Bayesian_Assignment

Getrude Gichuhi

106218

2023-02-14

Question One

a) Describe what a Markov Chain is and how Markov Chain Monte Carlo methods are used in statistical modeling.

Markov Chain: A Markov chain is a random process in which the future is independent of the past, giventhe present. Markov Chain Monte Carlo methods can be used to identify the posterior distribution and sample from it, A technique used in fitting models and drawing sample from posterior distribution of the parameters.

b)Explain how the accept-reject approach works.

The accept-reject method is a simple algorithm for generating random samples from a target probability distribution that may be difficult to sample from directly. The generated samples are either accepted or rejected.

Accept-Reject Approach:

- Choose a simpler probability distribution, for instance unniform distribution
- Generate a sample from the simpler distribution.
- Evaluate the ratio of the probability density function (PDF) to the simpler PDF for the generated sample. T
- Generate a uniform random variable between 0 and 1.
- If the generated uniform variable is less than or equal to the acceptance probability, then accept the sample. Otherwise, reject it and go back to step 2. -Repeate the steps till it closely approximates the target distribution

c) Explain how the Metropolis-hasting and the Gibbs sampler work.

Metropolis-hasting algorithm: Generates a Markov chain where each state is generated from previous state through a proposal distribution. Acceptance probability ensures algorithm satisfies the detailed balance distribution of a Markov chain converging towards the target distibution

Gibbs Sampler: Generates a markov process where each state is generated by sampling from the distribution of each component given the previous state. However it alternates between sampling each component, ensuring that the resulting Markov chain converges to the target distribution.

d) Explain what you understand by the terminologies: stationarity, irreducibility and ergodicity within the framework of MCMCs.

Stationarity: A MCMC is said to be stationary incase the distribution of the state at any given time is the same as the distribution of it's state at any other time. No change

Irreducibility: A MCMC is said to be irreducible if it is possible to get from any state to any other state in a finite number of steps. No absorbing states

Ergodicity: A Markov chain is said to be ergodic if it is both irreducible and aperiodic. Aperiodicity means that the chain does not exhibit any regular, repeating patterns.

e) Explain the meaning of the term Monte Carlo error (MC error) and its role in the analysis of MCMC outputs.

Monte Carlo error (MC error) is a measure of the error associated with Monte Carlo integration, which is a common method for approximating integrals in MCMC methods.

MC error is often used to estimate the accuracy of the generated samples. The error can be estimated by computing the standard deviation of the estimated quantity based on multiple independent runs of the MCMC algorithm. A smaller MC error implies that the generated samples are more accurate and closer to the true distribution.

The batch mean method is used to calculate the MC error. It works by first partitioning the resulting MCMC output sample into K batches.The number of batches K and the sample size of each batch $v = \frac{T\prime}{K}$ must be sufficiently large to allow one to estimate the variance consistently and remove auto correlations. To estimate the MC error of the posterior mean of $G(\theta)$, you first get the batch mean i.e $G(\theta)$ would be given by

$$\overline{G(\theta)} = \frac{1}{v} \sum_{t=(b-1)v+1}^{bv} G\left(\theta^{(t)}\right)$$

for each batch $b = 1, \ldots\ldots, K$ the overall mean is:

$$\overline{G(\theta)} = \frac{1}{T'} \sum_{t=1}^{T'} G\left(\theta^{(t)}\right) = \frac{1}{k} \sum_{b=1}^{k} \overline{G(\theta)_b}$$

Assumning all the observations from $\theta^1 \ldots\ldots, \theta^{T'}$ are retained, then the MC error estimate is given by the standard deviation of the batch means estimate $\overline{G(\theta)_b}$

$$MCE[G(\theta)] = \widehat{se}\left[\overline{G(\theta)}\right] = \sqrt{\frac{1}{k} \widehat{sd}\left[\overline{G(\theta)_b}\right]}$$

$$MCE[G(\theta)] = \sqrt{\frac{1}{k(k-1)}\sum_{b=1}^{k}\left(\overline{G(\theta)}_b\right) - \overline{G(\theta)})^2}$$

f) Explain what you understand by the terminologies: thinning and burn-in period within the framework of MCMCs

Thinning: This is a post-processing step that involves selecting a subset of the MCMC samples to reduce correlation among the generated samples. With this, only the K-th sample from the mCMC chain is reserved. This method is useful when the samples generated are highly correlated, however, there is a likelihood of an MC error of the remaining samples being not representative of the distribution.

