



# Hierarchical Models

# Where We Are

- We now know how to carry out a Bayesian analysis for simple models.
  - t-tests
  - regression
  - use MCMC
- But most real problems are more complicated
- Now we find out how to deal with a lot of them!

# A Problem

- How big are Swedish frogs?
  - and how does the size vary?
- Variation at several levels
  - between individuals
  - between populations (i.e. ponds)
  - between regions
- Look at the population and individual levels
  - sample several frogs from different ponds

# How to Model

- We want to model the variation between individuals and between ponds
- Treat the pond mean as coming from a random distribution
  - e.g. Normal
  - Random effect
- Quantify the amount of variation

# The Model

- Individual  $i$  from pond  $j$
- $x_{ij}$  – Size (“Snout Ventral Length”)

$$x_{ij} \sim N(\mu_j, \tau)$$

- Common variance ( $1/\tau$ )
- Means differ between ponds
- Model this as a Normal distribution:
- $\mu_j \sim N(\mu_0, \tau_0)$

# Something Interesting

- For the pond means, we have this:

$$\mu_j \sim N(\mu_0, \tau_0)$$

- If we knew the means, this would be a normal inference
  - treating the  $\mu_j$ s as data
- We don't know them, but we estimate them:

$$x_{ij} \sim N(\mu_j, \tau)$$

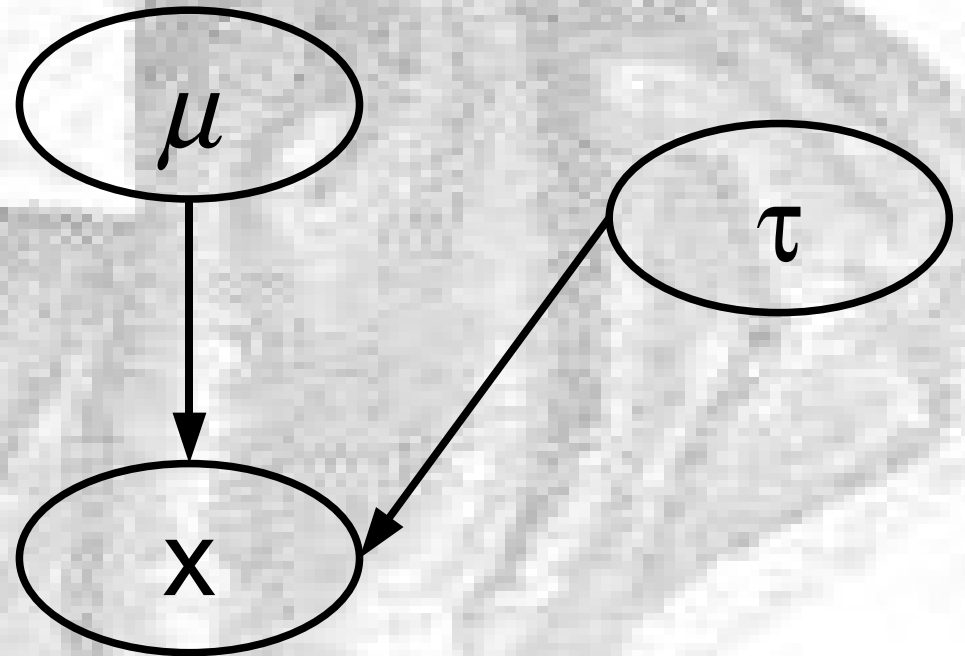
- The posterior for  $\mu_0$  and  $\tau$  is then integrated over the uncertainty in the  $\mu_i$ 's

# DAGs

- DAGs are very convenient ways of drawing hierarchical models

$$x \sim N(\mu, \tau)$$

Parent of  $x$



Child of  $\mu$  and  $\tau$

# Priors

- For the data ( $x_{ij}$ ), we can view the distribution of  $\mu_i$  as the prior for its mean
- So  $x_{ij}$  only depends on  $\mu_0$  and  $\tau$  through  $\mu_i$
- This makes things easier
  - we can follow the dependencies more easily
- This is a simple hierarchical model

# Hierarchical Models

- The essence of a hierarchical model is that the dependencies between variables can be defined easily
- Variables only depend on those below them
  - through the likelihood
- And those above them
  - through the prior
- This gives us a nice way of describing our models

