Project Bayesian Data Analysis

Bayesian Data Analysis DL (3792)
2nd year Master of Statistics and Data Science
Hasselt University
2023-2024

Group members:
Thanh Chuong Cao (2262621)
Joke Opdenacker (2261489)
Trang Tu Pham Hien (2261972)

Submission Date: 09 January, 2024

Lecturers:
Prof. Dr. Christel Faes
Dr. Oswaldo Gressani

1 PART 1

The Gibbs algorithm

An experiment is conducted in a laboratory to assess the strength of infection of $n = 8$ different artificial viruses on microorganisms. The number of infected organisms (y_i) during an exposure period (t_i) to a virus (v_i) is given below. Consider that the infections generated

Virus (v_i)	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
Infections (y_i)	4	1	5	14	3	19	7	6
Exposure period (ti)	95	16	63	126	6	32	16	19

by the ith virus follow a Poisson process with parameter λ_i, $i \in \{1, .., 8\}$, so that for an exposure time t_i, the number of infections y_i is Poisson distributed $y_i \sim \mathcal{P}(\lambda_i t_i)$ with mean $\mathbf{E}(y_i) = \lambda_i t_i$. In addition, assume the following Gamma prior distributions $\lambda_i \sim_{i.i.d} \mathcal{G}(\alpha, \beta)$ for $i = 1, .., n$ with $\alpha = 1.8$ and $\beta \sim \mathcal{G}(\gamma, \delta)$ with $\gamma = 0.01$ and $\delta = 1$, where $\mathcal{G}$ denotes a gamma distribution with the shape-rate parameterization.

Question 1

Using Bayes' theorem compute the joint posterior distribution $p(\lambda, \beta | \mathcal{D})$, where $\mathcal{D}$ denotes the observed data and $\lambda = (\lambda_1, ..., \lambda_{8]})^T$.

We specify the Baysian Poisson-gamma model based on information from the question:
Level 1: The number of infections y_i is Poisson distributed

$$y_i | \lambda_i \sim \mathcal{P}(\lambda_i t_i)$$

Level 2: Assuming the following Gamma prior distributions for λ_i for $i = 1, .., 8$

$$\lambda_i | \alpha, \beta \sim \mathcal{G}(\alpha, \beta)$$

with $\alpha = 1.8$; $\beta \sim \mathcal{G}(\gamma, \delta)$ where $\gamma = 0.01, \delta = 1$
With $\mathbf{y} = \{y_1, ..., y_n\}$, the joint posterior distribution function is:

$$p(\lambda, \beta | \mathcal{D}) \propto \prod_{i=1}^{8} p(y_i | \lambda_i, \beta) \prod_{i=1}^{8} p(\lambda_i | \beta) p(\beta) \tag{1}$$

Because of hierarchical independence of y_i, $p(y_i | \lambda_i, \beta) = p(y_i | \lambda_i)$ and $p(\lambda_i | \beta) = p(\lambda_i | \alpha, \beta)$, the notation becomes :

$$p(\lambda, \beta | \mathcal{D}) \propto \prod_{i=1}^{8} p(y_i | \lambda_i) \prod_{i=1}^{8} p(\lambda_i | \alpha, \beta) p(\beta)$$

Therefore, we have:

$$p(\lambda, \beta | \mathcal{D}) \propto \prod_{i=1}^{8} \frac{(\lambda_i t_i)^{y_i}}{y_i!} exp(-\lambda_i t_i). \prod_{i=1}^{8} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} exp(-\beta \lambda_i). \frac{\delta^\gamma}{\Gamma(\gamma)} \beta^{\gamma-1} exp(-\delta \beta)$$

Question 2

Compute the following conditional posterior distributions $p(\lambda_i | \lambda_{-\mathbf{i}}, \beta, \mathcal{D})$ for $i = 1, .., 8$ and $p(\beta | \lambda, \mathcal{D})$. To what well-known parametric family do these conditional posterior distributions belong?

The first conditional posterior distribution can be expressed as:

$$p(\lambda_i | \lambda_{-\mathbf{i}}, \beta, \mathcal{D}) = \frac{p(\lambda | \beta, \mathcal{D})}{p(\lambda_{-\mathbf{i}} | \beta, \mathcal{D})} = \frac{p(\lambda, \beta | \mathcal{D}) / p(\beta | \mathcal{D})}{p(\lambda_{-i}, \beta | \mathcal{D}) / p(\beta | \mathcal{D})} = \frac{p(\lambda, \beta | \mathcal{D})}{p(\lambda_{-i}, \beta | \mathcal{D})}$$

Then from the joint posterior kernel (1) we can obtain:

$$p(\lambda_i | \lambda_{-\mathbf{i}}, \beta, \mathcal{D}) = p(y_i | \lambda_i) p(\lambda_i | \beta)$$

1

Therefore, the notation for this conditional posterior distribution becomes:

$$p(\lambda_i|\lambda_{-\mathbf{i}}, \beta, \mathcal{D}) \propto \lambda_i^{y_i+\alpha-1} exp[-(t_i+\beta)\lambda_i]$$

For the second conditional posterior distribution, we have:

$$p(\beta|\lambda, \mathcal{D}) = \frac{p(\lambda, \beta|\mathcal{D})}{p(\lambda|\mathcal{D})} = \frac{p(\lambda, \beta|\mathcal{D})}{\int p(\lambda, \beta|\mathcal{D})d\beta}$$

We can compute the marginal posterior $p(\lambda|\mathcal{D})$ (the denominator) from doing integral the joint posterior density over β:

$$p(\lambda|\mathcal{D}) \propto \prod_{i=1}^{8} \frac{(\lambda_i t_i)^{y_i}}{y_i!} exp(-\lambda_i t_i) . \prod_{i=1}^{8} \frac{\lambda_i^{\alpha-1}}{\Gamma(\alpha)} . \frac{\delta^{\gamma}}{\Gamma(\gamma)} \int \beta^{8\alpha+\gamma-1} exp[-(\sum_{i=1}^{8} \lambda_i + \delta)\beta]d\beta$$

Since the integral component is against Gamma family function, the notation becomes:

$$p(\lambda|\mathcal{D}) \propto \prod_{i=1}^{8} \frac{(\lambda_i t_i)^{y_i}}{y_i!} exp(-\lambda_i t_i) . \prod_{i=1}^{8} \frac{\lambda_i^{\alpha-1}}{\Gamma(\alpha)} . \frac{\delta^{\gamma}}{\Gamma(\gamma)} . \frac{\Gamma(8\alpha+\gamma)}{(\sum_{i=1}^{8} \lambda_i + \delta)^{8\alpha+\gamma}}$$

The second conditional posterior distribution therefore has a kernel following:

$$p(\beta|\lambda, \mathcal{D}) \propto \beta^{8\alpha+\gamma-1} exp[-(\sum_{i=1}^{8} \lambda_i + \delta)\beta]$$

Conclude: These conditional posterior distributions belong to Gamma parametric family.

Question 3

Based on the full conditionals obtained in the previous step, write a pseudo-code to obtain a random sample of total size 35 000 (including a burn-in of size 5 000) from $p(\lambda, \beta|\mathcal{D})$ using the Gibbs sampler.

