1. BDA3 Problem 2.1

1. Posterior inference: suppose you have a Beta(4, 4) prior distribution on the probability θ that a coin will yield a 'head' when spun in a specified manner. The coin is independently spun ten times, and 'heads' appear fewer than 3 times. You are not told how many heads were seen, only that the number is less than 3. Calculate your exact posterior density (up to a proportionality constant) for θ and sketch it.

$\pi(\theta) \propto \theta^3 (1-\theta)^3$, $y \sim Bin(10, \theta)$, $\theta \in (0,1)$

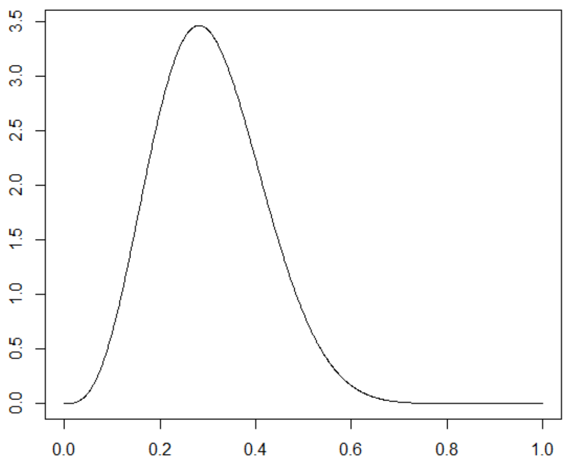
$p(y<3 \mid \theta) = \sum_{i=0}^{2} \binom{10}{i} \theta^i (1-\theta)^{10-i}$

$= (1-\theta)^{10} + 10\,\theta(1-\theta)^9 + 45\,\theta^2 (1-\theta)^8$

$\pi(\theta \mid y<3) \propto \pi(\theta) \cdot p(y<3 \mid \theta)$ Seems like a mixture model

$\propto \theta^3 (1-\theta)^{13} + 10\,\theta^4 (1-\theta)^{12} + 45\,\theta^5 (1-\theta)^{11}$

I used the "integrate" function in R to get the normalizing constant, after which, I got the posterior:



2. BDA3 Problem 2.5

5. Posterior distribution as a compromise between prior information and data: let y be the number of heads in n spins of a coin, whose probability of heads is θ.

(a) If your prior distribution for θ is uniform on the range $[0, 1]$, derive your prior predictive distribution for y,

$$\Pr(y = k) = \int_0^1 \Pr(y = k|\theta)d\theta,$$

for each $k = 0, 1, \ldots, n$.

(b) Suppose you assign a Beta(α, β) prior distribution for θ, and then you observe y heads out of n spins. Show algebraically that your posterior mean of θ always lies between your prior mean, $\frac{\alpha}{\alpha+\beta}$, and the observed relative frequency of heads, $\frac{y}{n}$.

(c) Show that, if the prior distribution on θ is uniform, the posterior variance of θ is always less than the prior variance.

(d) Give an example of a Beta(α, β) prior distribution and data y, n, in which the posterior variance of θ is higher than the prior variance.

(a) $P(y=k) = \int_0^1 P(y=k|\theta)\,\pi(\theta)\,d\theta = \int_0^1 P(y=k|\theta)\cdot 1\,d\theta$

$= \int_0^1 \binom{n}{k} \theta^k (1-\theta)^{n-k} d\theta$

$= \binom{n}{k} \int_0^1 \underbrace{\theta^k (1-\theta)^{n-k}}\, d\theta$

$\hookrightarrow$ It's a Beta$(k+1, n-k+1)$ kernel

$\Rightarrow \quad = \binom{n}{k} \cdot B(k+1, n-k+1)$

$= \dfrac{n!}{k!(n-k)!} \cdot \dfrac{(k+1)!\,(n-k+1)!}{(n+2)!} = \dfrac{(k+1)(n-k+1)}{(n+1)(n+2)}$

(b) $\pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$

$p(y|\theta) \propto \theta^{y}(1-\theta)^{n-y}$

$\pi(\theta|y) \propto \underbrace{\theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}}$

↳ It's kernel of $Beta(\alpha+y, \beta+n-y)$

with $E(\theta|y) = \dfrac{\alpha+y}{\alpha+\beta+n}$

$= \dfrac{\alpha}{\alpha+\beta} \cdot \dfrac{\alpha+\beta}{\alpha+\beta+n} + \dfrac{y}{n} \cdot \dfrac{n}{\alpha+\beta+n}$

$= \dfrac{\alpha}{\alpha+\beta} \cdot w_1 + \dfrac{y}{n} \cdot w_2$

with $w_1 + w_2 = 1$

$\Rightarrow E(\theta|y)$ is between $\dfrac{\alpha}{\alpha+\beta}$ and $\dfrac{y}{n}$.

