Assignment 2

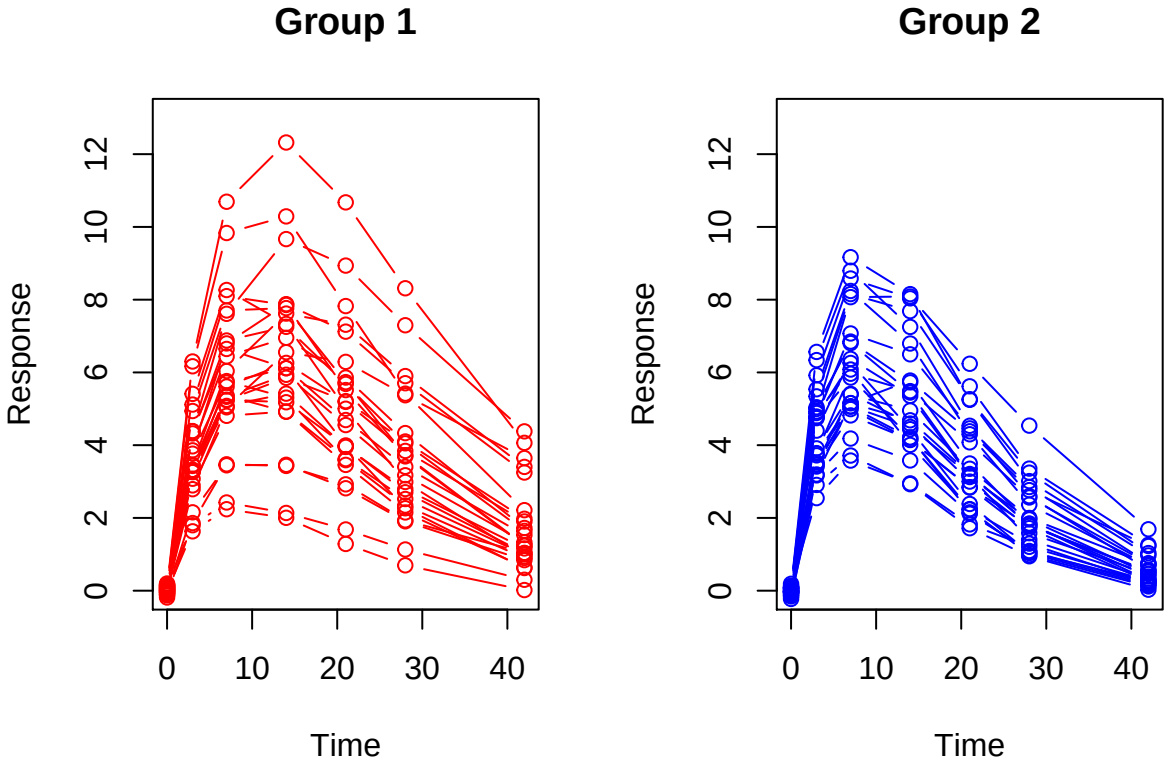
Tiffany Woodley

6/15/2020

In order to eliminate malaria and prevent the spread of drug resistance it is imperative to understand a populations infectivity. The main factor influencing a patient's infectivity is the density of gametocytes. The modelling process of gametocyte density using mixed effect models is outlined in *Nonlinear mixed effects modeling of gametocyte carriage in patients with uncomplicated malaria* (Distiller et al). This assignment aims to fit two alternate models using Bayesian analysis on a simulated dataset.

The log_2 gametocyte densities are modelled over a 42 day time period with measurements taken on days 0, 3, 7, 14, 21, 28 and 42 (denoted as t_j). By using the log_2 gametocyte densities it becomes easier to interpret as a change of one unit represents the density doubling or halving.

The simulated dataset can be seen below with 50 patient profiles taken from two distinct population groups.



Question 1

For the first part of the assigment individual patient profiles from group one were modeled.

For each of the profiles of group 1 (i = 1,. . . ,n = 25) the following model was fitted (for j = 1,. . . ,7)

$$y_{i,j} = \beta_0 + \sum_{k=1}^{L*} \phi_k b(|t_j - \zeta_k^{(1)}|) + e_{i,j}\ , \quad for \ \ j = 1, ..., 7$$

where

$$b(x) = xlog(x)$$

and

$$\zeta^{(1)} = [0.5, 10, 25]^T$$

It was assumed that $e_{i,j} \sim N(0, \sigma_e^2)$.

Let $\Phi = [\beta_0, \phi]^T$

The prior distributions were assumed to be

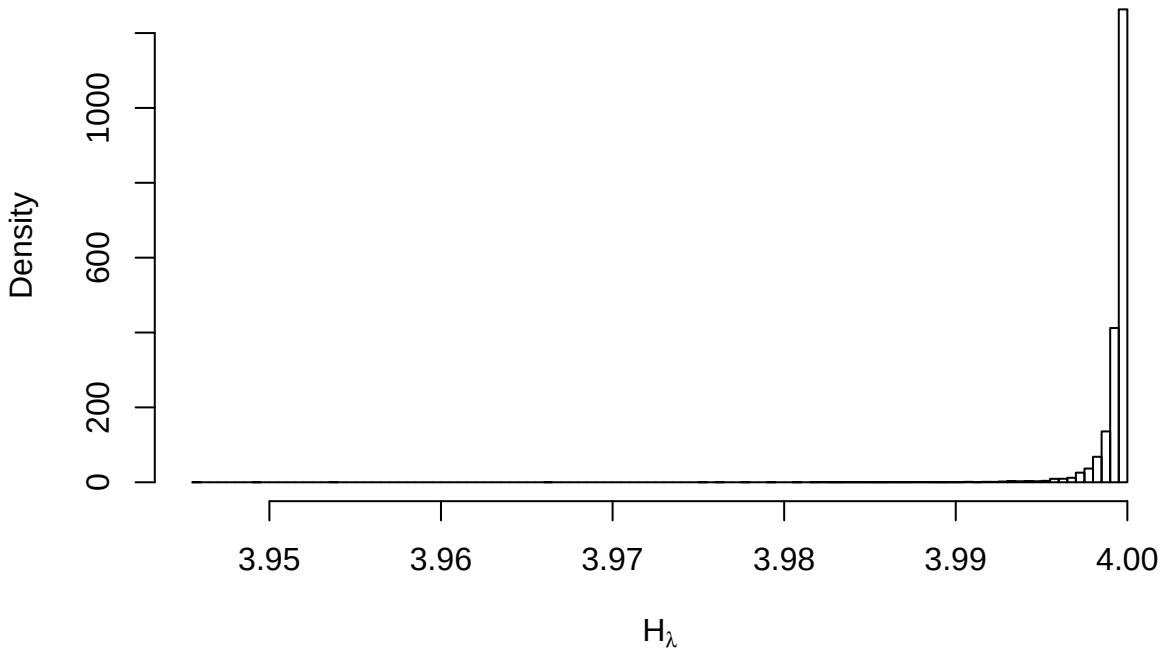
$$\Phi|\sigma_\phi^2 \sim (\sigma_\phi^2)^{-L*/2} exp(-\frac{1}{\sigma_\phi^2}\Phi^T D\Phi)$$

where

$$D = \begin{bmatrix} 0_{1\times 1} & 0_{1\times L*} \\ 0_{L*\times 1} & I_{L*} \end{bmatrix}$$

$$\sigma_e^2 \sim IG(i_1, i_2)$$
$$\sigma_\phi^2 \sim IG(i_3, i_4)$$

Since the varince of the ϕ coefficients come from an inverse gamma distribution and are not a fixed constant, they indicate that the ϕ coefficients are constrained. Due to these constraints, the prior coefficients $i_1, i_2, i_3 \& i_4$ were chosen based on the effective degrees of freedom(H_λ). Since we are fitting individual profiles and not trying to fit a model over all the profiles, the priors were chosen to create a better individual fit and so the constaints were kept low in the prior. With $\lambda = 0$ the maximum degrees of freedom the model could have is 4, therefore the distribution of the H_λ was chosen to have a greater mass closer to 4 indicating a small penalisation. This allows for the data to determine how constrained the coefficients should be. The chosen distribution has a strong prior belief that the effective degrees of freedom will be 4 and can be seen below, this was achieved with $i_1 = 2, i_2 = 5, i_3 = 2, i_4 = 100$.



