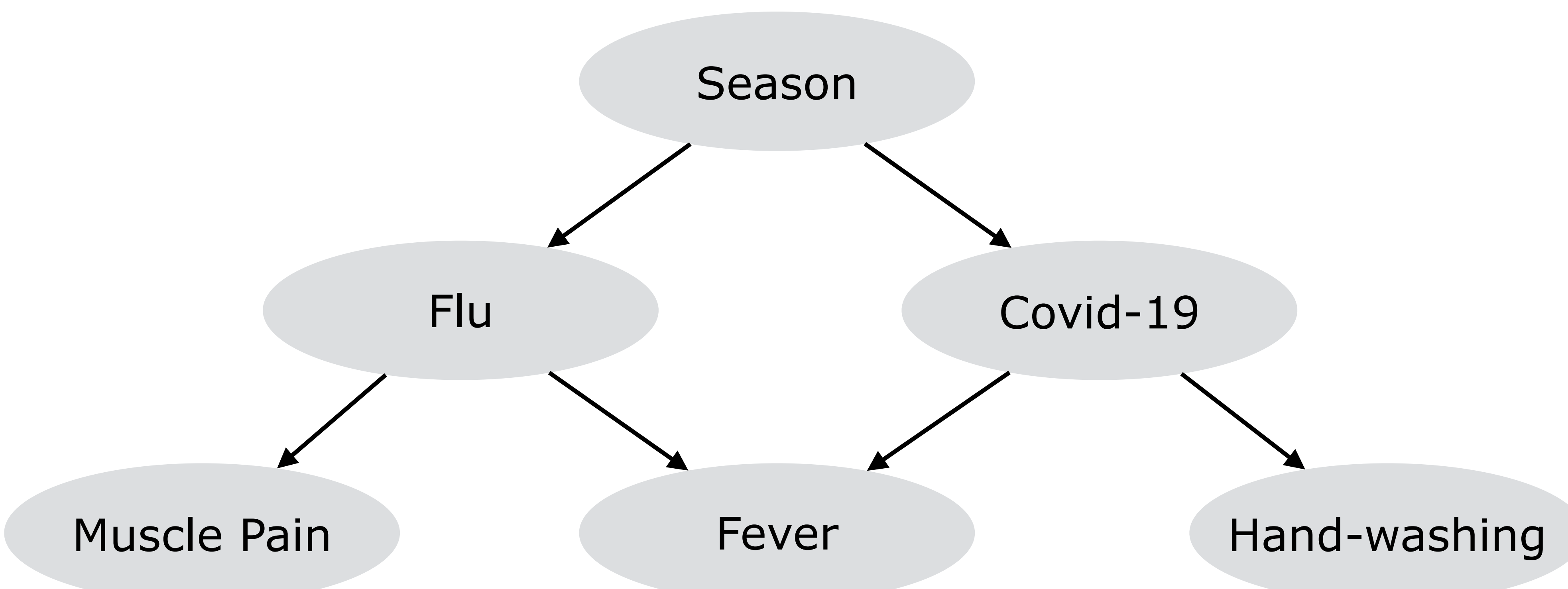
Graphical Models

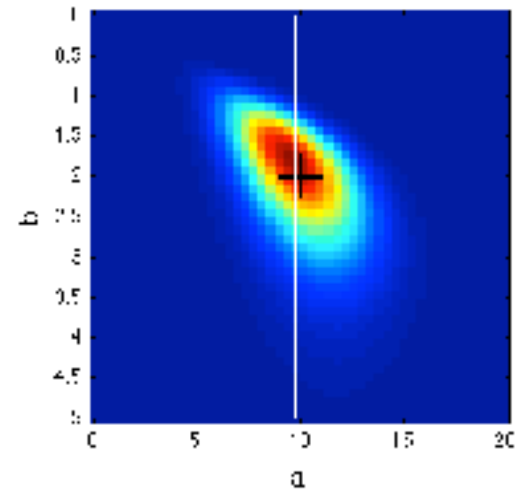


Outline

- Joint/Conditional/Marginal
- Bayesian graphical models
- Designing a model
- Bayes in FSL
- Derivation
  - Data fusion
  - Kalman Filter

p(A, B)

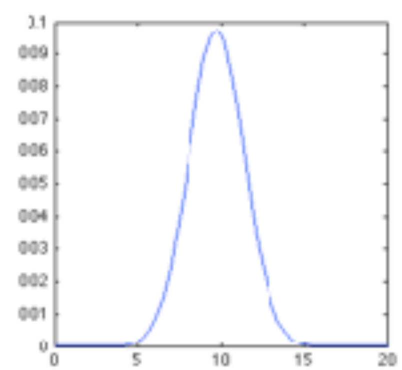
Joint distribution



$\int p(A,B)dAdB=1$

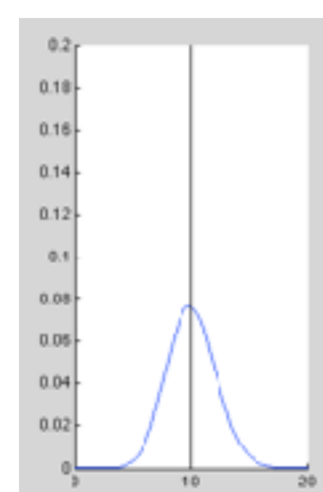
Conditioning (slice)

p(A | B)



Marginalisation (sum)

p(A, ~~B~~)



$p(A)=\int p(A,B)dB$

- Relationship between **joint, conditional, and marginal** probability

$$p(A, B) = p(A \mid B) * p(B)$$

product rule

joint conditional marginal

Bayes rule

$$p(A, B) = p(A \mid B) * p(B)$$ product rule

=

$$p(B, A) = p(B \mid A) * p(A)$$ product rule

↓

$$p(A \mid B) = p(B \mid A) * p(A) / p(B)$$

Bayesian modelling and inference

y : data
a : parameters of a model $y=F(a)$

Bayesian modelling and inference

y : data

a : parameters of a model y=F(a)

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

Bayesian modelling and inference

y : data

a : parameters of a model y=F(a)

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

$$\underbrace{p(y | a) * p(a)}_{\text{model}}$$

Bayesian modelling and inference

y : data

a : parameters of a model y=F(a)

$$\underbrace{p(a|y)}_{\text{infer}} = \underbrace{p(y|a) * p(a)}_{\text{model}} / p(y)$$

Outline

- Joint/Conditional/Marginal
- Bayesian graphical models
- Designing a model
- Bayes in FSL
- Derivation
  - Data fusion
  - Kalman Filter

Graphical models

Graphical models

p(a,b) start with a joint distribution

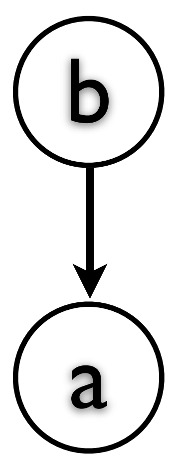
Graphical models

$p(a,b)$	start with a joint distribution
$p(a,b)=p(a\|b)p(b)$	decompose it using the product rule

Graphical models

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

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

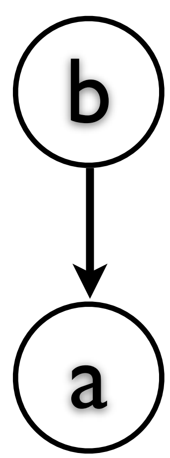


draw it using arrows instead of |

Graphical models

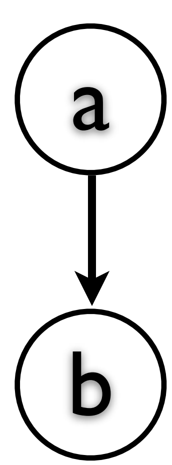
$p(a,b)$ start with a joint distribution

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

Graphical models

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

let’s try a longer joint distribution

Graphical models

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

let's try a longer joint distribution

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

I can decompose it like this

Graphical models

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

let's try a longer joint distribution

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

I can decompose it like this

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

Keep going

Graphical models

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

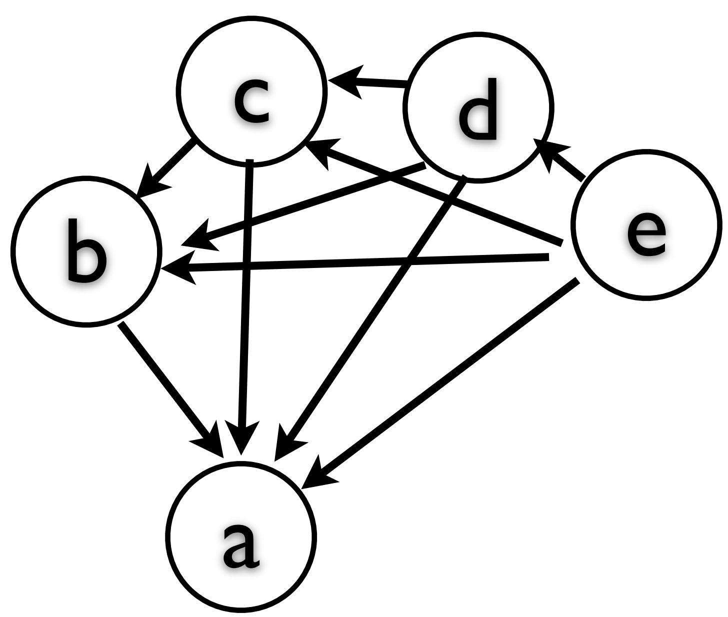
let's try a longer joint distribution

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

I can decompose it like this

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

Keep going



draw it using arrows instead of |

Graphical models

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

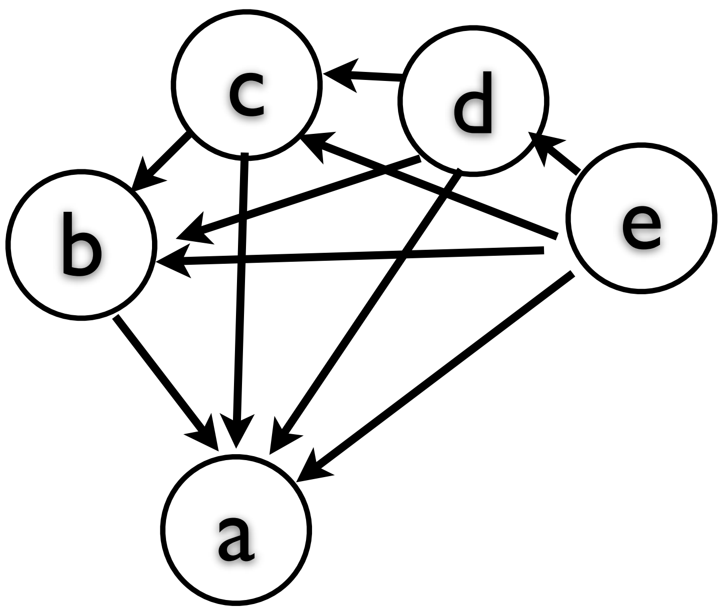
let's try a longer joint distribution

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

I can decompose it like this

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

Keep going



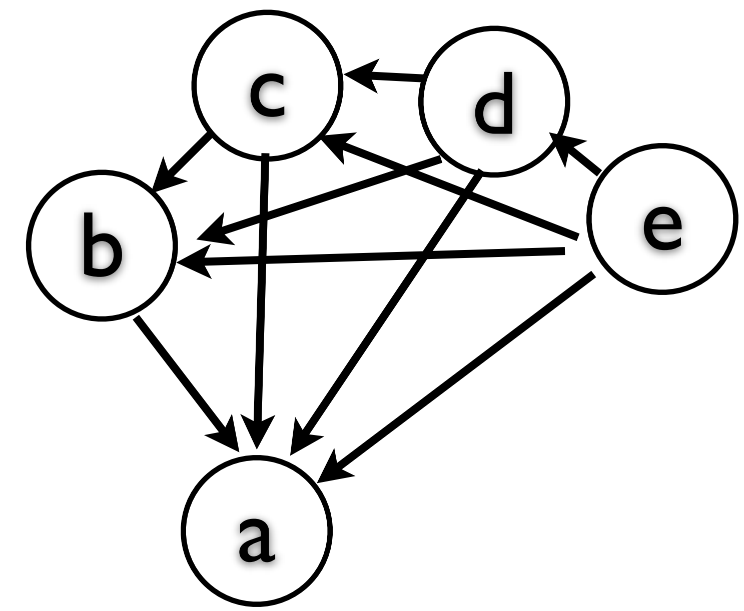
draw it using arrows instead of |

terminology

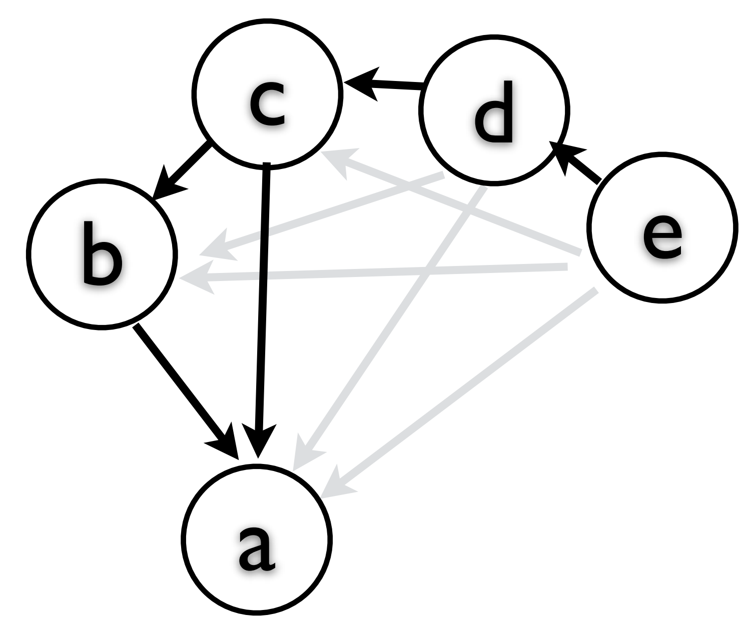
Directed acyclic graphs
Bayesian networks
Directed graphical models
Belief networks

Graphical models

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



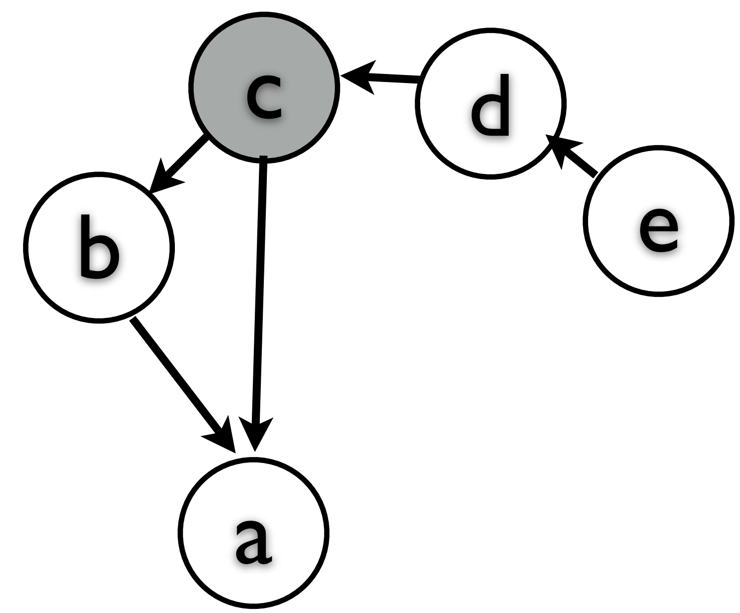
$$p(a,b,c,d,e)=p(a|b,c)p(b|c)p(c|d)p(d|e)p(e)$$



removing dependences
= creating causal
structure

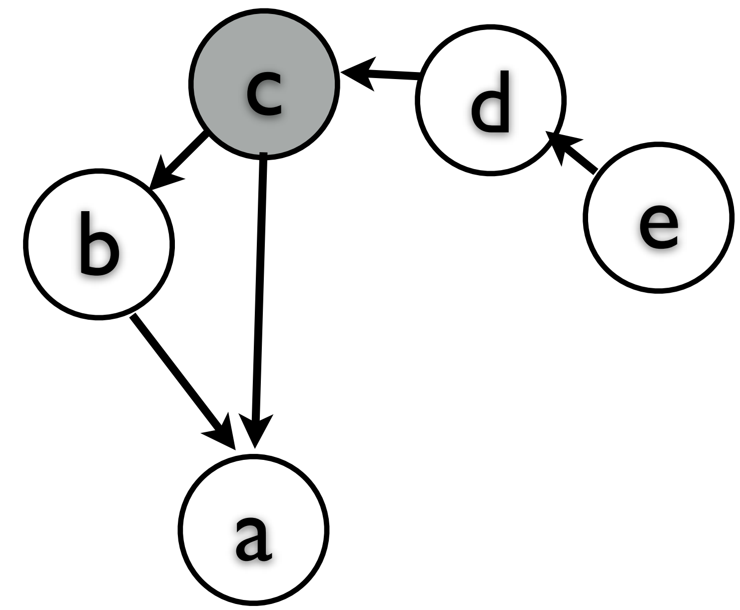
Graphical models

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



Graphical models

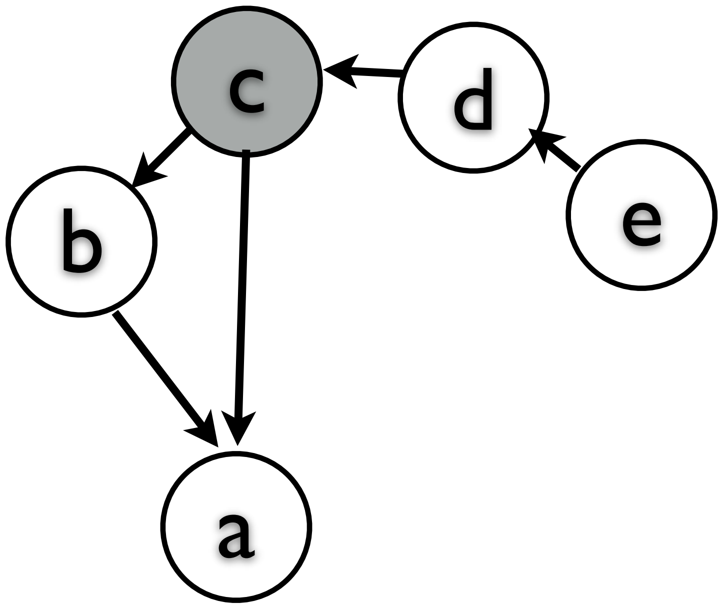
$p(a,b,c,d,e)=p(a|b,c)p(b|c)p(c|d)p(d|e)p(e)$



Observing one node

Graphical models

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

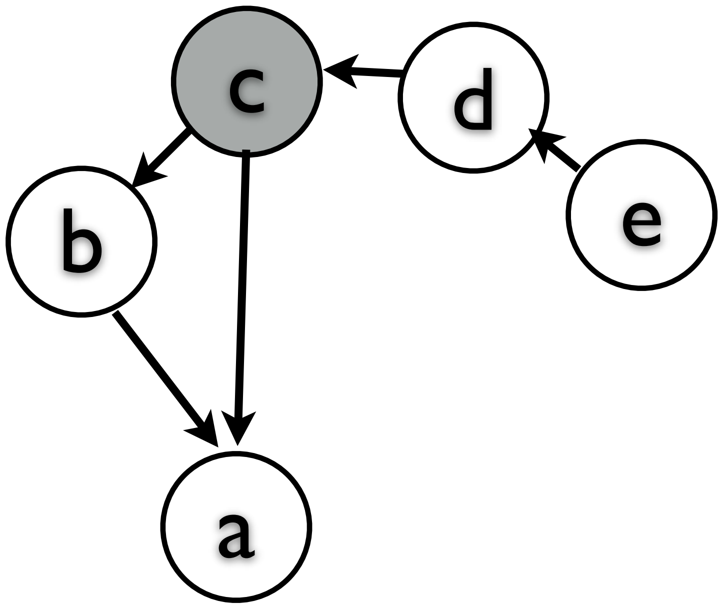


Observing one node

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

Graphical models

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

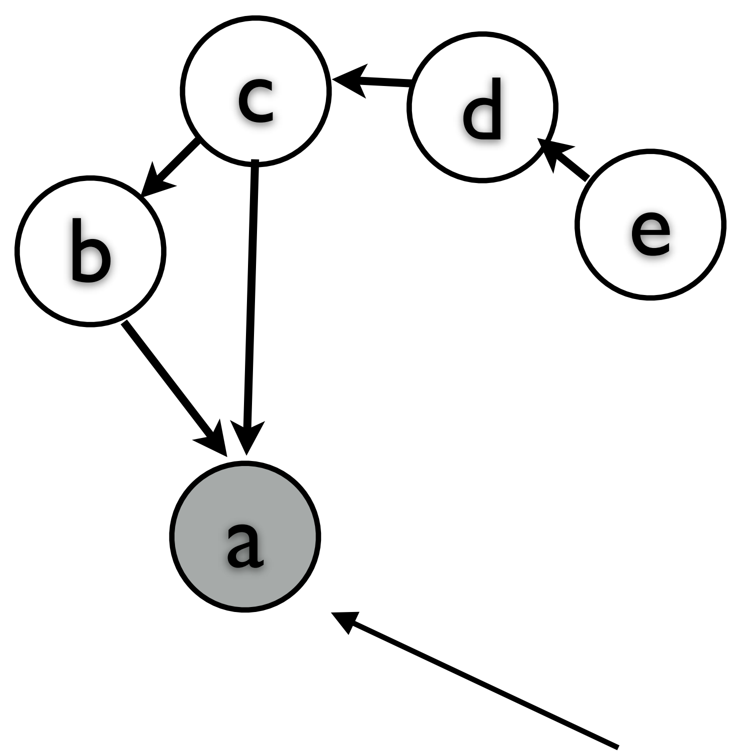


Observing one node

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

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

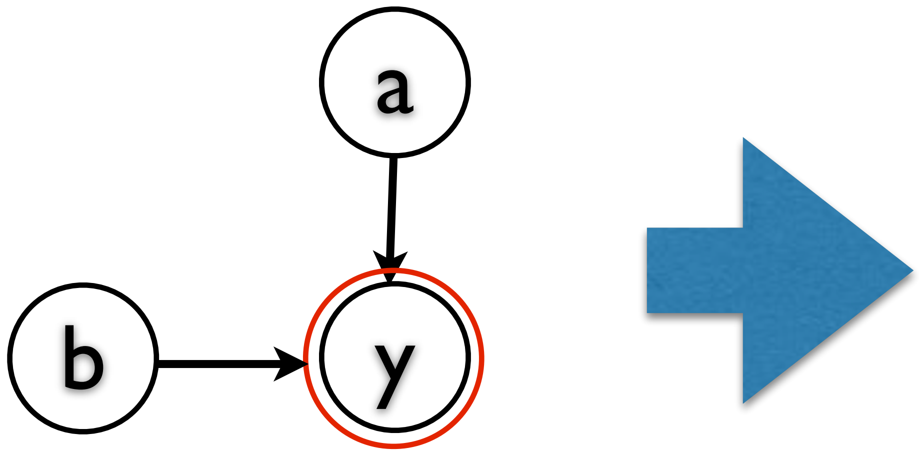
Generative models



where we observe the leaves

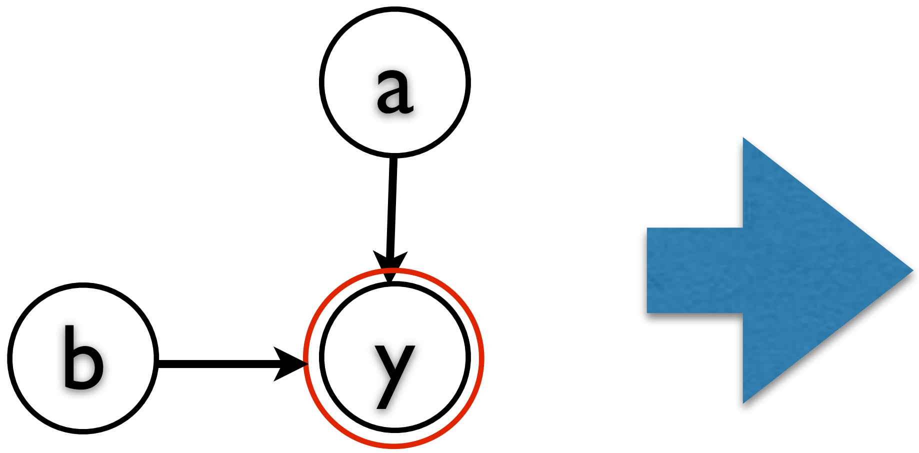
(and try to make inference on the rest)

An abstract example



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

An abstract example

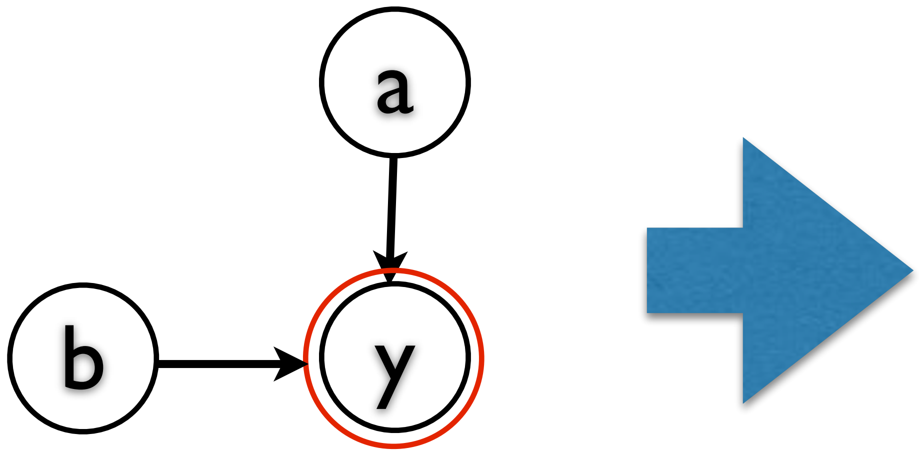


$p(y,a,b) = p(y|a,b)p(a)p(b)$

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

(product rule)

An abstract example



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

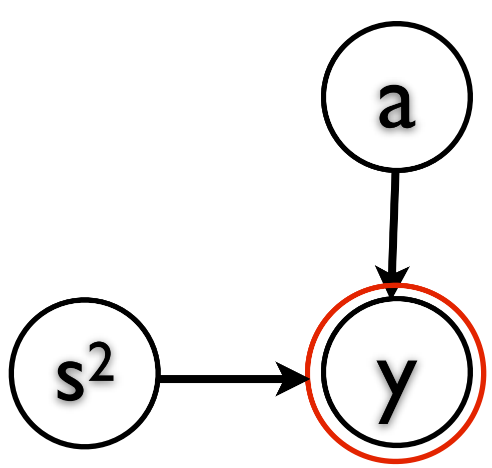
$p(a,b|y) = p(y,a,b)/p(y)$ (product rule)