(c) by (b): $\pi(\theta|y) \propto \theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}$, $\alpha=\beta=1$

$\Rightarrow \pi(\theta|y) \propto \theta^{y}(1-\theta)^{n-y}$, it's $Beta(y+1, n-y+1)$

$\Rightarrow Var(\theta|y) = \dfrac{(y+1)(n-y+1)}{(n+2)^2(n+3)}$, $Var(\theta) = \dfrac{1}{2^2 \times 3} = \dfrac{1}{12}$

with $n \geq 0$, $y \leq n$: $Var(\theta|y) = \dfrac{-y^2+ny+1}{(n+2)^2(n+3)}$

It's a quadratic function of y, which gets its maximum

when $y = \dfrac{n}{2}$ $\Rightarrow Var(\theta|y) \leq \dfrac{-\frac{n^2}{4} - \frac{n^2}{4} + 1}{(n+2)^2(n+3)} = \dfrac{1}{(n+2)^2(n+3)}$

But $n \geq 0$ $\Rightarrow Var(\theta|y) \leq \dfrac{1}{2^2 \times 3} = \dfrac{1}{12} = Var(\theta)$

(d) With $\pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$

$$\pi(\theta|y) \propto \theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}$$

$Var(\theta) = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$, $Var(\theta|y) = \dfrac{(\alpha+y)(\beta+n-y)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}$

$E(\theta) = \dfrac{\alpha}{\alpha+\beta}$, $E(\theta|y) = \dfrac{\alpha+y}{\alpha+\beta+n}$

If the prior is wrong and the data size is small, then I guess it could be the case. Also, the prior is "$\boxed{\text{wrong}}$".

ps: I encountered one case in stats 206 MCMC process, if the prior is very wrong but strong, it could make posterior have very large variance.

Set $\alpha = 100$, $\beta = 400$, $E(\theta) = 0.2$

Set $y = 10$, $n = 10$,

$Var(\theta) = \dfrac{100 \times 400}{500^2 \times 501} \approx 3.2 \times 10^{-4}$

$Var(\theta|y) = \dfrac{110 \times 400}{(510)^2 \times 511} \approx 3.3 \times 10^{-4}$

→ It's the case.

ps: It means, prior has to enlarge the variance to fit the data, when the $E(\theta)$ is very wrong.

3. BDA3 Problem 2.11

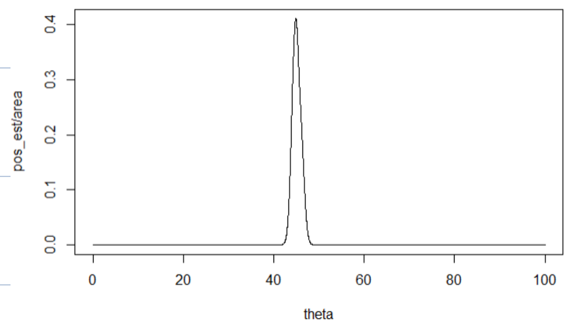
11. Computing with a nonconjugate single-parameter model: suppose $y_1, \ldots, y_5$ are independent samples from a Cauchy distribution with unknown center θ and known scale 1: $p(y_i|\theta) \propto 1/(1 + (y_i - \theta)^2)$. Assume, for simplicity, that the prior distribution for θ is uniform on $[0, 100]$. Given the observations $(y_1, \ldots, y_5) = (43, 44, 45, 46.5, 47.5)$:

    (a) Compute the unnormalized posterior density function, $p(\theta)p(y|\theta)$, on a grid of points $\theta = 0, \frac{1}{m}, \frac{2}{m}, \ldots, 100$, for some large integer m. Using the grid approximation, compute and plot the normalized posterior density function, $p(\theta|y)$, as a function of θ.

    (b) Sample 1000 draws of θ from the posterior density and plot a histogram of the draws.

    (c) Use the 1000 samples of θ to obtain 1000 samples from the predictive distribution of a future observation, y_6, and plot a histogram of the predictive draws.

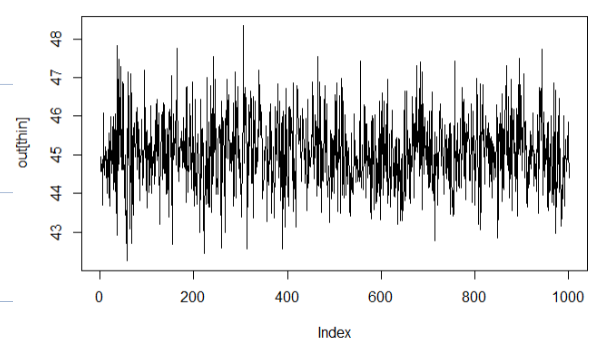
(a) $\pi(\theta|y) \propto \pi(\theta) \cdot f(y|\theta) = \frac{1}{100} \times \prod_{i=1}^{5} \frac{1}{1+(y_i-\theta)^2} \cdot \mathbb{I}(\theta \in [0,100])$



ps: Code is in another file.

