Applications

- Linear regression
- Nonlinear regression
- Generalized linear regression
  - Poisson
  - Binomial
- Hierarchical models
- These can contain lots of parameters, so the posterior distribution is always multidimensional.

Linear regression models

- **York rainfall data: x= in November, y= in December**

list(y = c(41,52,18.7,55,40,29.2,51,17.6,46.6,57),
x = c(23.9,43.3,36.3,40.6,57,52.5,46.1,142,112.6,23.7))

```
model{
for(i in 1:10){
y[i] ~ dnorm(mu[i],tau)
mu[i] <- beta[1] + beta[2]*x[i]                    # unstandardized x
# mu[i] <- beta[1]+ beta[2]*(x[i]-mean(x[]))   # standardized x
}
for(i in 1:2){
beta[i] ~ dnorm(0,0.001)
}
tau ~ dgamma(0.01,0.01)
# prediction with given fixed value xnew:
ynew ~ dnorm(munew,tau); munew <- beta[1] + beta[2]*xnew
# munew <-beta[1] + beta[2]*(xnew-mean(x[])) # standardized covariates
}
```

Interpretation of beta[1] in both cases? E(y | x=0) vs E(y|x=mean(x))

Linear regression models

- **Priors:**
- Uninformative, independent: $\pi(\beta_1)\ \pi(\beta_2)$ typically 'flat distributions'.
- Also possible: joint prior $\pi(\beta_1,\beta_2)$ could be multinormal distribution.
- Informative priors? Difficult to think directly regression parameters. Could think the observable outcome y* for a given explanatory variable x* and set a prior for this y* → solve regression parameters from this. → **'Induced prior for β'.** Needs as many priors as there are parameters.
- *Partially informative priors:* set standard uninformative priors for some parameters, but informative for others.
- Also: Can define functional constraints between parameters, and hierarchical structures.

Linear regression models

- **Standardization of explanatory variables X:**
- Can standardize as
  - (x-mean(x))
  - (x-mean(x))/sd(x)
- This can make Gibbs sampling more efficient, because it affects the posterior correlations of the regression parameters.

- See the effect in BUGS simulations…

Linear regression models

- **With prior $\pi(\beta,\tau) \propto 1/\tau$ the conditional posterior $\pi(\beta \mid \tau,X,Y)$ of regression parameters β is:**

Normal($(X^TX)^{-1} X^T Y$, $(X^TX)^{-1} \sigma^2$)

- Here X is the design matrix, and $\beta^* = (X^TX)^{-1} X^T Y$ is also the same as **least squares estimate** of regression parameters β.

- **Posterior $\pi(\tau \mid X,Y)$ of precision parameter τ is**

Gamma($(n-r)/2$, $(Y-X\beta^*)^T(Y-X\beta^*)/2$)

- This looks similar to the earlier shown posterior of μ,τ, based on normally distributed data X.

Linear regression models

- Sampling from the posterior could be done ’manually’ by simple Monte Carlo, in which τ is first sampled from this Gamma-density, and then β from the multivariate normal density, conditional on τ.
  - This could be done in R

- In BUGS, we can also try other priors which do not lead to the previous analytically solvable posterior…

Linear regression models

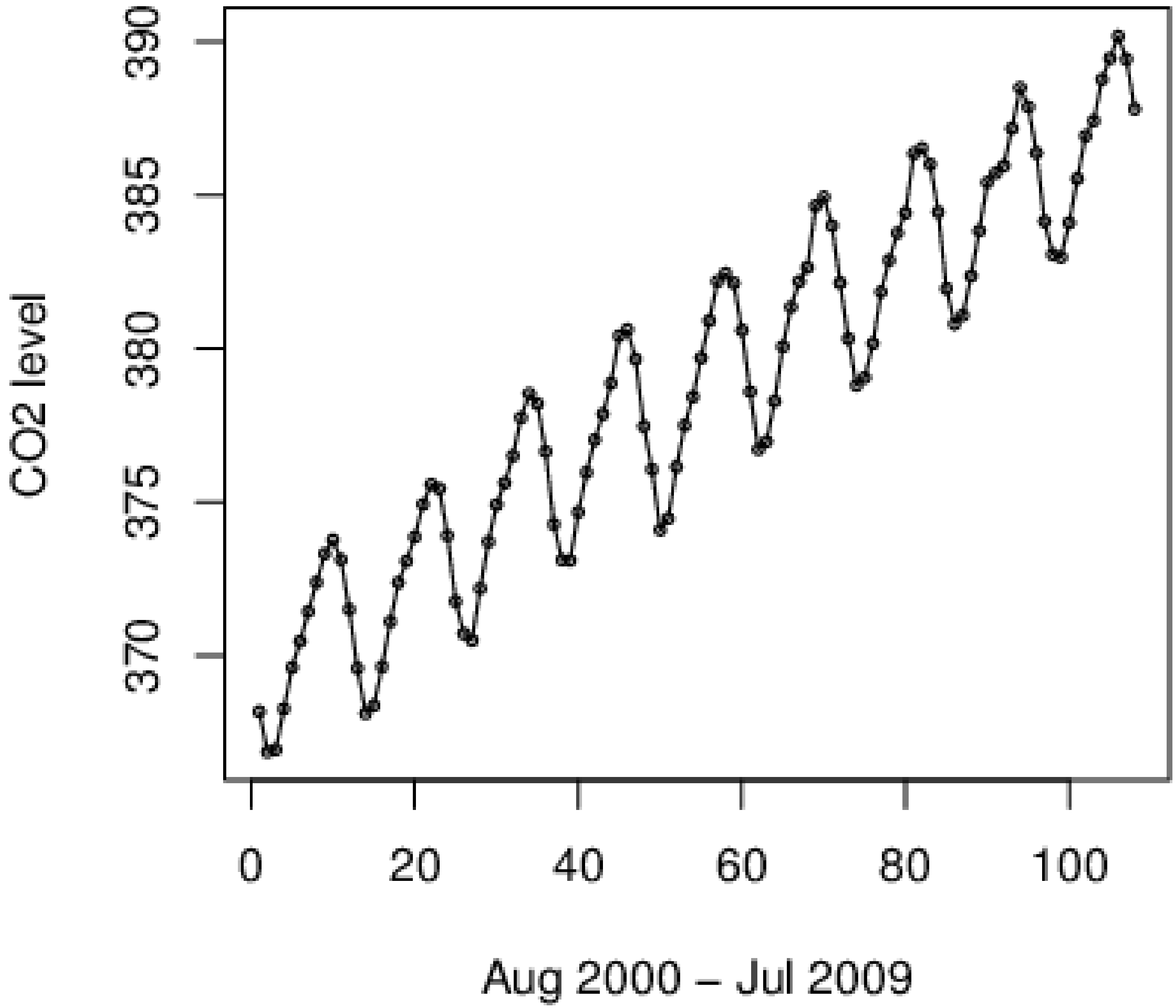
- Missing values occur in many application data sets!

