Probabilistic Modeling for Multi Systems Microarray Experiments

Peter Sykacek

Richard Clarkson

Department of Pathology

University of Cambridge

Statistical Test Based Approach

Inflammatory Mediator

Pos.	Affy Id.	p-value
	. . .	
15	583.s.at	1.76e-30
16	1006.at	1.51e-29
17	37388.at	3.93e-28
	. . .	
25	1052.s.at	6.61e-20
26	875.g.at	5.65e-19
27	38430.at	2.32e-18

Endothelial Cell Death

Pos.	Affy Id.	p-value
1	875.g.at	3.58e-39
2	867.s.at	2.72e-38
	. . .	
10	40490.at	8.74e-19
11	1006.at	1.10e-18
12	874.at	2.48e-18

Statistical Test Based Approach

Inflammatory Mediator

Pos.	Affy Id.	p-value
	. . .	
15	583.s.at	1.76e-30
16	1006.at	1.51e-29
17	37388.at	3.93e-28
	. . .	
25	1052.s.at	6.61e-20
26	875.g.at	5.65e-19
27	38430.at	2.32e-18

Endothelial Cell Death

Pos.	Affy Id.	p-value
1	875.g.at	3.58e-39
2	867.s.at	2.72e-38
	. . .	
10	40490.at	8.74e-19
11	1006.at	1.10e-18
12	874.at	2.48e-18

Pattern Matching − >
No principled Ranking!

Combined Result

875.g.at,   1006.at,

Probabilistic Approach

Thomas Bayes (1701 - 1763)
Learning from data based on a
decision theoretic framework

Probabilistic Approach

Thomas Bayes (1701 - 1763)
Learning from data based on a
decision theoretic framework

$$p(G|\mathcal{D}) = \frac{p(\mathcal{D}|G)p(G)}{p(\mathcal{D})}$$

First consequence: we must revise beliefs according to Bayes theorem

Probabilistic Approach

Thomas Bayes (1701 - 1763) Learning from data based on a decision theoretic framework

$$p(G|\mathcal{D}) = \frac{p(\mathcal{D}|G)p(G)}{p(\mathcal{D})}$$

First consequence: we must revise beliefs according to Bayes theorem

$\alpha_{opt} = \mathrm{argmax}_\alpha <u(\alpha)>$, where

$<u(\alpha)> = \int_G u(\alpha, G)p(G|\mathcal{D})dG.$

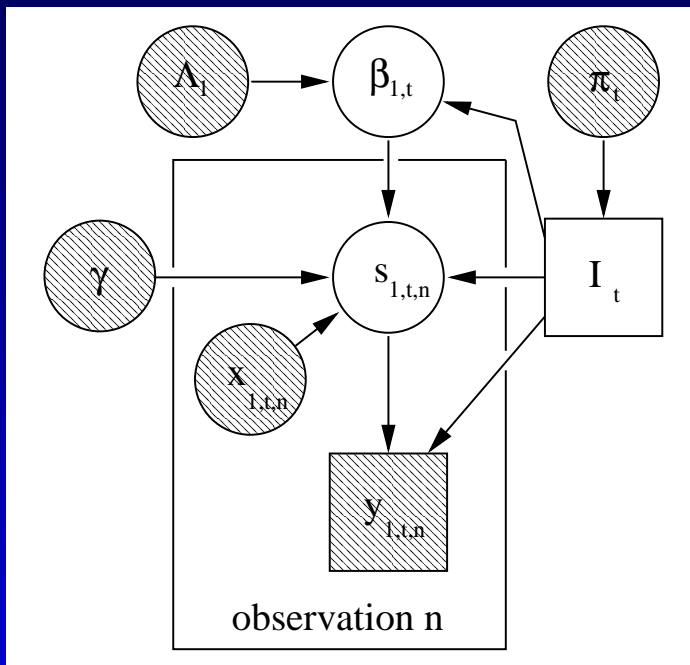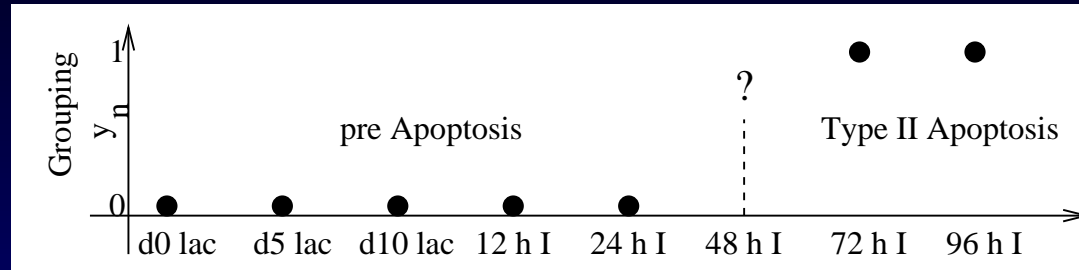Second consequence: Decisions by maximising expected utilities

