Homework #2

Stat4DS2+DS
https://elearning.uniroma1.it/course/view.php?id=7253

deadline 06/25/2019 (23:55)

Antonini Valerio 1611556

1a) Illustrate the characteristics of the statistical model for dealing with the *Dugong*'s data. Lengths (Y_i) and ages (x_i) of 27 dugongs (see cows) captured off the coast of Queensland have been recorded and the following (non linear) regression model is considered in Carlin and Gelfand (1991):

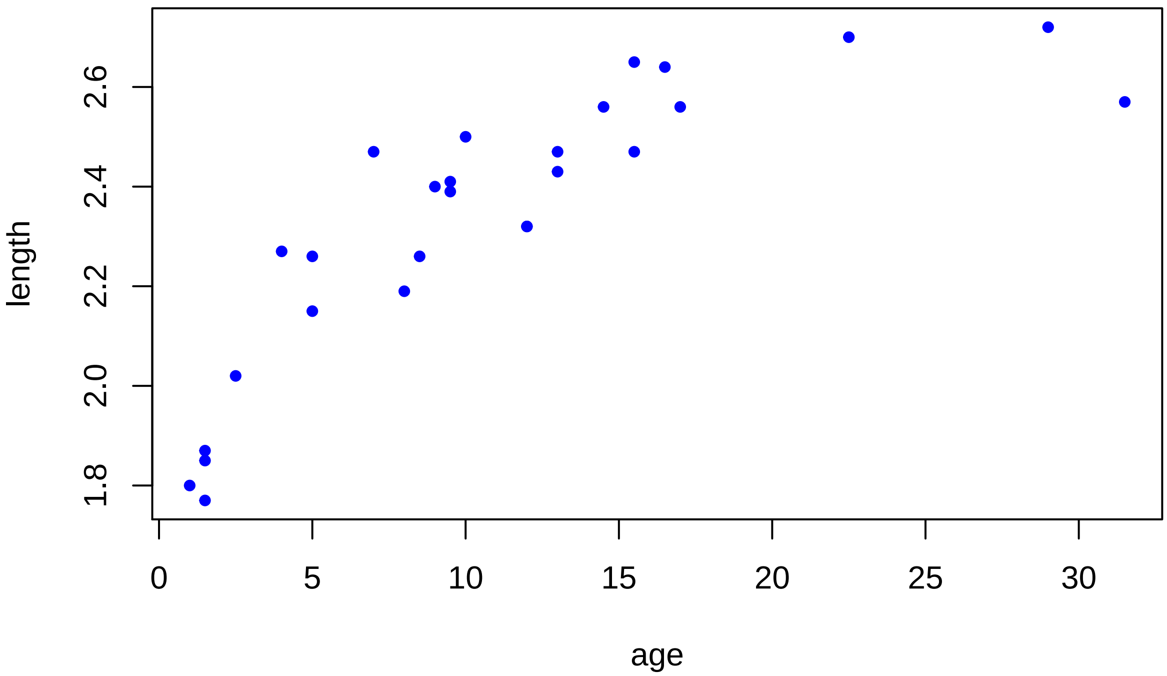
$$\begin{aligned} Y_i &\sim N(\mu_i, \tau^2) \\ \mu_i = f(x_i) &= \alpha - \beta\gamma^{x_i} \end{aligned}$$

Model parameters are $\alpha \in (1, \infty)$, $\beta \in (1, \infty)$, $\gamma \in (0, 1)$, $\tau^2 \in (0, \infty)$.

Let us consider the following prior distributions:

$$\begin{aligned} \alpha &\sim N(0, \sigma_\alpha^2) \\ \beta &\sim N(0, \sigma_\beta^2) \\ \gamma &\sim Unif(0, 1) \\ \tau^2 &\sim IG(a, b))(InverseGamma) \end{aligned}$$

```r
df = read.csv("dugong-data.txt", sep = "")
x = df$Age
y = df$Length
plot(x, y, xlab = "age", ylab = "length", pch = 20, col = "blue")
```



