

Bayesian Data Analysis

Daniel Paulin & Nicolò Margaritella

University of Edinburgh



Semester 2, 2020/2021

With thanks to Jonathan Gair, Rubén Amorós-Salvador, Ken Newman, Vanda Inácio and
Natalia Bochkina for much of the material

Outline

- 1 Review
- 2 Introduction
- 3 Salmon Survival Example
- 4 General Remarks
- 5 Salmon Survival Example (II)
- 6 Radon Levels Example
- 7 Identical Model
- 8 Independent Model
- 9 Hierarchical Model
- 10 Hierarchical Models: from JAGS to INLA

Section 1

Review

Hierarchical Models

Review

Structure of the Bayesian models considered thus far:

- 1 **The Likelihood** or **observation distribution** is a probability distribution for n *independent* observations, $\mathbf{y} = y_i, i=1, \dots, n$, that depends on one or more unknown parameters, $\theta = (\theta_1, \dots, \theta_q)$.

$$f(y_1, y_2, \dots, y_n | \theta) = \prod_{i=1}^n p(y_i | \theta) \quad (1)$$

- With frequentist inference, we treat those parameters as constants $f(y_1, y_2, \dots, y_n | \theta) = L(\theta | \mathbf{y})$. A typical point estimate of θ is the value of θ maximizing the “probability” of \mathbf{y} , the MLE.

$$\text{MLE: } \hat{\theta} = \underset{\theta}{\operatorname{argmax}} L(\theta | y_1, \dots, y_n) \quad (2)$$

Hierarchical Models

Review

The Bayesian models considered thus far have had the following structure:

- 2 **The prior distribution** for θ reflects our uncertainty about θ and is in turn a function of hyperparameters ψ (so far, we consider them fixed).

$$\pi(\theta|\psi) \quad (3)$$

- 3 **The posterior distribution** for θ , calculated via the Bayes Theorem, i.e., the conditional distribution for θ given \mathbf{y} .

$$p(\theta|\psi, \mathbf{y}) = \frac{p(\theta, \mathbf{y}|\psi)}{m(\mathbf{y})} = \frac{f(\mathbf{y}|\theta)\pi(\theta|\psi)}{m(\mathbf{y})} \propto f(\mathbf{y}|\theta)\pi(\theta|\psi) \quad (4)$$

- 4 Two special cases you know well:

$$y \sim \text{Binomial}(n, p), p \sim \text{Beta}(\alpha, \beta) \rightarrow p|y \sim \text{Beta}(\alpha + y, \beta + n - y)$$

$$y_i \sim \text{Poisson}(\mu), i = 1, \dots, n; \mu \sim \text{Gamma}(\alpha, \beta) \rightarrow \mu|y \sim \text{Gamma}\left(\alpha + \sum_{i=1}^n y_i, \beta + n\right)$$

Hierarchical Models

Review

- 5 **Observation Distribution given Covariates:** Sometimes, there are explanatory variables, \mathbf{x} , that affect the expected value of the observations: $\mathbb{E}[Y] = \mu = g^{-1}(\mathbf{x}, \theta)$.

Special Cases of distributions and link functions $g()$ that you've worked with:

$$\text{Normal Linear Model: } y|x, \beta_0, \beta_1 \sim \text{Normal}(\mu = \beta_0 + \beta_1 x, \sigma^2) \quad (5)$$

$$\text{Poisson 'Loglinear' Model: } y|x, \beta_0, \beta_1 \sim \text{Poisson}(\mu = \exp(\beta_0 + \beta_1 x)) \quad (6)$$

$$\text{Bernoulli 'Logit' Model: } y|x, \beta_0, \beta_1 \sim \text{Bernoulli}\left(\mu = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}\right) \quad (7)$$

Section 2

Introduction

Hierarchical Models

Introduction

Hierarchical models extend what we've done so far:

- ↪ There are j groups for our observations $y_{i,j}$.
- ↪ The behaviour of all groups is supposed to be similar (but not equal).
- ↪ The parameters θ_j for every group come from a common probability distribution (whose parameters need to be estimated) that can be considered a **Bayesian random effect**.

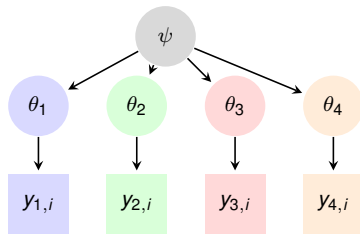
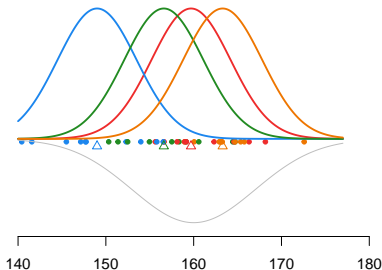
Examples:

- $y_{j,i}$ = household income in city j and household i in Scotland.
Household incomes within a city vary less than between cities.
- $y_{j,i}$ = score on a standardized mathematics exam in school j for a 10-year-old student i .
Scores vary more between different schools than between students in the same school.
- $y_{j,i}$ = indicator that infant has low birthweight ($<2000\text{mg}$) for infant i born in hospital j .
The probability that a newly born infant has a low birthweight differs between hospitals.

Hierarchical Models

Hierarchical Models

Another example: Height of women in different cities.



$$p(y_{j,i}|\theta_j)$$

$$p(\theta_j|\psi)$$

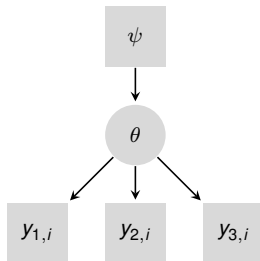
$$p(\psi)$$

- The height distribution for each city is represented on top in colours.
- The mean for each city (triangle) is a realization of a common distribution of the means (represented below in gray).