The final joint posterior is given as

$$p(\beta, \sigma_e^2, \sigma_\phi^2|X) \propto p(\sigma_e^2)p(\sigma_\phi^2)p(\beta|\sigma_\phi^2)L(X|\beta, \sigma_e^2)$$

To find this posterior distribution we use gibbs sampling to sample from the marginal posterior distributions given below,

$$\beta|X, \sigma_e^2, \sigma_\phi^2 \sim N((D\frac{\sigma_e^2}{\sigma_\phi^2} + X^T X)^{-1}X^T y, \frac{1}{2\sigma_e^2}(D\frac{\sigma_e^2}{\sigma_\phi^2} + X^T X))$$

$$\sigma_e^2|\beta, X, y \sim IG(i_1 + \frac{n}{2}, \frac{1}{2}e^T e + i_2)$$

where

$$e = y - X\beta$$

$$\sigma_\phi^2|\beta \sim IG(i_3 + \frac{d}{2}, \frac{1}{2}\beta^T D\beta + i_4)$$

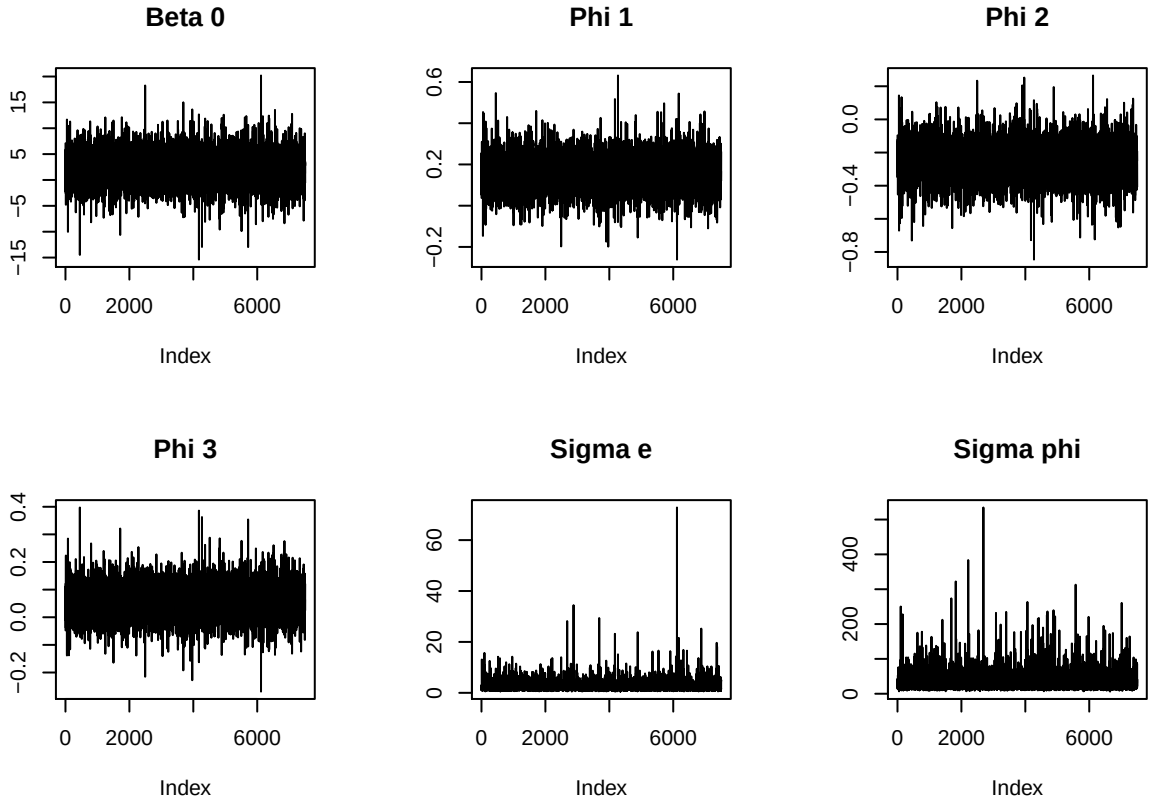
where

$$d = no.\ of\ columns\ in\ X$$

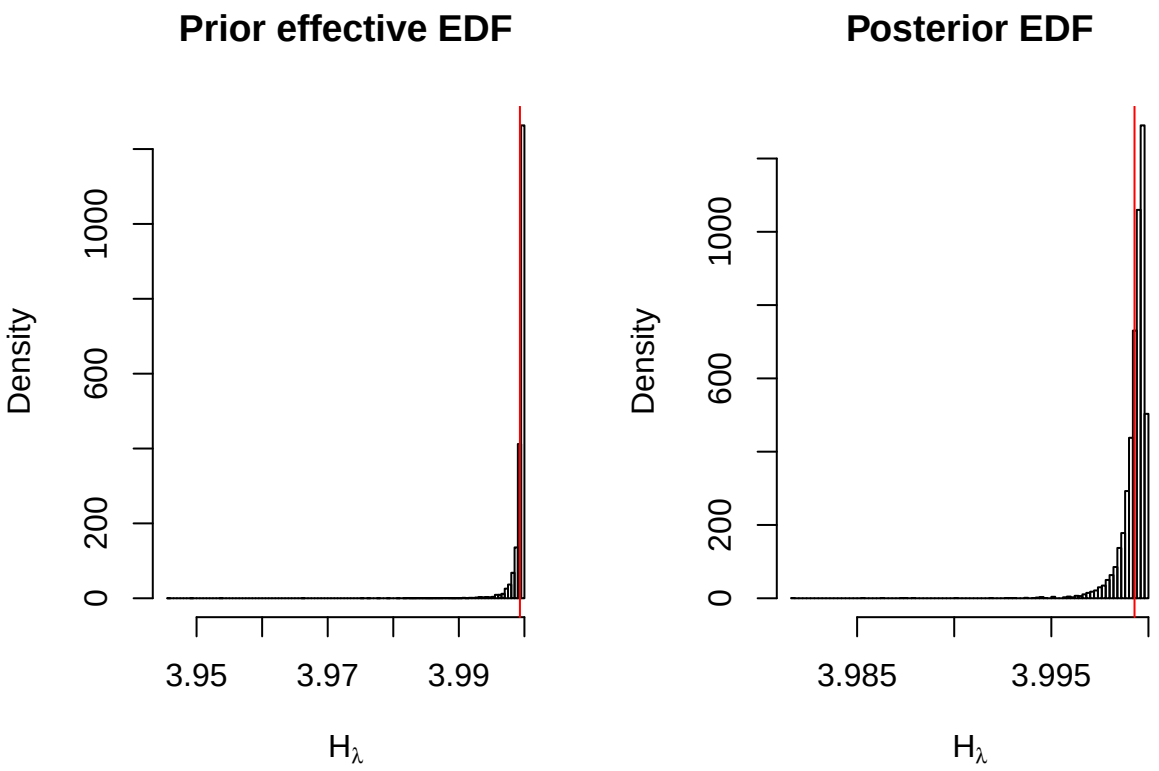
To visualise the process of fitting the model, the model was fit to patient 1.

The first half of the samples are discarded as the burn in sample so that all samples retained are from the posterior distribution. The retained samples are plotted below for all the coefficients and variance values,

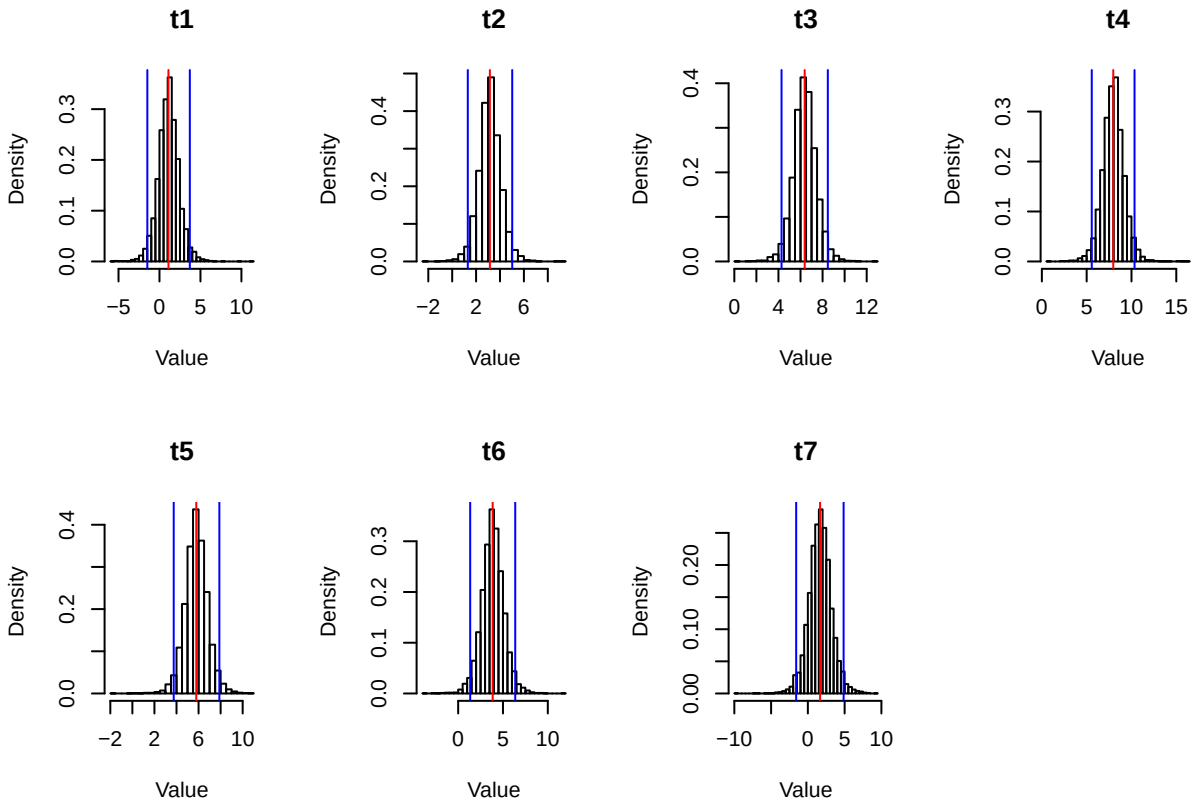
there is no inherent pattern in them, indicating the gibbs sampler had converged and we were sampling from the true posterior distrubution.