# The Maths

- The full probability model can be worked out from the graph:
- If  $v$  is a stochastic node
  - i.e. an oval
- and  $V$  is the set of all nodes
  - i.e. all stochastic parts of the model
- Then  $P(V)$  is the probability density for the model, and:

$$Pr(V) = \prod_{v \in V} P(v | Parents[v])$$

# The Bayesian Bit

$$Pr(V) = \prod_{V \in \mathcal{V}} P(v | Parents[v])$$

- For Bayesian analysis,  $v$  is a parameter, so  $P(v)$  is the posterior distribution

- As we all know:

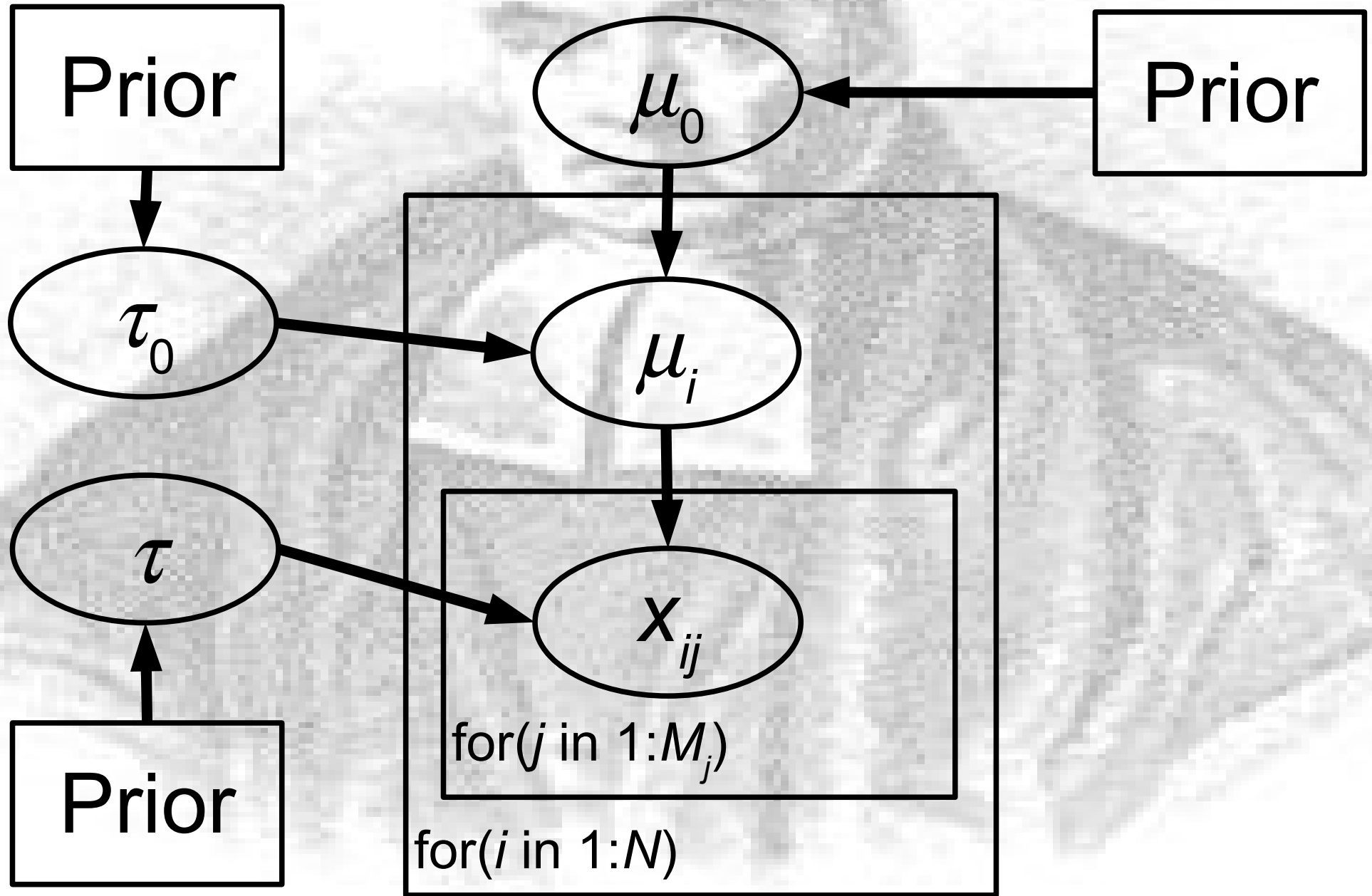
$$P(\theta | X) \propto P(X | \theta) P(\theta)$$