$p(a,b|y) = p(y|a,b)p(a)p(b)/p(y)$ (from the graph)

A concrete example

Generative model

y = a + noise

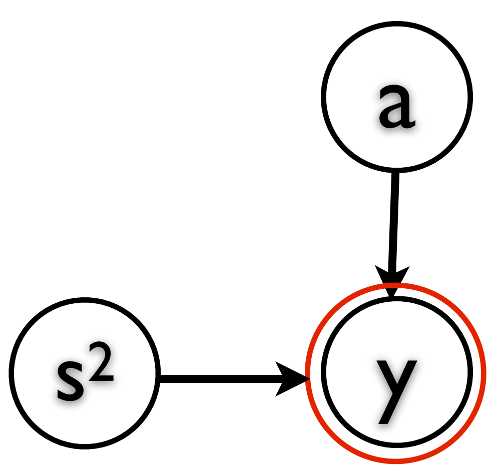


A concrete example

Generative model

y = a + noise

noise ~ Normal(0,s^2)



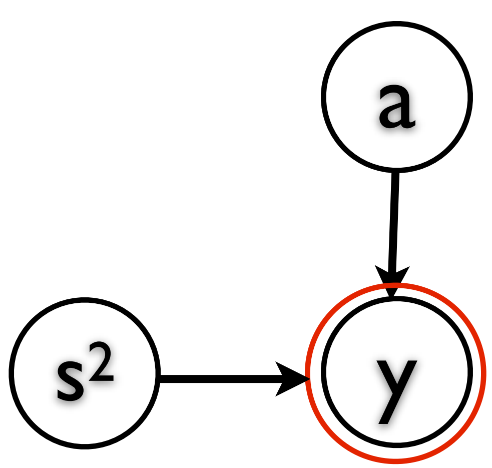
A concrete example

Generative model

y = a + noise

noise ~ Normal(0,s^2)

→ y ~ Normal(a,s^2)



A concrete example

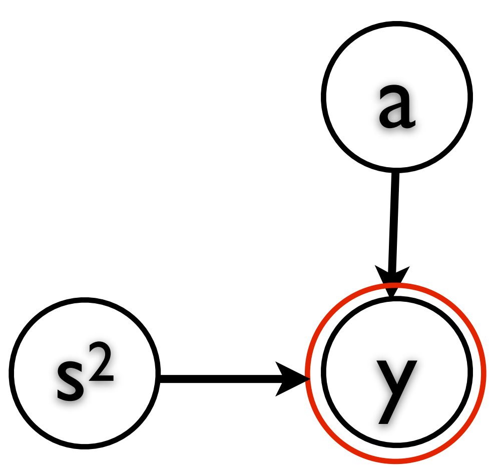
Generative model

$y = a + \text{noise}$

$\text{noise} \sim \text{Normal}(0,s^2)$

⟶ $y \sim \text{Normal}(a,s^2)$

prior for a = Uniform(0,1)



A concrete example

Generative model

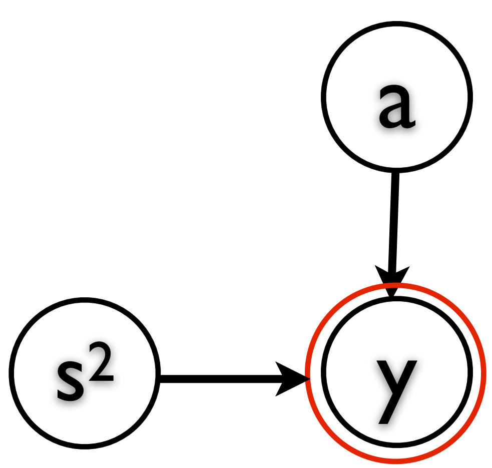
$y = a + \text{noise}$

$\text{noise} \sim \text{Normal}(0, s^2)$

→ $y \sim \text{Normal}(a, s^2)$

prior for a = Uniform(0,1)

prior for s^2 = Gamma(1,1)



A concrete example

Generative model

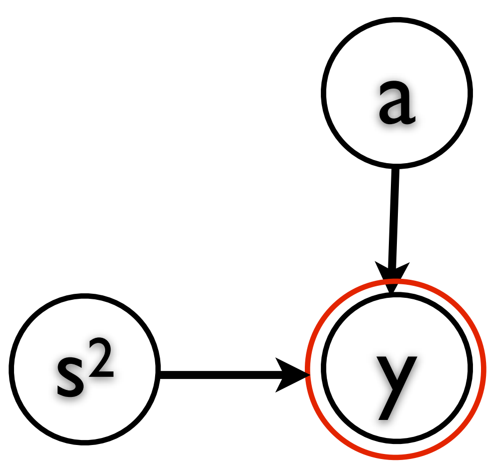
y = a + noise

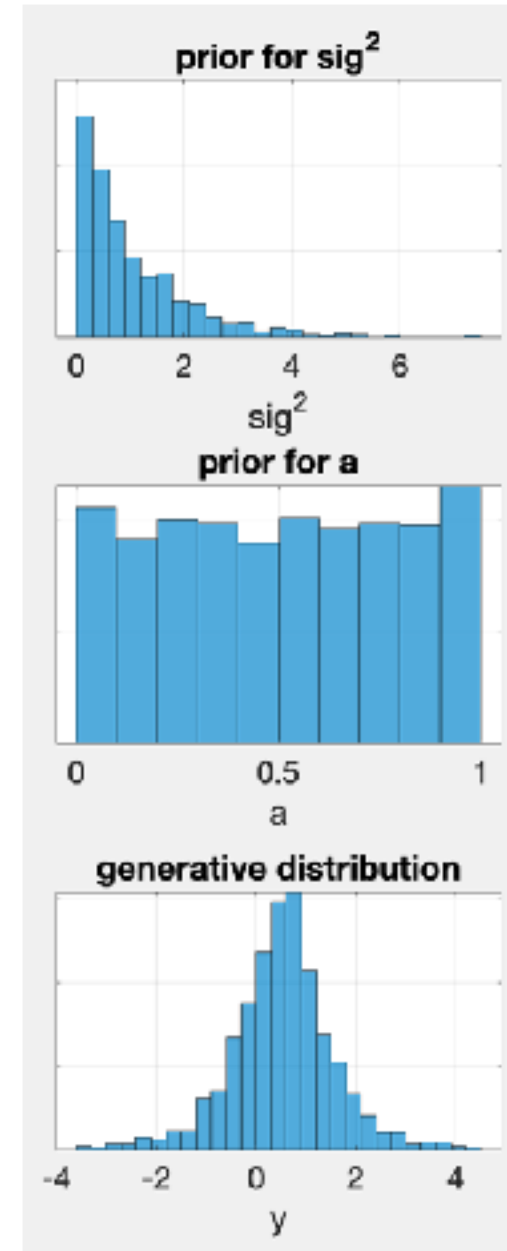
noise ~ Normal($0,s^2$)

→ y ~ Normal(a,s^2)

prior for a = Uniform(0,1)

prior for s^2 = Gamma(1,1)

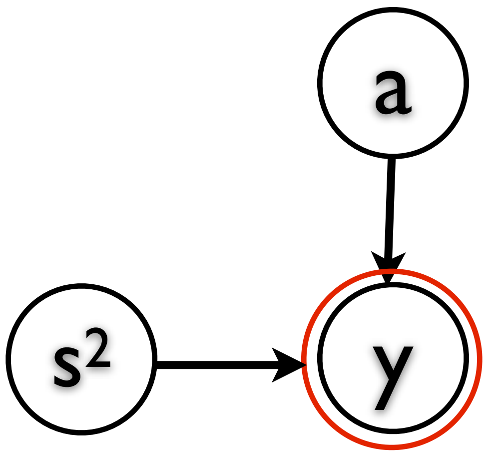





```
for i=1:1000
    sig2(i) = random('gamma',1,1);
    a(i)    = random('unif',0,1);
    y(i)    = random('normal',a(i),sqrt(sig2(i)));
end
```

A concrete example

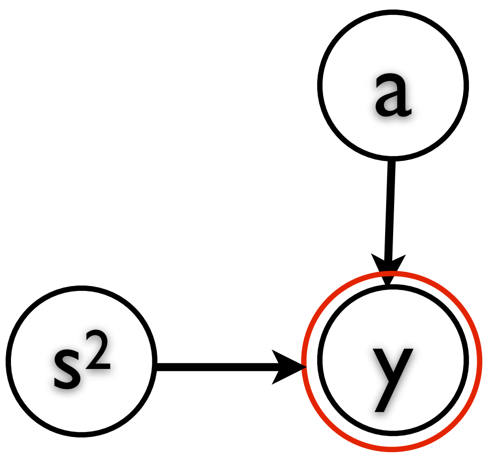
Inference



$$p(a,s^2|y) = p(y|a,s^2)p(a)p(s^2)/p(y)$$

A concrete example

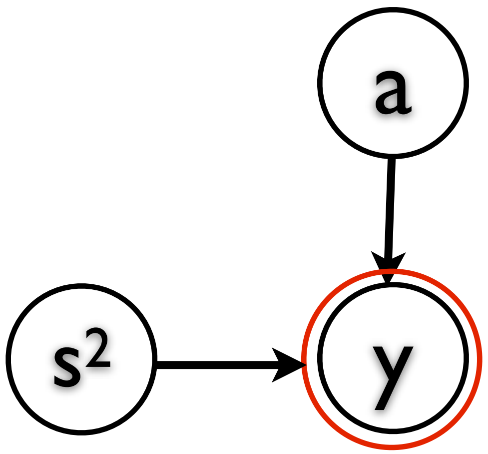
Inference



$$p(a,s^2|y) = p(y|a,s^2)p(a)p(s^2)/p(y)$$

A concrete example

Inference

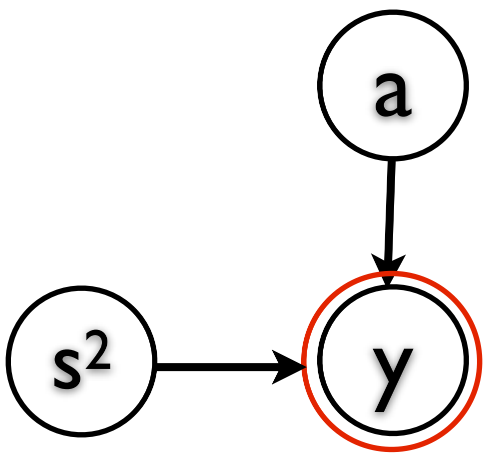


$$p(a,s^2|y) = p(y|a,s^2)p(a)p(s^2)/p(y)$$

$$\text{Normal}(y|a,s^2)^* \text{ Uniform}(0,1)^*\text{Gamma}(1,1)^*\text{cst}$$

A concrete example

Inference



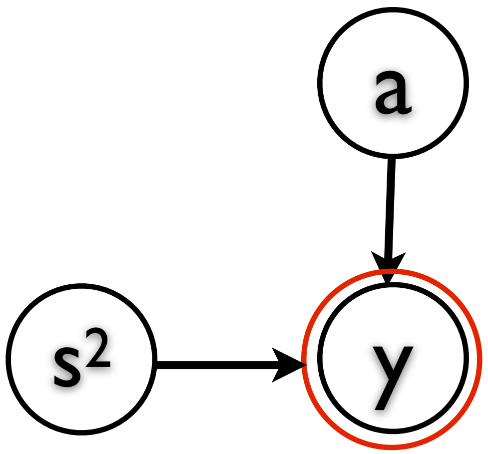
$$p(a,s^2|y) = p(y|a,s^2)p(a)p(s^2)/p(y)$$

$$\text{Normal}(y|a,s^2) * \text{Uniform}(0,1) * \text{Gamma}(1,1) * \text{cst}$$

$$\sqrt{\frac{1}{2\pi s^2}} \exp\left[-\frac{(x-a)^2}{2s^2}\right]$$

A concrete example

Inference



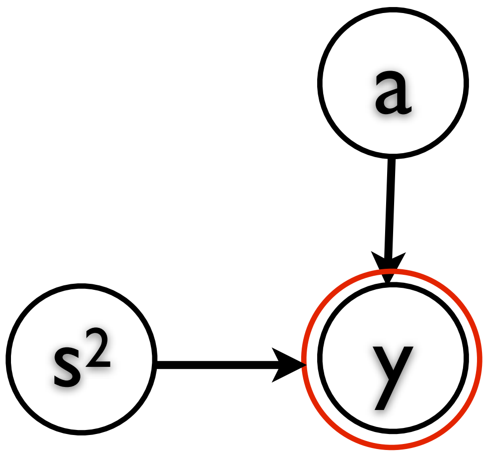
$$p(a,s^2|y) = p(y|a,s^2)p(a)p(s^2)/p(y)$$

$$\text{Normal}(y|a,s^2)^* \ \text{Uniform}(0,1)^*\text{Gamma}(1,1)^*\text{cst}$$

$$\sqrt{\frac{1}{2\pi s^2}} \exp\left[-\frac{(x-a)^2}{2s^2}\right] \qquad \exp\left(-s^2\right)$$

A concrete example

Inference



$$p(a,s^2|y) = p(y|a,s^2)p(a)p(s^2)/p(y)$$

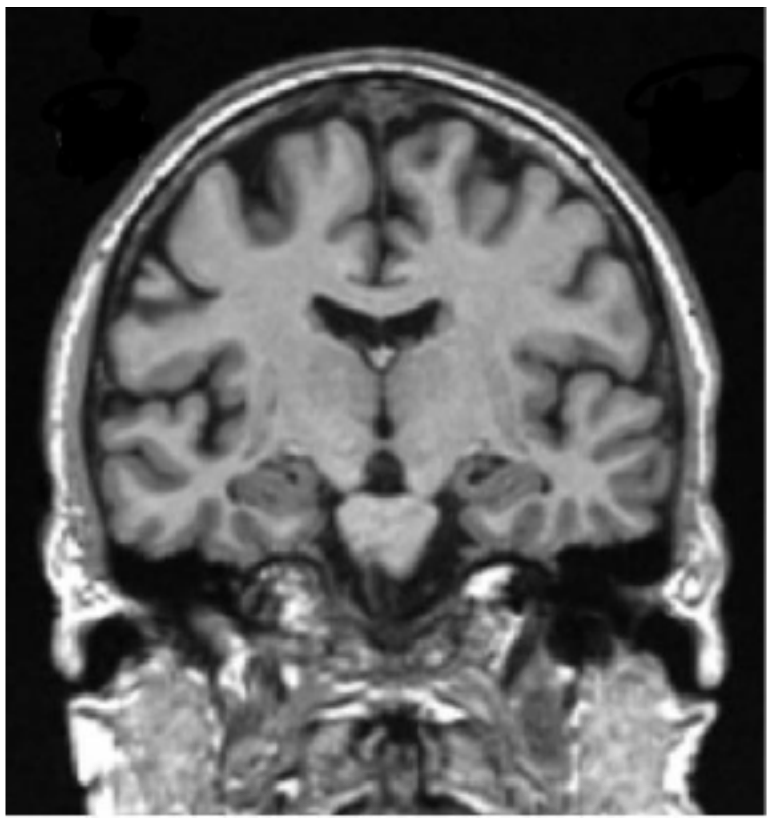
Normal(y|a,s²)* Uniform(0,1)*Gamma(1,1)*cst

$$\sqrt{\frac{1}{2\pi s^2}}\exp\left[-\frac{(x-a)^2}{2s^2}\right] \qquad 1 \qquad \exp\left(-s^2\right)$$

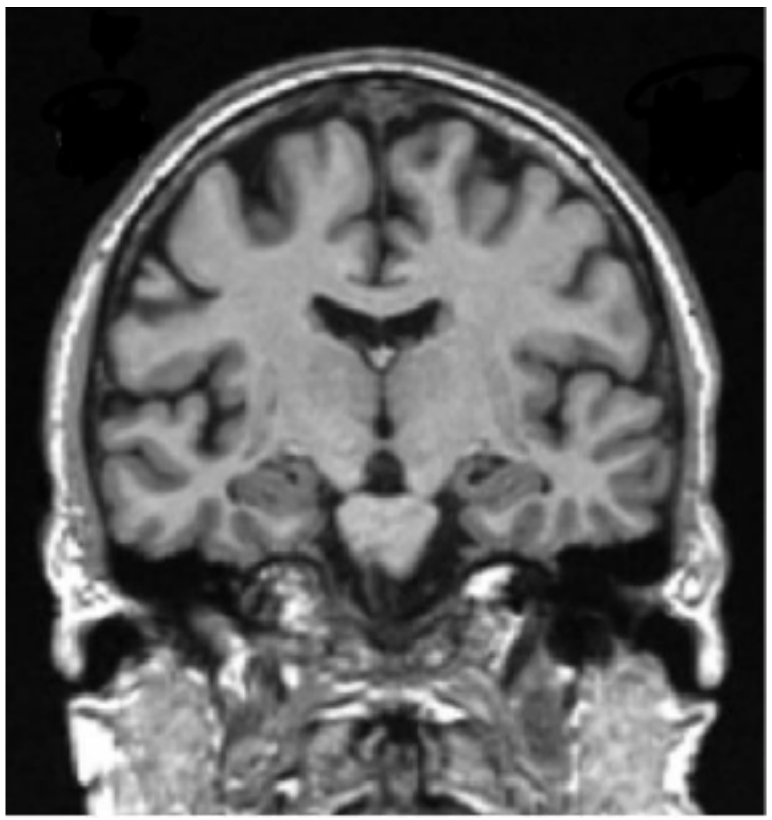
Outline

- Joint/Conditional/Marginal
- Bayesian graphical models
- Designing a model
- Bayes in FSL
- Derivation
  - Data fusion
  - Kalman Filter

Generative model for a T1w image

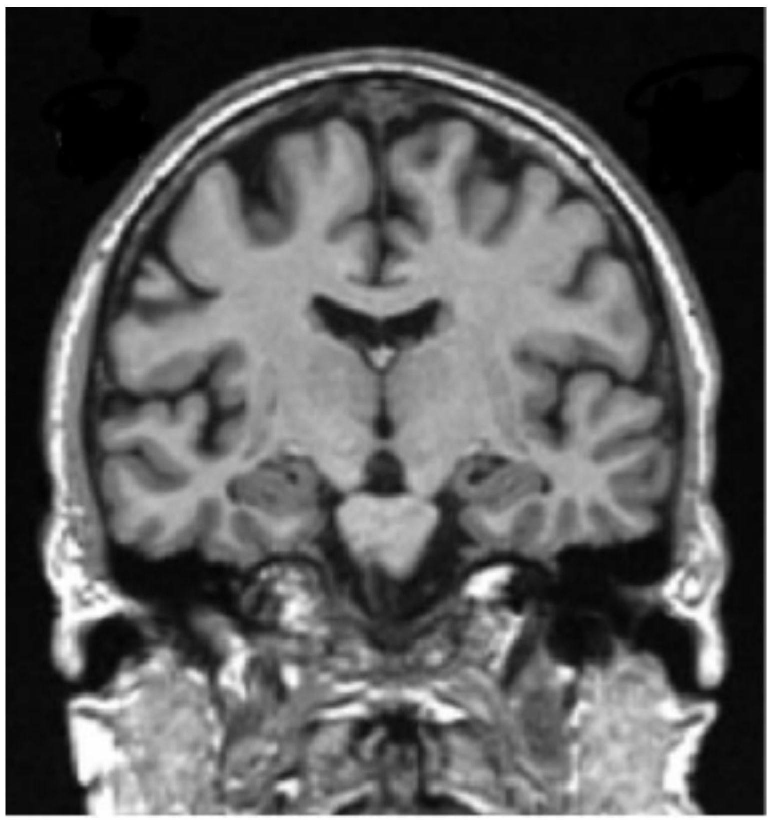


Generative model for a T1w image



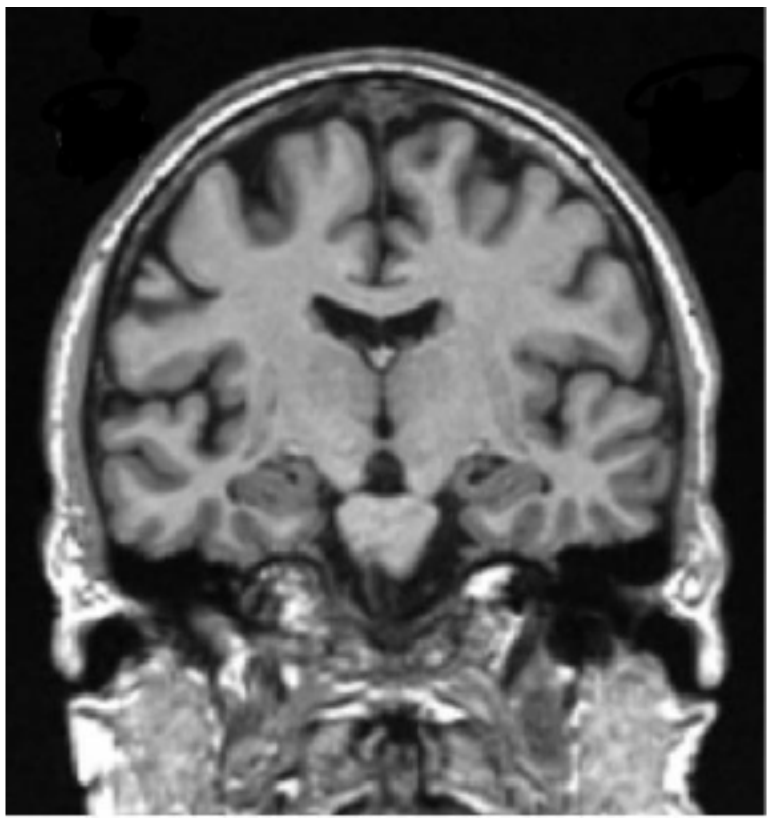
i = voxel

Generative model for a T1w image



i = voxel

y_i = intensity



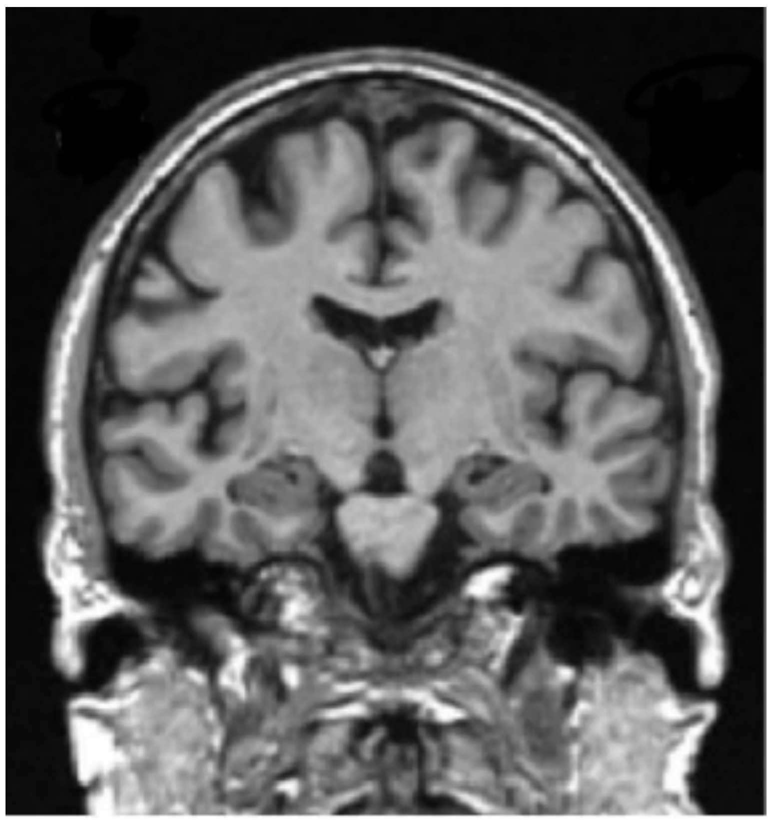
Generative model for a T1w image

i = voxel

y_i = intensity

$y_i \sim N(m_0, s_0^2)$ if i in class 0

Generative model for a T1w image



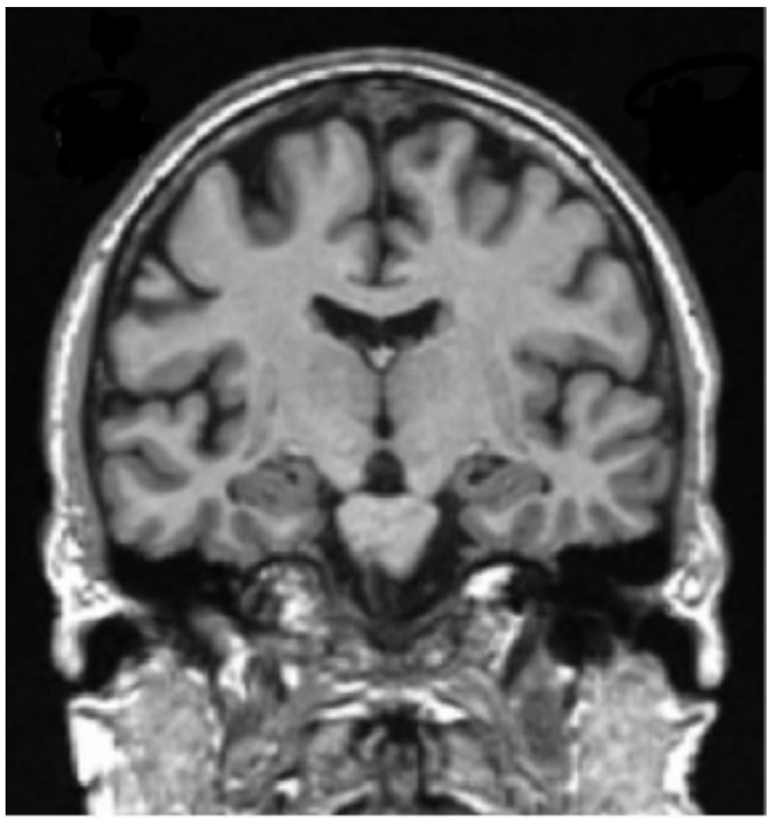
i = voxel

y_i = intensity

$y_i \sim N(m_0, s_0^2)$ if i in class 0

$y_i \sim N(m_1, s_1^2)$ if i in class 1

Generative model for a T1w image



i = voxel

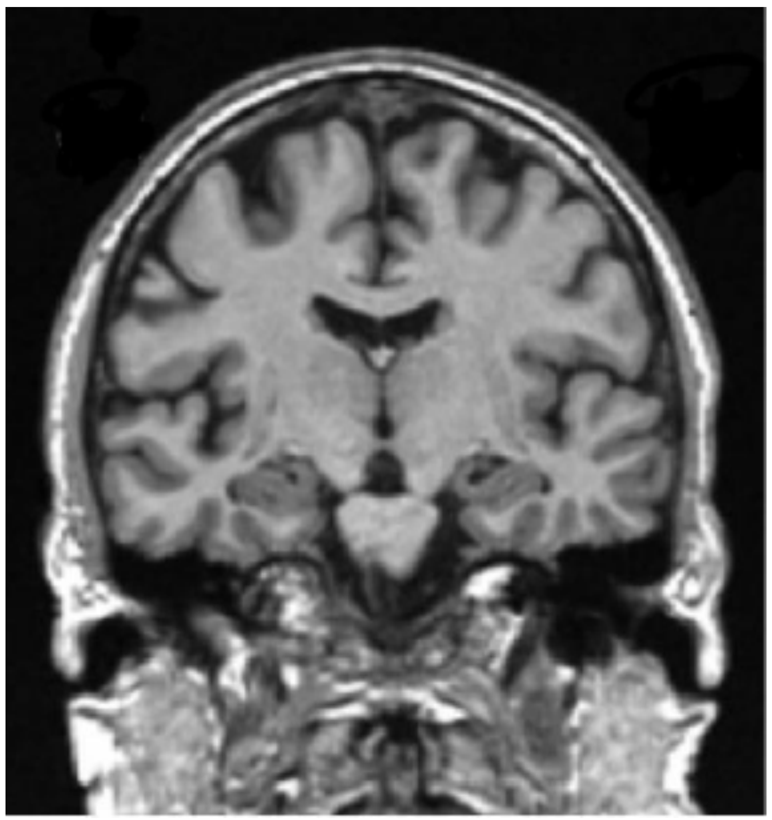
y_i = intensity

$y_i \sim N(m_0, s_0^2)$ if i in class 0

$y_i \sim N(m_1, s_1^2)$ if i in class 1

$y_i \sim N(m_2, s_2^2)$ if i in class 2

Generative model for a T1w image



i = voxel

y_i = intensity

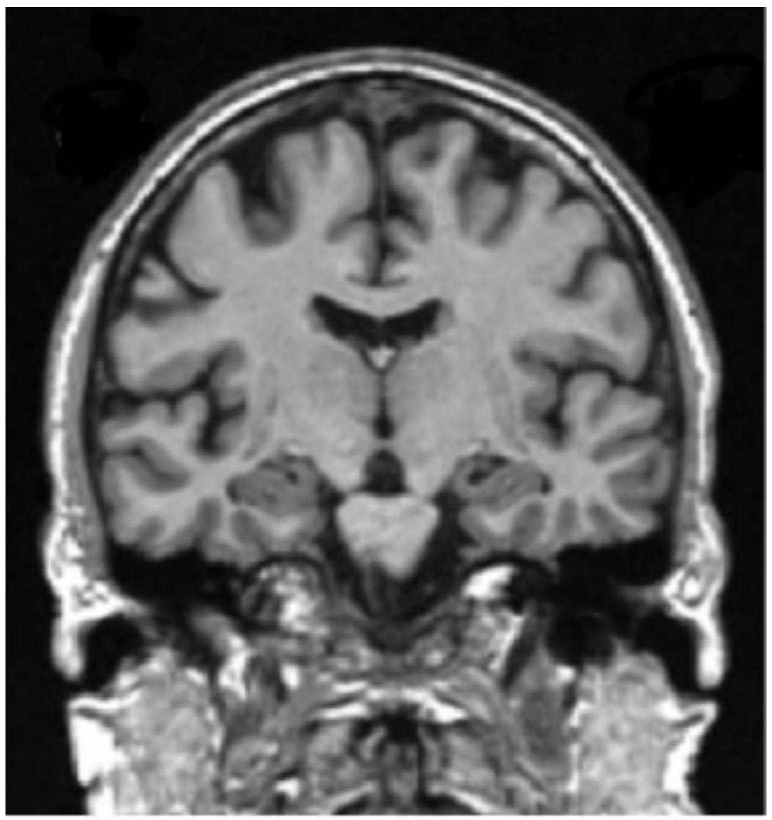
$y_i \sim N(m_0, s_0^2)$ if i in class 0

