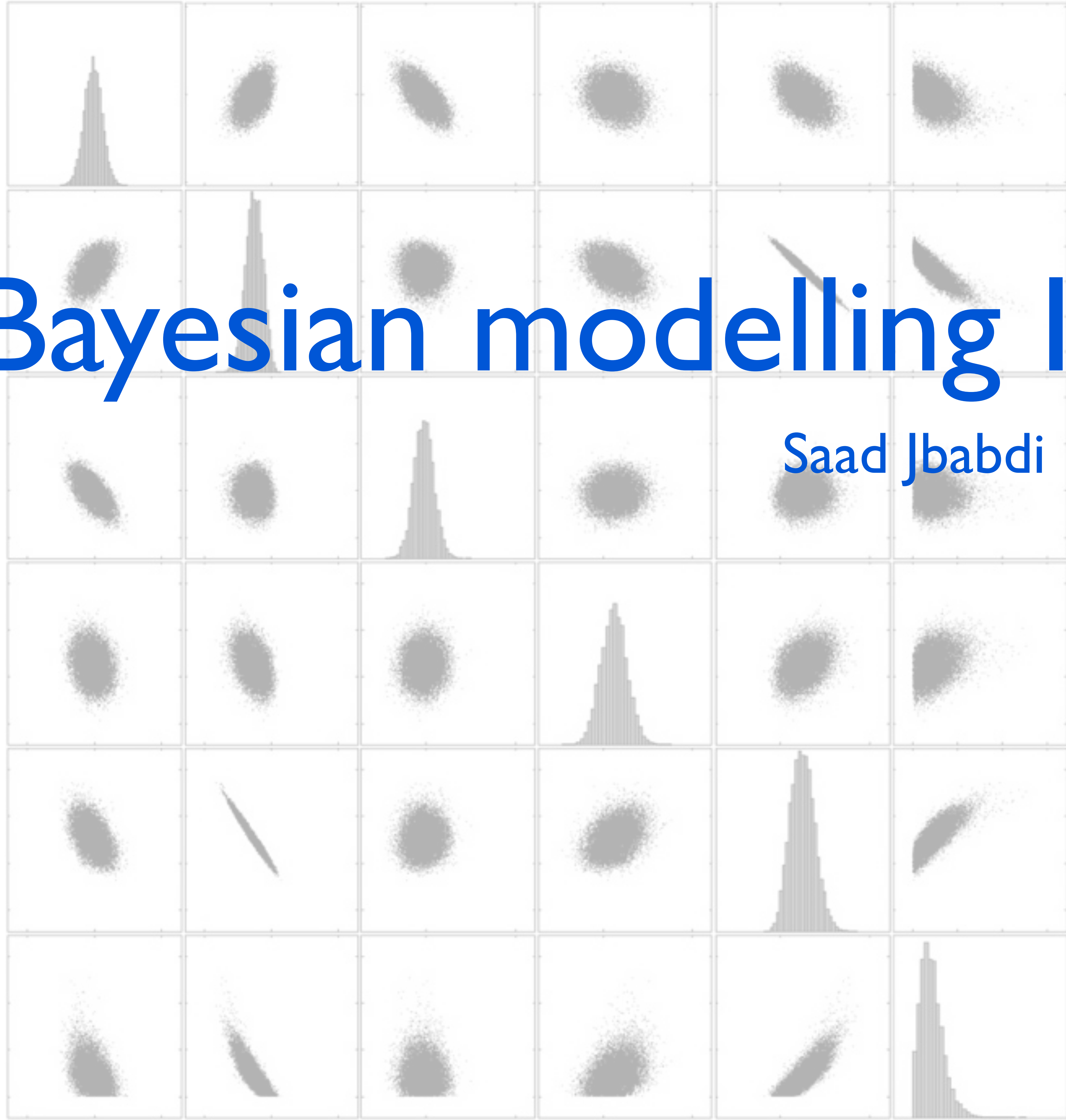
Bayesian modelling II

Saad Jbabdi



- Graphical models
  - Data fusion
  - Kalman Filter
- Model selection

- quick recap

Bayes = Conditional probabilities

how does p(a|b) relate to p(b|a)