- Missing values of Y are easy to handle. ('NA')
- Missing values of X would require an additional model structure, to give a well-defined conditional distribution for them.
- Bayesian "imputation technique" of missing values is to sample the missing values from the *joint posterior distribution, together with all other unknowns.*

Nonlinear regression models

- **An example with seasonal fluctuations: atmospheric CO_2, monthly, Mauna Loa, Hawaii**

list(N=120,x=c(368.18,366.87,366.94,368.27,369.62,370.47,
371.44,372.39,373.32,373.77,373.13,371.51,369.59,368.12,
368.38,369.64,371.11,372.38,373.08,373.87,374.93,375.58,
375.44,373.91,371.77,370.72,370.5,372.19,373.71,374.92,
375.63,376.51,377.75,378.54,378.21,376.65,374.28,373.12,
373.1,374.67,375.97,377.03,377.87,378.88,380.42,380.62,
379.66,377.48,376.07,374.1,374.47,376.15,377.51,378.43,
379.7,380.91,382.2,382.45,382.14,380.6,378.6,376.72,
376.98,378.29,380.07,381.36,382.19,382.65,384.65,
384.94,384.01,382.15,380.33,378.81,379.06,380.17,
381.85,382.88,383.77,384.42,386.36,386.53,386.01,
384.45,381.96,342,
385.72,385.96,387.18,388.5,387.88,386.38,384.15,
383.07,382.98,384.11,385.54,386.93,387.42,388.77,
389.46,390.18,389.43,387.81)


CO2 level
390
385
380
375
370
0
20
40
60
80
100
Aug 2000 – Jul 2009

Nonlinear regression models

- **Linear and nonlinear terms: trend + seasonality**

```
model{
tau ~ dgamma(0.01,0.01);
for(i in 1:5){a[i] ~ dnorm(0,0.001)}
for(i in 1:N){
month[i] <- i
x[i] ~ dnorm(mu[i],tau)
mu[i]<- a[1]+a[2]*i+a[3]*sin(2*pi*i/12)+a[4]*cos(2*pi*i/12)
}
pi <- 3.1415926
}
```

Generalized linear regression

- **Example of generalized linear Poisson modeling**
- Data:
  - Number of lung cancer cases $\mathbf{X}_{age,city}$
  - Population counts $\mathbf{pop}_{age,city}$
  - In age groups, in different cities, in 1968-1971.
- Model: (log-linear for $\lambda_{i,j}$ → link function)
  - Use the first age group in the first city as a reference, to define **age effects** and **city effects**
  - $\log(\lambda_{age,city}) = \mu_0 + \alpha_{age} + \alpha_{city}$, with $\alpha_{age=1} = \alpha_{city=1} = 0$
  - $\mathbf{X}_{age,city} \sim \mathbf{Poisson}(4\lambda_{age,city}\, \mathbf{pop}_{age,city})$

cases[] pop[] age[] city[]
11 3059 1 1
11 800 2 1
11 710 3 1
10 581 4 1
11 509 5 1
10 605 6 1
13 2879 1 2
6 1083 2 2
15 923 3 2 ← Case count (15) and population (923) in 3rd age group, in 2nd city
10 834 4 2
12 634 5 2
2 782 6 2
4 3142 1 3
8 1050 2 3
7 895 3 3
11 702 4 3
9 535 5 3
12 659 6 3
5 2520 1 4
7 878 2 4
10 839 3 4
14 631 4 4
8 539 5 4
7 619 6 4
END

design matrix X:

Parameter vector →

The first 10 rows of design matrix X would look like this.