Pseudo code: Gibbs sampler to draw from $p(\lambda, \beta|\mathcal{D})$

1. Fix initial value $\lambda^{(\mathbf{0})}$

2. for $m = 1, ..., 35000$ **do**:

3. $\beta^{(m)} \sim \mathcal{G}(8\alpha + \gamma, \sum_{i=1}^{8} \lambda_i + \delta)$

4. $\lambda_1^{(m)} \sim \mathcal{G}(y_1 + \alpha, t_1 + \beta^{(m)})$
   $\lambda_2^{(m)} \sim \mathcal{G}(y_2 + \alpha, t_2 + \beta^{(m)})$
   .
   .
   .
   $\lambda_8^{(m)} \sim \mathcal{G}(y_8 + \alpha, t_8 + \beta^{(m)})$

5. **end** for

6. Discard burn in parts: $\beta^{(1)}, \beta^{(2)}, ...\beta^{(5000)}, \lambda^{(1)}, \lambda^{(2)}, ...\lambda^{(5000)}$

Under mild regularity and if convergence occurs, $\lambda^{(m)}$, $\lambda^{(m+1)}$,...,$\beta^{(m)}$, $\beta^{(m+1)}$,... can be considered as drawn from the joint posterior distributions.

Question 4

Using R and without relying on external software/packages (except the *coda* package), write the Gibbs sampler based on your pseudo-code in step 3. Please specify *set.seed*(2025) at the beginning of your code for reproducibility of your results. What is your acceptance rate?

We performed Gibbs sampling with 35 000 iterations as specified in pseudo-code in step 3. We set initial values of $\lambda_0 = (0.1, 0.1, ..., 0.1)$. After discarding burn in parts (5 000 iterations), we summarised our results by histograms, trace plots and summary statistics of MCMC chains, including means, medians, 95% HDP intervals and 95% equal tail credible intervals (Figure
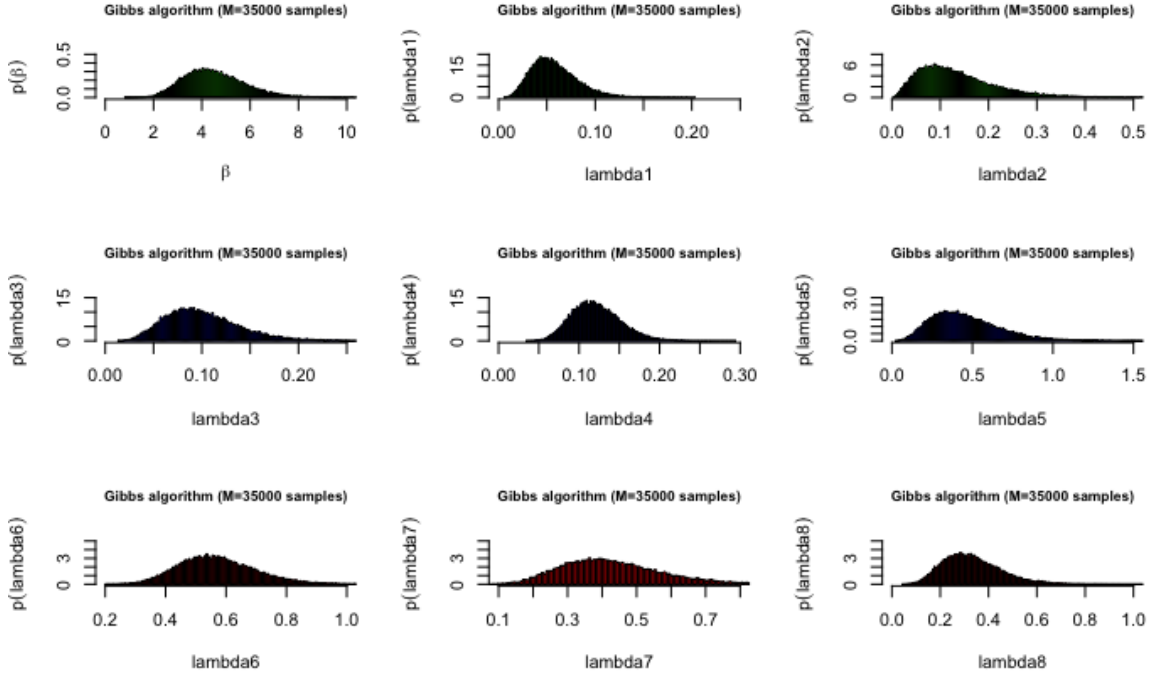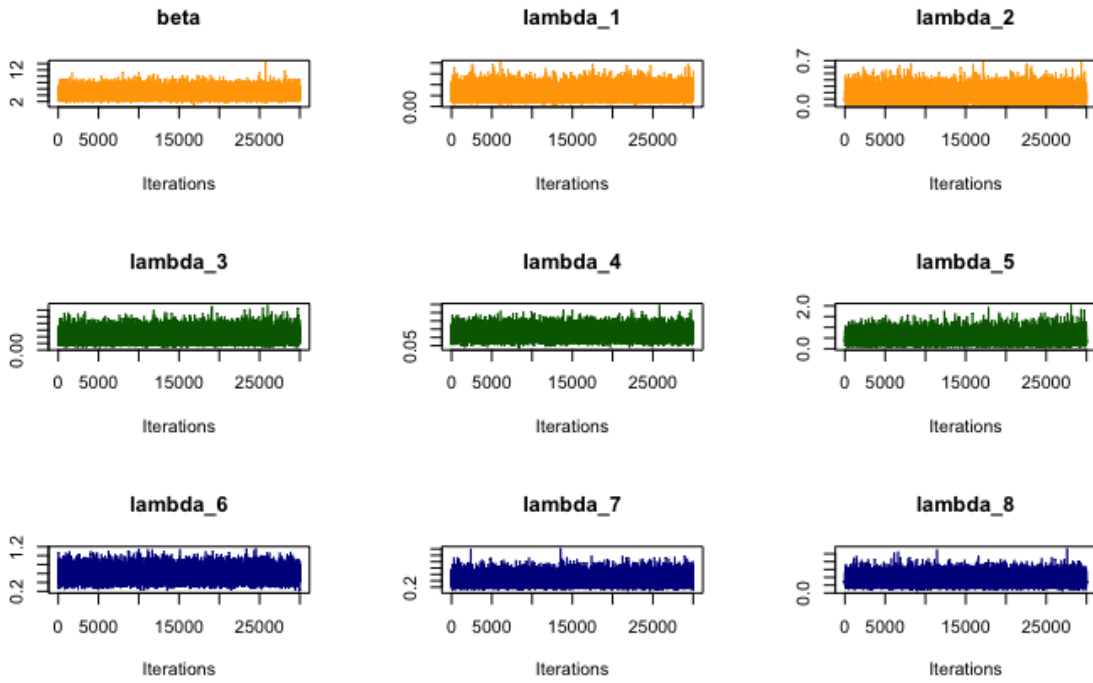
2

Figure 1: Histograms of sampling β and λ

Figure 2: Traceplots of β and λ chains

Table 1: Summary of Gibbs sampler chains

	Mean	Median	95% HDP interval	95% equal tail CI
β	4.5381	4.3965	(2.2586, 7.1863)	(2.4077, 7.4625)
λ_1	0.0581	0.0548	(0.0160, 0.1052)	(0.0210 0.1141)
λ_2	0.1367	0.1201	(0.0129, 0.2992)	(0.0263, 0.3403)
λ_3	0.1007	0.0957	(0.0324, 0.1770)	(0.0395, 0.1899)
λ_4	0.1212	0.1186	(0.0648, 0.1813)	(0.0692, 0.1872)
λ_5	0.4632	0.4280	(0.0951, 0.8980)	(0.1399, 0.9874)
λ_6	0.5710	0.5611	(0.3356, 0.8253)	(0.3513, 0.8483)
λ_7	0.4304	0.4130	(0.1700, 0.7279)	(0.1925, 0.7688)
λ_8	0.3317	0.3177	(0.1185, 0.5703)	(0.1403, 0.6103)

1, 2 and Table 1)

For Gibbs sampler techniques, which was based on full conditional distributions, the acceptance rate was indeed 100%.