Probabilistic Gene Ranking
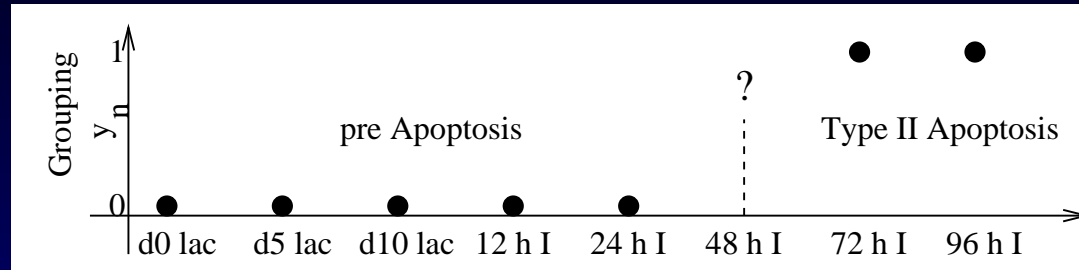
Model Type II Apop-
tosis in the Mouse
Mammary Gland

Probabilistic Gene Ranking

Model Type II Apoptosis in the Mouse Mammary Gland
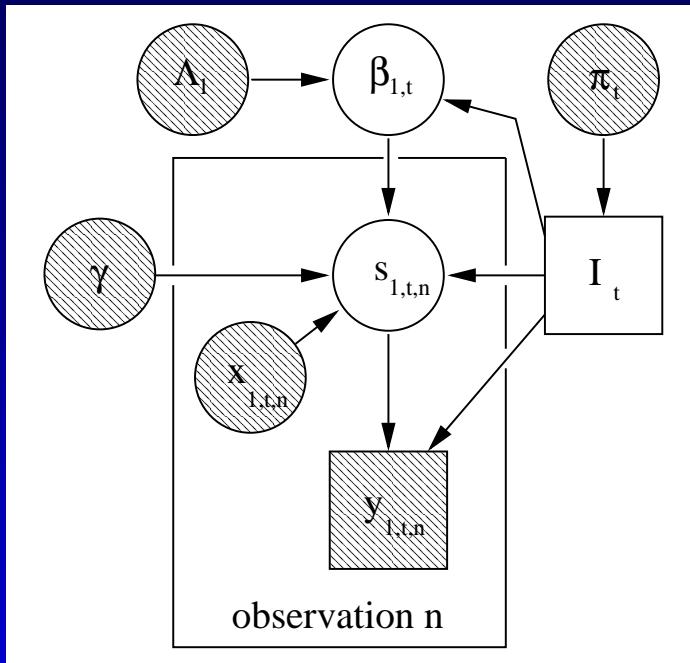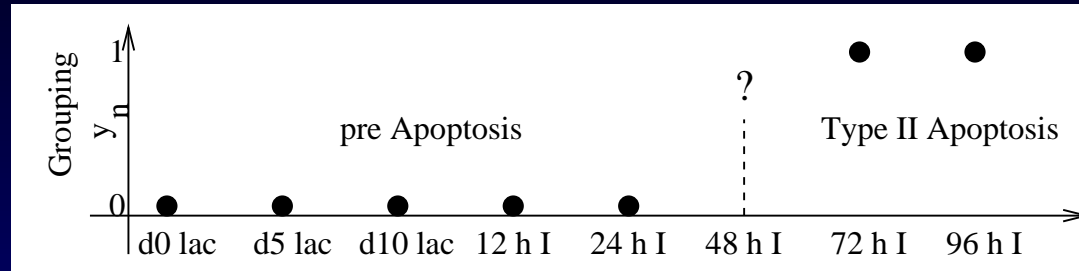