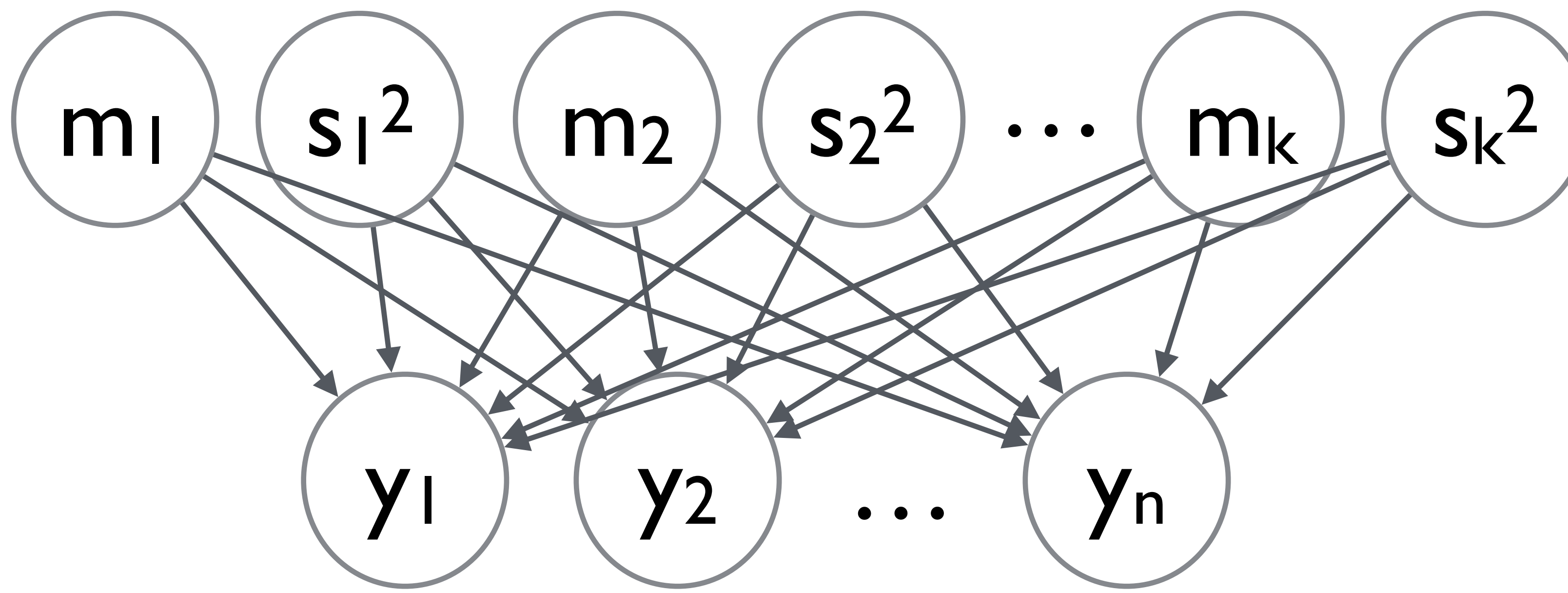
$y_i \sim N(m_1, s_1^2)$ if i in class 1

$y_i \sim N(m_2, s_2^2)$ if i in class 2

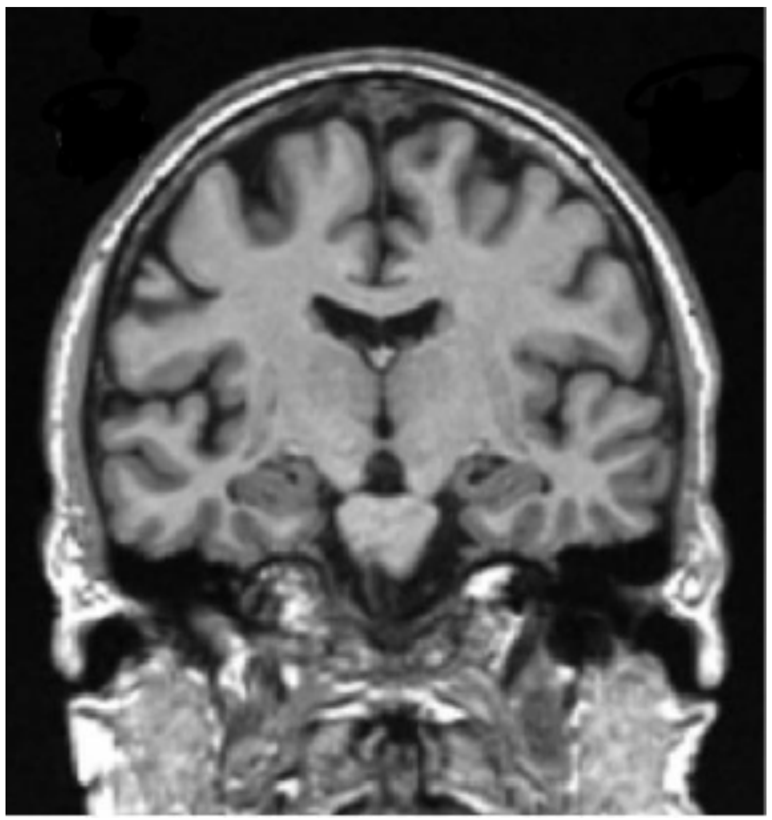
$y_i \sim N(m_k, s_k^2)$ if i in class k

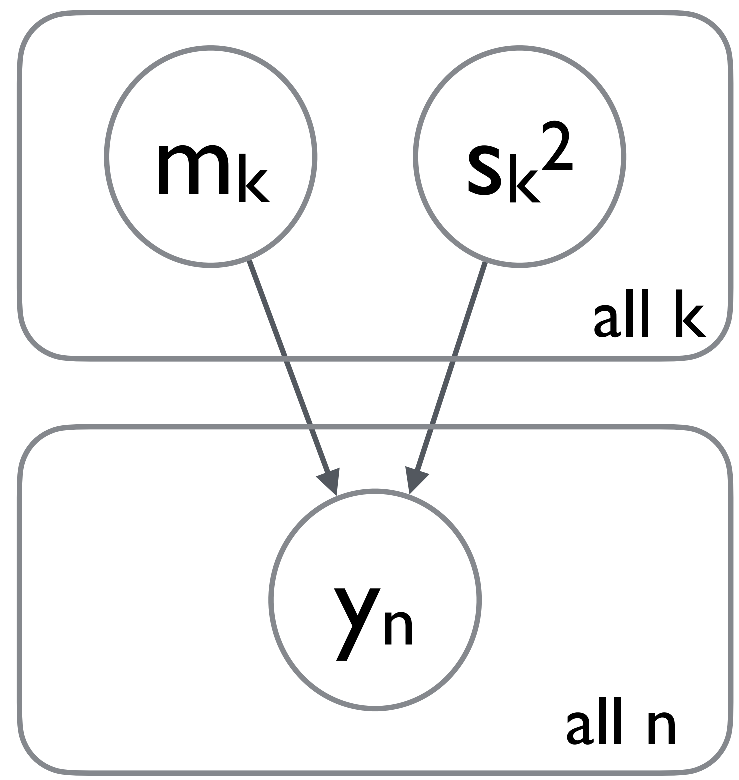
Generative model for a T1w image



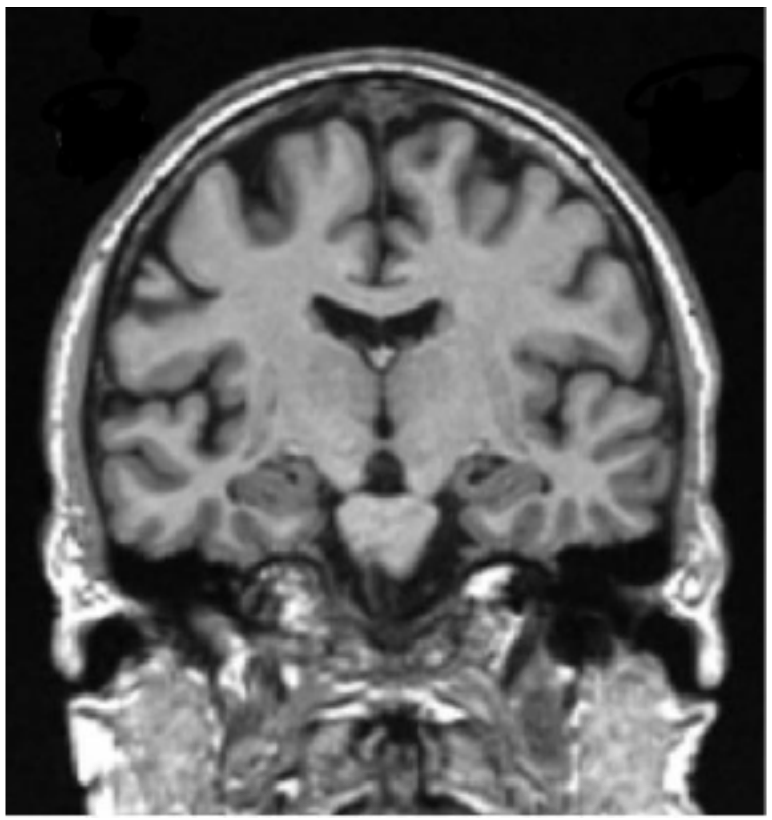


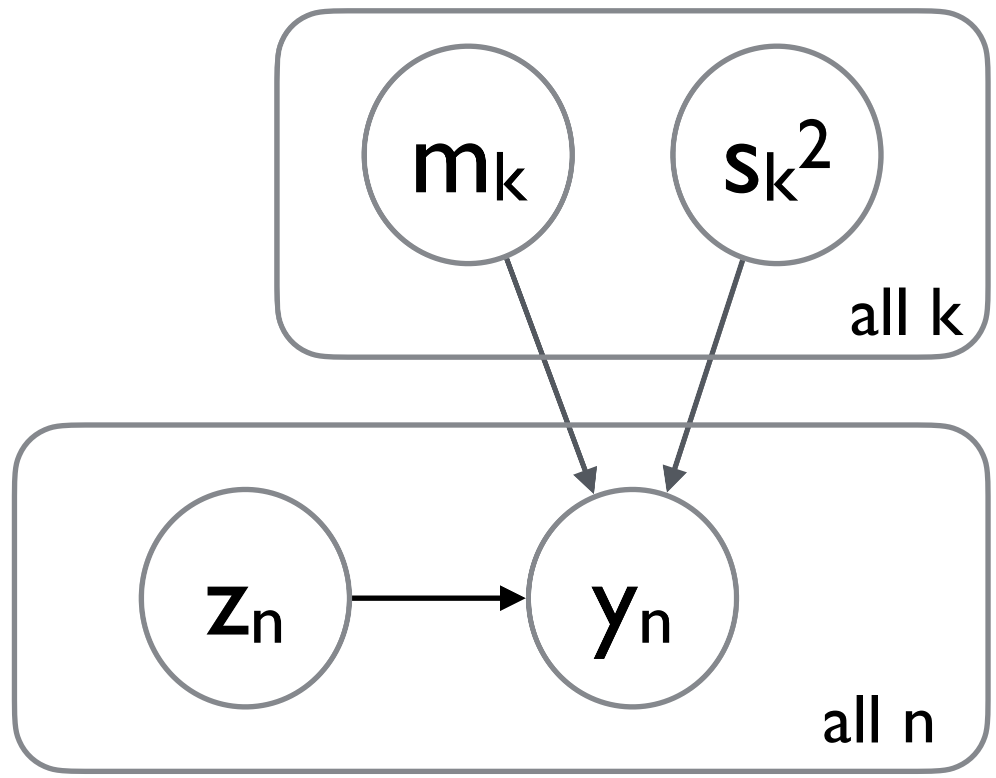
Generative model for a T1w image





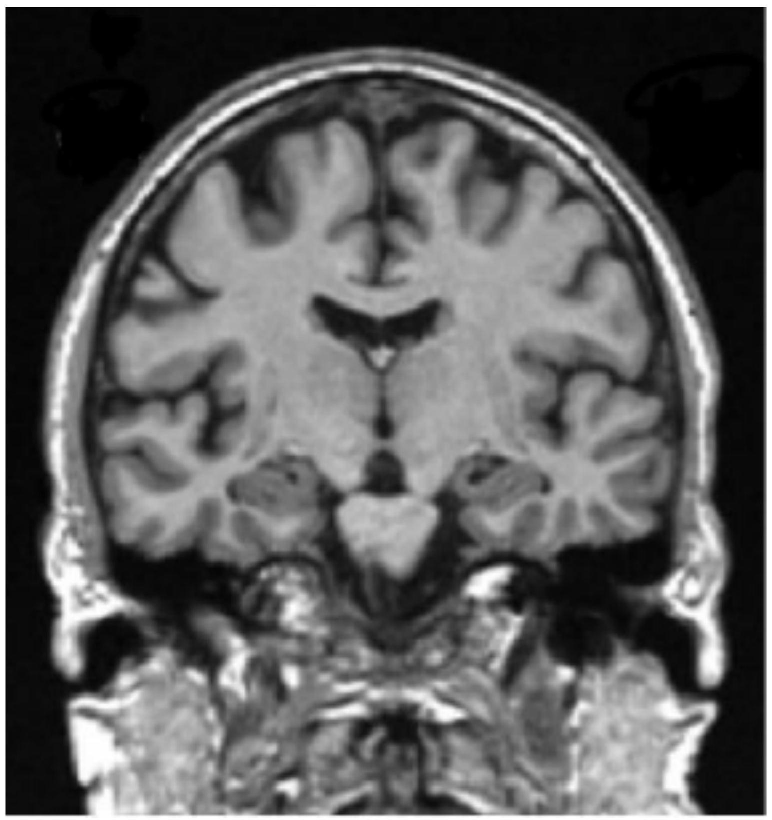
Generative model for a T1w image

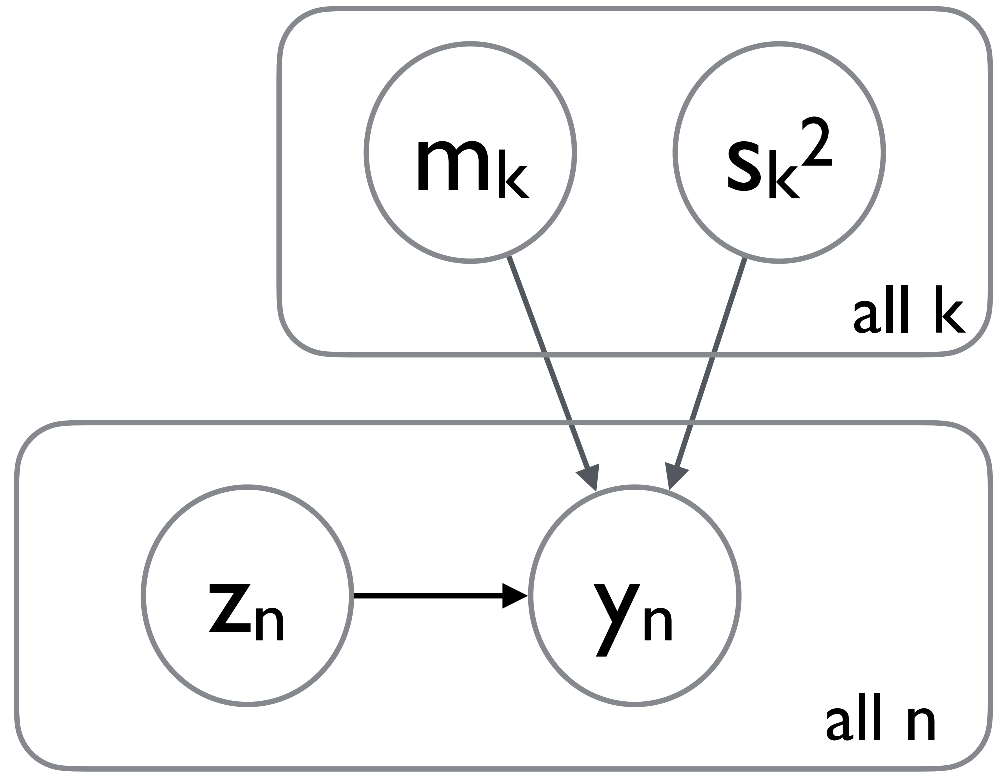




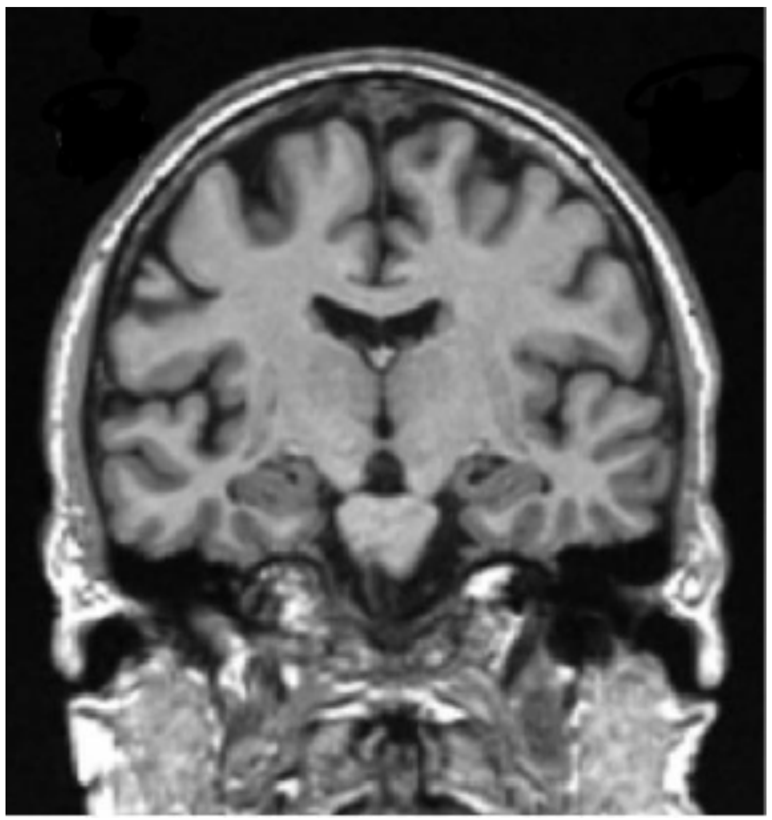
$$p(y_n|z_n=k,\{m's\},\{s's\}) = N(m_k,s_k^2)$$

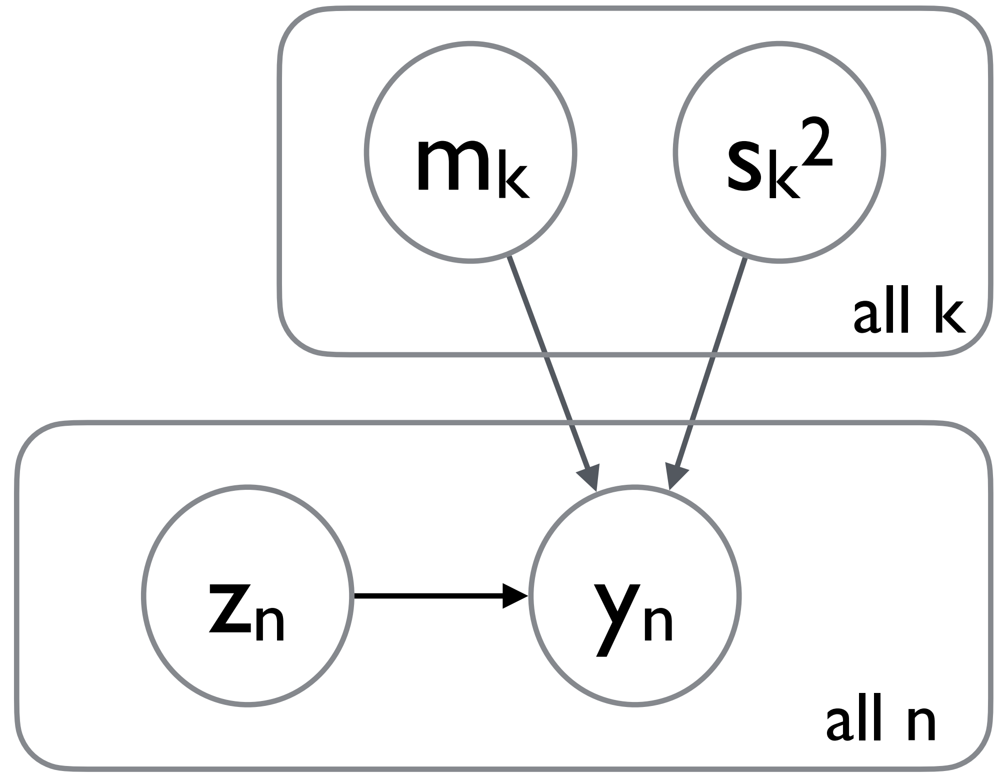
Example inference algorithm
(if uniform priors)





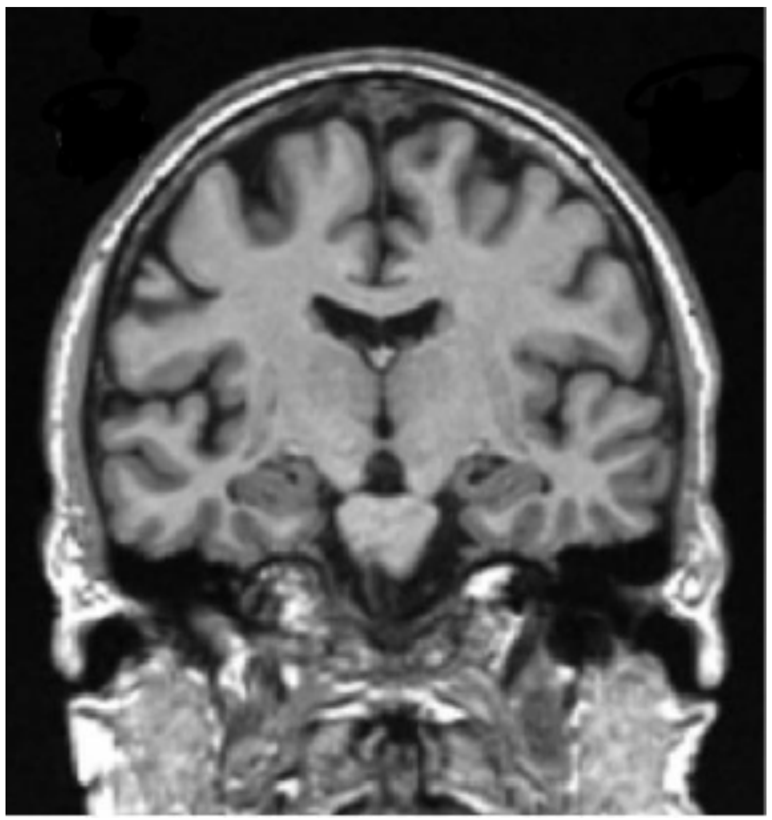
Example inference algorithm
(if uniform priors)

