Aims

- Hierarchical models for combining information and meta-analysis
- Rat-tumor example
  - analysis with a fixed prior
  - analysis with historical data
- Exchangeability (de Finetti's theorem)
- Posterior predictive distributions
- Fully Bayesian treatment of the hierarchical model
  - computation
  - Rat-tumor example (cont.)

Hierarchical or multi-stage models are a natural way to think about modeling information from partially exchangeable units

- they may be tailored to model both the properties about the units themselves

$$y_i^s \mid \theta^s \sim f(y_i^s \mid \theta^s),\ i = 1, \ldots, n^s$$

- and how these properties may among units

$$\theta^s \mid \theta^* \sim g(\theta^*),\ s = 1, \ldots, S$$

- along with a specification of prior distributions for the hyperparameters in the last stage

$$\theta^* \sim h(\theta_0)$$

Example

- Study of the effectiveness of cardiac treatments
- θ^j is the survival probability for patients in hospital j

  it might be reasonable to expect that estimates of the θ^j's, which represent a sample of hospitals, should be related each other — $\theta^s \sim g(\theta^*)$
- θ^* overall survival probability

1 estimate θ^j's borrowing strength information from all the other hospitals — $p(\theta^j \mid y^1, \ldots, y^J)$

2 estimate θ^* taking into account the variability among hospitals — $p(\theta^* \mid y^1, \ldots, y^J)$

Hierarchical models for meta-analysis

- If there are several studies that address the same research question, one might be interested in combining the information from the individual studies in order to draw some sort of overall conclusion about the research question of interest
- the studies may be thought of a belonging to a population of studies addressing the same research question, and the combining of individual studies in order to learn about the whole is referred to in the literature as *meta-analysis*

Analyzing a single experiment in the context of the historical data

- θ=probability of tumor in a population of female laboratory rats that receive zero dose of the drug
- y/n =number of rats with tumor/number of rats
- current experiment $y/n = 4/14$

Analysis with a fixed prior distribution

$$\begin{aligned} y \mid n, \theta &\sim \text{Bin}(n, \theta) \\ \theta &\sim \text{Beta}(\alpha, \beta) \\ \theta \mid y, n &\sim \text{Beta}(\alpha + 4, \beta + 10) \end{aligned}$$

Analysis using historical data

- historical experiments: $y_j \mid n_j,\ j = 1, \ldots, 70$
- $y_j \mid n_j, \theta_j \sim \text{Bin}(n_j, \theta_j)$
- θ_j is the study-specific mean
- $\theta_j \sim \text{Beta}(\alpha, \beta)$
- estimate α and β with the observed mean and std of $y_j/n_j,\ j = 1, \ldots, 70$

$$\begin{aligned} 0.136 &= \tfrac{1}{n_j}\textstyle\sum_{j=1}^{n_j} \tfrac{y_j}{n_j} &&= \hat{\mu} &&= \tfrac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}} \\ 0.103 &= \tfrac{1}{n_j}\textstyle\sum_{j=1}^{n_j} (\tfrac{y_j}{n_j} - \hat{\mu})^2 &&= \hat{\sigma}^2 &&= \tfrac{\hat{\alpha}\hat{\beta}}{(\hat{\alpha}+\hat{\beta})^2(\hat{\alpha}+\hat{\beta}+1)} \end{aligned}$$

- $\hat{\alpha} = 1.4,\ \hat{\beta} = 8.6$
- $\theta_j \sim \text{Beta}(1.4, 8.6),\ y_{71} = 4,\ n_{71} - y_{71} = 10$
- $\theta_j \mid y_{71}, n_{71},$ historical data $\sim \text{Beta}(5.4, 18.6)$

Results of an "empirical Bayes approach"

- $E[\theta_{71} \mid \text{historical data}] = \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}} = .136$
- $\hat{\theta}_{71} = \frac{y_{71}}{n_{71}} = .286$
- $E[\theta_{71} \mid y_{71}, n_{71},\ \text{historical data}] = .223$

this is because the weight of the experience indicates that the number of tumors in the current experiment is unusually high

This is not a Bayesian approach because it is not based on any specified full probability model. The estimate of α and β from y is simply a starting point from which one can explore the idea of estimating the parameters of the population distribution

Complete Bayesian Analysis

Define a probability model on the entire set of parameters and experiments and then perform a Bayesian analysis on the joint distribution of all model parameters

Empirical Bayes

Find point estimates of the population parameters using historical data. This approach can be viewed as approximation to the complete Bayesian analysis