Bayesian modelling and inference

y : data

a : parameters of a "generative" model a--->y

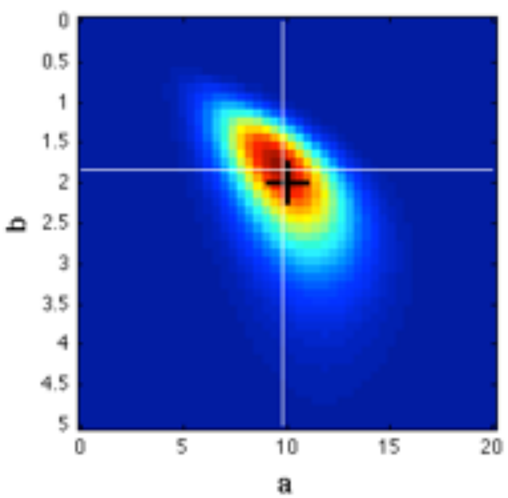
$$p(a|y) = p(y|a) * p(a) / p(y)$$

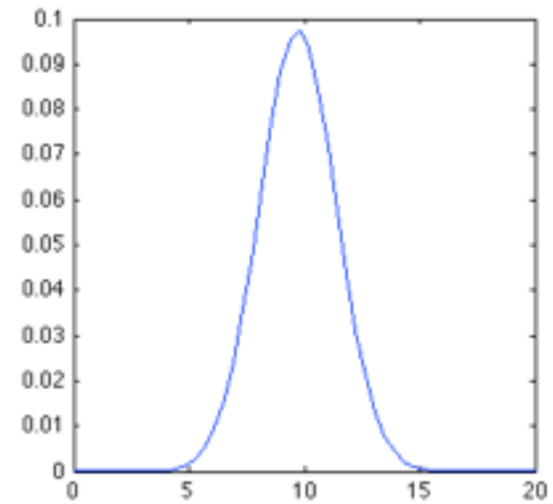
$\underline{p(a|y)}$: infer

$\underline{p(y|a) * p(a)}$: model

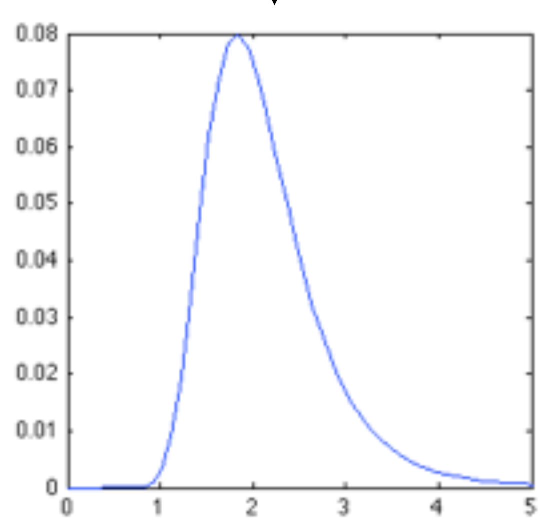
Joint- conditional - marginal posterior

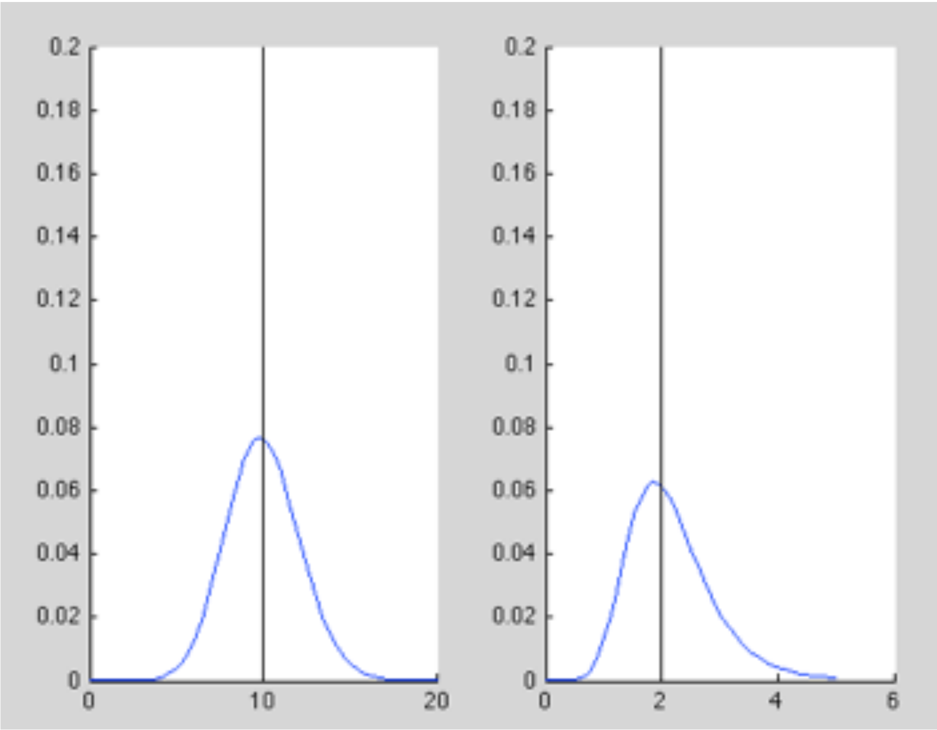
joint posterior





conditional posteriors





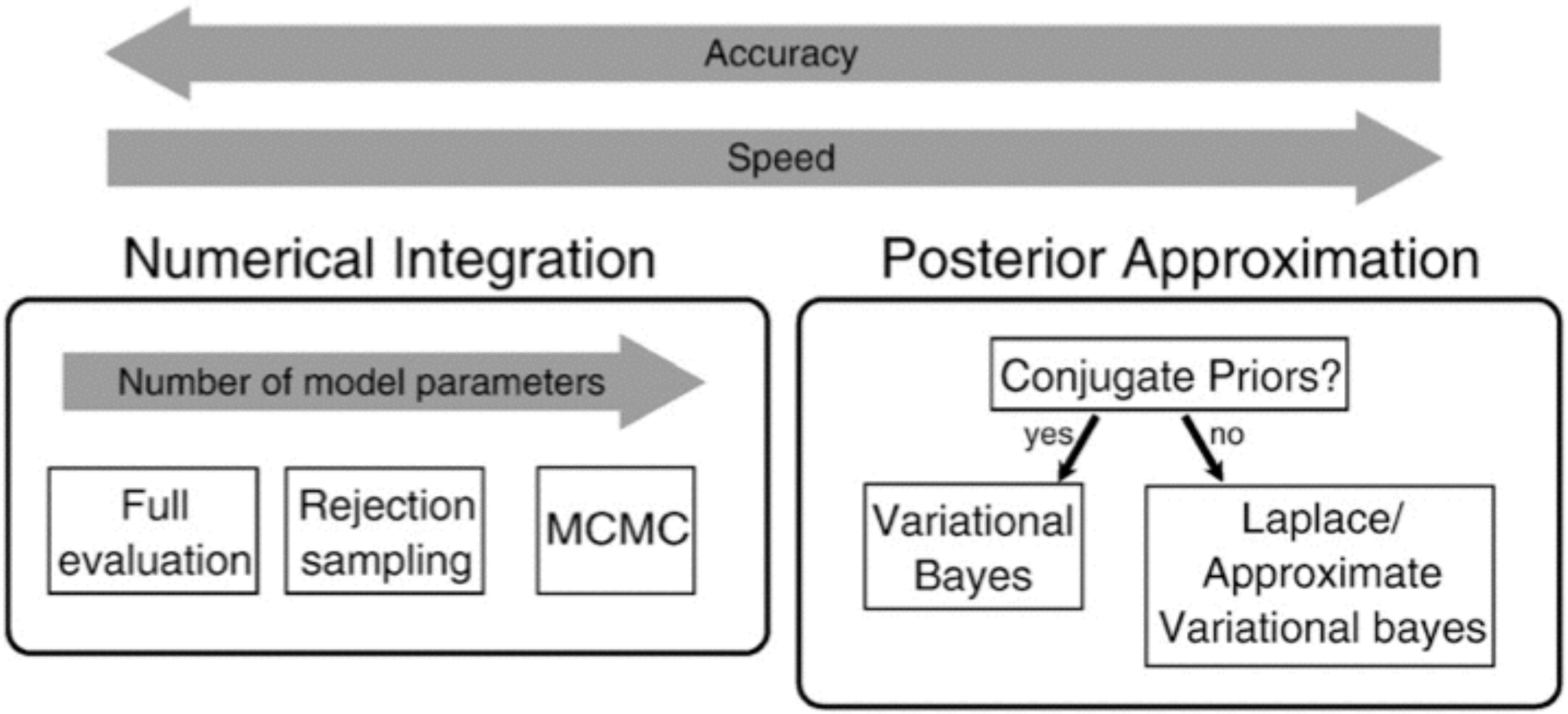
marginal posteriors

Bayesian Inference

- Exact inference

  Analytic integrals (gaussian - or nasty)
  Grid evalutation (1-5 parameters, max)

Bayesian Inference



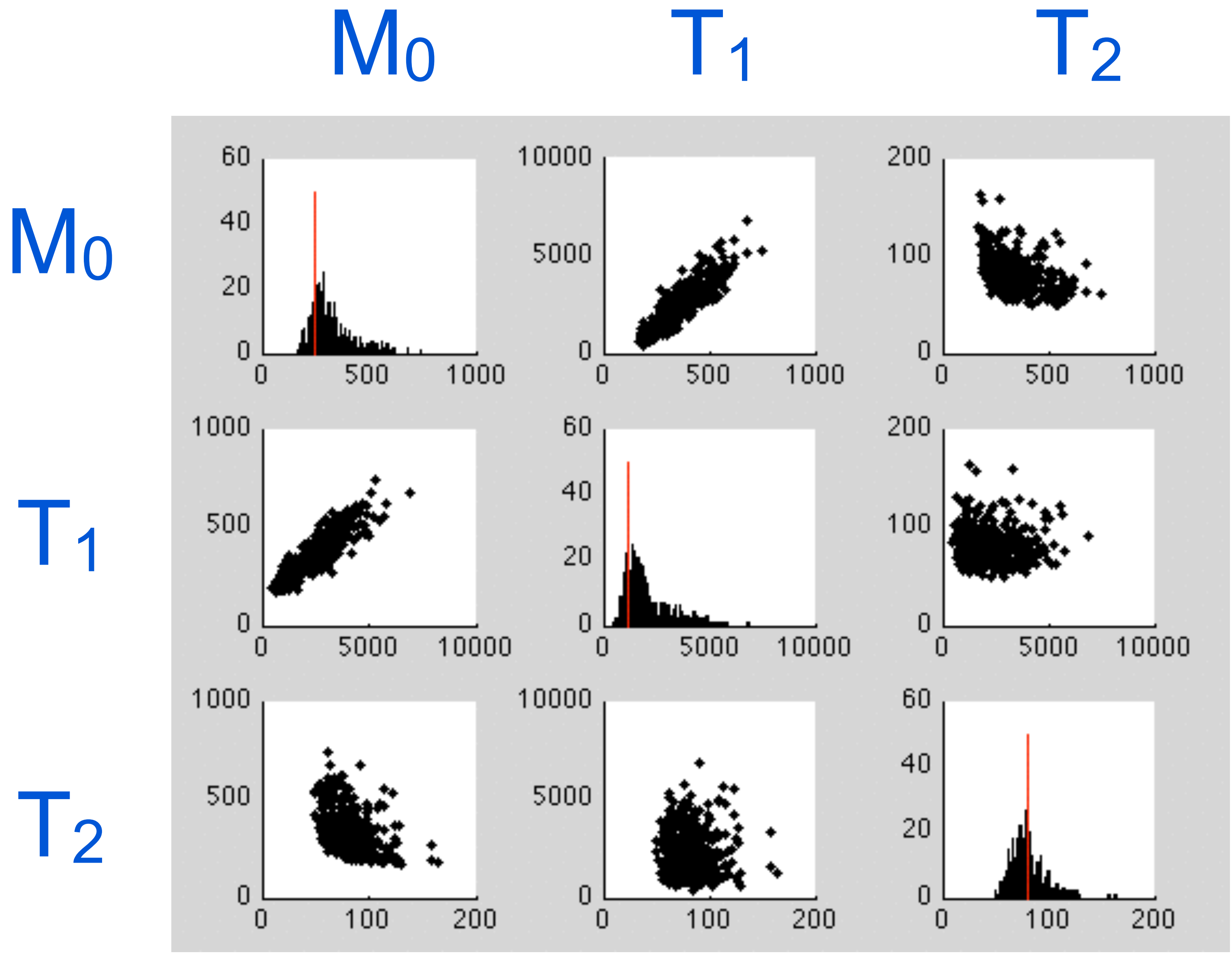
Woolrich et al. 2009

Metropolis Hastings Code

```
function samples = mh(varargin)
% res = mh(y,x0,@genfunc,[LB,UB,params])
%
% Compulsory Parameters
%   y        = data (Nx1)
%   x0       = initial parameters (Px1)
%   @genfunc = generative model
%
%   Output is nsamples*P where nsamples=njumps/sampleevery
%
% Example:
%   % define a forward model (here y=a*exp(-bx))
%   myfun=@(x,c)(exp(-x(1)*c)+x(2));
%   % generate some noisy data
%   true_x = [1;2];
%   c=linspace(1,10,100);
%   y=myfun(true_x,c) + .05*randn(1,100);
%   % estimate parameters
%   x0=[1;2]; % you can get x0 using nonlinear opt
%   samples=mh(y,x0,@(x)(myfun(x,c)));
%   figure,plot(samples)
%
% Other Parameters
%   LB = lower bounds on x (default=[-inf]*ones(size(x0)))
%   UB = upper bounds on x (default=[+inf]*ones(size(x0)))
%   params.burnin = #burnin iterations (default=1000)
%   params.njumps = #jumps (default=5000)
%   params.sampleevery = keep every n jumps (default=10)
%   params.update = update proposal every n jumps (default=20)
%
%
% S. Jbabdi 01/12
```

TR = 1, 1.6, 2.3, 4 s
TE = 30, 47, 65, 82, 100 ms

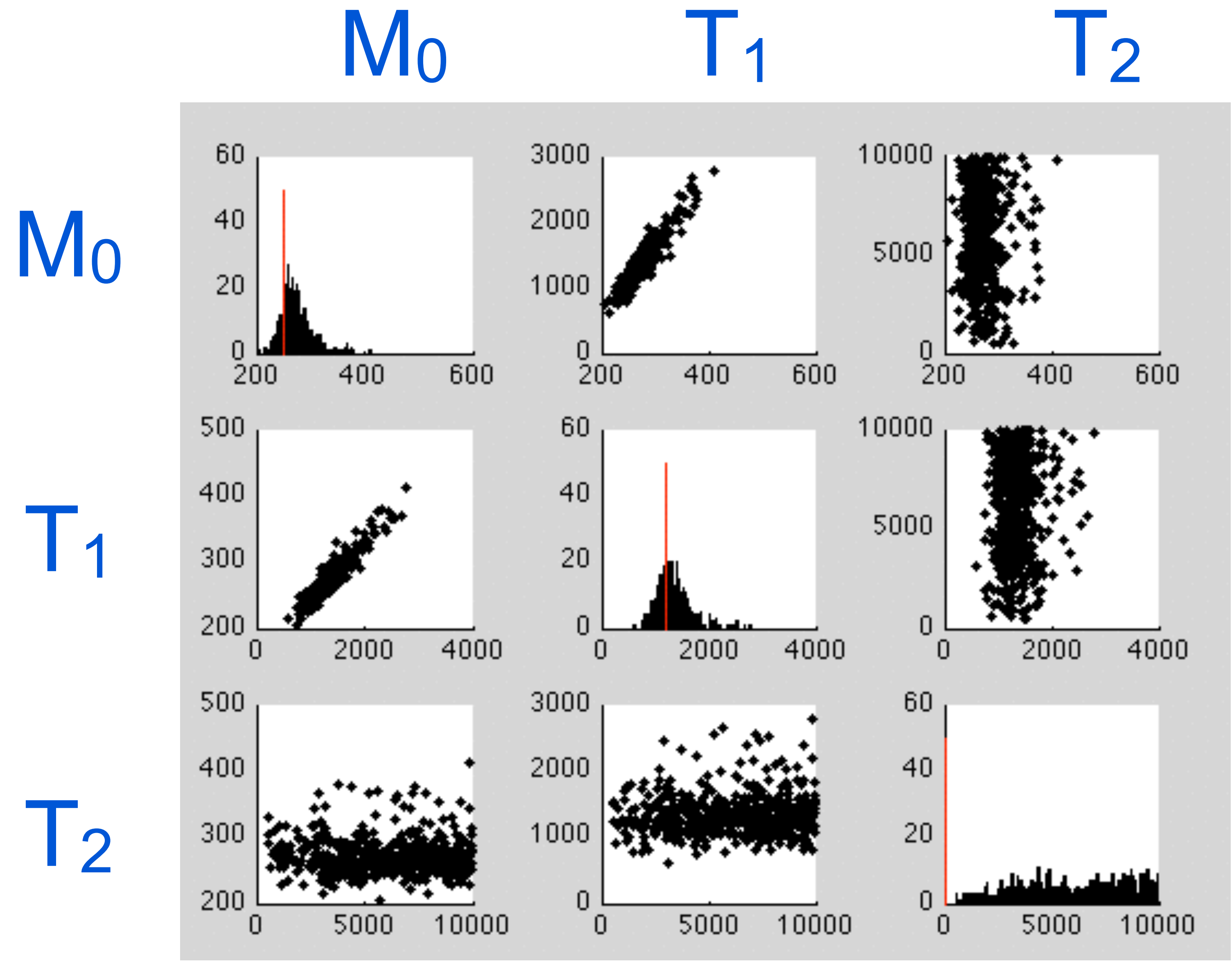
$Y = M_0 * \exp(-TE/T_2) * (1-\exp(-TR/T_1))$