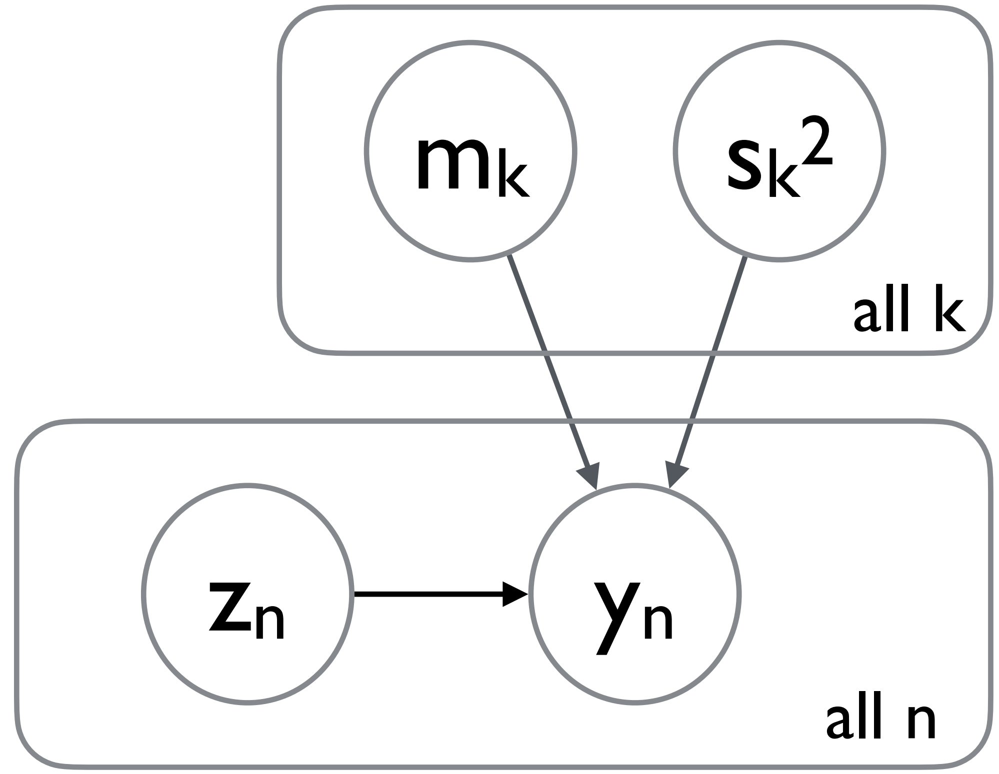




m_k = mean(y_i for i in class k)

Example inference algorithm
(if uniform priors)

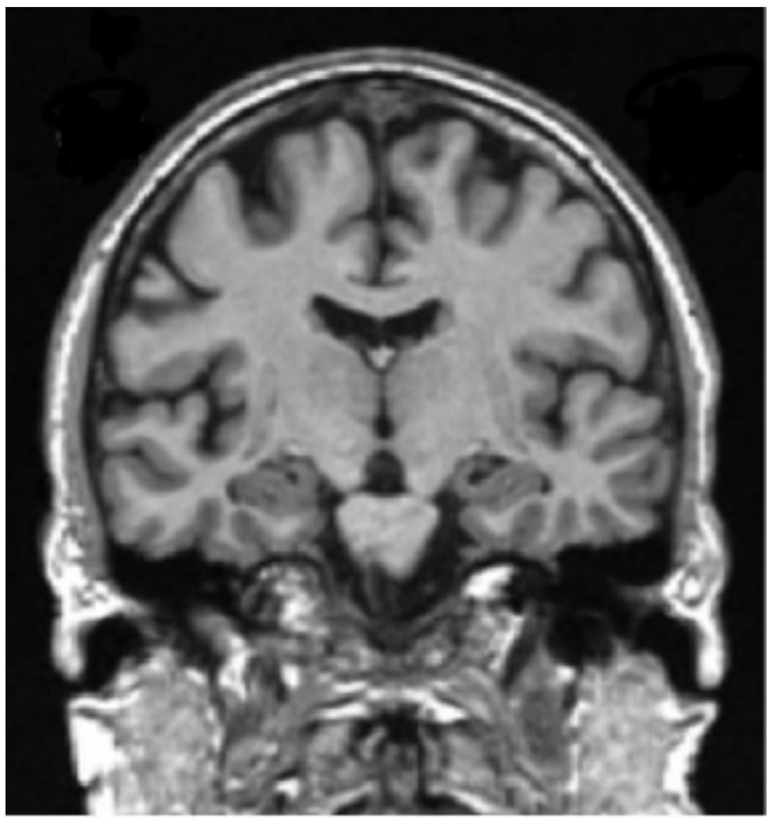


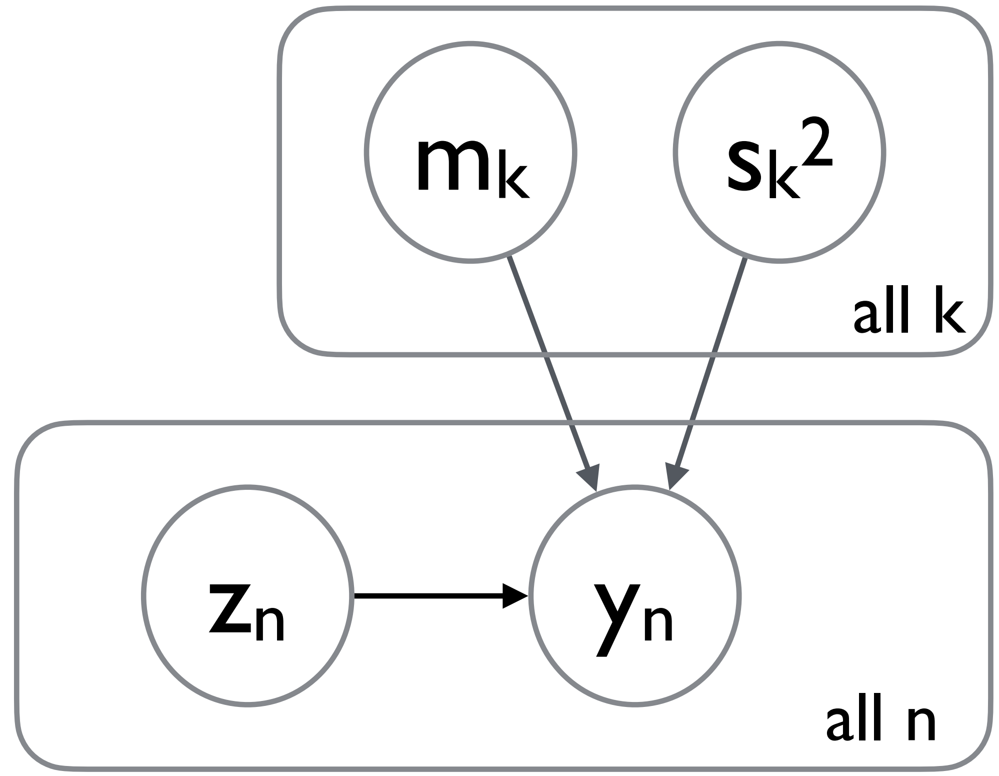


m_k = mean(y_i for i in class k)

s^2_k = var(y_i for i in class k)

Example inference algorithm
(if uniform priors)



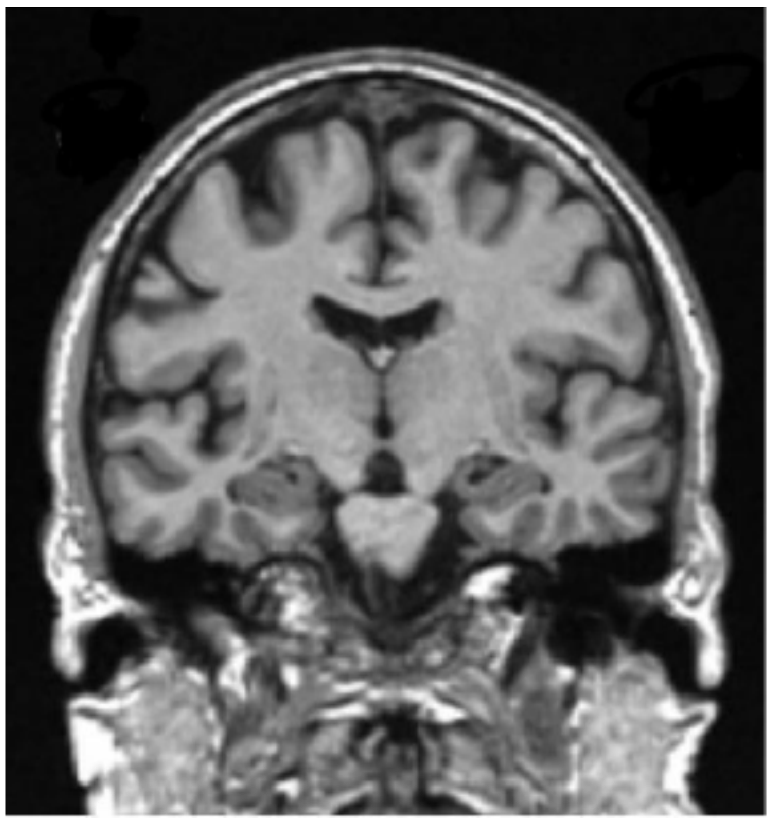


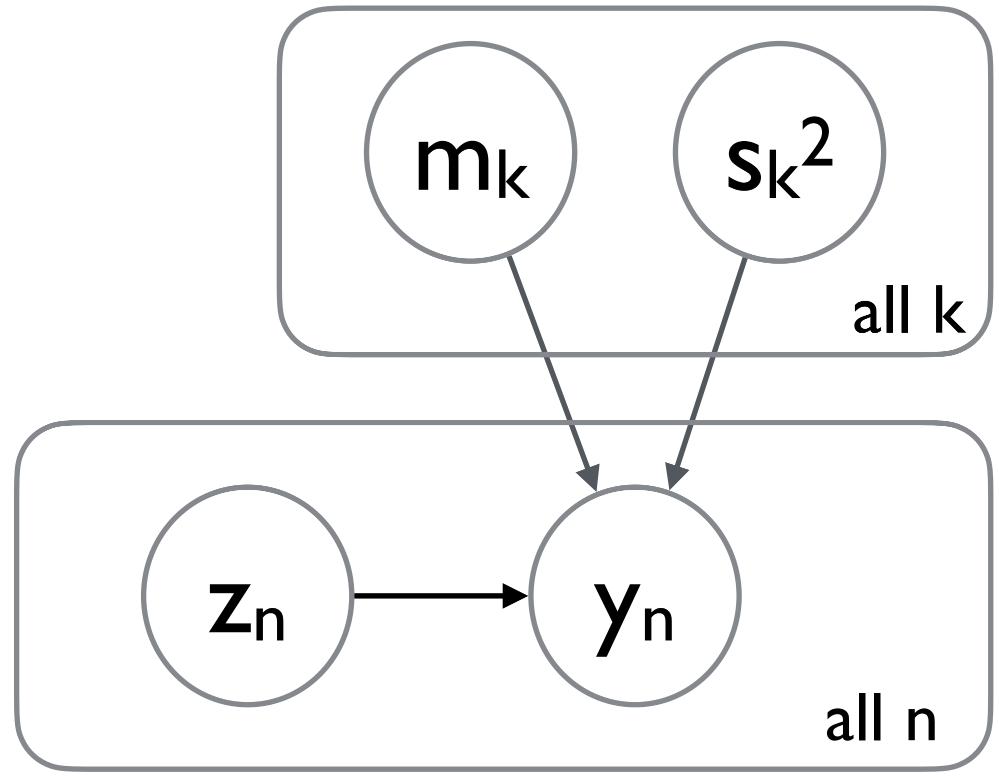
m_k = mean(y_i for i in class k)

s^2_k = var(y_i for i in class k)

zi = argmax_k [p(yi | z_i=k ,m_k,s_k)]

Example inference algorithm
(if uniform priors)





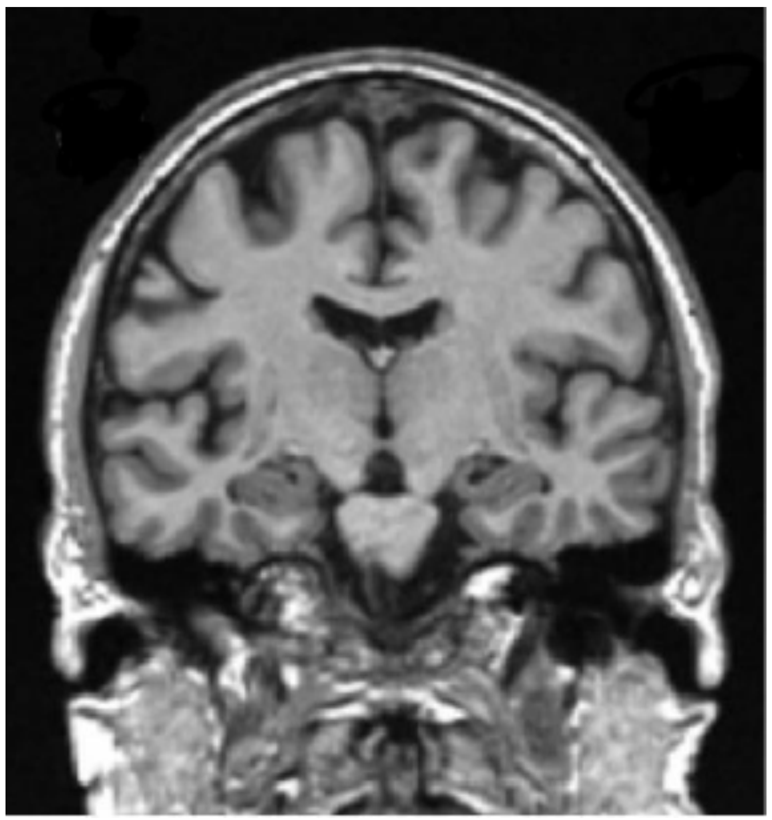
iterate

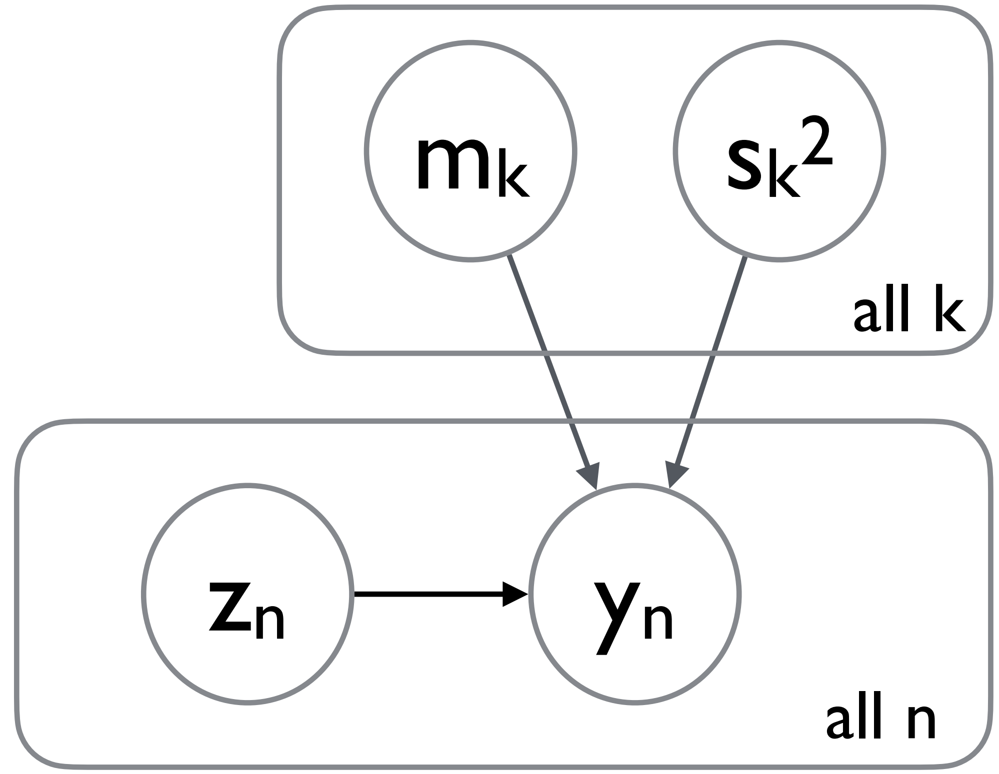
m_k = mean(y_i for i in class k)

s^2_k = var(y_i for i in class k)

zi = argmax_k [p(yi | z_i=k ,m_k,s_k)]

Example inference algorithm
(if uniform priors)





iterate

m_k = mean(y_i for i in class k)

s^2_k = var(y_i for i in class k)

$z_i = \text{argmax}_k$ [p(y_i | z_i=k ,m_k,s_k)]

Note: this only gives point-estimate, not distribution (not Bayesian!)

Outline

- Joint/Conditional/Marginal
- Bayesian graphical models
- Designing a model
- Bayes in FSL
- Derivation
  - Data fusion
  - Kalman Filter

FSL: FMRIB Software Library

Structural
- BET: *brain extraction*
- FAST: tissue *segmentation*
- FIRST: subcortical *segmentation*
- FLIRT: linear *registration*
- FNIRT: nonlinear *registration*
- FUGUE: EPI *unwarping*
- SIENA: *atrophy analysis*
- FSL-VBM: *grey matter density*
- MSM: *multimodal surface registration*

Diffusion
- Bedpostx: diffusion
- Probtrackx : *tractography*
- TBSS: *voxelwise DTI analysis*
- Eddy*: distortion correction*
- Topup*: distortion correction*
- XTRACT: *automated tractography*

Functional
- FEAT: model-based *FMRI analysis*
- MELODIC: model-free *FMRI analysis*
- FIX: *Artefact cleanup*
- FLOBS: optimal HRF basis functions
- FABBER: perfusion analysis
- PROFUMO: model-free *FMRI analysis*
- FLICA: *multimodal analysis*

Other tools
- FSLeyes: display tool
- Randomise: inference
- PALM: inference
- BIANCA: *lesion detection*
- Brain atlases
- POSSUM: FMRI simulator
- FSLUTILS misc. utilities
  e.g. fslmaths, fslstats, etc.

FSL: FMRIB Software Library

Structural
- BET: *brain extraction*
- FAST: tissue *segmentation*
- FIRST: subcortical *segmentation*
- FLIRT: linear *registration*
- FNIRT: nonlinear *registration*
- FUGUE: EPI *unwarping*
- SIENA: *atrophy analysis*
- FSL-VBM: *grey matter density*
- MSM: *multimodal surface registration*

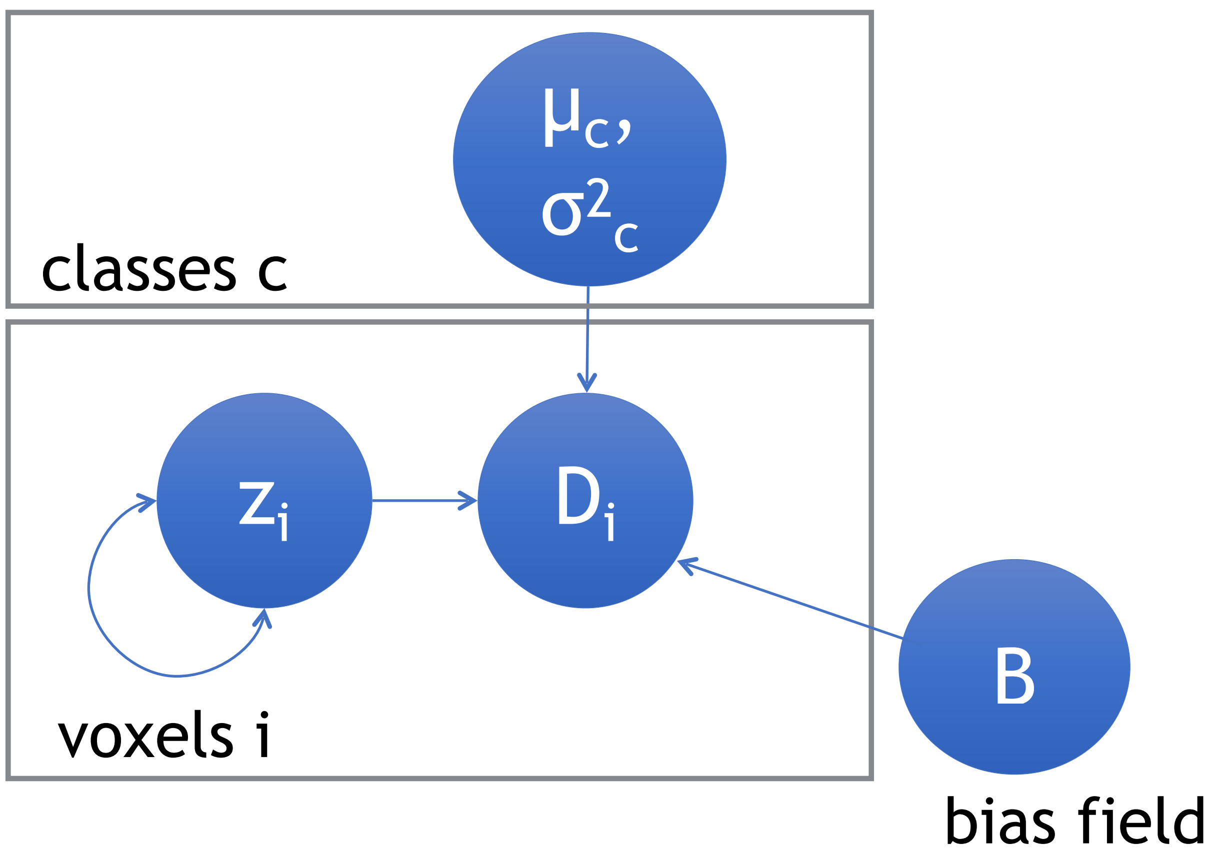
Diffusion
- Bedpostx: diffusion
- Probtrackx : *tractography*
- TBSS: *voxelwise DTI analysis*
- Eddy: *distortion correction*
- Topup: *distortion correction*
- XTRACT: *automated tractography*

Functional
- FEAT: model-based *FMRI analysis*
- MELODIC: model-free *FMRI analysis*
- FIX: *Artefact cleanup*
- FLOBS: optimal HRF basis functions
- FABBER: perfusion analysis
- PROFUMO: model-free *FMRI analysis*
- FLICA: *multimodal analysis*

Other tools
- FSLeyes: display tool
- Randomise: inference
- PALM: inference
- BIANCA: *lesion detection*
- Brain atlases
- POSSUM: FMRI simulator
- FSLUTILS misc. utilities
  e.g. fslmaths, fslstats, etc.

FAST - segmentation

Each voxel *i* is assigned a class $z_i = c$, where *c* is one of {CSF, GM, WM}



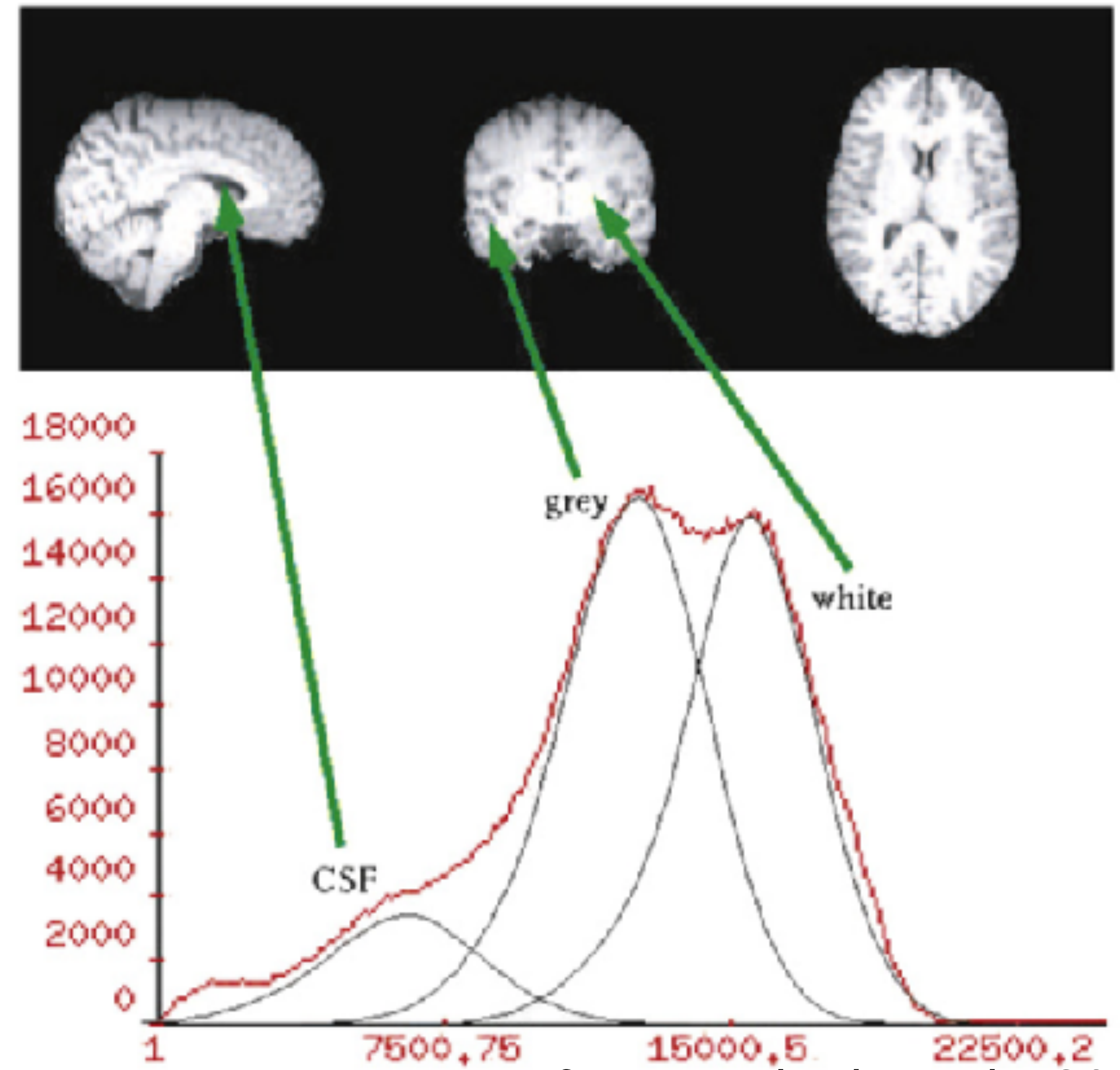
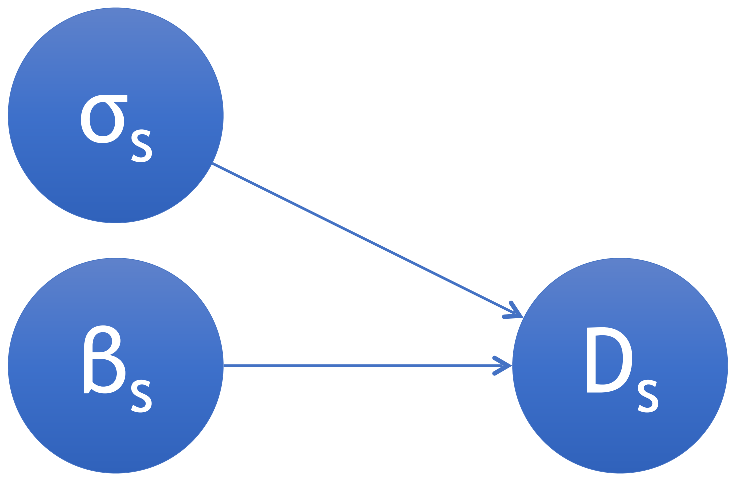