Burn-in Period: This is the initial period of MCMC iterations during which the chain is still adapting to the d specific istribution. It's purpose is ensuring the chain converges to the specific distribution and that the initial samples are not biased by the starting point of the chain. Samples generated are typically discarded and not used for subsequent analysis. When calculating the BigO notation, depends on the specific MCMC algorithm and the complexity of the target distribution.

g) Suppose that a random variable Y has a Poisson distribution. That is,

-Assuming a Gamma prior-distribution for derive expressions for the posterior density, posterior mean and variance for Y. Hint: The density of a Gamma variate X is:

$$f(y;\theta) = \frac{e^{-\theta}\theta^{y}}{y!}, y = 0,1, \ldots \ldots ..$$

Assuming a $Gamma(\alpha, \beta)$ prior-distribution for θ, derive expressions for the posterior density, posterior mean and variance for Y. using the gamma density given by:

$$f(x:\theta) = \frac{\beta^{\alpha}}{\Gamma\alpha} e^{-\beta x} x^{\alpha-1}, x > 0$$

$$f(\theta|y) = f(y|\theta) * y(\theta)$$

Getting the likelihood function for the poisson distribution:

$$f(y|\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{y_i}}{y_i!} = \frac{e^{-n\theta}\theta^{\sum_{i=1}^{n} yi}}{\prod_{i=1}^{n} y_i\,!}$$

$$f(\theta|y) = \frac{e^{-n\theta}\theta^{\sum_{i=1}^{n} yi}}{\prod_{i=1}^{n} y_i\,!} \times \frac{\beta^{\alpha}}{\Gamma\alpha} e^{-\beta\theta}\theta^{\alpha-1}$$

$$= e^{-n\theta}\theta^{\sum_{i=1}^{n} yi} * \beta^{\alpha} e^{-\beta\theta}\theta^{\alpha-1}$$

collecting the like terms together we get:

$$= e^{-\theta(n+\beta)}\theta^{n\bar{y}+\alpha}$$

Thus

$$\theta|y\ gamma(n\bar{y}+\alpha, n+\beta)$$

Posterior mean will then be:

$$E[\theta|y] = \frac{\alpha}{\beta} = \frac{n\bar{y}+\alpha}{n+\beta}$$

$$= \frac{n}{n+\beta}\bar{y} + \frac{\beta}{n+\beta}\left(\frac{\alpha}{\beta}\right)$$

$$= E(\theta|y) = w\bar{y} + (1-w)\frac{\alpha}{\beta}$$

since

$$E(\theta) = \frac{\alpha}{\beta}$$

:

$$E(\theta|y) = w\bar{y} + (1-w)E(\theta)$$

Posterior variance will be expressed as:

$$var[\theta|y] = \frac{\alpha}{\beta^2} = \frac{n\bar{y}+\alpha}{(n+\beta)^2}$$

$$= \frac{n}{(n+\beta)^2}\bar{y} + \frac{\alpha}{(n+\beta)^2}$$

$$= \frac{n^2}{(n+\beta)^2}\frac{\bar{y}}{n} + \frac{\beta^2}{(n+\beta)^2}\frac{\alpha}{\beta^2}$$

since

$$V(\theta) = \frac{\alpha}{\beta^2}$$

:

$$var[\theta|y] = w^2\frac{\bar{y}}{n} + (1-w)^2 V(\theta)$$

Question Two

a) Explain briefly how the inverse transform method can be used to generate pseudo-random numbers in monte-carlo simulation.

The method involves transforming uniform random variables on the interval 0,1, to random variables with desired distribution.