Age1 and City1 are reference categories (baseline) against which Age2,… and City2,… are 'effects'.

log-incidence in the group "City2,Age3" would be

$\alpha_0 + \alpha_2 + \alpha_6$

Therefore, incidence in this group is the baseline multiplied by effects:

$Exp(\alpha_0)Exp(\alpha_2)\ Exp(\alpha_6)$

Baseline	Parameters for age effects					..and for city effects		
α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
Base	age2	age3	age4	Age5	age6	city2	city3	city4
1	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0
1	0	0	0	1	0	0	0	0
1	0	0	0	0	1	0	0	0
1	0	0	0	0	0	1	0	0
1	1	0	0	0	0	1	0	0
1	0	1	0	0	0	1	0	0
1	0	0	1	0	0	1	0	0

Linear predictor for any group is found by multiplying the parameter vector and the corresponding row of X.

BUGS ’tricks’ using design matrix

```
model{  # design matrix X could also be written beforehand in data
        # but it is here constructed from ’age’ and ’city’.
        # The linear predictor can then be computed using inprod.
for(i in 1:24){
cases[i] ~ dpois(mu[i]); group[i] <- i
mu[i] <- pop[i]*4*lambda[i]       # lambda = incidence per year
LA[i] <- lambda[i]/100000         # LA = inc. per 10^5 per year
log(lambda[i]) <- inprod(alpha[],X[i,])    # link function
X[i,1] <- 1
for(k in 2:6){X[i,k] <- equals(age[i],k) }
for(k in 2:4){X[i,k+5] <- equals(city[i],k) }
}
for(k in 1:9){  alpha[k] ~ dnorm(0,0.001)  # priors for all effect-parameters
                A[k] <- exp(alpha[k])  # multiplicative effects
               }
}
```

Generalized linear: Binomial

- Explanatory variables X for p
  - $Y_i \sim \text{Bin}(p_i, n_i)$
  - For each group i, there are variables X which are thought to explain p.
  - This needs some *link function* between p and effects α, for example logit:

    $\text{logit}(p_i) = \log(p_i/(1-p_i)) = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i3}$.
  - Or probit:

    $\text{probit}(p_i) = \Phi^{-1}(p_i) = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i3}$.

    Φ^{-1} is the inverse of cumulative probability for N(0,1)
  - X could be categorical or continuous or both.
  - Priors are set for parameters α.

Generalized linear: Binomial

- With these link functions, the data model (likelihood) is either

$$\pi(y \mid \alpha) = \prod_{i=1}^{n} \binom{n_i}{y_i} \left(\frac{e^{\eta_i}}{1+e^{\eta_i}} \right)^{y_i} \left(\frac{1}{1+e^{\eta_i}} \right)^{n_i - y_i}$$

or

$$\pi(y \mid \alpha) = \prod_{i=1}^{n} \binom{n_i}{y_i} \left(\Phi(\eta_i) \right)^{y_i} \left(1 - \Phi(\eta_i) \right)^{n_i - y_i}$$

Here η_i is the linear expression (real number)
= $\alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i3}$.

Generalized linear: Binomial

- Linear term could be extended by *random effects*

  $\alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i3} + \beta_j$

  $\beta_j \sim N(0,\sigma^2)$
- With a prior on σ^2. This could describe group specific ’random’ differences that are not well explained by the ’systematic’ effects X.
- This makes already a hierarchical model.

Example: O-rings

- Model the space shuttle O-ring failures as a function of temperature at launch.

$$\text{Logit}(p_i) = \alpha_0 + \alpha_1 X_i$$

- Here X is temperature.
- The observations are interpreted as binary indicators (failure=yes/no) to describe if any of the O-rings failed, for each flight.

Failure	Temp (F)
1	53
1	57
1	58
1	63
0	66
0	67
0	67
0	67
0	68
0	69
0	70
0	70
1	70
1	70
0	72
0	73
0	75
1	75
0	76
0	76
0	78
0	79
0	81

O-ring data

BUGS:

```
model{
for(i in 1:23){
Fail[i] ~ dbern(p[i])
logit(p[i]) <- a[1]+a[2]*T[i]
}
for(i in 1:2){a[i]~dnorm(0,0.001)}
}
```

Standardized T:

```
Ts[i] <- (T[i]-mean(T[]))/sd(T[])
```

Example: O-rings

- Freezing point is at F=32.
- Make prediction of the proportion of failures under F=31. (Temperature when Challenger exploded).

```
logit(p31) <- a[1]+a[2]*31
```

- Lowest observed Temp was F=53, so prediction should be uncertain because we are extrapolating long way down.

Example: O-rings

- **Informative prior approach:**
- Expert assessment on p, considering two temperatures 55F and 75F
  - The chosen temperatures should be ’enough’ apart from each other, so we could have independent opinion on both situations.
  - The chosen temperatures should be meaningful to the expert, so that there is an opinion about p at those temps.
  - The resulting matrix X should be nonsingular, so it can be inverted.
- logit(p55)=a[1]+a[2]*55
- logit(p75)=a[1]+a[2]*75

→ solve a[1] and a[2] from this...