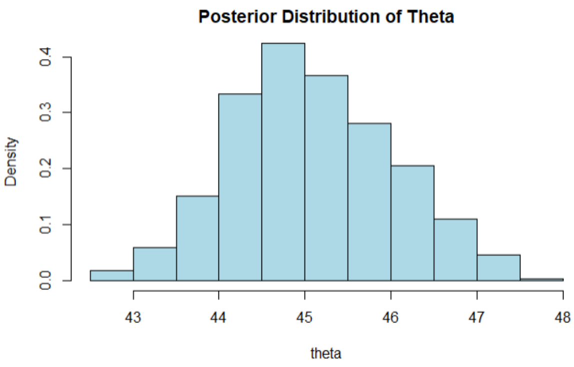
After normalizing, the pictures are like this

(b). I will consider to use Metropolis-Hasting with a random walk transit kernel. and the standard deviation of the Gaussian distribution is 2.



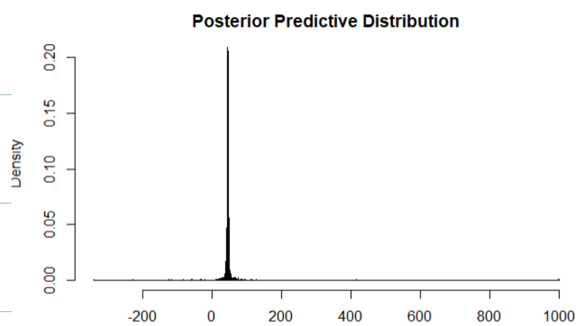
The chain is mixing well.

[After burn-in and thinning]



Here is the posterior distribution of the theta.

(c)



Here is the posterior predictive distribution.

ps: It's a t distribution so it has heavy tails, extreme values will be generated easily so the picture looks ugly but we can still recognize the shape, approximately.

And all codes are attached in the othe file.

4. Consider a distribution with p.d.f

$$p(x, y) \propto \mathbf{1}(|x - y| < c)\mathbf{1}(x, y \in (0, 1))$$

(a) Derive a Gibbs sampler to sample from this distribution.

(b) Implement the Gibbs sampler for 1000 iterations for each of the following: $c = 0.3, c = 0.05$ and $c = 0.01$.

(c) For each value of the c make a traceplot of x and a scatter plot of (x, y).

(d) What do you see as c gets smaller?

(a) Step 1: Get the full conditional distribution:

$$f(x|y) = \frac{f(x,y)}{\int f(x,y)\,dx} \qquad \text{since } y \text{ is given}$$

$\Rightarrow f(x|y) \propto f(x,y) \propto I(|x-y|<c)\, I(x,y \in (0,1))$

Given y: $f(x|y) \propto I(y-c < x < y+c)$ and $x,y \in (0,1)$

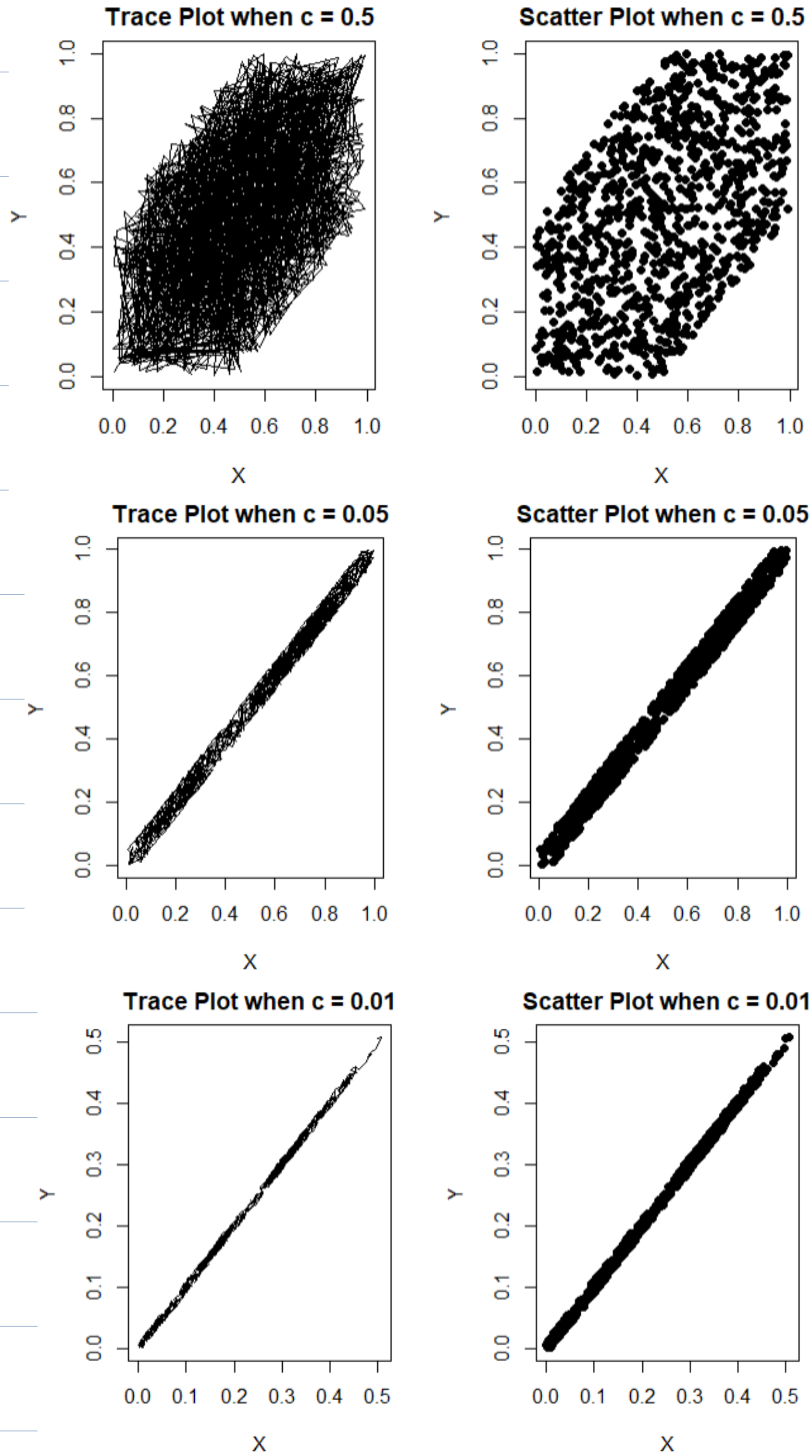
It's a $\text{Unif}(\max(0, y-c), \min(1, y+c))$

