

# Chapter 9: Hierarchical Models

ST4234: Bayesian Statistics

Semester 2, AY 2019/2020

Department of Statistics and Applied Probability

National University of Singapore

LI Cheng

[stalic@nus.edu.sg](mailto:stalic@nus.edu.sg)

# Outline

Introduction

Heart transplant mortality data

Individual and combined estimates

Equal mortality rates

Exchangeable model

Posterior inferences

- Shrinkage

- Comparing hospitals

- Bayesian Sensitivity Analysis

- Posterior predictive model checking

## 9.1 Introduction

- In many statistical problems, we are interested in learning about a large number of parameters which are connected in some way.
- In this chapter, we illustrate the concept of [hierarchical models](#), using a data example of heart transplant mortality data, following Chapter 7 of Jim Albert's book.
- For more examples of hierarchical models, please read [Chapter 8 in Peter Hoff's book](#), which provides another example of hierarchical normal models for grouped data.
- [Bayesian hierarchical modeling](#) has wide applications. For example, [Bayesian \(generalized linear\) mixed effects models](#) are useful in epidemiology, economics, psychology, etc. Bayesian topic models, such as [latent Dirichlet allocation](#), are important in machine learning and natural language processing.

## 9.1 Introduction

- For example, suppose one is interested in learning about the mortality rates due to heart transplant surgery for 94 hospitals.
- Each hospital has a true mortality rate  $\lambda_i$ , and it is reasonable to believe a priori that  $\lambda_1, \dots, \lambda_{94}$  are similar, which implies a dependence structure among these parameters.
- If one is told some information about a particular hospital's true rate, that information would likely affect one's belief about the rate of a second hospital.

## 9.1 Introduction

- In many-parameter situations such as this, it is natural to construct a prior distribution in a hierarchical fashion.
- In **hierarchical** models, the observations are given distributions conditional on parameters, and the parameters in turn have distributions conditional on additional parameters called hyperparameters.

## 9.1 Introduction

- Specifically, we begin by specifying a sampling model

$$y \sim p(y|\theta),$$

and the parameter  $\theta$  is assigned a prior distribution with unknown hyperparameter  $\lambda$ :

$$\theta \sim p(\theta|\lambda).$$

- The hyperparameter  $\lambda$  is in turn assigned another prior distribution

$$\lambda \sim p(\lambda).$$

More generally, the hierarchical structure may be extended to more than 2 levels.

## 9.1 Introduction

- A general way of constructing a hierarchical prior is based on the prior belief of exchangeability.
- A set of parameters  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  is **exchangeable** if the distribution of  $\boldsymbol{\theta}$  is unchanged when the parameter components are permuted. This implies that one's prior belief about  $\theta_j$  will be the same as one's belief about  $\theta_h$ .
- One can construct an exchangeable prior by assuming  $(\theta_1, \dots, \theta_k)$  are a random sample from a distribution  $p(\theta|\lambda)$  and the unknown hyperparameter  $\lambda$  is assigned a known prior at the second stage,  $\lambda \sim p(\lambda)$ .

# Outline

Introduction

Heart transplant mortality data

Individual and combined estimates

Equal mortality rates

Exchangeable model

Posterior inferences

- Shrinkage

- Comparing hospitals

- Bayesian Sensitivity Analysis

- Posterior predictive model checking



## 9.2 Heart transplant mortality data

- Let us consider a heart transplant mortality dataset, where the number of deaths within 30 days of heart transplant surgery is recorded for each of 94 hospitals. The expected number of deaths called **exposure** is also recorded.
- Let  $y_i$  and  $e_i$  denote the number of deaths and exposure for the  $i$ th hospital respectively. These are observed in the data.
- This dataset is contained in the **LearnBayes** package.

```
require(LearnBayes)
data(hearttransplants)
y <- hearttransplants$y
e <- hearttransplants$e
```

## 9.2 Heart transplant mortality data

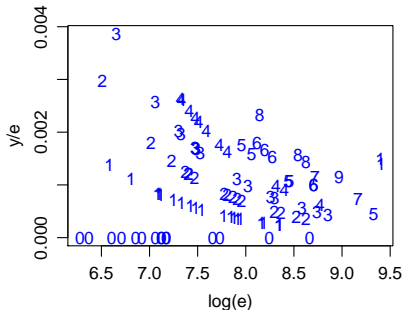
- A standard model assumes

$$y_i \sim \text{Poisson}(e_i \lambda_i)$$

and the objective is to estimate the mortality rate per unit exposure  $\lambda_i$ .

- The fraction  $y_i/e_i$  is the number of deaths per unit exposure and provides an estimate of the death rate for the  $i$ th hospital.

- The figure on the right plots  $\{y_i/e_i\}$  against  $\{\log e_i\}$  for all hospitals, where each point is labeled by  $y_i$ .
- The estimated rates are highly variable, especially for hospitals with small exposures.