Steps Involved:

- Develop and define a continous and strick cumulative distribution function (CDF) of the desired distribution.
- Generate a uniform random variable U on the interval.
- Compute the inverse of the CDF to obtain a value

$$X = F^{-1}(U), where F^{-1}$$

  is the inverse CDF of the desired distribution.

Repeat steps 2 and 3 as many times as necessary to obtain the desired number of pseudo-random values.

b) Explain how random numbers can be generated from discrete distributions.

Assuming that a random variable has a probability mass function of $P(X = x_i) = p_i$

- Generate

$$U\ Unif(0,1)$$

- Determine the index k such that

$$\sum_{j=1}^{k-1} p_j <= U < \sum_{j=1}^{k}$$

- Return random numbers

$$X = x_k$$

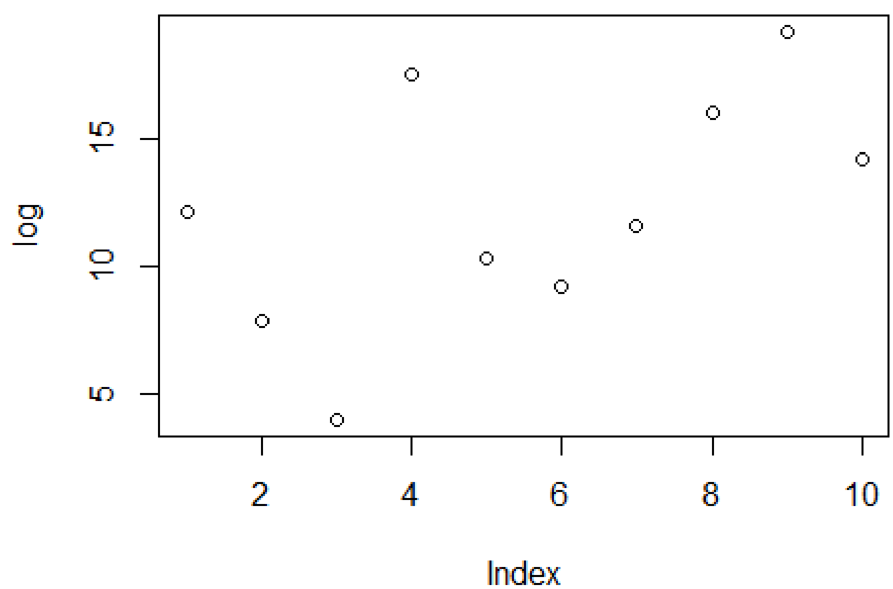
c) Based on the following 10 random numbers from Uniform distribution

0.63, 0.37, 0.18, 0.87, 0.52, 0.45, 0.60, 0.82, 0.91, 0.74

i) Generate 10 random numbers from the Logistic distribution.

```
# inverse CDF for Logistic(10,4)
cdf <- function(p, mu, s) {
  mu + log(p / (1 - p)) * s
}

# Generate 10 random numbers from Logistic(10, 4) using Inverse Transform
Method
```

```
uni <- c(0.63, 0.37, 0.18, 0.87, 0.52, 0.45, 0.60, 0.82, 0.91, 0.74)
log <- cdf(uni, 10, 4)
plot(log)
```



###ii) Generate 10 random numbers from the Binomial distribution.

```
# Generate 10 random numbers from Binomial(4, 0.2) using Inverse Transform
Method

unif <- c(0.63, 0.37, 0.18, 0.87, 0.52, 0.45, 0.60, 0.82, 0.91, 0.74)
bi <- qbinom(unif, 4, 0.2)
bi

##  [1] 1 0 0 2 1 1 1 2 2 1
```