Similarly, $f(y|x)$ is a $\text{Unif}(\max(0, y-c), \min(1, y+c))$

Since we are asked to make a trace plot, I think I'm not going to do the burn-in and thinning because I need to reveal how the chain converges.

(b) It's in the code file. We continuously sample $y|x$ and $x|y$ to get the sample

(c)



(d) As c becomes smaller, x and y are more concentrated around the line that $y = x$.

Because the pdf is $I(|x-y| < c)$ so as $c \downarrow$, it forces x to be closer and closer to y.

5. BDA3 Problem 3.1

1. Binomial and multinomial models: suppose data $(y_1, \ldots, y_J)$ follow a multinomial distribution with parameters $(\theta_1, \ldots, \theta_J)$. Also suppose that $\theta = (\theta_1, \ldots, \theta_J)$ has a Dirichlet prior distribution. Let $\alpha = \frac{\theta_1}{\theta_1+\theta_2}$.
   (a) Write the marginal posterior distribution for α.
   (b) Show that this distribution is identical to the posterior distribution for α obtained by treating y_1 as an observation from the binomial distribution with probability α and sample size $y_1 + y_2$, ignoring the data $y_3, \ldots, y_J$.

   This result justifies the application of the binomial distribution to multinomial problems when we are only interested in two of the categories; for example, see the next problem.

(a) $f(y|\theta) \propto \prod_{i=1}^{J} \theta_i^{y_i}$

$\pi(\theta) \propto \prod_{i=1}^{J} \theta_i^{\alpha_i - 1}$

$\pi(\theta|y) \propto f(y|\theta)\cdot\pi(\theta) \propto \prod_{i=1}^{J} \theta_i^{\alpha_i + y_i - 1}$

And $\sum_{i=1}^{J} \theta_i = 1$. We can regard:

$\theta_i = \dfrac{\beta_i}{\sum_{i=1}^{J} \beta_i}$, in which $\beta_i \overset{ind.}{\sim} \text{Gamma}(\alpha_i + y_i, \gamma)$

$$\alpha = \frac{\frac{\beta_1}{\sum_{i=1}^{J}\beta_i}}{\frac{\beta_1}{\sum_{i=1}^{J}\beta_i} + \frac{\beta_2}{\sum_{i=1}^{J}\beta_i}} = \frac{\beta_1}{\beta_1 + \beta_2} \sim \text{Beta}(\alpha_1 + y_1, \alpha_2 + y_2)$$

$\Rightarrow p(\alpha|y) \propto \alpha^{\alpha_1 + y_1 - 1}(1-\alpha)^{\alpha_2 + y_2 - 1}$

(b). Similar to a, $\pi(\alpha) \sim \text{Beta}(\alpha_1, \alpha_2) \propto \alpha^{\alpha_1 - 1}(1-\alpha)^{\alpha_2 - 1}$

$f(y|\alpha) \propto \alpha^{y}(1-\alpha)^{N-y}$, let $y = y_1$, $N = y_1 + y_2$

$\pi(\alpha|y) \propto \alpha^{\alpha_1 + y_1 - 1}(1-\alpha)^{\alpha_2 + y_2 - 1}$ same as (a)

6. BDA3 Problem 3.2

2. Comparison of two multinomial observations: on September 25, 1988, the evening of a presidential campaign debate, ABC News conducted a survey of registered voters in the United States; 639 persons were polled before the debate, and 639 different persons were polled after. The results are displayed in Table 3.2. Assume the surveys are independent simple random samples from the population of registered voters. Model the data with two different multinomial distributions. For $j = 1, 2$, let α_j be the proportion of voters who preferred Bush, out of those who had a preference for either Bush or Dukakis at the time of survey j. Plot a histogram of the posterior density for $\alpha_2 - \alpha_1$. What is the posterior probability that there was a shift toward Bush?