# Outline

Introduction

Heart transplant mortality data

Individual and combined estimates

Equal mortality rates

Exchangeable model

Posterior inferences

- Shrinkage

- Comparing hospitals

- Bayesian Sensitivity Analysis

- Posterior predictive model checking

## 9.3 Individual and combined estimates

- Suppose we are interested in simultaneously estimating the true mortality rates  $\{\lambda_i\}$  for all hospitals. One option is to use the individual death rates

$$y_1/e_1, \dots, y_{94}/e_{94}.$$

- These rates can be poor estimates, especially for hospitals with small exposures, as there is high variability.
- Some hospitals did not experience any deaths and the individual death rate  $y_i/e_i = 0$  would likely underestimate the hospital's true mortality rate.

## 9.3 Individual and combined estimates

- It seems desirable to **combine the individual estimates** in some way to obtain improved estimates. In Bayesian hierarchical modeling, it is a common practice to **“borrow strength”** from across different subgroups of data.
- Suppose we can assume that the true mortality rates are equal across hospitals;  $\lambda_1 = \dots = \lambda_{94}$ . Under this “equal-means” Poisson model, the estimate of the mortality rate for the  $i$ th hospital would be the pooled estimate

$$\frac{\sum_{i=1}^{94} y_i}{\sum_{i=1}^{94} e_i}.$$

- However, this estimate is based on the strong assumption that the **true mortality rates are equal across hospitals**, which is questionable since one would expect some variation in the true rates.

## 9.3 Individual and combined estimates

- A third possibility is the compromise estimate

$$(1 - \omega) \frac{y_i}{e_i} + \omega \frac{\sum_{j=1}^{94} y_j}{\sum_{j=1}^{94} e_j}.$$

This estimate shrinks the individual estimate  $y_i/e_i$  towards the pooled estimate  $\sum_{i=1}^{94} y_i / \sum_{i=1}^{94} e_i$  where the parameter  $0 < \omega < 1$  determines the size of the shrinkage.

- We will see that this shrinkage estimate is a natural by-product of the application of an exchangeable prior model to the true mortality rates.

# Outline

Introduction

Heart transplant mortality data

Individual and combined estimates

Equal mortality rates

Exchangeable model

Posterior inferences

- Shrinkage

- Comparing hospitals

- Bayesian Sensitivity Analysis

- Posterior predictive model checking

## 9.4 Equal mortality rates

- Before considering an exchangeable model, let's illustrate fitting and checking the model where the mortality rates are assumed equal.
- Let  $\mathbf{y} = (y_1, \dots, y_{94})$ . Suppose  $y_i \sim \text{Poisson}(e_i \lambda)$  for  $i = 1, \dots, 94$  and the common mortality rate  $\lambda$  is assigned a prior  $p(\lambda) \propto \frac{1}{\lambda}$ . Then

$$\begin{aligned} p(\lambda|\mathbf{y}) &\propto p(\lambda) \prod_{i=1}^n p(y_i|\lambda) \propto \frac{1}{\lambda} \prod_{i=1}^n [\lambda^{y_i} \exp(-e_i \lambda)] \\ &\propto \lambda^{\sum_{i=1}^n y_i - 1} \exp\left(-\lambda \sum_{i=1}^n e_i\right). \end{aligned}$$

- This is recognized as a  $\text{Gamma}(a, b)$  density, where

$$a = \sum_{i=1}^n y_i = 277, \quad b = \sum_{i=1}^n e_i = 294681.$$

- Note that  $E(\lambda|\mathbf{y}) = \sum_{i=1}^n y_i / \sum_{i=1}^n e_i$  (the pooled estimate).



## 9.4 Equal mortality rates

- To check the suitability of this fitted model, we consider the posterior predictive distribution. Let  $y_i^*$  denote the number of transplant deaths for hospital  $i$  with exposure  $e_i$  in a future sample.

$$p(y_i^*|\mathbf{y}, e_i) = \int \underbrace{p(y_i^*|e_i, \lambda)}_{\text{Poisson}(e_i\lambda)} \underbrace{p(\lambda|\mathbf{y})}_{\text{Gamma}(a,b)} d\lambda.$$

It can be shown that  $p(y_i^*|\mathbf{y}, e_i)$  is a negative binomial distribution,  $\text{NB}\left(a, \frac{b}{b + e_i}\right)$  (refer to Chapter 2).

- If the number of observed deaths  $y_i$  is in the middle of this predictive distribution, then **our observation is consistent with our model fit**.
- On the other hand, if the observed  $y_i$  is in the extreme tails of  $p(y_i^*|\mathbf{y}, e_i)$ , then **the model is inadequate in fitting this observation**.

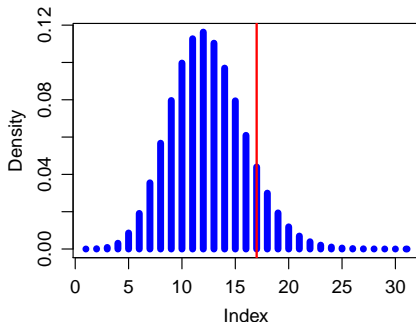
## 9.4 Equal mortality rates

- For illustration, let us consider hospital 94, for which  $e_{94} = 12131$ ,  $y_{94} = 17$ . Did this hospital have an unusually high number of deaths?
- The R-code below plots the posterior predictive distribution  $NB\left(277, \frac{294681}{294681+12131}\right)$  and adds a vertical red line showing the actual number of transplant deaths  $y_{94}$ .

```
a <- sum(y)
b <- sum(e)
i <- 94
grid <- seq(from=0,to=30,by=1)
plot(dnbinom(grid,size=a,prob=b/(b+e[i])),
     type="h",col="blue",lwd=6,ylab="Density")
abline(v=y[i],col="red",lwd=2)
```

## 9.4 Equal mortality rates

- Since the observed  $y_{94}$  is in the tail portion of the distribution, it seems inconsistent with the fitted model – it suggests that hospital 94 actually has a higher true mortality rate than estimated from this equal-rates model.



