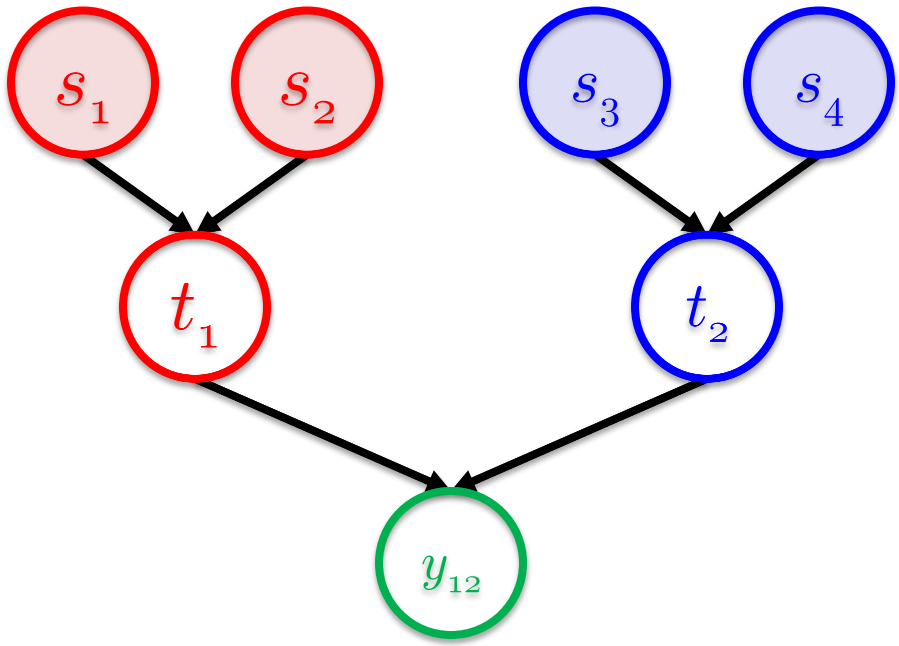
Two Player Match Outcome Model





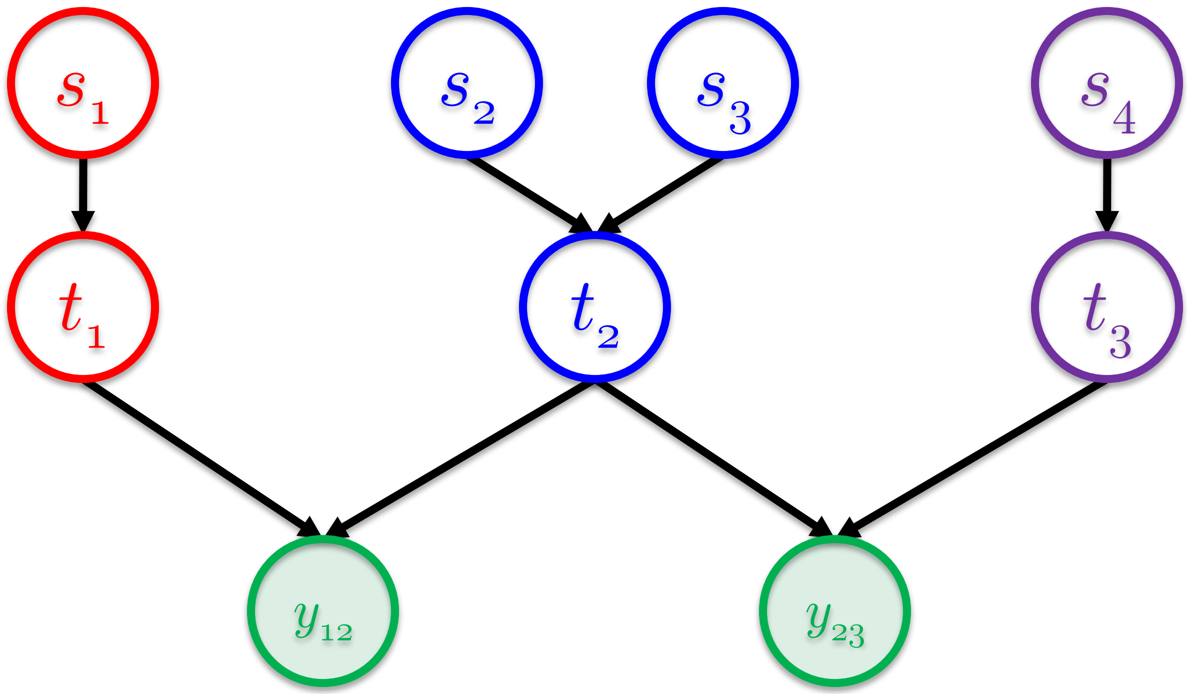
$$p\left(y_{12}=(1,2)|\pi_1,\pi_2\right)=\mathrm{I}(\pi_1>\pi_2)$$