Survey	Bush	Dukakis	No opinion/other	Total
pre-debate	294	307	38	639
post-debate	288	332	19	639

I will choose the prior $\pi(\theta_1, \theta_2, \theta_3) \propto \theta_1^{2-1}\, \theta_2^{2-1}\, \theta_3^{2-1}$ for both $j=1$ and $j=2$.

$J=1$: $\pi_1(\theta|y_1) \propto \theta_1^{294+2-1}\, \theta_2^{307+2-1}\, \theta_3^{38+2-1}$

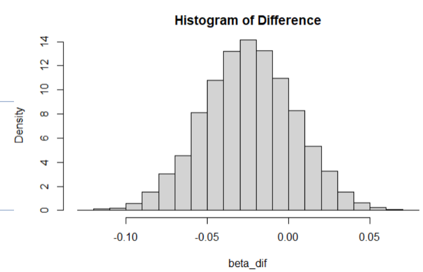
$\propto \theta_1^{295}\, \theta_2^{308}\, \theta_3^{39}$

By conclusion from the last question: $\alpha_1 \sim Beta(296, 309)$

$J=2$: $\pi_2(\theta|y_2) \propto \theta_1^{288+2-1}\, \theta_2^{332+2-1}\, \theta_3^{19+2-1}$

Similarly: $\alpha_2 \sim Beta(290, 334)$.

I will generate 10000 α_1, 10000 α_2, then use Monte Carlo simulation to calculate the probability.



The probability that beta_diff bigger than 0 is 0.1938

7. BDA3 Problem 3.5

5. Rounded data: it is a common problem for measurements to be observed in rounded form (for a review, see Heitjan, 1989). For a simple example, suppose we weigh an object five times and measure weights, rounded to the nearest pound, of 10, 10, 12, 11, 9. Assume the unrounded measurements are normally distributed with a noninformative prior distribution on the mean μ and variance σ^2.

   (a) Give the posterior distribution for (μ, σ^2) obtained by pretending that the observations are exact unrounded measurements.

   (b) Give the correct posterior distribution for (μ, σ^2) treating the measurements as rounded.

   (c) How do the incorrect and correct posterior distributions differ? Compare means, variances, and contour plots.

   (d) Let $z = (z_1, \ldots, z_5)$ be the original, unrounded measurements corresponding to the five observations above. Draw simulations from the posterior distribution of z. Compute the posterior mean of $(z_1 - z_2)^2$.

(a) Assume $\pi(\mu|\sigma^2) \propto 1 \quad \pi(\sigma^2) \propto 1/\sigma^2$

$$f(y|\mu,\sigma^2) \propto \prod_{i=1}^{5} (\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right)$$

$$\pi(\mu,\sigma^2|y) \propto (\sigma^2)^{-\frac{5}{2}-1} \exp\left(-\frac{\sum_{i=1}^{5}(y_i-\mu)^2}{2\sigma^2}\right)$$

(b) $$f(y_i|\mu,\sigma^2) = \propto \int_{y_i-0.5}^{y_i+0.5} (\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

$$f(\mu,\sigma^2|y) \propto (\sigma^2)^{-1} \prod_{i=1}^{5} \int_{y_i-0.5}^{y_i+0.5} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt$$

(c) for (a) case: I will use a Gibbs sampler.

$$\sigma^2|\mu \sim IG\left(\frac{5}{2}, \frac{\sum_{i=1}^{5}(y_i-\mu)^2}{2}\right)$$

$$P(\mu|\sigma^2) \propto \exp\left(-\frac{5\mu^2-2\sum_i y_i\mu}{2\sigma^2}\right) \propto \exp\left(-\frac{(\mu-\bar{y})^2}{2\times\sigma^2/5}\right) \sim N\left(\bar{y}, \frac{1}{5}\sigma^2\right)$$

for (b) case, I will use random walk proposal for μ and independent MH for σ^2.

Not considering rounding effects:

Not Considering Rounding Effect – Gibbs Sampler

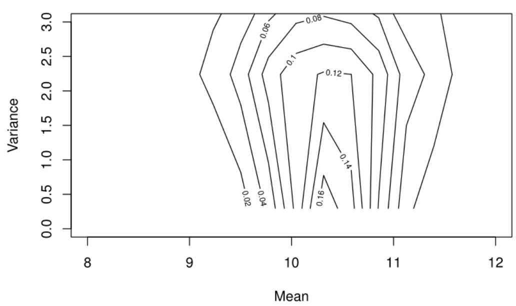


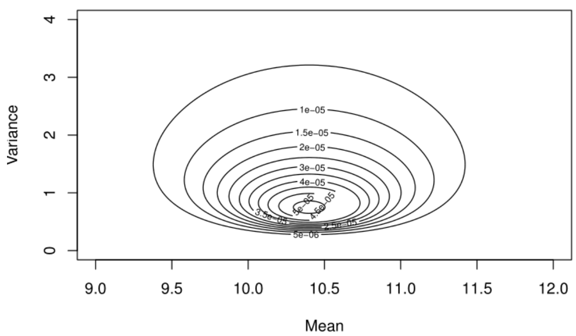
Table 1: Not Considering Rounding Effect - Gibbs Sampler