The posterior of H_λ for patient 1 is shown below, the mean is extremely similiar for both the prior and posterior, allowing the model to be closer to an unconstrained model with the effective degrees of freedom being close to 4.

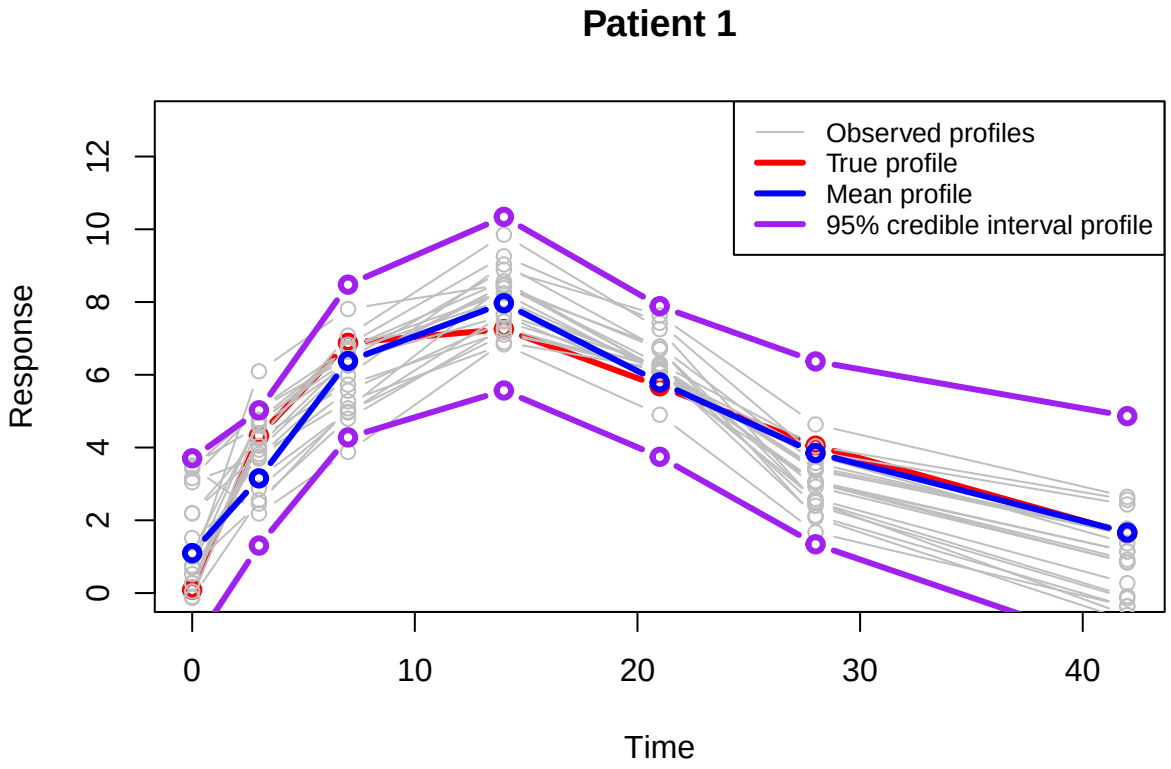


The posterior of the gametocyte densities at different times(days) for patient 1 can be seen below.

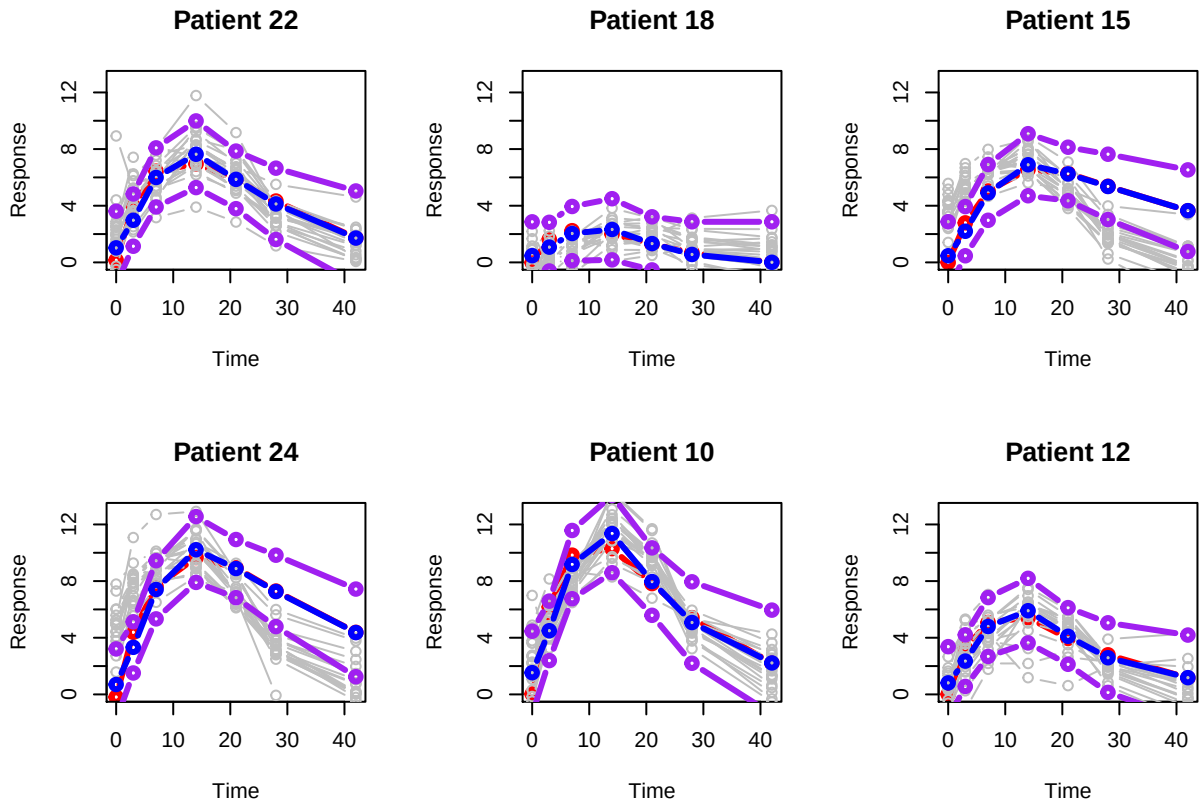


20 of the observed profiles are plotted below with the mean profile superimposed in blue, and the 95% credible interval for the fitted profiles shown in purple. The true profile, in red, lies within the 95% credible interval

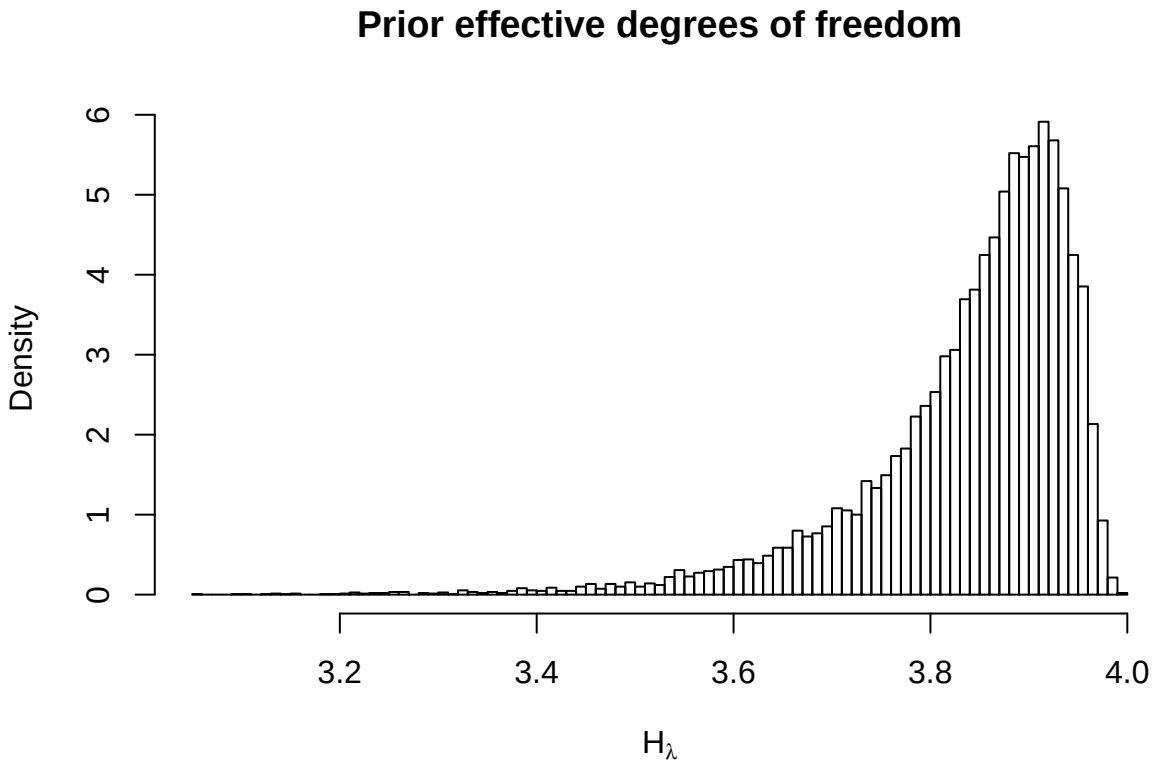
and is close to the mean profile and so the posterior distributions make sense given the patient's true data. Since this is such a good fit to the specific patient, chances at generalisation for other patients is not great. But since this model is modeling seperate patients and not taking into account inter-patient variation it makes sense to have a less constrained model.