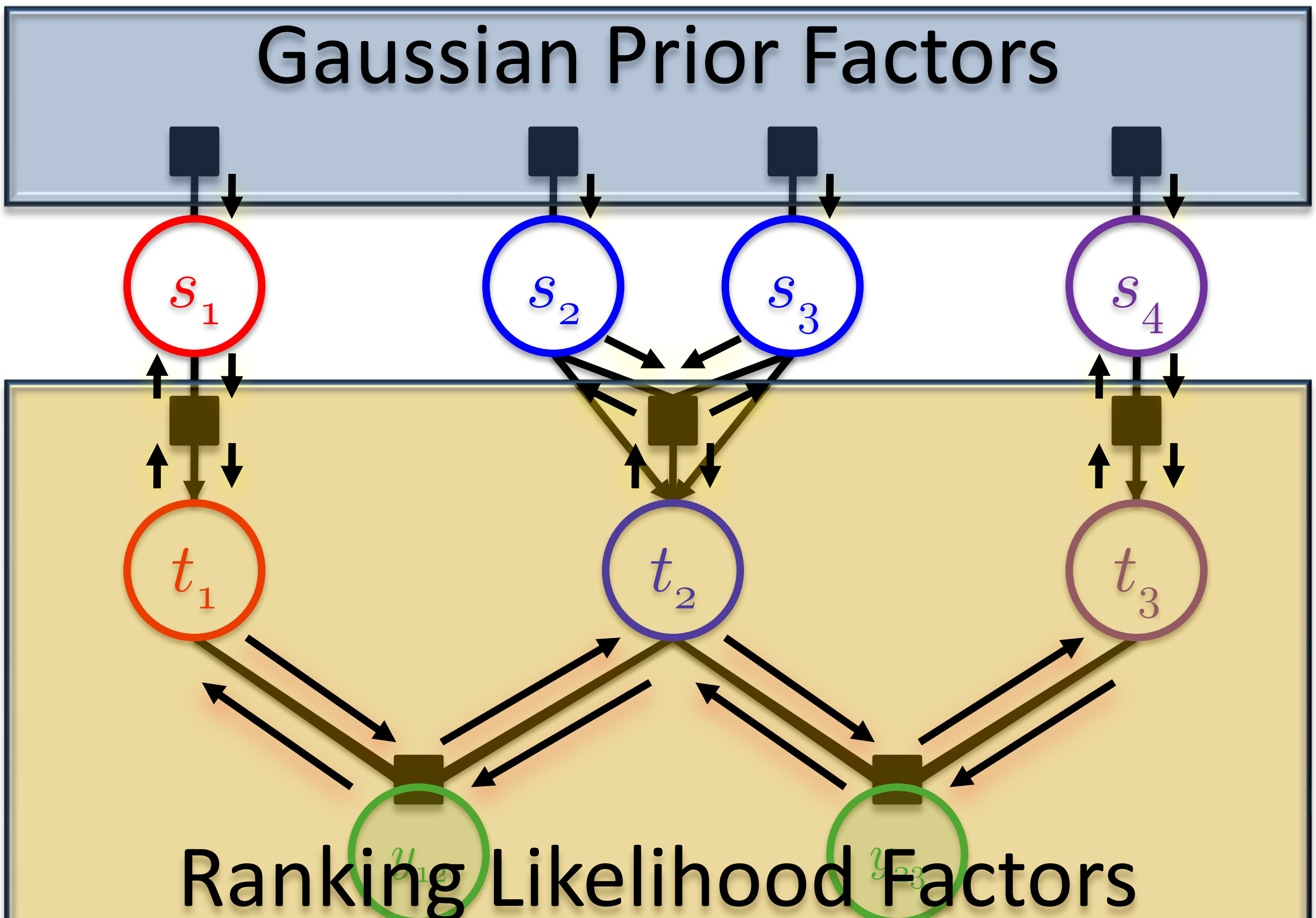
Two Team Match Outcome Model