Example: O-rings

- Solving the equations leads to

  a[1]=(75/20)*logit(p55)-(55/20)*logit(p75)

  a[2]=(-1/20)*logit(p55)+(1/20)*logit(p75)

- In matrix notation: $\alpha = X'^{-1} F^{-1} (p')$, where X' is the design matrix with chosen X-values, and p' is the corresponding vector of p, for which expert opinion is obtained, and F is the link function.
- Setting a prior on those p', *induces* a prior on paramerers α.

Example: O-rings

- As a result: we might have expert opinion which gives priors
  - p55 ~ Beta(1.6,1)
  - p75 ~ Beta(1,1.6)
- In BUGS, just write these priors for p55 and p75, and the parameters a[] are then simply a function of these

  a[1]← … and a[2]← …

Example: O-rings

- With the original x values, we solve a[] from $$\text{logit}\begin{pmatrix} p55 \\ p75 \end{pmatrix} = \begin{bmatrix} 1 & 55 \\ 1 & 75 \end{bmatrix}\alpha = X'\alpha$$

- With standardized values Z=(x-mean(x))/sd(x) we solve b[] from $$\text{logit}\begin{pmatrix} p55 \\ p75 \end{pmatrix} = \begin{bmatrix} 1 & -2.06 \\ 1 & 0.77 \end{bmatrix}\beta = Z'\beta$$

Because now the model is written with parameters β corresponding to the standardized values.

Default priors?

- Uninformative priors?
  - When no substantial prior knowledge available
  - Could use vague priors for probabilities p, corresponding to selected value combinations of explanatory variables, which induces prior for the regression parameters α.
  - For all k regression parameters, need k equations to be solved! (transform from $p_1,\ldots,p_k$ to $\alpha_1,\ldots,\alpha_k$)
  - Could use vague prior for regression parameters α
  - With small sample and/or true p near 0 or 1, different priors could cause bigger difference in posterior.

Small data & true p near 0 or 1?

-See effect with basic model-

```
model{
x ~ dbin(p[1],n[1]);     p[1] ~ dbeta(1,1)
y ~ dbin(p[2],n[2])
logit(p[2]) <- theta;
                theta ~ dnorm(0,tau); tau <-1/2.71
z ~ dbin(p[3],n[3])
logit(p[3]) <- eta;
                eta ~ dnorm(0,0.001)
}
list(x=4,y=4,z=4,n=c(10,10,10))
list(x=0,y=0,z=0,n=c(10,10,10))
```

See the effect of priors with different data

Usual Prior choices for α

- Improper flat priors $\pi(\alpha_i) \propto 1$ for all i.
- Vague normal priors $\pi(\alpha_i)$ =N(0,0.001) for all i. ($\tau=0.001$)
- Vague multinormal priors $\pi(\alpha_i,\ldots,\alpha_k)$ = MN(0,T)

- As a result, with logit(p) transformation these priors put most of the prior probability near 0 and near 1.
- *Usually* not much effect on posterior, but check this with sensitivity analysis.
- Possible recommendation: Normal priors for α with such variance that the induced prior on p will be closely uniform.

Hierarchical models

- **Example: hierarchical binomial model**
- Could be constructed in different ways:
- Basic model for observations is $X_i \sim \text{bin}(n_i,p_i)$ in groups i=1,…,n

- With prior for p:
- $p_i \sim \text{beta}(a,b)$ # variation between groups
- $\pi(a)$, $\pi(b)$ are some hyper prior densities.

- Or with prior for logit(p):
- $\text{logit}(p_i) \sim N(\mu,\sigma^2)$ # variation between groups
- $\pi(\mu)$, $\pi(\tau)$ are some hyper prior densities.