## 9.4 Equal mortality rates

- We can check the consistency of the observed  $y_i$  with its posterior predictive distribution for all hospitals. For each hospital, we compute the probability that the future observation  $y_i^*$  is at least as extreme as  $y_i$ :

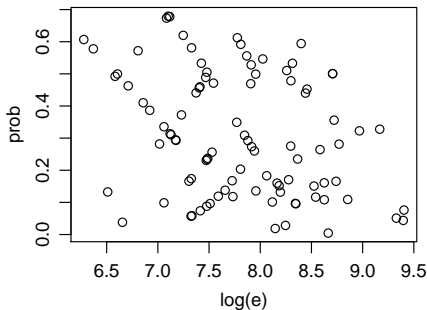
$$\min \{P(y_i^* \leq y_i), P(y_i^* \geq y_i)\}.$$

- The R-code below computes the probabilities of “at least as extreme” for all observations and plots the probabilities against the log exposures.

```
> prob1 <- pnbinom(y,size=a,prob=b/(b+e))
> prob2 <- 1- prob1 + dnbinom(y,size=a,prob=b/(b+e))
> prob <- apply(cbind(prob1,prob2),1,min)
> plot(log(e),prob)
> sum(prob<0.1)
[1] 15
```

## 9.4 Equal mortality rates

- A number of these tail probabilities appear small (15 are smaller than 0.10). Hence, the equal-rates model is inadequate for explaining the distribution of mortality rates for the group of 94 hospitals.
- We will have to assume differences between the true mortality rates, which will be modeled by the exchangeable model described next.



# Outline

Introduction

Heart transplant mortality data

Individual and combined estimates

Equal mortality rates

Exchangeable model

Posterior inferences

- Shrinkage

- Comparing hospitals

- Bayesian Sensitivity Analysis

- Posterior predictive model checking

## 9.5 Exchangeable model

- At the first stage of the prior, we assume that conditional on  $\alpha$  and  $\mu$ ,

$$\lambda_1, \dots, \lambda_{94} | \alpha, \mu \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha, \alpha/\mu).$$

The prior conditional mean and variance of  $\lambda$  are given by  $\mu$  and  $\mu^2/\alpha$  respectively.

- At the second stage of the prior, the hyperparameters  $\mu$  and  $\alpha$  are assumed independent, with

$$p(\mu) \propto \frac{1}{\mu}, \quad \mu > 0$$

and

$$p(\alpha) = \frac{z_0}{(\alpha + z_0)^2}, \quad \alpha > 0.$$

The parameter  $z_0$  is the prior median of  $\alpha$ . In this example, we let  $z_0 = 0.53$  (we will discuss the suitability of this value later).

## 9.5 Exchangeable model

- The joint posterior distribution is given by

$$p(\lambda_1, \dots, \lambda_{94}, \alpha, \mu | \mathbf{y}) \propto p(\mu)p(\alpha) \prod_{i=1}^{94} [p(y_i | \lambda_i) p(\lambda_i | \alpha, \mu)].$$

- The posterior distribution of  $\lambda_1, \dots, \lambda_{94}$ , conditional on  $\mu$  and  $\alpha$ , is

$$\begin{aligned} p(\lambda_1, \dots, \lambda_{94} | \mathbf{y}, \alpha, \mu) &\propto \prod_{i=1}^{94} [p(y_i | \lambda_i) p(\lambda_i | \alpha, \mu)] \\ &\propto \prod_{i=1}^{94} [e^{-e_i \lambda_i} \lambda_i^{y_i} \cdot \lambda_i^{\alpha-1} e^{-\alpha \lambda_i / \mu}] \propto \prod_{i=1}^{94} [e^{-\lambda_i (\alpha / \mu + e_i)} \lambda_i^{(\alpha + y_i) - 1}]. \end{aligned}$$

Therefore, conditional on  $\mu$  and  $\alpha$ ,  $\lambda_1, \dots, \lambda_{94}$ , have independent posterior distributions and

$$\lambda_i | y_i, \alpha, \mu \sim \text{Gamma}(\alpha + y_i, \alpha / \mu + e_i).$$



## 9.5 Exchangeable model

- This result is due to the conditionally independent structure of the hierarchical model and the choice of a conjugate prior at stage 2.
- It follows that

$$E(\lambda_i | y_i, \alpha, \mu) = \frac{\alpha + y_i}{\alpha/\mu + e_i} = B_i \mu + (1 - B_i) \frac{y_i}{e_i},$$

where  $B_i = \frac{\alpha}{\alpha + e_i \mu}$ .