- So,  $P(v)$  depends on its parents...
  - through  $P(\theta)$
- ... and its children
  - through  $P(X | \theta)$

# Fitting the Model

- We can calculate  $P(v)$  [=  $P(\theta|X)$ ] conditionally on its parents and children
- Given these, it is independent of the rest of the model
- Therefore, we can use a Gibbs sampler to estimate the parameters
  - Gibbs sampler = cunning MCMC method
  - see last lecture

# Frogs: The Model



# Summarising the Results

- We can summarise the amount of variation that is regional by the proportion of total variation due to the region =  $\rho$ :

$$\rho = \frac{1/\tau_0}{1/\tau + 1/\tau_0}$$

– “intra-class correlation”

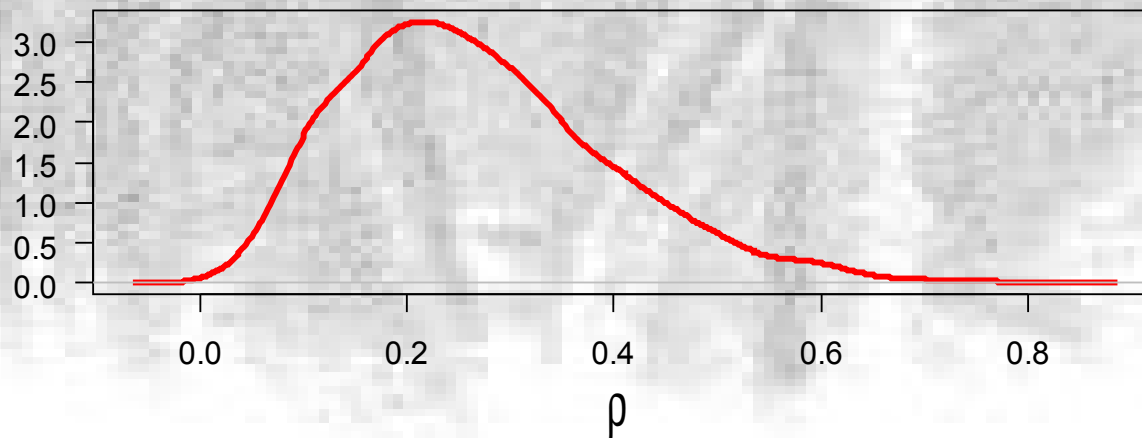
- To calculate this we use each draw from the posterior for  $\tau$  and  $\tau_0$  and calculate it from this...

# The Estimation

- If we have  $\tau$  and  $\tau_0$ , we calculate  $\rho$ :

$\tau$	$\tau_0$	$\rho$
1.380	4.347	0.241
1.748	3.421	0.338
1.842	1.939	0.487
1.598	3.429	0.318
1.947	3.084	0.387

- And this gives the correct posterior:



# More Models

- The response does not have to be normally distributed
- Some alternatives:
  - Binomial
  - Poisson
  - Negative Binomial
  - Gamma
- Likelihood different, but use same approach
  - extend generalised linear models