1b) Derive the corresponding likelihood function

$$L_y(\alpha,\beta,\gamma,\tau^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\tau^2}} \exp\{-\frac{1}{2}(\frac{(y_i-\alpha+\beta\gamma^{y_i})^2}{\tau^2})\} I_{(0,\infty)}(\alpha) I_{(0,\infty)}(\beta) I_{(0,1)}(\gamma) I_{(0,\infty)}(\tau^2) =$$

$$= \frac{1}{(2\pi\tau^2)^{\frac{n}{2}}} \exp\{-\frac{1}{2\tau^2}\sum_{i=1}^{n}(y_i-\alpha+\beta\gamma^{y_i})^2\} I_{(0,\infty)}(\alpha) I_{(0,\infty)}(\beta) I_{(0,1)}(\gamma) I_{(0,\infty)}(\tau^2)$$

1c) Write down the expression of the joint prior distribution of the parameters at stake and illustrate your suitable choice for the hyperparameters.

First, let's define the marginal prior distributions of the each parameters:

$$\alpha \sim \mathsf{N}(0,\sigma_\alpha^2) --> \pi(\alpha) = \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\{-\frac{1}{2}(\frac{\alpha}{\sigma_\alpha^2})^2\}$$

$$\beta \sim \mathsf{N}(0,\sigma_\beta^2) --> \pi(\beta) = \frac{1}{\sqrt{2\pi\sigma_\beta^2}} \exp\{-\frac{1}{2}(\frac{\beta}{\sigma_\beta^2})^2\}$$

$$\gamma \sim \mathsf{Unif}(0,1) --> \pi(\gamma) = I_{(0,1)}(\gamma)$$

$$\tau^2 \sim \mathsf{IG}(a,b) --> \pi(\tau^2) = \frac{b^a}{\Gamma(a)} \tau^{2^{(-a-1)}} exp\left\{-\frac{b}{\tau^2}\right\} I_{(0,\infty)}(\tau^2) \propto \tau^{2^{(-a-1)}} exp\left\{-\frac{b}{\tau^2}\right\} I_{(0,\infty)}(\tau^2)$$

Prior on $\alpha,\beta,\gamma,\tau^2$

$$\pi(\alpha,\beta,\gamma,\tau^2) = (\frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\{-\frac{1}{2}(\frac{\alpha}{\sigma_\alpha^2})^2\})(\frac{1}{\sqrt{2\pi\sigma_\beta^2}} \exp\{-\frac{1}{2}(\frac{\beta}{\sigma_\beta^2})^2\})(\frac{b^a}{\Gamma(a)} \tau^{2^{(-a-1)}} exp\left\{-\frac{b}{\tau^2}\right\}$$

$$\propto \tau^{2^{(-a-1)}} exp\Bigg\{-\frac{b}{\tau^2}\Bigg\} I_{(0,\infty)}(\alpha) I_{(0,\infty)}(\beta) I_{(0,1)}(\gamma) I_{(0,\infty)}(\tau^2) =$$

$$= \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \frac{1}{\sqrt{2\pi\sigma_\beta^2}} \frac{b^a}{\Gamma(a)} \exp\{-\frac{1}{2}(\frac{\alpha}{\sigma_\alpha^2})^2(\frac{\beta}{\sigma_\beta^2})^2\}\tau^{2(-a-1)} \exp\{-\frac{b}{\tau^2}\}$$

$$\propto \exp\{-\frac{1}{2}(\frac{\alpha}{\sigma_\alpha^2})^2(\frac{\beta}{\sigma_\beta^2})^2\}\tau^{2(-a-1)} \exp\{-\frac{b}{\tau^2}\}$$

The best choice for the hyperparameters is the one that guarantee high variance. In order to get that, i choose: $\sigma_\alpha = 10000$, $\sigma_\beta = 10000$, $a = 0.001$, $b = 0.001$

1d) Derive the functional form (up to proportionality constants) of all *full-conditionals*

