Comparison of discrete measurements by directed graphical models using Gibbs Sampling

Højbjerre, Malene

Comparison of Discrete Measurements by Directed Graphical Models Using Gibbs Sampling

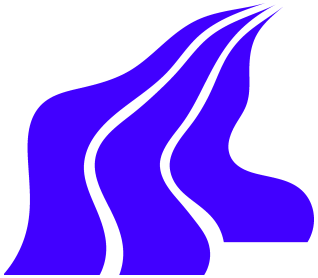
Malene Højbjerre

Department of Mathematical Sciences

Aalborg University

malene@math.auc.dk

www.math.auc.dk/~malene



7th Valencia International Meeting on Bayesian Statistics

June 2002

Motivation

Screening for Cervical Cancer:

Human Papilloma Virus (HPV) in uterine cervix has shown causative effect on developing cervical cancer, i.e. an important factor when screening for cervical cancer.

For 106 women HPV in uterine cervix detected by 4 screening methods:

- ➤ smear analysed in microscopy
- ➤ biopsy analysed in microscopy
- ➤ smear analysed by DNA
- ➤ biopsy analysed in DNA

Which method is best to screen for cervical cancer?

Problems:
- ➤ True class unknown
- ➤ Two levels of uncertainty
- ➤ Test results dependent

➜

Solution:
Bayesian approach to a two stage latent structure model/ directed graphical model

Directed Graphical Models

Directed Graphical Model:

Defined by

- Directed Acyclic Graph $\mathcal{G} = (\mathrm{V}, \mathrm{E})$

  V: set of vertices

  E: set of directed edges

- joint distribution of V is directed Markov wrt. to $\mathcal{G}$

Recursive factorization:

$$\mathrm{p}(\mathrm{V}) = \prod_{v \in \mathrm{V}} \mathrm{p}(v | \, \mathrm{pa}(v))$$

Directed local Markov property:

$$v \perp\!\!\!\perp \mathrm{nd}(v) \mid \mathrm{pa}(v)$$

Lauritzen(1996)

Terminology:

Vertices

$$\mathrm{V} = \mathrm{X} \cup \mathrm{Y} \cup \Theta$$

where

X: observed data

Y: unobserved data and/or latent variables

Θ: parameters

Inference about Θ

Bayesian Inference by MCMC Methods

Bayesian Inference: all quantities random

Prior $\quad p(\theta)$

Posterior $\quad p(\theta|x) = \int p(y, \theta|x)dy$ intractable integral

Gibbs Sampling:

Successively simulate values from the full conditionals

$$p(v|V \setminus v) \propto p(v|\,\mathrm{pa}(v)) \prod_{w:v\in \mathrm{pa}(w)} p(w|\,\mathrm{pa}(w)), \quad v \in Y \cup \Theta$$

Converges to a Markov chain with stationary distribution $p(y, \theta|x)$

- Marginalize by considering only parts of simulated values
- Inference is based on summary statistics of simulated values

Spiegelhalter(1998)

Bayesian Inference by MCMC Methods

Software:

- BUGS (Bayesian inference Using Gibbs Sampling)

  Spiegelhalter et al (1996)

- CODA (Convergence Diagnostics and Output Analysis)

  Best et al (1996)

Prior: a prerequisite

- information - a strength
- no/little information - a prior with large variance chosen almost per default

Influence of prior? $\rightarrow$ Prior sensitivity analysis by likelihood inference

Likelihood Inference by MCMC Methods

Likelihood Inference:

Likelihood in θ_0 (specific parameter value)

$$
\begin{aligned}
L(\theta_0|\mathbf{x}) &= \int p(\mathbf{x},\mathbf{y}|\theta_0)d\mathbf{y} && \text{intractable integral}\\
&= \iint p(\mathbf{x},\mathbf{y}|\theta_0)p(\theta)d\mathbf{y}d\theta && (\textstyle\int p(\theta)d\theta = 1)\\
&= \iint \frac{p(\mathbf{x},\mathbf{y}|\theta_0)}{p(\mathbf{x},\mathbf{y}|\theta)}p(\mathbf{x},\mathbf{y}|\theta)p(\theta)d\mathbf{y}d\theta\\
&= p(\mathbf{x}) \iint \frac{p(\mathbf{x},\mathbf{y}|\theta_0)}{p(\mathbf{x},\mathbf{y}|\theta)}p(\mathbf{y},\theta|\mathbf{x})d\mathbf{y}d\theta && (p(\mathbf{x},\mathbf{y}|\theta)p(\theta) = p(\mathbf{y},\theta|\mathbf{x})p(\mathbf{x}))
\end{aligned}
$$