# Frog Survival

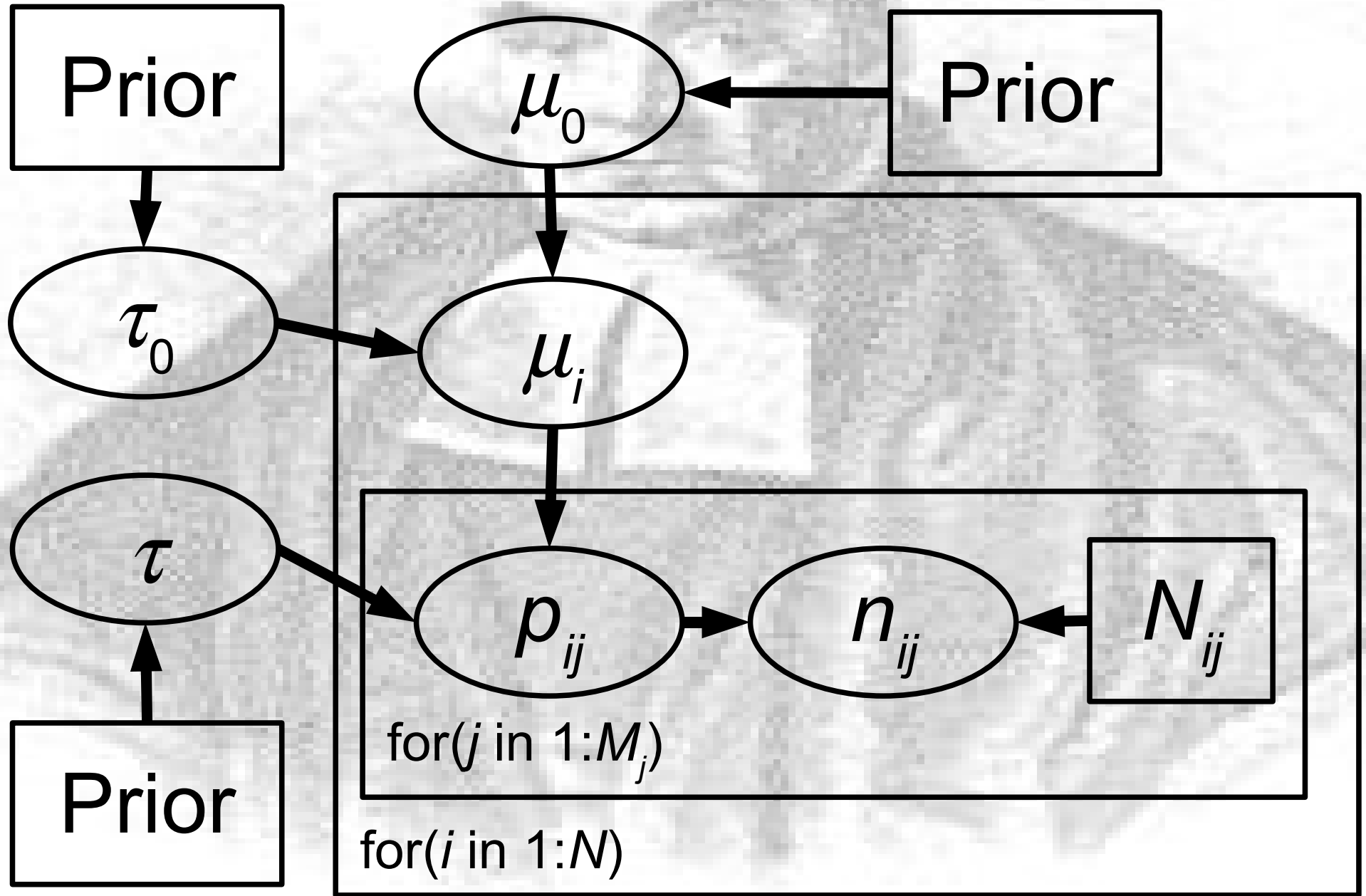
- Rather than look at size, we can look at survival of offspring
- Treat this as a trait, like size
  - each offspring has a probability  $p$  of surviving
- Of  $N_{ij}$  offspring from individual  $i$ , from population  $j$ ,  $n_{ij}$  survive

$$n_{ij} \sim \text{Bin}(N_{ij}, p_{ij})$$

- Then model logit  $p_{ij}$ :