Exchangeability

- $j = 1, \ldots, J$ is a set of experiments
- $y_1, \ldots, y_J$ data
- $\theta_1, \ldots, \theta_J$ experiment-specific parameters
- $p(y_1 \mid \theta_1), \ldots, p(y_J \mid \theta_J)$ specific likelihood

$\theta_1, \ldots, \theta_J$ are exchangeable iid $p(\theta_1, \ldots, \theta_J)$ is invariant to permutations of indexes $j = 1, \ldots, J$

Ignorance $\rightarrow$ Exchangeability

if no ordering or grouping of the parameters can be made, one must assume symmetry among the parameters in their prior distribution

- θ_j is an independent draw from a population distribution governed by some unknown parameter ϕ

$$\begin{aligned} p(\theta \mid \phi) &= \prod_{j=1}^{J} p(\theta_j \mid \phi) \\ p(\theta) &= \int \left[\prod_{j=1}^{J} p(\theta_j \mid \phi)\right] p(\phi) d\phi \end{aligned}$$

- $p(\theta)$ is a mixture of iid distributions

De Finetti's Theorem

As $J \rightarrow \infty$, any "suitable well-behaved" exchangeable distribution on $\theta_1, \ldots, \theta_J$ can be written in the iid mixture form $p(\theta)$

- Statistically, the iid mixture model characterizes parameters θ from a common "superpopulation" that it is determined by the unknown hyperparameter ϕ

Exchangeability when additional information is available on the units

- $x = (x_1, \ldots, x_J)$: experiment-specific covariates
- $\theta = (\theta_1, \ldots, \theta_J)$: experiment-specific parameters
- $\theta_1 \mid x_1, \ldots, \theta_J \mid x_J$: are exchangeable
- $p(\theta \mid x) = \int \left[\prod_{j=1}^{J} p(\theta_j \mid \phi, x_j)\right] p(\phi \mid x) d\phi$

Objections to exchangeable models

- In rat tumor example, the 71 experiments were performed at different times, on different rats, and in different laboratories.
- Is it the exchangeable assumption acceptable?
- *that the experiments differ implies that the θ's differ, but it might be perfectly acceptable to consider them as if drawn from a common distribution*

Full Bayesian treatment of the hierarchical model

- Posterior distribution of (ϕ, θ)

$$\begin{aligned} p(\phi, \theta) &= p(\phi)p(\theta \mid \phi) \\ p(\phi, \theta \mid y) &\propto p(\phi, \theta)p(y \mid \phi, \theta) \\ &\propto p(\phi, \theta)p(y \mid \theta) \end{aligned}$$

Posterior predictive distributions

1. distribution of a future observations $\tilde{y}$ corresponding to an existing θ_j (*additional rats from an existing experiment*)
2. distribution of a future observations $\tilde{y}$ corresponding to future θ_j's drawn from the same superpopulation (*results from a future experiment*)

1. posterior predictive draws $\tilde{y}$ are based on the posterior draws θ_j for the existing experiment
2. posterior predictive draws $\tilde{y}$ are based on simulated $\tilde{\theta}$:

   draw $\tilde{\phi}$ from $p(\phi \mid y)$

   draw $\tilde{\theta}$ from $p(\theta \mid \tilde{\phi}, y)$

   draw $\tilde{y}$ from $p(y \mid \tilde{\theta})$

Computation with Hierarchical Models

- θ vector of parameters of interest
- ϕ vector of nuisance parameters

AIM: to obtain simulations from the joint posterior distribution $p(\theta, \phi \mid y)$ when the population distribution $p(\theta, \phi)$ is conjugate to the likelihood $p(y \mid \theta)$.

Analytic derivation

1. write $p(\phi, \theta \mid y)$ in unnormalized form (*immediate*)

$$p(\theta, \phi \mid y) \propto p(\phi)p(\theta \mid \phi)p(y \mid \theta)$$

2. determine analytically $p(\theta \mid \phi, y)$ — (*easy for conjugate normal model*)

3. find $p(\phi \mid y) = \int p(\theta, \phi \mid y)d\theta$

Drawing simulations from the posterior distribution

1. draw $\phi^* \sim p(\phi \mid y)$

2. draw $\theta^* \sim p(\theta \mid \phi^*, y)$

3. if the factorization $p(\theta \mid \phi, y) = \prod p(\theta_j \mid \phi, y)$ holds, then the components θ_j can be drawn independently, one at a time