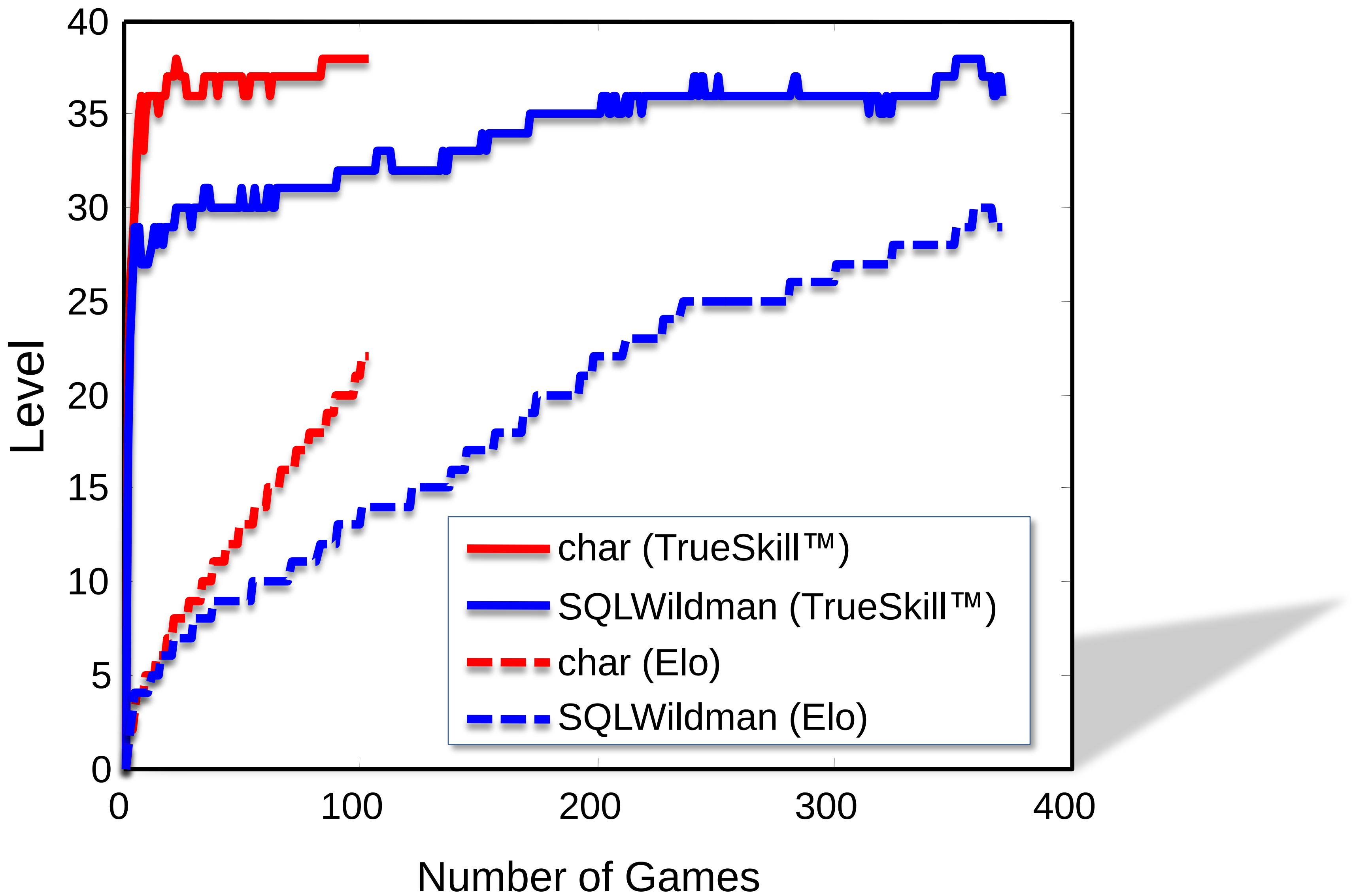
$$p(t_1|s_1, s_2) = \mathcal{N}(t_1|s_1 + s_2, 2\beta^2)$$