Hierarchical Models

Common effect vs random effects vs fixed effects

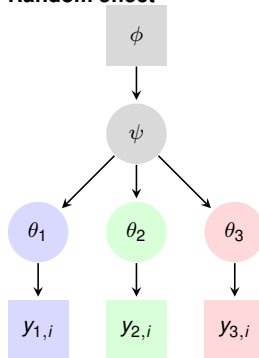
Common effect



$$y_{j,i}|\theta \sim \text{dist}(\theta)$$

$$\theta|\psi \sim \text{dist}(\psi)$$

Random effect

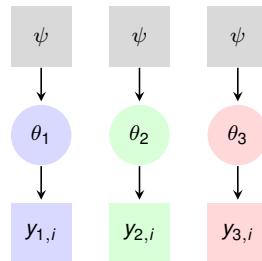


$$y_{j,i}|\theta_j \sim \text{dist}(\theta_j)$$

$$\theta_j|\psi \sim \text{dist}(\psi)$$

$$\psi|\phi \sim \text{dist}(\phi)$$

Fixed effect



$$y_{j,i}|\theta_j \sim \text{dist}(\theta_j)$$

$$\theta_j|\psi \sim \text{dist}(\psi)$$

Section 3

Salmon Survival Example

Hierarchical Models

Salmon Survival Example

- ↪ In a given year, several fish hatcheries located along rivers in Washington state, USA raise coho salmon from eggs to a juvenile stage. **Each hatchery releases a batch of juvenile fish** into the rivers. The fish then travel to the ocean and **some of them return** to the hatchery 3 years later.
- ↪ The **probability** that a juvenile salmon returns **varies between hatcheries** due to different hatchery practices and river conditions at the point of release.
- ↪ Let $J = \text{\#Hatcheries}$ and $p_j = \Pr(\text{Salmon from hatchery } j \text{ returns})$.
- ↪ Assume the p_j are independent random variables from a $\text{Beta}(\alpha, \beta)$, where α and β are unknown (NOT known and specified hyperparameters).



Photo from USGS.

Hierarchical Models

Salmon Survival Example

- Let n_j be the number of salmon released from hatchery j . Let y_j be the number of salmon that return 3 years later.
- This leads to a hierarchical model.

Hyperparameters α, β	:	?
Parameters $p_j, j = 1, \dots, J$	$\stackrel{iid}{\sim}$	Beta (α, β)
Observations $I_{j,i}, j = 1, \dots, J, i = 1, \dots, n_j$	$\stackrel{indep}{\sim}$	Bernoulli (p_j)

where $I_{j,i}=1$ if fish returns, 0 otherwise.

Note: assuming independence between fish

$$y_j = \sum_{i=1}^{n_j} I_{j,i} \sim \text{Binomial}(n_j, p_j)$$

Hierarchical Models

Salmon Survival Example

- ↪ The Bayesian solution adds another layer, *hyperprior distributions* for the hyperparameters α and β . Let $h(\alpha, \beta|\phi)$ denote the joint prior for α and β with ϕ denoting the specified “hyper”-hyperparameters
- ↪ Then

$$\begin{array}{llll}
 \text{Hyperparameters } \alpha, \beta & \sim & h(\alpha, \beta|\phi) \\
 \text{Parameters } p_j, j = 1, \dots, J & \stackrel{iid}{\sim} & \text{Beta}(\alpha, \beta) \\
 \text{Observations } y_j, j = 1, \dots, J & \stackrel{indep}{\sim} & \text{Binomial}(n_j, p_j)
 \end{array}$$

Section 4

General Remarks

Hierarchical Models

General Remarks

- ↪ Note that there is a hierarchy to this model, hence the name.
- ↪ The hierarchy can be viewed from the observed outcomes up:

$$\begin{array}{ccccc} \text{Data \& Likelihood} & & \text{Random Params \& Prior} & & \text{Hyperparameters \& Hyperprior} \\ f(y|\theta) & \Rightarrow & g(\theta|\psi) & \Rightarrow & h(\psi|\phi) \end{array}$$

- ↪ More levels can be imagined: person \rightarrow household \rightarrow city \rightarrow country
- ↪ Another name: Multi-Level Models
- ↪ Classical Frequentist Framework: Random Effects Models.

$$\begin{aligned} y_{j,i} &\sim \text{Normal}(\mu_i, \sigma^2), j = 1, \dots, J, \quad i = 1, \dots, n_j \\ \mu_i &\sim \text{Normal}(\psi_1, \psi_2) \end{aligned}$$

where the μ_i are called random effects.

Hierarchical Models

General Remarks

- ↪ The model for the observations can include both random and non-random parameters; e.g.,

$$y_{j,i} \sim \text{Normal}(\theta_j + \beta x_{j,i}, \sigma^2)$$

where $\theta_j \sim \text{Normal}(\mu, \eta^2)$ and β is a constant. This is sometimes called a *Mixed Effects* model.

Hierarchical Models

General Remarks

- ↪ When the hierarchy for a model is Normal for the observations and Normal for the random parameters, there is often interest in the relative magnitude of the variances in both models:

$$y_{j,i} \sim \text{Normal}(\theta_j, \sigma_y^2), j = 1, \dots, J, i = 1, \dots, n_j$$

$$\theta_j \sim \text{Normal}(\mu, \sigma_\theta^2), j = 1, \dots, J$$

E.g., is the between observation variance, σ_y^2 , much larger than the between group variance, σ_θ^2 ? This is sometimes expressed by the *intraclass correlation*:

$$ICC = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_y^2} \quad (8)$$

“If members of a group are unrelated, $ICC \rightarrow 0$ (the grouping contains no information); if members of a group are identical, $ICC \rightarrow 1$ (the grouping contains all the information)”
Such an analysis is sometimes called *Variance Components analysis*.

Section 5

Salmon Survival Example (II)

Hierarchical Models

Returning to the Salmon Survival Example

A simulation example will be used because one can see how the estimated results differ from the true values.

