UNIVERSITY OF EDINBURGH
SCHOOL OF MATHEMATICS
BAYESIAN DATA ANALYSIS

Assignment 1 — Solutions

1. (a) The conjugate prior to a Normal distribution is a Normal distribution (**1 mark**). The expert prior could be interpreted as a uniform distribution on $[0, 2]$, which has mean 1 and variance $1/3$. The Normal distribution with this mean and variance is $\mathrm{N}(1, 1/3)$ and so that is a good choice of prior. It is not the only choice. Anything of the form $N(1, k)$ with $k \sim 1$, e.g., $k = 0.5, 1, 2$ is OK since the expert opinion is vague. However a prior with $k \ll 1$ or $k \gg 1$ would not respect the expert opinion and a truncated distribution would not be conjugate (**1 mark for prior and 1 mark for justification**). The posterior for a Normal-Normal model with known measurement variance σ^2 and prior $N(\mu_0, \sigma_0^2)$ is

$$\mathrm{N}\left(\frac{n\bar{y}\sigma_0^2 + \mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}, \frac{\sigma^2\sigma_0^2}{n\sigma_0^2 + \sigma^2}\right).$$

This data has $n = 10$, $\sigma^2 = 30$ and $\bar{y} = 1.6116$ so for the $\mathrm{N}(1, 1/3)$ prior the posterior is $\mathrm{N}(1.06, 0.3)$ (**2 marks**).

(b) As in part (a) there are several ways to interpret the US expert's information. Following the procedure above the US expert prior can be interpreted as $\mathrm{N}(5, 4/3)$ (**1 mark**). A suitable mixture prior is of the form $p(\mu) = wp_1(\mu) + (1 - w)p_2(\mu)$ where $p_1(\mu)$ and $p_2(\mu)$ are the prior from the UK and US experts respectively and w is the weight for prior $p_1(\mu)$. A suitable choice is $w = 2/3$ since there are twice as many US experts (**1 mark**). In this case we have $p_1(\mu) = N(\mu_1, \sigma_1^2)$ and $p_2(\mu) = N(\mu_2, \sigma_2^2)$. The posterior can be found to be

$$w'\mathrm{N}\left(\frac{n\bar{y}\sigma_1^2 + \mu_1\sigma^2}{n\sigma_1^2 + \sigma^2}, \frac{\sigma^2\sigma_1^2}{n\sigma_1^2 + \sigma^2}\right) + (1 - w')\mathrm{N}\left(\frac{n\bar{y}\sigma_2^2 + \mu_2\sigma^2}{n\sigma_2^2 + \sigma^2}, \frac{\sigma^2\sigma_2^2}{n\sigma_2^2 + \sigma^2}\right), \tag{1}$$

where

$$w' = \frac{k_1 w}{k_1 w + k_2(1 - w)}, \qquad k_i = \frac{1}{\sqrt{\sigma^2 + n\sigma_i^2}} \exp\left[-\frac{1}{2}\left(\frac{n(\bar{y} - \mu_i)^2}{\sigma^2 + n\sigma_i^2}\right)\right] \tag{2}$$

(**2 marks**). In this case we find $w' = 0.890$ and the posterior is
$0.890N(1.06, 0.3) + 0.110N(3.95, 0.923)$ (**1 mark**).

(c) We need to choose a suitable prior on the precision $\tau = 1/\sigma^2$ and we use $\Gamma(0.01, 0.01)$ (**1 mark**). The posterior distribution obtained using JAGS is shown in Figure 1 (**4 marks**). The posterior mean, median, lower and upper quartiles for the mean μ are 1.255, 1.177, 0.8206 and 1.548 respectively. For the standard deviation σ these are 3.592, 3.425, 2.939 and 4.055 respectively. Note that you will not get exactly these values due to sampling error, but your values should be close to these (**3 marks**).

(d) The probability that $\mu < 1$ can be found by integrating the posterior for μ from $-\infty$ to 1. This can be done by including a line like

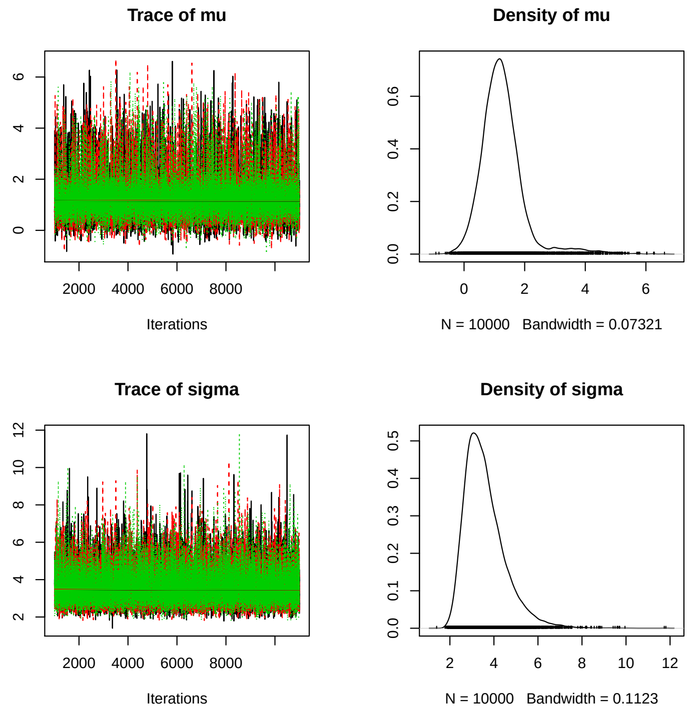

Figure 1: Posterior distributions (right) and trace plots (left) for the mean μ and standard deviation σ of the log concentration of the chemical.

```
frac<-step(1.0-mu)
```

in the JAGS model definition and looking at the posterior mean of the new variable "frac". We obtain an estimate $p = 0.366$ (**2 marks**).