Question 5

Use the Geweke convergence diagnostic test and the Heidelberger-Welch stationarity test on the generated chains and interpret your results.

Geweke test is a formal convergence diagnostics, which can be used for checking stationary. Geweke test statistic was calculated to compare means of an early part (10%) to late part (50%) of chain: $Z = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\frac{\hat{\tau}_A^2}{n_A} + \frac{\hat{\tau}_B^2}{n_B}}}$, where $n_A = 0.1 * 30000 = 3000$, $n_B = 0.5 * 30000 = 15000$.

We also applied Z test on $\frac{K-m}{K}\%$ last last iterations of the chain where $m = (0, 1, ..., K), K = 20$ to produce Z_m test statistics and plotted in a time series.

As can be seen from Table 2, there was no any Z value falling outside the acceptance region at 5% level of significance. Therefore, we have no evidence to reject the null hypothesis that mean of the parameter in late part is difference from the early part. The dynamic Geweke test shown very few values of $\beta, \lambda_1, \lambda_2, \lambda_6$ exceeded this normal range, suggesting chains all achieved good convergence.

Table 2: Geweke convergence diagnostic test results

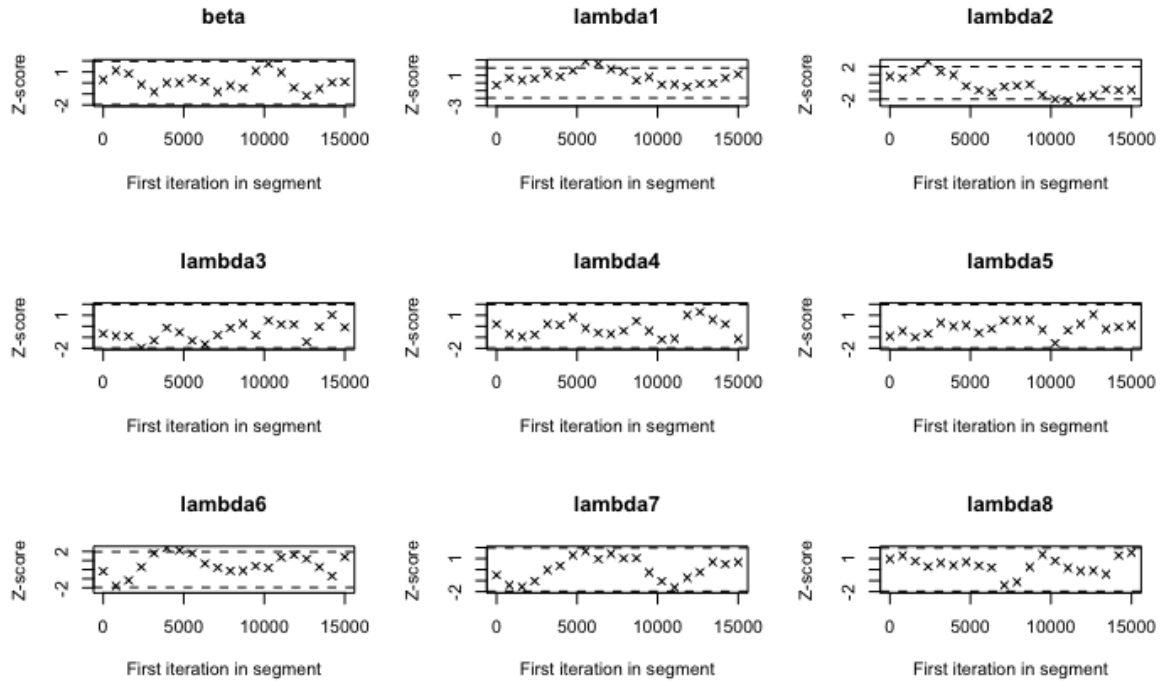|	1st window (%)	2nd window (%)	Z value	Acceptance region ($\alpha = 0.05$)
β	0.1	0.5	0.2944	(-1.96,1.96)
λ_1	0.1	0.5	-0.2912	(-1.96,1.96)
λ_2	0.1	0.5	0.7852	(-1.96,1.96)
λ_3	0.1	0.5	-0.6844	(-1.96,1.96)
λ_4	0.1	0.5	0.1914	(-1.96,1.96)
λ_5	0.1	0.5	-0.9028	(-1.96,1.96)
λ_6	0.1	0.5	-0.187	(-1.96,1.96)
λ_7	0.1	0.5	-0.4995	(-1.96,1.96)
λ_8	0.1	0.5	0.9646	(-1.96,1.96)

Figure 3: Time series plot of Z_m Geweke test statistic

The Heidelberger-Welch convergence diagnostic have two parts:

Part 1: We tested null hypothesis that the chain was from a stationary distribution. We pre-specified a level of significance at 5%. At first, this test calculated the Cramer von Mise test statistic for whole chains. As can be seen from Table 3, all p-values were larger than 0.05, therefore we did not reject any null hypothesis. It suggested that there was no part discarded

4

Table 3: Heidelberger-Welch stationary diagnostic test results

	Start iteration	p-value	Mean	Halfwidth	Ratio
β	1	0.9710	4.5400	0.0168	0.004
λ_1	1	0.1700	0.0581	0.0003	0.005
λ_2	1	0.2430	0.1370	0.0009	0.007
λ_3	1	0.2450	0.1010	0.0004	0.004
λ_4	1	0.7470	0.1210	0.0003	0.002
λ_5	1	0.6370	0.4630	0.0027	0.006
λ_6	1	0.1760	0.5710	0.0015	0.003
λ_7	1	0.9130	0.4300	0.0018	0.004
λ_8	1	0.5970	0.3320	0.0014	0.004

and all single MCMC chains passed the first part after just one iteration. They were then used in the second part.

Part 2: We estimated mean of parameters from 9 single MCMC chains (for 9 parameter) passing part 1. Since there was no discarded, these estimates coincided with results from table 1 (except some rounded values). As we specified the target value for the ratio of halfwidth to sample mean of 0.1, results from table shown all chains passed the halfwidth test.

Two parts of Heidelberger-Welch convergence diagnostic demonstrated the single MCMC chains for 9 parameters achieved good convergence.

Question 6

Based on the posterior mean of your generated chain, provide a point es- timate and 95% quantile-based credible interval for $\mathbf{E}(y_6)$, i.e. the average number of infections generated by virus v_6 for $t_6 = 32$ minutes. (Round your results to the nearest integer).

Since the number of infections y_6 is Poisson distributed: $y_6 \sim \mathcal{P}(\lambda_6 t_6)$, the expectation is expressed as: $\mathbf{E}(y_6) = \lambda_6 t_6$.

Because λ_6 is a stochastic variable, $\mathbf{E}(y_6)$ is a function λ_6. Therefore, mean of $\mathbf{E}(y_6)$ can be expressed as: $t_6 E(\lambda_6)$.

Due to monotonic increasing nature of that function, a transformation from λ_6 to $t_6\lambda_6$ is in- variant. Consequently, we can use the 95% quantile-based credible interval of λ_6 to compute the 95% quantile-based credible interval of $\mathbf{E}(y_6)$

Conclusion:

Mean of $\mathbf{E}(y_6)$ was rounded to: 18 (18.27043)

95% equal tail CI was rounded to: (11, 27) (11.2426, 27.1455)

Question 7

Using your generated chain, estimate the probability $\mathbf{P}(\lambda_6 > 0.53)$

We calculated p Monte Carlo value for which $\lambda_6 > 0.53$:

$p_{mc} = \frac{1+sum(\lambda_6 > 0.53)}{M+1}$, where $M = 30000$.