Latent variable probit GLM.

$$\text{if } G = \begin{cases} 1 : s_{1,n} \sim 1 + x_n \\ 0 : s_{1,n} \sim 1 \end{cases}$$

$s_{1,n}$ is a one dimensional Gaussian random variable with mean $\boldsymbol{\beta}_1^T \boldsymbol{x}_n$ and precision 1.

Probabilistic Gene Ranking

Model Type II Apoptosis in the Mouse Mammary Gland
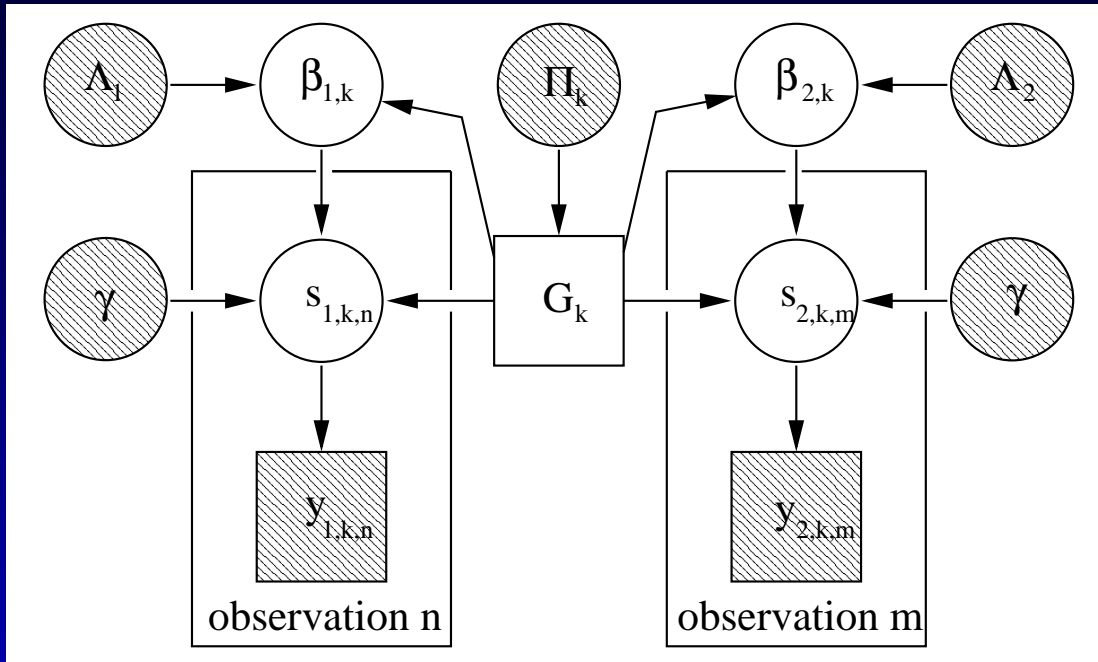

Latent variable probit GLM.

$$\text{if } G = \begin{cases} 1 : s_{1,n} \sim 1 + x_n \\ 0 : s_{1,n} \sim 1 \end{cases}$$

$s_{1,n}$ is a one dimensional Gaussian random variable with mean $\boldsymbol{\beta}_1^T \boldsymbol{x}_n$ and precision 1.

As a rank measure, we infer $P(G|\mathcal{D}_1)$, the posterior probability of genes being differentially expressed.