Full conditional for α:

$$\pi(\alpha|\beta,\gamma,\tau^2,x,y) = \frac{L(\alpha,\beta,\gamma,\tau^2|Y)\pi(\alpha)}{f(y)} \propto L(\alpha,\beta,\gamma,\tau^2|Y)\pi(\alpha)$$

$$\propto \frac{1}{(2\pi\tau^2)^{n/2}} \exp\{-\frac{1}{2}\sum_{i=1}^n (y_i-\mu_i)^2\} \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\{-\frac{1}{2}(\frac{\alpha}{\sigma_\alpha^2})^2\} I_{(0,\infty)}(\alpha)$$

$$\propto \exp\{-\frac{1}{2}(\sum_{i=1}^n (\frac{y_i-\alpha+\beta\gamma^{x_i}}{\tau^2})^2 - (\frac{\alpha}{\sigma_\alpha^2})^2\} I_{(0,\infty)}(\alpha)$$

$$= \exp\{\frac{-\sigma_\alpha^2\sum_{i=1}^n (y_i-\alpha+\beta\gamma^{x_i})^2 + \gamma^2\alpha^2}{2\tau^2\sigma_\alpha^2}\} I_{(0,\infty)}(\alpha)$$

$$= \exp\{\frac{-\sigma_\alpha^2\sum_{i=1}^n (\alpha^2-2\alpha y_i-2\alpha\beta\gamma^{x_i})^2 + \gamma^2\alpha^2}{2\tau^2\sigma_\alpha^2}\} I_{(0,\infty)}(\alpha)$$

$$= \exp\{\frac{n\sigma_\alpha^2\alpha^2 - 2\alpha\sigma_\alpha^2\sum_{i=1}^n y_i - 2\alpha\beta\sigma_\alpha^2\sum_{i=1}^n \gamma^{x_i-\tau^2\alpha^2}}{2\tau^2\sigma_\alpha^2}\} I_{(0,\infty)}(\alpha)$$

$$= \exp\{\alpha^2\frac{n\sigma_\alpha^2+\tau^2}{2\tau^2\sigma_\alpha^2} + \frac{\alpha\sum_{i=1}^n (y_i+\beta\gamma^{x_i})}{\tau^2}\} I_{(0,\infty)}(\alpha)$$

Full conditional for β:

$$\pi(\beta|\alpha,\gamma,\tau^2,x,y) = \frac{L(\alpha,\beta,\gamma,\tau^2|Y)\pi(\beta)}{f(y)} \propto L(\alpha,\beta,\gamma,\tau^2|Y)\pi(\beta)$$

$$\propto \frac{1}{(2\pi\tau^2)^{n/2}} \frac{1}{\sqrt{2\pi\sigma_\beta^2}} \exp\{-\frac{1}{2}\frac{\sum_{i=1}^n (y_i-\alpha+\beta\gamma^{x_i})^2}{\tau^2}\} \exp\{-\frac{1}{2}(\frac{\beta^2}{\sigma_\beta^2})\} I_{(0,\infty)}(\beta)$$

$$= \exp\{-\frac{1}{2}\frac{\sigma_\beta^2\sum_{i=1}^n (y_i-\alpha+\beta\gamma^{x_i})^2 + \tau^2\beta^2}{\tau^2\sigma_\beta^2}\} I_{(0,\infty)}(\beta)$$

$$= \exp\Bigg\{-\frac{\beta^2}{2}\frac{\tau^2+\sigma_\beta^2\sum\gamma^{2x_i}}{\tau^2\sigma_\beta^2} + \beta\frac{\sigma_\beta^2\sum\gamma^{x_i}(\alpha-y_i)}{\tau^2\sigma_\beta^2}\Bigg\}$$

Full conditional for γ:

$$\pi(\gamma|\alpha,\beta,\tau^2,x,y) = \frac{1}{(2\pi\tau^2)^{\frac{n}{2}}}\exp\{-\frac{1}{2\tau^2}\sum_{i=1}^{n}(y_i-\mu_i)^2 I_{(n,2)}(\gamma)$$

$$\propto \exp\{\frac{1}{2\tau^2}\sum_{i=1}^{n}(y_i-\alpha+\beta\gamma^{x_i})^2\}I_{(n,2)}(\gamma)$$