The model is then fit to a random sample of patients and the fitted profiles can be seen below. The fitted profiles have large variation across patients but fit well to the individual patients.

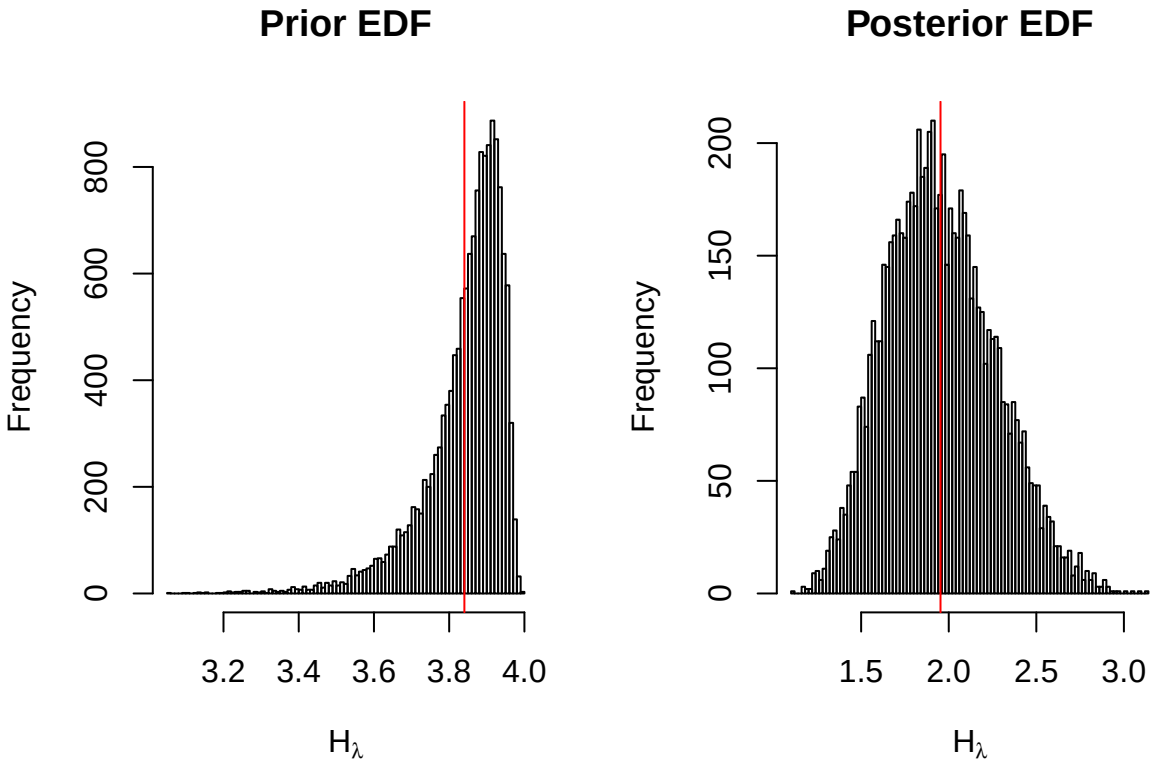


To check the validity of the results we test how sensitive the model is to the chosen prior values $i_1, i_2, i_3 \& i_4$. We change the prior belief to a more constrained model to see the effect it has on the posterior. The new prior has less of a strong belief and has a larger range of values in it's distribution.



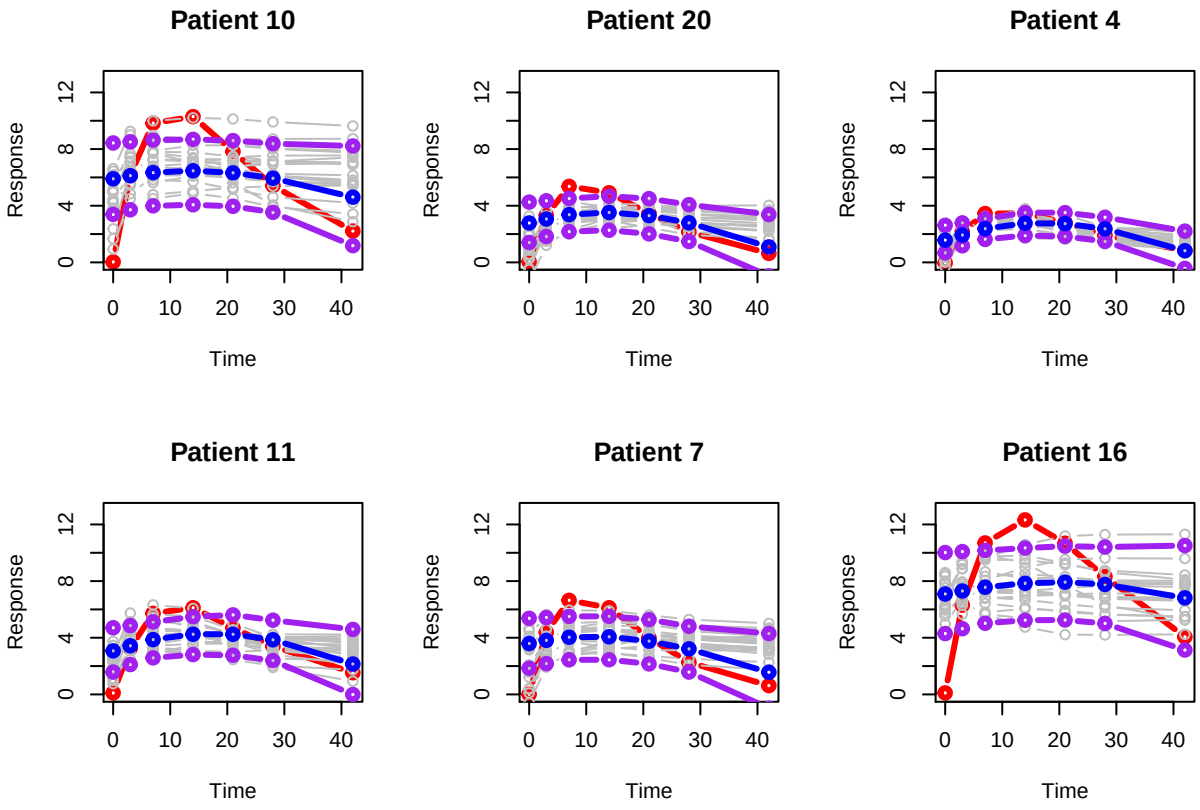
The posterior of H_λ for patient 1 is shown below, the mean has decreased substantially indicating a more

constrained model with close to two degrees of freedom, meaning the model only had the freedom to fit something close to a linear line.



The mean fits are close to linear and do not capture the true values well. This was expected given the posterior effective degrees of freedom. The true values do not all lie within the patient's credible intervals. The smoother patient profiles have better fitting observations as the more constrained model struggles to capture the curvier patient profiles. This shows that different profiles require different degrees of freedom(priors) in order to attain suitable observations.

It also shows the posterior is very sensitive to the prior and it is important to have a well informed prior to attain accurate results. It is also important to note if intra-patient variation is important in the model when choosing the prior, if it is the prior should be chosen to be more constrained to lead to better generalisation. Since this question is focused on individual patient fits, the first model with less constraints makes more sense as it gives a better fit to the individual patients and will be more accurate in extrapolating values for each patient between the measured time frames.



Question 2

For the second part of the assignment we model profiles from both groups allowing the ϕ coefficients to vary for individual patients but keeping the β coefficients constant among all patients.

All of the data (both group 1 and group 2 simultaneously) was used and the following model was fit (for j = 1,. . . ,7):

$$y_{i,j} = \beta_0 + \beta_1 group_i + \sum_{k=1}^{L} \phi_k b(|t_j - \zeta_{i,k}^{(1)}|) + e_{i,j} \ , \quad for \ \ j = 1, ..., 7$$

$$y_{i,j} = \beta_0 + \beta_1 group_i + \sum_{k=1}^{L} \phi_k b(|t_j - \zeta_{i,k}^{(2)}|) + e_{i,j} \ , \quad for \ \ j = 1, ..., 7$$

where $\zeta^{(1)} = [0.5, 10, 25]^T$ and $\zeta^{(2)} = [0.5, 5, 25]^T$. The variable $group_i$ is an indicator variable which is coded 1 for group 1 and 0 for group 2. Take note that β_0 and β_1 are the same for both groups although the spline regression coefficients are allowed to be different for each profile.

It was assumed that both σ_e^2 and σ_ϕ^2 are common for both groups (i.e. $\sigma_e^2 \sim IG(i1, i2)$ and $\sigma_\phi^2 \sim IG(i3, i4)$). Further it was assumed that each $e_{i,j}$ is independently and identically distributed (for all i, j).

In order to fit a model having constant β values with different ϕ values for the different patients a basis matrix was created where the columns for a paticular patient's ϕ values were only non-zero for the specific patient resulting in a 350×152 matrix.