4. draw $\tilde{y} \sim p(y \mid \theta^*)$

5. repeat the steps L times in order to obtain a set of L draws

Fully Bayesian analysis of rat tumors example

$j = 1, \ldots, 71$ experiments (Tarone 1982)

$$\begin{aligned} y_j \mid \theta_j, n_j &\sim \text{Bin}(\theta_j, n_j) \\ \theta_j \mid \alpha, \beta &\sim \text{Beta}(\alpha, \beta) \\ \alpha, \beta &\sim \text{non-informative} \end{aligned}$$

1. $p(\alpha, \beta, \theta \mid y) \propto p(\alpha, \beta) p(\theta \mid \alpha, \beta) p(y \mid \alpha, \beta)$
2. $\theta_j \mid \alpha, \beta, y \sim \text{Beta}(\alpha + y_j, \beta + n_j - y_j)$
3. simulate from $p(\alpha, \beta \mid y)$

   3.1) valuate $p(\alpha, \beta \mid y)$ over a grid of points

   3.2) approximate it as a step-function

   3.3) sample $\alpha^l \sim p(\alpha \mid y)$ and $\beta^l \sim p(\beta \mid \alpha^l, y)$ using inverse cdF method

Setting up a non informative prior distribution

Aim:found a diffuse hyperprior distribution for α and β

- let $\gamma = \log \frac{\alpha}{\beta}$
- let $\delta = \log(\alpha + \beta)$
- $p(\gamma, \delta) \propto$ constant

  this prior leads an improper posterior

- let $\gamma_1 = \frac{\alpha}{\alpha+\beta}$
- let $\delta_1 = (\alpha + \beta)^{-1/2}$

  this prior leads a proper posterior, and yields:

$$\begin{aligned} p(\alpha, \beta) &\propto (\alpha + \beta)^{-5/2} \\ p(\gamma_1, \delta_1) &\propto \alpha\beta(\alpha + \beta)^{-5/2} \end{aligned}$$

Computing the marginal posterior density

- Contour plot of the unnormalized marginal posterior $p(\gamma, \delta \mid y)$
- draw 1000 random samples from the joint posterior $p(\alpha, \beta, \theta_1, \ldots, \theta_{71} \mid y)$ as follows:

1) simulate γ^l, δ^l from $p(\gamma, \delta \mid y)$ with the discrete-grid sampling procedure

   [1.1] for $l = 1, \ldots 1000$

   [1.2] transform $\gamma^l, \delta^l \to \alpha^l, \beta^l$

2) for each l draw $\theta^l \sim \text{Beta}(\alpha^l + y_j, \beta^l + n_j - y_j)$
3) displays the results, for example histogram of $ED50 = \alpha/\beta$


```
#Example:analysis o biossay experiment
#Gelman and Rubin book, pag 121, 71 experiments, Tarone(1982) data
#plot joint posterior of gamma and delta
log.post.tarone_function(gamma,delta){
  tarone _ matrix(scan("tarone.txt"),byrow=T,ncol=2)
  deaths _ tarone[,1]
  rats   _ tarone[,2]
  alpha  _ exp(gamma+delta)/(1+exp(gamma))
  beta   _ exp(delta)/(1+exp(gamma))
  ldens  _ 0
  for (i in 1:length(tarone[,1])){
    ldens  _ ldens +
      (lgamma(alpha+beta)+lgamma(alpha+deaths[i])+lgamma(beta+rats[i]-deaths[i]))-
        (lgamma(alpha)+lgamma(beta)+lgamma(alpha+beta+rats[i]))
  }
  ldens - 5/2*log(alpha+beta)+log(alpha)+log(beta)
}
plot.joint.post.tarone_function(w=0,ll=200,sim){
  if (w == 1) postscript("/home/biostats/fdominic/course/jposterior.tarone.ps")
  gamma _ seq(-2.3,-1.3,length=ll)
  delta  _ seq(1,5,length=ll)
  contours _ seq(.05,.95,.1)
  logdens _outer(gamma,delta,log.post.tarone)
  dens_exp(logdens-max(logdens))
  contour(gamma,delta,dens,levels=contours,xlab="log(alpha/beta)",ylab="log(alpha+beta)",la
points(log(sim$aaa/sim$bbb),log(sim$aaa+sim$bbb))
 mtext("Posterior density",3,line=1,cex=1.2)
par(oma=c(0,0,0,0))
par(mfrow=c(1,1))
if (w == 1) dev.off()}
#grid of values of the joint posterior
grid.value.tarone_function(ll=10){
  gamma _ seq(-2.3,-1.3,length=ll)
  delta  _ seq(1,5,length=ll)
  PP    _ matrix(NA,ll,ll)
  for(i in 1:ll){
    for(j in 1:ll){
```

```
      PP[i,j]_log.post.tarone(gamma[i],delta[j])
    }}
  MM _ max(PP)
  PP_exp(PP-MM)
  ccc _ sum(PP)
  PP  _ PP/ccc
  return(PP,gamma,delta)
}
#draw posterior sample of alpha,beta, and theta using the inverse cdf method
        sampling.tarone_function(NN=1000,II,gamma,delta){
          ggg   _ aaa _ ddd _ bbb _ NULL
          gamma.mar_apply(II,1,sum)
          delta.mar_apply(II,2,sum)
          tarone _ matrix(scan("tarone.txt"),byrow=T,ncol=2)
          deaths _ tarone[,1]
          rats   _ tarone[,2]
          ttt_matrix(NA,length(rats),NN)
          for(l in 1:NN){
            uuu_runif(1,0,1)
            ggg[l]_max(gamma[cumsum(gamma.mar) < uuu])
            junk_(1:length(gamma))[gamma==ggg[l]]
            conditional_II[junk,]/sum(II[junk,])
            uuu_runif(1,0,1)
            ddd[l]_max(delta[cumsum(conditional) < uuu])
            aaa[l]  _ exp(ggg[l]+ddd[l])/(1+exp(ggg[l]))
            bbb[l]  _ exp(ddd[l])/(1+exp(ggg[l]))
    for(i in 1:length(deaths)){
 ttt[i,l]_rbeta(1,aaa[l]+deaths[i],bbb[l]+rats[i]-deaths[i])   }}
 return(aaa,bbb,ttt)  }
```