SNR=20

Uncertainty and modelling

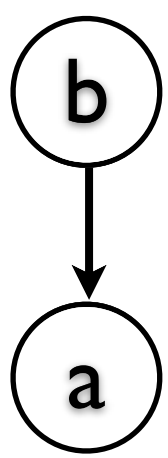
- Uncertainty doesn't only come from noise, it also comes from fitting the wrong model



Graphical models

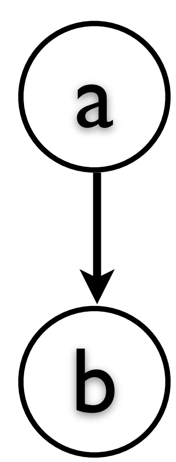
$p(a,b)$ start with a joint distribution

$p(a,b)=p(a|b)p(b)$ decompose it using the product rule



draw it using arrows instead of |

$p(a,b)=p(b|a)p(a)$



note that I could have done this

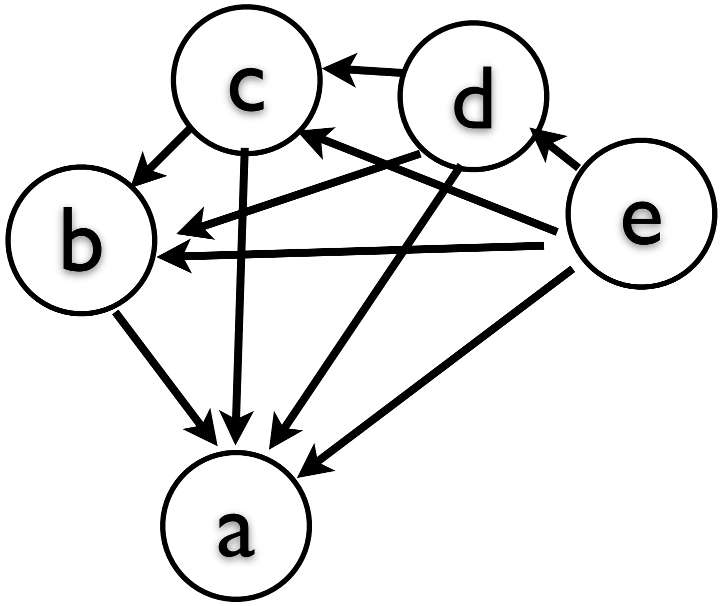
Graphical models

p(a,b,c,d,e)

let's try a longer joint distribution

$p(a,b,c,d,e)=p(a|b,c,d,e)p(b|c,d,e)p(c|d,e)p(d|e)p(e)$

I can decompose it like this



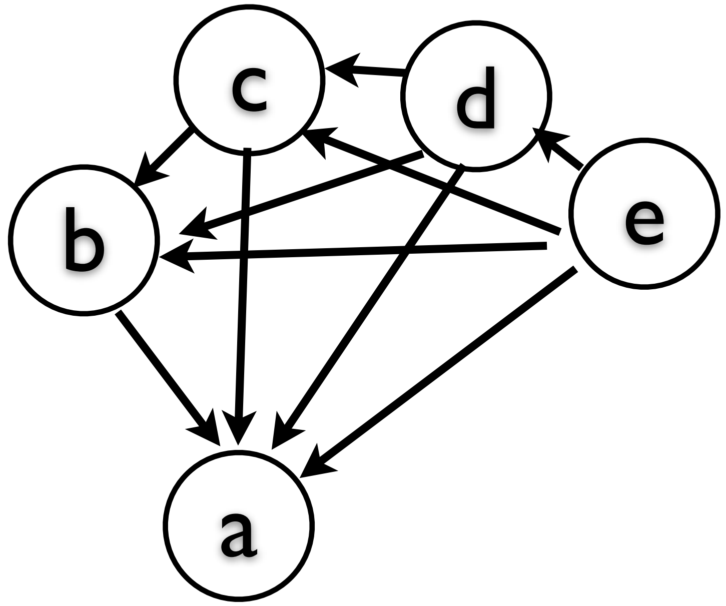
draw it using arrows instead of |

terminology

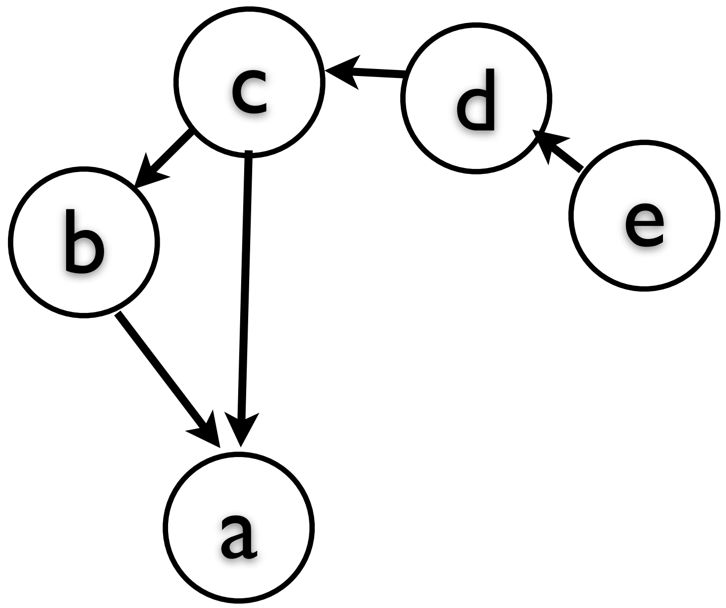
Directed acyclic graphs
Bayesian networks
Directed graphical models
Belief networks

Graphical models

p(a,b,c,d,e)=p(a|b,c,d,e)p(b|c,d,e)p(c|d,e)p(d|e)p(e)

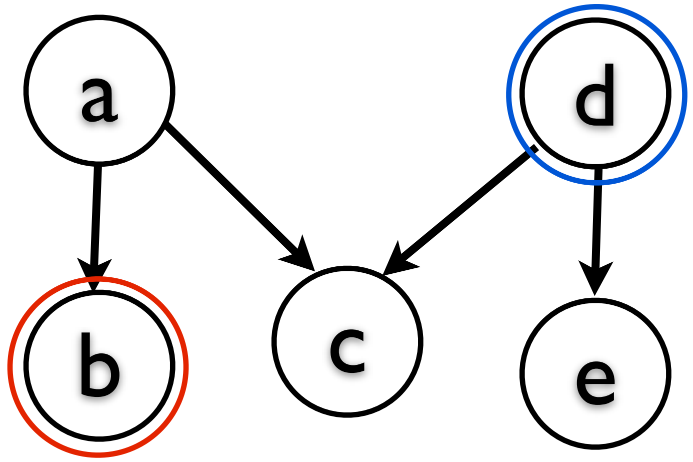


p(a,b,c,d,e)=p(a|b,c)p(b|c)p(c|d)p(d|e)p(e)



choose conditional independences

Graphical models



$p(a,b,c,d,e)=p(b|a)p(c|a,d)p(e|d)p(a)p(d)$

I observe b.
What is $p(d|b)$?

$$p(\mathbf{d}|\mathbf{b}) = \frac{p(\mathbf{d},\mathbf{b})}{p(\mathbf{b})}$$

get rid of conditionals

$$p(d|b) = \frac{\int p(a,b,c,d,e)\,dadcde}{\int p(a,b,c,d,e)\,dadcdedd}$$

definition of marginals

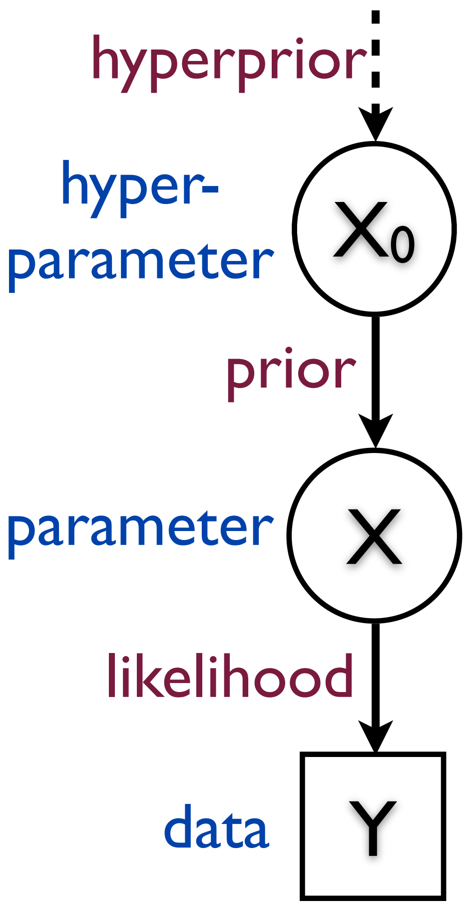
$$p(d|b) = \frac{p(d)\int p(b|a)p(c|a,d)p(e|d)p(a)\,dadcde}{\int p(b|a)p(c|a,d)p(e|d)p(a)p(d)\,dadcddde}$$

(“e” will disappear because $\int p(e|d)de=1$)

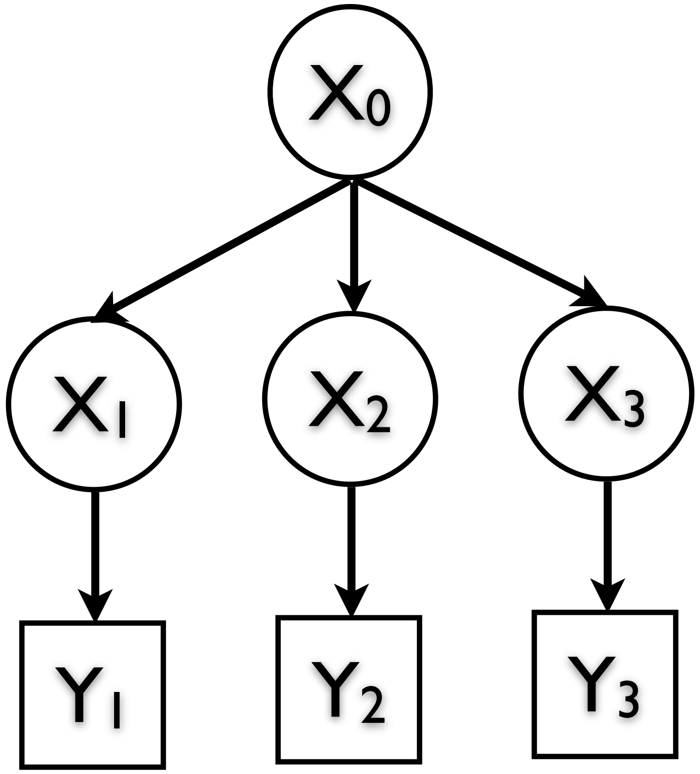
$$p(d|b) \propto p(d)\int p(b|a)p(c|a,d)p(a)\,dadc$$

information “propagates” through a and c

Graphical models

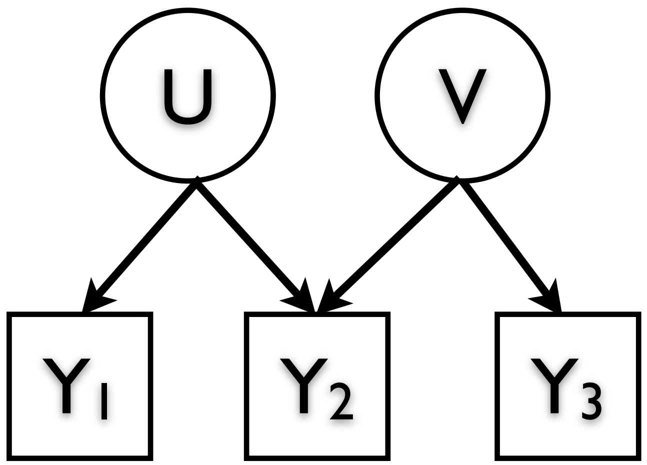


- arrows point at data - generative model
- data are observed - so we can condition on them
- graph allows us to visualise conditional dependencies