Full conditional for τ^2

$$\pi(\tau^2|\alpha,\beta,\gamma,x,y) = \frac{1}{\tau^{2(\frac{n}{2})}\tau^{2(a+1)}}\exp\{\frac{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\alpha+\beta\gamma^{x_i})^2-b}{\tau^2}\}I_{(0,\infty)(\tau)}$$

$$\pi(\tau^2|\alpha,\beta,\gamma,x,y) \sim \mathsf{IG}(a+\frac{n}{2}, b+\frac{1}{2}\sum_{i=1}^{n}(y_i-\alpha+\beta\gamma^{x_i})^2)$$

1e) Which distribution can you recognize within standard parametric

families so that direct simulation from full conditional can be easily implemented ?

$$\pi(\alpha|\beta,\gamma,\tau^2,x,y) \sim \mathsf{N}(\frac{\sigma_\alpha^2\sum_{i=1}^{n}(y_i+\beta\gamma^{x_i})}{n\sigma_\alpha^2+\gamma^2}, \frac{\tau^2\sigma_\alpha^2}{n\sigma_\alpha^2+\tau^2})$$

$$\pi(\beta|\alpha,\gamma,\tau^2,x,y) \sim \mathsf{N}(\frac{\sigma_\beta^2\sum_{i=1}^{n}\gamma^{x_i}(\alpha-y_i)}{\tau^2+\sigma_\beta^2\sum_{i=1}^{n}\gamma^{2x_i}}, \frac{\tau^2\sigma_\beta^2}{\tau^2+\sigma_\beta^2\sum_{i=1}^{n}\gamma^{2x_i}})$$

Both α and β are truncated Normal.

$$\pi(\tau^2|\alpha,\beta,\gamma,x,y) \sim \mathsf{IG}(a+\frac{n}{2}, b+\frac{1}{2}\sum_{i=1}^{n}(y_i-\alpha+\beta\gamma^{x_i})^2)$$

Instead the distribution of γ is an unknown distribution.