To compute the probability that a single future measurement will yield a negative log-concentration, we first need to compute $p_{-,1}$, the posterior predictive probability of obtaining a negative measurement in a single future observation. This is accomplished by adding these lines to the JAGS model definition

```
ynew~dnorm(mu,tau)
prob<-step(0.0-ynew)
```

and looking at the posterior mean for "prob". This gives $p_{-,1} \approx 0.363$ (**2 marks**).

The probability that at least one of N future measurements yields a value less than 0 is one minus the probability that none of them yield a value less than 0 which can be calculated as $p_{-,5} = 1 - (1 - p_{-,1})^N$. For $N = 5$ and $p_{-,1} = 0.363$ we find $p_{-,5} \approx 0.895$ (**2 marks**).

(e) If we include w as a parameter with a flat prior in the range $[0, 1]$ the posterior on (μ, w) is given by Eq. (1) above, but with w' and $(1 - w')$ replaced by

$$w' \to \frac{2k_1 w}{k_1 + k_2}, \qquad 1 - w' \to \frac{2k_2(1-w)}{k_1 + k_2},$$

with k_i as defined in Eq. (2). In this case we find the joint posterior is

$$1.561 w p_G(\mu; 1.06, 0.3) + 0.439(1 - w) p_G(\mu; 3.95, 0.923),$$

where $p_G(x; \mu, \sigma^2)$ denotes the pdf of an $N(\mu, \sigma^2)$ distribution (**2 marks**).

The marginal distribution on μ is found by integrating over w

$$p(\mu|\mathbf{d}) = 0.781 p_G(\mu; 1.06, 0.3) + 0.219 p_G(\mu; 3.95, 0.923),$$

(**1 mark**).

The marginal distribution on w is found by integrating over μ

$$p(w|\mathbf{d}) = 0.439 + 1.122w,$$

(**1 mark**).

The marginalisation distribution on μ is the same distribution that would be obtained using equal weights on the two priors in the mixture, i.e., $w = 1/2$. This is because $w = 1/2$ is the prior expectation value for a $U[0,1]$ and the w prior is a hyperprior, i.e., the prior on a parameter that describes a prior on other parameters. The marginal on w is a straight line. It is rising, meaning that the mode of the posterior is $w = 1$, i.e., we favour the prior from the UK experts. We have weak evidence to suggest the UK experts are better at predicting than the US experts, but this is perhaps unsurprising given that the data is being collected in the UK. A straight line posterior does not indicate a strong constraint on the parameter. This is because the w parameter only enters once, as a prior on the mean that is common to all the subsequent observations. As we make more observations we expect to measure μ better and better, but there will be no strong change in our ability to measure w, since it only enters once. If we imagine a scenario in which we collect sets of data in multiple different sites, and we suppose the mean at each site is different, drawn from the prior described by w, then as we add more and more sites we would start to see a concentration in the w prior and stronger evidence that one set of experts is correct. (**2 marks**)

2. (a) The information available before O1 indicates that the rate is uncertain over orders of magnitude. Under these circumstances it is reasonable to suppose that the the log of the rate is uniform in some range. So, we represent the prior as

$$\log_{10}(\lambda) \sim U[-2, 3].$$

(**2 marks**). This prior has an expectation value of $999.99/\ln(10^5) = 86.858$ and variance of $999999.9999/2\ln(10^5) - 86.86^2 = 35885.13$ (**1 mark**). The conjugate distribution to a Poisson model is a Gamma distribution, $\Gamma(a, b)$, for which the mean and variance are a/b and a/b^2 respectively (**1 mark**). Matching the mean and variance we find $b = 1/413.15$ and $a = 0.210$ (**1 mark**). We use a conjugate distribution since we then know the posterior will also be in the conjugate family and so it is computationally convenient (**1 mark**). [**Note:** marks were awarded for reasonable prior choices with justification. It must be wide and flat over several decades and make some use of the prior information.)

(b) We note first that all of the observation runs are different lengths. The rate λ was quoted in units of yr^{-1}. Poisson processes are additive, i.e., if the rate in time period T is λ, the rate in time period kT is $\tilde{\lambda} = k\lambda$. If the prior on λ is $\Gamma(a, b)$ then we have

$$p(\tilde{\lambda}) = k^{-1}\frac{b^a}{\Gamma(a)}\left(\frac{\tilde{\lambda}}{k}\right)^{a-1} e^{-b\tilde{\lambda}/k} = \frac{(b/k)^a}{\Gamma(a)}\tilde{\lambda}^{a-1}e^{-(b/k)\tilde{\lambda}},$$

i.e., the prior on $k\lambda$ is $\Gamma(a, b/k)$ (**2 marks**). As three events are observed in O1, we can write down the posterior distribution on $\tilde{\lambda}$ as $\Gamma(a + 3, b/k + 1)$ and the posterior distribution on λ is $\Gamma(a + 3, b + k)$. In this case using the conjugate prior derived above we have $\Gamma(3.21, 0.252)$ (**2 marks**). The posterior mean and standard deviation are 12.717 and 7.098 respectively, a 95% symmetric confidence interval is $(2.817, 29.934)$ and the posterior distribution is shown in Figure 2 (**2 marks** Note: credit was also given for solutions obtained numerically using JAGS or similar).