Likelihood Approximation by Gibbs Sampling:

Sample $(\mathbf{y}^{(1)},\theta^{(1)}),(\mathbf{y}^{(2)},\theta^{(2)}),\ldots,(\mathbf{y}^{(N)},\theta^{(N)})$ from $p(\mathbf{y},\theta|\mathbf{x})$ [BUGS]

$$
\begin{aligned}
\tilde{L}(\theta_0|\mathbf{x}) &\propto \sum_{j=1}^{N} \frac{p(\mathbf{x},\mathbf{y}^{(j)}|\theta_0)}{p(\mathbf{x},\mathbf{y}^{(j)}|\theta^{(j)})}\\
&= \sum_{j=1}^{N} \prod_{x,y\in ch(\Theta)} \frac{p(x|\,pa(x)^{(\mathbf{x},\mathbf{y}^{(j)},\theta_0)})p(y^{(j)}|\,pa(x)^{(\mathbf{x},\mathbf{y}^{(j)},\theta_0)})}{p(x|\,pa(x)^{(\mathbf{x},\mathbf{y}^{(j)},\theta^{(j)})})p(y^{(j)}|\,pa(x)^{(\mathbf{x},\mathbf{y}^{(j)},\theta^{(j)})})}
\end{aligned}
$$

Likelihood Inference by MCMC Methods

Profile Log-likelihood:

$$\log \hat{\mathrm{L}}\left(\theta_i|\boldsymbol{x}\right) = \sup_{\Theta\setminus i} \log \mathrm{L}\left(\boldsymbol{\theta}|\boldsymbol{x}\right)$$

Profile Log-likelihood Approximation by Gibbs Sampling:

1. Compute $\log \tilde{\mathrm{L}}\left(\boldsymbol{\theta}_0|\boldsymbol{x}\right)$ in grid formed by quantiles of Gibbs output $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}), \ldots, \boldsymbol{\theta}^{(\mathrm{N})}$
2. Maximize over the grid to approximate profile log-likelihood

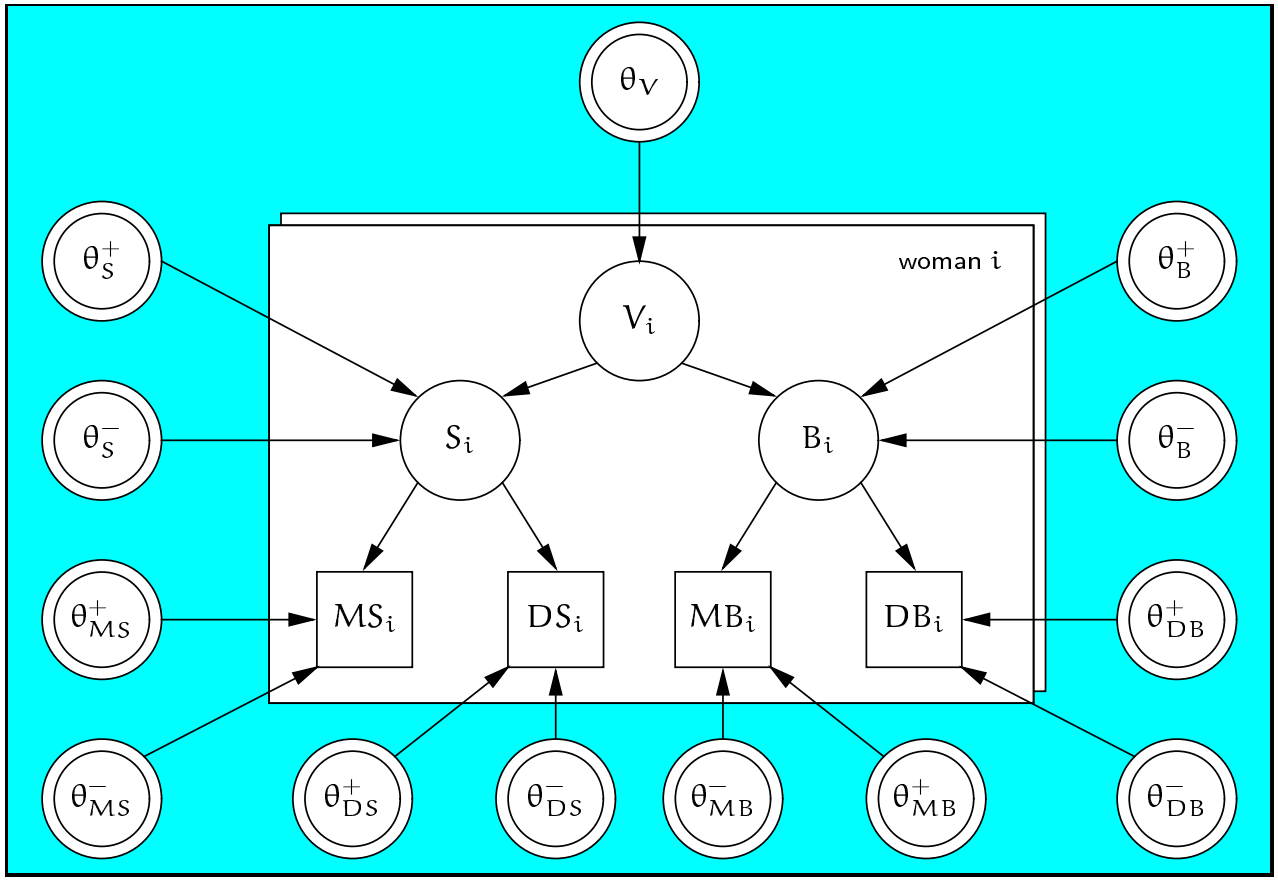
Profile Log-likelihood Approximation of Function of Parameters:

1. Compute function value of each grid point and pair this with corresponding log-likelihood approximation
2. Bin pairs wrt. function value
3. Maximize wrt. log-likelihood value over bin to approximate profile log-likelihood of function

Højbjerre (2002)

Screening for Cervical Cancer

Directed Graphical Model:



Højbjerre (2001)

Screening for Cervical Cancer

For woman i, $i = 1, 2, \ldots, 106$

$$V_i = \begin{cases} 1 & \text{if HPV in uterine cervix} \\ 0 & \text{if HPV not in uterine cervix} \end{cases}$$

$$S_i = \begin{cases} 1 & \text{if HPV in smear} \\ 0 & \text{if HPV not in smear} \end{cases}$$

$$B_i = \begin{cases} 1 & \text{if HPV in biopsy} \\ 0 & \text{if HPV not in biopsy} \end{cases}$$

$$MS_i = \begin{cases} 1 & \text{if HPV in smear by microscopy} \\ 0 & \text{if HPV not in smear by microscopy} \end{cases}$$

$$DS_i = \begin{cases} 1 & \text{if HPV in smear by DNA} \\ 0 & \text{if HPV not in smear by DNA} \end{cases}$$

$$MB_i = \begin{cases} 1 & \text{if HPV in biopsy by microscopy} \\ 0 & \text{if HPV not in biopsy by microscopy} \end{cases}$$

$$DB_i = \begin{cases} 1 & \text{if HPV in biopsy by DNA} \\ 0 & \text{if HPV not in biopsy by DNA} \end{cases}$$

$$\theta_V = P(V_i = 1) \qquad V_i|\theta_V \sim \text{Bern}(\theta_V)$$

$$\theta_S^+ = P(S_i = 1|V_i = 1)$$

$$\theta_S^- = P(S_i = 1|V_i = 0) \qquad S_i|V_i, \theta_S^+, \theta_S^- \sim \text{Bern}\left(\theta_S^+ V_i + \theta_S^-(1 - V_i)\right)$$

$$\theta_{MS}^+ = P(MS_i = 1|S_i = 1)$$

$$\theta_{MS}^- = P(MS_i = 1|S_i = 0) \qquad MS_i|S_i, \theta_{MS}^+, \theta_{MS}^- \sim \text{Bern}\left(\theta_{MS}^+ S_i + \theta_{MS}^-(1 - S_i)\right)$$

$$\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots$$

Screening for Cervical Cancer

Quantities of interest:

$$
\begin{aligned}
\text{sen}_{\text{MS}} &= P(MS_i = 1 | V_i = 1) \\
&= \theta^-_{MS} + \theta^+_S (\theta^+_{MS} - \theta^-_{MS}) \\
\text{spe}_{\text{MS}} &= P(MS_i = 0 | V_i = 0) \\
&= 1 - \theta^-_{MS} - \theta^-_S (\theta^+_{MS} - \theta^-_{MS}) \\
\text{sen}_{\text{DS}} &= P(DS_i = 1 | V_i = 1) \\
&\vdots
\end{aligned}
$$