$$\text{logit}(p_{ij}) \sim \text{N}(\mu_i, \tau)$$

# Frog Survival



# Quantitative Genetics

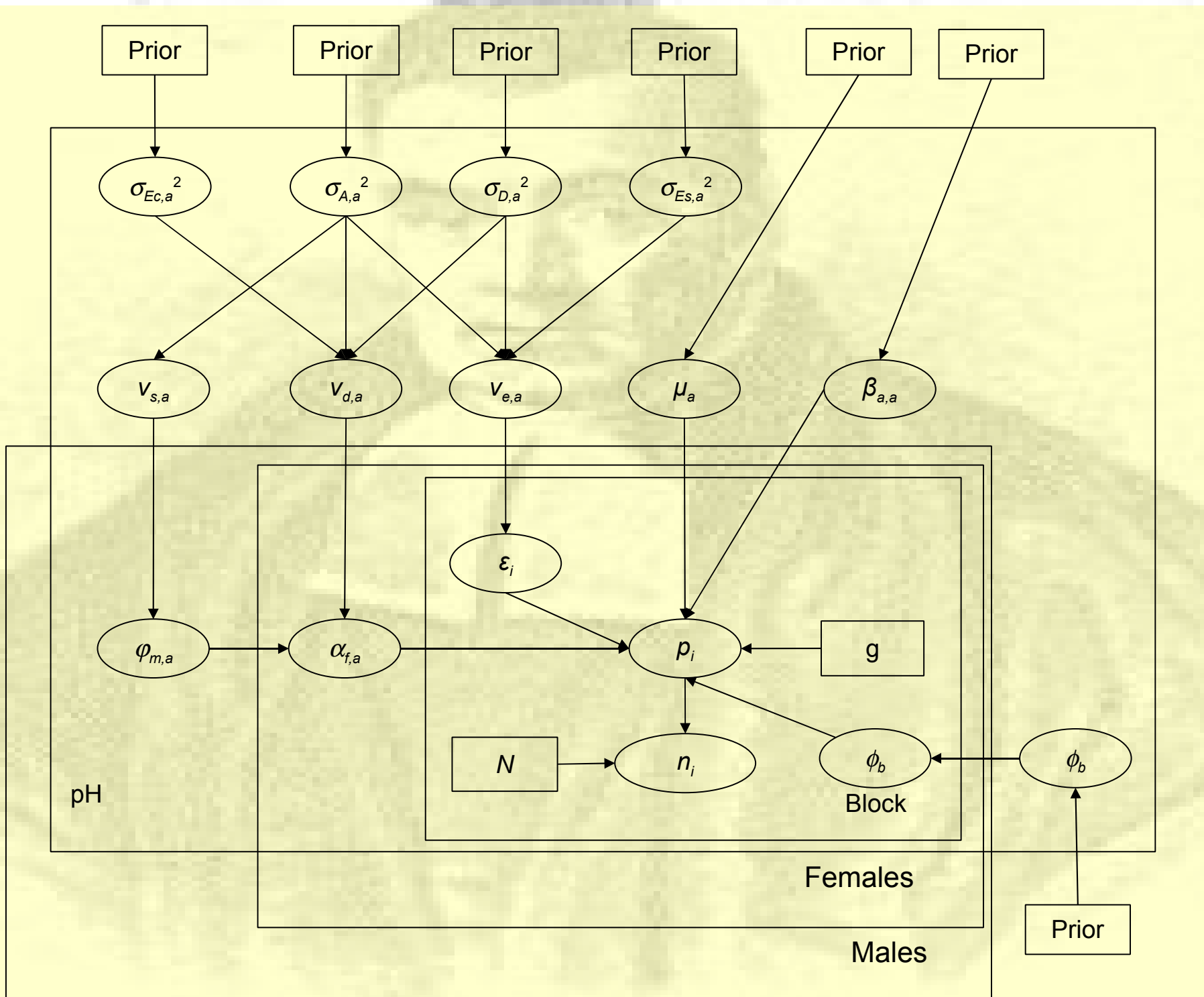
- Interested in genetics of variation in traits
- Set up crosses, look at offspring
- Partition variation into different components
  - Genetic (additive, dominance)
  - Maternal
  - Environmental
- Traditionally calculate by subtraction
- Bayesian approach: use all the information

# The Model

- The trait has a value  $\eta_i$  for individual  $i$ 
  - trait either measured, e.g. size
  - or latent, e.g. probability of surviving
- Look at variation between crosses
- Partition the total variation into variation
  - between female parents
  - between male parents
  - between full sibs

# Genetics Variances

- Each component of experimental variation is a sum of causal components
  - e.g. Sire = Additive/4
  - Dam = Additive/4 + Dominance/4 + Maternal
- We then put priors on Additive etc.
- Use all of the information
  - variation in posterior correct
- And now – the full model



# The Message

- Hierarchical models can get complicated
- We can keep control with the DAG
  - still able to fit them to the data
- Great flexibility available
- When do we stop adding to the model?
  - when we can no longer interpret the results
  - when the data is not rich enough
    - e.g. too few values