Conclusion: $\mathbf{P}(\lambda_6 > 0.53) = 60.0247\%$

The Metropolis algorithm

Let $\theta = (\theta_1, \theta_2)^\top$, and assume the following posterior distribution, with $A = 1.90$, $B = 0.54$, $C_1 = 0.5$ and $C_2 = -1.4$

$$p(\theta|\mathcal{D}) \propto exp(-\frac{1}{2}(A\theta_1^2\theta_2^2 + \theta_1^2 + \theta_2^2 - 2B\theta_1\theta_2 - 2C_1\theta_1 - 2C_2\theta_2))$$

Question 1

Write the log posterior distribution log $p(\theta|\mathcal{D})$ and obtain analytically the gradient $\nabla_\theta logp(\theta|\mathcal{D}) = (\partial \log p(\theta|\mathcal{D})/\partial\theta_1; \partial\log p(\theta|\mathcal{D})/\partial\theta_2)$ and Hessian matrix.

The log posterior distribution could be written as:

$$\log p(\theta|\mathcal{D}) \propto -\frac{1}{2}(A\theta_1^2\theta_2^2 + \theta_1^2 + \theta_2^2 - 2B\theta_1\theta_2 - 2C_1\theta_1 - 2C_2\theta_2)$$

By taking the first and second-order partial derivatives with regard to θ_1 and θ_2, of log $p(\theta|\mathcal{D})$, we obtain the gradient as follow:

$$\nabla_\theta logp(\theta|\mathcal{D}) = (-A\theta_1\theta_2^2 - \theta_1 + B\theta_2 + C_1; \ -A\theta_1^2\theta_2 - \theta_2 + B\theta_1 + C_2)^\top$$

And the Hessian matrix:

$$\nabla_\theta^2 logp(\theta|\mathcal{D}) = \begin{pmatrix} -A\theta_2^2 - 1 & -2A\theta_1\theta_2 + B \\ -2A\theta_1\theta_2 + B & -A\theta_1^2 - 1 \end{pmatrix}$$

Question 2

Using the gradient and Hessian matrix obtained in the previous step, implement a Newton-Raphson algorithm in R to find the posterior mode of $p(\theta|\mathcal{D})$ and denote by $\theta_{NR}^* = (\theta_1^*; \theta_2^*)^\top$ the mode after convergence of the algorithm.

The Pseudo-code of Newton-Raphson algorithm to find mode of $p(\theta|\mathcal{D})$ is as follow:

1. Set tolerance ϵ, initial distance d, initialze $\theta^{(0)}$ and set $m = 0$

2. **while** $d > \epsilon$ **do**

3. $\theta^{m+1} = \theta^m - (\nabla^2 \log p(\theta^{(m)}|\mathcal{D}))^{-1}\nabla\log p(\theta^{(m)}|\mathcal{D})$

4. Compute distance $d = ||\theta^{m+1} - \theta^{(m)}||$

5. **end while**

6. Return $theta_M$

The following R code was implemented to obtain the posterior mode:

```
  NewtonRaphson <- function(theta0=c(1,2), tolerance = 1e-10, iteration = 1000){
# Define constants
A <- 1.9
B <- 0.54
C1 <- 0.5
C2 <- -1.4
#Calculate gradients
gradient_theta <- function(theta=c(1,2))
{
  t1 <- theta[1]
  t2 <- theta[2]
  grad1 <- -A*t1*t2^2 -t1 + B*t2 + C1
  grad2 <- -A*t1^2*t2 -t2 + B*t1 + C2
  return(c(grad1, grad2))
}
 #Calculate Hessian matrix
 hessian_theta <- function(theta=c(1,2))
 {
   t1 <- theta[1]
```

```
    t2 <- theta[2]
    h11 <- -A*t2^2 - 1
    h12 <- h21 <- -2*A*t1*t2 + B
    h22 <- -A*t1^2 - 1
    return(matrix(c(h11,h12, h21, h22), nrow=2))
  }
  for (i in 1: iteration) {
    grad_theta0 <- gradient_theta(theta0)
    hes_theta0 <- hessian_theta(theta0)
    theta1 <- theta0 - solve(hes_theta0,grad_theta0) #calculate next value of theta
    distance <- sum(abs(theta1 - theta0))
    if (distance < tolerance) { #Once distance gets sufficiently small, output result
      result <- round(theta1, 5)
      return(list("mode theta" = result, paste("Converged at iteration", i)))
    }
    # If convergence not reached set theta1 as theta0 and continue
    theta0 <- theta1
  }
  print('Newton-Raphson algorithm did not converge. Choose larger number of iteration')
}
```

Using the above algorithm, The posterior mode of $p(\theta|\mathcal{D})$ was found to locate at $\theta^*_{NR} = (-0.0553; -1.4216)^\top$

Question 3

Compute a Laplace approximation to $p(\theta|\mathcal{D})$ around θ^*_{NR}. Report the covariance matrix Σ^* of the Laplace approximation.

Using Laplace approximation around θ^*_{NR}, the posterior distribution $p(\theta|\mathcal{D})$ is approximated by a Gaussian distribution with

- Mean $\mu_0 = \theta^*_{NR} - (\nabla^2 \log p(\theta^*_{NR}|\mathcal{D}))^{-1} \nabla \log p(\theta^*_{NR}|\mathcal{D}))$

- Covariance matrix $\Sigma^* = (\nabla^2 \log p(\theta^*_{NR}|\mathcal{D}))^{-1}$

With $\nabla \log p(\theta^*_{NR}|\mathcal{D})$ and $\nabla^2 \log p(\theta^*_{NR}|\mathcal{D}))^{-1}$ the gradient and Hessian, respectively, of the log target posterior distribution, evaluated at θ^*_{NR}.

Using the calculation from Question 1, the covariance matrix Σ^* is:

$$\Sigma^* = \begin{pmatrix} 0.20912 & 0.05016 \\ 0.05016 & 1.00626 \end{pmatrix}$$

And the mean μ_0:

$$\mu_0 = \begin{pmatrix} -0.0553 \\ -1.4216 \end{pmatrix}$$

Thus the distribution of the approximated posterior distribution is:

$$\tilde{p}_G(\theta|\mathcal{D}) \sim N(\begin{pmatrix} -0.0553 \\ -1.4216 \end{pmatrix}, \begin{pmatrix} 0.20912 & 0.05016 \\ 0.05016 & 1.00626 \end{pmatrix})$$

Question 4

In R write a random-walk Metropolis algorithm to explore the joint posterior $p(\theta|\mathcal{D})$ using a Gaussian proposal with covariance matrix $\tilde{\Sigma} = c\Sigma^*$. Use a chain of length $M = 50\ 000$ and tune c to (approximately) reach the optimal acceptance rate of 23%. Specify set.seed(1993) for reproducibility.

The pseudo-code of the random-walk Metropolis algorithm is as follow:

1. Choose initial value θ^0 satisfying $p(\theta^0|\mathcal{D}) > 0$

2. For m in 1 to M do:

2.1 sample $\tilde{\theta}^{(m)} \sim q(\cdot | \theta^{(m-1)})$

2.2 Compute the log ratio $log\varrho(\theta^{(m-1)}, \tilde{\theta}^{(m)}) = log\frac{p(\tilde{\theta}^{(m)}|\mathcal{D}}{p(\theta^{(m-1)})|\mathcal{D}}$

2.3 Sample $u \sim U(0,1)$

2.4 If $log(\varrho) \geq 0 \parallel u \leq exp(log(\varrho))$, set $\theta^{(m)} \leftarrow \tilde{\theta}^{(m)}$ (accept candidate)

2.5 else $\theta^{(m)} \leftarrow \tilde{\theta}^{(m-1)}$ (reject candidate)

3. end for.