1f) Using a suitable Metropolis-within-Gibbs algorithm simulate a Markov chain

($T = 10000$) to approximate the posterior distribution for the above model

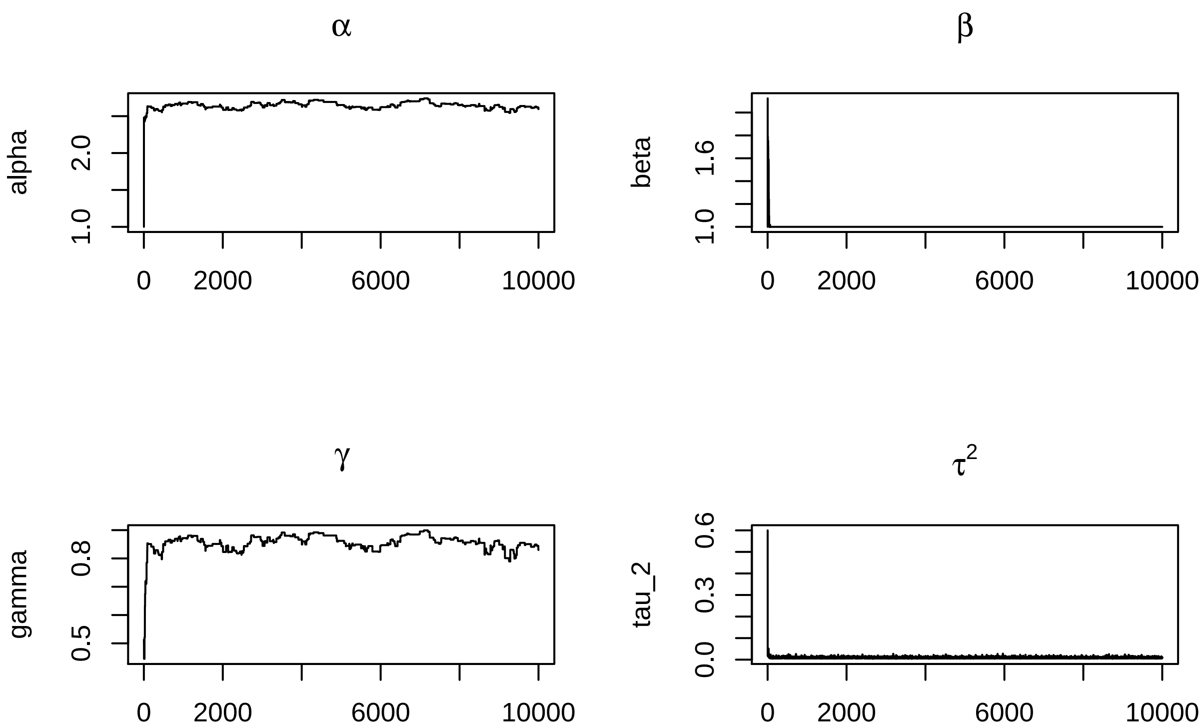
```
library(truncnorm)
n = length(x)
iter = 10000

alpha = rep(NA, iter + 1)
beta = rep(NA, iter + 1)
gamma = rep(NA, iter + 1)
tau_2 = rep(NA, iter + 1)

full_cond_gamma = function(gamma, alpha, beta, tau_2) {
    return(exp(-1/(2 * tau_2) * sum((y - alpha + beta * gamma^x)^2)))

}
```

```
alpha[1] = 1
beta[1] = 1
gamma[1] = 0.5
tau_2[1] = 0.6

for (i in 1:iter) {


    alpha[i + 1] = rtruncnorm(1, (10000 * sum(y + beta[i] * gamma[i]^x))/(n *
        10000 + tau_2[i]), sqrt(tau_2[i] * 10000)/(n * 10000 +
        tau_2[i]), a = 1, b = Inf)


    beta[i + 1] = rtruncnorm(1, (10000 * sum((alpha[i + 1] -
        y) * gamma[i]^x))/(10000 * sum(alpha[i + 1] - y) * gamma[i]^(2 *
        x) + tau_2[i]), sqrt(tau_2[i] * 10000)/(10000 * sum(gamma[i]^(2 *
        x)) + tau_2[i]), a = 1, b = Inf)


    proposal = runif(1)

    ratio = full_cond_gamma(proposal, alpha[i + 1], beta[i +
        1], tau_2[i])/full_cond_gamma(gamma[i], alpha[i + 1],
        beta[i + 1], tau_2[i])

    gamma[i + 1] = ifelse(runif(1, 0, 1) <= ratio, proposal,
        gamma[i])

    tau_2[i + 1] = invgamma::rinvgamma(1, n/2 + 0.001, 0.001 +
        1/2 * sum((y - alpha[i + 1] + beta[i + 1] * (gamma[i +
            1]^x))^2))


}

MG = cbind(alpha, beta, gamma, tau_2)
head(MG, 10)
```

```
##          alpha     beta     gamma      tau_2
##  [1,] 1.000000 1.000000 0.5000000 0.60000000
##  [2,] 2.404134 2.124182 0.5000000 0.03159836
##  [3,] 2.482262 1.146443 0.5132697 0.03249949
##  [4,] 2.417873 1.793405 0.4452364 0.03225700
##  [5,] 2.437054 1.627528 0.4452364 0.01972263
##  [6,] 2.427588 1.756834 0.4452364 0.02273766
##  [7,] 2.435061 1.650393 0.4452364 0.02095459
##  [8,] 2.428893 1.739850 0.4452364 0.01726833
##  [9,] 2.434025 1.664455 0.4452364 0.01525005
## [10,] 2.429608 1.729518 0.4452364 0.02116133
```