- Thus the conditional posterior mean of  $\lambda_i$  can be viewed as a **shrinkage estimator**, where  $B_i$  is the shrinkage fraction of the posterior mean away from the estimate  $y_i/e_i$  toward the prior mean  $\mu$ .

## 9.5 Exchangeable model

- Since a conjugate model structure was used, the marginal posterior density of  $\mu$  and  $\alpha$  can be obtained (up to a proportionality constant) by integrating out the rates  $\lambda_1, \dots, \lambda_{94}$  from the joint posterior density:

$$\begin{aligned} p(\mu, \alpha | \mathbf{y}) &= \int p(\lambda_1, \dots, \lambda_{94}, \alpha, \mu | \mathbf{y}) \, d\lambda_1 \dots, d\lambda_{94} \\ &\propto \frac{1}{\mu(\alpha + z_0)^2} \prod_{i=1}^{94} \left\{ \frac{(\alpha/\mu)^\alpha}{\Gamma(\alpha)} \int e^{-\lambda_i(\alpha/\mu + e_i)} \lambda_i^{(\alpha+y_i)-1} d\lambda_i \right\} \\ &= \frac{1}{\mu(\alpha + z_0)^2} \prod_{i=1}^{94} \left\{ \frac{(\alpha/\mu)^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + y_i)}{(\alpha/\mu + e_i)^{\alpha+y_i}} \right\}. \end{aligned}$$

## 9.5 Exchangeable model

- The joint posterior density can be written as

$$p(\lambda_1, \dots, \lambda_{94}, \alpha, \mu | \mathbf{y}) = p(\lambda_1, \dots, \lambda_{94} | \alpha, \mu, \mathbf{y}) p(\alpha, \mu | \mathbf{y}).$$

Thus we can simulate random draws from the joint posterior by using the composition method:

### Simulating from the joint posterior

For  $t = 1, \dots, T$ ,

1. simulate  $(\alpha^{(t)}, \mu^{(t)}) \sim p(\alpha, \mu | \mathbf{y})$ ;
2. simulate  $(\lambda_1^{(t)}, \dots, \lambda_{94}^{(t)}) \sim p(\lambda_1, \dots, \lambda_{94} | \alpha^{(t)}, \mu^{(t)}, \mathbf{y})$ .

The second step can be performed easily since  $\lambda_1, \dots, \lambda_{94}$  have independent posterior distributions conditional on  $\mu$  and  $\alpha$ . Thus we can just simulate

$$\lambda_i^{(t)} | \alpha^{(t)}, \mu^{(t)} \sim \text{Gamma}(\alpha^{(t)} + y_i, \alpha^{(t)} / \mu^{(t)} + e_i), \text{ for } i = 1, \dots, 94.$$

## 9.5 Exchangeable model

- But first we need to simulate from the marginal posterior density of  $(\mu, \alpha)$ . Since both parameters are positive, we first transform each to the real-valued parameters,  $\theta_1 = \log \alpha$ ,  $\theta_2 = \log \mu$ .
- Applying the change of variables formula, the marginal posterior of  $\theta = (\theta_1, \theta_2)$  is given by

$$\begin{aligned} p(\theta|\mathbf{y}) &\propto \frac{1}{\mu(\alpha + z_0)^2} \prod_{i=1}^{94} \left\{ \frac{(\alpha/\mu)^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + y_i)}{(\alpha/\mu + e_i)^{\alpha+y_i}} \right\} \cdot \alpha\mu \\ &= \underbrace{\frac{\alpha}{(\alpha + z_0)^2} \prod_{i=1}^{94} \left\{ \frac{(\alpha/\mu)^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + y_i)}{(\alpha/\mu + e_i)^{\alpha+y_i}} \right\}}_{\exp(\ell(\theta))}. \end{aligned}$$

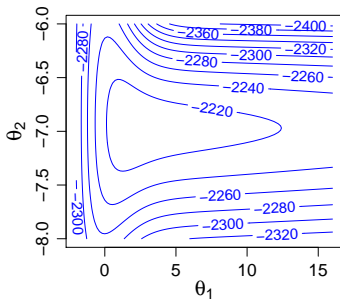
## 9.5 Exchangeable model

- The R-code below stores  $z_0$  and the  $y$  and  $e$  values in the list `data`, and writes a function `logpost` for computing  $\ell(\theta)$ . Note that we use the function `lgamma` to compute the log of the gamma function.

```
z0 <- 0.53
data <- list(z0=z0,y=y,e=e)
logpost <- function(theta, data){
  y <- data$y
  e <- data$e
  z0 <- data$z0
  n <- length(y)
  alpha <- exp(theta[1]); mu <- exp(theta[2])
  L <- (theta[1] - 2*log(alpha+z0) - n*lgamma(alpha)
        + sum(lgamma(alpha+y)) + 94*alpha*(theta[1]-theta[2])
        - sum((alpha+y)*log(alpha/mu+e)))
  return(L)
}
```

## 9.5 Exchangeable model

- The figure shows a contour plot of the log posterior  $\ell(\theta)$ . We see that the mode is close to  $(2, -7)$  and we use this value as the starting value in `optim`.