(c) The probability that the rate exceeds 15 can be computed from the cumulative density function of the gamma distribution (**1 mark**). This is $\gamma(\alpha, \beta x)/\Gamma(\alpha)$, where $\gamma(\alpha, x)$ is the incomplete gamma function. In this case we need $1 - \gamma(3.21, 3.78)/\Gamma(3.21) = 0.312$ (**1 mark**).

(d) The Jeffrey's prior for the Poisson distribution is the improper prior $p(\lambda) \propto \lambda^{-1/2}$ (**1 mark**). This can be approximated by a $\Gamma(1/2, \beta)$ distribution with $\beta \to 0$. The posterior on λ from the O1 data with the Jeffrey's prior is therefore $\Gamma(3.5, 0.25)$ (**1 mark**). The posterior is shown as a red line in Figure 2, the posterior mean and standard deviation are 14 and 7.48 and a 95% symmetric confidence interval is given by $(3.38, 32.0)$ (**1 mark**). The probability that the rate exceeds 15 is now 0.379 (**1 mark**). So, the results change a little bit and in particular the Jefffrey's prior favours slightly higher rates than the conjugate prior, but there is not a very big difference between the two (**1 mark**).

(e) The second science run, O2, lasts 6 months, so the Poisson rate is 0.5λ. The posterior for O1 therefore gives a prior on the rate in O2 of $\Gamma(3.21, 0.504)$ using the conjugate prior (**1 mark**). For a posterior of the form $\Gamma(\alpha, \beta)$, the posterior predictive probability of seeing n events in O2 is therefore

$$p(n|d_{O1}) = \int_0^\infty \frac{\lambda^n e^{-\lambda}}{n!}\frac{\beta^\alpha\lambda^{\alpha-1}e^{-\beta\lambda}}{\Gamma(\alpha)}d\lambda = \frac{1}{n!}\frac{\beta^\alpha}{(\beta+1)^{\alpha+n}}\frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} \qquad (3)$$

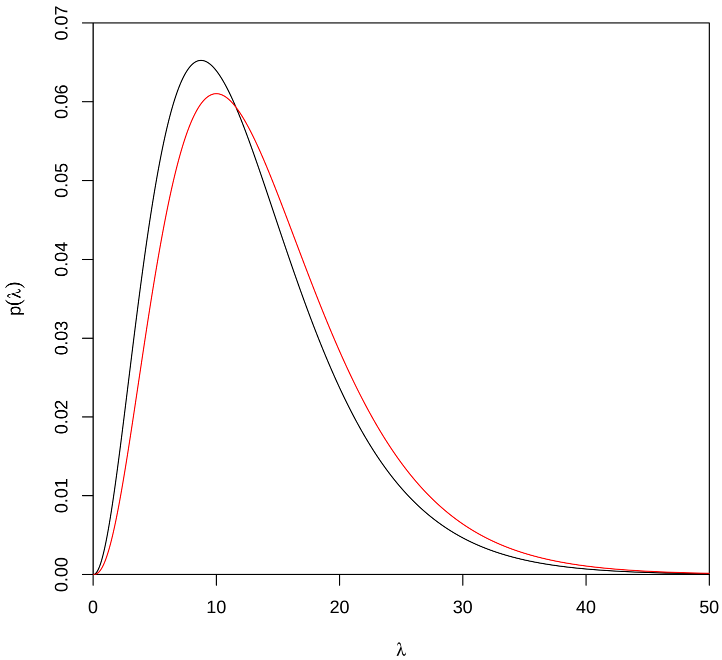

Figure 2: Posterior distribution for the rate per year, λ, after observing the O1 data for the conjugate prior (black line) and the Jeffrey's prior (red line).

(**1 mark**). This can be recognised as a negative binomial distribution. We compute the probability of seeing 6 or more events in O2 as $1 - \sum_{n=0}^{5} p(n|d_{O1}) = 0.504$ (**1 mark**). The posterior for the rate in the first 5 months of O2 is $\Gamma(3.21, 0.608)$ and the probability of seeing 1 or fewer events in 5 months is given by computing $p(0|d_{O1}) + p(1|d_{O1}) = 0.131$ using this α and β in Eq. (3) (**1 mark**). The posterior probability for the rate in 1 month is $\Gamma(3.21, 3.024)$, from which we compute the probability of seeing 5 or more events in one month of O2 as $p_1 = 1 - \sum_{n=0}^{4} p(n|d_{O1}) = 0.015$. The probability of seeing 5 or more events in at least one month of O2 is given by $1 - (1 - p_1)^6 = 0.086$ (**1 mark**). With the Jeffrey's prior the posterior distributions on the rate in 6 months, 5 months and 1 month are $\Gamma(3.5, 0.5)$, $\Gamma(3.5, 0.6)$ and $\Gamma(3.5, 3.0)$ respectively. The probabilities of seeing 6 or more events in O2, 1 or fewer in 5 months of O2, 5 or more in the last month of O2 and seeing 5 or more in at least 1 month of O2 are 0.564, 0.103, 0.019 and 0.110 respectively (**2 marks**). The probability that the last month would contain the number of events that were seen is significantly small (at a 2% confidence level). However, there is no reason to single out the last month *a priori* and the probability that one month would be at least this exceptional is only around 10%, which is small but not sufficiently significant to be a cause for concern. The choice of prior does not significantly influence this, indicating that we are data dominated and the conclusion is robust. So, based on O2 we cannot conclude the rate is inhomogeneous in time, but the significance is high enough that we should collect more data and see if the next science run shows any evidence for a time-dependent rate (**1 mark**).