Compare for all methods:

- ➤ sensitivity
- ➤ specificity

Prior:

	Prior 1	*Prior 2*	*Prior 3*
θ_V	$\text{Be}(6, 14)$	$\text{Be}(1, 1)$	$\text{Be}(1, 1)$
θ^+_S	$\text{Be}(112, 6)$	$\text{Be}(2, 1)$	$\text{Be}(2, 1)$
θ^-_S	$\text{Be}(9, 581)$	$\text{Be}(9, 581)$	$\text{Be}(9, 581)$
θ^+_B	$\text{Be}(3.0, 1.3)$	$\text{Be}(2, 1)$	$\text{Be}(2, 1)$
θ^-_B	$\text{Be}(9, 581)$	$\text{Be}(9, 581)$	$\text{Be}(9, 581)$
θ^+_{MS}	$\text{Be}(13.8, 9.2)$	$\text{Be}(2, 1)$	$\text{Be}(2, 1)$
θ^-_{MS}	$\text{Be}(3.5, 31.5)$	$\text{Be}(1, 2)$	$\text{Be}(1, 2)$
θ^+_{DS}	$\text{Be}(69.6, 2.2)$	$\text{Be}(2, 1)$	$\text{Be}(2, 1)$
θ^-_{DS}	$\text{Be}(9, 581)$	$\text{Be}(1, 2)$	$\text{Be}(9, 581)$
θ^+_{MB}	$\text{Be}(13.8, 9.2)$	$\text{Be}(2, 1)$	$\text{Be}(2, 1)$
θ^-_{MB}	$\text{Be}(3.5, 31.5)$	$\text{Be}(1, 2)$	$\text{Be}(1, 2)$
θ^+_{DB}	$\text{Be}(54.4, 4.7)$	$\text{Be}(2, 1)$	$\text{Be}(2, 1)$
θ^-_{DB}	$\text{Be}(9, 581)$	$\text{Be}(1, 2)$	$\text{Be}(9, 581)$