- The survival probabilities are independent Beta random variables with an average survival probability, $\mathbb{E}[p]=0.15$, and a Coefficient of Variation (CV) of 0.30. Thus $p \sim \text{Beta}(9.3, 52.7)$.
- There are $J = 10$ hatcheries which have released a different number of fish n_j .
- The numbers surviving from hatchery j , $y_j \sim \text{Binomial}(n_j, p_j)$.

```
set.seed(9931)
true.a <- 9.3; true.b <- 52.7
J <- 10
n.vector <- sample(x=50:100, size=J, replace=TRUE)
p.vector <- rbeta(J, shape1=true.a, shape2=true.b)
y.vector <- rbinom(n=J, size=n.vector, prob=p.vector)
# n = 86 83 71 60 96 85 58 95 63 100
# y = 8 12 10 8 16 14 10 24 12 17
```

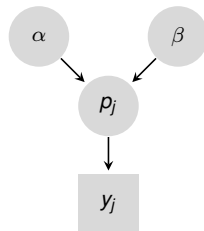
- The resulting values of p and the empirical survival probabilities, $\hat{p} = y/n$:

```
# p.vector      0.11 0.15 0.11 0.11 0.22 0.20 0.13 0.18 0.18 0.22
# phat.vector  0.09 0.14 0.14 0.13 0.17 0.16 0.17 0.25 0.19 0.17
```

Hierarchical Models

Salmon Survival Example

Hyperparameters $\alpha, \beta \sim h(\alpha, \beta | \psi)$
 Parameters $p_j, j = 1, \dots, J \stackrel{iid}{\sim} \text{Beta}(\alpha, \beta)$
 Observations $y_j, j = 1, \dots, J \stackrel{indep}{\sim} \text{Binomial}(n_j, p_j)$



Setting priors for the hyperparameters:

- ↪ Prior distribution $h(\alpha, \beta | \psi)$ is not a trivial decision.
- ↪ Can have a joint distribution or two independent marginal distributions.
- ↪ α and β must stay positive; e.g., lognormal, gamma, exponential or uniform (lowbounded by 0) priors are some possibilities.
- ↪ How to parameterize the prior distribution? Speculate on the induced values of p
- ↪ Suppose $\mathbb{E}[p] = 0.25$ with a coefficient of variation, $CV = \sqrt{\text{Var}(p)} / \mathbb{E}[p]$, of 0.30. Thus $\text{Var}(p)$ is $(\mathbb{E}[p]^2 CV^2) = (0.25 * 0.30)^2 = 9e-04$.
- ↪ Given $p \sim \text{Beta}(\alpha, \beta)$ and $\mathbb{E}[p]$ and $\text{Var}(p)$, can solve for α and β : $\alpha=8.1$ and $\beta=21.3$.

Hierarchical Models

Salmon Survival Example

Parameterize the lognormal prior distributions for α and β such that the median values are 8.1 for α and 21.3 for β . Then choose “relatively” large CV for the lognormal, e.g., $CV=0.8$, then lognormal $\sigma = \sqrt{\ln(CV^2 + 1)} = 0.70$:¹

$$\alpha \sim \text{Lognormal}(\ln(8.1), 0.70) \quad (9)$$

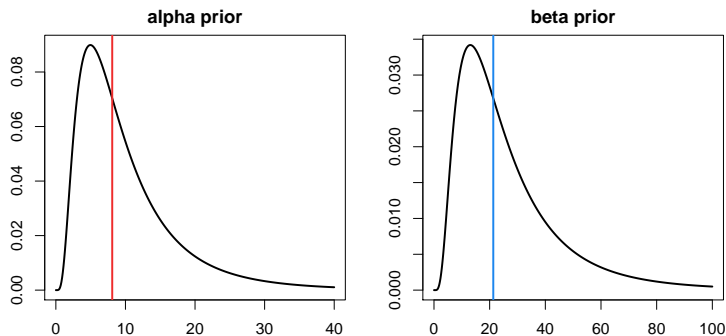
$$\beta \sim \text{Lognormal}(\ln(21.3), 0.70) \quad (10)$$

¹If $X \sim \text{Lognormal}(\mu, \sigma^2) \rightarrow CV(X) = \sqrt{\exp(\sigma^2) - 1}$

Hierarchical Models

Salmon Survival Example

The resulting priors for α and β with median denoted by red and blue vertical lines.



Hierarchical Models

Salmon Survival Example

The R and JAGS code

- ↪ Data Block: passing hyperparameters, too

```
salmon.data <- list(J=10, n=n.vector, y=y.vector,
                   ln.mu.alpha=log(8.1), ln.mu.beta=log(21.3),
                   sigma.theta=0.70)
```

- ↪ Initial Values Block: 5 IVs for α and β are chosen systematically from the lognormal prior distributions: the 0.05, 0.30, 0.50, 0.80, and 0.95 quantiles

```
prob.vec <- c(0.05, 0.30, 0.50, 0.80, 0.95)
alpha.IVs <- qlnorm(p=prob.vec, meanlog=log(8.1), sdlog=0.70)
# 2.6 5.6 8.1 14.6 25.6
beta.IVs <- qlnorm(p=prob.vec, meanlog=log(21.3), sdlog=0.70)
# 6.7 14.8 21.3 38.4 67.4
salmon.inits <- list()
for(i in 1:5) {
  salmon.inits[[i]] <- list(alpha=alpha.IVs[i],
                           beta=beta.IVs[i],
                           p=rbeta(J, 2, 3)) #initial values for p's
}
```


Hierarchical Models

Salmon Survival Example

↪ Model Block

```
salmon.model <- "model{  
  
  #-- Prior for alpha and beta  
  tau.theta    <- 1/pow(sigma.theta,2)  
  alpha ~ dlnorm(ln.mu.alpha,tau.theta)  
  beta  ~ dlnorm(ln.mu.beta,tau.theta)  
  
  #-- pdf for survival probabilities  
  #   and the likelihood for the observed returns  
  for(j in 1:J) {  
    p[j] ~ dbeta(alpha,beta)  
    y[j] ~ dbin(p[j],n[j])  
  }  
  
  expect.survival <- alpha/(alpha+beta)  
} "
```

Hierarchical Models

Salmon Survival Example

↪ Call to JAGS:

```
burnin          <- 5000
inference.length <- 50000
salmon.res.A <- jags.model(file=textConnection(salmon.model),
                           data=salmon.data,  inits=salmon.inits,
                           n.chains=length(salmon.inits),
                           quiet = TRUE)
update(salmon.res.A, n.iter=burnin)
salmon.res.B <- coda.samples(salmon.res.A,
                             variable.names=c("alpha", "beta", "expect.survival", "p"),
                             n.iter=inference.length)
```

Hierarchical Models

Salmon Survival Example

↪ Diagnostics:

```
# BGR statistic
gelman.diag(salmon.res.B)
# Number of effective size of the simulations
effectiveSize(salmon.res.B)
# Trace plots
plot(salmon.res.B)
# BGR plots
gelman.plot(salmon.res.B)
# Autocorrelation plots
autocorr.plot(salmon.res.B)
```

Hierarchical Models

Salmon Survival Example

↪ Summary statistics

```
summary(salmon.res.B)
# ...
#               Mean          SD   Naive SE Time-series SE
# alpha           9.5578   4.38067 8.761e-03      6.834e-02
# beta           47.8301  22.34306 4.469e-02      3.489e-01
# expect.survival 0.1682   0.02141 4.281e-05      7.209e-05
# ...
```

Compared to the true values of $\alpha = 9.3$, $\beta = 52.7$, and $E[p] = 0.15$.