Question Three

Let us consider the exponential distribution with density function and an i.i.d. sample exponential for

a) Show that a gamma prior distribution is conjugate for

$$f(\theta|y) = f(y|\theta) * f(\theta)$$

$$f(y|\theta) = \prod_{i=1}^{n} \theta\, e^{-\theta y_i}$$

$$= f(y|\theta) = \theta^n e^{-\theta \sum_{i=1}^{n} y_i}$$

$$f(\theta) = \frac{\beta^\alpha}{\Gamma\alpha} e^{-\beta\theta} \theta^{\alpha-1}$$

$$= f(\theta|y) = \theta^n e^{-\theta \sum_{i=1}^{n} y_i} \times \frac{\beta^\alpha}{\Gamma\alpha} e^{-\beta\theta} \theta^{\alpha-1}$$

$$= \theta^n e^{-\theta \sum_{i=1}^{n} y_i} * \beta^\alpha e^{-\beta\theta} \theta^{\alpha-1}$$

$$f(\theta|y) = \theta^{n+\alpha} * e^{-\theta(n\bar{y}+\beta)}$$

thus:

$$\theta|y - gamma(n\bar{y} + \beta, n + \alpha)$$

b) Calculate the posterior mean and variance under this conjugate prior.

Posterior mean will be given as:

$$E(\theta|y) = \frac{\alpha}{\beta} = \frac{n\bar{y} + \beta}{n + \alpha}$$

$$= \frac{n}{n+\alpha}\bar{y} + \frac{\beta}{n+\alpha}$$

$$= \frac{n}{n+\alpha}\bar{y} + \frac{\alpha}{n+\alpha}\frac{\beta}{\alpha}$$

$$= w\bar{y} + (1-w)\frac{\beta}{\alpha}$$

$$= E(\theta|y) = w\bar{y} + (1-w)E(\theta)^{-1}$$

Posterior variance is given as:

$$var(\theta|y) = \frac{\alpha}{\beta^2} = \frac{n\bar{y} + \beta}{(n+\alpha)^2}$$

$$= \frac{n}{(n+\alpha)^2}\bar{y} + \frac{\beta}{(n+\alpha)^2}$$

$$= \frac{n^2}{(n+\alpha)^2}\frac{\bar{y}}{n} + \frac{\alpha^2}{(n+\alpha)^2}\frac{\beta}{\alpha^2}$$

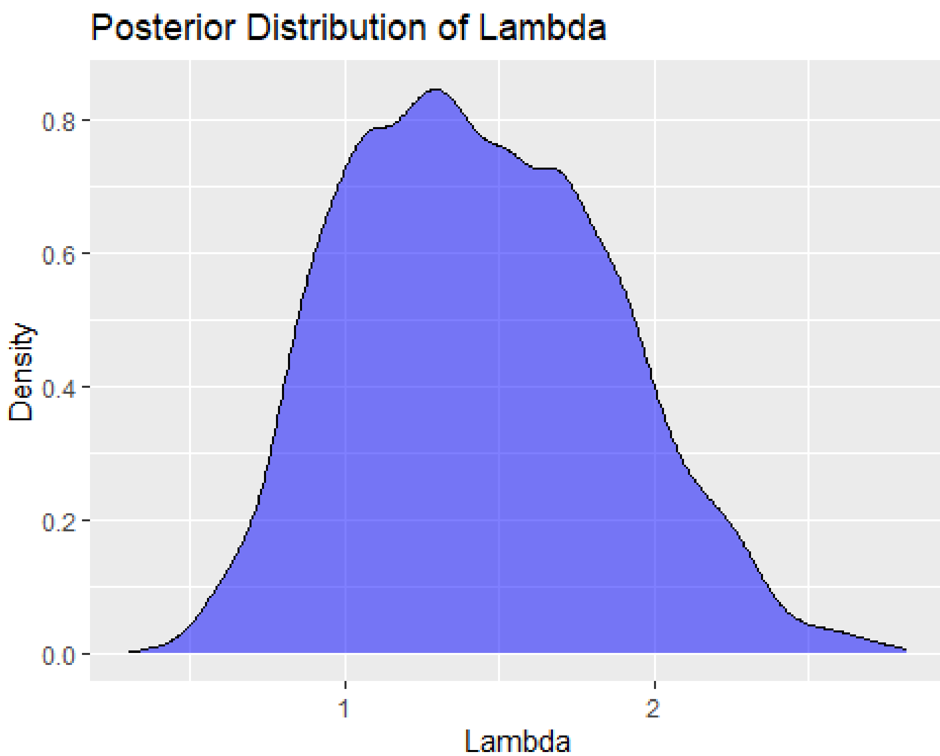
$$V(\theta|y) = w^2\frac{\bar{y}}{n} + (1-w)^2\frac{\beta}{\alpha^2}$$