Hierarchical Normal Model

- variance known (sec. 5.4)

$$
\begin{aligned}
y_{ij} &\sim N(\theta_j, \sigma^2) \quad , \; i = 1, \ldots, n_j, \; j = 1, \ldots, J \\
\theta_j &\sim N(\mu, \tau^2) \\
\mu, \tau^2 &\sim p(\mu, \tau^2)
\end{aligned}
$$

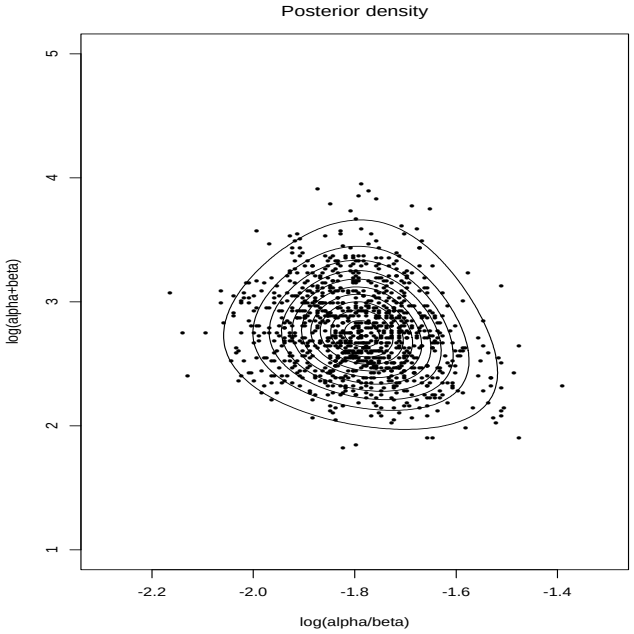
  SAT coaching experiment

- variance unknown (sec. 9.8)

$$
\begin{aligned}
y_{ij} &\sim N(\theta_j, \sigma^2), \; i = 1, \ldots, n_j, \; j = 1, \ldots, J \\
\theta_j &\sim N(\mu, \tau^2) \\
\mu, \sigma^2, \tau^2 &\sim p(\mu, \sigma^2, \tau^2)
\end{aligned}
$$

  DIET measurements

- approximating joint posterior distribution with direct simulation and with a Gibbs sampler