	2.5% Lower	Mean	97.5% Upper	Variance
mu	8.91	10.36	11.76	0.47
sigma	0.48	2.46	10.49	10.72

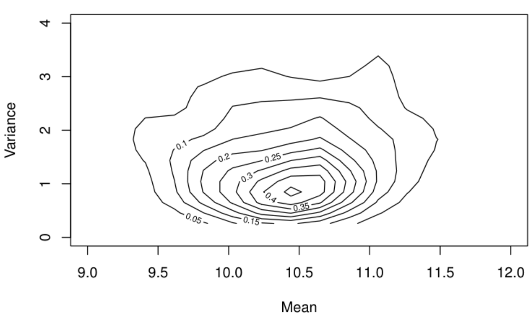
Table 2: Not Considering Rounding Effect - Grid Approximation

	2.5% Lower	Mean	97.5% Upper	Variance
mu	9.34	10.43	11.69	0.34
sigma	0.45	1.69	4.32	1.06

Not Considering the Rounding Effect – Grid Approximation



Not Considering the Rounding Effect – Empirical Sampling Distributi



I used two methods, Gibbs and grid approximation. Here is the output of these two olgrithms and some figures

Considering Rounding Effects:

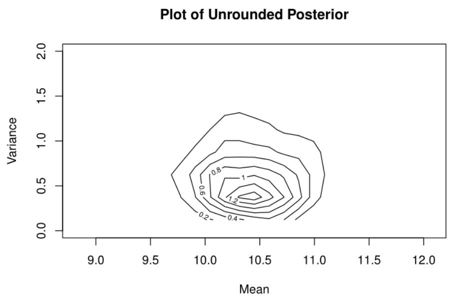
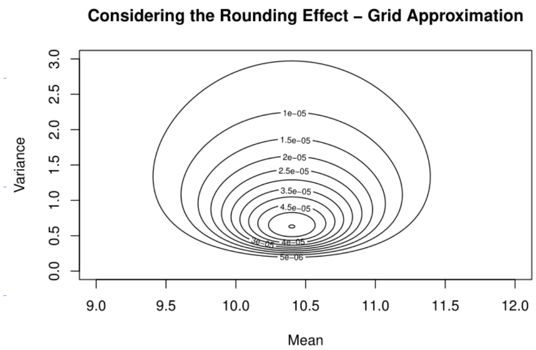


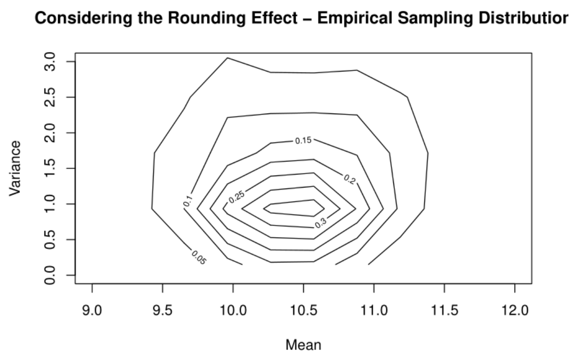
Table 3: Considering Rounding Effect - MH

	2.5% Lower	Mean	97.5% Upper	Variance
mu	9.59	10.40	11.21	0.17
sigma	0.18	0.73	2.13	0.31

Table 4: Considering Rounding Effect - Grid Approximation

	2.5% Lower	Mean	97.5% Upper	Variance
mu	9.22	10.42	11.71	0.43
sigma	0.38	2.14	9.19	5.43





Comparing with the last output, it's almost the same about the mean of μ and σ^2, but the variance of ~~both~~ both of them enlarges a little.

(d) They follows truncated normal distribution $N(10.4, 2.14)$. Lower $= y_i - 0.5$ upper $= y_i + 0.05$

Code is attached, the posterior mean of $(z_1 - z_2)^2$ is approximately 0.1609

8. BDA3 Problem 3.9

9. Conjugate normal model: suppose y is an independent and identically distributed sample of size n from the distribution $\mathrm{N}(\mu, \sigma^2)$, where the prior distribution for (μ, σ^2) is $\text{N-Inv-}\chi^2(\mu, \sigma^2|\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2)$; that is, $\sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$ and $\mu|\sigma^2 \sim \mathrm{N}(\mu_0, \sigma^2/\kappa_0)$. The posterior distribution, $p(\mu, \sigma^2|y)$, is also normal-inverse-χ^2; derive explicitly its parameters in terms of the prior parameters and the sufficient statistics of the data.