The following Metropolis algorithm was implemented in R to explore $p(\theta|\mathcal{D})$:

```
  Metropolis <- function(M =60000, burn_in = 10000, seed = 1993,
                     theta_start = c(-0.0553, -1.4216), c_tune=6.2 )
{
  set.seed(seed)

  # Joint log posterior distribution
  log_posterior <- function(theta)
  {
    t1 <- theta[1]
    t2 <- theta[2]
    return(-1/2*(A*t1^2*t2^2 + t1^2 + t2^2 - 2*B*t1*t2 - 2*C1*t1 -2*C2*t2))
  }

  # Starting value for theta 1 and theta 2
  theta <- array(dim=c(2,M))
  theta[, 1] <- theta_start
  n_accept <- 0

  # Gaussian proposal
  # Covariance matrix
  theta_NR_star <- NewtonRaphson()$'mode theta' #The NewtonRaphson function built in Q2
  sigma_star <- round(solve(-hessian_theta(theta = theta_NR_star)),5)
  sigma_tilde <- c_tune*sigma_star

  # Metropolis loop
  for (i in 2:M)
  {
    # New proposed theta
    theta_prop <- theta[,i-1] + rmvnorm(n = 1,
                                        mean = c(0,0), sigma = sigma_tilde)
    # Log ratio for accept-reject decision
    logr <- log_posterior(theta_prop) - log_posterior(theta[,i-1])

    # Accept or reject
    u <- runif(1)
    if(logr >= 0 || u <= exp(logr))
    { #Accept
      theta[,i] <- theta_prop
      n_accept <- n_accept + 1
    }
    else
    { #Reject, stay where theta is
      theta[,i] <- theta[, i-1]
    }
  }
  # Exclude burn-in
  theta <- theta[,-c(1:burn_in)]
  # Output
  accept_rate <- round(n_accept/(M-1), digits = 3) * 100
```

```
rownames(theta) <- c("Theta1","Theta2")
output <- list(theta= theta,
               M = M,
               burn_in = burn_in,
               log_posterior = log_posterior)
print(paste("Acceptance rate:", accept_rate,"%"))
return(invisible(output))
}
```

Our MCMC chain had 60000 iterations, in which the first 10000 were burn-in period. The starting point of theta was set at θ_{NR}^*. The tuning factor c was set at 6.2 to give the acceptance rate of 23.2%. The trace plots, Q-Q plots, running mean plots and Geweke diagnostic (Figure 4, 5, 6 and 7) showed good sign of convergence.
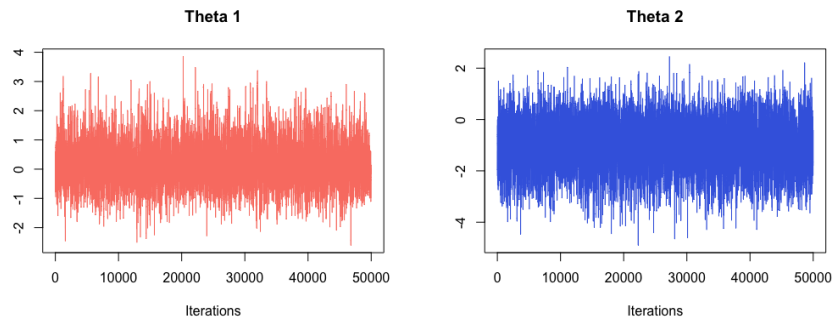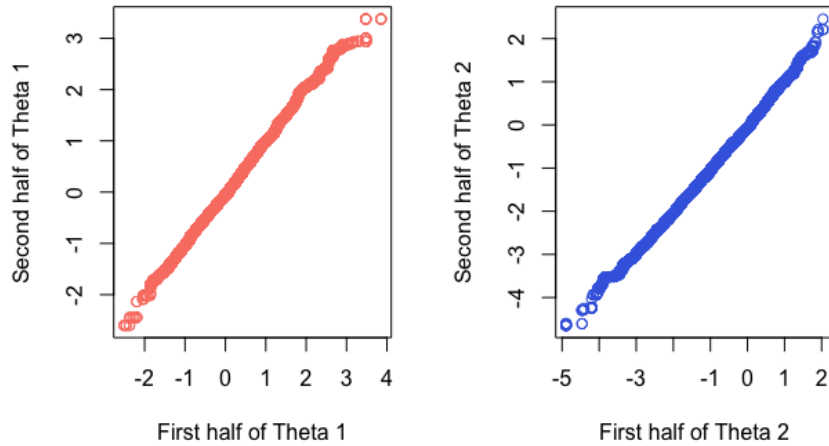

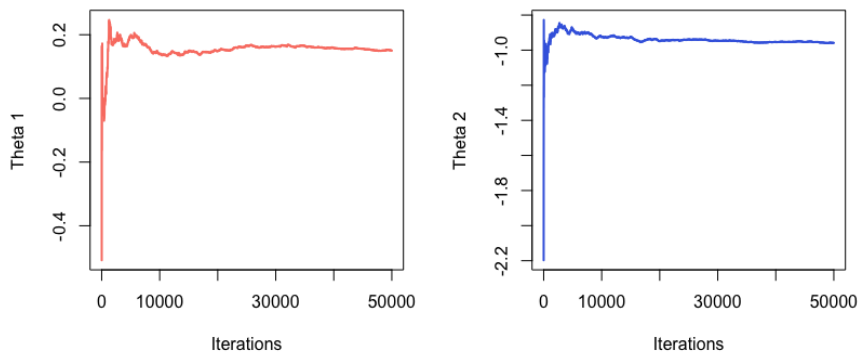Figure 4: Trace plot of θ_1 and θ_2

Figure 5: Q-Q plots of θ_1 and θ_2

Figure 6: Running mean plot of θ_1 and θ_2

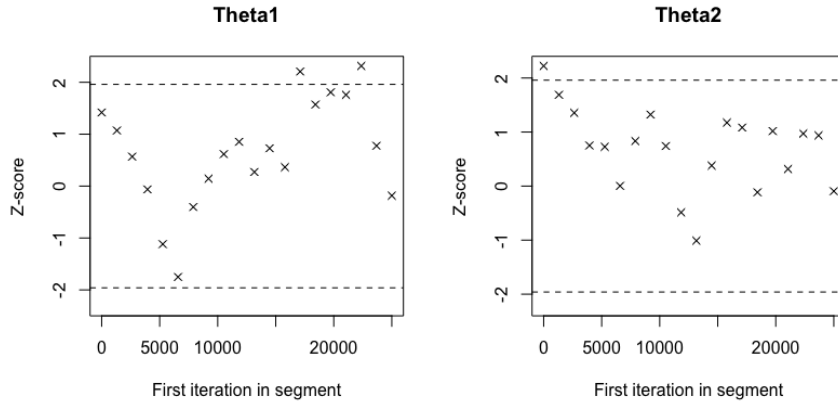The summary statistics of posterior θ_1 and θ_2 is given in Table 4.