Graphical models

Data fusion



$Y_1 = U$ +noise

$Y_2 = U + V$ +noise

$Y_3 = V$ +noise

$$p(U|Y_1, Y_2, Y_3) = ?$$

$$p(Y_1|U) = N(U|Y_1, 1/b_1)$$

$$p(Y_2|U,V) = N(U+V|Y_2, 1/b_2)$$

$$p(Y_3|V) = N(V|Y_3, 1/b_3)$$

For simplicity: $p(U)=p(V)=1$ (uniform priors)

Gaussian $N(x|m,1/b)$

[b=precision=1/variance]

$N(x|m,1/b) \sim b.\exp\{-0.5b(x-m)^2\}$

Two useful facts about the Gaussian distribution

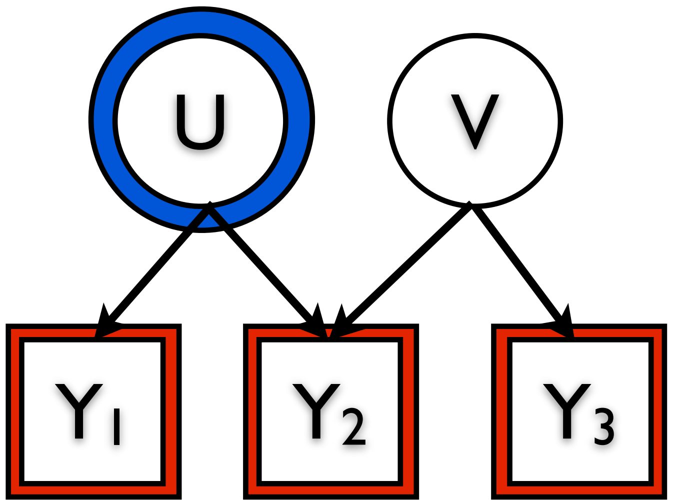
$N(\mathbf{x}|a, 1/b) = N(a|\mathbf{x}, 1/b)$

$N(\mathbf{x}|a, 1/b)N(\mathbf{x}|c, 1/d) = N(\mathbf{x}|..., b+d)N(a|c, 1/b+1/d)$

If interested in x, then it's a Gaussian
If you're not interested in x, then the integral is a Gaussian

Graphical models

Data fusion



$Y_1 = U$ +noise
$Y_2 = U + V$ +noise
$Y_3 = V$ +noise

$p(Y_1|U) = N(Y_1|U, 1/b_1)$
$p(Y_2|U+V) = N(Y_2|U+V, 1/b_2)$
$p(Y_3|V) = N(Y_3|V, 1/b_3)$

$$p(U,V,Y_1,Y_2,Y_3) = p(Y_1|U)p(Y_2|U,V)p(Y_3|V)p(U)p(V)$$

$$p(U|Y_1,Y_2,Y_3) \propto \int p(U,V,Y_1,Y_2,Y_3)dV$$

$$p(U|Y_1,Y_2,Y_3) \propto \int p(Y_1|U)p(Y_2|U,V)p(Y_3|V)dV$$

$$p(U|Y_1,Y_2,Y_3) \propto \int N(Y_1|U,1/b_1)N(Y_2|U+V,1/b_2)N(Y_3|V,1/b_3)dV$$

$$p(U|Y_1,Y_2,Y_3) \propto N(Y_1|U,1/b_1)\int N(V|Y_2-U,1/b_2)N(V|Y_3,1/b_3)dV$$

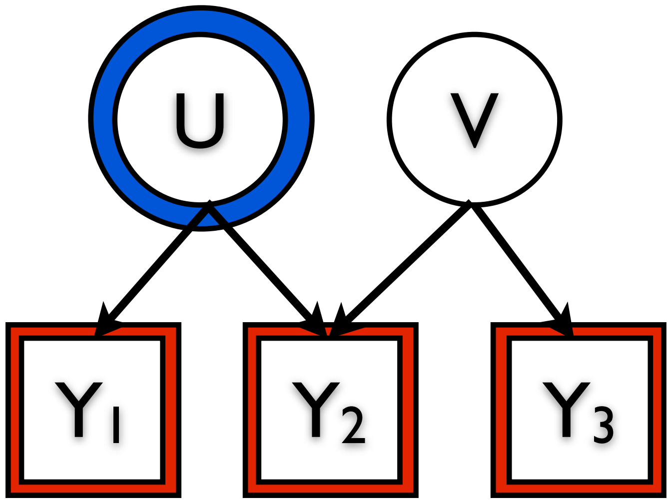
$$p(U|Y_1,Y_2,Y_3) \propto N(Y_1|U,1/b_1)\int N(V|\ldots,\ldots)N(U|Y_2-Y_3,1/b_2+1/b_3)dV$$

$$p(U|Y_1,Y_2,Y_3) \propto N(U|Y_1,1/b_1)N(U|Y_2-Y_3,1/b_2+1/b_3)$$

$$=N(U|(b_1Y_1+(Y_2-Y_3)b_{23})/(b_1+b_{23}), b_1+b_{23})$$

$b_{23} = b_2b_3/(b_2+b_3)$

Let's stare at this posterior



$Y_1 = U$ +noise
$Y_2 = U + V$ +noise
$Y_3 = V$ +noise

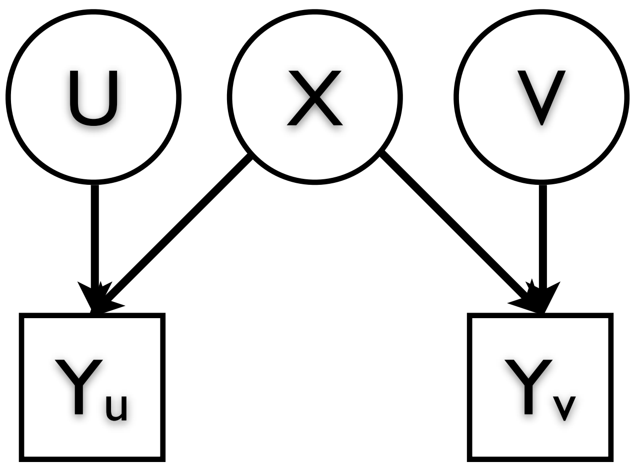
$$p(U|Y_1,Y_2,Y_3) = N\{U|\frac{b_1 Y_1 + (Y_2 - Y_3)b_{23}}{b_1 + b_{23}}, 1/(b_1 + b_{23})\}$$

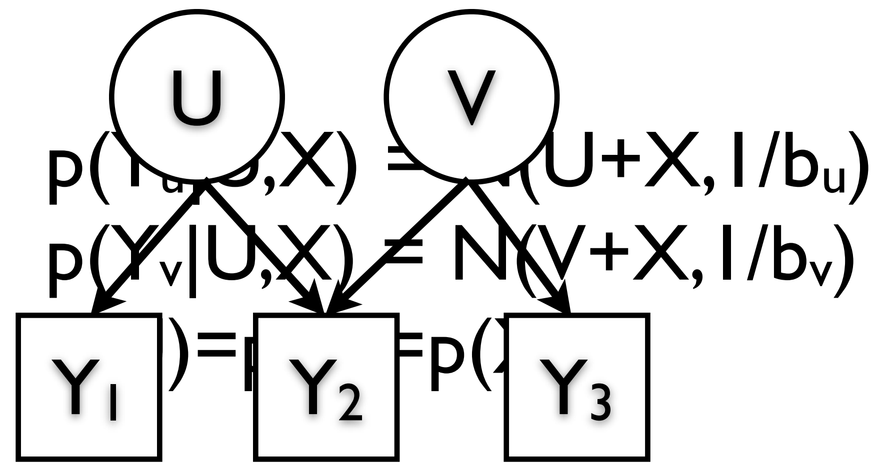
$b_{23} = b_2 b_3/(b_2 + b_3)$

imagine Y_3 has lots of noise: $b_3=0$, i.e. $b_{23}=0$, i.e. $p(U|Y_1,Y_2,Y_3)=N(U|Y_1,1/b_1)$

Graphical models

Data fusion





$$p(Y_u,Y_v,U,V,X) = p(Y_u|U,X)p(Y_v|V,X)p(U)p(V)p(X)$$

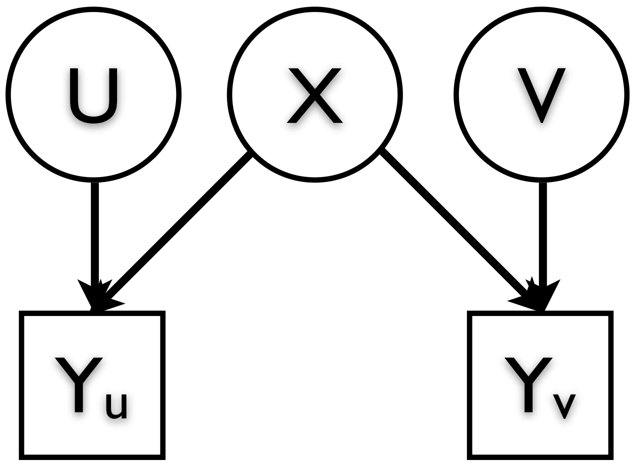
$$p(U|Y_u,Y_v) = p(U)\int p(Y_u|U,X)p(Y_v|V,X)dXdV$$