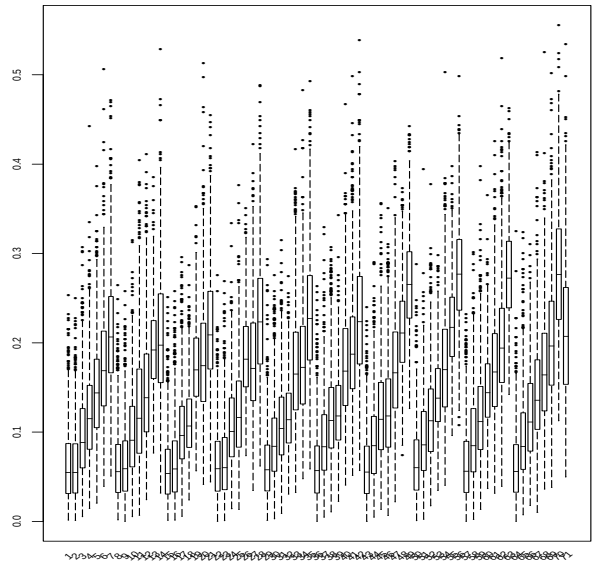


Figure 1: Marginal posterior distributions of $\theta_1, \ldots, \theta_7 1$

Classical random-effect analysis of the variance

$$\begin{aligned} y_{ij} &\sim N(\theta_j, \sigma^2) \\ y_{.j} &= \tfrac{1}{n_j}\textstyle\sum_{i=1}^{n_j} y_{ij} \\ \sigma_j^2 &= \tfrac{\sigma^2}{n_j} \\ y_{.j} &\sim N(\theta_j, \sigma_j^2) \end{aligned}$$

How do we estimate θ_j? Two possibilities

1. Separate estimates: $y_{.j}$ *(bad choice when $J = 20$ and $n_j = 2$)*
2. Pooled estimates: $y_{..} = (\sum_j \frac{1}{\sigma_j^2} y_{.j})/(\sum_j \frac{1}{\sigma_j^2})$
3. Weighted combination:
   $\hat{\theta}_j = \lambda_j y_{.j} + (1 - \lambda_j) y_{..}, \quad 0 \le \lambda \le 1$

keep in mind that $y_{.j} \sim N(\theta_j, \sigma_j^2)$ can be a good approximation of the likelihood function even when y_{ij} are not normal!

Classical approach perform an analysis of variance F test for differences among means.

$$F = \frac{MS(\text{between})}{MS(\text{within})}$$

- if $F >> 1$ we reject $H_0 : \tau^2 = 0$, and we use $y_{.j}$'s
- if $F << 1$ we accept $H_0 : \tau^2 = 0$, and we use $y_{..}$
- $y_{.j}$ is the posterior mean of θ_j if $\theta_j \sim U(\infty, \infty)$
- $y_{..}$ is the posterior mean of θ if $\theta_1 = \ldots = \theta_J = \theta$ and $\theta \sim U(\infty, \infty)$
- $\lambda_j y_{.j} + (1 - \lambda_j) y_{..}$ is the posterior mean of θ_j if $\theta_j \sim N(\mu, \tau^2)$
- $\lambda_j = 1 \rightarrow \ \tau^2 = \infty$ *separate*
- $\lambda_j = 0 \rightarrow \ \tau^2 = 0$ *pooling*

Computations

- let $\theta = (\theta_1, \ldots, \theta_J)$
- joint prior $p(\mu, \tau^2) = p(\mu \mid \tau^2) \times p(\tau^2)$
- joint posterior

$$\begin{aligned} p(\theta, \mu, \tau \mid y) &= p(\tau \mid y) p(\mu \mid \tau, y) p(\theta \mid \mu, \tau, y) \\ &\propto p(\tau \mid y) N(\hat{\mu}, V_\mu) \textstyle\prod_{j=1}^{J} N(\hat{\theta}_j, V_j) \end{aligned}$$

where

$$\begin{aligned} \hat{\mu} &= \frac{\sum_j \frac{1}{\sigma_j^2 + \tau^2} y_{.j}}{\sum_j \frac{1}{\sigma_j^2 + \tau^2}} \\ V_\mu^{-1} &= \textstyle\sum_j \frac{1}{\sigma_j^2 + \tau^2} \\ \hat{\theta}_j &= \frac{\frac{1}{\sigma_j^2} y_{.j} + \frac{1}{\tau^2}\mu}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}} \\ V_j^{-1} &= \frac{1}{\sigma_j^2} + \frac{1}{\tau^2} \end{aligned}$$

- $p(\tau \mid y)$ is not available in closed form
- $p(\log \tau) \propto 1$ lead to an improper posterior!!