9

Figure 7: Time series plot of Geweke test statistic for θ_1 and θ_2

Table 4: Summary measures of the posterior distribution of θ_1 and θ_2

Parameter	Posterior mean	Posterior median	Posterior SD	95% equal tail CI	95% HPD interval
θ_1	0.150	0.064	0.728	-1.160 ; 1.895	-1.317 ; 1.664
θ_2	-0.958	-0.897	0.936	-2.930 ; 0.687	-2.829 ; 0.733

Question 5

Using your generated chains, estimate of the probability $P((\theta_1/\theta_2) > 0.45)$

The following codes were implemented in R to obtain $P((\theta_1/\theta_2) > 0.45)$

```
MCMC_run1 <- Metropolis(c_tune = 6.2, M = 60000, burn_in = 10000, seed = 1993,
                        theta_start = c(-0.0553, -1.4216))
theta_run1 <- MCMC_run1$theta
ratio_theta12 <- theta_run1[1,]/ theta_run1[2,]
p_ratio_045 <- sum(ifelse(ratio_theta12 > 0.45, 1, 0))/ length(ratio_theta12)
p_ratio_045
```

The probability $P((\theta_1/\theta_2) > 0.45)$ is calculated as the proportion of the MCMC chain that gives the ratio $\theta_1/\theta_2 > 0.45$. The probability $P((\theta_1/\theta_2) > 0.45)$ is estimated to be **0.237**.

2 PART 2

The German socioeconomic panel study data was taken from the first twelve an- nual waves (1984 through 1995) of the German Socioeconomic Panel (GSOEP) which surveys a representative sample of East and West German households. The data provide detailed information on the utilization of health care facilities, characteristics of current employment, and the insurance schemes under which individuals are covered. We consider a random sample of 100 individuals aged 25 through 65 from the West German subsample and of German nationality, which participated throughout all years. We want to investigate whether there is a link between employment status and individual characteristics (age, gender, health status, etc), and the evolution of the employment status over time.

Question 1

Fit a logistic regression model and investigate the link between working status and the individual variables. Take vague priors for all model parameters. Select the most important variables explaining the working status of an individual using DIC and/or WAIC. Note: standardize variables to improve convergence of the MCMC procedure.

We fitted a logistic model with working status as a response variable and age (age), gender (female), health satisfaction (hsat), degree of handicap in percent (handper) , household nominal monthly income in German marks/1000 (hhninc), children under age 16 in the household (hhkids), years of schooling (educ), marital status (married), number of doctors vists in the last three months (docvis), the number of hospital visits in the last calendar year (hospvis), and insurance status (public) as covariates. The continuous covariates were standardized and we let the binary covariates take values of -0.5 and 0.5 instead of 0 and 1. The prior distributions are then symmetric in the sense that the logits for each group have the same prior variability as well as the same prior means, yet the beta's still have the usual interpretation of a conditional log odds ratio [1]. The model was fitted in Openbugs with 3 MCMC chains of 10000 iterations of which 5000 were discarded as burnin.

Model M_1:

$$logit[P(Working status = 1)] = \beta_1 + \beta_2 * age + \beta_3 * female + \beta_4 * hsat$$
$$+ \beta_5 * handper + \beta_6 * hhninc + \beta_7 * hhkids + \beta_8 * educ + \beta_9 * married + \beta_{10} * docvis$$
$$+ \beta_{11} * hospvis + \beta_{12} * public$$

where $\beta_i \sim \mathcal{N}(0, \sigma_i^2)$ with $\sigma_i^2 = 10000$(vague priors)
The history plots (Figure 8) and auto-correlation plots (Figure 9) of this model indicate that the model parameters converged well and rapidly. In the following models, the covariate with the highest p-value, according to a logistic regression in SAS, was excluded from the model and the model was fitted again. Table 5 gives an overview of the covariates that were included in each model. In Table 6 the goodness of fit measures of the different models are shown.
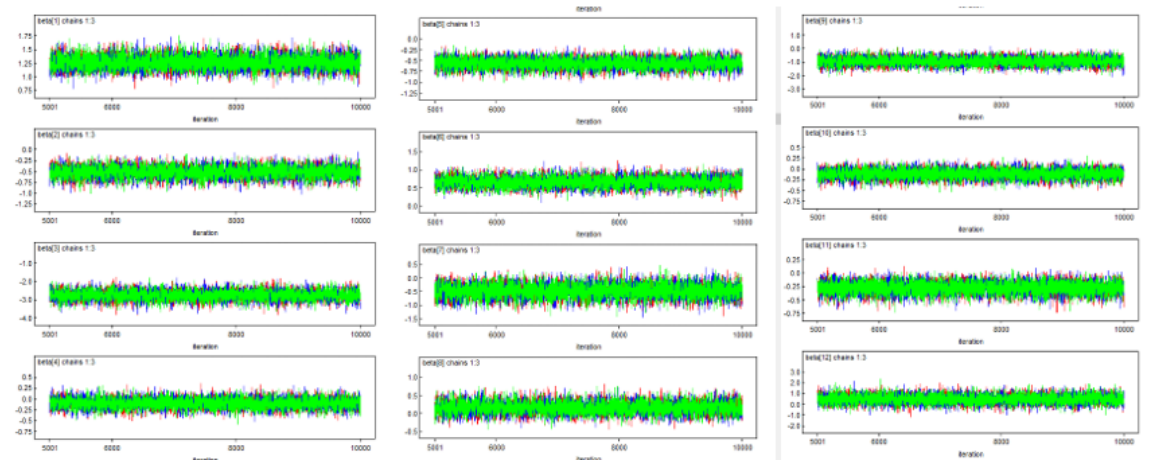

Figure 8: History plots Model 1

The result in Table 6 show that Model 5 had the best fit. The different parameters in this model showed good convergence with history and auto-correlation plots similar to Figure 8 and Figure 9. This is also the model that includes all the covariates that had a significant
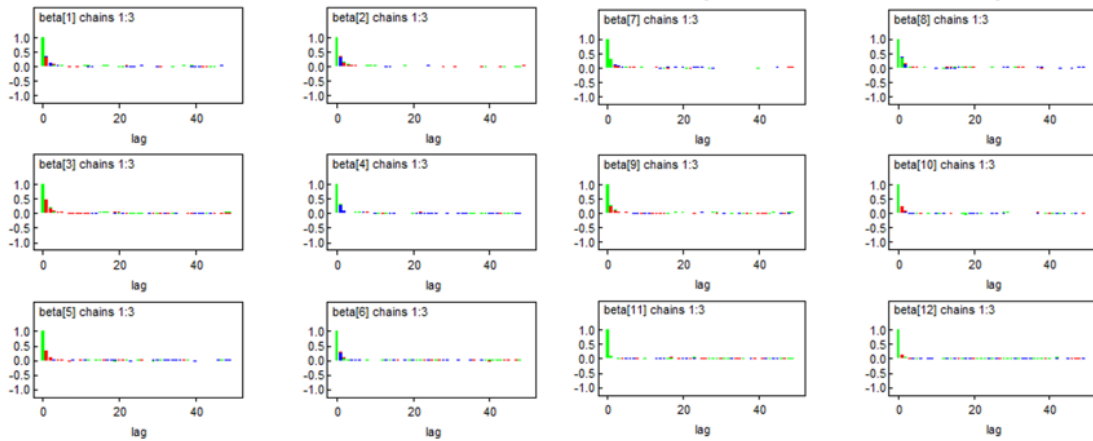
Figure 9: Auto-correlation plots Model 1

Table 5: Covariates in each model

Models	Covariates in the model
M_1	age, female, hsat, handper, hhinc, hhkids, educ, married, docvis, hospvis, public
M_2	age, female, handper, hhinc, hhkids, educ, married, docvis, hospvis, public
M_3	age, female, handper, hhinc, hhkids, married, docvis, hospvis, public
M_4	age, female, handper, hhinc, hhkids, married, hospvis, public
M_5	age, female, handper, hhinc, hhkids, married, hospvis
M_6	age, female, handper, hhinc, married, hospvis

Table 6: Comparing the fit of different models

Model	Dbar	Dhat	p_D	DIC
M_1	624.727	612.685	12.043	636.770
M_2	624.402	613.384	11.018	635.420
M_3	623.950	613.926	10.023	633.973
M_4	623.748	614.759	8.990	632.738
M_5	623.775	615.742	8.033	631.809
M_6	627.458	620.480	6.978	634.436

effect on the working status in the logistic regression in SAS. The posterior summary measures for Model 5 are shown in Table 7. These results indicate that being older, begin a female, being more disabled, having children under 16 at home, being married and the number of hospital visits in a year all had a negative influence on the probability to work. Only the monthly household had a positive effect and the probability to work.