For example,

$$X_{1,j} = [1, group_i, b(t_j - knot_1), b(t_j - knot_2), b(t_j - knot_3), 0, 0, 0, ..., 0]$$

A gibbs sampler is used to sample from the following marginal posterior distributions(see calculation in the appendix)

$$\beta|X, \sigma_e^2, \sigma_\phi^2 \sim N((D\frac{\sigma_e^2}{\sigma_\phi^2} + X^T X)^{-1} X^T y, \frac{1}{2\sigma_e^2}(D\frac{\sigma_e^2}{\sigma_\phi^2} + X^T X))$$

$$\sigma_e^2|\beta, X, y \sim IG(i_1 + \frac{n}{2}, \frac{1}{2}e^T e + i_2)$$

where

$$e = y - X\beta$$

$$\sigma_\phi^2|\beta \sim IG(i_3 + \frac{d}{2}, \frac{1}{2}\beta^T D\beta + i_4)$$

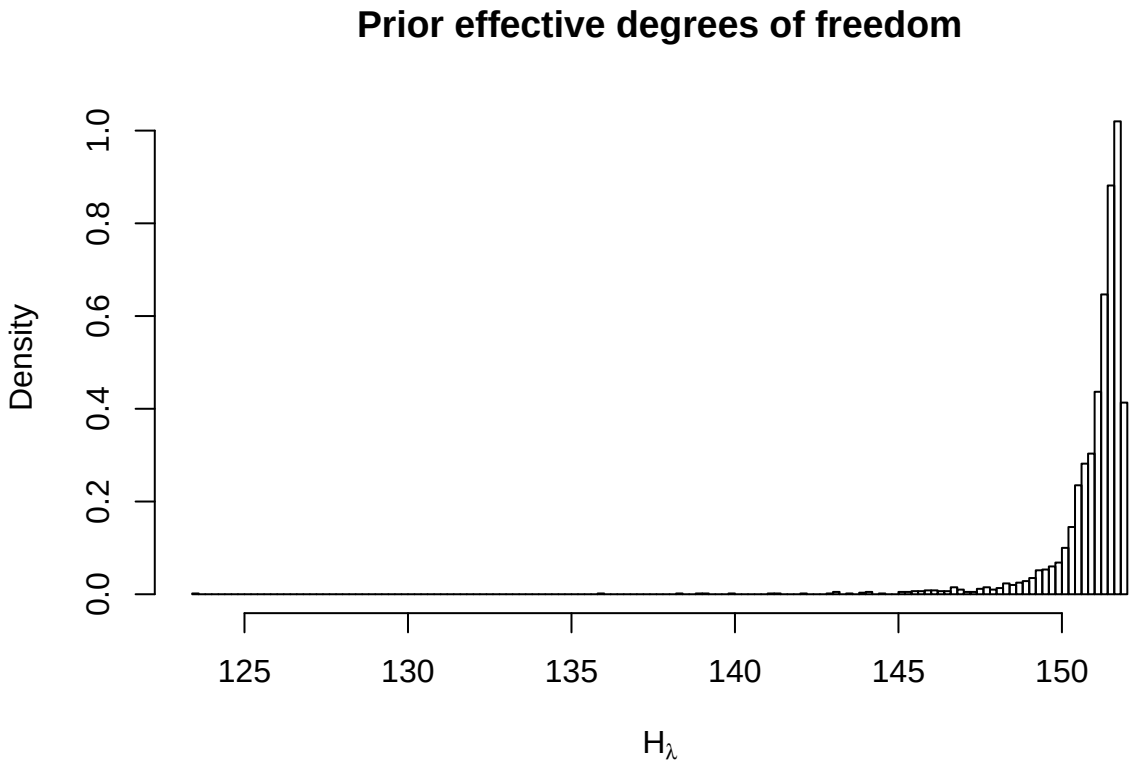
where

$$d = no.\ of\ columns\ in\ X$$

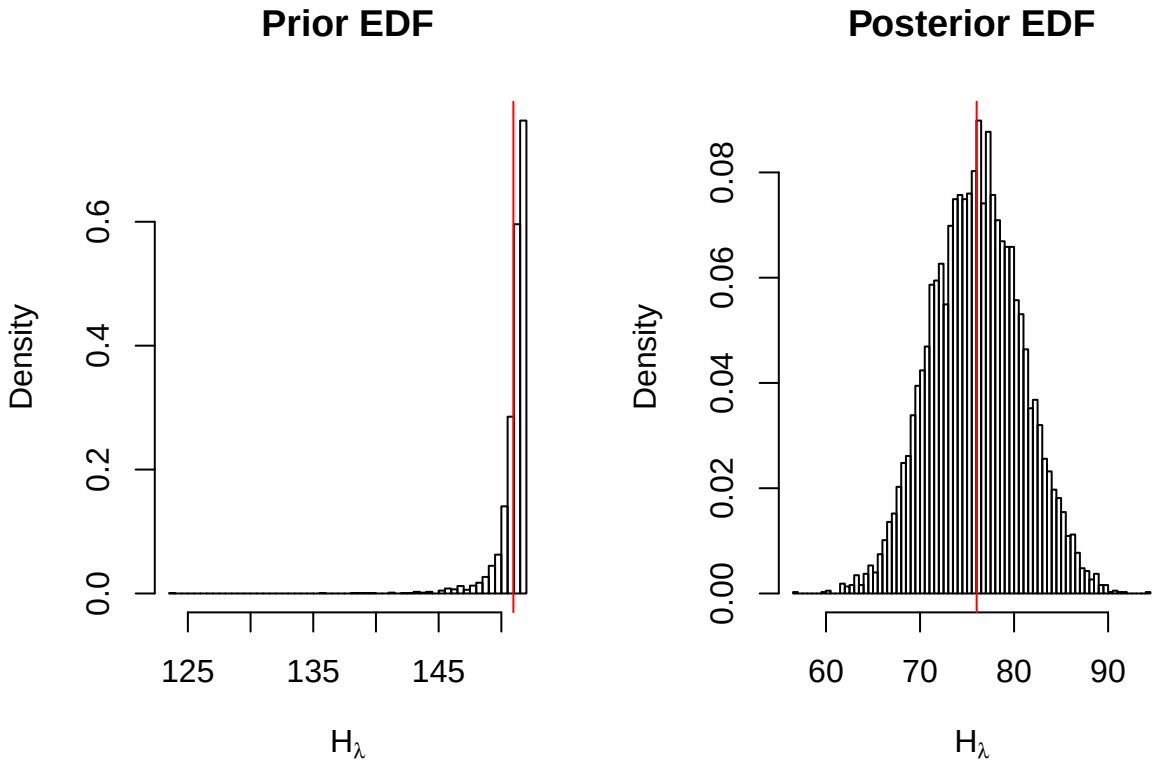
Since both β_0 and β_1 are unconstrained the D matrix becomes

$$D = \begin{bmatrix} 0_{2\times 2} & 0_{2\times L*} \\ 0_{L*\times 2} & I_{L*} \end{bmatrix}$$

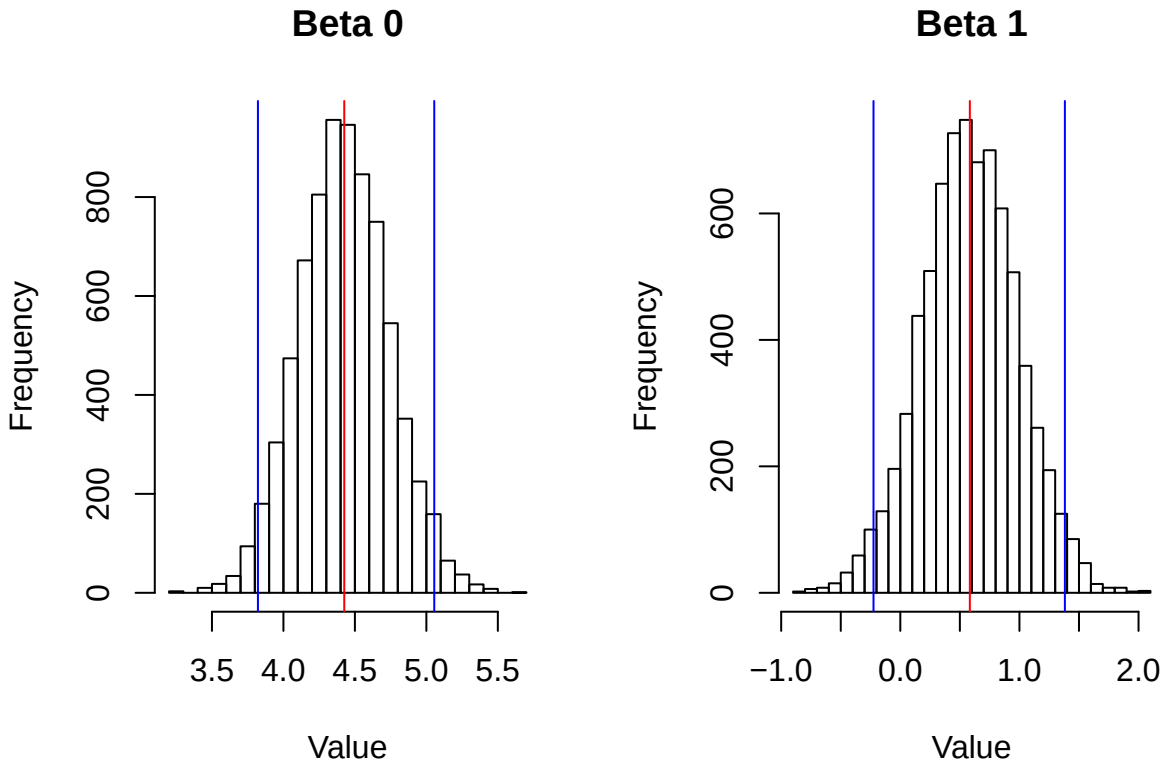
We start with a strong belief that the model is unconstrained, to allow for better individual fits.