1g) Show the 4 univariate trace-plots of the simulations of each parameter

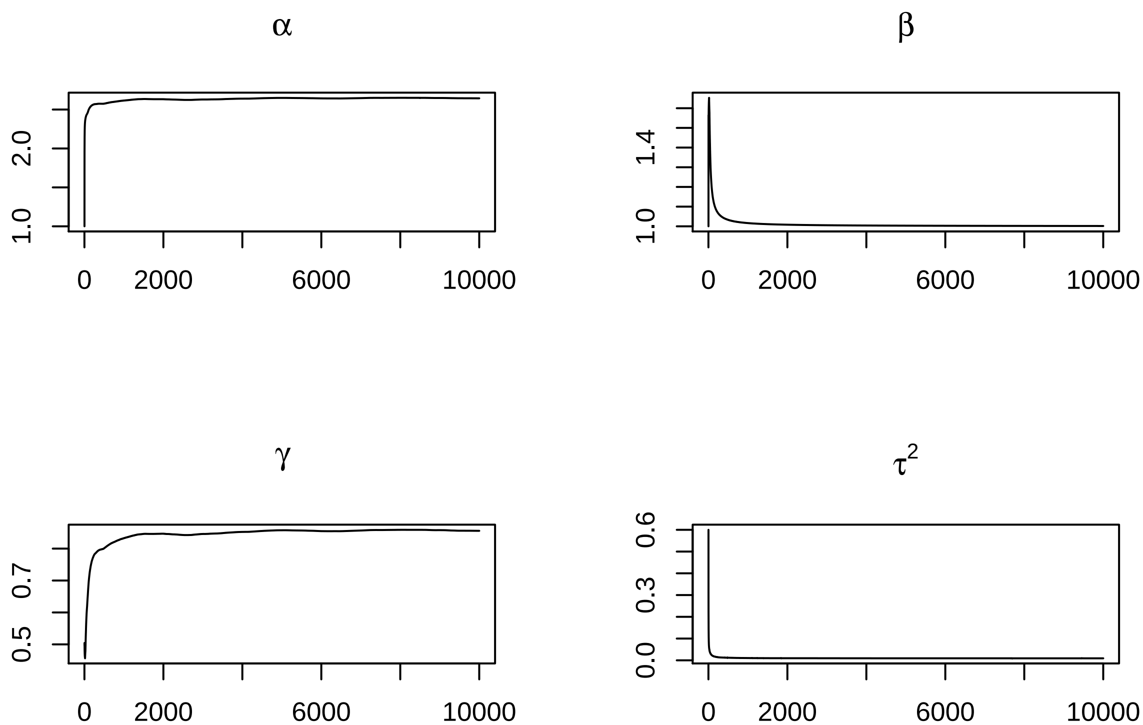
```r
par(mfrow = c(2, 2))
plot(alpha, xlab = "", main = expression(alpha), type = "l")
plot(beta, xlab = "", main = expression(beta), type = "l")
plot(gamma, xlab = "", main = expression(gamma), type = "l")
plot(tau_2, xlab = "", main = expression(tau^2), type = "l")
```



1h) Evaluate graphically the behaviour of the empirical averages

$\hat{I}_t$ with growing $t = 1, ..., T$

```r
par(mfrow = c(2, 2))
plot(cumsum(alpha)/(1:length(alpha)), xlab = "", ylab = "", main = expression(alpha),
    type = "l")
plot(cumsum(beta)/(1:length(beta)), xlab = "", ylab = "", main = expression(beta),
    type = "l")
plot(cumsum(gamma)/(1:length(gamma)), xlab = "", ylab = "", main = expression(gamma),
    type = "l")
plot(cumsum(tau_2)/(1:length(tau_2)), xlab = "", ylab = "", main = expression(tau^2),
    type = "l")
```



1i) Provide estimates for each parameter together with the

approximation error and explain how you have evaluated such error

The expression for the approximation error is: $\sigma^2_{\hat{I}_t} = Var[\hat{I}_t] = \frac{Var_\pi[h(X_i)]}{t_{eff}}$.

with $t_{eff} = \frac{t}{1+s\sum_{k=1}^{\infty}\rho_k}$

```
alpha_hat = mean(alpha)
beta_hat = mean(beta)
gamma_hat = mean(gamma)
tau_hat = mean(tau_2)
cbind(alpha_hat, beta_hat, gamma_hat, tau_hat)
```

```
##      alpha_hat beta_hat gamma_hat     tau_hat
## [1,]  2.644263 1.001636 0.8557039 0.009239816
```

```
approx_err_alpha = var(alpha)/length(alpha)
approx_err_beta = var(beta)/length(beta)
approx_err_gamma = var(gamma)/length(gamma)
approx_err_tau_2 = var(tau_2)/length(tau_2)
rbind(approx_err_alpha, approx_err_beta, approx_err_gamma, approx_err_tau_2)
```

```
##                          [,1]
## approx_err_alpha 2.166384e-07
## approx_err_beta  1.015870e-07
## approx_err_gamma 9.367974e-08
## approx_err_tau_2 4.327165e-09
```

1l) Which parameter has the largest posterior uncertainty? How did

you measure it?