Figure from Woolrich et al. (2009)

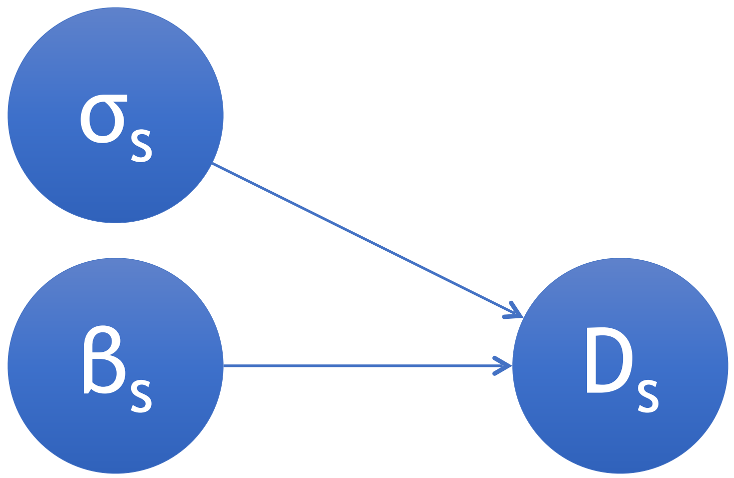
FEAT - first level



joint posterior | likelihood (GLM) | priors

$$p(\beta_s, \sigma_s^2 | D_s) \propto p(D_s | \beta_s, \sigma_s^2) p(\beta_s) p(\sigma_s^2)$$

FEAT - first level



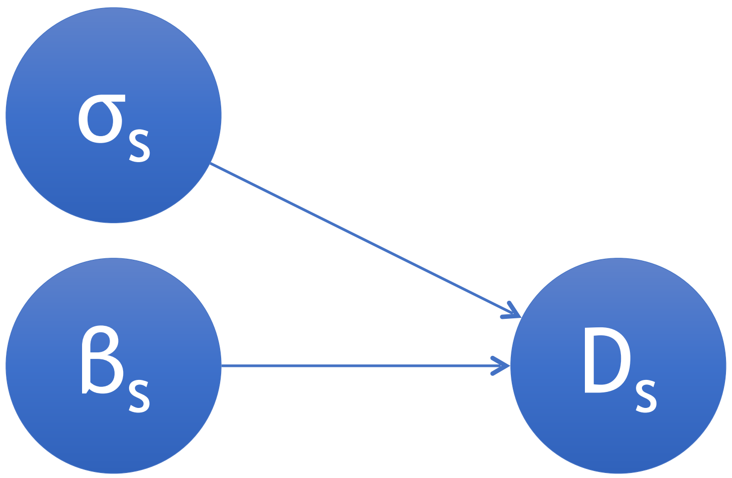
joint posterior   likelihood (GLM)   priors

$$p(\beta_s, \sigma_s^2 | D_s) \propto p(D_s | \beta_s, \sigma_s^2) p(\beta_s) p(\sigma_s^2)$$

$$p(\beta_s | D_s) = \int p(\beta_s, \sigma_s^2 | D_s) p(\beta_s) \mathrm{d}\beta_s$$

Marginal posterior for beta (cope)

FEAT - first level



joint posterior — likelihood (GLM) — priors

$$p(\beta_s, \sigma_s^2 | D_s) \propto p(D_s | \beta_s, \sigma_s^2) p(\beta_s) p(\sigma_s^2)$$

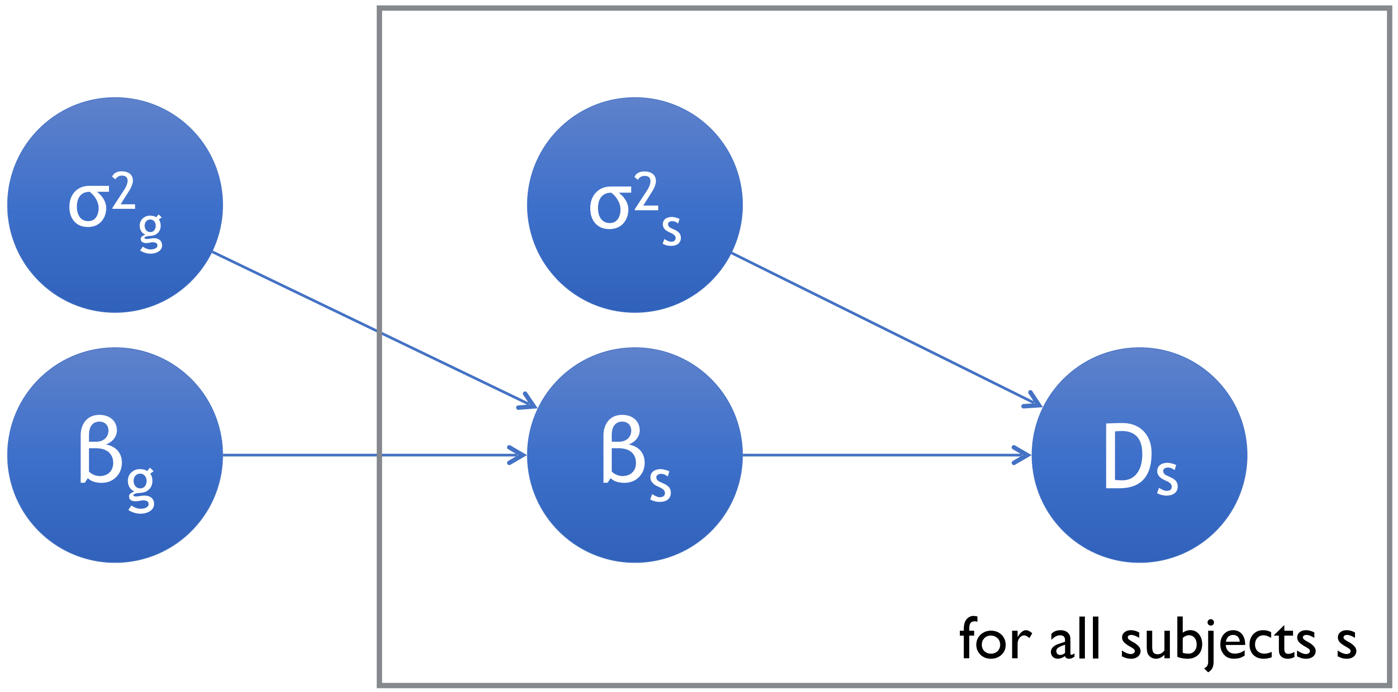
$$p(\beta_s | D_s) = \int p(\beta_s, \sigma_s^2 | D_s) p(\beta_s) \mathrm{d}\beta_s$$

Marginal posterior for beta (cope)

$$p(\sigma_s^2 | D_s) = \int p(\beta_s, \sigma_s^2 | D_s) p(\sigma_s^2) \mathrm{d}\sigma_s^2$$

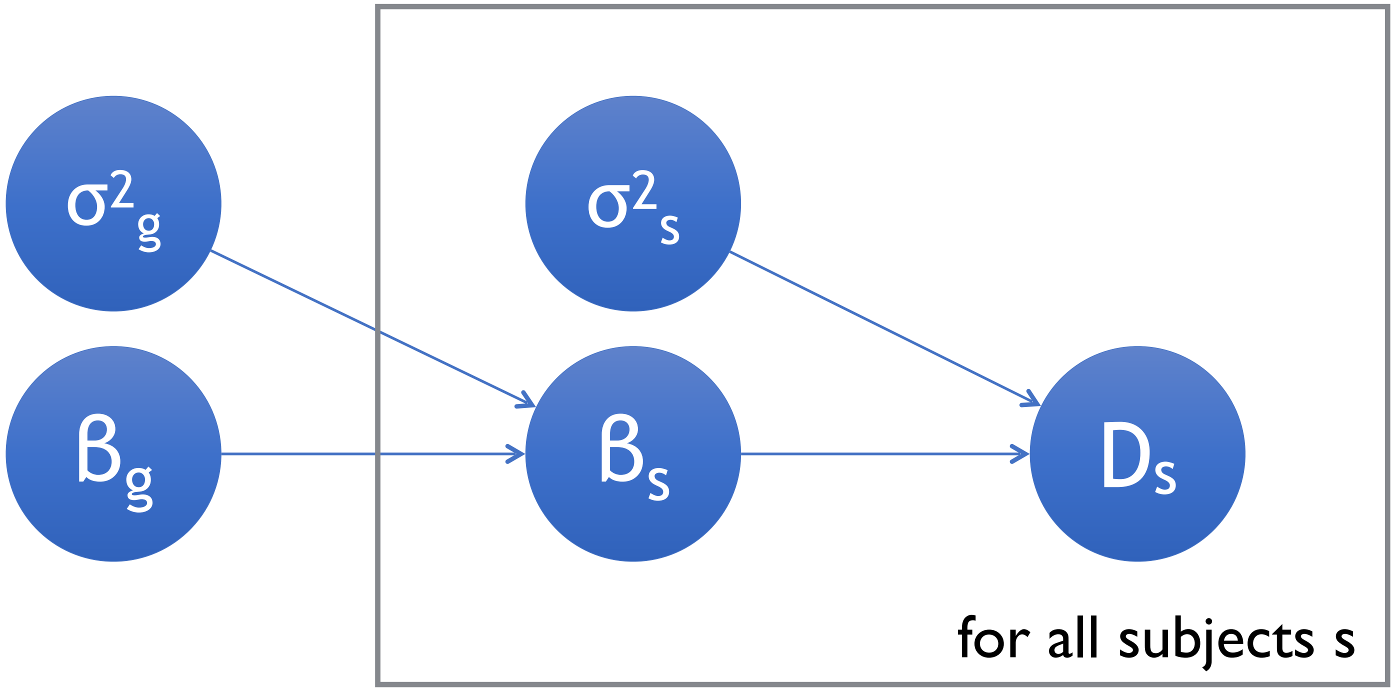
Marginal posterior for sigma^2 (varcope)

FEAT - second level



$$p(\text{all params}|\text{all data}) \propto p(\beta_g)p(\sigma_g^2)\prod_s p(D_s|\beta_s,\sigma_s^2)p(\beta_s|\beta_g,\sigma_g^2)p(\sigma_s^2)$$

FEAT - second level

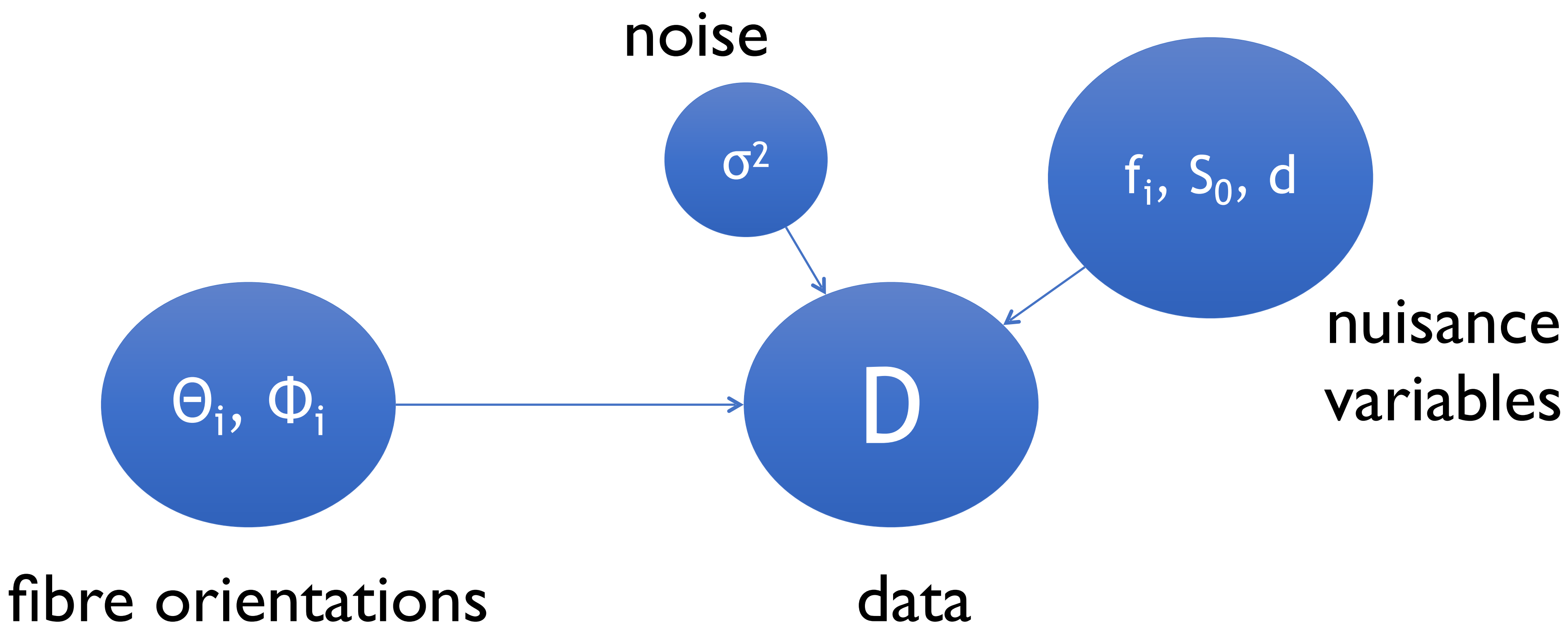


$$p(\text{all params}|\text{all data}) \propto p(\beta_g)p(\sigma_g^2)\prod_s p(D_s|\beta_s,\sigma_s^2)p(\beta_s|\beta_g,\sigma_g^2)p(\sigma_s^2)$$

Integrate over subjects using summary statistics from first level

BEDPOSTX - diffusion model

- For every voxel:

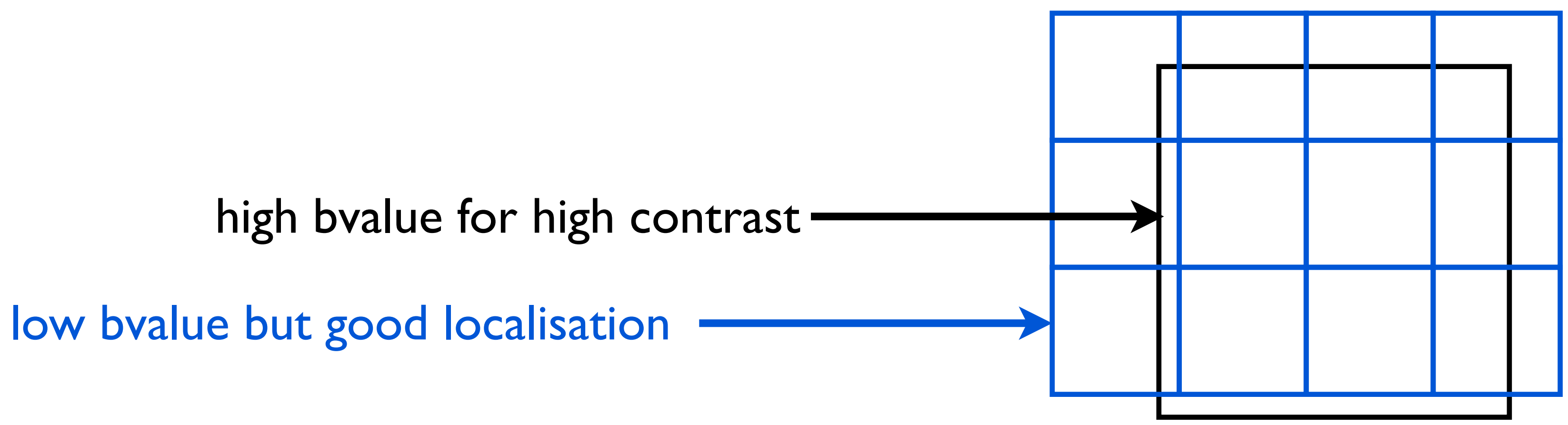


RUBIX - data fusion

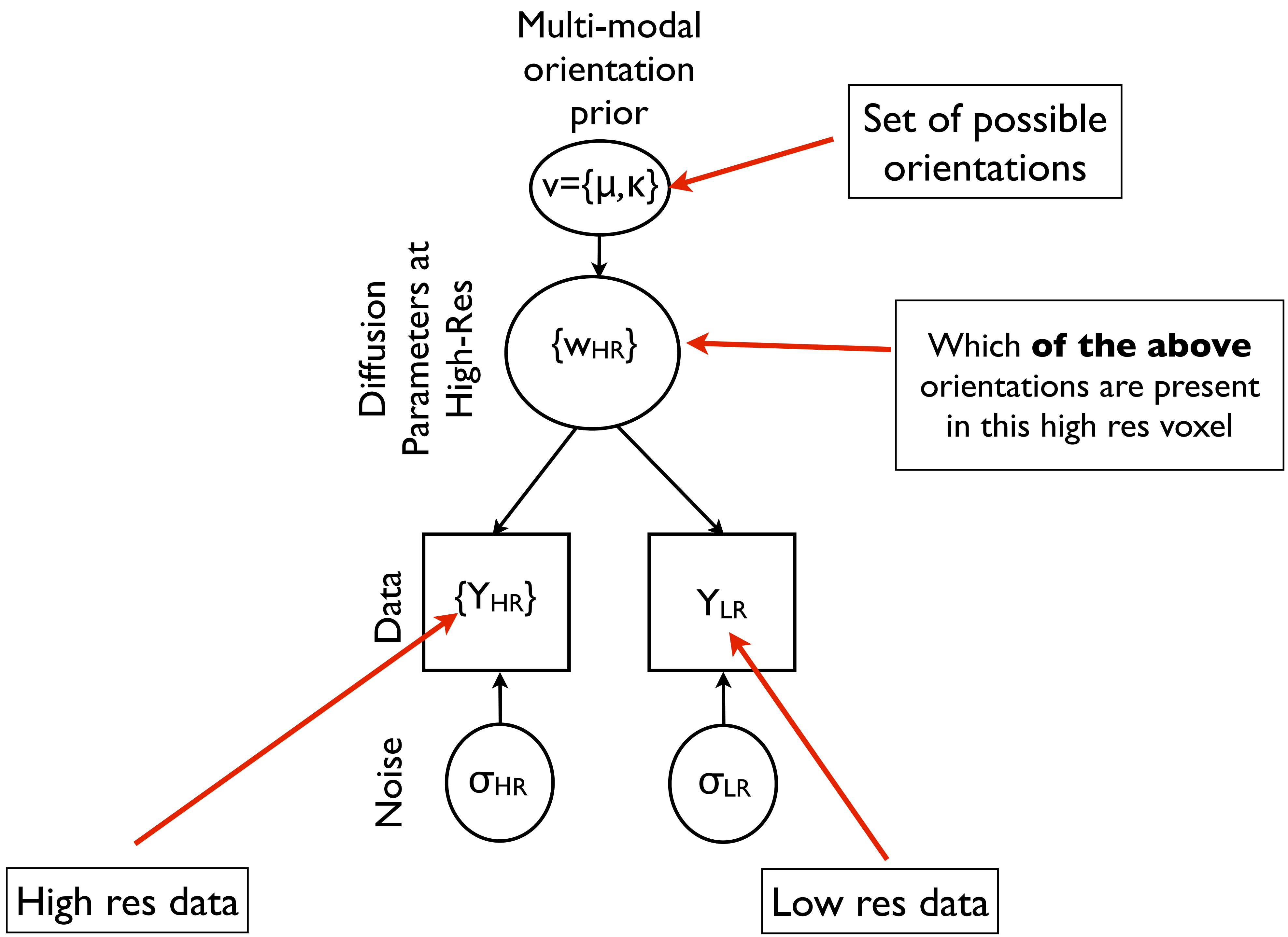
- RUBIX - Resolutions unified with Bayesian inference on Xing fibres

RUBIX - data fusion

- RUBIX - Resolutions unified with Bayesian inference on Xing fibres



RUBIX - data fusion



Sotiropoulos, Jbabdi, Behrens et al. IEEE TMI 2013

ASL: biophysical priors

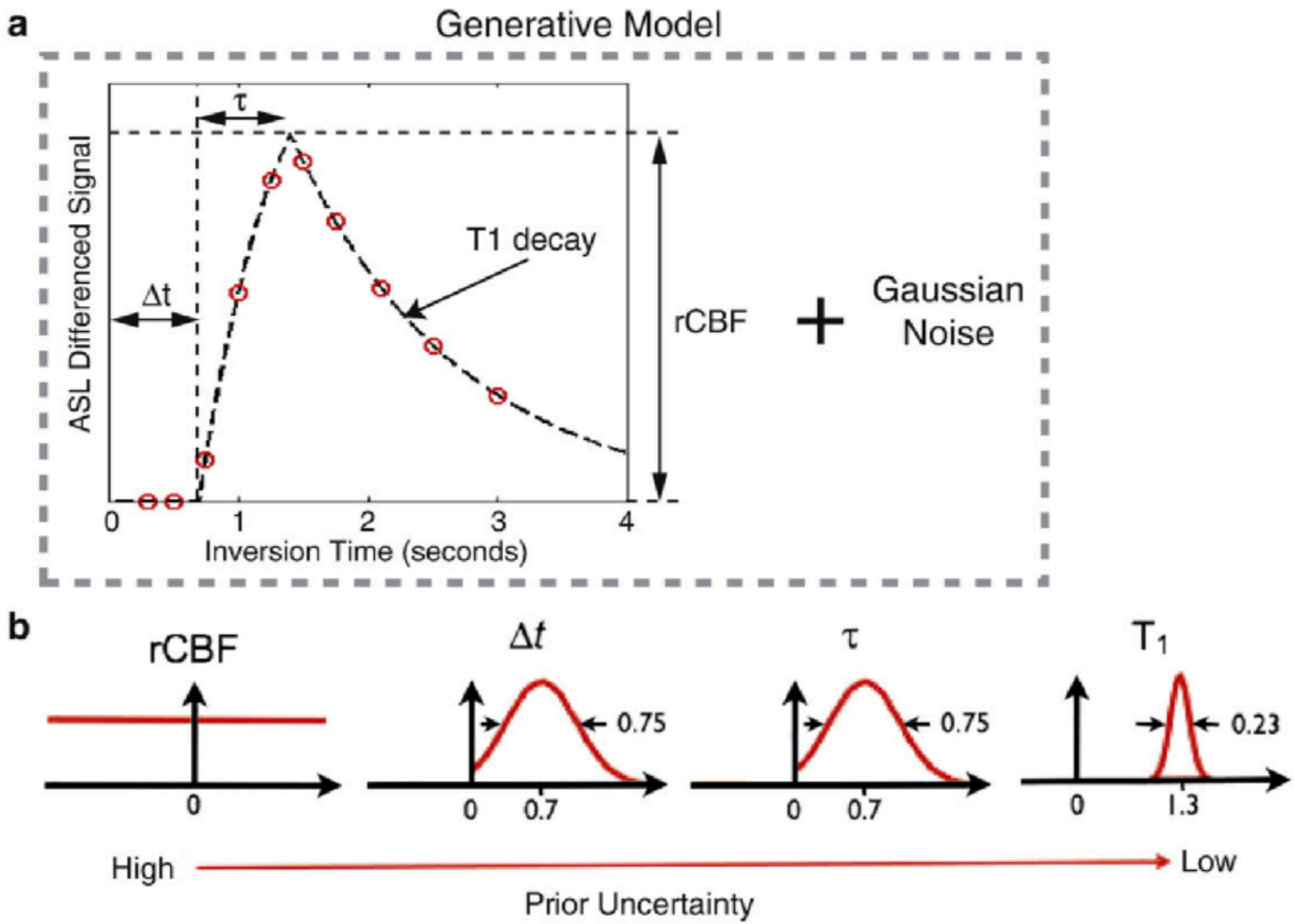
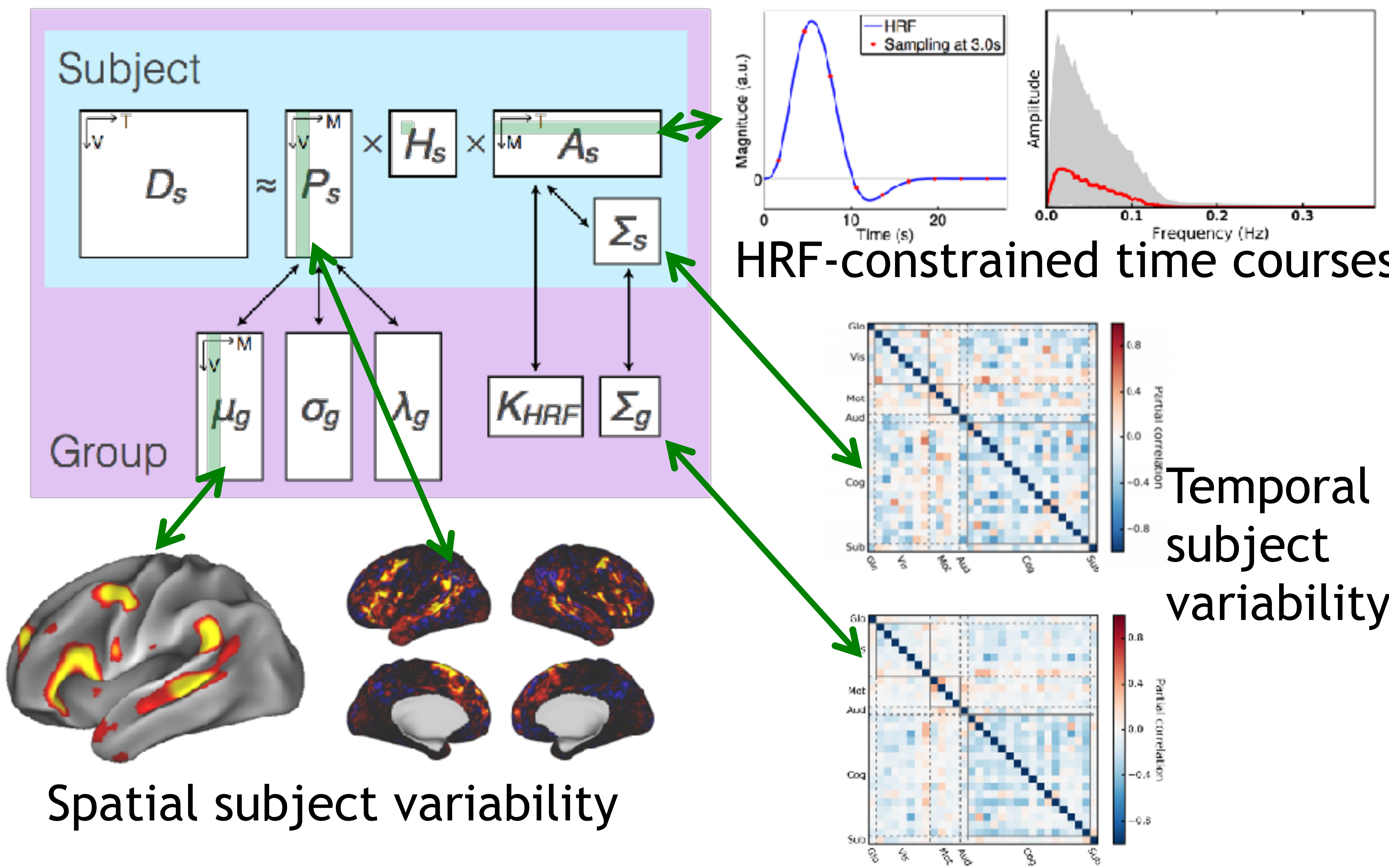
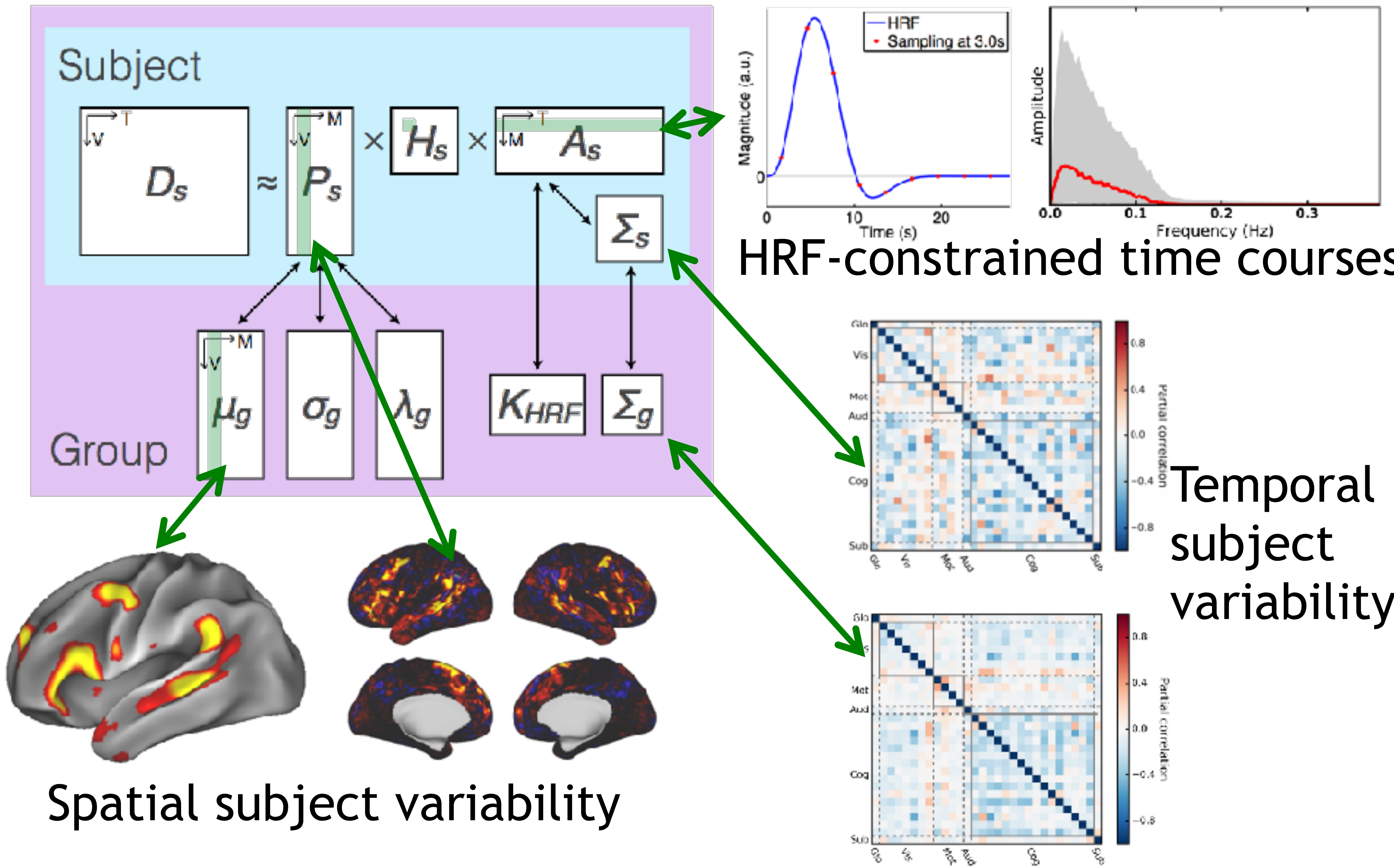