Combined System Analysis
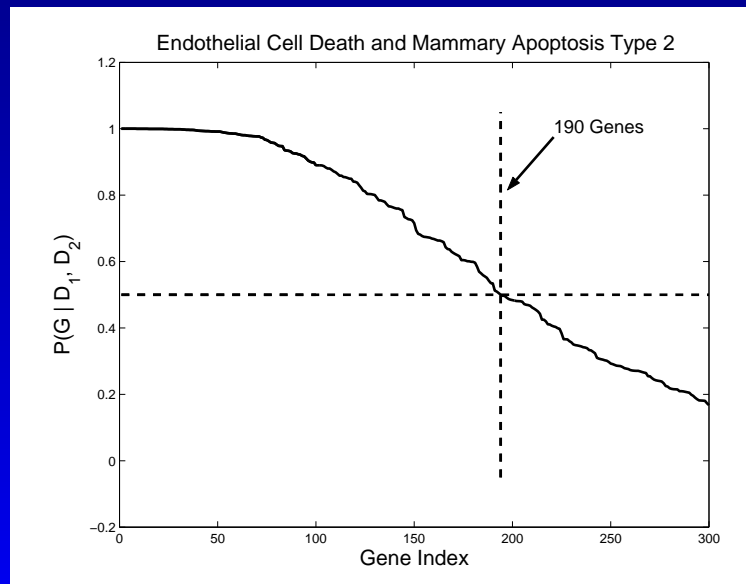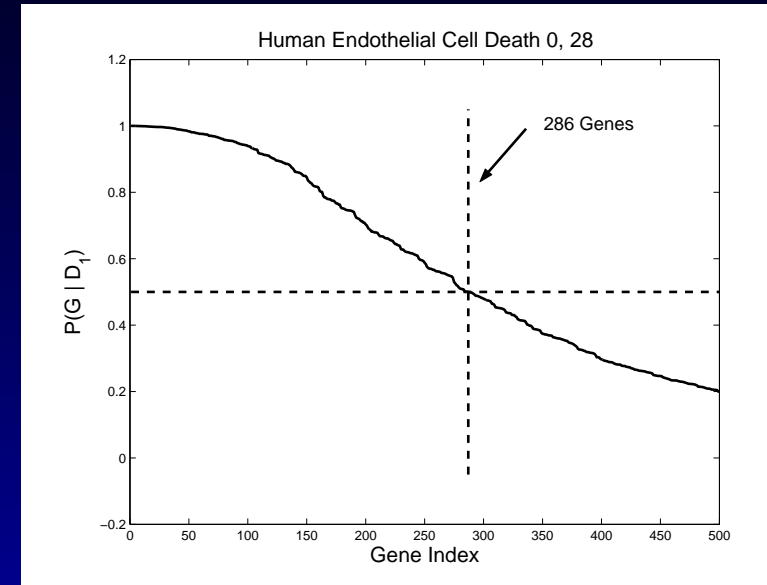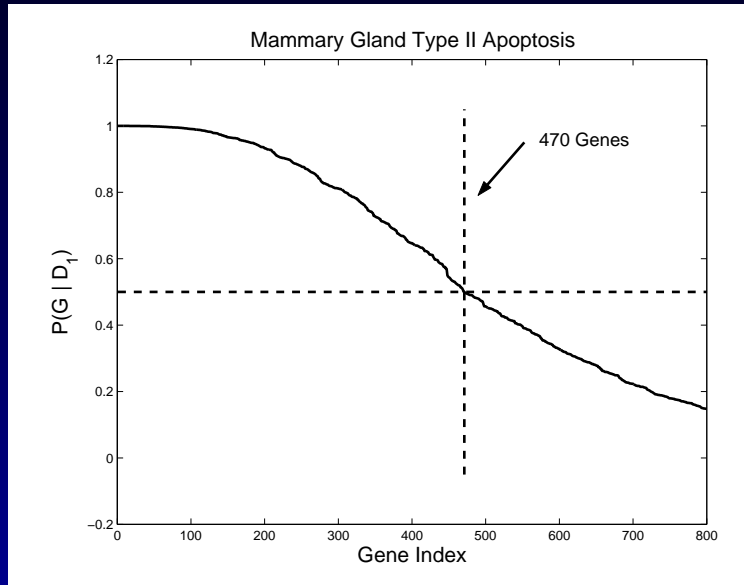
Include Information about Endothelial Cell Death