Table 7: Posterior summary measures Model 5

Effect	mean	sd	MC error	val 2.5%	median	val 97.5%
Intercept	1.238	0.1207	0.001402	1.008	1.237	1.478
Age	-0.5466	0.1237	0.001405	-0.7913	-0.5466	-0.3065
Female	-2.791	0.253	0.003188	-3.297	-2.787	-2.312
Handper	-0.5537	0.1109	0.00128	-0.774	-0.5518	-0.3392
Hhninc	0.6661	0.1381	0.001476	0.4023	0.6644	0.9489
Hhkids	-0.5249	0.2387	0.00295	-0.9969	-0.5258	-0.06205
Married	-0.9669	0.2943	0.003025	-1.55	-0.9636	-0.401
Hospvis	-0.2869	0.1061	0.000913	-0.4977	-0.2863	-0.0824

Question 2

Fit a logistic mixed effects model with a random intercept and/or slope to the longitudinal profiles. Assume normality for the random effects. Take vague priors for all model parameters.

A logistic mixed effect model with random intercepts was compared to a logistic mixed effect model with random intercepts and random slopes. The covariates that were included in the final model of question 1 were also included in this model. To assess the evolution over time we subtracted 1984 from the response on the variable 'year', resulting in a time variable with the following values for each participant: 0, 1, 2, 3, 4, 7, 10. The interactions of time with age and time with household income were also included in the model because these interactions appeared to be significant in a GEE in SAS. Normality was assumed for the random effects ($b[0,1] \sim dnorm(0, tau[0,1])$) and we took vague priors for all model parameters ($beta[1:11] \sim dnorm(0.0, 0.0001)$). The models were fitted in OpenBUGS. The continuous covariates were standardized and we let the binary covariates take values of -0.5 and 0.5 instead of 0 and 1. To get good convergence we used 3 MCMC chains of 100 000 iterations, with a burnin of 50000, thinning of 4 and over-relaxation. The posterior summary measures were thus based on 37500 samples. These are the models that we fitted:

Model with random intercepts:

$$logit[P(Working status = 1)] = \beta_1 + \beta_2 * age + \beta_3 * female + + \beta_4 * handper + \beta_5 * hhninc$$
$$+ \beta_6 * hhkids + \beta_7 * married + \beta_8 * hospvis + \beta_8 * time + \beta_8 * time * age + \beta_8 * time * hhinc + b_0$$

where $\beta_i \sim \mathcal{N}(0, \sigma_i^2)$; $b_0 \sim \mathcal{N}(0, 1)$

Model with random intercepts and random slopes:

$$logit[P(Working status = 1)] = \beta_1 + \beta_2 * age + \beta_3 * female + + \beta_4 * handper + \beta_5 * hhninc$$
$$+ \beta_6 * hhkids + \beta_7 * married + \beta_8 * hospvis + \beta_8 * time + \beta_8 * time * age + \beta_8 * time * hhinc + b_0 + b_1 * time$$

where $\beta_i \sim \mathcal{N}(0, \sigma_i^2)$; $b_0 \sim \mathcal{N}(0, 1)$; $b_1 \sim \mathcal{N}(0, 1)$

Table 8 gives an overview of the fit of these two models. The model with random intercepts and random slopes gave a better fit to the data, so this model was also used in the following questions. The posterior summary results for this model are presented in Table 9. Because this a model has random intercepts and random slopes the regression coefficients in this model are subject specific.

Table 8: Comparing the fit of models with and without random slopes

Model	Dbar	Dhat	DIC	p_D
Random intercepts	298.4	234.8	362.1	63.64
Random intercepts and random slopes	244.6	165.4	323.8	79.17

Table 9: Posterior summary measures model with random intercepts and random slopes

Effect	mean	sd	MC error	2.5%	median	97.5%
Intercept	3.902	1.038	0.0184	2.08	3.826	6.171
Age	-0.556	0.791	0.0134	-2.231	-0.5214	0.9234
Female	-8.66	1.904	0.0475	-12.91	-8.48	-5.474
Handper	-0.818	0.405	0.0037	-1.653	-0.805	-0.057
Hhninc	2.888	0.669	0.0082	1.657	2.859	4.295
Hhkids	-1.684	0.862	0.0095	-3.482	-2.265	-0.037
Married	-2.31	1.22	0.0167	-4.837	-2.265	-0.0376
Hospvis	-0.174	0.244	0.0014	-0.667	-0.171	0.298
Time	0.261	0.148	0.0024	-0.008	0.2531	0.576
Time*Age	-0.450	0.108	0.0017	-0.684	-0.443	-0.261
TIme*Hhninc	-0.196	0.081	0.0007	-0.361	-0.195	-0.039
σ_{b_0}	5.099	1.032	0.0253	3.147	4.982	7.477
σ_{b_1}	0.462	0.139	0.0033	0.222	0.451	0.766

Question 3

Check the posterior predictive performance to evaluate the model.

If the model is adequate, new data sets generated form the model should look like the observed data [1]. We therefore added a statement to the model in OpenBUGS to predict to probability to work for existing subjects and compared this predictions to the observed data. We again used 3 MCMC chains of 100 000 iterations, with a burnin of 50000, thinning of 4 and over-relaxation. To determine the posterior predicted values, the medians of the MCMC chains were taken for each observation. This gave a value of 0 or 1 that could be compared to the observed values. Figure 10 shows the number of predictions that were in line with the observations in green and the number of predictions that were not in line with the observed values in red. Overall only 35 predictions were not in line with the observations which is exactly 5% of all cases. This indicates that the posterior predictive performance of the model was good.
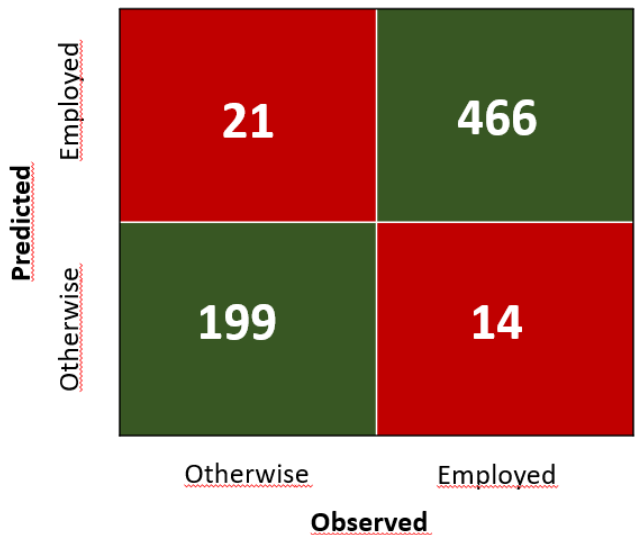

Figure 10: Observed versus predicted working status

Question 4

Check whether the logistic mixed effects model can be improved by changing the distribution of the random effects.

The logistic mixed effect model wit random intercepts and a random slopes was fitted assuming the random effects followed a student t-distribution instead of a normal distribution ($b[0, 1] \sim dt(0, tau[0, 1], 3)$). We again used 3 MCMC chains of 100 000 iterations, with a burnin of 50000, thinning of 4 and over-relaxation. Table 10 gives the goodness of fit values for the two models. The difference in DIC between the models is less than 5 which indicates that both models fit the data equally well so the fit of the logistic mixed effects model was not improved by changing the distribution of the random effects.