The prior effective degrees of freedom distribution shows a much more constrained model with the mean of the effective degrees of freedom sitting at close to 75.

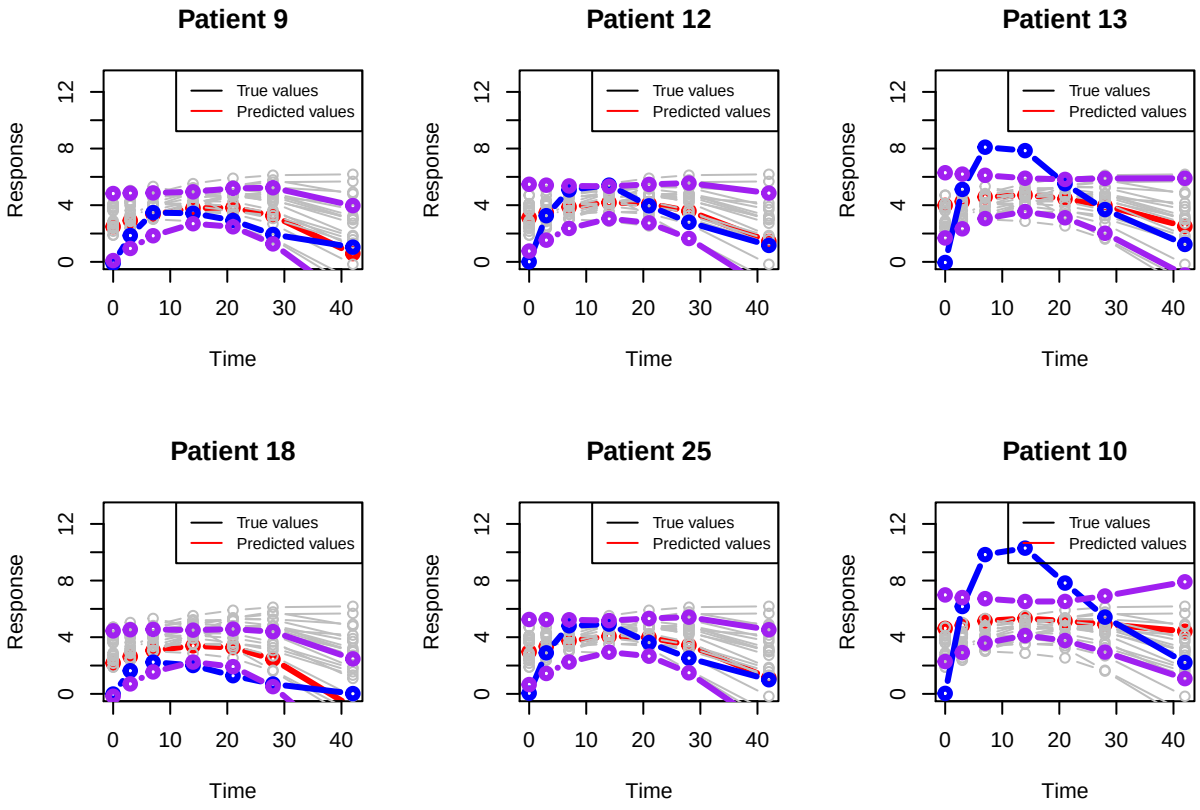


The posterior of the shared coefficients can be seen below with the posterior mean in red and the 95% credible interval between the two blue lines

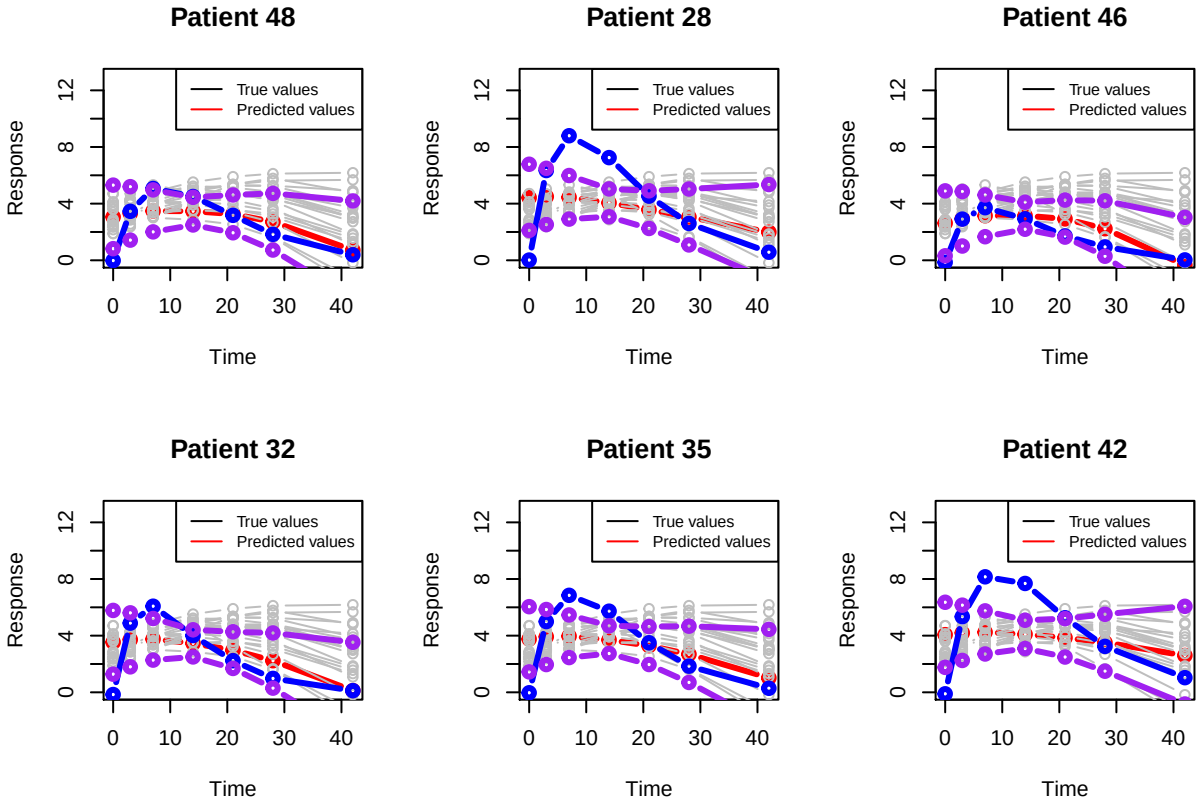


The fitted profiles for group 1 and 2 can be seen below, the observed profiles do not capture the true values well as the observed profiles are a lot flatter due to the constraints. The smoother patient profiles are captured better than the curvier patient profiles.