$$p(U|Y_u,Y_v) = \int\int N(Y_u|U+X,1/b_u)N(Y_v|V+X,1/b_v)dVdX$$

$$p(U|Y_u,Y_v) = \int N(X|U-Y_u,1/b_u)\{\int N(V|X-Y_v,1/b_v)dV\}dX = 1 \ !!!$$

Graphical models

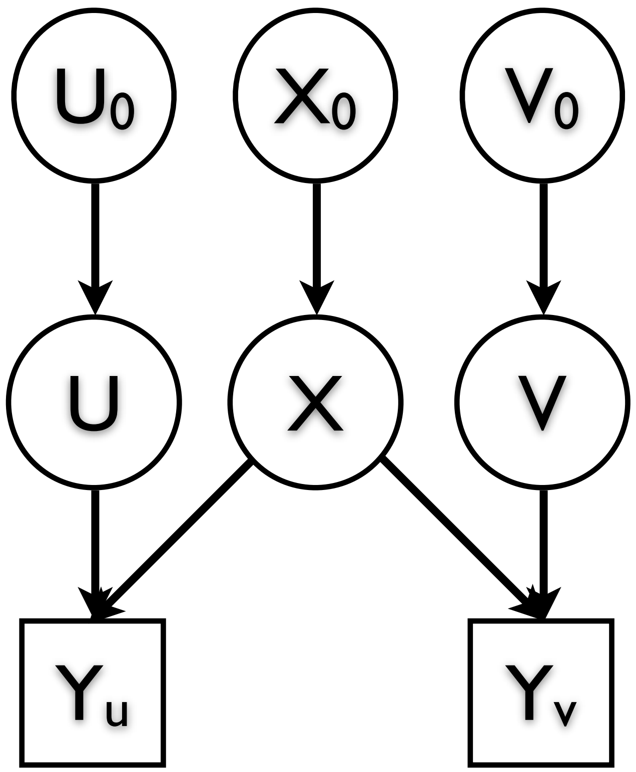
Data fusion



$p(Y_u|U,X) = N(U+X, 1/b_u)$
$p(Y_v|U,X) = N(V+X, 1/b_v)$
$p(U)=p(V)=p(X)=1$

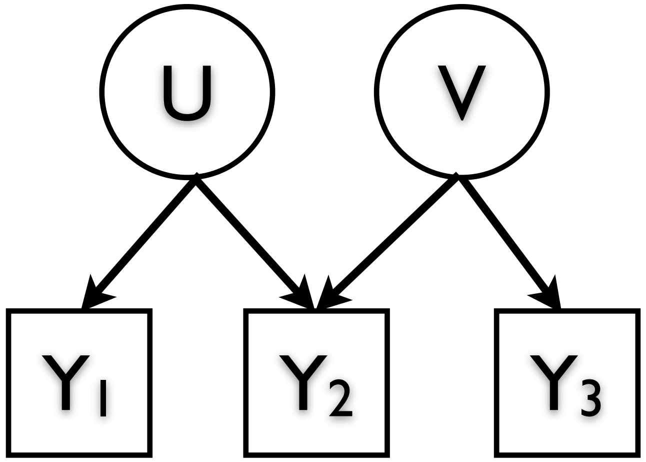
$Y_u = U+X$ + noise
$Y_v = V+X$ + noise
$p(U)=p(V)=p(X)=1$

under-determined...



better behaved (but more complex)

Note: data fusion??? why not just use the GLM?



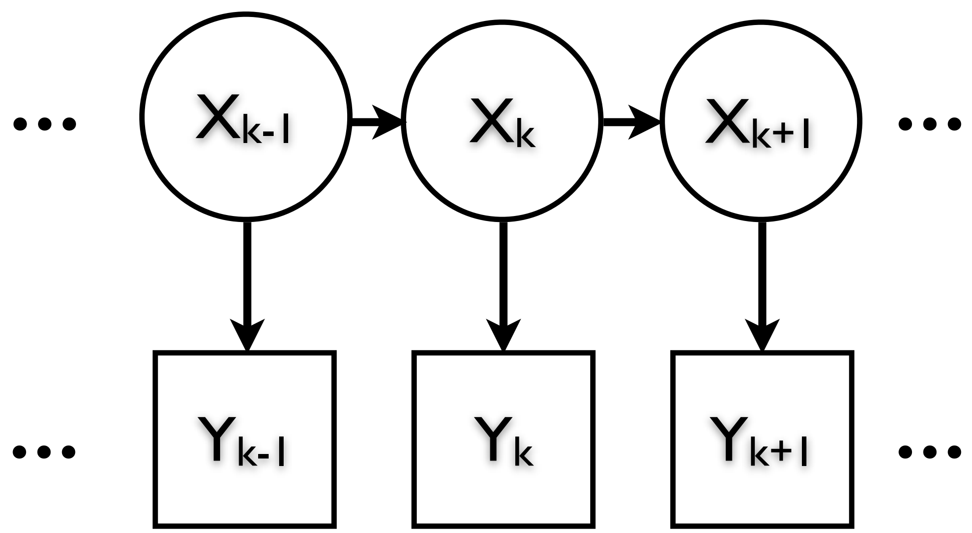
$Y_1 = aU + bV$ + noise

$Y_2 = cU + dV$ +noise

$Y_3 = eU + fV$ +noise

. can be non-linear
. estimate noise parameters
. not necessarily same data (or even same type of data)
. can be made more and more complex
...

Markov models

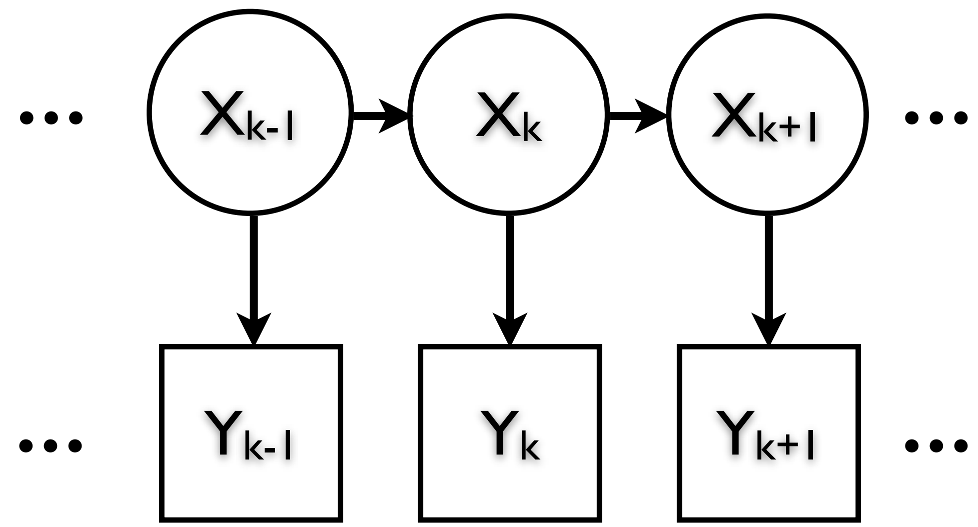


hidden states

observations

Joint (from model structure) : $p(\text{all}) = \prod p(Y_k|X_k)p(X_k|X_{k-1})$

Markov models



$p(Y_k|X_k) = N(Y_k \mid X_k, 1/b)$
$p(X_k|X_{k-1}) = N(X_k \mid X_{k-1}, 1/b_0)$

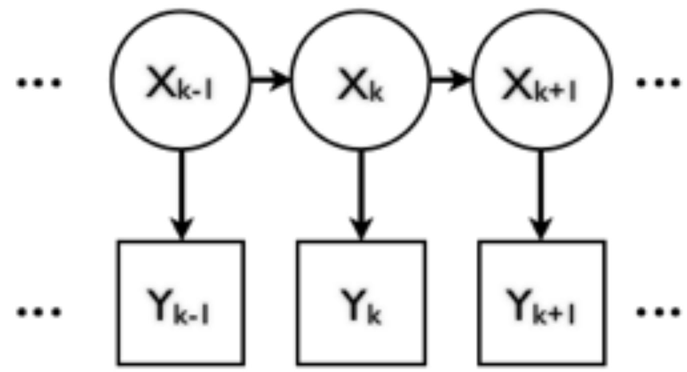
equivalently

$Y_k = X_k + \text{noise}(0,1/b)$ "observation eq"
$X_k = X_{k-1} + \text{noise}(0,1/b_0)$ "state evolution eq"

$p(X_k|Y_1,...,Y_k) = ?$

What is the posterior distribution of the states X_k given the data so far?

(1) $p(Y_k|X_k) = N(Y_k \mid X_k, 1/b)$
(2) $p(X_k|X_{k-1}) = N(X_k \mid X_{k-1}, 1/b_0)$



$$p(X_k|Y_1,\ldots,Y_{k-1},Y_k) = p(Y_k|X_k,Y_1,\ldots,Y_{k-1})\, p(X_k|Y_1,\ldots,Y_{k-1})$$ **Bayes' rule**

$$p(X_k|Y_1,\ldots,Y_{k-1},Y_k) = p(Y_k|X_k,\cancel{Y_1,\ldots,Y_{k-1}})\, p(X_k|Y_1,\ldots,Y_{k-1})$$ **Graph**

$$p(X_k|Y_1,\ldots,Y_{k-1},Y_k) = \underbrace{p(Y_k|X_k)}_{\text{update}}\, \underbrace{p(X_k|Y_1,\ldots,Y_{k-1})}_{\text{prediction}}$$

$$p(X_k|Y_1,\ldots,Y_{k-1}) = \int p(X_k|X_{k-1},Y_1,\ldots,Y_{k-1})p(X_{k-1}|Y_1,\ldots,Y_{k-1})\, dX_{k-1}$$ **Sum rule**

$$p(X_k|Y_1,\ldots,Y_{k-1}) = \int \underbrace{p(X_k|X_{k-1})}_{\text{Markov}}\underbrace{p(X_{k-1}|Y_1,\ldots,Y_{k-1})}_{\text{previous posterior}}\, dX_{k-1}$$ **Graph**

Let's pretend that $p(X_{k-1}|Y_1,\ldots,Y_{k-1}) = N(X_{k-1}, a_{k-1}, 1/b_{k-1})$

$$p(X_k|Y_1,\ldots,Y_{k-1}) = \int p(X_k|X_{k-1})p(X_{k-1}|Y_1,\ldots,Y_{k-1})\, dX_{k-1}$$

$$= \int N(X_{k-1}|X_k, 1/b_0)N(X_{k-1}|a_{k-1}, 1/b_{k-1})\, dX_{k-1}$$

$$= \int N(X_{k-1}|\ldots,\ldots)N(X_k|a_{k-1}, 1/b_0+1/b_k)\, dX_{k-1}$$

$$= N(X_k|a_{k-1}, \underbrace{1/b_{0}+1/b_{k-1}}_{1/\beta_{k-1}})$$

In other words: prediction has mean a_{k-1} (previous state) and variance = $1/b_{k-1}+1/b_0$

$$p(X_k|Y_1,\ldots,Y_{k-1},Y_k) = p(Y_k|X_k)\, p(X_k|Y_1,\ldots,Y_{k-1})$$

$$= N(X_k|Y_k|1/b)\, N(X_k|a_{k-1}, 1/\beta_{k-1})$$

$$= N(X_k|\underbrace{(bY_k + \beta_{k-1}a_{k-1})/(b+\beta_{k-1})}_{a_k}, \underbrace{1/(b+\beta_{k-1})}_{1/b_k})$$