$$= w^2 \frac{\bar{y}}{n} + (1-w)^2 V(\theta|y)^{-1}$$

c) For the following data 0.4, 0.0, 0.2, 0.1, 2.1, 0.1, 0.9, 2.4, 0.1, 0.2 use the exponential distribution and

i) Plot the posterior distribution for gamma prior parameters $\alpha = \beta = 0.001$.

```
## Loading required package: coda

## Loading required package: MASS

## ##
## ## Markov Chain Monte Carlo Package (MCMCpack)

## ## Copyright (C) 2003-2023 Andrew D. Martin, Kevin M. Quinn, and Jong Hee
Park

## ##
## ## Support provided by the U.S. National Science Foundation

## ## (Grants SES-0350646 and SES-0350613)
## ##

##
##
## @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
## The Metropolis acceptance rate was 0.96753
## @@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
```



ii) Perform sensitivity analysis for various values of a and b. Produce related plots depicting changes on the posterior mean and variance.

```
# Prior parameters to test
#alpha_values <- seq(0.01, 10, by = 0.1)
#beta_values <- seq(0.01, 10, by = 0.1)

# Compute posterior mean and variance for each pair of alpha and beta values
#post_mean_var <- t(sapply(alpha_values, function(a) {
 # sapply(beta_values, function(b) {
  #  posterior_mean_var(data, a, b)
  #})
#}))

# Plot posterior mean and variance as a function of alpha and beta
#post_mean_plot <- ggplot(data.frame(alpha = alpha_values, beta =
beta_values, mean = post_mean_var[, 1]),
                         #aes(x = alpha, y = beta, z = mean)) +
 # geom_contour(aes(color = ..level..), bins = 10) +
  #scale_color_gradient(low = "white", high = "blue") +
  #labs(title = "Posterior Mean", x = "alpha", y = "beta")

#post_var_plot <- ggplot(data.frame(alpha = alpha_values, beta = beta_values,
var = post_mean_var[, 2]),
                   #     aes(x = alpha, y = beta, z = var)) +
  #geom_contour(aes(color = ..level..), bins = 10) +
  #scale_color_gradient(low = "white", high = "blue") +
```

```
  #labs(title = "Posterior Variance", x = "alpha", y = "beta")

#grid.arrange(post_mean_plot, post_var_plot, nrow = 1)
```

Question Four

Assume that a friend tosses a coin 10 times and tells you that “heads” appeared less than 4 times.

a)Calculate the posterior density for the success probability π using a Beta(α,α) prior density.

$$f(\pi|y) = f(y|\pi) * f(\pi)$$

$$f(y|\pi) = \prod_{i=1}^{n} \binom{N_i}{y_i} \pi^{y_i}(1-\pi)^{n-y_i}$$

$$f(y|\pi) = \pi^{\sum_{i=1}^{n} y_i}(1-\pi)^{N-\sum_{i=1}^{n} y_i}$$

$$= f(y|\pi) = \pi^{n\bar{y}}(1-\pi)^{N-n\bar{y}}$$

$$f(\pi) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}$$

$$f(\pi|y) = \prod_{i=1}^{n} \binom{N_i}{y_i} \pi^{y_i}(1-\pi)^{n-y_i} * \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1}$$

$$= \pi^{\sum_{i=1}^{n} y_i}(1-\pi)^{N-\sum_{i=1}^{n} y_i} * \pi^{\alpha-1}(1-\pi)^{\beta-1}$$