(f) In total over O1 and O2 we see 9 events and the total observing time is 0.75 years. Therefore the

R	$p(r = R \mid d_{1+2})$	$p(r \leq R \mid d_{1+2})$	R	$p(r = R \mid d_{1+2})$	$p(r \leq R \mid d_{1+2})$
0	0.123	0.123	6	0.048	0.935
1	0.229	0.352	7	0.029	0.965
2	0.194	0.546	8	0.018	0.983
3	0.158	0.704	9	0.009	0.992
4	0.109	0.813	10	0.003	0.995
5	0.075	0.888	11	0.002	0.997

Table 1: Posterior predictive probability of the absolute difference in the number of events detected in the first and second 6 month periods of the O3 science run. The columns give the difference in the number of events, the posterior probability of observing that difference and the cumulative posterior probability of observing a difference less than or equal to that value.

combined posterior is $\Gamma(9.21, 0.752)$ using the conjugate prior derived in (a) or $\Gamma(9.5, 0.75)$ using the Jeffrey's prior. The posterior predictive distribution for the rate in a given 6 month period of O3 is $\Gamma(9.21, 1.504)$ or $\Gamma(9.5, 1.5)$ respectively (**2 marks**). The distribution of the difference $r = |n_1 - n_2|$ of the number of events observed in two independent samples from a Poisson distribution with rate θ is given by the *Skellam distribution* with pmf

$$p(r|\theta) = \begin{cases} \mathrm{e}^{-2\theta} I_0(2\theta) & r = 0 \\ 2\mathrm{e}^{-2\theta} I_r(2\theta) & r = 1, 2 \ldots, \end{cases}$$

where $I_k(x)$ is the modified Bessel function of the first kind. Hence the posterior predictive distribution on r is

$$p(r|d_{1+2}) = \int_0^\infty \mathrm{e}^{-2\lambda} I_r(2\lambda) \frac{\beta^\alpha \lambda^{\alpha-1} \mathrm{e}^{-\beta\lambda}}{\Gamma(\alpha)} \mathrm{d}\lambda$$

for a $\Gamma(\alpha, \beta)$ posterior distribution on the rate. Note that this can also be written as the difference between two independent negative binomial variables, which follow a *generalised discrete Laplace distribution*, but the expressions that must be evaluated are no easier than this integral (**2 marks** Note: marks were also awarded for a correct numerical derivation of the distribution).

Table 1 lists the cumulative posterior density for the difference r. We see that there is less than 5% probability of seeing a difference of 7 or more events. Therefore a difference of this size or larger would be significant at a 5% level (**2 marks**).

There are a number of other ways in which this question could be addressed. For example, we could look at the number of events in each month and set a threshold, based on the posterior predictive distribution, on the difference between the largest and smallest monthly count. Alternatively, we could model the rates in each one month period as being potentially different, with λ_i denoting the rate in month i. These rates can be connected by a hyperprior, e.g., $\lambda_i \sim \Gamma(\alpha, \beta)$, and the parameters of that hyperprior constrained from the data. Alternatively, the rates can be modelled parametrically, e.g., $\lambda_i = a + bi$, and the parameters of the parametric model constrained from the data. If the posterior on the slope parameter, b, is inconsistent with 0 there is evidence for an evolving rate. Similarly if the parameters of the hyperprior are inconsistent with a constant rate there is evidence for evolution. The advantage of these kind of approaches is that the results of the analysis give an estimate of the nature and size of the effect, not just the presence of the effect. (**2 marks**)