Table 10: Comparing the fit of models with a different distribution for the random effects

Model	Dbar	Dhat	DIC	p_D
Normal distribution for random effects: $b_i \sim \mathcal{N}(0, 1)$	244.6	165.4	323.8	79.17
Student t-distribution for random effects: $b_i \sim t_3$	245.2	163.9	326.4	81.25

Question 5

Perform a sensitivity analysis on the selected priors. Check what happens to the mixing of the chains and the posterior results.

To assess the influence of the prior on the results we compared models with different priors to each other. Table 11 gives an overview of the priors we used in these models. Model 1 is the reference model that was used in the previous questions. In models 2 and 3 we changed the uninformative priors for the β's to respectively diffuse and informative priors [1]. In model 4 we changed the priors for the variance of the random effects to inversed gamma instead

of uniform priors. The goodness of fit of these models can be found in Table 12. In Figure 11 and 12 the mixing of the chains for some variables are visualised in the history and auto correlation plots. Finally Table shows the posterior results for the different models. For all models we used 3 MCMC chains of 100 000 iterations, with a burnin of 50000, thinning of 4 and over-relaxation.

Table 11: Priors in each model

Models	Prior for β's	Prior for σ
Model 1	Non-informative (sd = 1000)	Uniform
Model 2	Diffuse (sd = 10)	Uniform
Model 3	Informative (sd = 1)	Uniform
Model 4	Non-informative (sd = 1000)	Inversed gamma

Table 12: Comparing the fit of models with different priors

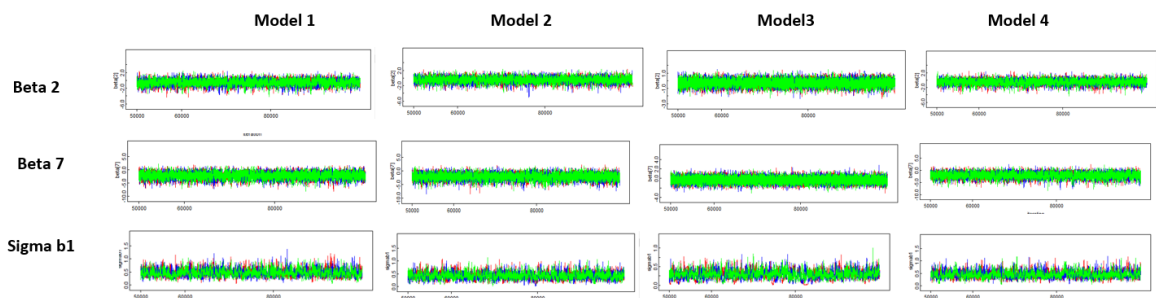| Model	Dbar	Dhat	DIC	p_D
Model 1	244.6	165.4	323.8	79.17
Model 2	247.1	167.6	326.5	79.46
Model 3	280.5	198.0	363.1	82.58
Model 4	250.4	171.2	329.7	73.26

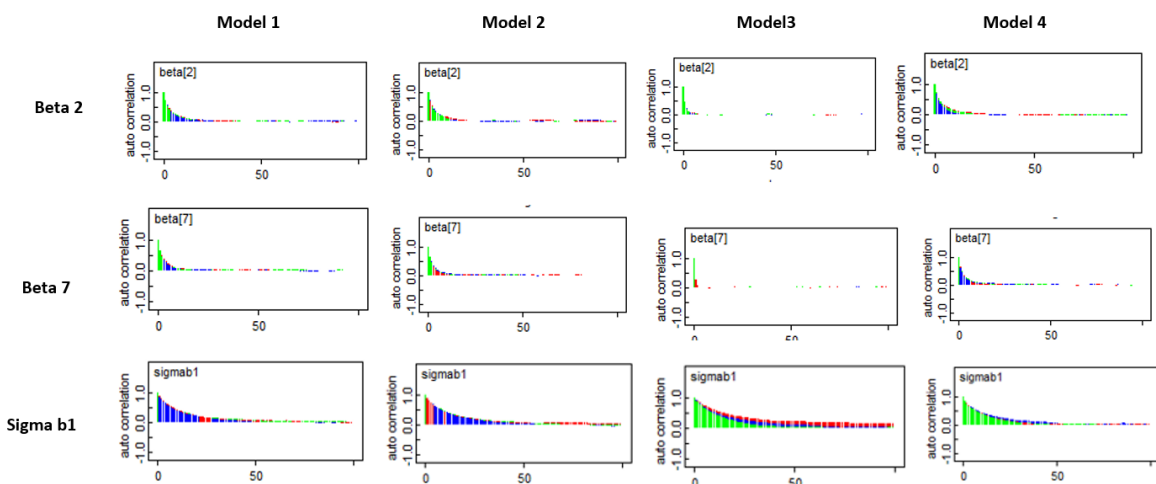Figure 11: History plots models with different priors

Figure 12: Auto-correlation plots models with different priors

As can be seen from Table 12 the fit of the models was more or less comparable for the models 1, 2 and 4. Only model 3 with the informative prior for the beta's had a significantly worse fit. This is also visible in Figures 11 and 12 since the mixing of the chains for sigmab1 is worse in model 3 compared to the other models. However the mixing of the chains for the beta's was better in model 3 compared to the other models. These difference between model 3 with the informative prior and the other models also resulted in different posterior measures for model 3 as can be seen in Table 13. When we for example look at the regression coefficients for the gender effect, we can see that this coefficient was quite different in model 3, compared to the other models. Changing the priors of a model can thus have quite some effect on the

16

mixing and the posterior results of a model, although this of course depends on the nature of the priors.

Table 13: Posterior summary measures models with different priors

Effect	Model 1 Mean (sd)	Model 2 Mean (sd)	Model 3 Mean (sd)	Model 4 Mean (sd)
Intercept	3.902 (1.038)	3.774 (0.983)	1.794 (0.500)	3.758 (1.011)
Age	-0.556 (0.791)	-0.492 (0.754)	-0.288 (0.45)	-0.504 (0.748)
Female	-8.666 (1.904)	-8.281 (1.784)	-3.32 (0.671)	-8.243 (1.857)
Handper	-0.819 (0.405)	-0.804 (0.399)	-0.504 (0.281)	-0.797 (0.388)
Hhninc	2.888 (0.669)	2.829 (0.647)	1.754 (0.423)	2.821 (0.649)
Hhkids	-1.684 (0.862)	-1.624 (0.829)	-0.985 (0.511)	-1.161 (0.816)
Married	-2.310 (1.220)	-2.201 (1.182)	-0.355 (0.637)	-2.174 (1.19)
Hospvis	-0.1745 (0.245)	-0.177 (0.240)	-0.186 (0.193)	-0.187 (0.239)
Time	0.262 (0.148)	0.253 (0.146)	0.228 (0.103)	0.239 (0.140)
Time*Age	-0.451 (0.108)	-0.441 (0.106)	-0.333 (0.073)	-0.429 (0.103)
TIme*Hhninc	-0.197 (0.081)	-0.191 (0.079)	-0.094 (0.057)	-0.193 (0.078)
σ_{b_0}	5.099 (1.032)	4.915 (0.971)	3.447 (0.6131)	4.7683 (0.9705)
σ_{b_1}	0.462 (0.139)	0.446 (0.137)	0.2789 (0.1118)	0.4064 (0.1357)

References

[1] Agresti A. Categorical data analysis. John Wiley & Sons; 2012.