The Bayesian analysis under a normal hierarchical model provides a compromises that combines information from all the experiments without assuming all the θ_js to be equal

Hierachical structuring of the models is an essential tool for achieving *partial pooling* of estimates, and compromising in a scientific way between alternative sources of information

Table 1: *Observed effects of special preparation on SAT-V scores in eight randomized experiments. Rubin (1981)*

School	Estimated treatment effect, y_j	Standard error of effect estimate, σ_j
A	28.39	14.9
B	7.94	10.2
C	-2.75	16.3
D	6.82	11.0
E	-0.64	9.4
F	0.63	11.4
G	18.01	10.4
H	12.16	17.6

SAT coaching experiment

Separate randomized experiments were performed to estimate the effect of coaching programs for the SAT-V in each of eight schools

- y_j = estimated coaching effects
- σ_j^2 = sampling variances

Separate Estimates

- Looking at the table we see that it is difficult statistically to distinguish between any of the experiments
- treating each experiment separately, and applying the simple normal analysis, yields 95% posterior intervals that all overlap

Difficulties

- $\theta_A \mid y \sim N(28.4, 14.9^2)$
- $P(\theta_A > 28.4 \mid y_4) = \frac{1}{2}$ *a doubtful statement, considering the results fronm the other seven schools*

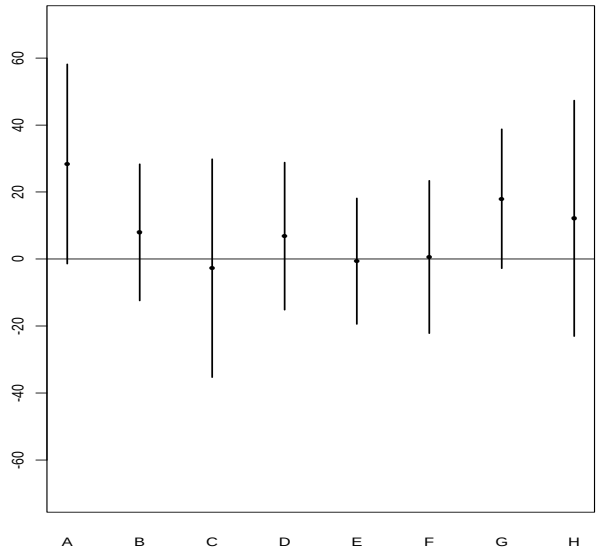


Figure 2: 95% confidence intervals of the estimated treatment effects

Pooled Estimate

Under the hyphotesis that all the experiments have the same effect and produce independent estimates of this common effect, we can suppose that $\theta_1 = \ldots = \theta_8$,

$$E[\theta \mid y] = \left(\sum_j y_j/\sigma_j^2\right)\left(\sum_j 1/\sigma_j^2\right)^{-1} = 7.9$$
$$Var[\theta \mid y] = \left(\sum_j 1/\sigma_j^2\right)^{-1} = 17.4$$

Difficulties

- $\theta_A \mid y \sim N(7.9, 17.4^2)$
- $\theta_C \mid y \sim N(7.9, 17.4^2)$
- $P(\theta_A - \theta_C < 0 \mid y) = \frac{1}{2}$ *It is difficult to justify from the table*

Hierarchical Model with known variance

$$
\begin{aligned}
y_j &\sim N(\theta_j, \sigma_j^2) \\
\theta_j &\sim N(\mu, \tau^2) \\
\mu &\sim N(0, 100) \\
\tau^2 &\sim IG(AA, BB)
\end{aligned}
$$

Two Options:

- Direct sampling:

$$p(\tau, \mu, \theta \mid y) \propto p(\tau \mid y) p(\mu \mid \tau, y) \prod_{j=1}^{J} p(\theta \mid \mu, \tau, y)$$

- Gibbs sampling:

write down

$$
\begin{aligned}
p(\theta_j \mid \mu, \tau, y) &= ? \\
p(\mu \mid \theta_1, \ldots, \theta_J, \tau, y) &= ? \\
p(\tau \mid \theta_1, \ldots, \theta_J, \mu, y) &= ?
\end{aligned}
$$

Table 2: *Coagulation time in seconds for blood drawn from 24 animals randomly allocated to four different diets. Different treatments have different numbers of observations because the randomization was unrestricted. (Box, Hunter,(1978))*

Diet	Measurements
A	62,60,63,59
B	63,67,71,64,65,66
C	68,66,71,67,68,68
D	56,62,60,61,63,64,63,59