Figure from Woolrich et al. (2009)

PROFUMO

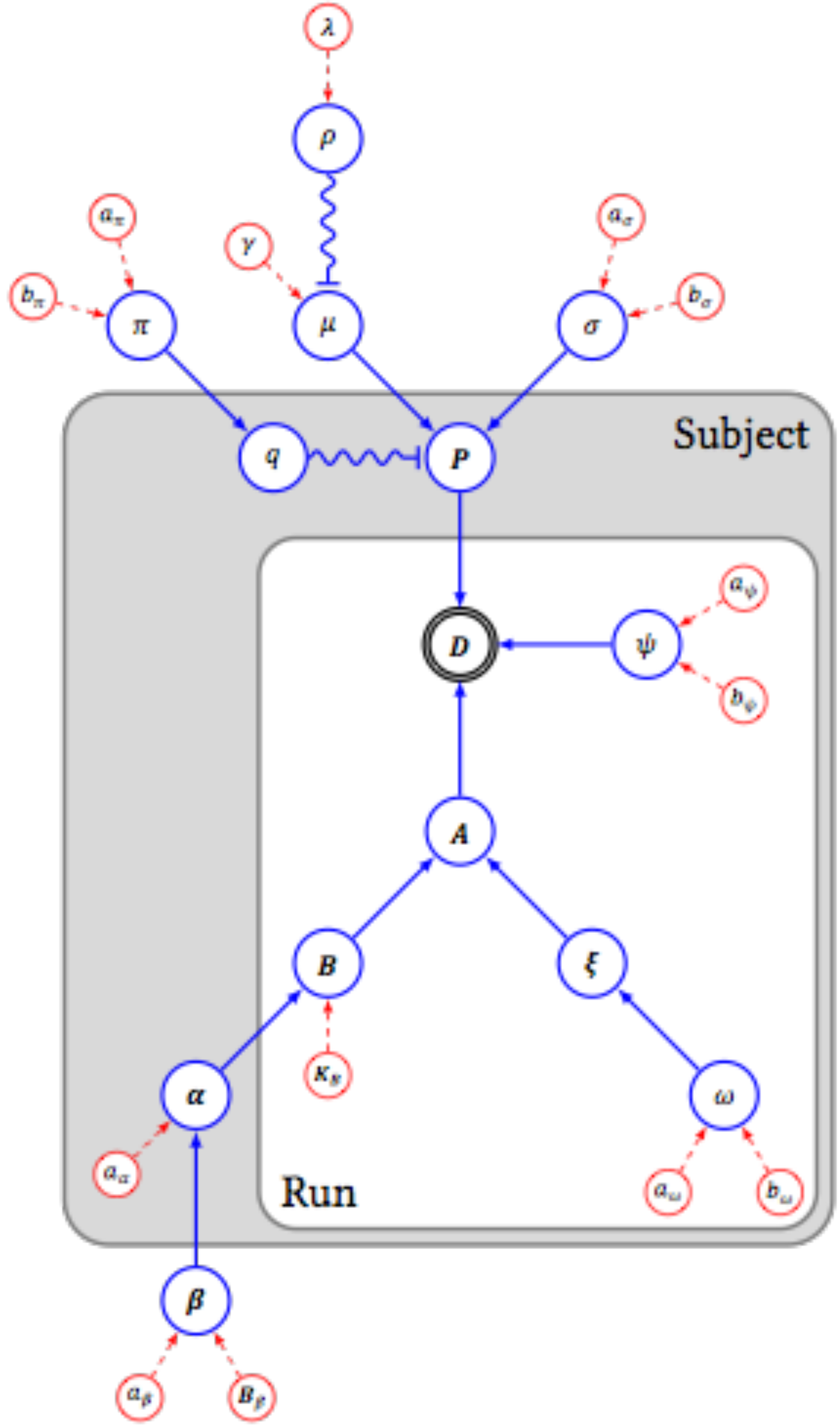

Subject
$D_s \approx P_s \times H_s \times A_s$
Σ_s
Group
μ_g
σ_g
λ_g
K_{HRF}
Σ_g
HRF
Sampling at 3.0s
Magnitude (a.u.)
Time (s)
Amplitude
Frequency (Hz)
HRF-constrained time courses
Temporal subject variability
Partial correlation
Spatial subject variability

PROFUMO

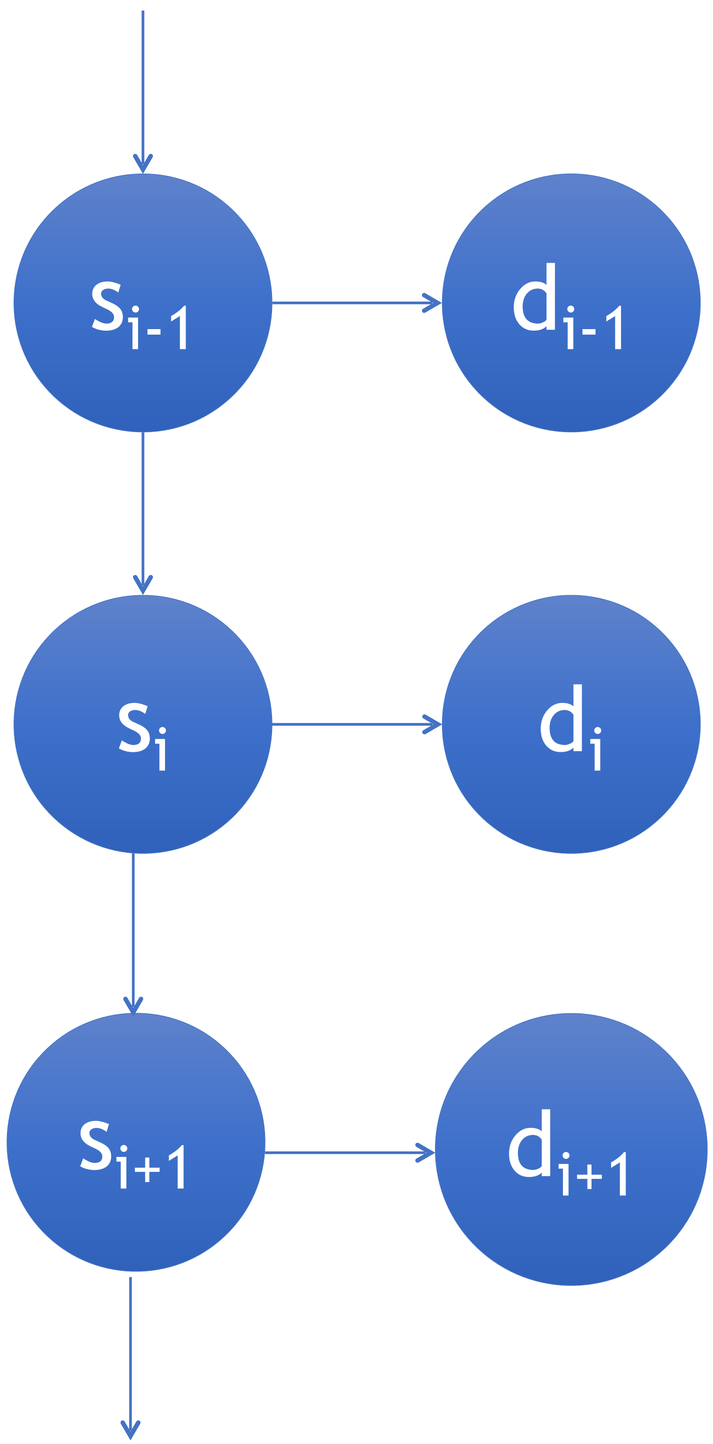


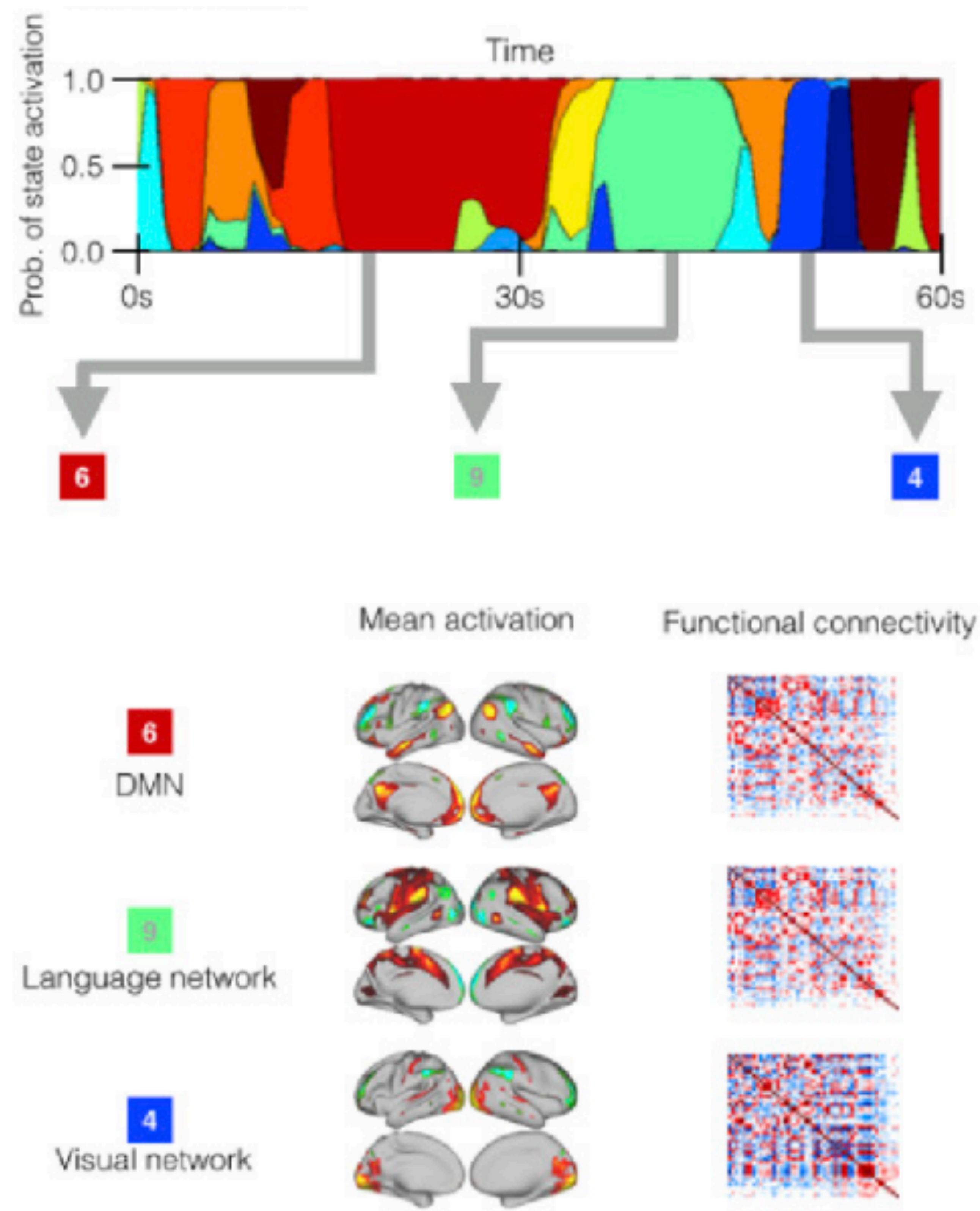
Subject
Group
HRF
Sampling at 3.0s
Magnitude (a.u.)
Time (s)
Amplitude
Frequency (Hz)
Partial correlation
HRF-constrained time courses
Temporal subject variability
Spatial subject variability

PROFUMO



MEG: Diego's Hidden Markov Models





FLAME: biophysically constrained HRF

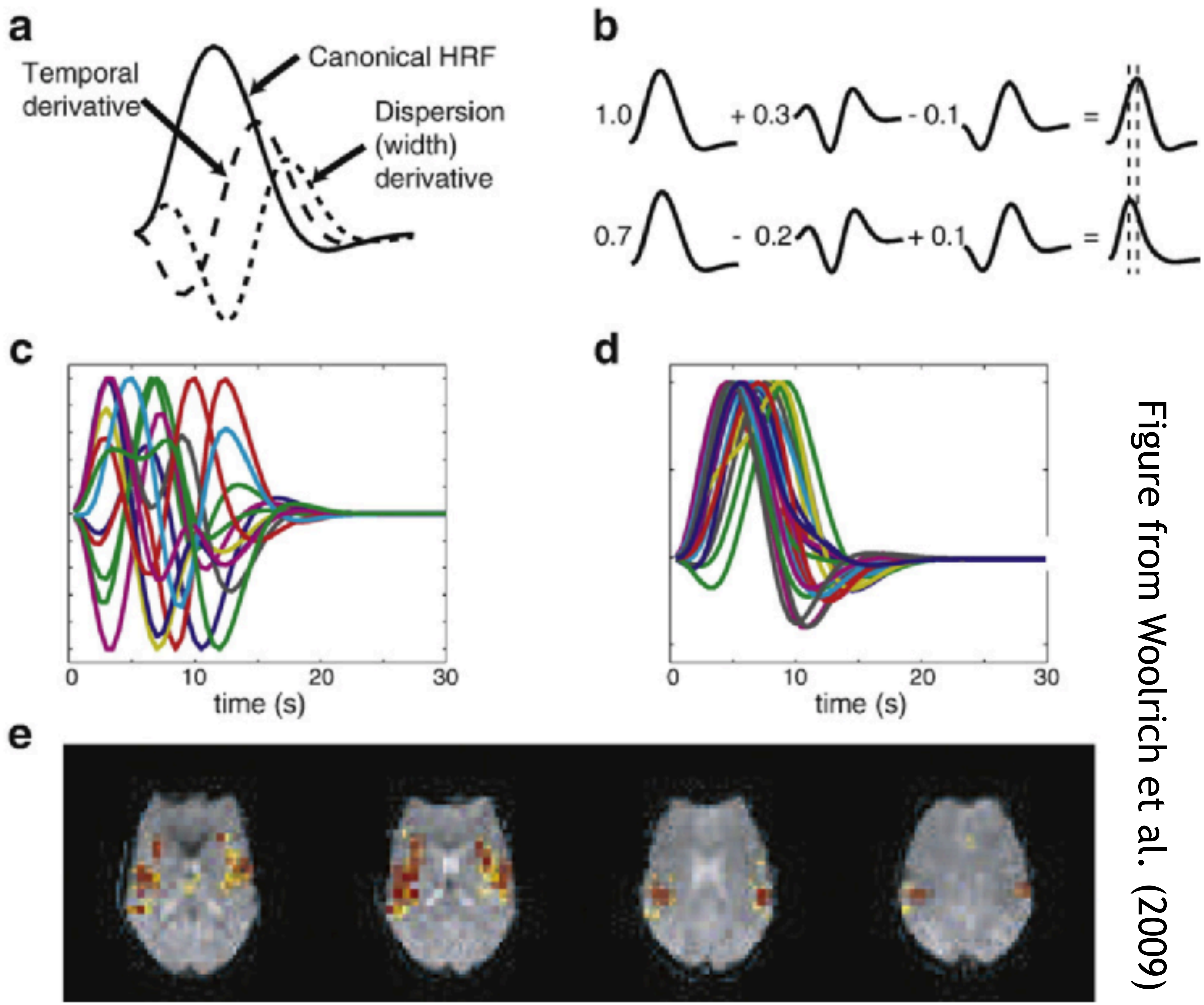


Figure from Woolrich et al. (2009)

Outline

- Joint/Conditional/Marginal
- Bayesian graphical models
- Designing a model
- Bayes in FSL
- Derivation
  - Data fusion
  - Kalman Filter

I am going to be using Gaussian distributions from now on

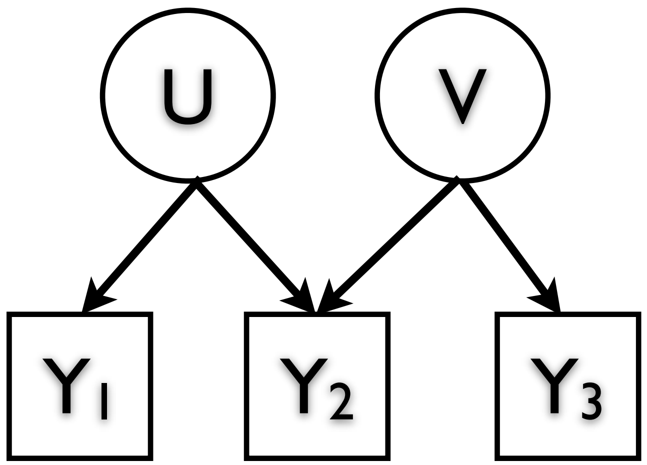
- Notation: $p(x|m,s^2) = N(x|m,s^2)$
- variance s^2 = 1/precision = 1/b
- $N(x|m,s^2) = N(x|m,1/b)$

$$\mathrm{N}(x|m,s^2) = \sqrt{\frac{1}{2\pi s^2}} \exp\left[-\frac{(x-m)^2}{2s^2}\right]$$

$$\mathrm{N}(x|m,1/b) = \sqrt{\frac{b}{2\pi}} \exp\left[-\frac{b(x-m)^2}{2}\right]$$

Graphical models

Data fusion



$Y_1 = U$ +noise

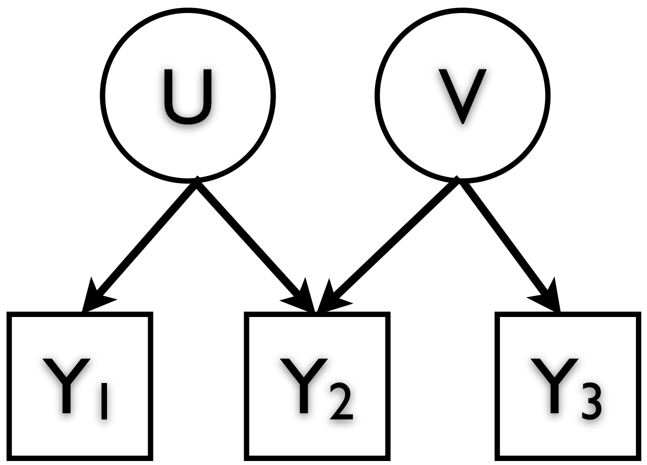
$Y_2 = U + V$ +noise

$Y_3 = V$ +noise

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

Graphical models

Data fusion



$Y_1 = U$ +noise

$Y_2 = U + V$ +noise

$Y_3 = V$ +noise

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

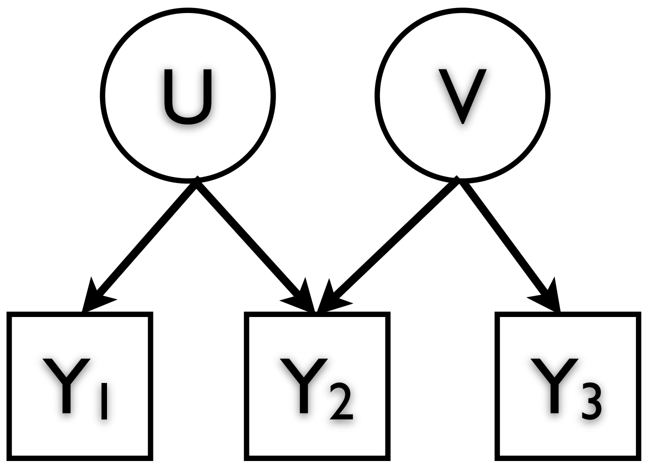
$$p(Y_1|U) = N(U|Y_1, 1/b_1)$$

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

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

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)

Two useful facts about the Gaussian distribution

Two useful facts about the Gaussian distribution

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

Two useful facts about the Gaussian distribution

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

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

Two useful facts about the Gaussian distribution

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

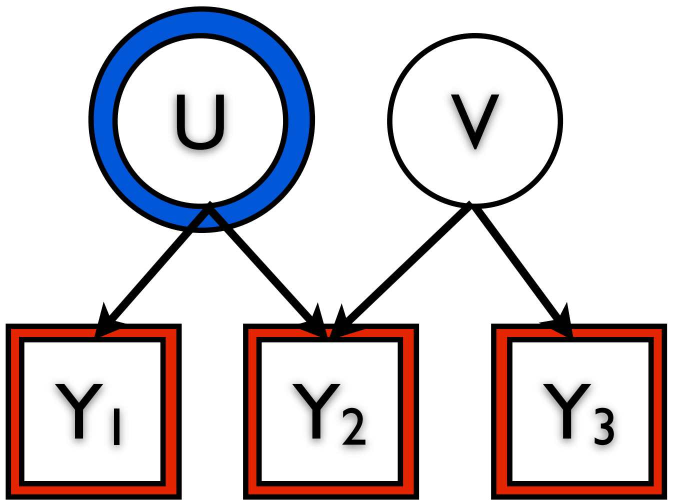
$N(\mathbf{x}|a, 1/b)N(\mathbf{x}|c, 1/d) = N(\mathbf{x}|..., b+d)N(a|c, 1/b+1/d)$