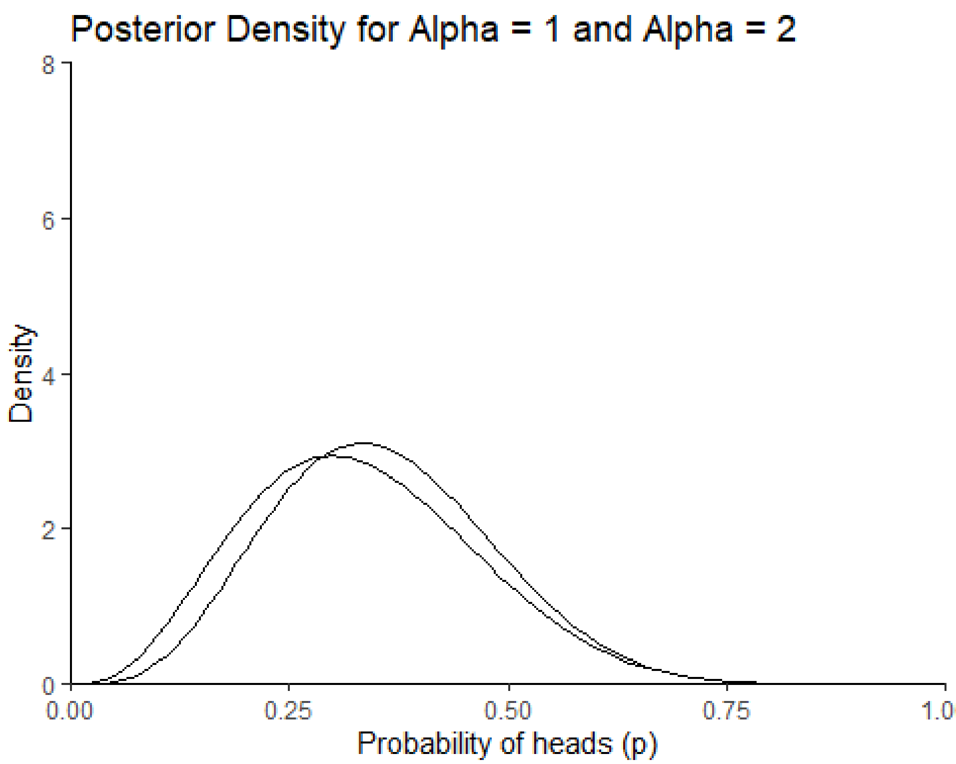
$$= \pi^{n\bar{y}+\alpha-1}(1-\pi)^{N-n\bar{y}+\beta-1}$$

$$\pi|y - beta(n\bar{y}+\alpha, N-n\bar{y}+\beta)$$

b) Plot the posterior density for α=1 and α=2.

```
# calculate posterior density
post_den <- function(x, alpha, beta, k, n) {
  dbeta(x, alpha + k, beta + n - k)
}

# Plot the posterior density for alpha = 1 and alpha = 2
x <- seq(0, 1, length.out = 100)
ggplot(data.frame(x = x), aes(x = x)) +
  stat_function(fun = post_den, args = list(alpha = 1, beta = 1, k = 3, n =
10)) +
  stat_function(fun = post_den, args = list(alpha = 2, beta = 2, k = 3, n =
10)) +
  xlab("Probability of heads (p)") +
  ylab("Density") +
```

```
  ggtitle("Posterior Density for Alpha = 1 and Alpha = 2") +
  scale_x_continuous(limits = c(0, 1), expand = c(0,0)) +
  scale_y_continuous(limits = c(0, 8), expand = c(0,0)) +
  theme_classic() +
  theme(legend.position = "none")
```



C)Calculate the posterior mean of π.

$$E(\pi|y) = \frac{\alpha}{\alpha+\beta} = \frac{n\bar{y}+\alpha}{N-n\bar{y}+\beta+n\bar{y}+\alpha} = \frac{n\bar{y}+\alpha}{N+\beta+\alpha}$$