DIET-measurements example

Hierarchical Normal model with unknown variance

$$
\begin{aligned}
y_{ij} &\sim N(\theta_j, \sigma^2),\ i = 1, \ldots, n_j \\
\theta_j &\sim N(\mu, \tau^2),\ j = 1, \ldots, J,\ n = \textstyle\sum n_j \\
\mu &\sim N(0, 100) \\
\sigma^2 &\sim IG(AA1, BB1) \\
\tau^2 &\sim IG(AA2, BB2)
\end{aligned}
$$

- if we were assign a uniform prior to $\log \tau$, the posterior would be improper,(sec. 5)

- crude estimates

$$
\begin{aligned}
\hat{\theta}_j &= \bar{y}_{.j} \\
\hat{\sigma}^2 &= \textstyle\sum_j \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{.j})^2 \\
\hat{\mu} &= \frac{1}{J} \textstyle\sum_j^J \hat{\theta}_j \\
\hat{\tau}^2 &= \frac{1}{J-1} \textstyle\sum_j^J (\hat{\theta}_j - \hat{\mu})^2
\end{aligned}
$$

Implement a Gibbs Sampler

List of the full conditional distributions

$$
\begin{aligned}
\theta_j \mid y, \mu, \sigma, \tau &\sim N\left((\tfrac{\mu}{\tau^2} + \tfrac{n_j \bar{y}_j}{\sigma^2})(\tfrac{1}{\tau^2} + \tfrac{n_j}{\sigma^2})^{-1}, (\tfrac{1}{\tau^2} + \tfrac{n_j}{\sigma^2})^{-1} \right) \\
\mu \mid y, \theta_j, \sigma, \tau &\sim N(\tfrac{1}{J} \textstyle\sum \theta_j, \tfrac{\tau^2}{J}) \\
\sigma^2 \mid y, \theta_1, \ldots, \theta_J \mu, \tau &\sim IG(AA1 + \tfrac{n}{2}, \tfrac{1}{2} \textstyle\sum_{ij} (y_{ij} - \theta_j) \\
\tau^2 \mid y, \theta_1, \ldots, \theta_J, \mu, \sigma &\sim IG(AA2 + \tfrac{J}{2}, \tfrac{1}{2} \textstyle\sum_j (\theta_j - \mu)^2)
\end{aligned}
$$

```
#Approximating the joint posterior distribution p(theta,mu,sigma,tau|data) for the hierarch
#list of functions
##LIST OF THE FUNCTION--------------------------------
D_data()
I_initial.values(D)
I_one.iteration(I,D)
#gibbs.chain_markov(2000,100,2,I)
#-------------------------------------------------
##DATA AND STARTING POINTS-------------------------
data_function(){
  J_4
                                      #data-set
  yy_rbind(c(62,60,63,59,NA,NA,NA,NA),
           c(63,67,71,64,65,66,NA,NA),  #use scan command for bigger
           c(68,66,71,67,68,68,NA,NA),  #data-sets
           c(56,62,60,61,63,64,63,59))


                #prior hyperparameters
AA1          _ 3 #tausqr and sigmasqr have a IG dist
AA2          _ 3  #with mean close to rough estimates a
BB1          _ 26 # a big variance
BB2          _ 10
  output_list(J=J,yy=yy,AA1=AA1,BB1=BB1,AA2=AA2,BB2=BB2)
return(output)
}

initial.values_function(D){
  yy_D$yy
  J_D$J
  AA1_D$AA1
  AA2_D$AA2
  BB1_D$AA1
  BB2_D$BB1
  theta_apply(yy,1,mean,na.rm=T)
  sigmasqr_mean(c(var(yy[1,!is.na(yy[1,])]),
                  var(yy[2,!is.na(yy[2,])]),
```



```
                  var(yy[3,!is.na(yy[3,])]),
                  var(yy[4,!is.na(yy[4,])])))
  mu_mean(theta)
  tausqr_var(theta)

  output_list(theta=theta,mu=mu,sigmasqr=sigmasqr,tausqr=tausqr)