↪ Monte Carlo simulation error relative to the standard deviations of the parameters.

$< 1/20 = 0.05?$

alpha MC/SD = 0.016,

beta MC/SD = 0.015,

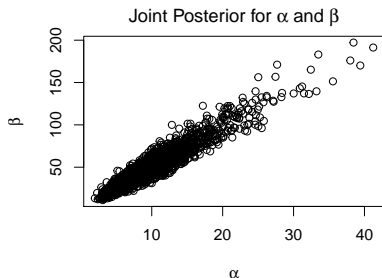
exp.S MC/SD = 0.0025

Hierarchical Models

Salmon Survival Example

The parameters α and β may be correlated which can affect the mixing rates.

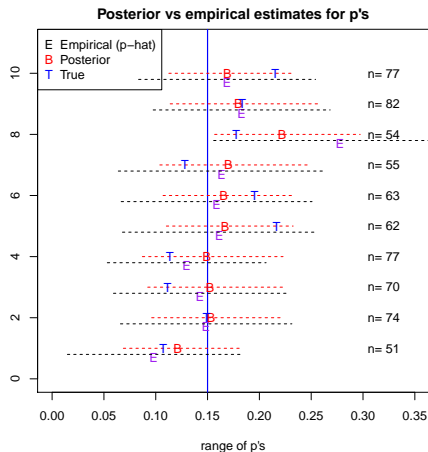
```
# Another way of pooling information from all 5 chains
salmon.output <- do.call(rbind.data.frame,salmon.res.B)
# Just work with a random sample from the joint posterior
x <- sample(1:nrow(salmon.output),2000)
plot(salmon.output$alpha[x],salmon.output$beta[x],
     xlab=expression(alpha),ylab=expression(beta),
     main=expression(paste("Joint Posterior for ",alpha," and ",beta)))
```



Hierarchical Models

Salmon Survival Example

Contrast the posterior estimates of p to the individual MLEs



Hierarchical Models

Salmon Survival Example

This plot demonstrates some important features of Bayesian HM:

Shrinkage: Compared to the individual group MLEs, the posterior means for the random parameters move toward their expected values.

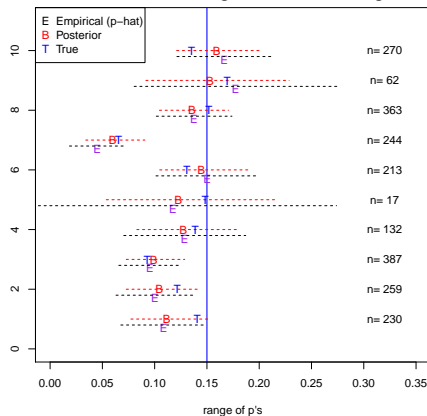
Increased Precision: The 95% (posterior) intervals are narrower than the 95% (confidence) intervals for the individual group-based estimates. This increased precision is often most noticeable for groups with smaller sample sizes.

Borrowing of Strength: this is a related point—an estimate of p_j for a group j with smaller n_j is indirectly getting additional information about p_j from the estimated distribution of all the p_j 's (which is based on all the data from the J hatcheries).

Hierarchical Models

Salmon Survival Example

Demonstration of Shrinkage and Borrowing of Strength with smaller n_j s:



Section 6

Radon Levels Example

Hierarchical Models

Radon Levels: Hierarchical Normal

- ↪ “Radon is a radioactive gas, we can’t see, smell or taste it: you need special equipment to detect it. It comes from the rocks and soil found everywhere in the UK. The radon level in the air we breathe outside is very low but can be higher inside buildings. [. . .] Radon is everywhere but some parts of the country are more likely to have high levels in buildings than others. [. . .] Any exposure to this type of radiation is a risk to health - radiation is a form of energy and can cause damage in living tissues increasing the risk of cancer.” ²
- ↪ The data we will examine here comes from a large United States survey and we just examine radon measurements from houses in counties in the state of Minnesota. ³

²From Public Health England: <http://www.ukradon.org/information/>

³Data source and examples from Patrick Breheny, University of Iowa

Hierarchical Models

Radon Levels: Hierarchical Normal

- ↪ 919 observations from randomly chosen houses from 85 of the 87 counties in Minnesota (1986-1988).
- ↪ Radon measurements (in pCi/l: picoCuries per liter) were made by “using charcoal canisters deployed during the winter months on the lowest livable area of the house”.
- ↪ The first few records of the reduced data set:

```
radon.df <- read.csv(file="Minnesota_radon_data.csv",header=TRUE)
radon.df$logradon <- pmax(log(0.1), log(radon.df$radon) )
radon.df$j.county <- as.numeric(as.factor(radon.df$county))
head(radon.df)
```

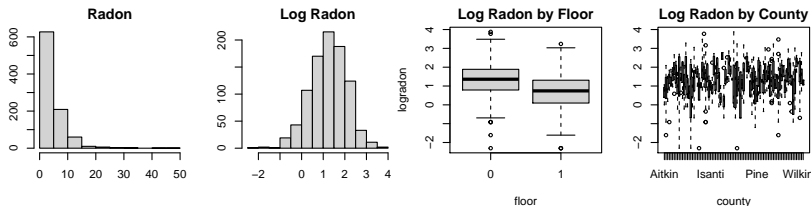
##	radon	county	floor	room	logradon	j.county
## 1	2.2	Aitkin	1	3	0.7884574	1
## 2	2.2	Aitkin	0	4	0.7884574	1
## 3	2.9	Aitkin	0	4	1.0647107	1
## 4	1.0	Aitkin	0	4	0.0000000	1
## 5	3.1	Anoka	0	4	1.1314021	2
## 6	2.5	Anoka	0	4	0.9162907	2

- ↪ The floor variable = 0 if the lowest level was a basement and = 1 if it was the ground floor.