- We note that the posterior density of  $(\theta_1, \theta_2)$  is non-normal in shape, especially in the direction of  $\theta_1$ . A normal approximation to the posterior is likely inadequate and we obtain a simulated sample of  $\theta$  by using the Metropolis within Gibbs algorithm via the function `gibbs`.

## 9.5 Exchangeable model

- The R-code below uses the function `optim` to obtain the posterior mode, which is used as the starting value in the Metropolis within Gibbs algorithm.

```
> require(LearnBayes)
> start <- c(2,-7)
> out <- optim(par=start,fn=logpost,hessian=TRUE,
+             control=list(fnscale=-1),data=data)
> (post.mode <- out$par)
[1] 1.883954 -6.955446
> (post.cov <- -solve(out$hessian))
           [,1]      [,2]
[1,] 0.233694945 -0.003086655
[2,] -0.003086655 0.005866020
> post.sd <- sqrt(diag(post.cov))
```

## 9.5 Exchangeable model

- The scale is set at 2.5 times the posterior standard deviation of  $(\theta_1, \theta_2)$  in the normal approximation. The acceptance rates are about 45%. We can use other methods such as `rwmetrop` and `indepmetrop` as well.

```
> T <- 10000
> set.seed(1)
> fit <- gibbs(logpost,start=start,m=T,
               scale=2.5*post.sd,data=data)
> fit$accept
      [,1] [,2]
[1,] 0.4509 0.4301
```

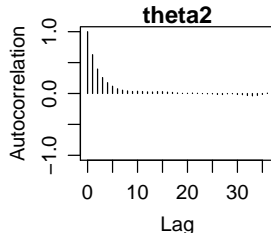
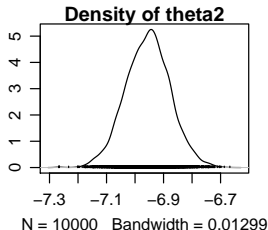
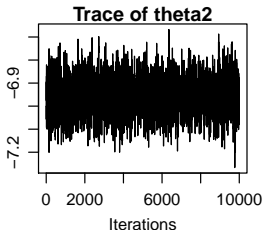
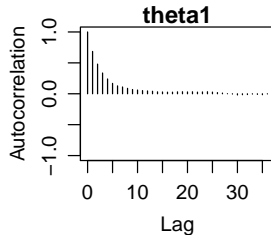
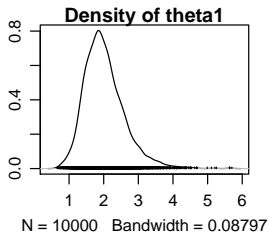
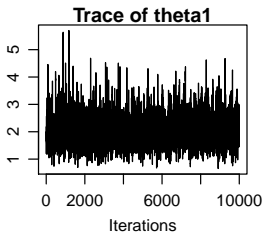
- Using the `coda` package, we obtain trace plots and autocorrelation plots.

```
require(coda)
mcmcobj <- mcmc(fit$par)
colnames(mcmcobj) <- c("theta1","theta2")
plot(mcmcobj)
autocorr.plot(mcmcobj, auto.layout=FALSE)
```



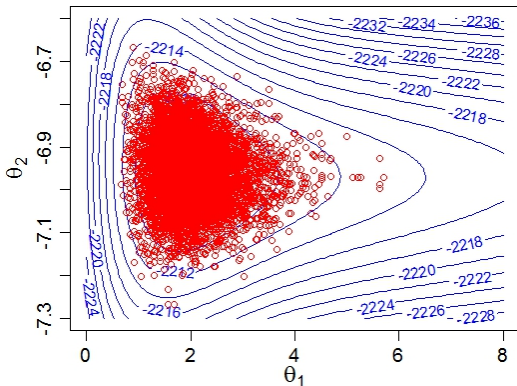
## 9.5 Exchangeable model

- The MCMC chains are mixing well. Estimates of the marginal posterior densities of  $\theta_1$  and  $\theta_2$  are also shown.



## 9.5 Exchangeable model

- The figure below shows a simulated sample of size 10000 placed on top of the contour graph. The algorithm appears to give a representative sample from the marginal posterior distribution of  $\theta_1$  and  $\theta_2$ .



## 9.5 Exchangeable model

- Now, we can simulate  $\lambda_1, \dots, \lambda_{94}$  from their posterior distributions:

$$\lambda_i^{(t)} | \alpha^{(t)}, \mu^{(t)} \sim \text{Gamma}(\alpha^{(t)} + y_i, \alpha^{(t)} / \mu^{(t)} + e_i), \text{ for } i = 1, \dots, 94.$$

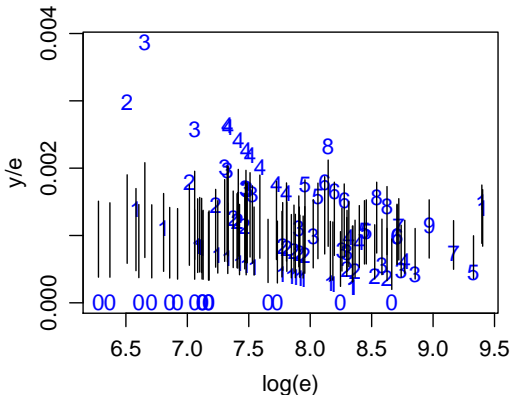
- The following R-code uses the `rgamma` function to simulate  $\lambda_i^{(t)}$  from the gamma distribution for  $t = 1, \dots, T$  and  $i = 1, \dots, 94$ .
- After obtaining a simulated sample of size 10000 for each  $\lambda_i$ , we summarize our results by computing a quantile-based 90% confidence interval for each  $\lambda_i$ .
- We graph these 90% confidence intervals as vertical lines on the original graph of the log exposures and the individual rates.