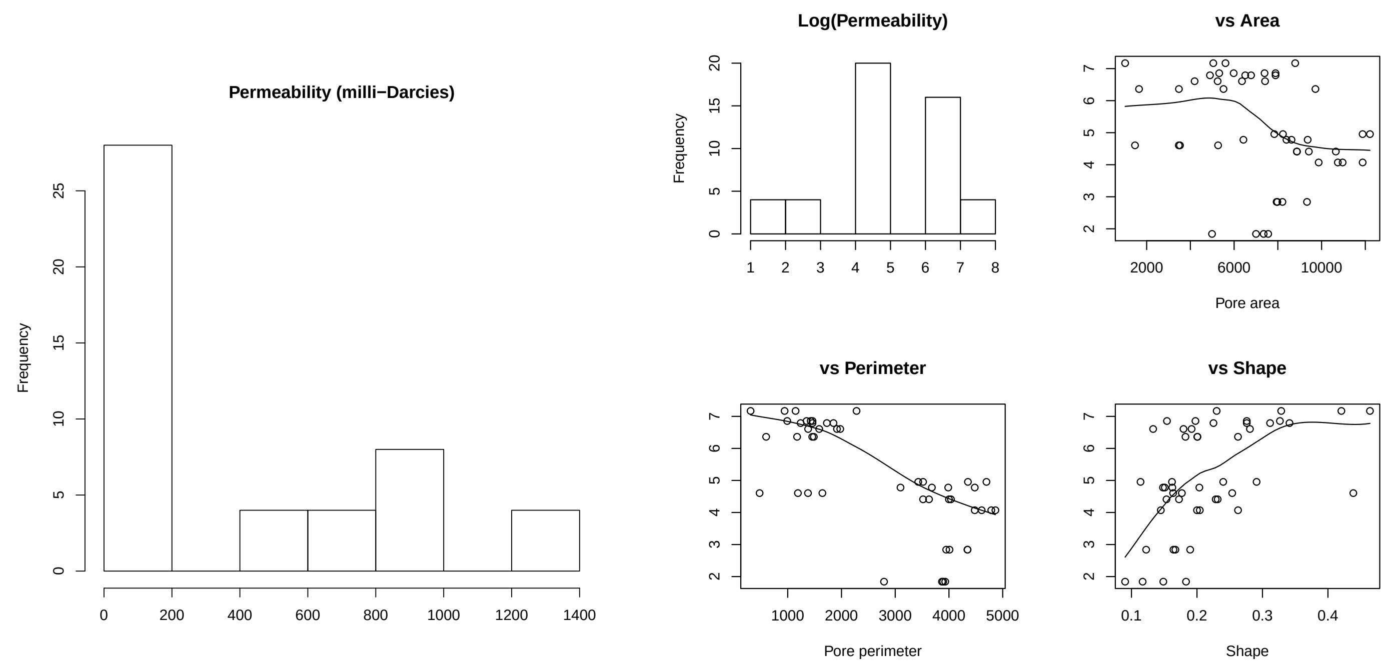

Figure 3: Histogram of permeability dataset (left). Histogram of the log transformed permeability dataset and scatter plots of log-permeability versus the various covariates (right).

3. (a) We first look at the data. The left panel of Figure 3 shows a histogram of the data and this exhibits a significant right skew. We therefore transform the data using a log-transform, which looks better. We then carry out some exploratory data analysis on the log-transformed data. The histogram of log-permeability and scatter plots of the log permeability against the three covariates are shown in the right panel of Figure 3. There is a lot of scatter but it looks like a linear relationship might be a reasonable model (**5 marks** for exploratory data analysis).

   (b) Now we fit a Bayesian linear model to the data with likelihood

$$p(\mathbf{d}|\vec{\beta},\tau) = \prod_{i=1}^{n} \frac{\sqrt{\tau}}{\sqrt{2\pi}} \exp\left[-\frac{\tau}{2}\left(d_i - \beta_0 - \beta_1 \mathbf{x}_{1i} - \beta_2 \mathbf{x}_{2i} - \beta_3 \mathbf{x}_{3i}\right)^2\right]$$

   where $\mathbf{d}$ denotes the set of measurements of log-permeability, indexed by i, n is the number of data points, $\vec{\beta}$ is the vector of linear model parameters, x_{1i} denotes the centred value of the pore area for the i'th data point, x_{2i} denotes the centred value of the pore perimeter for the i'th data point, x_{3i} denotes the centred value of the pore shape for the i'th data point and τ denotes the precision of the measurements. We use normal priors on the linear model parameters, $\beta_i \sim N(\mu_0, 1/\tau_0)$ and a gamma prior on the precision, $\tau \sim \Gamma(a_\tau, b_\tau)$. We choose $\mu_0 = 0$, $\tau_0 = 10^{-6}$ and $a_\tau = b_\tau = 0.01$. Note that we are using centred covariates, found for example using

```
data(rock)
area<-rock$area
```

```
area.ctr <- area-mean(area)
```

as this helps with chain mixing. (**5 marks** for model specification)

(c) We fit the model using JAGS which leads to the posterior distributions and trace plots shown in Figure 4 and the posterior summary statistics below

```
1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                 Mean        SD  Naive SE Time-series SE
beta.area   0.0004848 8.834e-05 5.100e-07      1.302e-06
beta.peri  -0.0015264 1.806e-04 1.042e-06      2.770e-06
beta.shape  1.7642818 1.795e+00 1.036e-02      1.752e-02
beta0       5.1079246 1.257e-01 7.255e-04      7.273e-04
sigma       0.8673506 9.460e-02 5.461e-04      6.218e-04

2. Quantiles for each variable:

                 2.5%        25%        50%        75%      97.5%
beta.area   0.0003084  0.0004264  0.0004857  0.0005436  0.0006562
beta.peri  -0.0018783 -0.0016466 -0.0015294 -0.0014073 -0.0011687
beta.shape -1.7609837  0.5761853  1.7644753  2.9403106  5.3525078
beta0       4.8564923  5.0245234  5.1079841  5.1914466  5.3522960
sigma       0.7086047  0.8005816  0.8585609  0.9238689  1.0764513
```

We see that all the linear model parameters are significant, i.e., different from 0, except the coefficient of the "shape" parameter. The shape parameter could be eliminated from the model. (**5 marks** for model fit and posterior summary statistics)