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

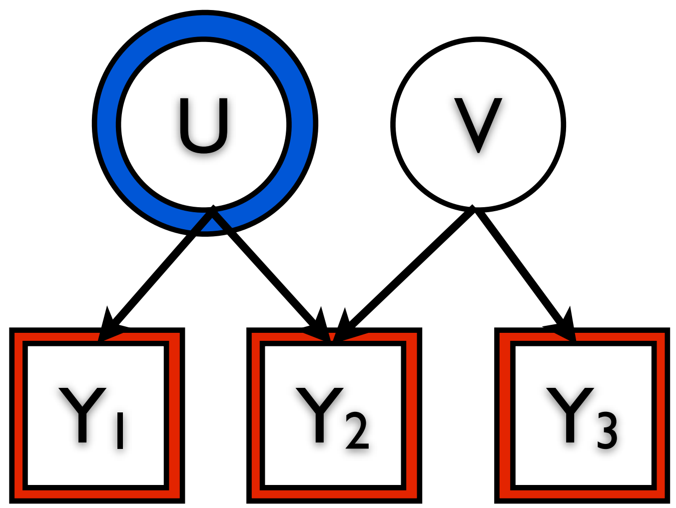


$Y_1 = U$ +noise
$Y_2 = U + V$ +noise
$Y_3 = V$ +noise

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

Graphical models

Data fusion



$Y_1 = U$ +noise

$Y_2 = U + V$ +noise

$Y_3 = V$ +noise

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

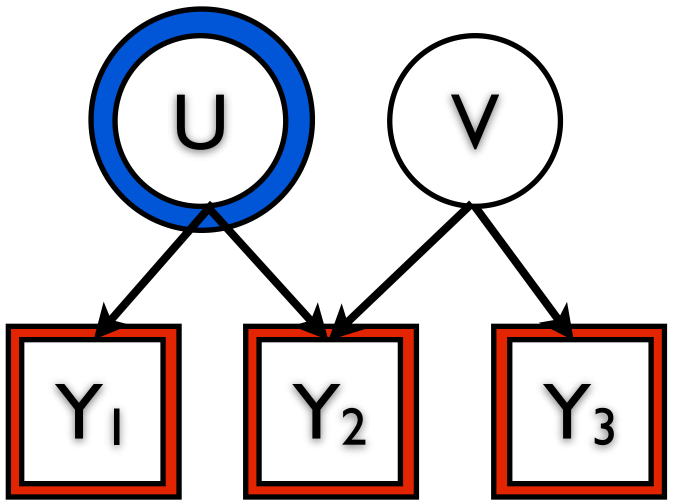
$p(Y_2|U+V) = N(Y_2|U+V, 1/b_2)$

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

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

Graphical models

Data fusion



$Y_1 = U$ +noise

$Y_2 = U + V$ +noise

$Y_3 = V$ +noise

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

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

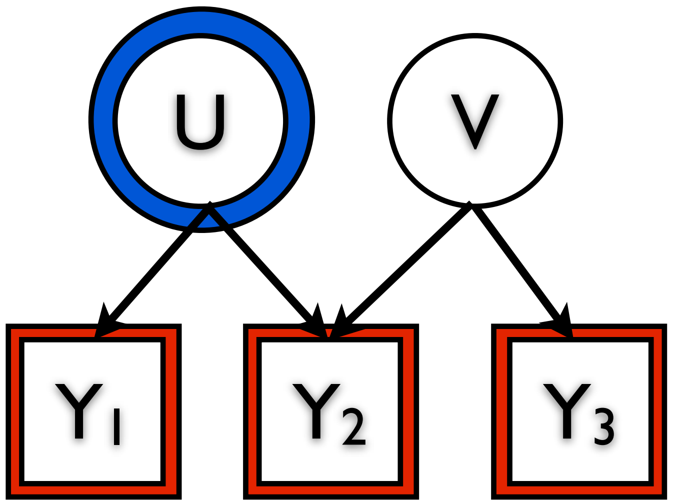
$p(Y_3|V) = N(Y_3|V, 1/b_3)$

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

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

Graphical models

Data fusion



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

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

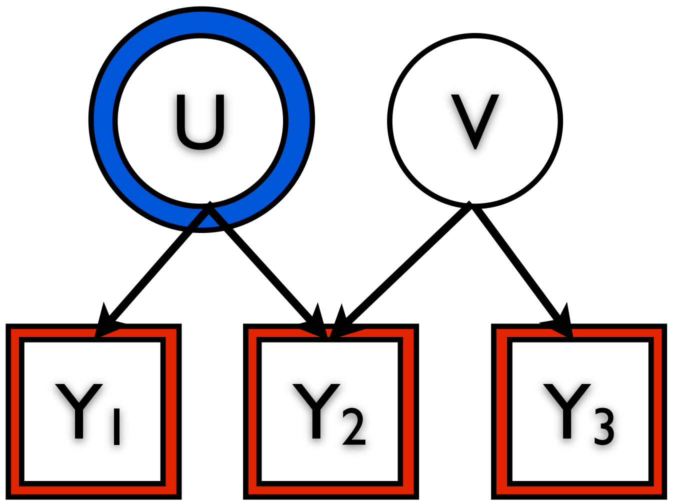
$$p(U,V,Y_1,Y_2,Y_3) = p(Y_1|U)p(Y_2|U,V)p(Y_3|V)p(U)p(V)$$

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

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

Graphical models

Data fusion



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

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

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

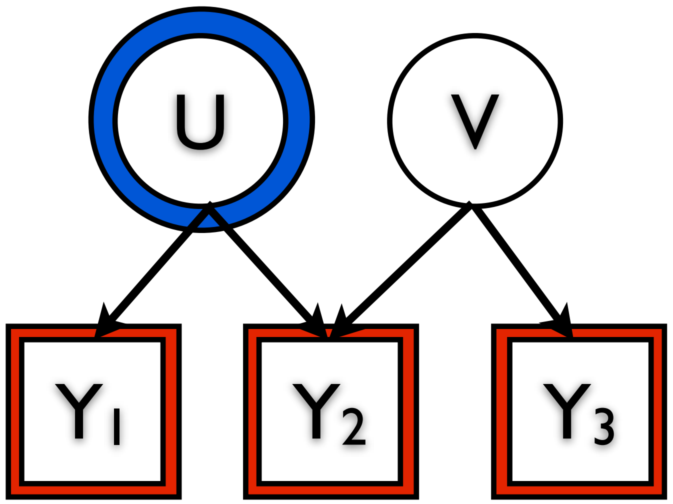
$$p(U|Y_1,Y_2,Y_3) \propto \int p(U,V,Y_1,Y_2,Y_3)dV$$

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

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

Graphical models

Data fusion



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

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

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

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

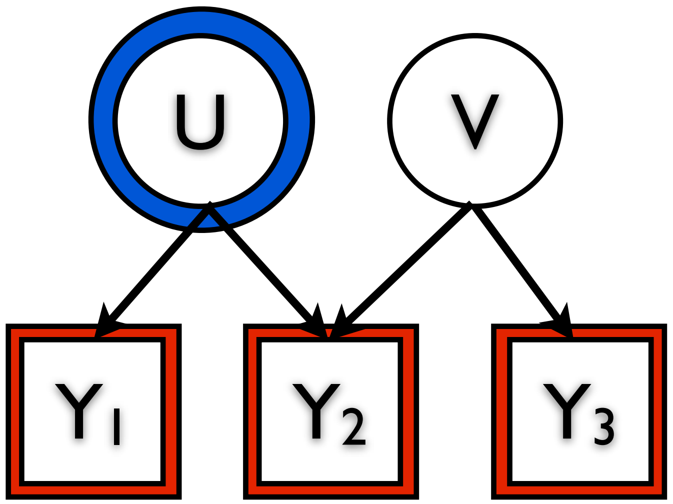
$$p(U|Y_1,Y_2,Y_3) \propto \int p(Y_1|U)p(Y_2|U,V)p(Y_3|V)dV$$

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

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

Graphical models

Data fusion



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

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

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

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

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

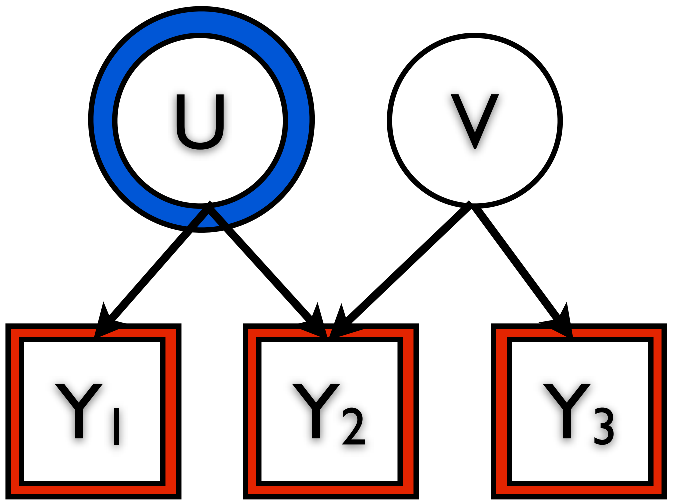
$$p(U|Y_1,Y_2,Y_3) \propto \int N(Y_1|U,1/b_1)N(Y_2|U+V,1/b_2)N(Y_3|V,1/b_3)dV$$

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

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

Graphical models

Data fusion



$Y_1 = U$ +noise

$Y_2 = U+V$ +noise

$Y_3 = V$ +noise

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

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

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

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

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

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

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

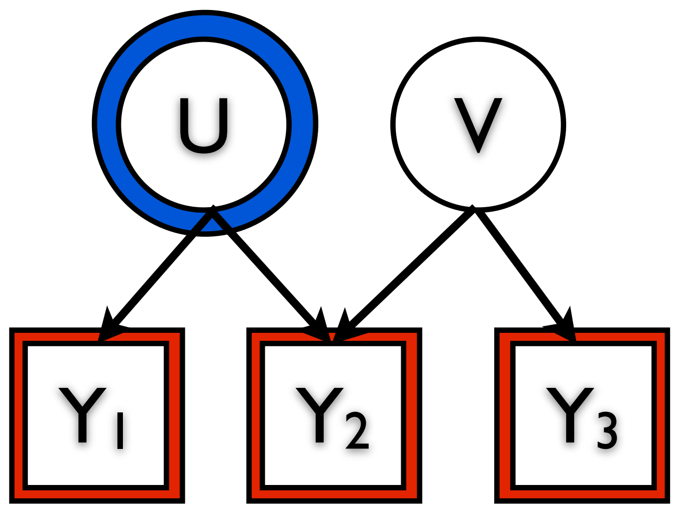
$$p(U|Y_1,Y_2,Y_3) \propto N(Y_1|U,1/b_1)\int N(V|Y_2-U,1/b_2)N(V|Y_3,1/b_3)dV$$

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

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

Graphical models

Data fusion



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

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

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

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

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

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

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

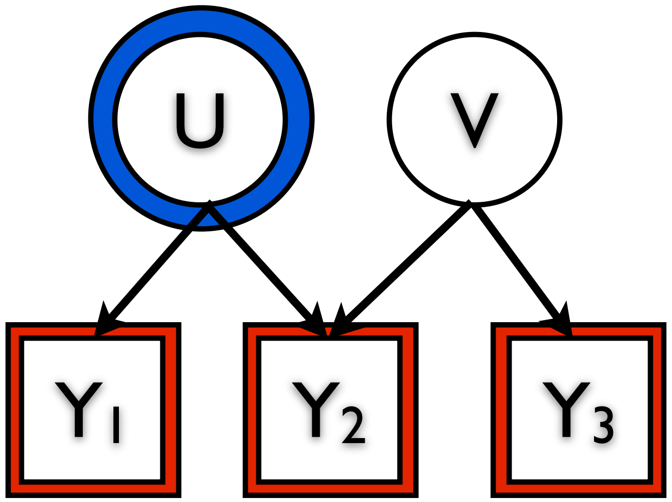
$$p(U|Y_1,Y_2,Y_3) \propto N(Y_1|U,1/b_1)\int N(V|\ldots,\ldots)N(U|Y_2\text{-}Y_3,1/b_2+1/b_3)dV$$

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

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

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

Let's stare at this posterior



$Y_1 = U$ +noise

$Y_2 = U + V$ +noise

$Y_3 = V$ +noise

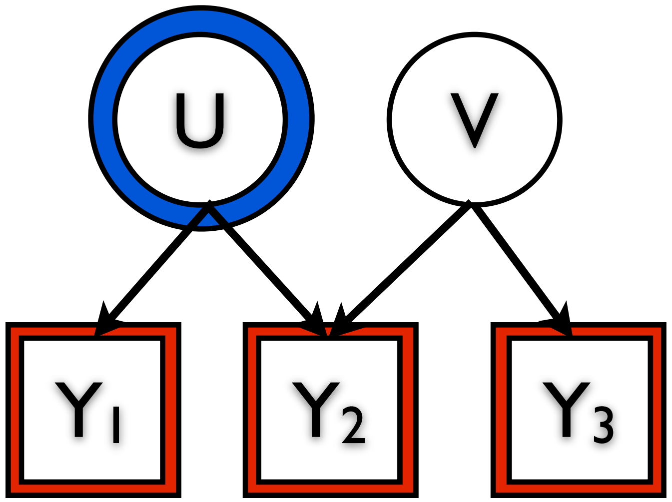
$$p(U|Y_1, Y_2, Y_3)$$

$$=$$

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

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

Let's stare at this posterior



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

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

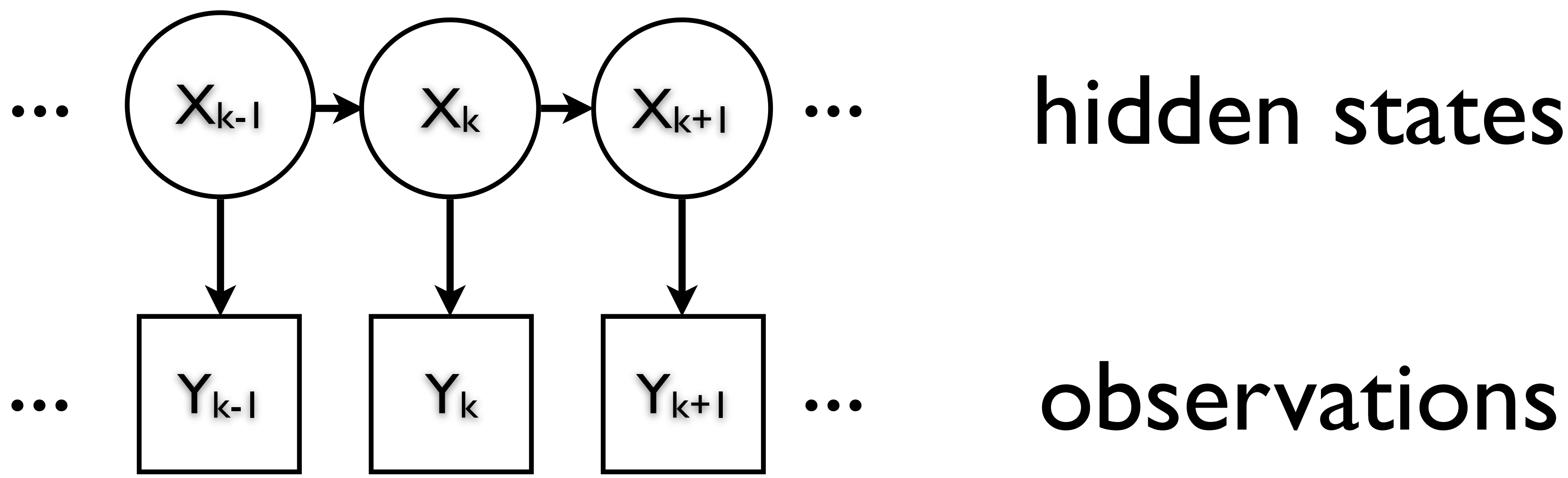
$b_{23} = b_2 b_3 / (b_2 + b_3)$

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

Outline

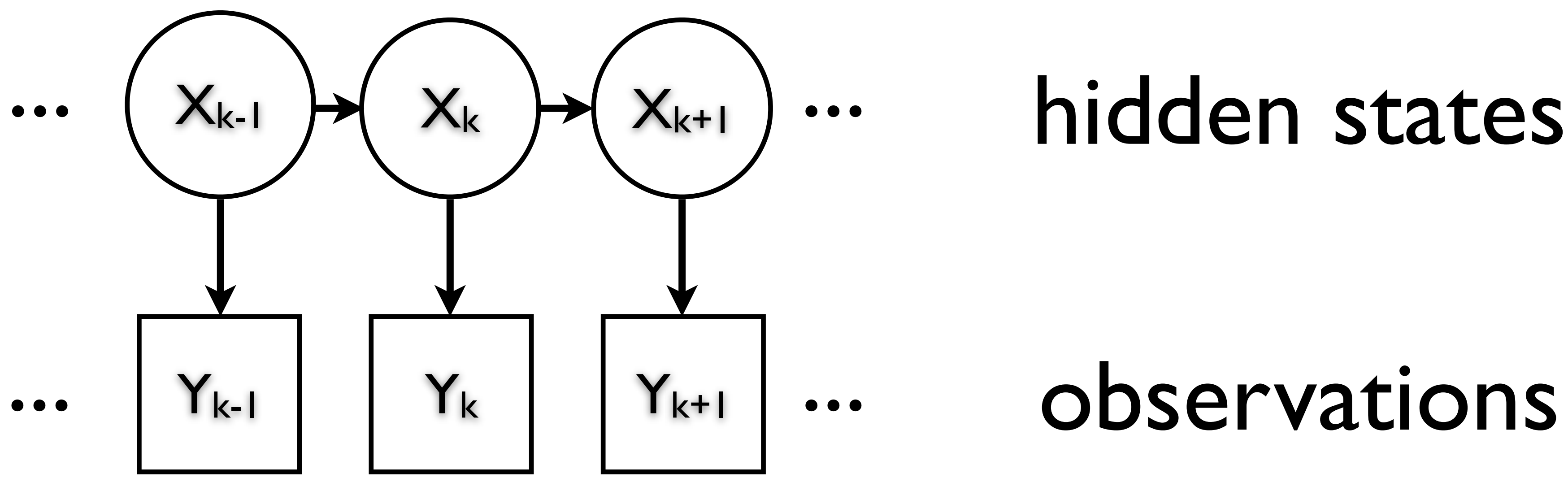
- Joint/Conditional/Marginal
- Bayesian graphical models
- Designing a model
- Bayes in FSL
- Derivation
    - Data fusion
    - Kalman Filter

Markov models



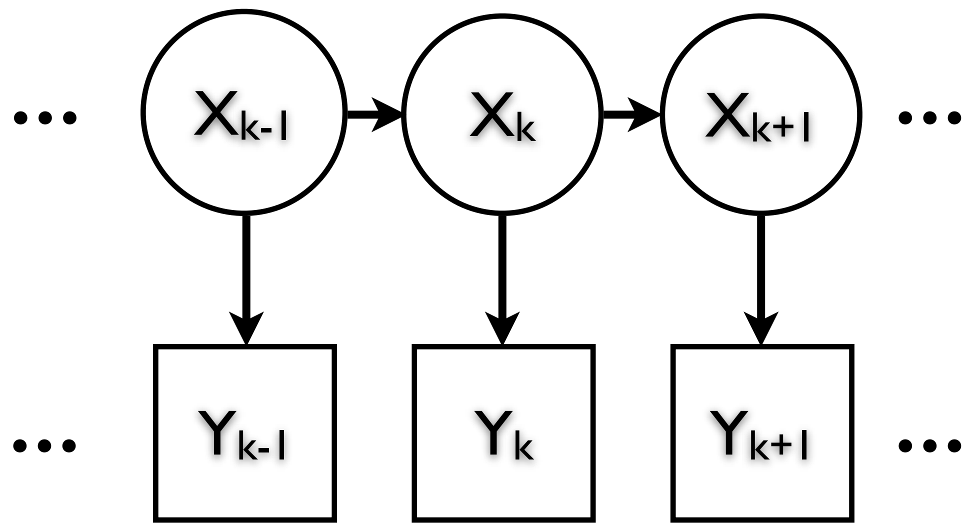
x_k discrete : Hidden Markov Model (HMM)
x_k continuous : State-Space Model

Markov models



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

Markov models

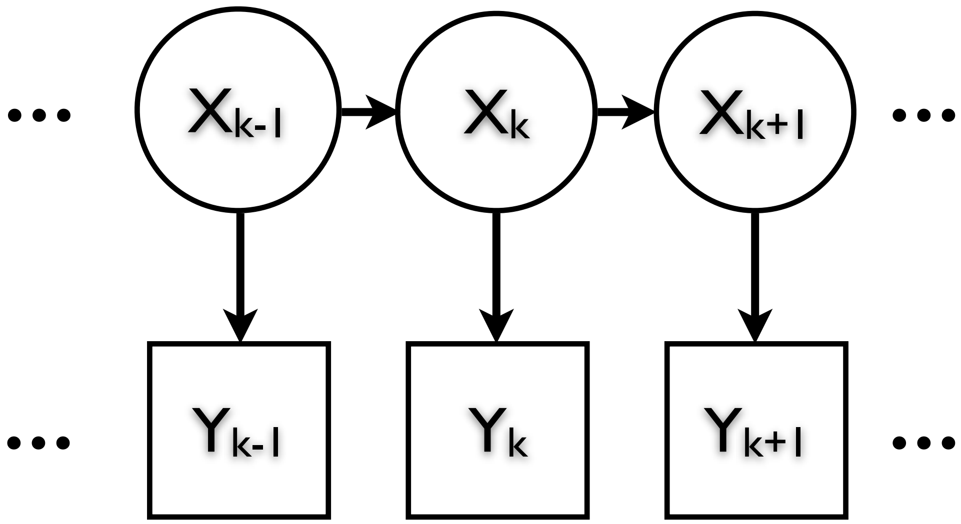


$p(Y_k|X_k) = N(Y_k \mid X_k, 1/b)$
$p(X_k|X_{k-1}) = N(X_k \mid X_{k-1}, 1/b_0)$

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"

Markov models



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

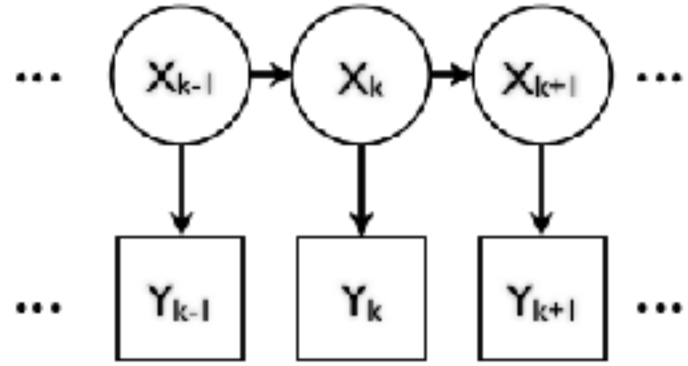
equivalently

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

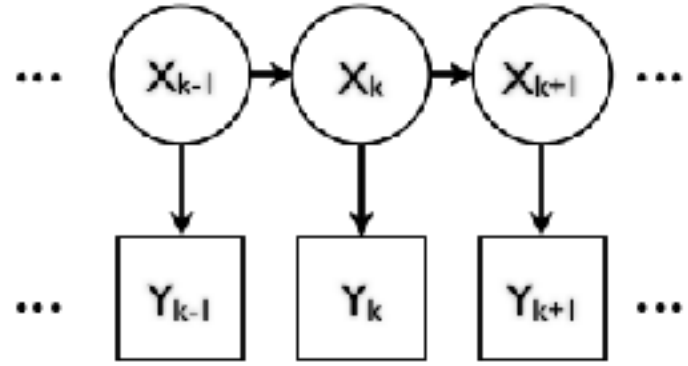
$p(X_k|Y_1,\ldots,Y_k) = ?$

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



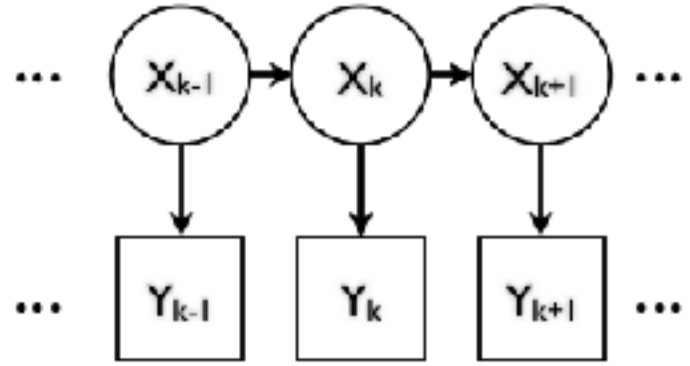
(1) $p(Y_k|X_k) = N(Y_k \mid X_k, 1/b)$
(2) $p(X_k|X_{k-1}) = N(X_k \mid X_{k-1}, 1/b_0)$



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

Bayes' rule

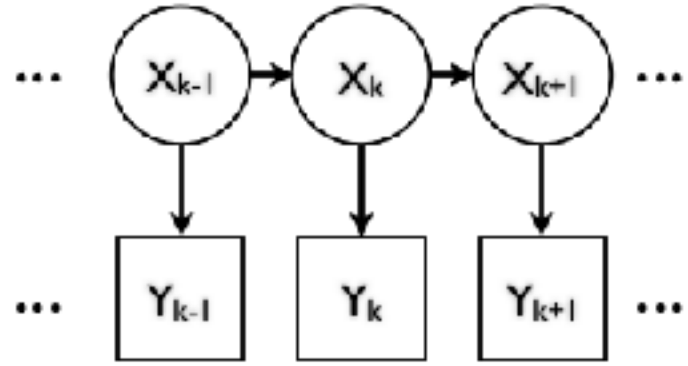
(1) $p(Y_k|X_k) = N(Y_k \mid X_k, 1/b)$
(2) $p(X_k|X_{k-1}) = N(X_k \mid X_{k-1}, 1/b_0)$



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

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

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

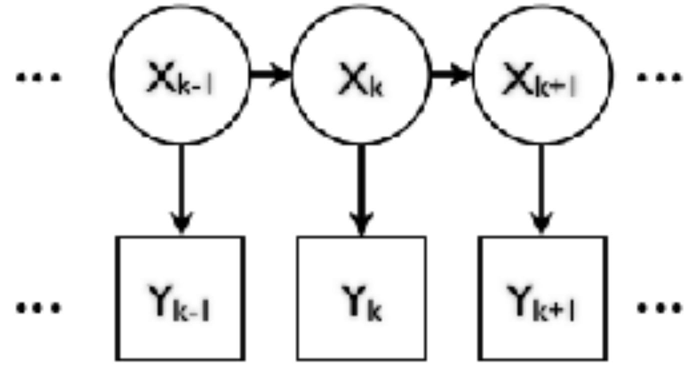


$$p(X_k|Y_1,\ldots,Y_{k-1},Y_k) = p(Y_k|X_k,Y_1,\ldots,Y_{k-1})\, p(X_k|Y_1,\ldots,Y_{k-1})$$ Bayes' rule

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

(1) $p(Y_k|X_k) = N(Y_k \mid X_k, 1/b)$

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

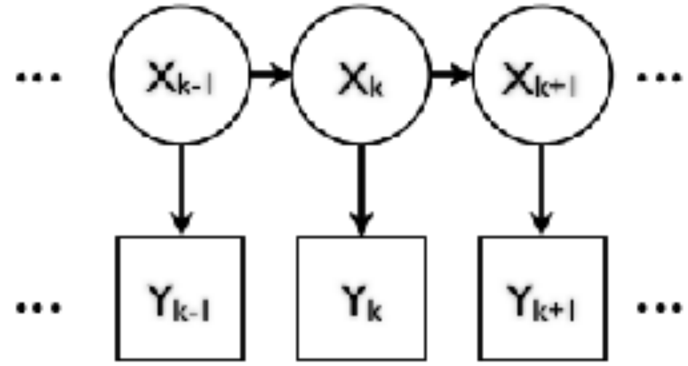


$$p(X_k|Y_1,\ldots,Y_{k-1},Y_k) = p(Y_k|X_k,Y_1,\ldots,Y_{k-1})\, p(X_k|Y_1,\ldots,Y_{k-1})$$ Bayes' rule