## 9.5 Exchangeable model

```
alpha_draws <- exp(fit$par[,1])
mu_draws <- exp(fit$par[,2])
lambda_draws <- matrix(0,T,94)
lambdaCI <- matrix(0,94,2)
for (i in 1:94){
  lambda_draws[,i] <- rgamma(10000,alpha_draws+y[i],
                             alpha_draws/mu_draws+e[i])
  lambdaCI[i,] <- quantile(lambda_draws[,i],c(0.05,0.95))
}
plot(log(e),y/e, pch=as.character(y))
for (i in 1:94){ lines(rep(log(e[i]),2),lambdaCI[i,]) }
```

## 9.5 Exchangeable model

- In contrast to the wide variation in the observed death rates, note the similarity in the locations of the probability intervals for the true rates. This indicates that these Bayesian estimates are shrinking the individual rates toward the pooled estimate.



# Outline

Introduction

Heart transplant mortality data

Individual and combined estimates

Equal mortality rates

Exchangeable model

Posterior inferences

- Shrinkage

- Comparing hospitals

- Bayesian Sensitivity Analysis

- Posterior predictive model checking

## 9.6 Posterior inferences

- Once a simulated sample of  $\{\lambda_1, \dots, \lambda_{94}, \mu, \alpha\}$  have been generated from the joint posterior distribution, we can use this sample to perform various types of inferences.

## 9.6.1 Shrinkage

- Previously, we have seen that

$$E(\lambda_i | y_i, \alpha, \mu) = B_i \mu + (1 - B_i) \frac{y_i}{e_i}, \quad \text{where } B_i = \frac{\alpha}{\alpha + e_i \mu}.$$

- We apply [the law of total expectation](#), i.e.  $E(Y) = E[E(Y|X)]$ . The posterior mean of  $\lambda_i$  is

$$\begin{aligned} E(\lambda_i | \mathbf{y}) &= E_{\alpha, \mu | \mathbf{y}} [E(\lambda_i | y_i, \alpha, \mu)] \\ &= E_{\alpha, \mu | \mathbf{y}} (B_i \mu) + (1 - E_{\alpha, \mu | \mathbf{y}} (B_i)) \frac{y_i}{e_i} \\ &\approx E_{\alpha, \mu | \mathbf{y}} (B_i) \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n e_i} + (1 - E_{\alpha, \mu | \mathbf{y}} (B_i)) \frac{y_i}{e_i}. \end{aligned}$$

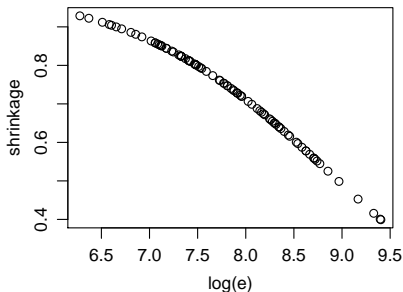
Note that  $\mu$  is the prior mean of  $\lambda$  and an estimate of  $\mu$  is  $\frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n e_i}$ .



## 9.6.1 Shrinkage

- In the R-code below, we compute the posterior mean of the shrinkage sizes  $\{B_i\}$  for all 94 hospitals and plot them against the logarithms of the exposures.

```
B <- rep(0,94)
for (i in 1:94){
  B[i] <- mean(alpha_draws /
    (alpha_draws + e[i]*mu_draws))
}
plot(log(e),B,ylab="shrinkage")
```



For hospitals with small exposures, the Bayesian estimate shrinks the individual estimate by 90% toward the combined estimate. In contrast, for large hospitals with high exposures, the shrinkage size is closer to 50%.

## 9.6.2 Comparing hospitals

- Suppose one is interested in comparing the true mortality rates of the hospitals, say by comparing the “best hospital” against the others.
- First, let us find the hospital with the smallest estimated mortality rate. In the R-code below, we compute the posterior mean of each  $\lambda_i$  and arrange them in increasing order.

```
> lambda_post.mean <- apply(lambda_draws,2,mean)
> order(lambda_post.mean)[1]
[1] 85
```
- Hospital 85 has the smallest estimated true mortality rate (in terms of the posterior mean).