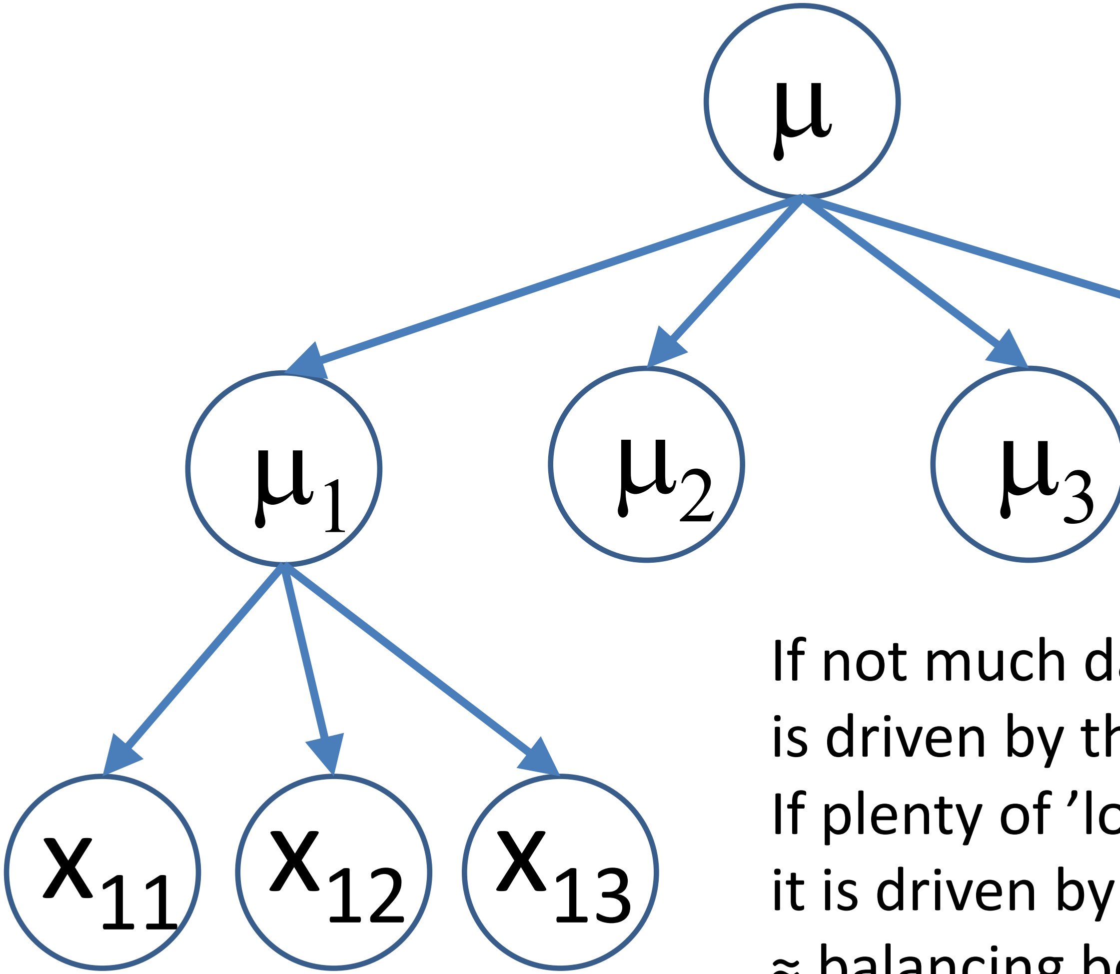
- The parameters for the hyper prior distribution are also unknown and to be estimated with all other parameters.

Hierarchical models

- **Example: hierarchical normal model**
- $X_i \sim N(\mu_i, \sigma^2_i)$
- $\mu_i \sim N(\mu, \sigma^2_0)$
- $\pi(\mu)$, $\pi(\sigma^2_0)$ are some hyper prior densities.

- Here, μ is the global (grand) mean, and μ_i is the mean of group i.
- Variance parameters describe between group variation and within group variation.
- Can make predictions for new groups, or new individuals within groups.
- By integrating over μ_i with respect to $N(\mu, \sigma^2_0)$ we get $X_i \sim N(\mu, \sigma^2_i + \sigma^2_0)$ so that $\sigma^2_i + \sigma^2_0$ = total variance.

Hierarchical models



If not much data for μ_3, its estimate is driven by the global information.
If plenty of 'local' data for μ_3,
it is driven by that.
≈ balancing between local data,
and global 'prior'.

Hierarchical models

- If not hierarchical model for hierarchical data, then what?
  - Could analyze each group *separately*
  - Could analyze all groups as *pooled*
  - Either way we *lose information*.
- Hierarchical model accounts for group specific differences, but borrows strength from all data.

  → e.g. evidence synthesis from multiple sources, meta-analyses, spatial smoothing, etc.

Hierarchical normal

- Assuming $\sigma^2_i=\sigma^2$, within all groups, so that mean(x_i) ~ N(μ_i, σ^2/n_i) and using new notation $\sigma^2/n_i = \sigma^2_i$, the structure is:

$$\text{level1}: \ N(x_{ij} \mid \mu_i, \sigma^2) \quad \text{, that is}: \ N(\bar{x}_i \mid \mu_i, \sigma_i^2), \ \text{where} \ \sigma_i^2 = \sigma^2 / n_i$$

$$\text{level 2}: \ N(\mu_i \mid \mu, \sigma_0^2)$$

- For simplicity, assume first that within group variance σ^2 is <u>known</u>.
- Posterior is then of the form:

$$\pi(\mu_1,\ldots,\mu_I,\mu,\sigma_0^2 \mid x) \propto \pi(\mu,\sigma_0^2)\prod_{i=1}^{I} N(\mu_i \mid \mu,\sigma_0^2)\prod_{i=1}^{I} N(\bar{x}_i \mid \mu_i,\sigma_i^2)$$

Hierarchical normal

- Note: although prior is hierarchical, this follows from Bayes theorem again.
- With these assumptions, some analytic results can be found:
  - The conditional distribution:

    $\pi(\mu_i \mid \sigma^2, \sigma^2_0, \mu, x) = N(\mu^*_i, V_i)$

$$\mu^*_i = \frac{\frac{1}{\sigma_i^2}\bar{x}_i + \frac{1}{\sigma_0^2}\mu}{\frac{1}{\sigma_i^2} + \frac{1}{\sigma_0^2}} \qquad V_i = \frac{1}{\frac{1}{\sigma_i^2} + \frac{1}{\sigma_0^2}}$$

  - It shows that the conditional expectation of group mean is a weighted average of μ and sample mean of the group (conditionally on σ^2, σ^2_0, μ, x).