$$\pi(\mu|\sigma^2) \propto (\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{(\mu-\mu_0)^2}{2\sigma^2/\kappa_0}\right)$$

$$\pi(\sigma^2) \propto (\sigma^2)^{-\frac{\nu_0}{2}-1} \exp\left(-\frac{\sigma_0^2}{2\sigma^2}\right)$$

$$\pi(\mu,\sigma^2) \propto (\sigma^2)^{-\left(\frac{\nu_0+1}{2}\right)-1} \exp\left(-\frac{1}{2\sigma^2}\left(\frac{(\mu-\mu_0)^2}{\kappa_0}+\sigma_0^2\right)\right)$$

$$f(y|\mu,\sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (y_i-\mu)^2}{2\sigma^2}\right)$$

$$\begin{cases} s^2 = \sum_{i=1}^n (y_i-\bar{y})^2 \\ \bar{y} = \sum_{i=1}^n y_i/n \end{cases}$$

$$\pi(\mu,\sigma^2|y) \propto f(y|\mu,\sigma^2)\cdot\pi(\mu|\sigma^2)\cdot\pi(\sigma^2)$$

$$\propto (\sigma^2)^{-\frac{n+\nu_0+1}{2}-1} \exp\left(-\frac{s^2+n(\bar{y}-\mu)^2+(\mu-\mu_0)^2/\kappa_0+\sigma_0^2}{2\sigma^2}\right)$$

$$\pi(\mu|\sigma^2,y) \propto \pi(\mu,\sigma^2|y) \propto \exp\left(-\frac{n(\bar{y}-\mu)^2+(\mu-\mu_0)^2/\kappa}{2\sigma^2}\right)$$

$$\propto \exp\left(-\frac{(n+1/\kappa)\mu^2 - 2\mu\cdot(n\bar{y}+\mu_0/\kappa)}{2\sigma^2}\right)$$

$$\propto \exp\left(-\frac{\left(\mu - \frac{n\bar{y}+\mu_0/\kappa}{n+1/\kappa}\right)^2}{2\sigma^2/(n+1/\kappa)}\right)$$

It's a kernel of $N\left(\frac{n\bar{y}+\mu_0/\kappa}{n+1/\kappa}, \frac{\sigma^2}{n+1/\kappa}\right)$

$$\pi(\sigma^2|y) = \int_{-\infty}^{\infty} \pi(\mu,\sigma^2)\,d\mu \propto (\sigma^2)^{-\frac{n+\nu_0+1}{2}-1} \exp\left(-\frac{s^2+\sigma_0^2}{2\sigma^2}\right)\cdot$$

$$\underbrace{\int_{-\infty}^{\infty} \exp\left(-\frac{(n+1/\kappa)\mu^2-2\mu(n\bar{y}+\mu_0/\kappa)+n\bar{y}^2+\mu_0^2/\kappa}{2\sigma^2}\right)d\mu}_{\text{denote to } \phi}$$

$$\star = \exp\left(-\frac{n\bar{y}^2 + \mu_0^2/k_0}{2\sigma^2}\right)\int_{-\infty}^{\infty}\exp\left(-\frac{\left(\mu-\frac{n\bar{y}+\mu_0/k_0}{n+1/k_0}\right)^2-\left(\frac{n\bar{y}+\mu_0/k_0}{n+1/k_0}\right)^2}{2\sigma^2/(n+1/k_0)}\right)d\mu$$

$$\propto \underbrace{\exp\left(-\frac{n\bar{y}^2+\mu_0^2/k_0-\frac{(n\bar{y}+\mu_0/k_0)^2}{n+1/k_0}}{2\sigma^2}\right)}_{\text{denote to }\star\star}\cdot\left(\frac{\sigma^2}{n+1/k}\right)^{\frac{1}{2}}$$

$$\star\star = \exp\left(-\frac{\cancel{n^2\bar{y}^2}+\frac{n}{k_0}\bar{y}^2+\frac{n\mu_0^2}{k_0}+\cancel{\frac{\mu_0^2}{k_0^2}}-\cancel{n^2\bar{y}^2}-\frac{2n\bar{y}\mu_0}{k_0}-\cancel{\frac{\mu_0^2}{k_0^2}}}{2\sigma^2(n+1/k_0)}\right)$$

$$= \exp\left(-\frac{n/k_0\left(\mu_0^2-2\bar{y}\mu_0+\bar{y}^2\right)}{2\sigma^2(n+1/k_0)}\right)$$

$$= \exp\left(-\frac{(\mu_0-\bar{y})^2}{2\sigma^2(k_0+\frac{1}{n})}\right)$$

$$\Rightarrow$$

$$\pi(\sigma^2|y)\propto(\sigma^2)^{-\frac{n+\nu_0+1}{2}-1}(\sigma^2)^{\frac{1}{2}}\exp\left(-\frac{1}{\sigma^2}\left(\frac{s^2+\sigma_0^2+\frac{(\mu_0-\bar{y})^2}{k_0+1/n}}{2}\right)\right)$$

$$\sim \text{Inv-}\chi^2\left(n+\nu_0,\ s^2+\sigma_0^2+\frac{(\mu_0-\bar{y})^2}{k_0+1/n}\right)$$

9. Implement the analysis of the Slovenia opinion poll data (from lecture 2) to obtain the fraction of voters who will both attend the plebiscite AND vote in favor of independence. You can assume missing at random for the missing responses as in the lecture slides.