Multiple Team Match Outcome Model



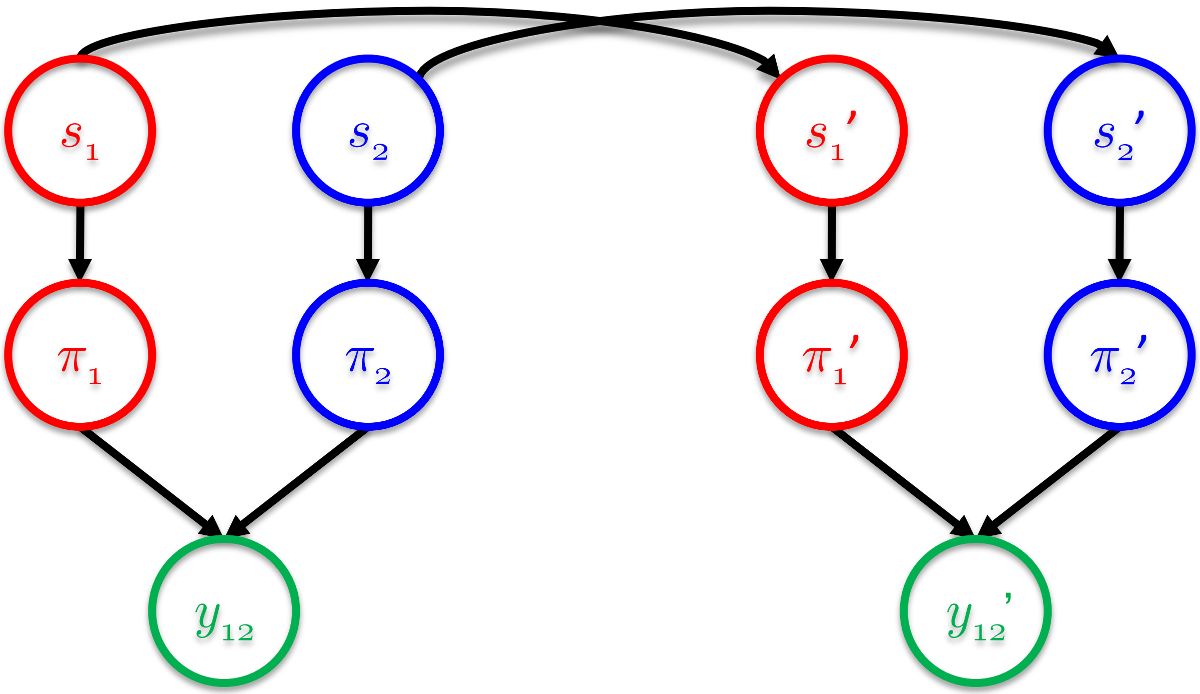


Gaussian Prior Factors
s_1
s_2
s_3
s_4
t_1
t_2
t_3
Ranking Likelihood Factors



Level
Number of Games
0
5
10
15
20
25
30
35
40
0
100
200
300
400
char (TrueSkill™)
SQLWildman (TrueSkill™)
char (Elo)
SQLWildman (Elo)