Group 1 fitted profiles

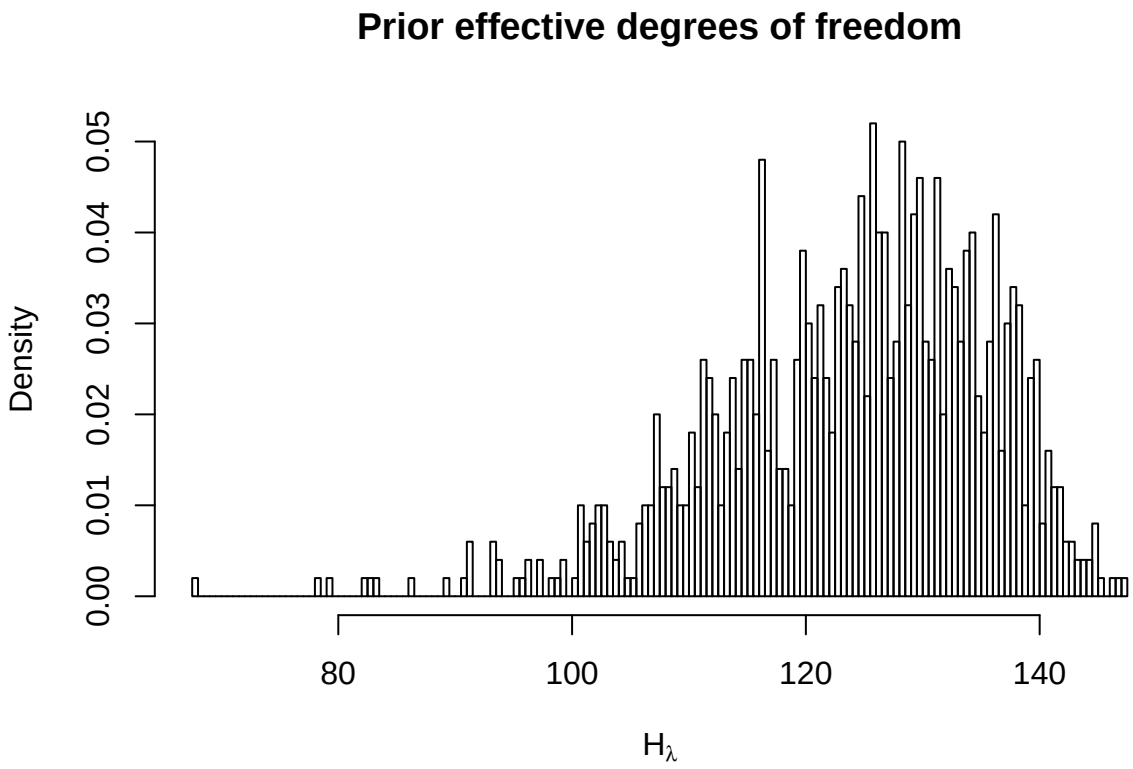


Group 2 fitted profiles

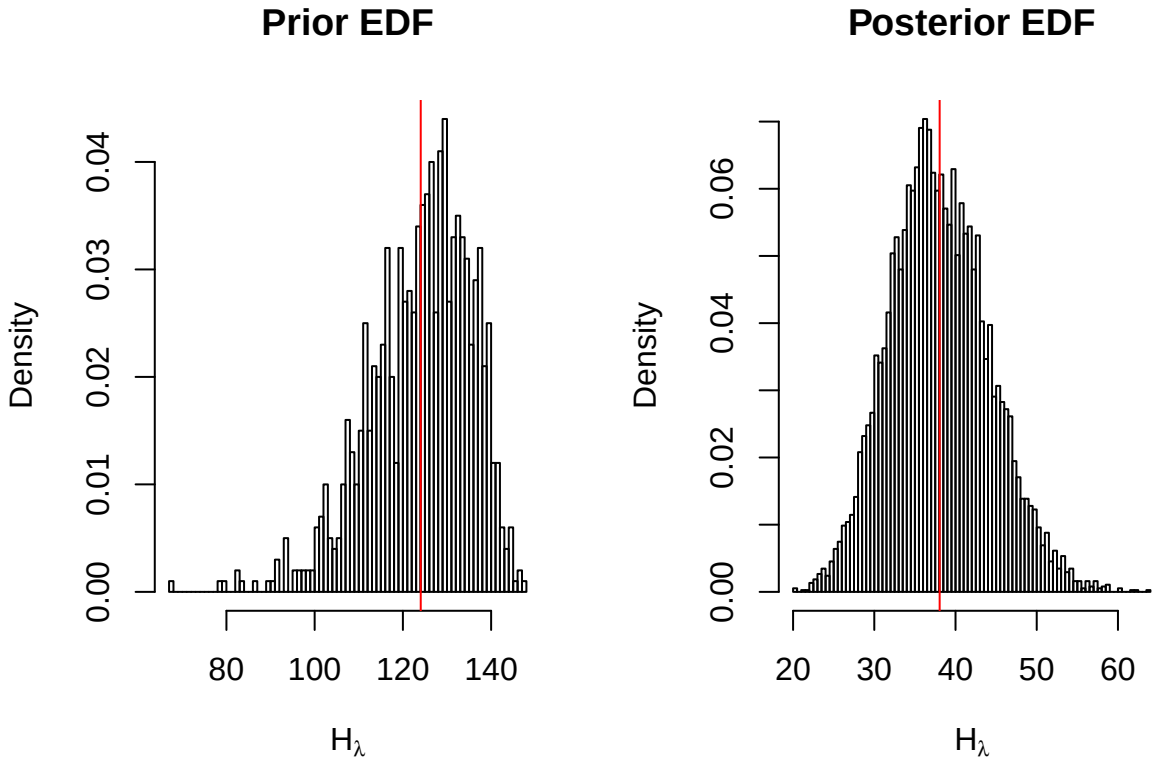


This model struggles to capture all the patient's fits, this is because all profiles are fit together, meaning the lambda constraint would have to be constant over all profiles. To account for different patients fits (smoother and curvier profiles) it would have to over constrain some and under constrain others to find a balance between all the patients.

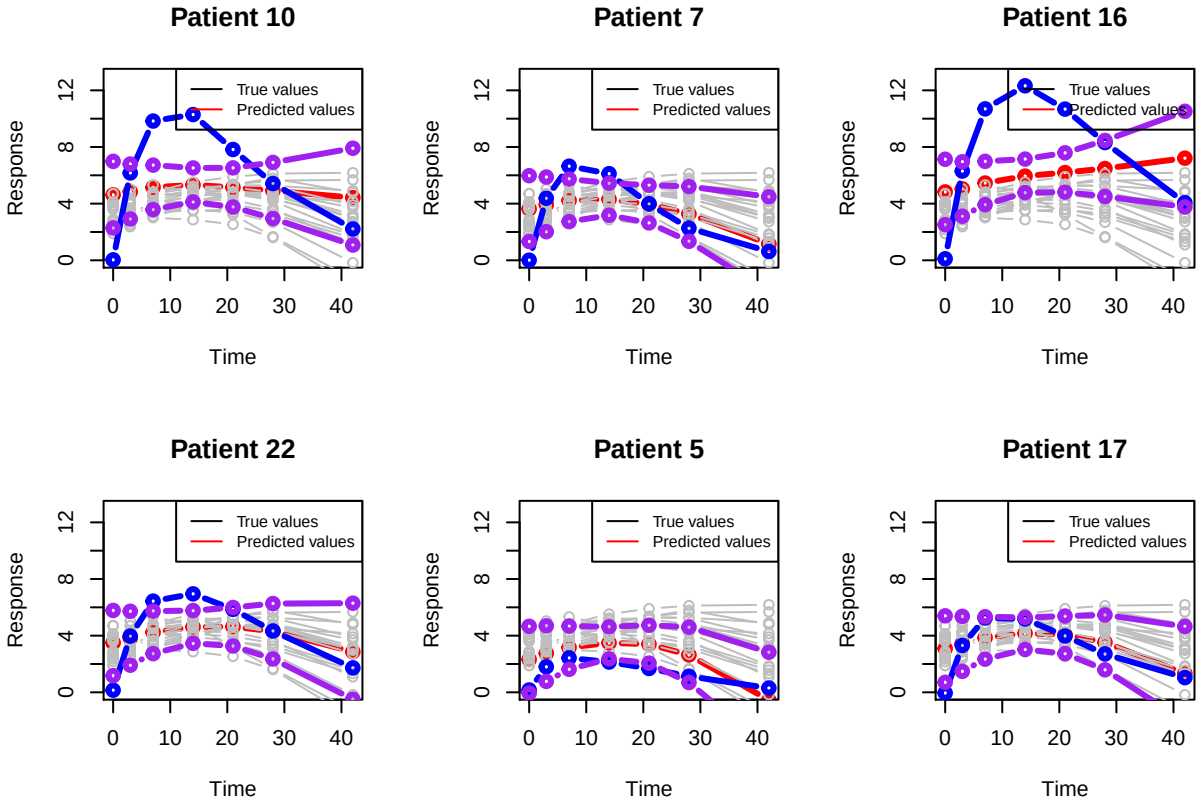
To check the validity of the results we test how sensitive the model is to the chosen prior values $i_1, i_2, i_3 \& i_4$. We change the prior belief to a more constrained model to see the effect it has on the posterior.



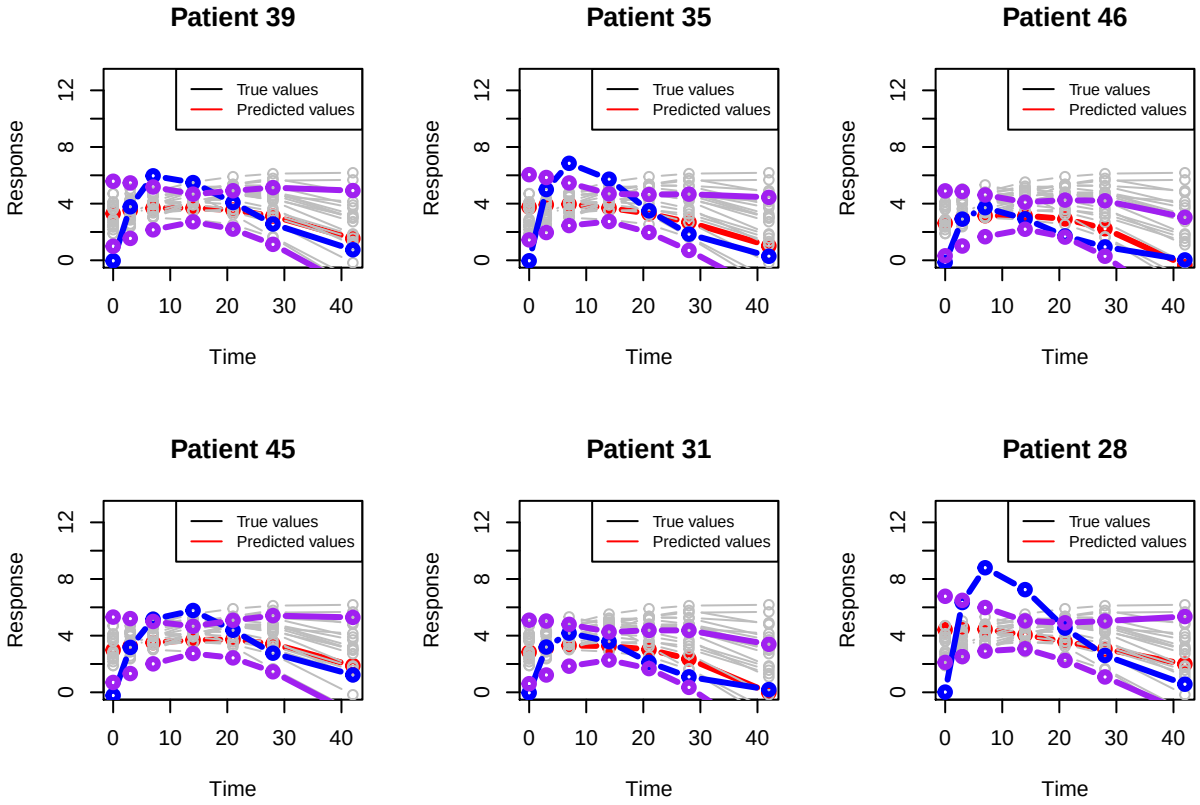
This prior results in an posterior with a lower mean effective degrees of freedom, sitting at around 38.