## 9.6.2 Comparing hospitals

- Suppose we wish to compare hospital  $i$  with hospital  $j$ . Then

$P(\text{hospital } i \text{ has smaller mortality rate than hospital } j)$

$$= P(\lambda_i < \lambda_j) \approx \frac{1}{T} \sum_{t=1}^T \mathbb{1}\{\lambda_i^{(t)} < \lambda_j^{(t)}\}$$

In the R-code below, we compute the comparison probabilities for all pairs of hospitals and store the results in the matrix `compare`. The probability that hospital  $i$ 's rate is smaller than hospital  $j$ 's rate is stored in the  $(i,j)$  element of `compare`.

```
compare <- matrix(0,94,94)
for (i in 1:93){
  for (j in (i+1):94){
    compare[i,j] <- sum(lambda_draws[,i] < lambda_draws[,j])/T
    compare[j,i] <- 1 - compare[i,j]
  }
}
```

## 9.6.2 Comparing hospitals

- To compare hospital 85 with the rest, we look at the 85th column of the matrix `compare`. This gives the probabilities  $P(\lambda_i < \lambda_{85})$  for all  $i$ . We display these probabilities for the first 24 hospitals.

```
[1] 0.1864 0.1904 0.0797 0.1276 0.1354 0.2285 0.2061  
[8] 0.1601 0.0542 0.2116 0.1938 0.1754 0.2008 0.0955  
[15] 0.0657 0.2234 0.2305 0.0826 0.2672 0.1506 0.1424  
[22] 0.1479 0.0481 0.0769
```

- Hospital 85 is better than most of these hospitals since most of the posterior probabilities are close to zero.

## 9.6.3 Bayesian Sensitivity Analysis

- In any Bayesian analysis, it is important to assess the sensitivity of any inferences with respect to changes in the model assumptions, including assumptions about the sampling density and the prior.
- Here we briefly explore the sensitivity of our posterior inferences with respect to the choice of parameters in the prior distribution.
- In our prior, we assume  $\lambda_1, \dots, \lambda_{94} \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha, \alpha/\mu)$ , where  $p(\mu) \propto 1/\mu$  and  $p(\alpha) = z_0/(\alpha + z_0)^2$ .
- $z_0$  is the prior median of  $\alpha$ . Since  $\alpha$  controls the shrinkage of the individual estimates toward the pooled estimate, it is natural to ask about the sensitivity of the posterior of  $\alpha$  to specification of  $z_0$ .

## 9.6.3 Bayesian Sensitivity Analysis

- We focus on the posterior of  $\theta_1 = \log \alpha$  since the distribution of  $\theta_1$  is approximately symmetric and more amenable to inspection.
- The induced prior for  $\theta_1$  is

$$p(\theta_1|z_0) = \frac{z_0}{(e^{\theta_1} + z_0)^2} \cdot e^{\theta_1}.$$

- Suppose we set  $z_0 = 5$  instead of 0.53. Will this tenfold increase in the prior median of  $\alpha$  have a substantial impact on the posterior distribution of  $\theta_1$ ?

## 9.6.3 Bayesian Sensitivity Analysis

- The SIR algorithm provides a convenient way of converting simulated draws of  $\theta_1$  from one posterior distribution to a new distribution. The weights are given by

$$\begin{aligned}w(\theta_1) &= \frac{p(\theta_1|y, z_0 = 5)}{p(\theta_1|y, z_0 = 0.53)} \\&= \frac{p(y|\theta_1)p(\theta_1|z_0 = 5)/p(y)}{p(y|\theta_1)p(\theta_1|z_0 = 0.53)/p(y)} \\&= \frac{p(\theta_1|z_0 = 5)}{p(\theta_1|z_0 = 0.53)}.\end{aligned}$$

- We resample from the original posterior sample of  $\theta_1$  with sampling probabilities proportional to the weights to obtain the new posterior sample.

## 9.6.3 Bayesian Sensitivity Analysis

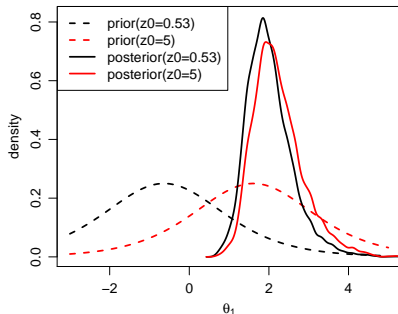
- The R-code below implements the SIR algorithm. The original draws of  $\theta_1$  are stored in `theta1_draws` and we write a function `logprior` for computing  $\log(p(\theta_1|z_0))$ . The log of the weights  $w(\theta_1)$  are stored in `logw`. The normalized weights are stored in `W` and we resample from the original draws using `W` as the sampling probabilities. The new draws of  $\theta_1$  are stored in `theta1_new`.

```
theta1_draws <- fit$par[,1]
logprior <- function(theta1,z0){
  log(z0) + theta1 - 2*log(z0 + exp(theta1))
}
logw <- logprior(theta1_draws,5)-logprior(theta1_draws,0.53)
w <- exp(logw-max(logw))
W <- w/sum(w)
idx <- sample(1:T, size=10000,replace=TRUE, prob=W)
theta1_new <- theta1_draws[idx]
```



## 9.6.3 Bayesian Sensitivity Analysis

- The figure below shows the prior and posterior densities for  $\theta_1$  corresponding to  $z_0 = 0.53$  and  $z = 5$ , and illustrates the impact of the choice of prior on the posterior inference of  $\theta_1$ .
- While the original and new priors are substantially different in location, the posteriors are similar in location.
- Thus the choice of  $z_0$  has only a modest effect on the posterior shrinkage of the model. This posterior inference appears to be robust to the change in prior specification of the median of  $\alpha$ .