Hierarchical Models

Radon Levels: Hierarchical Normal

The radon data are strongly right-skewed and a natural log transformation was applied (after adding 0.1 to all measurements to handle the 3 cases with values = 0).



- Radon levels tend to be higher in houses with basements
- High variability in radon levels within the counties.
- Considerable variability number of houses sampled per county.

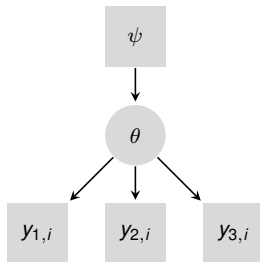
```
summary(as.numeric(table(radon.df$county)))
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.00	3.00	5.00	10.81	10.00	116.00

Hierarchical Models

Common effect vs random effects vs fixed effects

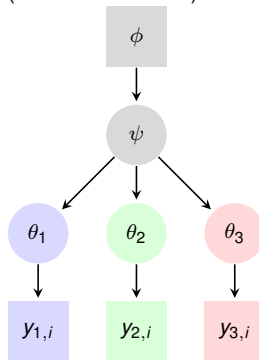
Common effect
(Identical model)



$$y_{j,i} | \theta \sim \text{dist}(\theta)$$

$$\theta | \psi \sim \text{dist}(\psi)$$

Random effect
(Hierarchical model)

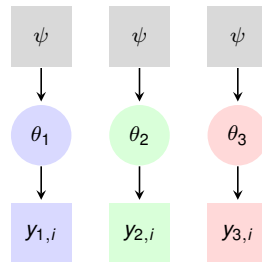


$$y_{j,i} | \theta_j \sim \text{dist}(\theta_j)$$

$$\theta_j | \psi \sim \text{dist}(\psi)$$

$$\psi | \phi \sim \text{dist}(\phi)$$

Fixed effect
(Independent model)



$$y_{j,i} | \theta_j \sim \text{dist}(\theta_j)$$

$$\theta_j | \psi \sim \text{dist}(\psi)$$

Hierarchical Models

Radon Levels: Hierarchical Normal

Mimicking Breheny, we now modeled log radon, y , as a function of the ground floor indicator in three different models. All of these assume an effect of not having basement ($I_{ground.floor} = 1$):

↪ **Identical model:** Same intercept for all the counties. Only non-basement has an effect.

$$y_i \stackrel{indep}{\sim} \text{Normal}(\alpha + \beta I_{ground.floor,i}, \sigma^2) \quad (11)$$

↪ **Hierarchical model:** Each county has a different intercept but the intercepts are random draws from the same distribution.

$$y_{j,i} \stackrel{indep}{\sim} \text{Normal}(\alpha_j + \beta I_{ground.floor,j,i}, \sigma^2), i = 1, \dots, n_j \quad (12)$$

$$\alpha_j \stackrel{indep}{\sim} \text{Normal}(\mu_\alpha, \sigma_\alpha^2), j = 1, \dots, 85 \quad (13)$$

↪ **Independent model:** Each county has a different conditionally independent intercept.

$$y_{j,i} \stackrel{indep}{\sim} \text{Normal}(\mu_j + \beta I_{ground.floor,j,i}, \sigma^2), j = 1, \dots, 85, i = 1, \dots, n_j \quad (14)$$

Section 7

Identical Model

Hierarchical Models

Radon Levels: Identical Model

Identical Model. The model and remarks.

There is no distinction according to county, the focus is on the general effect of having a basement or not.

$$y_i \stackrel{\text{indep}}{\sim} \text{Normal} \left(\alpha + \beta I_{\text{ground.floor},i}, \sigma^2 \right), i = 1, \dots, N \quad (15)$$

Hierarchical Models

Radon Levels: Identical Model

Identical Model. Data and Initials.

Priors for 3 parameters:

- $\alpha \sim \text{Normal}(0, \tau=0.01)$
- $\beta \sim \text{Normal}(0, \tau=0.01)$
- $\sigma \sim \text{Uniform}(0, 10)$. NOTE: specifying prior on σ not τ .

```
# Data block #includes hyperparameters
tau.alpha <- tau.beta <- 0.01
sigma.ub <- 10
radon.ident.data <- list(n=nrow(radon.df), log.radon=radon.df$logradon,
                        floor=radon.df$floor, sigma.ub=sigma.ub,
                        tau.alpha=tau.alpha, tau.beta=tau.beta)

# Initial values
radon.ident.inits <- function(){
  list(alpha=rnorm(1, 0, 1/sqrt(tau.alpha)),
        beta=rnorm(1, 0, 1/sqrt(tau.beta)),
        sigma=runif(1, 0, sigma.ub))
}
```

Hierarchical Models

Radon Levels: Identical Model

Identical Model. Model.

```
# Model
radon.ident.model <- "model {
  #likelihood
  for(i in 1:n) {
    log.radon[i] ~ dnorm(mu[i],tau)
    mu[i] <- alpha + beta*floor[i]
  }

  # priors for intercept, slope and sigma
  alpha ~ dnorm(0,tau.alpha)
  beta  ~ dnorm(0,tau.beta)
  tau   <- 1/pow(sigma,2)
  sigma ~ dunif(0,sigma.ub)
}"
```

Hierarchical Models

Radon Levels: Identical Model

Identical Model. Inference.

```
radon.ident.res.A <- jags.model(file=textConnection(radon.ident.model),  
                               data=radon.ident.data, inits=radon.ident.inits, n.chains=3,  
                               quiet = TRUE)  
update(radon.ident.res.A, n.iter=2000)  
radon.ident.res.B <- coda.samples(radon.ident.res.A,  
                                 variable.names=c("alpha", "beta", "sigma"), n.iter=10000)
```

Hierarchical Models

Radon Levels: Identical Model

Identical Model. Results.