Hierarchical normal

- Furthermore:
  - Level 2 -parameters μ and σ_0 have posterior of the form $\pi(\mu,\sigma_0|x) = \pi(\mu,\sigma_0)\pi(x|\mu,\sigma_0)/c$
  - Here the likelihood term can be difficult in general, (because it involves integration over unknown group means μ_i), but with Normal-models the following result applies: $\pi(\text{mean}(x_i))=N(\mu, \sigma^2_i+\sigma^2_0)$, so we can write $\pi(x|\mu,\sigma_0)$ as a product of these group specific likelihoods.
  - Using that form, and exploiting product rule which says $\pi(\mu,\sigma_0|x) = \pi(\mu|\sigma_0,x)\,\pi(\sigma_0|x)$, and with some manipulations, we find a solution for $\pi(\mu|\sigma_0,x)$

Hierarchical normal

- The solution is: $\pi(\mu \mid \sigma_0, x) = N(\mu^*, V)$ where

$$\mu^* = \frac{\sum \frac{\bar{x}_i}{\sigma_i^2 + \sigma_0^2}}{\sum \frac{1}{\sigma_i^2 + \sigma_0^2}} \qquad V^{-1} = \sum \frac{1}{\sigma_i^2 + \sigma_0^2}$$

- It shows the conditional expectation of grand mean μ is a weighted average of group specific sample means.

- Finally: the marginal density of between group variance σ_0^2 does not come out as a standard density. As an uninformative prior we could use $\pi(\sigma_0)$ = const, but the prior $\pi(\log(\sigma_0))$=const leads to improper posterior. → A prior $\tau_0 \sim$ Gamma(0.001,0.001) is nearly the same but (barely) proper. Some problems could occur if number of groups is small or if between group variance is small. Then: recommended to use e.g. flat prior for σ_0.

Hierarchical binomial

- For the **hierarchical binomial model**, with beta-prior for p_i, similar issues:
  - Joint distribution of hyper parameters α,β is of the form $\pi(\alpha,\beta|x) = \pi(\alpha,\beta)\pi(x|\alpha,\beta)/c$
  - The 2nd term (likelihood) can even be expressed as

$$\prod \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+x_i)\Gamma(\beta+n_i-x_i)}{\Gamma(\alpha+\beta+n_i)}$$

Hierarchical binomial

- A possible prior (by Gelman et al.) would be to set prior for logit($\alpha/(\alpha+\beta)$)=log(α/β) and log($\alpha+\beta$).

- But an improper uniform prior on these yields an improper posterior.

- Practical approach: check numerically by plotting the contours of the joint posterior, or by trying to simulate from it. If improper, this should be noticed → countour lines drift to infinity, simulations do not converge…. (note that also a proper distribution can be almost improper if the tails of the distribution go to zero very slowly, …too slowly)

Hierarchical normal example

- log-bacteria counts in 7 samples from each of the 15 batches:

  (simulated data based on real data)

batch specific observations (7 per batch, from 15 batches):

x[,1]	x[,2]	x[,3]	x[,4]	x[,5]	x[,6]	x[,7]	x[,8]	x[,9]	x[,10]	x[,11]	x[,12]	x[,13]	x[,14]	x[,15]
1.9	2.1	1.1	2.5	3.2	3.2	2.9	2.8	3.4	2.3	2.3	2.5	2.1	2.4	1.4
2.6	2.9	1.3	3.1	1.8	2.6	3.6	2.7	3.5	2.7	3.0	3.1	2.7	3.5	2.0
2.9	1.8	1.6	2.4	3.3	3.6	2.1	2.0	2.7	3.0	2.5	2.1	2.6	3.3	1.5
1.8	1.4	1.8	2.8	3.6	2.9	2.5	2.6	3.5	2.4	3.1	2.4	3.2	2.7	1.5
2.8	2.0	0.8	2.7	3.3	2.8	2.0	2.5	3.9	2.8	2.5	2.5	2.7	2.4	1.6
2.2	2.6	2.3	3.2	3.5	3.0	3.1	2.7	2.7	2.5	2.8	2.9	2.6	2.4	0.2
1.1	1.3	2.4	3.4	1.3	2.5	3.5	2.7	3.3	2.2	2.4	1.9	2.7	2.6	1.3
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

(NAs added for prediction)