Skill Dynamics



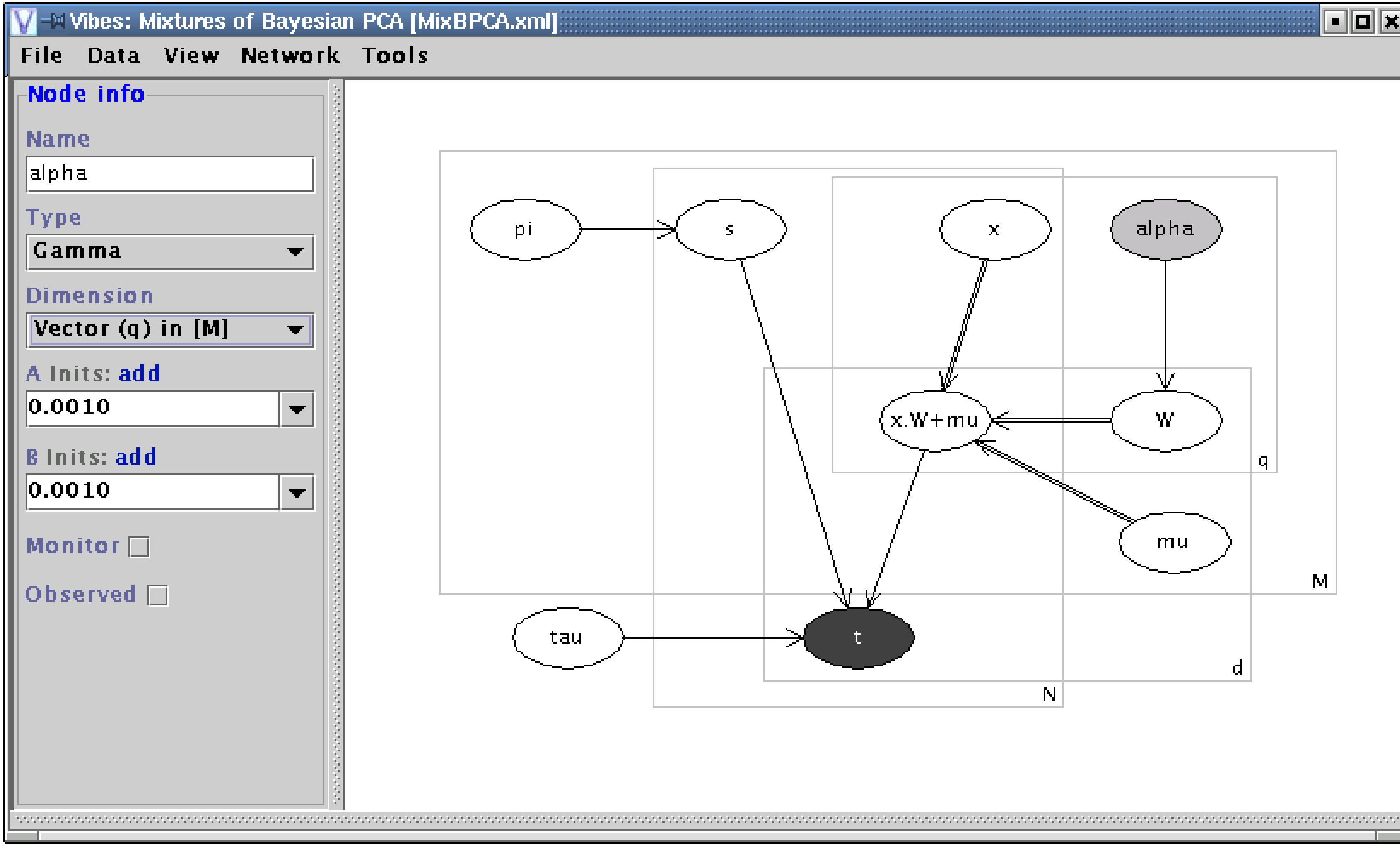
$$p(s_i'|s_i) = \mathcal{N}(s_i'|s_i, \tau^2)$$

TrueSkill™

Xbox 360 Live: launched September 2005

- *TrueSkill*™ for ranking and to match players
- 10M active users, 2.5M matches per day