The uncertainty is evaluated computing the standard deviation of the parameters

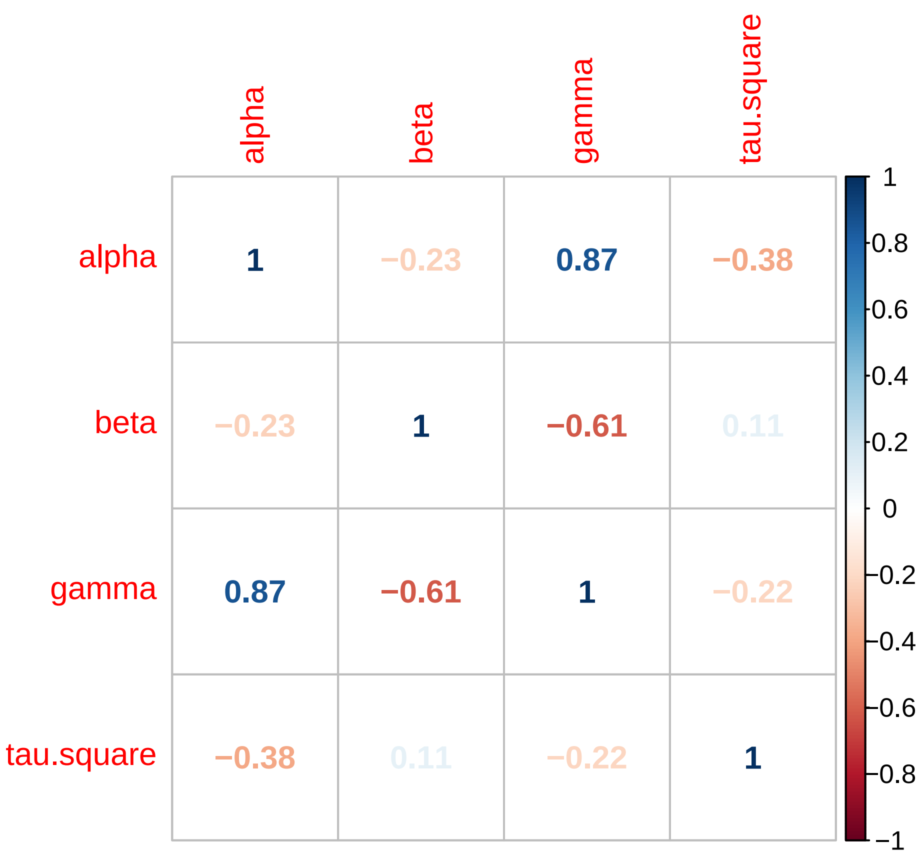
```
variance_alpha = sd(alpha)
variance_beta = sd(beta)
variance_gamma = sd(gamma)
variance_tau = sd(tau_2)

rbind(variance_alpha, variance_beta, variance_gamma, variance_tau)
```

```
##                       [,1]
## variance_alpha 0.046546755
## variance_beta  0.031874315
## variance_gamma 0.030608677
## variance_tau   0.006578448
```

1m) Which couple of parameters has the largest correlation (in

absolute value)?

```
library(corrplot)
colnames(MG) <- c("alpha", "beta", "gamma", "tau.square")
corrplot(cor(MG), method = "number")
```



The parameters with the higest correlation are α and γ.

1n) Use the Markov chain to approximate the posterior predictive

distribution of the length of a dugong with age of 20 years.

```
post_pred_distr_20 = rep(NA, 10000)
for (i in 1:10000) {
    mu = alpha[i] - beta[i] * gamma[i]^20
    post_pred_distr_20[i] <- rnorm(1, mu, sqrt(tau_2[i]))
}
mean(post_pred_distr_20)
```

```
## [1] 2.594433
```

1o) Provide the prediction of a different dugong with age 30

```
post_pred_distr_30 = rep(NA, 10000)
for (i in 1:10000) {
    mu = alpha[i] - beta[i] * gamma[i]^30
    post_pred_distr_30[i] <- rnorm(1, mu, sqrt(tau_2[i]))
}
mean(post_pred_distr_30)
```

```
## [1] 2.633466
```