END

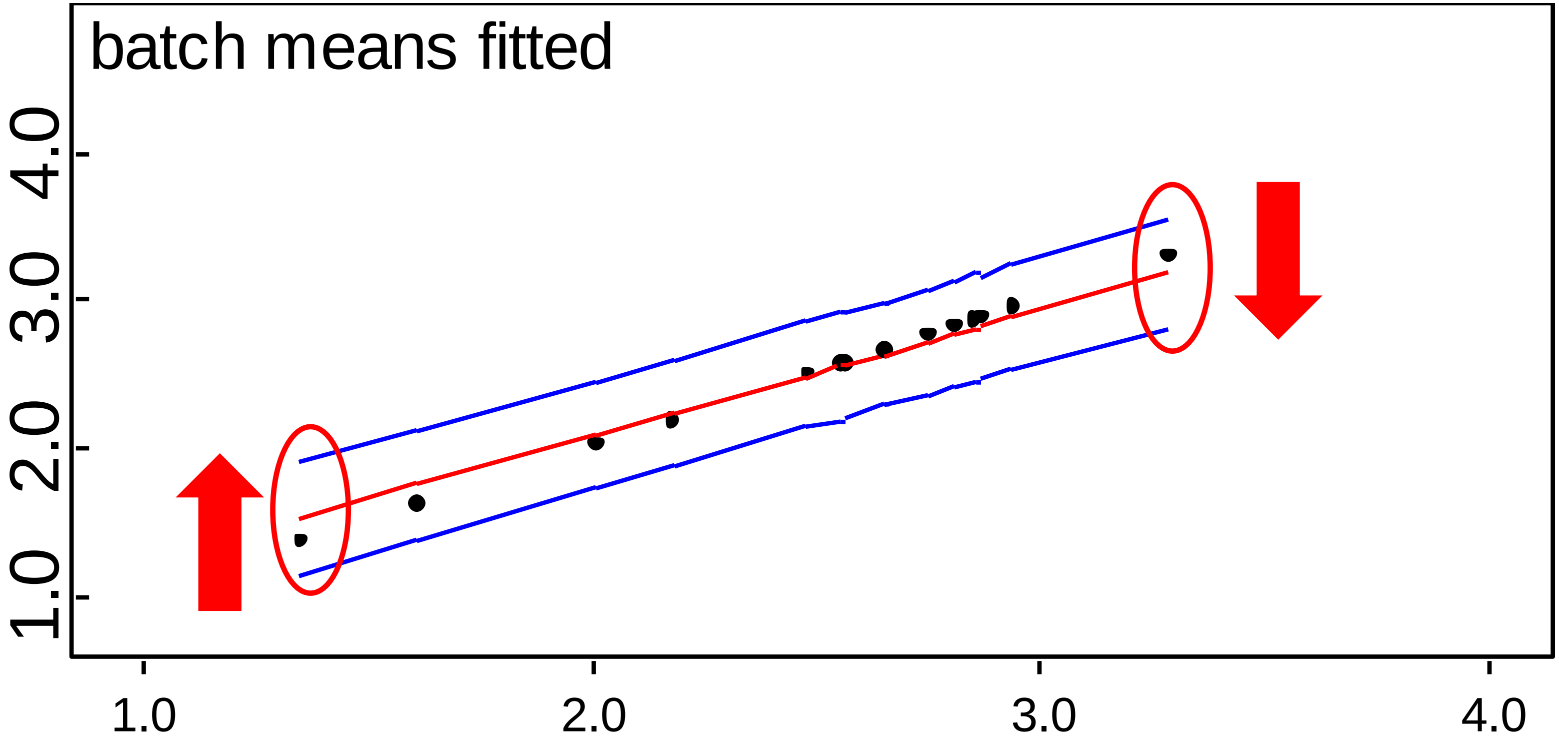
Hierarchical normal example

```
model{
for(i in 1:15){
mu[i] ~ dnorm(mu0,tau0)
for(j in 1:8){
x[j,i] ~ dnorm(mu[i],tau)
}
}
mu0 ~  dunif(-10,10)
tau0 ~ dgamma(0.01,0.01); var0 <- 1/tau0; sigma0 <- sqrt(var0)
tau ~ dgamma(0.01,0.01); var <- 1/tau; sigma <- sqrt(var)

# percentage of between variance from total variance:
r <- 100*var0/(var0+var)
}
```

Hierarchical normal example

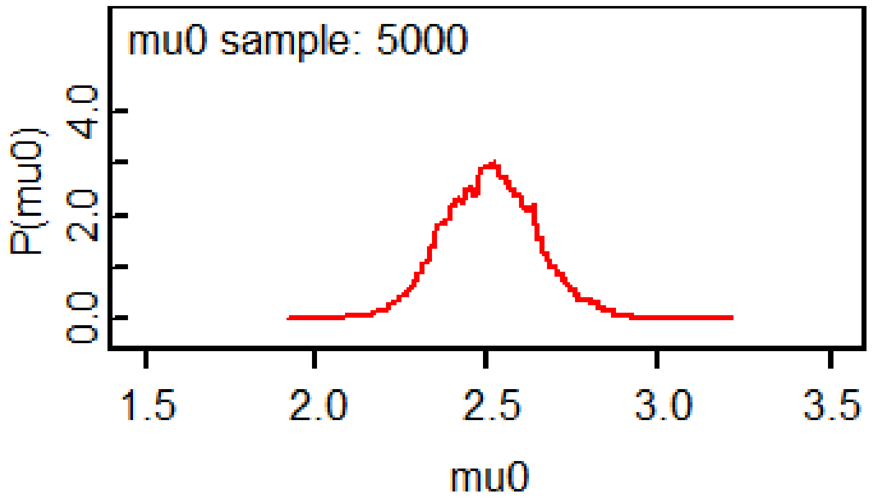
- Comparison of observed batch means ('dots') and estimated batch means μ_i (95% CIs)



- Note: shrinkage to the overall mean μ_0.
- The more data in a group, the less shrinkage to μ_0.