(d) The trace plots show good mixing, but we can check convergence by looking at the autocorrelation functions, shown in Figure 5, by computing the effective sample sizes which, for a run of 10000 samples, were 10000, 1516, 1386, 3742, 7735 and by computing the Gelman-Rubin statistic, shown in Figure 6. The autocorrelation falls off quickly and the Gelman-Rubin statistics are all close to 1, with potential scale reduction factors of 1. (**5 marks**)

(e) We perform prior robustness checks by choosing different values for $(\mu_0, \tau_0, a_\tau, b_\tau)$ and comparing posterior distributions and summary statistics. For example $\mu_0 = 0.1$, $\tau_0 = 10^{-3}$, $a_\tau = b_\tau = 0.1$ gives

```
Iterations = 1001:11000
Thinning interval = 1
Number of chains = 3
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:
```

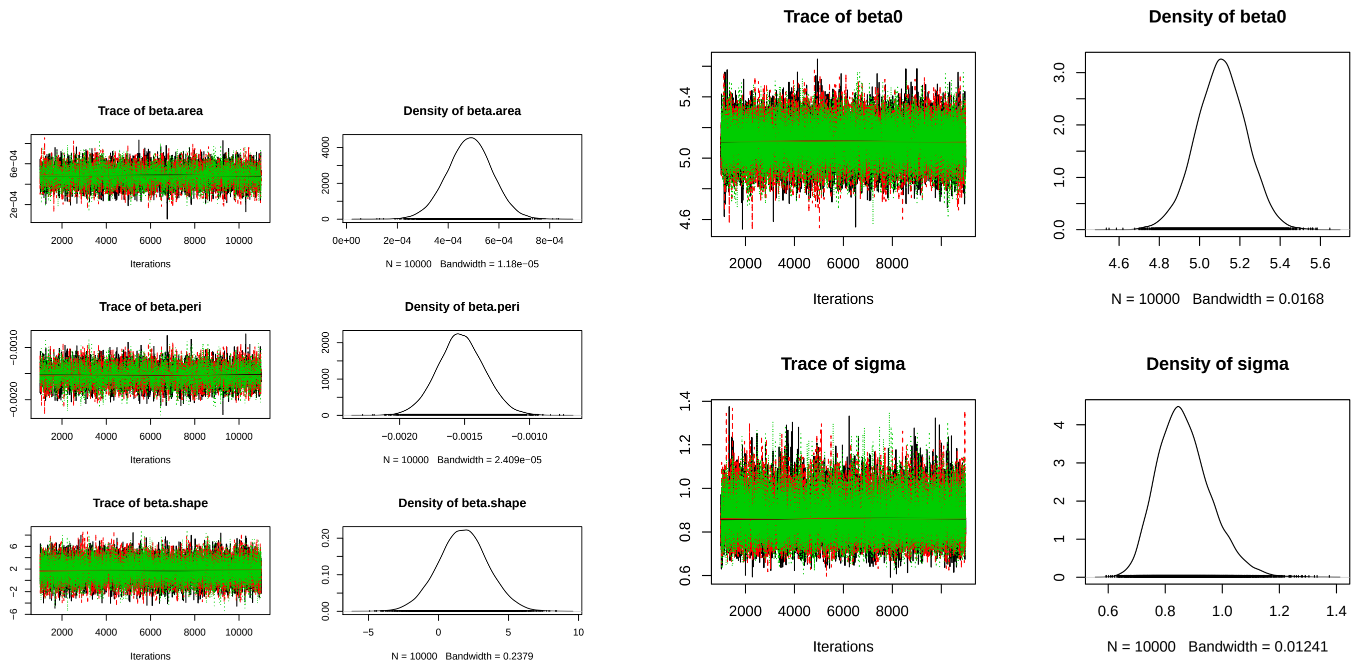

Figure 4: Posterior distributions and trace plots for the Bayesian model fit to the rock permeability data.

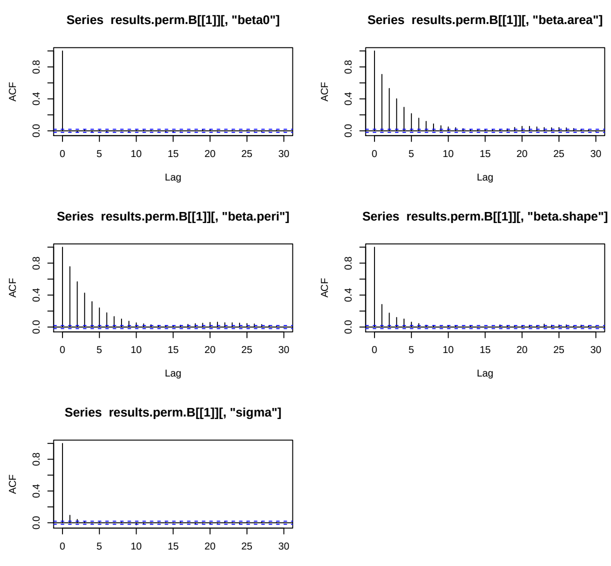

Figure 5: Autocorrelation function plots for the fit to the rock permeability data.