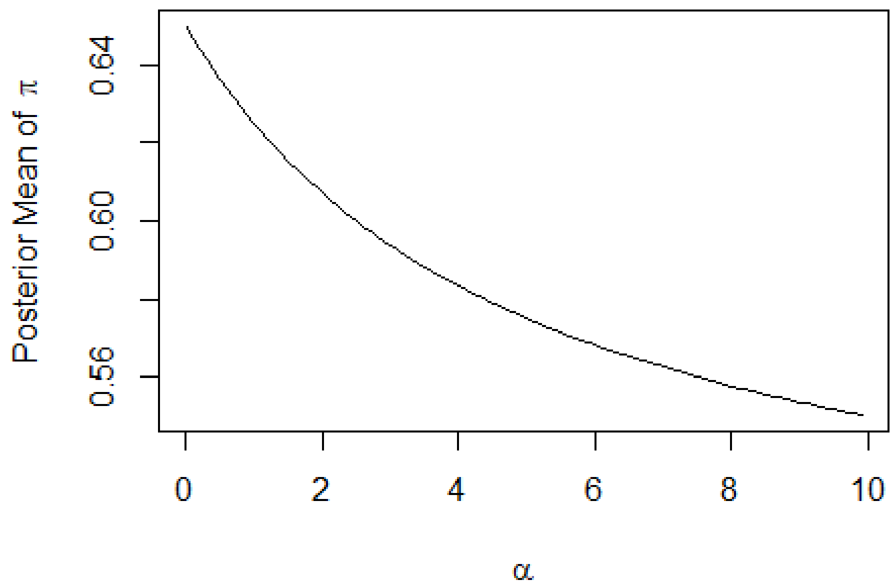
$$\bar{y} = np$$

$$= \frac{10*10*0.5+1}{10+1+2} = \frac{51}{13}$$

d) Plot the posterior mean over different values of the prior parameter π and examine the sensitivity of the results.

```
# posterior mean
alpha_range <- seq(0.01, 10, by = 0.1)
n=10

# Calculate the posterior mean for each value of alpha
post_mean <-sapply (alpha_range, function(alpha) {
 (sum(data) + alpha) / (n + 2*alpha)
})
```

```r
# Plot the posterior mean as a function of alpha
plot(alpha_range, post_mean, type = "l",
     xlab = expression(alpha), ylab = "Posterior Mean of " ~ pi)
```



Question Five

Consider the following data: (0.671, 1.412, -2.119, 1.224, -1.168, -0.860, 1.936, 3.396, 4.808, -1.259, 0.271, 1.820, 2.417, 2.929, 7.020, 0.483, 6.483, 2.966, 0.942, 2.846, -3.398, 3.840, 6.640, 1.018, 2.747, 1.857, 7.270, 2.734, 4.325, -1.222)

a) Assuming the normal distribution for these data and the corresponding conjugate prior, calculate the posterior mean and variance.

```r
# Data
df <- c(0.671, 1.412, -2.119, 1.224, -1.168, -0.860, 1.936, 3.396, 4.808, -
1.259,
          0.271, 1.820, 2.417, 2.929, 7.020, 0.483, 6.483, 2.966, 0.942,
2.846,
          -3.398, 3.840, 6.640, 1.018, 2.747, 1.857, 7.270, 2.734, 4.325, -
1.222)

# Sample mean and variance
n <- length(df)
x_bar <- mean(df)
s2 <- var(df)
```

```
cat("Sample mean: ", x_bar, "\n")

## Sample mean:  2.067633

cat("Sample variance: ", s2, "\n")

## Sample variance:  7.322025

# Prior parameters
mu_0 <- 0
sigma_0 <- 1

# Known variance
sigma2 <- 1

# Posterior parameters
tau2 <- sigma2/sigma_0^2
mu_n <- ((n*x_bar*sigma_0^2)/sigma2 + mu_0/tau2)/((n*sigma_0^2)/sigma2 +
1/tau2)
sigma_n2 <- 1/((1/sigma2) + (n/sigma_0^2))

cat("Posterior mean: ", mu_n, "\n")

## Posterior mean:  2.000935

cat("Posterior variance: ", sigma_n2, "\n")

## Posterior variance:  0.03225806
```

b) Produce graphical representations of the marginal posterior distributions for μ and σ^2 as well as their corresponding joint posterior distribution.