"Planet-scale" application of Bayesian methods



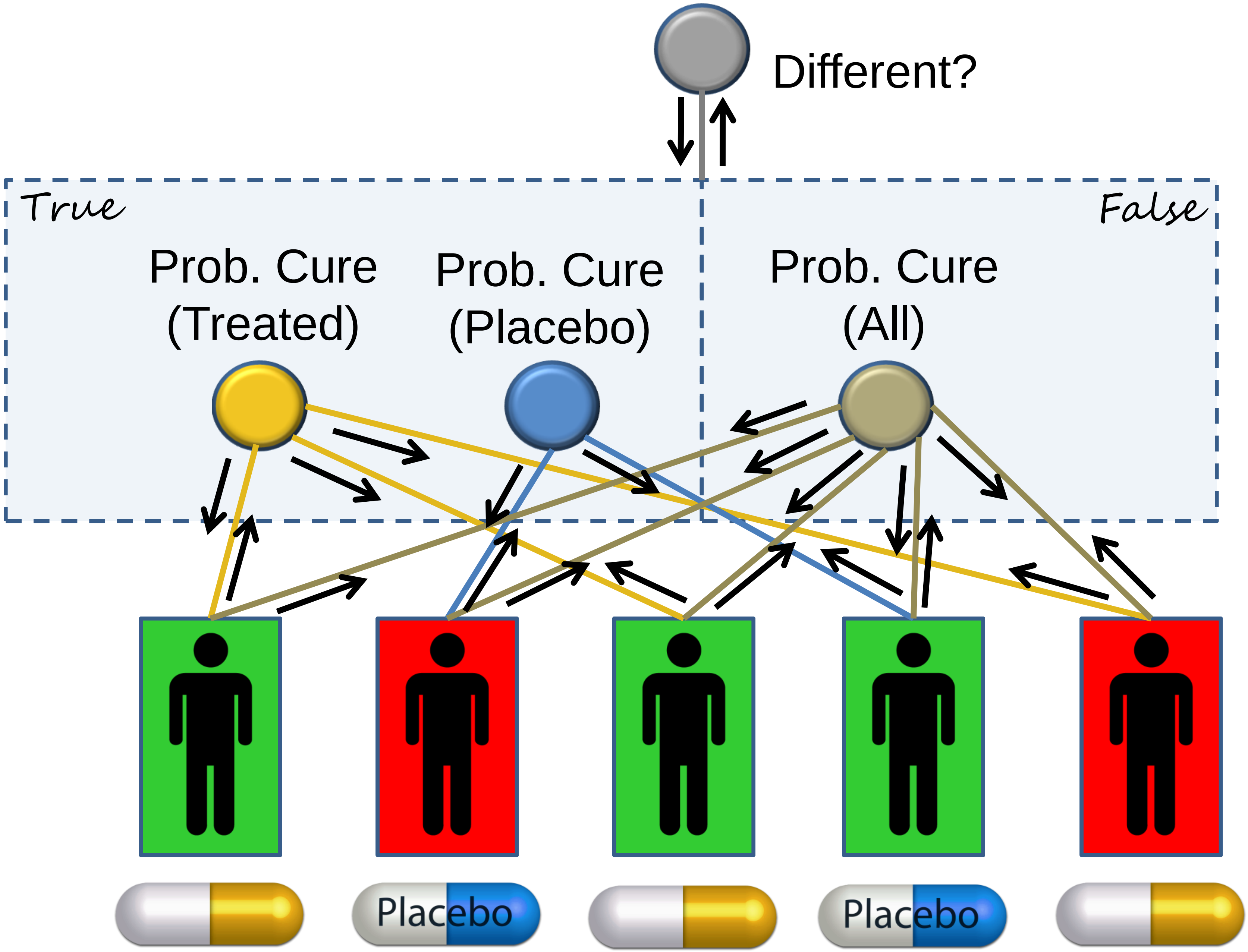
Vibes: Mixtures of Bayesian PCA [MixBPCA.xml]
File
Data
View
Network
Tools
Node info
Name
alpha
Type
Gamma
Dimension
Vector (q) in [M]
A Inits: add
0.0010
B Inits: add
0.0010
Monitor
Observed
pi
s
x
alpha
x.W+mu
W
q
mu
M
tau
t
d
N

John Winn

research.microsoft.com/infernet

Tom Minka
John Winn
John Guiver
Anitha Kannan

Infer.Net demonstration


Different?
True
False
Prob. Cure (Treated)
Prob. Cure (Placebo)
Prob. Cure (All)
Placebo
Placebo