  return(output)
}

#####ONE ITERATION OF THE GIBBS-SAMPLER----------------------------------------
one.iteration_function(I,D){
  J_D$J
  yy_D$yy
  AA1_D$AA1
  AA2_D$AA2
  BB1_D$BB1
  BB2_D$BB2
  theta_I$theta
  mu_I$mu
  sigmasqr_I$sigmasqr
  tausqr_I$tausqr
  n_NULL
                                  #Draw thetaj's

  for(j in 1:J){
    n[j]_length(yy[j,!is.na(yy[j,])])
    vvv   _ (1/tausqr+n[j]/sigmasqr)^(-1)
    mmm   _ (mu/tausqr+sum(yy[j,!is.na(yy[j,])])/sigmasqr)*vvv
    theta[j]_rnorm(1,mmm,sqrt(vvv))
  }


                                        #Draw sigmasqr
  num_NULL
  for(j in 1:J){
   yysqr_sum(yy[j,]^2,na.rm=T)
    num[j]_yysqr+n[j]*theta[j]^2-2*sum(yy[j,],na.rm=T)*theta[j]
```



```
  }


num_sum(num)
  AA1_AA1+sum(n)/2
  BB1_BB1+num/2
  sigmasqr   _ 1/(rgamma(1,AA1)/BB1)

#Draw mu

  mu_rnorm(1,mean(theta),sqrt(tausqr/J))

 #Draw tausqr
  AA2_AA2  + J/2
  BB2_ BB2 + .5*sum(theta-mu)^2
  tausqr_1/(rgamma(1,AA2)/BB2)
  return(theta,sigmasqr,mu,tausqr)
}

#MARKOV CHAIN
markov _ function(NN,burnin,skip,I) {
# I: list of parameter and data value as in the output of gibbs.city
# skip: number of iteration between two recorded iteration (must be >= 0)
# NN total number of recorded iteration.
# total length of the chain is burnin + skip * NN
  theta _ matrix(NA,length(I$theta),NN)
  sigmasqr_rep(NA,NN)
  mu_rep(NA,NN)
  tausqr_rep(NA,NN)
  for (m in 1:burnin) { I _ one.iteration(I,D) }
  for (m in 1:NN) {
# if (skip > 0)
    for (s in 1:skip) { I _ one.iteration(I,D) }
    I _ one.iteration(I,D)
    theta[,m]   _ I$theta
    sigmasqr[m] _ I$sigmasqr
    mu[m]       _ I$mu
```



```
    tausqr[m]   _ I$tausqr
  }

return(theta,sigmasqr,mu,tausqr)
}

plotmeans_function(w=0,I=gibbs.chain,cc=2){
if (w == 1) postscript("/home/biostats/fdominic/course/plotmeans.ps")
par(oma=c(0,0,2,0))
boxplot(I$theta[1,],I$theta[2,],I$theta[3,],I$theta[4,],I$mu,
names=c("theta1","theta2","theta3","theta4","mu"),style.bxp = "old",cex=cc)
par(oma=c(0,0,0,0))
par(mfrow=c(1,1))
if (w == 1) dev.off()

}
plotsigma_function(w=0,I=gibbs.chain,cc=2){
if (w == 1) postscript("/home/biostats/fdominic/course/plotsigma.ps")
par(oma=c(0,0,2,0))
hist(sqrt(I$sigmasqr),xlab="sigma",density=-1,yaxt="n",cex=cc)
abline(v=mean(I$sigma),lty=2)
par(oma=c(0,0,0,0))
par(mfrow=c(1,1))
if (w == 1) dev.off()

}
plottau_function(w=0,I=gibbs.chain,cc=2){
if (w == 1) postscript("/home/biostats/fdominic/course/plottau.ps")
par(oma=c(0,0,2,0))
hist(sqrt(I$tausqr),xlab="tau",nclass=20,density=-1,yaxt="n",xlim=c(0,15),cex=cc)
par(oma=c(0,0,0,0))
par(mfrow=c(1,1))
if (w == 1) dev.off()

}
```

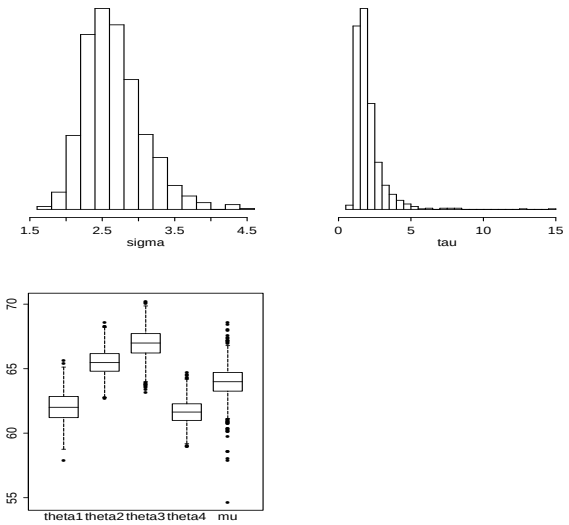


Figure 3: Posterior distributions of σ^2, τ^2, $\theta_1, \ldots, \theta_j$ and μ