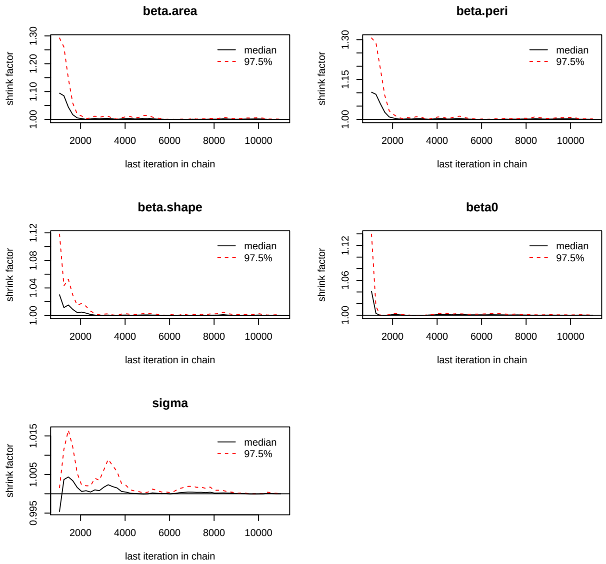

Figure 6: Plots of the Gelman-Rubin statistic for the fit to the rock permeability data.

```
                   Mean        SD  Naive SE Time-series SE
beta.area    0.0004852 8.841e-05 5.105e-07      1.288e-06
beta.peri   -0.0015277 1.806e-04 1.043e-06      2.692e-06
beta.shape   1.7510902 1.792e+00 1.035e-02      1.686e-02
beta0        5.1084410 1.252e-01 7.230e-04      7.230e-04
sigma        0.8667420 9.421e-02 5.439e-04      6.096e-04

2. Quantiles for each variable:

                  2.5%       25%        50%        75%     97.5%
beta.area    0.0003116  0.000427  0.0004856  0.0005433  0.000658
beta.peri   -0.0018830 -0.001647 -0.0015273 -0.0014090 -0.001172
beta.shape  -1.7898424  0.558516  1.7505669  2.9375347  5.256450
beta0        4.8633358  5.024347  5.1076797  5.1924800  5.356873
sigma        0.7074132  0.800324  0.8595123  0.9244228  1.074699
```

Similarly, we could try changing the prior distribution, e.g., using a student-t distribution with 5 degrees of freedom for the linear model parameters and a log-normal distribution for the precision, τ. This gives

```
Iterations = 2001:12000
Thinning interval = 1
Number of chains = 3
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
```

```
   plus standard error of the mean:

                 Mean        SD  Naive SE Time-series SE
beta.area   0.0004835 8.944e-05 5.164e-07      1.433e-06
beta.peri  -0.0015223 1.829e-04 1.056e-06      3.090e-06
beta.shape  1.7925099 1.799e+00 1.038e-02      2.024e-02
beta0       5.1063383 1.259e-01 7.269e-04      9.159e-04
sigma       0.8667029 9.547e-02 5.512e-04      7.964e-04

2. Quantiles for each variable:

                 2.5%        25%       50%        75%      97.5%
beta.area   0.0003072  0.0004244  0.000483  0.0005433  0.0006606
beta.peri  -0.0018862 -0.0016432 -0.001522 -0.0014017 -0.0011588
beta.shape -1.7528893  0.5900453  1.798074  3.0048406  5.2719718
beta0       4.8557938  5.0233656  5.106779  5.1903618  5.3532085
sigma       0.7035797  0.7994135  0.858336  0.9254219  1.0759625
```

In both cases the results do not change too much, showing that we are insensitive to the prior choice. (**5 marks** For some prior robustness checks.)

(f) Finally, we carry out some posterior predictive checking of the model by computing studentised residuals. In Figure 7 we show the studentised residuals, a QQ plot based on these residuals, the residuals versus fitted value and the posterior predictive distribution of the minimum and maximum value in a dataset of this size. These plots all look reasonable, indicating that the model is good and fits the data well. The maximum value in the data set looks low relative to the posterior predictive distribution, perhaps indicating that the normal error approximation needs modifying, but even that value is consistent with the predictive distribution. (**5 marks** For posterior predictive checks.)

(g) The above is a sufficient fit for this problem. The remaining (**5 marks**) were given for a discussion of the results and some discussion of additional investigations that could be performed. For example, a comparison to the result of a standard regression analysis using the LM command, which gives the following

```
Call:
lm(formula = log(perm) ~ area + peri + shape, data = rock)

Residuals:
    Min      1Q  Median      3Q     Max
-1.8092 -0.5413  0.1734  0.6493  1.4788

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.333e+00  5.487e-01   9.720 1.59e-12 ***
area         4.850e-04  8.657e-05   5.602 1.29e-06 ***
```