## 9.6.3 Bayesian Sensitivity Analysis

- The figure can be obtained using the R-code below.

```
theta1_grid <- seq(from=-3,to=5,by=0.1)

plot(theta1_grid,exp(logprior(theta1_grid,0.53)),type="l",
     lty=2,ylim=c(0,0.82),lwd=2)
points(theta1_grid,exp(logprior(theta1_grid,5)),type="l",
      lty=2,col="red",lwd=2)
points(density(theta1_draws),type="l",lwd=2)
points(density(theta1_new),type="l",lwd=2,col="red")

legend("topleft",legend=c("prior(z0=0.53)","prior(z0=5)",
      "posterior(z0=0.53)","posterior(z0=5)"),lty=c(2,2,1,1),
      col=c("black","red","black","red"),lwd=2)
```

## 9.6.4 Posterior predictive model checking

- We have used the posterior predictive distribution to examine the suitability of the “equal-rates” model where  $\lambda_1 = \dots = \lambda_{94}$  and saw that this model seemed inadequate. Here we use the same methodology to check the appropriateness of the exchangeable model.
- Let us consider hospital 94 again, which experienced 17 deaths. To simulate from the predictive distribution of  $y_{94}$ , note that

$$p(y_{94}^*|\mathbf{y}) = \int p(y_{94}^*|\lambda_{94}) p(\lambda_{94}|\mu, \alpha, \mathbf{y}) p(\mu, \alpha|\mathbf{y}) d\mu d\alpha d\lambda_{94}.$$

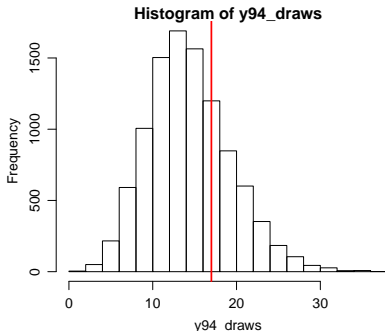
Since we have already simulated draws of  $(\mu, \alpha)$  from  $p(\mu, \alpha|\mathbf{y})$  and  $\lambda_{94}$  from  $p(\lambda_{94}|\mu, \alpha, \mathbf{y})$ , it remains to simulate  $y_{94}$  from  $p(y_{94}^*|\lambda_{94})$ .

## 9.6.4 Posterior predictive model checking

- This can be performed using the R-code below. We also create a histogram of  $y_{94}^*$  and add a vertical line on top, corresponding to the observed  $y_{94} = 17$ .

```
y94_draws <- rpois(T, e[94]*lambda_draws[,94])  
hist(y94_draws)  
abline(v=y[94],col="red",lwd=2)
```

- The observed number of deaths is in the middle of the predictive distribution, which indicates agreement of this observation with the fitted model.



## 9.6.4 Posterior predictive model checking

- We can also check the consistency of the observed  $y_i$  with its posterior predictive distribution for all hospitals.
- In the R-code below, we compute the probability that the future observation  $y_i^*$  is at least as extreme as  $y_i$  for all observations.

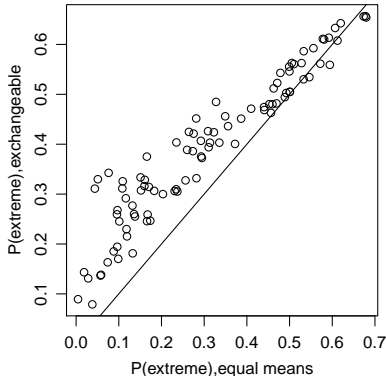
```
prob.exchange <- rep(0,94)
for (i in 1:94){
  yi_draws <- rpois(T, e[i]*lambda_draws[,i])
  prob1 <- sum(yi_draws <= y[i])/T
  prob2 <- sum(yi_draws >= y[i])/T
  prob.exchange[i] <- min(prob1,prob2)
}
```

## 9.6.4 Posterior predictive model checking

- To compare the goodness of fits of the two models, we plot these probabilities against those in the equal-means model (stored in `prob`). A comparison line  $y = x$  is then placed on top.

```
plot(prob, prob.exchange)  
abline(0,1)
```

- The probabilities of extreme for the exchangeable model are larger, indicating that the observations are more consistent with the exchangeable fitted model.
- Only two observations have a probability smaller than 0.1 for the exchangeable model, indicating general agreement of the observed data with this model.



## 9.6.4 Posterior predictive model checking

### Reading Assignment

- Please read **Chapter 8 in Peter Hoff's book**, which is another example of hierarchical models. The model involves group comparisons, in which each group is modeled by a normal distribution.
- The take-home final and the data project will cover the topics in Chapters 1–9.
- Due to the COVID-19 situation, our teaching activity has to end by Week 12. I will not have time to cover Bayesian linear regression, model comparison, variational Bayes, and software implementation such as Stan. I will upload some slides on these topics to a separate folder of **"Chapter X"** for your reference.