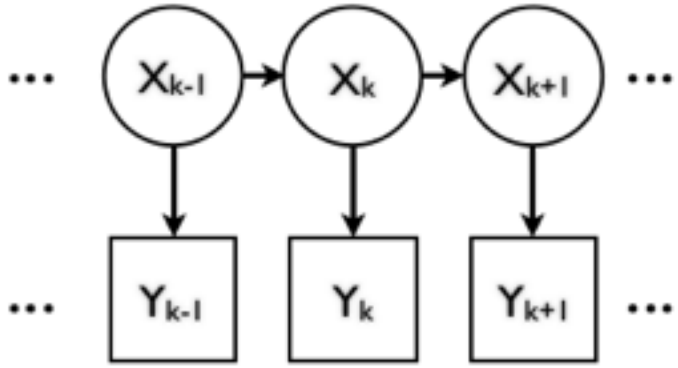
$$p(X_k|Y_1,\dots,Y_{k-1},Y_k) = p(Y_k|X_k)\, p(X_k|Y_1,\dots,Y_{k-1})$$

$$= N(X_k|\underbrace{(bY_k + \beta_{k-1}a_{k-1})/(b+\beta_{k-1})}_{a_k},\ \underbrace{1/(b+\beta_{k-1})}_{1/b_k})$$

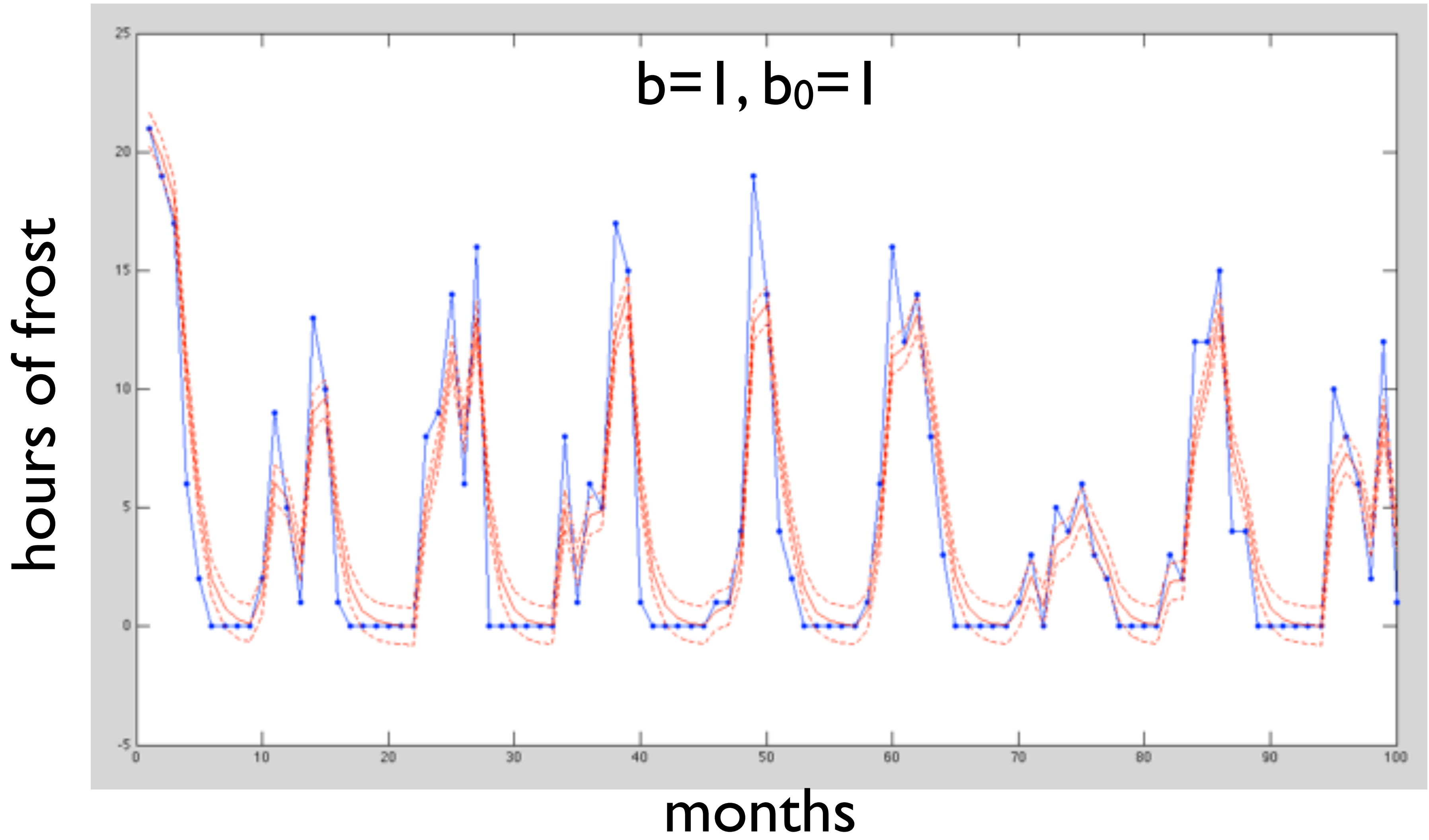
$a_k = a_{k-1} + b/b_k(Y_k - a_{k-1})$ update: new mean is previous mean + weighted error term.

$b_k = b + b_0 b_{k-1}/(b_0 + b_{k-1})$

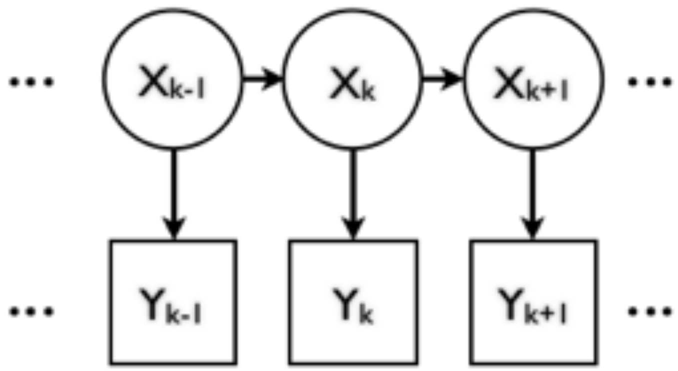
(1) $p(Y_k|X_k) = N(Y_k \mid X_k, 1/b)$
(2) $p(X_k|X_{k-1}) = N(X_k \mid X_{k-1}, 1/b_0)$



weather data - Oxfordshire (1929-1949)



(1) $p(Y_k|X_k) = N(Y_k \mid X_k, 1/b)$
(2) $p(X_k|X_{k-1}) = N(X_k \mid X_{k-1}, 1/b_0)$

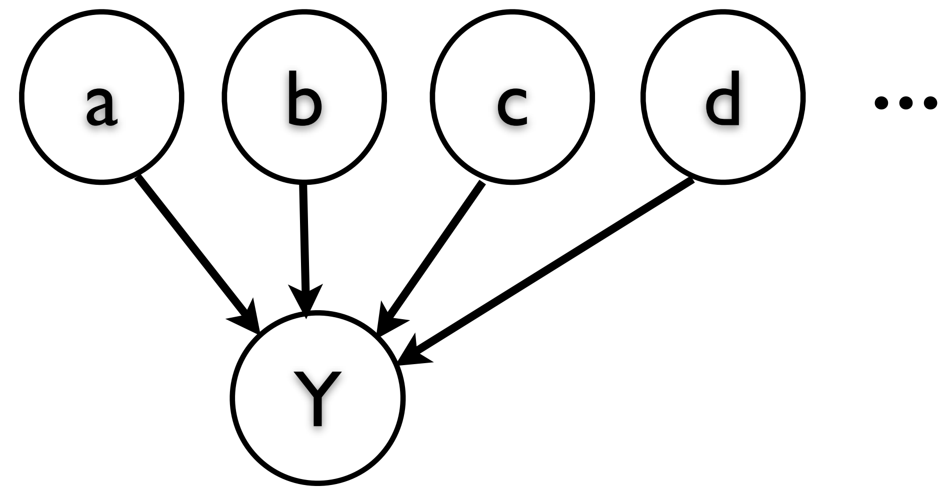


Ok - lots of integrals, what have we learnt?

- Graphs => conditional posteriors
- Exact inference requires integration (Gaussians are helpful)

Metropolis Hastings code

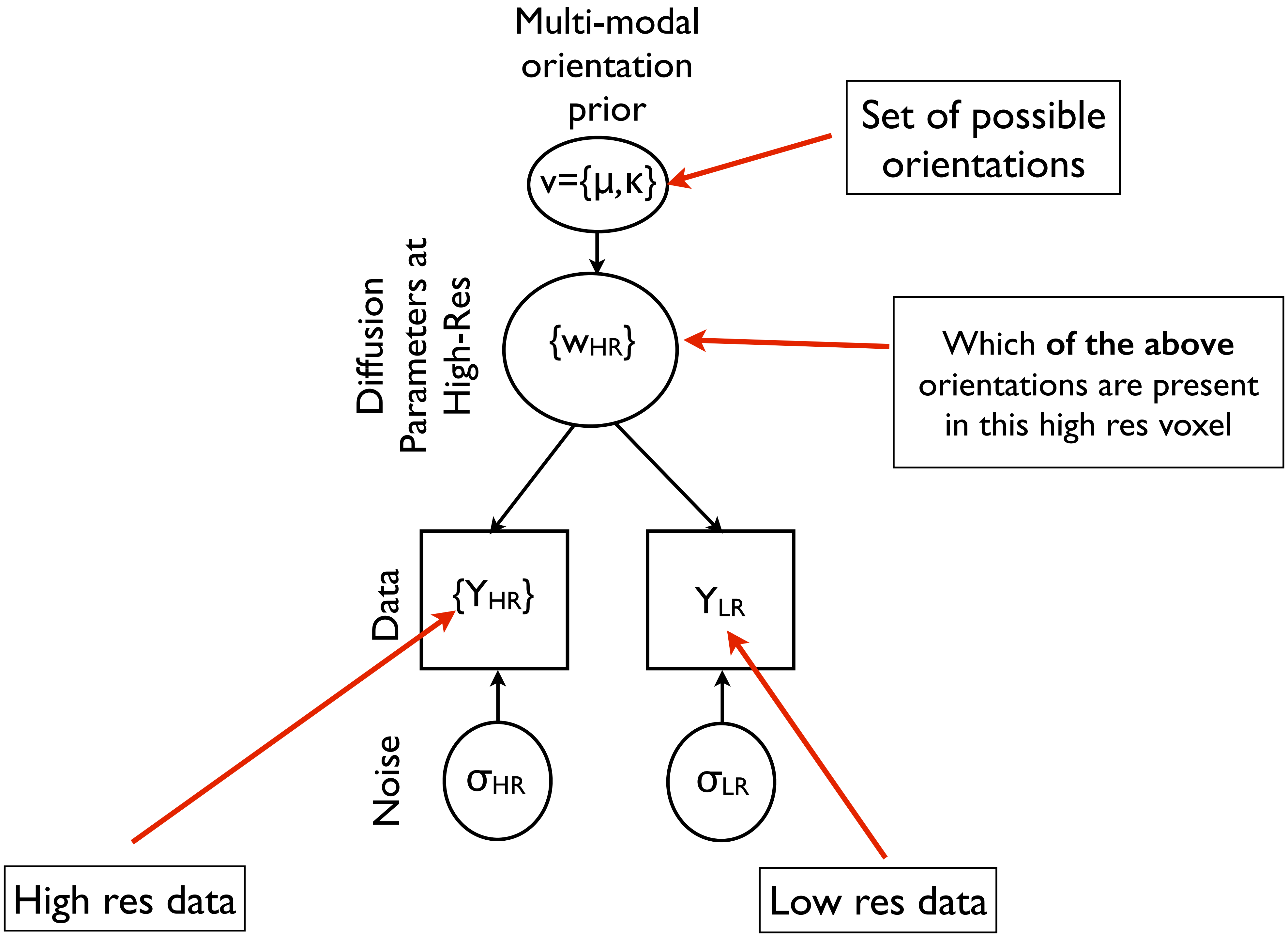
The mh.m function assumes the following model:



You don't have to restrict yourself to this model! All you need in MH is the joint distribution.