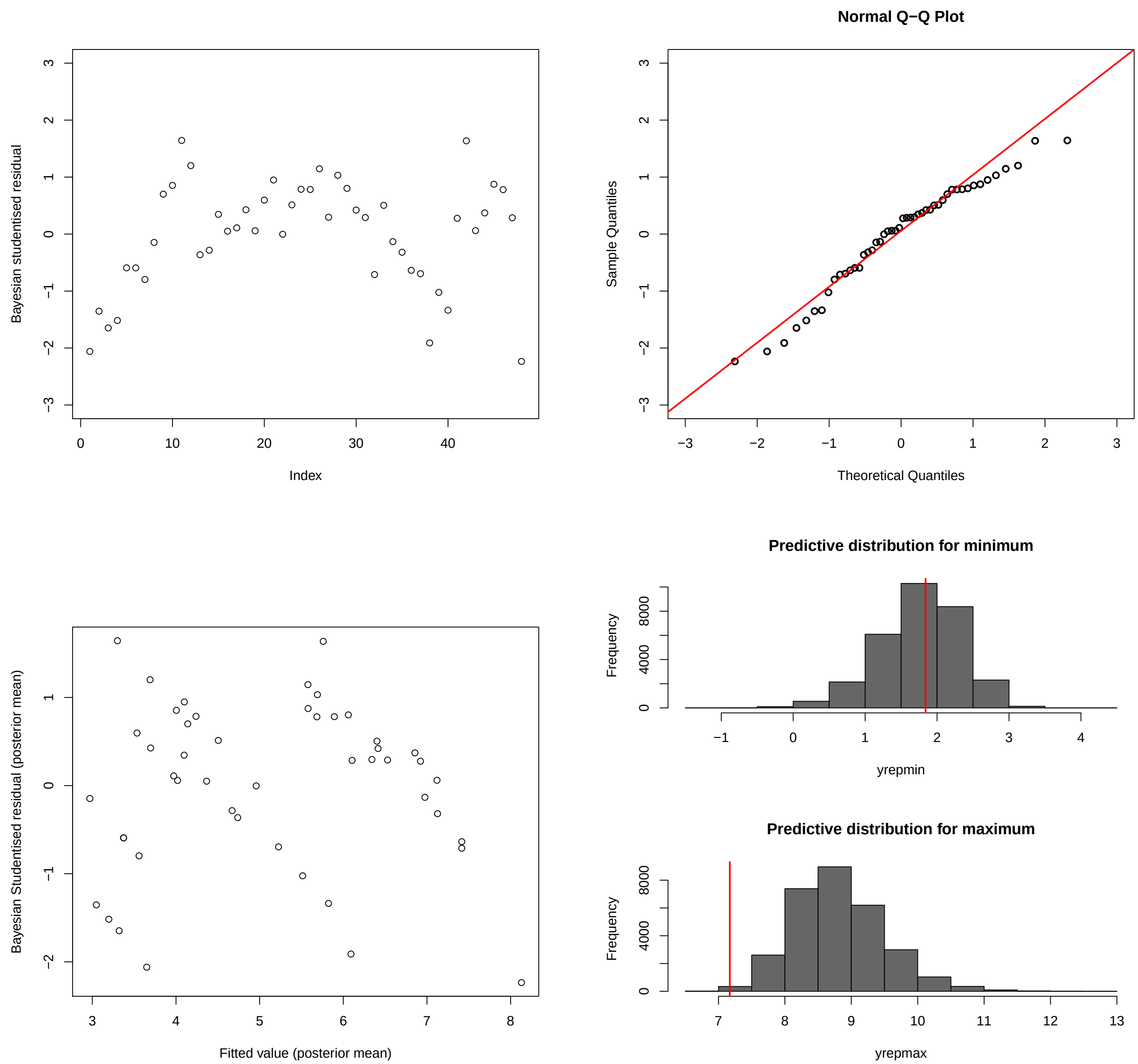

Figure 7: Posterior predictive plots: a) studentised residuals versus index (top left); b) Q-Q plot of studentised residuals (top right); c) studentised residuals versus fitted value; d) posterior predictive distribution of the maximum and minimum log-permeability, with observed values marked.

```
peri        -1.527e-03  1.770e-04  -8.623 5.24e-11 ***
shape        1.757e+00  1.756e+00   1.000    0.323
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8521 on 44 degrees of freedom
Multiple R-squared:  0.7483,Adjusted R-squared:  0.7311
F-statistic:  43.6 on 3 and 44 DF,  p-value: 3.094e-13
```

This gives very similar results to the Bayesian fit, in particular it gives the same conclusion that the shape parameter can be dropped from the model. Note that the value of the intercept has changed since this fit is not using centred covariates.

Alternatively, one could assess robustness to outliers by changing the measurement error distribution from Gaussian to a student-t. Doing this with the degrees of freedom, ν, assigned a prior of $\Gamma(0.01, 0.01)$ yields the following

```
Iterations = 2001:12000
Thinning interval = 1
Number of chains = 3
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                 Mean        SD  Naive SE Time-series SE
beta.area   0.0004858 9.953e-05 5.746e-07      1.847e-06
beta.peri  -0.0015371 1.976e-04 1.141e-06      3.757e-06
beta.shape  1.6300188 1.834e+00 1.059e-02      2.240e-02
beta0       5.1376812 1.322e-01 7.631e-04      1.189e-03
sigma       0.8148645 1.103e-01 6.366e-04      1.341e-03

2. Quantiles for each variable:

                 2.5%        25%        50%        75%      97.5%
beta.area   0.0002907  0.0004197  0.0004848  0.0005518  0.0006804
beta.peri  -0.0019311 -0.0016672 -0.0015354 -0.0014049 -0.0011576
beta.shape -1.9740418  0.4199686  1.6382097  2.8435686  5.2376135
beta0       4.8818035  5.0496499  5.1364463  5.2250233  5.4000731
sigma       0.6004362  0.7441767  0.8121035  0.8841363  1.0380799
```

So the conclusions and parameter estimates are essentially unchanged.