```
summary(radon.ident.res.B)
```

#	Mean	SD	Naive SE	Time-series SE
# alpha	1.3270	0.02937	0.0001696	0.0001994
# beta	-0.6129	0.07299	0.0004214	0.0005010
# sigma	0.8238	0.01920	0.0001109	0.0001383

- ↪ Expected log radon in basement $\rightarrow E(\alpha|y) = 1.33$ piC/l
- ↪ Expected log radon in ground floor $\rightarrow 1.33 - 0.61 = 0.71$ piC/l
- ↪ Large relative variations:
 - CV in basement $\sigma/1.33 = 62\%$
 - CV in ground floor $\sigma/0.71 = 115\%$

Section 8

Independent Model

Hierarchical Models

Radon Levels: Independent Model

Independent Model. The model and remarks.

Each county has a different intercept, i.e. its own basement level radon, but the ground floor effect is identical.

$$y_{j,i} \stackrel{\text{indep}}{\sim} \text{Normal} \left(\alpha_j + \beta I_{\text{ground.floor},j,i}, \sigma^2 \right), j = 1, \dots, 85, \quad i = 1, \dots, n_j \quad (16)$$

Instead of creating 85 dummy variables, we will use the index j to define all the α_j with a `for` loop in JAGS.

Hierarchical Models

Radon Levels: Independent Model

Independent Model. Data and Initial values.

Priors for $85+2=87$ parameters:

- $\mu_j \sim \text{Normal}(0, \tau=0.01)$, $j=1, \dots, 85$
- $\beta \sim \text{Normal}(0, \tau=0.01)$
- $\sigma \sim \text{Uniform}(0, 10)$.

```
# Data block #includes hyperparameters
radon.indep.data <- list(n=nrow(radon.df), log.radon=radon.df$logradon,
                        floor=radon.df$floor, sigma.ub=10,
                        tau.alpha=0.01, tau.beta=0.01,
                        j.county=radon.df$j.county,
                        J=max(radon.df$j.county))

# Initial values
radon.indep.inits <- function(){
  list(alpha=rnorm(max(radon.df$j.county), 0, 1/sqrt(tau.alpha)),
        beta=rnorm(1, 0, 1/sqrt(tau.beta)),
        sigma=runif(1, 0, sigma.ub))
}
```

Hierarchical Models

Radon Levels: Independent Model

Independent Model. Model.

```
# Model
radon.indep.model <- "model {
  #likelihood
  for(i in 1:n) {
    log.radon[i] ~ dnorm(mu[i],tau)
    mu[i] <- alpha[j.county[i]] + beta*floor[i]
  }

  # priors for independent intercepts per county
  for(j in 1:J){
    alpha[j] ~ dnorm(0,tau.alpha)
  }

  # priors for intercept, slope and sigma
  beta ~ dnorm(0,tau.beta)
  tau <- 1/pow(sigma,2)
  sigma ~ dunif(0,sigma.ub)
}"
```


Hierarchical Models

Radon Levels: Independent Model

Independent Model. Inference.

```
radon.indep.res.A <- jags.model(file=textConnection(radon.indep.model),  
                               data=radon.indep.data, inits=radon.indep.inits, n.chains=3,  
                               quiet = TRUE)  
update(radon.indep.res.A, n.iter=2000)  
radon.indep.res.B <- coda.samples(radon.indep.res.A,  
                                 variable.names=c("alpha", "beta", "sigma"), n.iter=10000)
```

Hierarchical Models

Radon Levels: Independent Model

Independent Model. Results.

```
summary(radon.indep.res.B)
```

#	Mean	SD	Naive SE	Time-series SE
# alpha[1]	0.8392	0.37854	0.0021855	0.0021957
# alpha[2]	0.8749	0.10514	0.0006070	0.0006115
# alpha[3]	1.5219	0.43982	0.0025393	0.0025874
# ...				
# alpha[85]	1.1830	0.53301	0.0030773	0.0030774
# beta	-0.7190	0.07322	0.0004227	0.0005843
# sigma	0.7573	0.01856	0.0001071	0.0001461

- ↪ Expected difference of log radon in ground floor with respect of the basement → 0.72 piC/l, compared to 0.61 piC/l for the Identical Model.
- ↪ Mean of the expected log radon in basement for all counties → $\frac{1}{J} \sum_j E(\alpha_j|y) = 1.50$ piC/l, compared to 1.33 piC/l for the Identical Model.
- ↪ Reduction in relative variation: CV in basement 50% (vs 62% IM); CV in ground floor 97% (vs 115% IM)

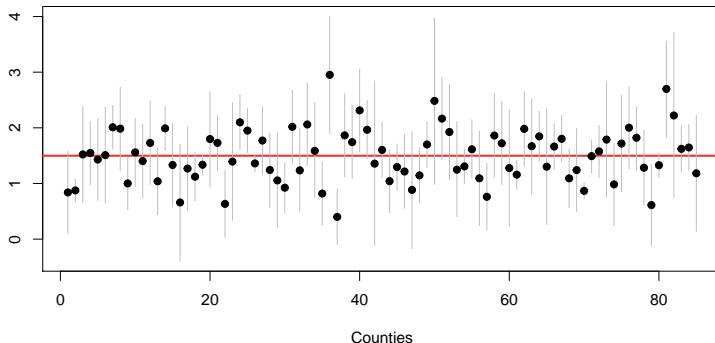
Hierarchical Models

Radon Levels: Independent Model

Independent Model. Results (II).

There is large variability among counties for $\alpha_j|y$ (posterior log radon in basements).

County intercepts: Independent Model



Section 9

Hierarchical Model

Hierarchical Models

Radon Levels: Hierarchical Model

Hierarchical Model.

- Each county has a different intercept but the intercepts are random draws from the same distribution.

$$y_{j,i} \stackrel{\text{indep}}{\sim} \text{Normal} \left(\alpha_j + \beta I_{\text{basement},j,i}, \sigma_{\text{obs}}^2 \right), i = 1, \dots, n_j \quad (17)$$

$$\alpha_j \stackrel{\text{indep}}{\sim} \text{Normal} \left(\mu_{\alpha}, \sigma_{\alpha}^2 \right), j = 1, \dots, 85 \quad (18)$$

- Now there are 4 parameters needing priors:

- $\mu_{\alpha} \sim \text{Normal}(0, \tau=0.01)$
- $\sigma_{\alpha} \sim \text{Uniform}(0, 20)$
- $\beta \sim \text{Normal}(0, \tau=0.01)$
- $\sigma_{\text{obs}} \sim \text{Uniform}(0, 20)$

Hierarchical Models

Radon Levels: Hierarchical Model

Hierarchical Model. Data and Initial values.