   (a) Write down the statistical model, i.e., the assumed data generating process and priors.

   (b) Write down how you will sample from the posterior and what parameters of the model is of interest.

   (c) Implement your sampler and plot the tarceplot and posterior marginal distribution of the parameter of interest.

Attendance	Yes	No	DK
Yes	1,439	78	159
No	16	16	32
DK	144	54	136

(a) The true value of the people who attend and responded "Yes"

$Y_{11} = x_{11} + x_{1\cdot}^{(11)} + x_{\cdot 1}^{(11)} + x_{\cdot\cdot}^{(11)}$. Same for other Y_{ij}'s.

$x_{1\cdot}^{(11)}, x_{1\cdot}^{(10)} \mid x_{1\cdot} \sim Mult\left(x_{1\cdot}, \frac{\theta_{11}}{\theta_{11}+\theta_{10}}, \frac{\theta_{10}}{\theta_{11}+\theta_{10}}\right)$ $\quad S = \theta_{11}+\theta_{10}+\theta_{01}+\theta_{00}$

$x_{\cdot 1}^{(11)}, x_{\cdot 1}^{(01)} \mid x_{\cdot 1} \sim Mult\left(x_{\cdot 1}, \frac{\theta_{11}}{\theta_{11}+\theta_{01}}, \frac{\theta_{01}}{\theta_{11}+\theta_{01}}\right)$ $\quad \uparrow$

$x_{\cdot\cdot}^{(11)}, x_{\cdot\cdot}^{(10)}, x_{\cdot\cdot}^{(01)}, x_{\cdot\cdot}^{(00)} \mid x_{\cdot\cdot} \sim Mult\left(x_{\cdot\cdot}, \frac{\theta_{11}}{S}, \frac{\theta_{10}}{S}, \frac{\theta_{01}}{S}, \frac{\theta_{00}}{S}\right)$

$\pi(\vec{\theta}) \sim$ Dirichlet $(1, 1, 1, 1)$

$f(\vec{Y} \mid \vec{\theta}) \sim Mult(N, \theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})$

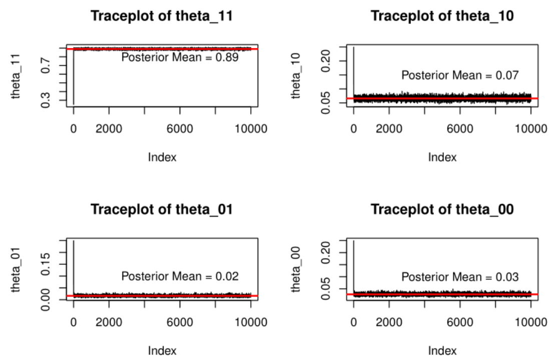
(b) Step 1: Initialize $\theta^{(0)} = (\theta_{11}^{(0)}, \theta_{10}^{(0)}, \theta_{01}^{(0)}, \theta_{00}^{(0)})$

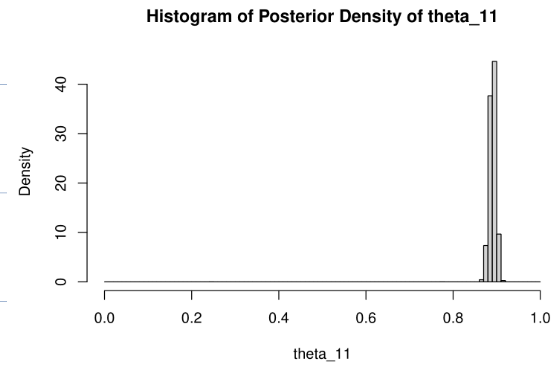
Step 2: Generate $x_{1\cdot}^{(11)}, x_{\cdot 1}^{(11)}, x_{\cdot\cdot}^{(11)}$ based on $x_{1\cdot}, x_{\cdot 1}, x_{\cdot\cdot}$ and $\vec{\theta}$

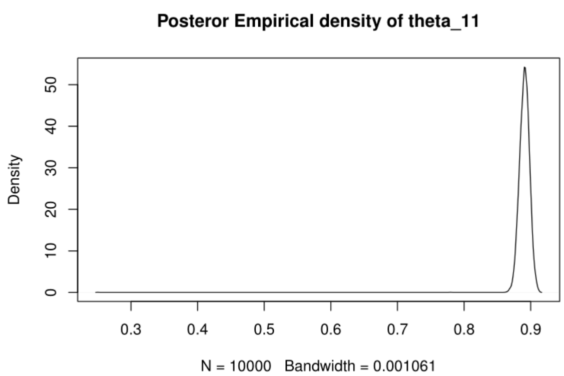
Step 3: $\pi(\theta \mid y) \sim$ Dirichlet $(y_{11}+1, y_{10}+1, y_{01}+1, y_{00}+1)$

Step 4: Repeatedly do step 2 and step 3.

(c)







ps: All the codes are in another file.