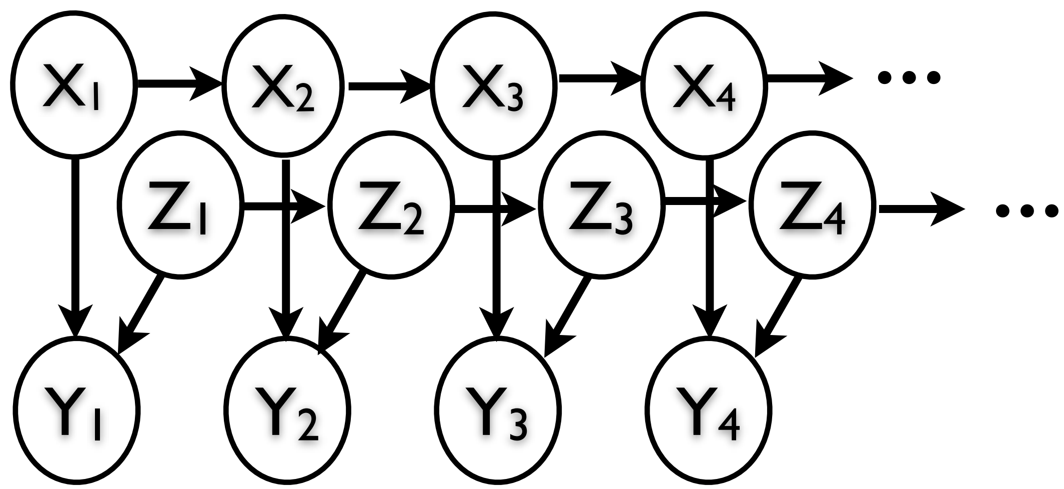
Other examples

RUBIX generative model

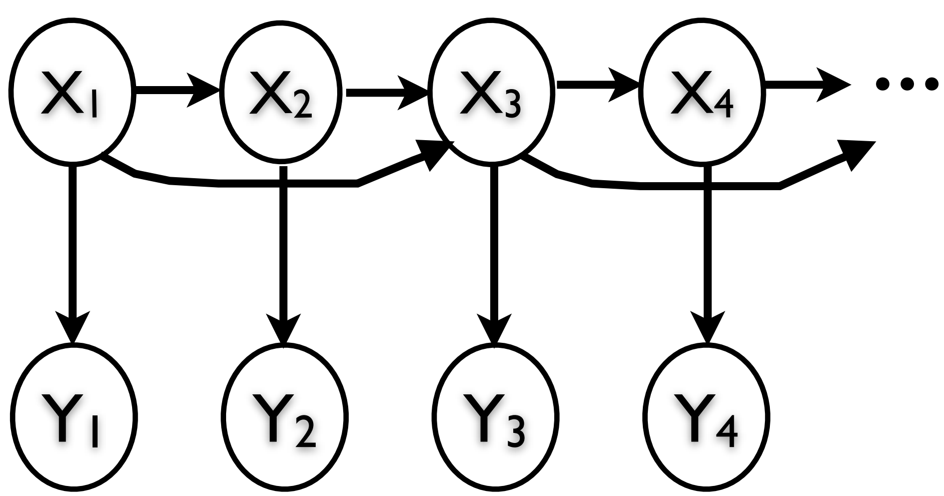


Sotiropoulos, Jbabdi, Behrens et al. IEEE TMI 2013

Other examples
Markov Models

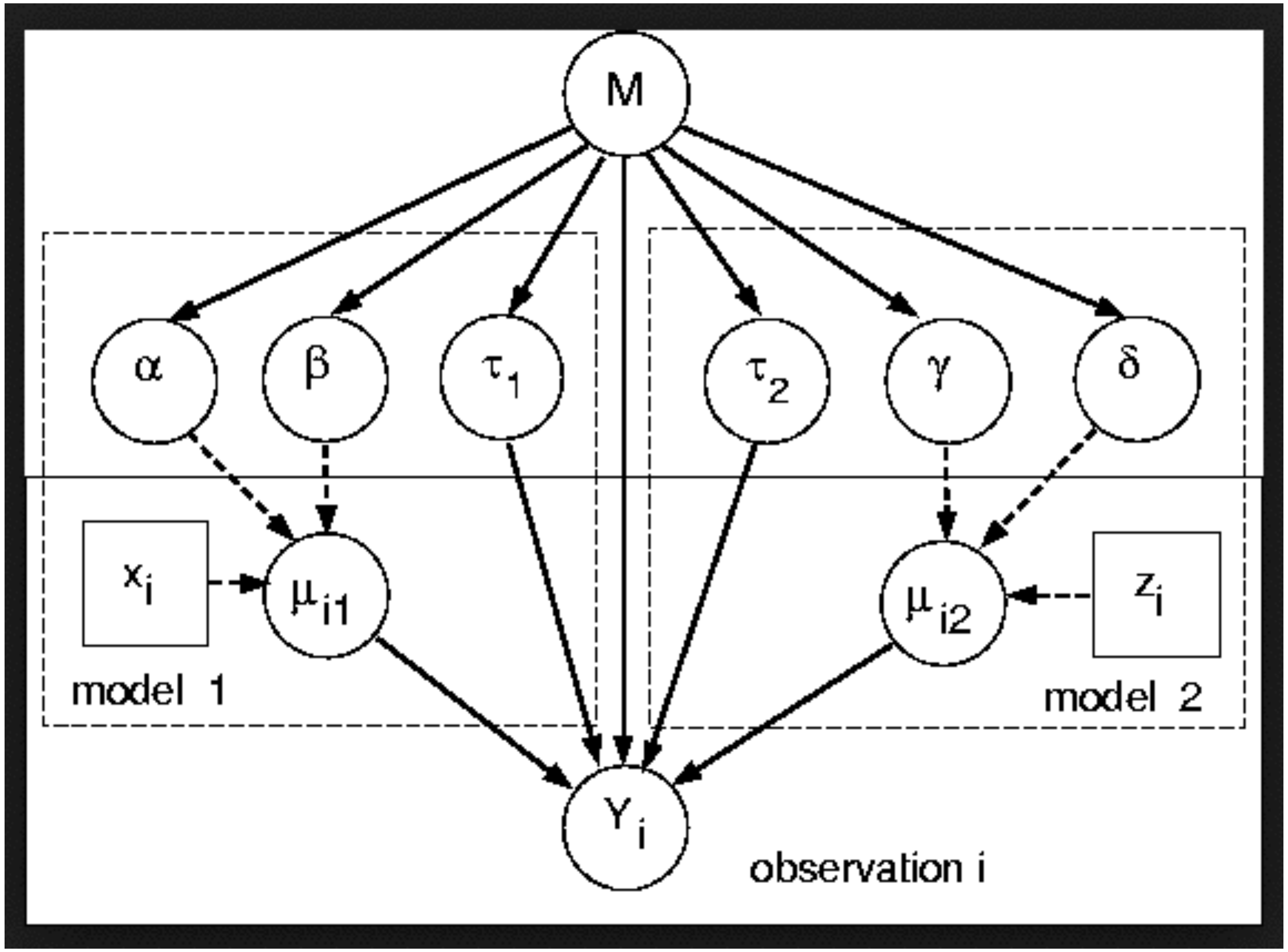


Factorial HMM



Memory?