```
# Data block #includes hyperparameters
radon.hier.data <- list(n=nrow(radon.df), log.radon=radon.df$logradon,
                        floor=radon.df$floor,
                        sigma.ub=10, sigma.alpha.ub=20,
                        tau.mu.alpha=0.01, tau.beta=0.01,
                        j.county=radon.df$j.county,
                        J=max(radon.df$j.county))

# Initial values
radon.hier.inits <- function(){
  list(mu.alpha=rnorm(1, 0, 1/sqrt(0.01)),
       beta=rnorm(1, 0, 1/sqrt(0.01)),
       sigma=runif(1, 0, 10),
       sigma.alpha=runif(1, 0, 20))
}
```

Hierarchical Models

Radon Levels: Hierarchical Model

Hierarchical Model. Model.

```
# Model
radon.hier.model <- "model {
###likelihood
  for(i in 1:n) {
    log.radon[i] ~ dnorm(mu[i],tau)
    mu[i] <- alpha[j.county[i]] + beta*floor[i]
  }
#### priors for independent intercepts per county
  for(j in 1:J){ alpha[j] ~ dnorm(mu.alpha, tau.alpha) }
#### Hyperpriors for mu.alpha and tau.alpha
  mu.alpha ~ dnorm(0, tau.mu.alpha)
  tau.alpha <- 1/pow(sigma.alpha,2)
  sigma.alpha ~ dunif(0, sigma.alpha.ub)
#### priors for intercept, slope and sigma
  beta ~ dnorm(0,tau.beta)
  tau <- 1/pow(sigma,2)
  sigma ~ dunif(0,sigma.ub)
}"
```

Hierarchical Models

Radon Levels: Hierarchical Model

Hierarchical Model. Inference.

```
radon.hier.res.A <- jags.model(file=textConnection(radon.hier.model),  
                             data=radon.hier.data, inits=radon.hier.inits, n.chains=3,  
                             quiet = TRUE)  
update(radon.hier.res.A, n.iter=2000)  
radon.hier.res.B <- coda.samples(radon.hier.res.A,  
                                variable.names=c("alpha", "mu.alpha", "sigma.alpha", "beta", "sigma"),  
                                n.iter=10000)
```


Hierarchical Models

Radon Levels: Hierarchical Model

Hierarchical Model. Results.

```
summary(radon.hier.res.B)
```

#	Mean	SD	Naive SE	Time-series SE
# alpha[1]	1.1853	0.25442	0.0014689	0.0015879
# alpha[2]	0.9282	0.10094	0.0005828	0.0005978
# alpha[3]	1.4811	0.26858	0.0015506	0.0016270
# ...				
# alpha[85]	1.3829	0.28523	0.0016468	0.0016915
# beta	-0.6935	0.07136	0.0004120	0.0006041
# mu.alpha	1.4624	0.05340	0.0003083	0.0005780
# sigma	0.7566	0.01849	0.0001068	0.0001437
# sigma.alpha	0.3359	0.04682	0.0002703	0.0007318

- The effects on radon of the ground floor are quite the same as before (-0.69 vs -0.71 for Ind.M), as is the expected basement effect (1.46 vs 1.50 for Ind.M).
- The Intraclass Correlation Coefficient:

$$ICC \approx \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_{obs}^2} = \frac{0.3359^2}{0.3359^2 + 0.7566^2} = 0.165$$

indicating that the between county variation is not as high as the within county variation. 🔍 🔄 📄

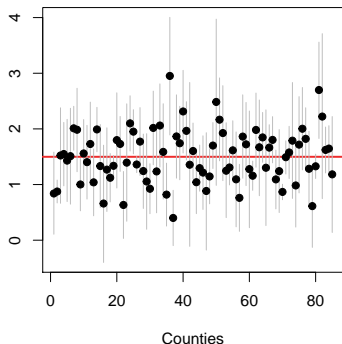
Hierarchical Models

Radon Levels: Hierarchical Model

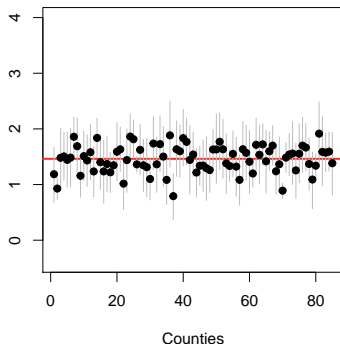
Hierarchical Model. Results (II).

The 95% credible intervals for the county effects are generally much narrower for the hierarchical model (right plot) than for the Independence model (left plot), with the same y-axis ranges.

County intercepts: Independent Model



County intercepts: Hierarchical Model



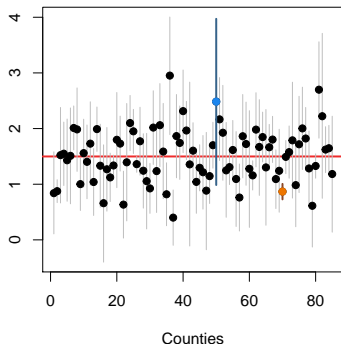
Hierarchical Models

Radon Levels: Hierarchical Model

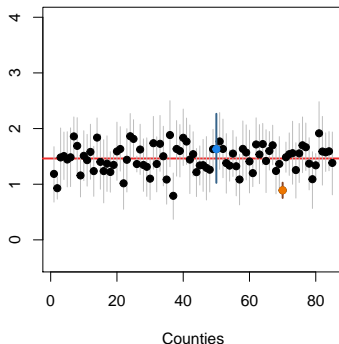
Hierarchical Model. Results (III).

The Shrinkage and Borrowing of Strength varies as a function of sample size as can be seen by two extreme cases below: a county with 1 measurement (Murray in blue) vs a county with 113 measures (St Louis in orange).

County intercepts: Independent Model



County intercepts: Hierarchical Model



Section 10

Hierarchical Models: from JAGS to INLA

Hierarchical Models

From JAGS to INLA: Hierarchical Model

- ↪ We repeat the analysis of Radon data with the hierarchical model.
- ↪ Recall that in this model each county has a random intercept, but these share the same mean and variance, and we use some priors on these hyperparameters (i.e. the common mean and variance).
- ↪ Recall also that INLA stores log-transformed precision parameters hence we need to define σ_α and σ_{obs} , in terms of $\log(\tau)$.

As in INLA we need to specify the prior on $\theta = \log(\tau)$ for $\tau = \frac{1}{\sigma^2}$, a transformation of r.v. is needed. As explained in Section 5.3.2 of Bayesian Inference in INLA (<https://becarioprecario.bitbucket.io/inla-gitbook/ch-priors.html>), the corresponding log-density $\log(\pi(\theta))$ on $\theta = \log(\tau) = -2 \log(\sigma)$ can be written as

$$\begin{aligned}\log(\pi(\theta)) &= \log(\pi_\sigma(\sigma = f(\theta)) | df(\theta)/d\theta|) \\ &= \log(\pi_\sigma(\exp(-\theta/2))) - \theta/2 - \log(2).\end{aligned}$$

In the specific case of $\sigma \sim U[0, b]$, we have $\pi_\sigma(x) = \frac{1}{b}$ for $0 \leq x \leq b$, and 0 elsewhere. So after rearrangement, it follows that

$$\log(\pi(\theta)) = \begin{cases} -\log(b) - \theta/2 - \log(2) & \text{if } \theta \geq -2\log(b) \\ -\infty & \text{if } \theta < -2\log(b) \end{cases}$$

Hierarchical Models

From JAGS to INLA: Hierarchical Model

- ↔ We can input this prior in INLA using the "expression:" string, which allows us to specify the log-density of the prior in a format that is similar to R