Model 0 hrs. vs. 24 hrs. as latent variable Probit GLM. Calculate $P(\mathcal{D}_2|G)$, the marginal likelihood.
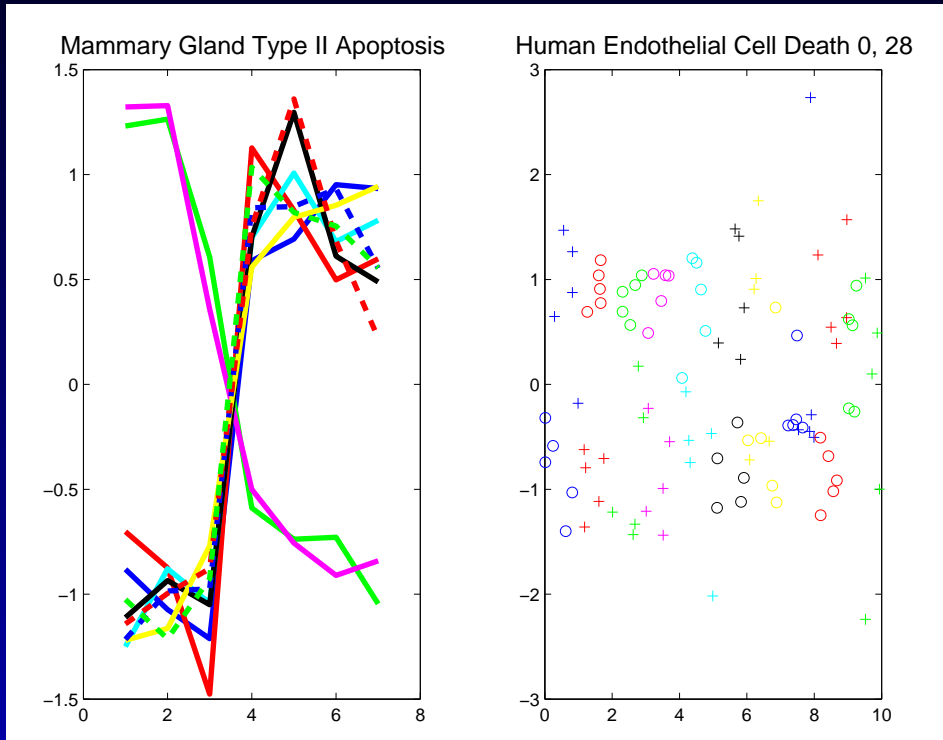
Bayes theorem gives a *principled* measure for ranking

$$P(G|\mathcal{D}_1, \mathcal{D}_2) = \frac{P(G|\mathcal{D}_1)p(\mathcal{D}_2|G)}{p(\mathcal{D}_2|\mathcal{D}_1)}$$

Combining Systems

Combined Result

Mammary Gland Type II Apoptosis

Human Endothelial Cell Death 0, 28

Top 10 $P(G = 1|\mathcal{D}_1, \mathcal{D}_2)$ for Mammary Type II Apoptosis *and* Endothelial cell death.

Gene Symbol	$P(G|\mathcal{D}_1, \mathcal{D}_2)$
TRIP6	1.00
PSMC2	1.00
PRDX4	1.00
BTG3	1.00
SCARB2	1.00
ZSWIM3	1.00
CD63	1.00
PDK4	1.00
GSTP1	1.00
TMSB10	1.00

Conclusion

- Lists of p-values are not capable of combining information from multi system microarray experiments.

- Probabilistic gene ranking provides an alternative to listing p-values.

- Extension to multi systems experiments is easily obtained via Bayes theorem.

- A combined analysis of two experiments demonstrates the feasibility of the proposed approach.

Table of Contents

- Multi Systems Analysis: Current Practise

- Motivation of Probabilistic Concepts

- Probabilistic Gene Ranking

- Extension to Multi Systems

- Combining Systems

- Combined Results

- Conclusion