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

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

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



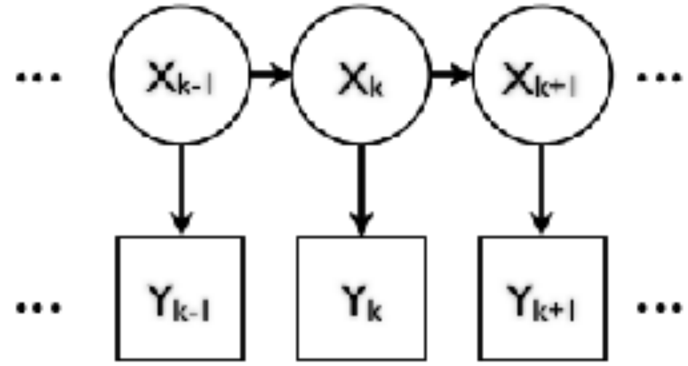
$$p(X_k|Y_1,\dots,Y_{k-1},{\color{red}Y_k}) = p({\color{red}Y_k}\,|X_k,Y_1,\dots,Y_{k-1})\; p(X_k|\,Y_1,\dots,Y_{k-1})$$ Bayes' rule

$$p(X_k|Y_1,\dots,Y_{k-1},{\color{red}Y_k}) = p({\color{red}Y_k}\,|X_k,\sout{Y_1,\dots,Y_{k-1}})\; p(X_k|\,Y_1,\dots,Y_{k-1})$$ Graph

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

(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

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

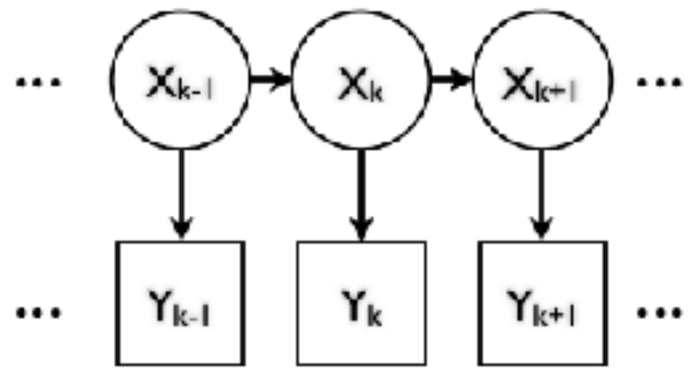
Graph

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

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

Sum rule

(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

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

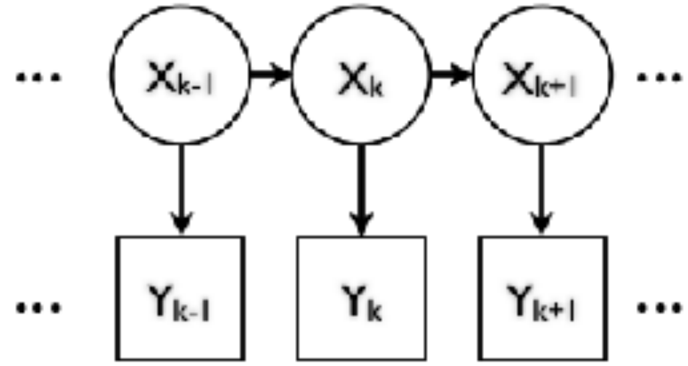
$$p(X_k|Y_1,\ldots,Y_{k-1}) = \int p(X_k|X_{k-1},Y_1,\ldots,Y_{k-1})p(X_{k-1}|Y_1,\ldots,Y_{k-1})\, dX_{k-1}$$

Sum rule

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

Graph

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



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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$
\begin{aligned}
p(X_k|Y_1,\ldots,Y_{k-1}) &= \int p(X_k|X_{k-1})p(X_{k-1}|Y_1,\ldots,Y_{k-1})\, dX_{k-1} \\
&= \int N(X_{k-1}|X_k,1/b_0)N(X_{k-1}|a_{k-1},1/b_{k-1})\, dX_{k-1} \\
&= \int N(X_{k-1}|\ldots,\ldots)N(X_k|a_{k-1},1/b_0+1/b_k)\, dX_{k-1} \\
&= N(X_k|a_{k-1},1/b_0+1/b_{k-1})
\end{aligned}
$$

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

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

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

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

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

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 = $1/b_{k-1}+1/b_0$

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 = $1/b_{k-1}+1/b_0$

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

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 = $1/b_{k-1}+1/b_0$

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

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

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

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

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

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

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

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

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

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

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

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 = $1/b_{k-1}+1/b_0$

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

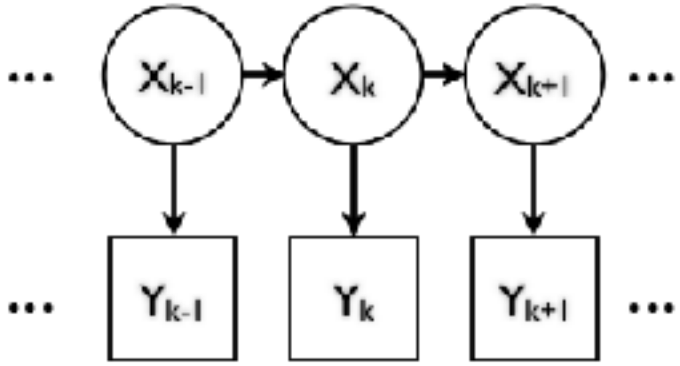
$$= N(X_k|Y_k|1/b)\, N(X_k|a_{k-1}, 1/\beta_{k-1})$$

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

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

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

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



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

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

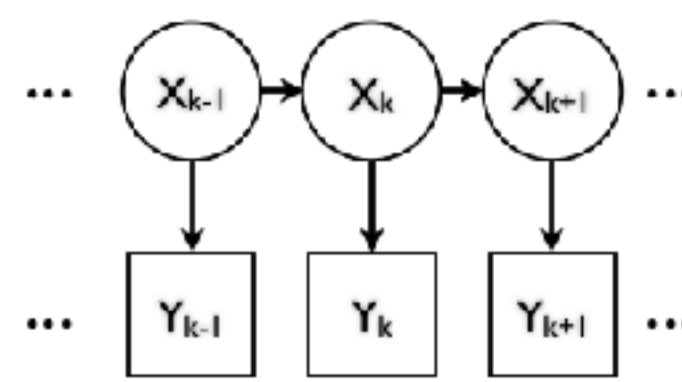
$$a_k = a_{k-1} + b/b_k(Y_k - a_{k-1})$$

update: new mean is previous mean + weighted error term.

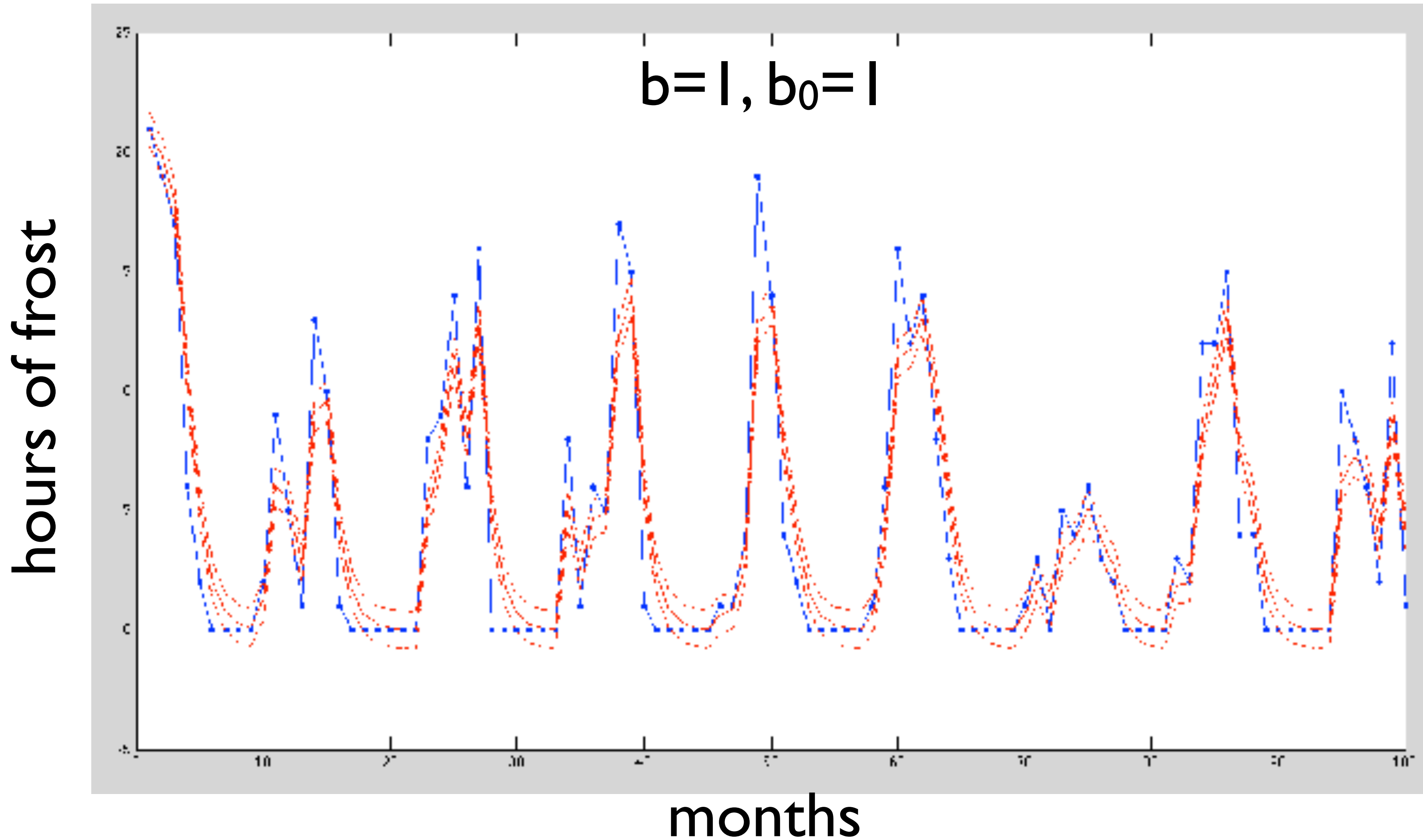
$$b_k = b + b_0 b_{k-1}/(b_0 + b_{k-1})$$

(1) $p(Y_k|X_k) = N(Y_k \mid X_k, 1/b)$

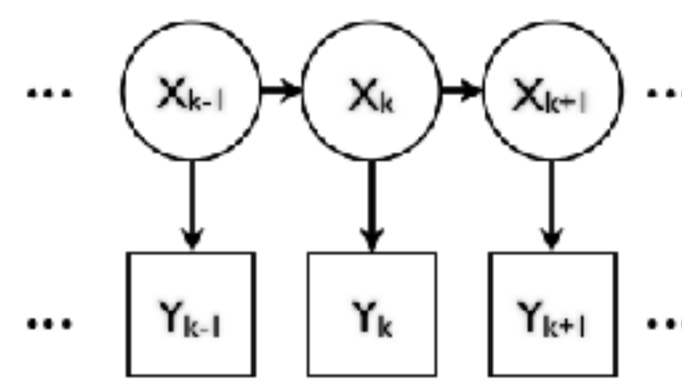
(2) $p(X_k|X_{k-1}) = N(X_k \mid X_{k-1}, 1/b_0)$



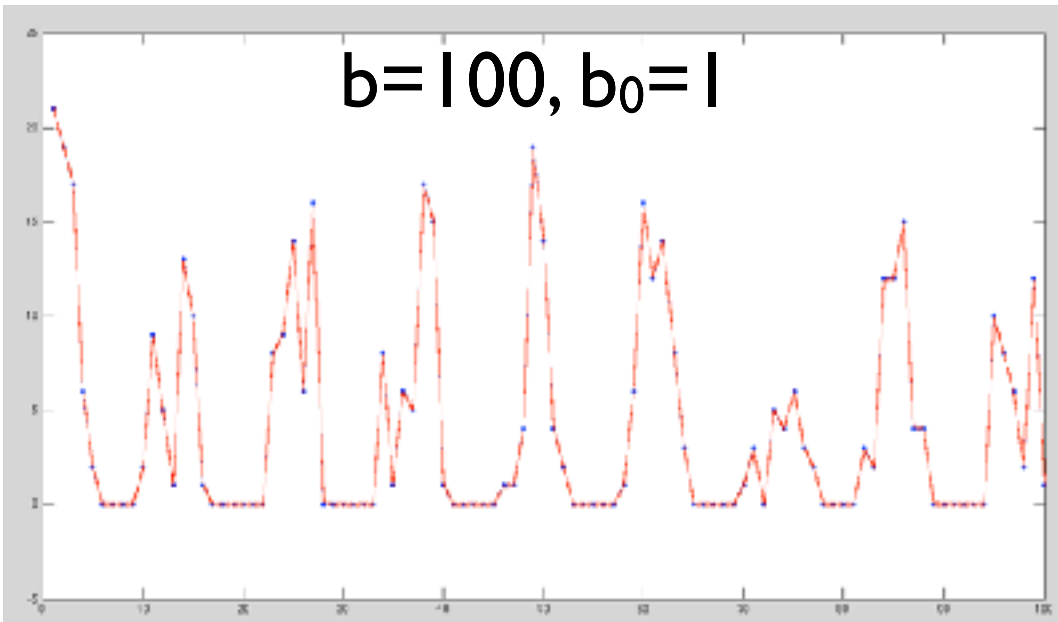
weather data - Oxfordshire (1929-1949)

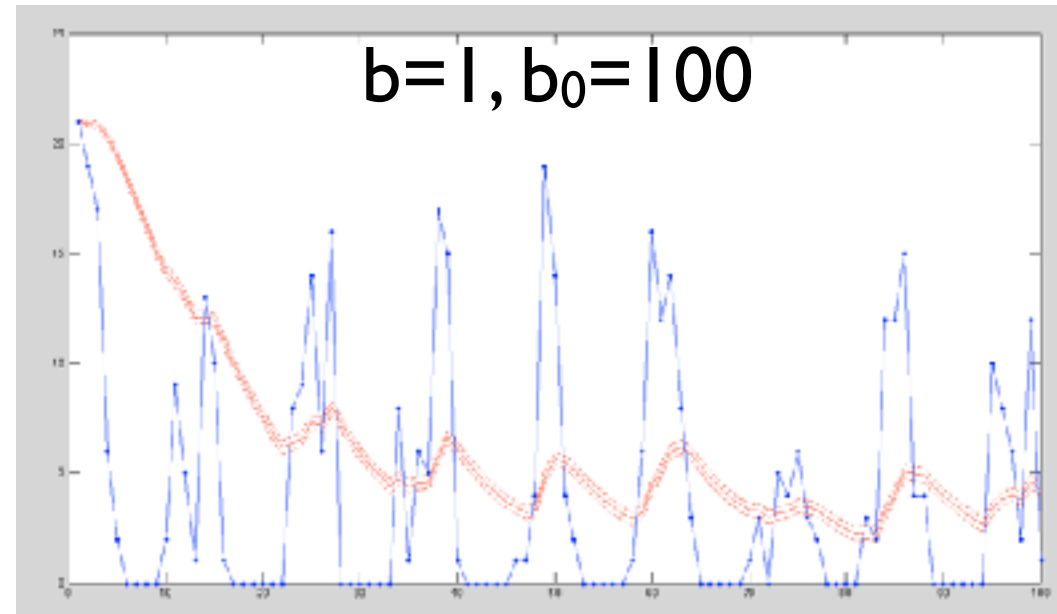


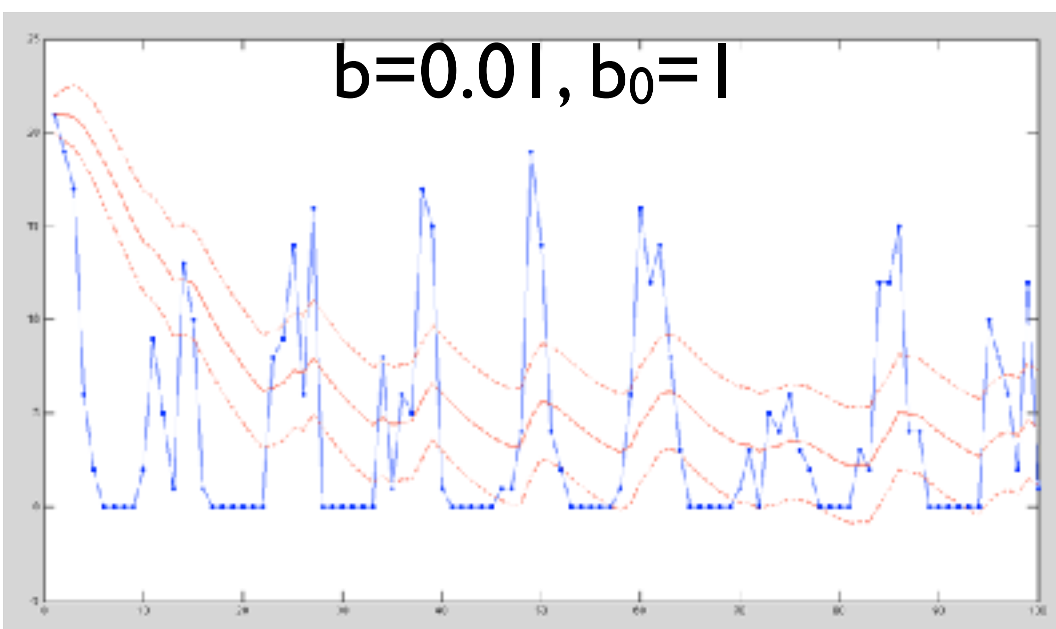
(1) $p(Y_k|X_k) = N(Y_k \mid X_k, 1/b)$
(2) $p(X_k|X_{k-1}) = N(X_k \mid X_{k-1}, 1/b_0)$

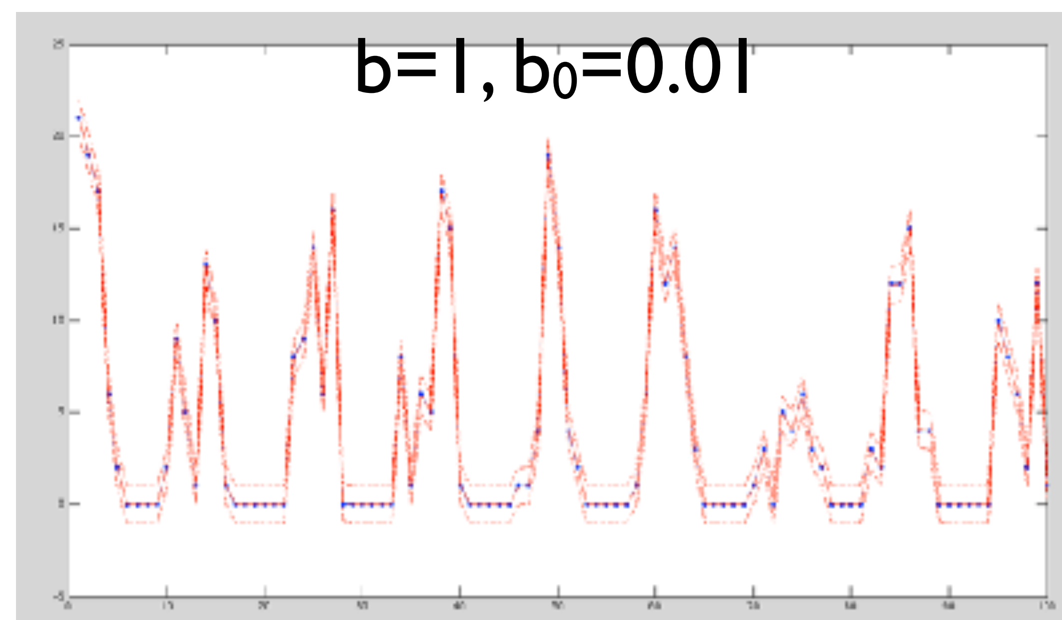


b=measurement noise precision
b0=state noise precision





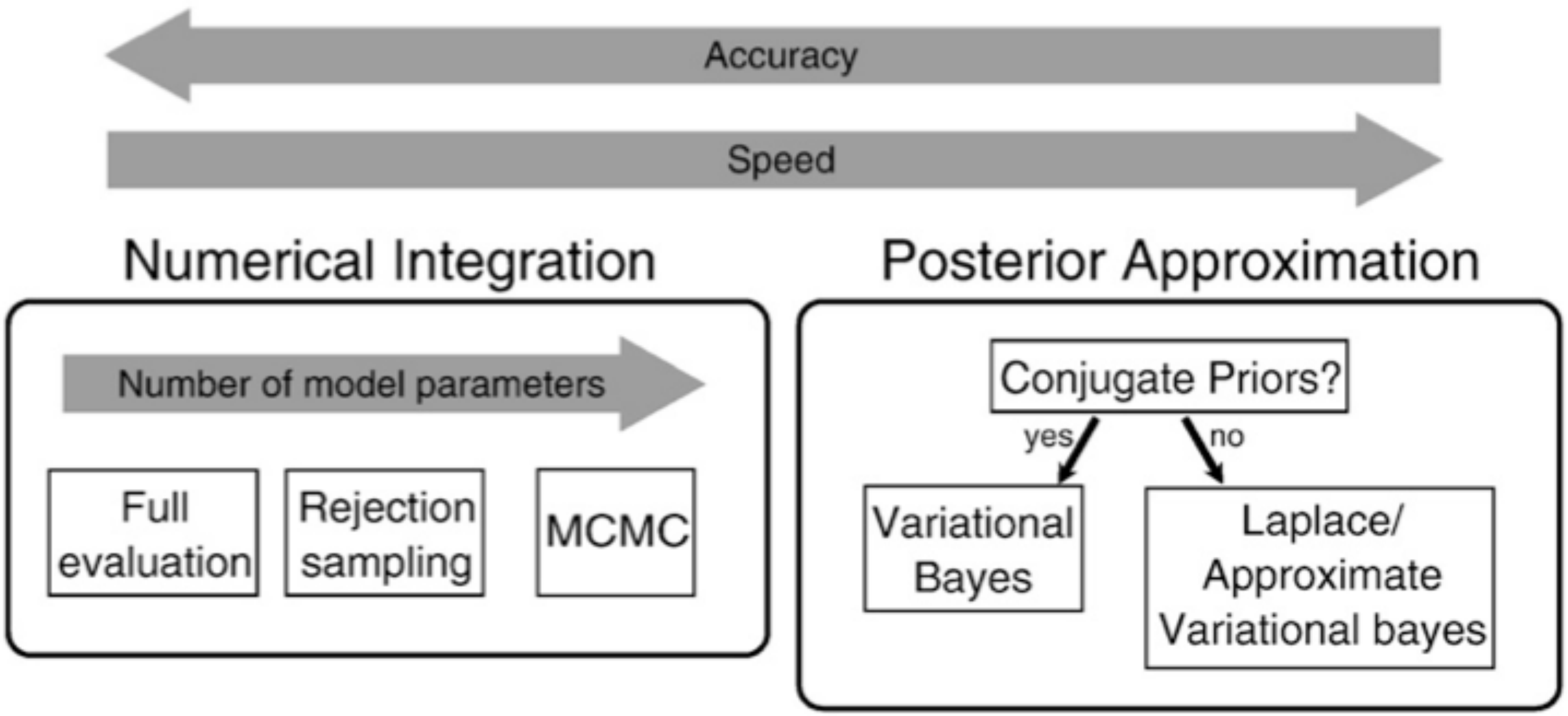




Ok - lots of integrals, what have we learnt?

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

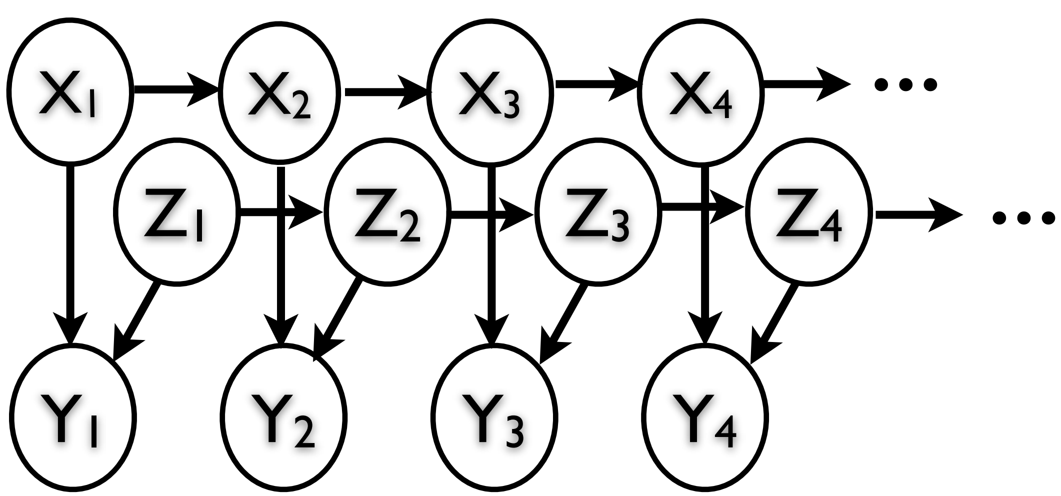
How to do inference?



Woolrich et al. 2009

Keep your model and your inference separate from each other!

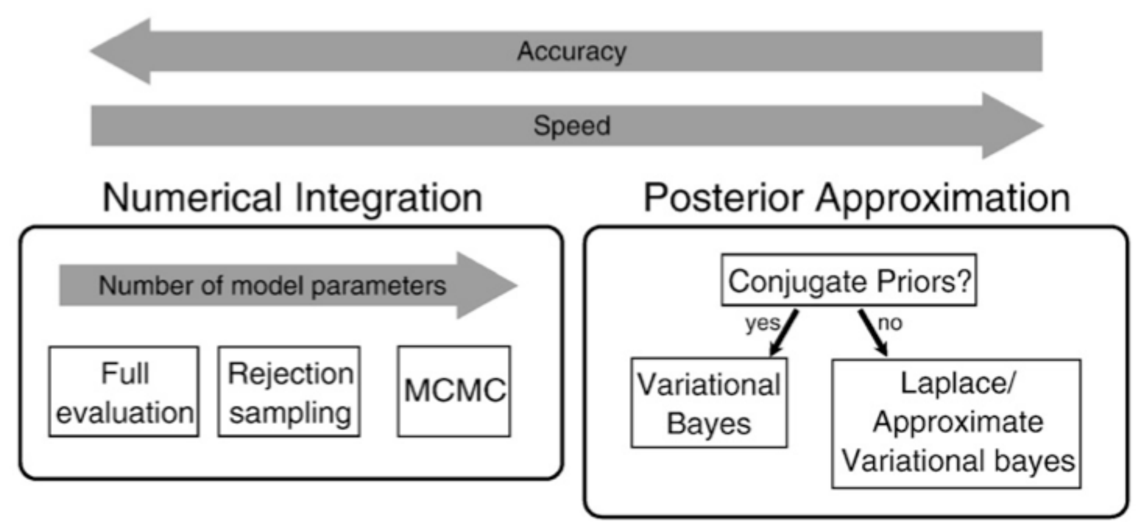
design model



write desired conditionals

e.g. $p(X_k|Y_1 \ldots Y_k)$

THEN decide on inference



The end.