Group 1 fitted profiles



Group 2 fitted profiles



The observed profiles still capture the true values poorly. The model needs to be unconstrained in order to better fit the individual patient profiles

References

Distiller, G.B., Little, F. and Barnes, K.I., 2010. Nonlinear mixed effects modeling of gametocyte carriage in patients with uncomplicated malaria. Malaria Journal, 9(60).

Appendix

Consider the regression model

$$y_{i,j} = \beta_0 + \beta_1 group_i + \sum_{j=1}^{p} \phi_{i,j}(x_i - \zeta_j) + e_i$$

where $e_i \sim N(0, \sigma_e^2), \zeta_1, ..., \zeta_p$ are knots defined on the range of x. Parameter estimation is undertaken by constraining the coefficients associated with the $\phi_j (j = 1, ..., p)$ regression coefficients as

$$\sum_{j=1}^{p} \phi_j^2 \leq c$$

for some c.

Let X be a fixed matrix of basis functions where a row resembles a matrix similar to $[1, group_i, (x-\zeta_1), .., (x-\zeta_p), 0, ..., 0]$

with $\Phi = [\beta_0, \beta_1, \phi_1, .., \phi_p]$

To find Φ we need to minimize

$$L = (y - \beta_0 + X_1\beta_1 + X_{2:p}\phi)^T(y - \beta_0 + X_1\beta_1 + X_{2:p}\phi) + 2\lambda(\phi^2)$$

This is similar to solving a ridge regression problem without the constraint on β_0 or β_1.

$$\beta_{ridge} = (X^TX + \lambda I_p)^{-1}X^Ty$$

This can be done by setting

$$\lambda I_p = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix}$$

Assuming the prior distributions for σ_e^2 , σ_ϕ^2 and Φ are:

$$p(\sigma_e^2) \sim IG(i_1, i_2)$$

$$p(\sigma_\phi^2) \sim IG(i_3, i_4)$$

$$p(\Phi|\sigma_\phi^2) \sim N(0, Q^2)$$

where

$$QI_p = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\phi & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \sigma_\phi \end{bmatrix}$$

where $i_1, i_2, i_3 \& i_4$ are the prior distribution parameters

The joint posterior distribution of parameters is then:

$$p(parameters|data) \propto prior(parameters)likelihood(data|parameters)$$

$$p(\Phi, \sigma_\phi^2, \sigma_e^2|data) \propto p(\sigma_e^2)p(\sigma_\phi^2)p(\Phi|\sigma_\phi^2)L(x|\Phi, \sigma_e^2)$$

A gibbs sampler can be set up to sample from

$$\sigma_e^2|\sigma_\phi^2, \Phi, x$$

$$\sigma_\phi^2|\sigma_e^2, \Phi, x$$

$$\Phi|\sigma_e^2, \sigma_{phi}^2, x$$

To calculate the marginal posterior of Φ:

$$[\Phi|\sigma_e^2, \sigma_\phi^2, x] \propto exp(-\frac{\Phi^T D\Phi}{2\sigma_\phi^2})exp(-\frac{(y-X\Phi)^T(y-X\Phi)}{2\sigma_e^2})$$

where D resembles

$$D = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

To solve for the exponent of the marginal posterior

$$L = \frac{\Phi^T D\Phi}{\sigma_\phi^2} + \frac{(y-X\Phi)^T(y-X\Phi)}{\sigma_e^2}$$

$$L = \frac{1}{\sigma_\phi^2}\Phi^T D\Phi + \frac{1}{\sigma_e^2}(y^T y - 2y^T X\Phi + \Phi^T X^T X\Phi)$$

We can remove any terms not a function of Φ

$$L = \Phi^T(\frac{D}{\sigma_\phi^2} + \frac{1}{\sigma^2}X^T X)\Phi - \frac{1}{\sigma_e^2}2y^T X\Phi$$

By completing the square we get

$$L = \frac{(\Phi - (\frac{D}{\sigma_\phi^2} + \frac{1}{\sigma_e^2}X^T X)^{-1}\frac{1}{\sigma_e^2}X^T y)^T(\Phi - (\frac{D}{\sigma_\phi^2} + \frac{1}{\sigma_e^2}X^T X)^{-1}\frac{1}{\sigma_e^2}X^T y)}{(\frac{D}{\sigma_\phi^2} + \frac{1}{\sigma_e^2}X^T X)} + ...$$

$$\Phi|\sigma_e^2, \sigma_\phi^2, x \propto exp(-L)$$

$$\Phi|\sigma_e^2, \sigma_\phi^2, x \sim N((\frac{D}{\sigma_\phi^2} + \frac{1}{\sigma_e^2}X^T X)^{-1}\frac{1}{\sigma_e^2}X^T y, \frac{1}{2}(\frac{D}{\sigma_\phi^2} + \frac{1}{\sigma_e^2}X^T X))$$

$$\Phi|\sigma_e^2, \sigma_\phi^2, x \sim N(\sigma_e^2(D\frac{\sigma_e^2}{\sigma_\phi^2} + X^T X)^{-1}\frac{1}{\sigma_e^2}X^T y, \frac{1}{2\sigma_e^2}(D\frac{\sigma_e^2}{\sigma_\phi^2} + X^T X))$$

$$\Phi|\sigma_e^2, \sigma_\phi^2, x \sim N(\sigma_e^2(D\frac{\sigma_e^2}{\sigma_\phi^2} + X^T X)^{-1}\frac{1}{\sigma_e^2}X^T y, \frac{1}{2\sigma_e^2}(D\frac{\sigma_e^2}{\sigma_\phi^2} + X^T X))$$

$$\Phi|\sigma_e^2, \sigma_\phi^2, x \sim N((D\frac{\sigma_e^2}{\sigma_\phi^2} + X^T X)^{-1} X^T y, \frac{1}{2\sigma_e^2}(D\frac{\sigma_e^2}{\sigma_\phi^2} + X^T X))$$

$$\Phi|\sigma_e^2, \sigma_\phi^2, x \sim N((\lambda I_p + X^T X)^{-1} X^T y, \frac{1}{2\sigma_e^2}(\lambda I_p + X^T X))$$

where $\lambda I_p = D\frac{\sigma^2}{\tau^2}$

To calculate the marginal posterior of σ_e^2:

$$\sigma_e|\Phi, \sigma_\phi^2, x \propto (\sigma_e^2)^{-(i_1+1)} e^{-\frac{i_2}{\sigma_e^2}} (\sigma_e^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma_e^2}(y-x\Phi)^T(y-x\Phi)}$$

$$\sigma_e|\Phi, \sigma_\phi^2, x \propto (\sigma_e^2)^{-(i_1+1)-\frac{n}{2}} e^{-\frac{i_2}{\sigma_e^2}-\frac{1}{2\sigma_e^2}(y-x\Phi)^T(y-x\Phi)}$$

$$\sigma_e|\Phi, \sigma_\phi^2, x \propto IG(i_1 + \frac{n}{2}, \frac{1}{2}(y-x\Phi)^T(y-x\Phi) + i_2)$$

where n is the number of rows in x

To calculate the marginal posterior of σ_e^2:

$$\sigma_\phi|\Phi, \sigma_e^2, x \propto (\sigma_\phi^2)^{-(i_3+1)} e^{-\frac{i_4}{\sigma_\phi^2}} (\sigma_\phi^2)^{-\frac{d}{2}} e^{-\frac{1}{2\sigma_\phi^2}(\Phi^T\Phi)}$$

$$\sigma_\phi|\Phi, \sigma_e^2, x \propto (\sigma_\phi^2)^{-(i_3+1)-\frac{d}{2}} e^{-\frac{i_4}{\sigma_\phi^2}-\frac{1}{2\sigma_\phi^2}\Phi^T\phi}$$

$$\sigma_\phi|\Phi, \sigma_e^2, x \propto IG(i_3 + \frac{d}{2}, \frac{1}{2}\Phi^T\Phi + i_4)$$

where d is the number of columns in x