Question Eleven

Let y_{it} denote the outcome of a random variable for the i-th individual at time and let x_{it} denote a covariate measure on the i-th individual at time t, t, i=1,…,n,t=1,…,T. The data could be modeled using the simple linear regression model: $y_{it} = \beta_0 + \beta_1 x_t + \epsilon_{it}, \epsilon_{it} \sim N(0, \tau)$

a) Explain why using this approach would be inapprpriate

There is a fixed linear relationship that is assumed, thus the model fails to consider individual effects.

b) Write a mathematical expression of the random intercept model and the random coefficients models as alternatives to this model.

In an attempt to repair the model to one that accomodates for the random intercept models, we rewrite the simple linear regression into two:

- Random intercept Model

$$y_i = \beta_0 + \beta_1 x_i + u + i$$

where

$$y_i$$

is the outcome of the i-th individual and

$$x_i$$

is a vector of predictor variables, β_0 is the overall intercept and β_1 is a vector of coefficients for the predictor variables and u_i is a random error term that represents the individual-specific intercept such that

$$u_i \sim N(0, \sigma^2)$$

- Random coefficients model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{iT}$ where y_i is the dependent variable for the i-th individual, x_{ij} is the j-th predictor variable for the i-th individual, β_0 is the overall intercept, β_t is a vector of coefficients for the j-th predictor variable, and u_i is a random error term that represents the individual-specific coefficients. The assumption is that the error term $u_i \sim N(0, \sigma^2)$

There is incoporation of unobserved individuals through the intercept or the coefficients.

c) Provide an expression for the proportion of variance attributable to each of the variance components in the random coefficients models.

In a random coefficients model, the total variance of the dependent variable can be partitioned into two components: the within-group variance and the between-group variance.

The within-group variance can be expressed as:

$$Var(u_i) = \sigma^2$$

where u_i is the individual-specific error term and σ^2 is the constant within-group variance. This component captures the variation in the outcome variable within each individual, after controlling for the predictor variables.

The between-group variance can be expressed as:

$$Var(\beta_j) = \sum_j$$

where β_j is a vector of coefficients for the j-th predictor variable and $\sum_j$ is the covariance matrix of the coefficients.

The proportion of variance attributable to the within-group variance and the between-group variance can be expressed as:

Proportion of variance attributable to within-group variance = $\frac{\sigma^2}{(\sigma^2+\sum_j)}$

Proportion of variance attributable to between-group variance = $\frac{\sum_j}{(\sigma^2+\sum_j)}$

These proportions help to understand the level of variability in the relationships between the predictor factors and the result variable by quantifying the relative contribution of each component to the overall variance in the dependent variable.

d) Write out an expression for the likelihood, priors and posterior for the model in part

The likelihood function for the random coefficients model is given by:

$$L(y|\beta, X) = (2\pi\sigma^2)^{(-n/2)} * \exp(-1/2\sigma^2 * \sum(y_i - X_i\beta)^2)$$

where y is the vector of dependent variables, β is the vector of coefficients, X is the design matrix of predictor variables, σ^2 is the within-group variance, and n is the number of observations.

The priors for the coefficients β and the within-group variance σ^2 can be specified using a multivariate normal distribution and an inverse-gamma distribution, respectively:

$$\beta \sim N\left(\mu_\beta, \sum_\beta\right)$$

$$\sigma^2 \sim IG(\alpha, \beta)$$

where $\mu_\beta$and $\sum_\beta$ are the mean vector and covariance matrix of the coefficients, and α and β are the shape and rate parameters of the inverse-gamma distribution.

The posterior distribution can then be calculated using Bayes' theorem:

$$p(\beta, \sigma^2|y, X) = L(y|\beta, X) * p(\beta) * p(\sigma^2)/p(y|X)$$

where $p(y|X)$ is the marginal likelihood.

e) Write out a WinBUGS code for this problem.