```
# sigma_obs:
sigma.unif.prior = "expression:
  b = 20;
  log_dens= (theta>=(-2*log(b)))*(-log(b)-theta/2-log(2))+
  (theta<(-2*log(b)))*(-Inf); return(log_dens);"

#
#sigma_alpha:
sigma.unif.prior.random.eff = "expression:
b = 20;
log_dens = (theta>=(-2*log(b)))*(-log(b)-theta/2-log(2)) +
(theta<(-2*log(b)))*(-Inf); return(log_dens);"

#
b=20;
prec.prior <- list(prec=list(prior = sigma.unif.prior,
initial = -2*log(b)+1,fixed = FALSE))
prec.prior.random.eff <- list(prec=list(
prior = sigma.unif.prior.random.eff, initial = -2*log(b)+1, fixed = FALSE))
```

Hierarchical Models

From JAGS to INLA: Hierarchical Model

- The random effects are added as **f(county,model="iid",hyper=prec.prior.random.eff)**, where "iid" defines a Gaussian random effect.

```
#
#prec.prior.random.eff encodes the prior for the random effect
#hyperparameter theta=log(tau.alpha)
#
#prec.prior encodes the prior on the likelihood precision tau.obs
#
#We set a Gaussian prior with mean 0 and precision 0.01 on the intercept
#(mu_alpha) and the regression coefficient (beta) using prior.beta
prior.beta <- list(mean.intercept = 0, prec.intercept = 0.01,
                   mean = 0, prec = 0.01)
#
m.radon.hierarchical=inla(logradon~1+floor+f(county,model="iid",hyper=
prec.prior.random.eff),data=radon.df, family="gaussian",
control.family=list(hyper=prec.prior),
control.fixed=prior.beta,control.compute=list(config=TRUE))
```

Hierarchical Models

From JAGS to INLA: Hierarchical Model

↪ Now look at the summary of the posteriors:

```
summary(m.radon.hierarchical)
```

##	mean	sd	0.025quant	0.5quant	0.975quant	mode
## (Intercept)	1.462	0.053	1.359	1.461	1.566	1.460
## floor	-0.693	0.071	-0.832	-0.693	-0.554	-0.693
## Prec.obs.	1.75	0.085	1.59	1.75	1.92	1.75
## Prec.alpha	9.46	2.699	5.31	9.08	15.82	8.35

↪ The summary statistics are essentially identical to what we got from JAGS once the precision for the observations (Prec.obs.) and the county (Prec.alpha) are back-transformed into standard deviations.

Hierarchical Models

From JAGS to INLA: Hierarchical Model

```
# sigma obs
marginal.tau=m.radon.hierarchical$
  marginals.hyperpar$`Precision for the Gaussian observations`
#We could also obtain the same marginal by
#marginal.tau=m.radon.hierarchical$marginals.hyperpar[[1]]
marginal.sigma <- inla.tmarginal(function(tau) tau^(-1/2),marginal.tau)
inla.zmarginal(marginal.sigma)
#Sigma.obs Mean 0.756436, 95%CI [0.721341,0.793192]
#
#sigma.alpha (county)
marginal.tau2=m.radon.hierarchical$
marginals.hyperpar$`Precision for county`
#We could also obtain the same marginal by
#marginal.tau2=m.radon.hierarchical$marginals.hyperpar[[2]]
marginal.sigma.alpha<-inla.tmarginal(function(tau) tau^(-1/2),marginal.tau2)
inla.zmarginal(marginal.sigma.alpha)
#Sigma.alpha Mean 0.334687, 95%CI [0.251692,0.432703]
```

The R code containing the analyses in Lecture 5 using JAGS and INLA is available on Learn. 

Hierarchical Models

Further readings on Hierarchical Models

- ↪ Kruschke JK. Doing Bayesian Data Analysis, second edition. Academic Press, 2015. Chapter 9 (the introduction and sections 9.2, 9.2.2, 9.2.3, 9.2.4, 9.3 and 9.5)
- ↪ Reich BJ, Ghosh SK. Bayesian Statistical Methods. CRC press, 2019. Chapter 6, Section 6.1 and examples 6.3 and 6.4