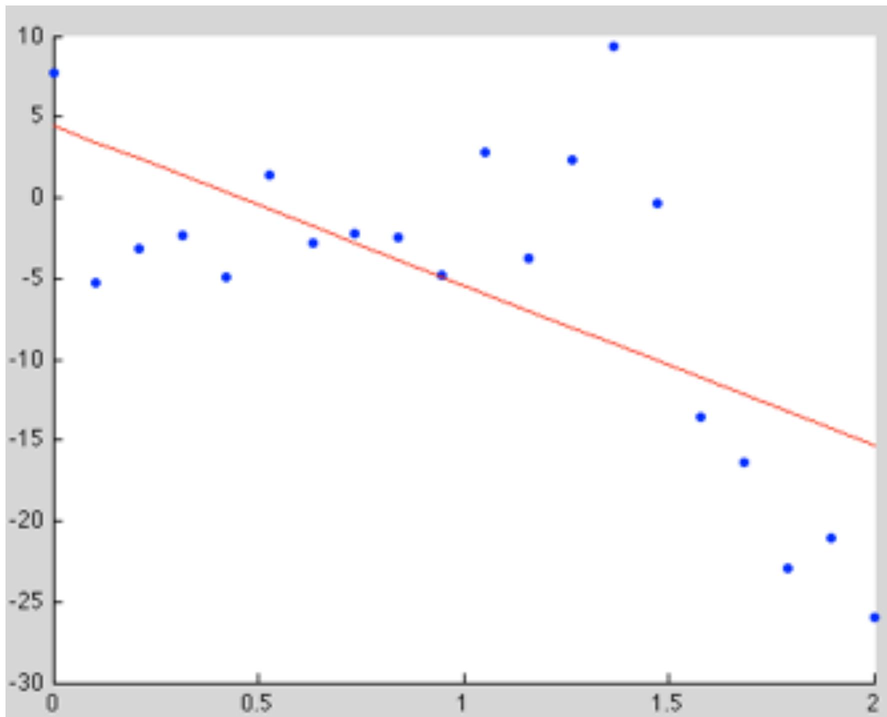
Other examples
Mixtures of experts

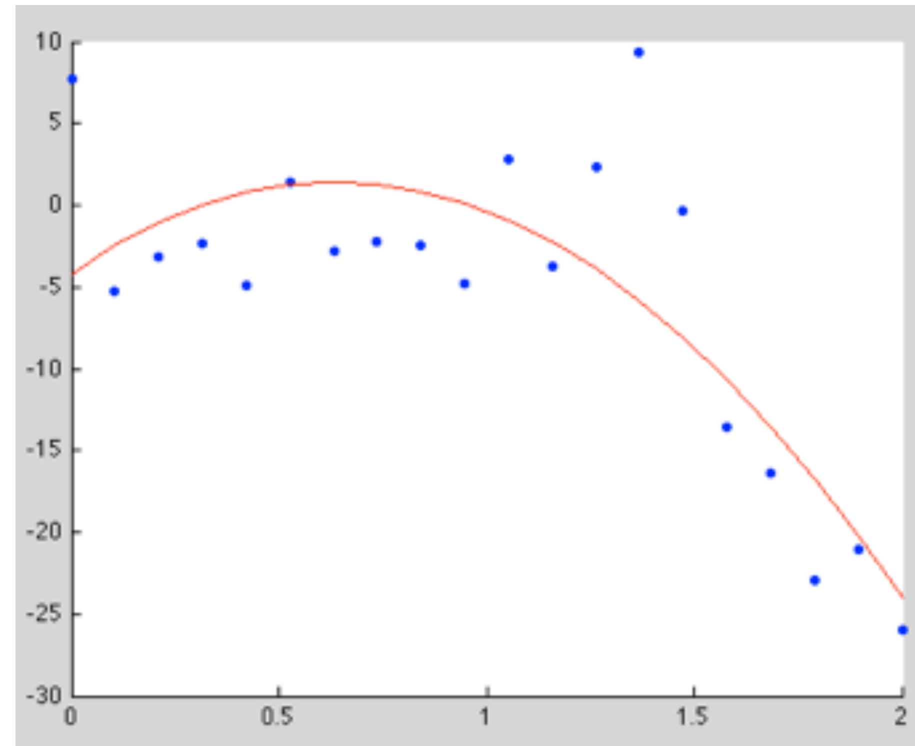


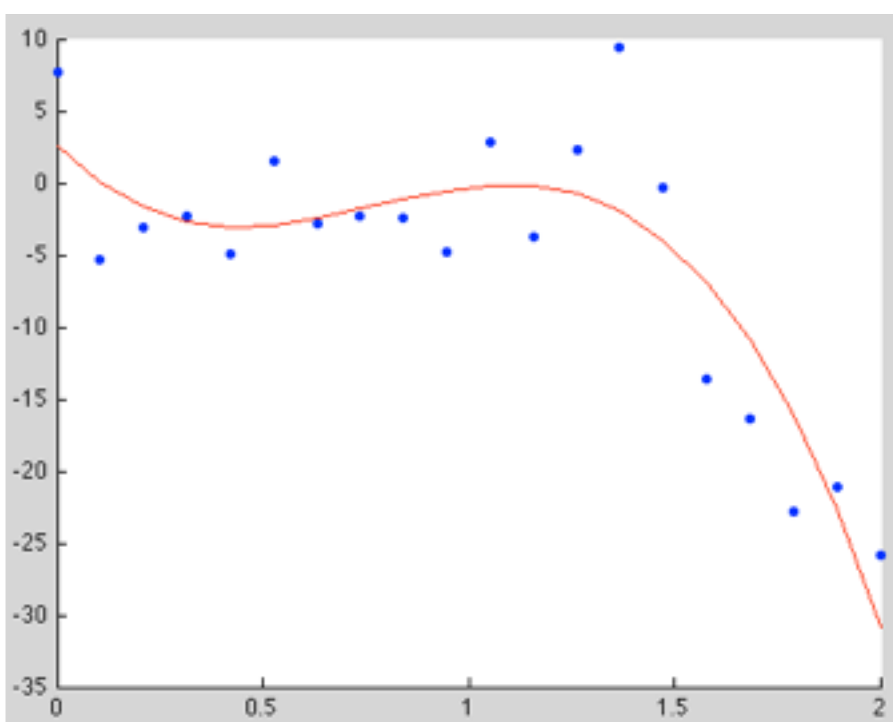
why graphical models?

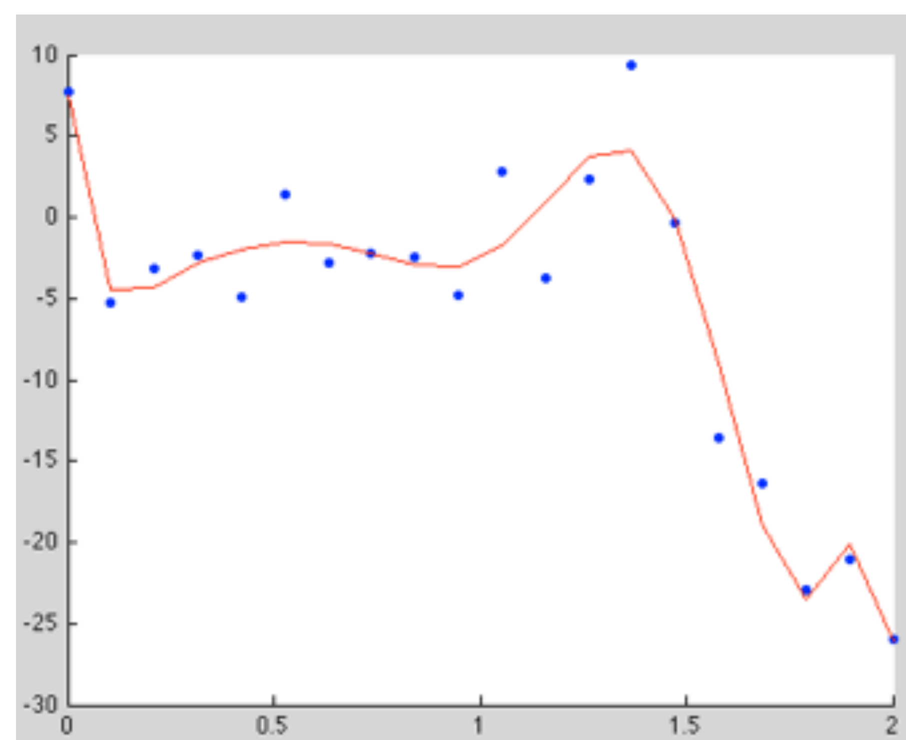
- visualise model structure
- understand how information propagates through
- handle complex models

Model selection









Model selection

$$p(a|y)=p(y|a)p(a)/p(y)$$

$$p(a|y,M)=p(y|a,M)p(a|M)/p(y|M)$$

$$p(y|M)=\int p(y|a,M)p(a|M)da$$

(model evidence)
(marginal likelihood)

$$\frac{p(M_1|y)}{p(M_2|y)} = \frac{p(M_1)}{p(M_2)} \frac{p(y|M_1)}{p(y|M_2)}$$

Bayes factor

Model selection

Gaussian example

$p(y|a)=N(y|a,1/b)$ likelihood

$p(a)=N(a|a_0,1/b_0)$ prior

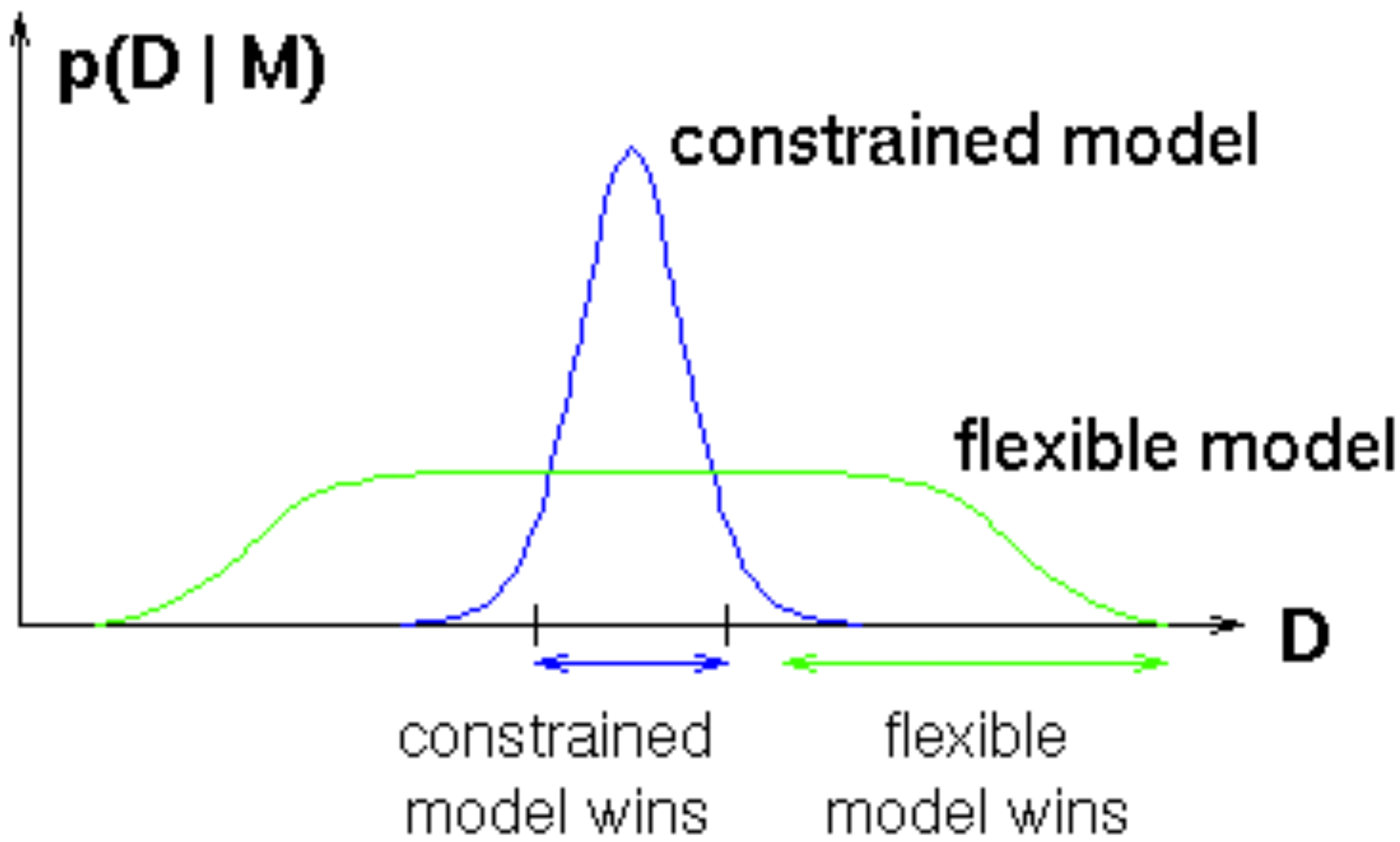
$$p(y|a)*p(a)=N(y|a,1/b)*N(a|a_0,1/b_0)$$
$$=N(a|...,...)*N(y|a_0,1/b+1/b_0)$$

$$p(y)=\int p(y|a)*p(a)da=N(y|a_0,1/b+1/b_0)$$

$$\log\{p(y)\}=\log\{bb_0/(b+b_0) - 0.5(bb_0/(b+b_0))(y-a_0)^2$$

"complexity" "data fit"

Model selection



That’s all folks

(don’t forget to look at the matlab examples: **data_fusion.m** and **kalman_filter.m**)