1p) Which prediction is less precise?

```
uncertainty_20 = sd(post_pred_distr_20)
c("uncertainty for prevision on the duogong of 20 age:", uncertainty_20)
```

```
## [1] "uncertainty for prevision on the duogong of 20 age:"
## [2] "0.0994996895296603"
```

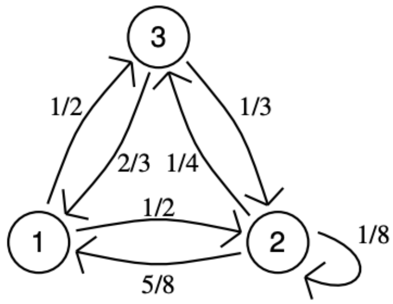
```
uncertainty_30 = sd(post_pred_distr_30)
c("uncertainty for prevision on the duogong of 30 age:", uncertainty_30)
```

```
## [1] "uncertainty for prevision on the duogong of 30 age:"
## [2] "0.101239450583985"
```

The prediction less precise is one about duogong with 30 age

Part 2)

Let us consider a Markov chain $(X_t)_{t \geq 0}$ defined on the state space $\mathcal{S} = \{1, 2, 3\}$ with the following transition



2a) Starting at time $t = 0$ in the state $X_0 = 1$ simulate the Markov chain with distribution assigned as ####above for $t = 1000$ consecutive times

```
states <- c(1, 2, 3)
transition_matrix <- matrix(data = c(0, 1/2, 1/2, 5/8, 1/8, 1/4,
    2/3, 1/3, 0), byrow = TRUE, nrow = 3)
t0 <- 1
nsample <- 1000
MChain <- rep(NA, nsample + 1)
MChain[1] <- t0
for (t in 1:nsample) {
    MChain[t + 1] <- sample(states, size = 1, prob = transition_matrix[MChain[t],
        ])
}
```

2b) compute the empirical relative frequency of the two states in your simulation

```
prop.table(table(MChain))
```

```
## MChain
##         1         2         3
## 0.3916084 0.3226773 0.2857143
```

2c) repeat the simulation for 500 times and record onlythe final state at time $t = 1000$ for each of the 500 simulated chains. Compute the relative frequency of the 500 final states.

What distribution are you approximating in this way?

Try to formalize the difference between this point and the previous point.

```
t0 <- 1
ntimes <- 500
nsample <- 1000
t1000 <- rep(NA, ntimes)
for (i in 1:ntimes) {
```

```r
    MChain <- rep(NA, nsample + 1)
    MChain[1] <- t0
    for (t in 1:nsample) {
        MChain[t + 1] <- sample(states, size = 1, prob = transition_matrix[MChain[t],
            ])
    }
    t1000[i] <- MChain[nsample + 1]
}
prop.table(table(t1000))
```

```
## t1000
##     1     2     3
## 0.416 0.324 0.260
```

2d) compute the theoretical stationary distribution π and explain how you have obtained it

π is one of the eigenvectors coming from solving the system $(P???\pi I)\pi = 0$ We can used one eigenvector and normalizing it, so we get:

```r
pi <- eigen(t(transition_matrix))$vector[, 1]/sum(eigen(t(transition_matrix))$vector[,
    1])
```

2e) is it well approximated by the simulated empirical relative frequencies computed in (b) and (c)?

```r
pi
```

```
## [1] 0.3917526 0.3298969 0.2783505
```

It seems very similar with respect to the previous results

2f) what happens if we start at $t = 0$ from state $X_0 = 2$ instead of $X_0 = 1$?

```r
MChain <- rep(NA, nsample + 1)
MChain[1] <- 2
for (i in 1:nsample) {
    MChain[i + 1] <- sample(states, size = 1, prob = transition_matrix[MChain[i],
        ])
}
prop.table(table(MChain))
```

```
## MChain
##         1         2         3
## 0.3966034 0.3156843 0.2877123
```

The new chain get results very near to the previous ones.

```
## This homework will be graded and it will be part of your final evaluation
##
##
## Last update by LT: Tue Jun 25 11:51:26 2019
```