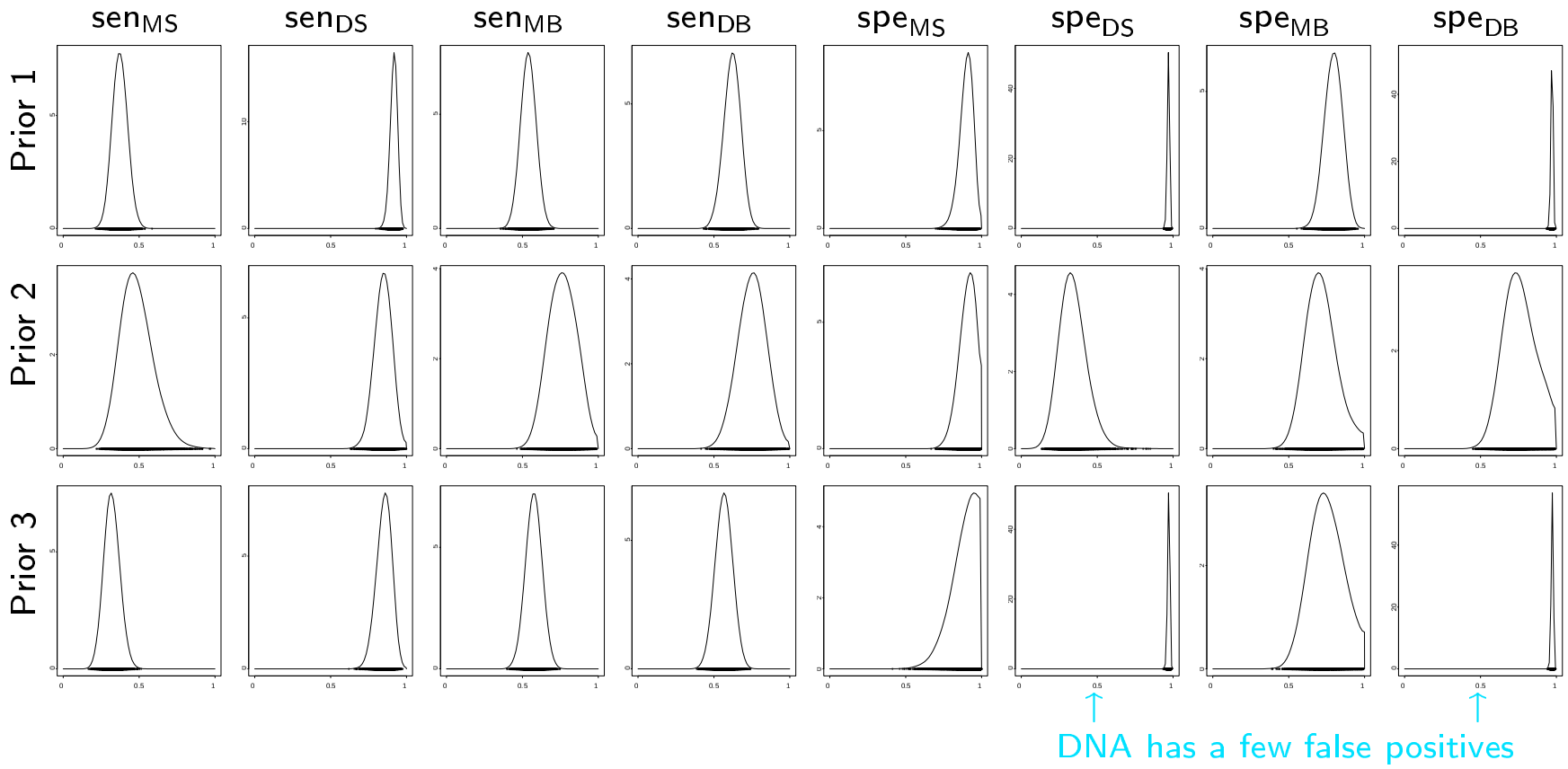
$\theta \sim \text{Be}(9, 581)$: $\mathbb{E}(\theta) = 0.015$ $2\sqrt{\mathbb{V}(\theta)} = 0.01$

Specifies that HPV not in smear or biopsy, if not in cervix and DNA has few false positives

Screening for Cervical Cancer

Bayesian Analysis:

Posterior:



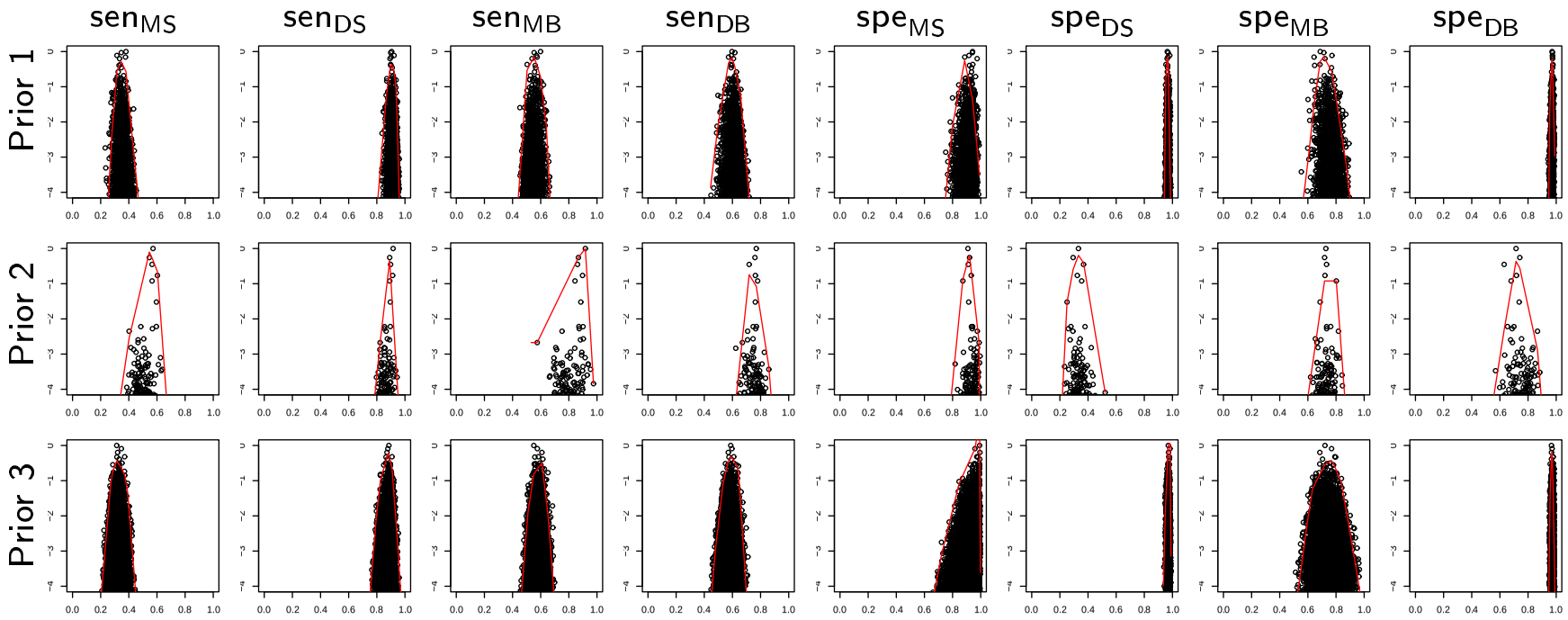
- Prior great influence
- Prior 2 (large variance) contradicts the well-known fact that DNA has few false positives

$\rightsquigarrow$ prior sensitivity analysis by approximating profile log-likelihood

Screening for Cervical Cancer

Likelihood Analysis:

Projected log-likelihood:



- Conclusions very dependent on prior
- Likelihood analysis reveals problems with default prior

Screening for Cervical Cancer

Summary: (Prior 1 and Prior 3)

Parameter	Posterior mean	95% Cred. interval	MLE	95% Conf. interval
sen_{MS}	0.32	0.23 – 0.42	0.31	0.26 – 0.40
sen_{DS}	0.85	0.76 – 0.93	0.89	0.80 – 0.93
sen_{MB}	0.57	0.48 – 0.67	0.55	0.51 – 0.68
sen_{DB}	0.58	0.46 – 0.67	0.59	0.50 – 0.67
spe_{MS}	0.88	0.68 – 0.99	0.97	0.82 – 0.99
spe_{DS}	0.97	0.96 – 0.98	0.97	0.94 – 0.99
spe_{MB}	0.75	0.55 – 0.96	0.72	0.60 – 0.87
spe_{DB}	0.97	0.96 – 0.98	0.97	0.95 – 0.99

- Smear analysed by DNA is most sensitive and most specific - based on prior information that DNA has no false positives

Discussion:

- Prior sensitivity analysis is possible by MCMC likelihood inference
- Likelihood analysis reveals problems with default priors - a supplement to the Bayesian analysis
- Analysis forms basis for a general method to compare discrete measurements where true class unknown

References

Best, N. G., Cowles, M. K. & Vines, K. (1996). *CODA, Convergence Diagnosis and Output Analysis Software for Gibbs Sampling Output, Version 0.30*, MRC Biostatistics Unit, Cambridge.

Højbjerre, M. (2001a). Comparison of four screening methods for cervical cancer analysed with directed graphical models by Gibbs sampling, *Technical Report R-01-2030*, Department of Mathematical Sciences, Aalborg University, Aalborg, Denmark.

Højbjerre, M. (2002). Profile likelihood in directed graphical models from BUGS output, *to appear in Statistics and Computing*.

Lauritzen, S. L. (1996). *Graphical Models*, Clarendon Press, Oxford, UK.

Spiegelhalter, D. J. (1998). Bayesian graphical modelling: a case-study in monitoring health outcomes, *Applied Statistics* **47**: 115–133.

Spiegelhalter, D. J., Thomas, A., Best, N. G. & Gilks, W. (1996). *BUGS 0.5 Bayesian inference Using Gibbs sampling Manual (version ii)*, MRC Biostatistics Unit, Cambridge.