Hierarchical normal example

- Comparison of observed overall mean (2.509) and estimated overall mean μ_0



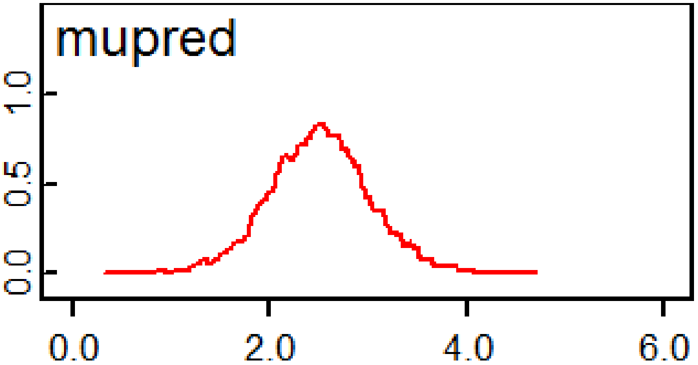
	mean	sd	val2.5pc	median	val97.5pc
mu0	2.508	0.1447	2.222	2.508	2.801

- In this case: all groups had same number of observations. If different, the group with most observations would have more weight. $$\text{weight}_\text{i} = \frac{1}{\sigma^2 / n_i + \sigma_0^2}$$

Hierarchical normal example

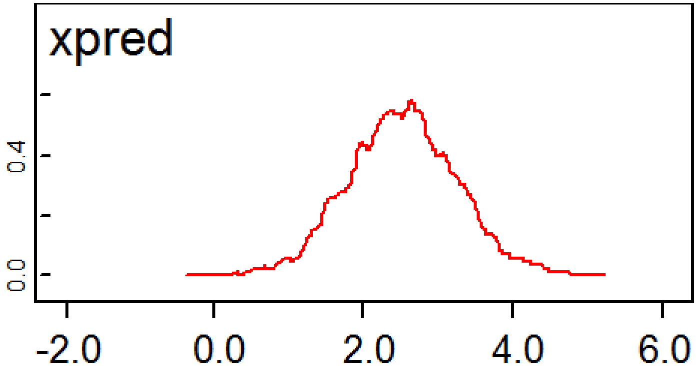
- Could make predictions for new group means.
- $\mu_k \sim N(\mu_0, \sigma_0^2)$

	mean	**sd**
mupred	2.518	0.5308



- Could make predictions for new units within groups
- $x_{jk} \sim N(\mu_k, \sigma^2)$

	mean	**sd**
xpred	2.51	0.7471



Hierarchical normal example

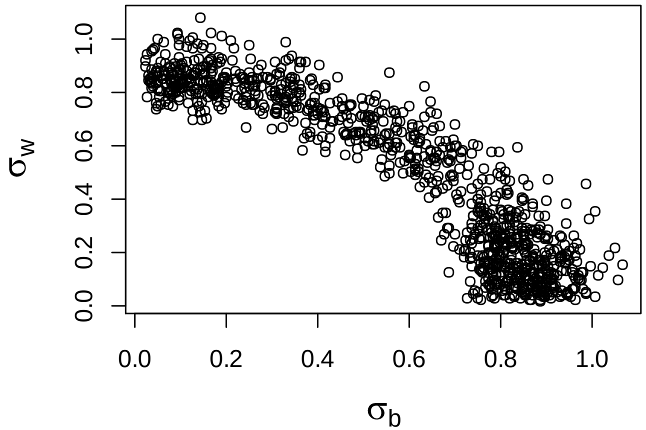
- Could estimate variance components to study between group variance versus within group variance.
- Could combine several data sources for evidence synthesis.
  - Some data could represent better samples within group
  - Some data could represent better samples between groups.

- Combining different data formats with different coarsity: e.g. individual unit samples and summary data
- Meta-analysis of several studies each with different strengths and weaknesses.

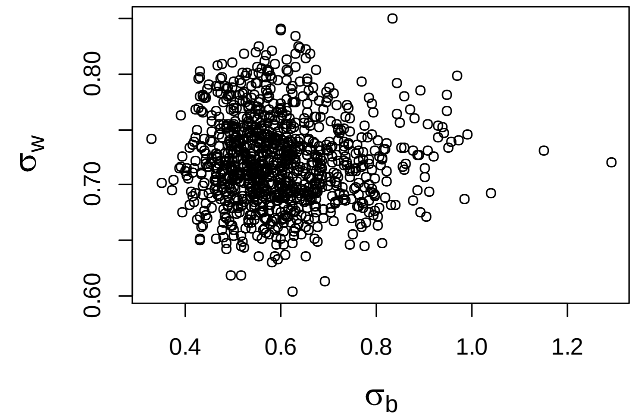
Hierarchical normal example

- Results for variance components from two data sources:

Posterior from Lindblad et al. data



Posterior from Hansson et al. data



Posterior from combined data