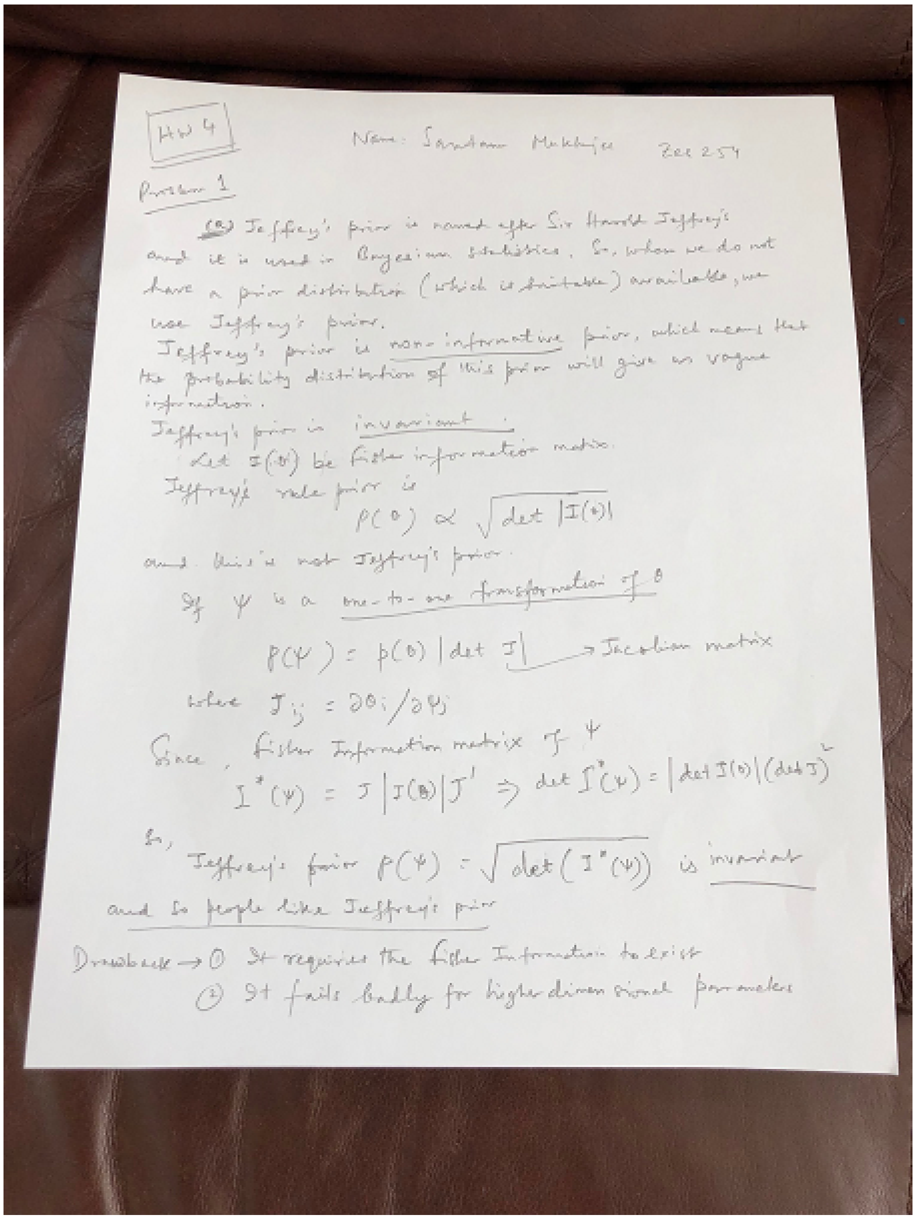
HW4 R Markdown

Santanu Mukherjee

11/7/2021

R Markdown

Answer 1a



HW 4

Name: Santanu Mukherjee Zee254

Problem 1

(a) Jeffrey's prior is named after Sir Harold Jeffrey's and it is used in Bayesian statistics. So, when we do not have a prior distribution (which is suitable) available, we use Jeffrey's prior.

Jeffrey's prior is non-informative prior, which means that the probability distribution of this prior will give us vague information.

Jeffrey's prior is invariant.

Let $I(\theta)$ be fisher information matrix.

Jeffrey's rule prior is

$$p(\theta) \propto \sqrt{\det |I(\theta)|}$$

and this is not Jeffrey's prior.

If ψ is a one-to-one transformation of θ

$$p(\psi) = p(\theta)\,|\det J| \rightarrow \text{Jacobian matrix}$$

where $J_{ij} = \partial\theta_i / \partial\psi_j$

Since, fisher Information matrix of ψ

$$I^*(\psi) = J\,|I(\theta)|\,J' \Rightarrow \det I^*(\psi) = |\det I(\theta)|(\det J)^2$$

So,

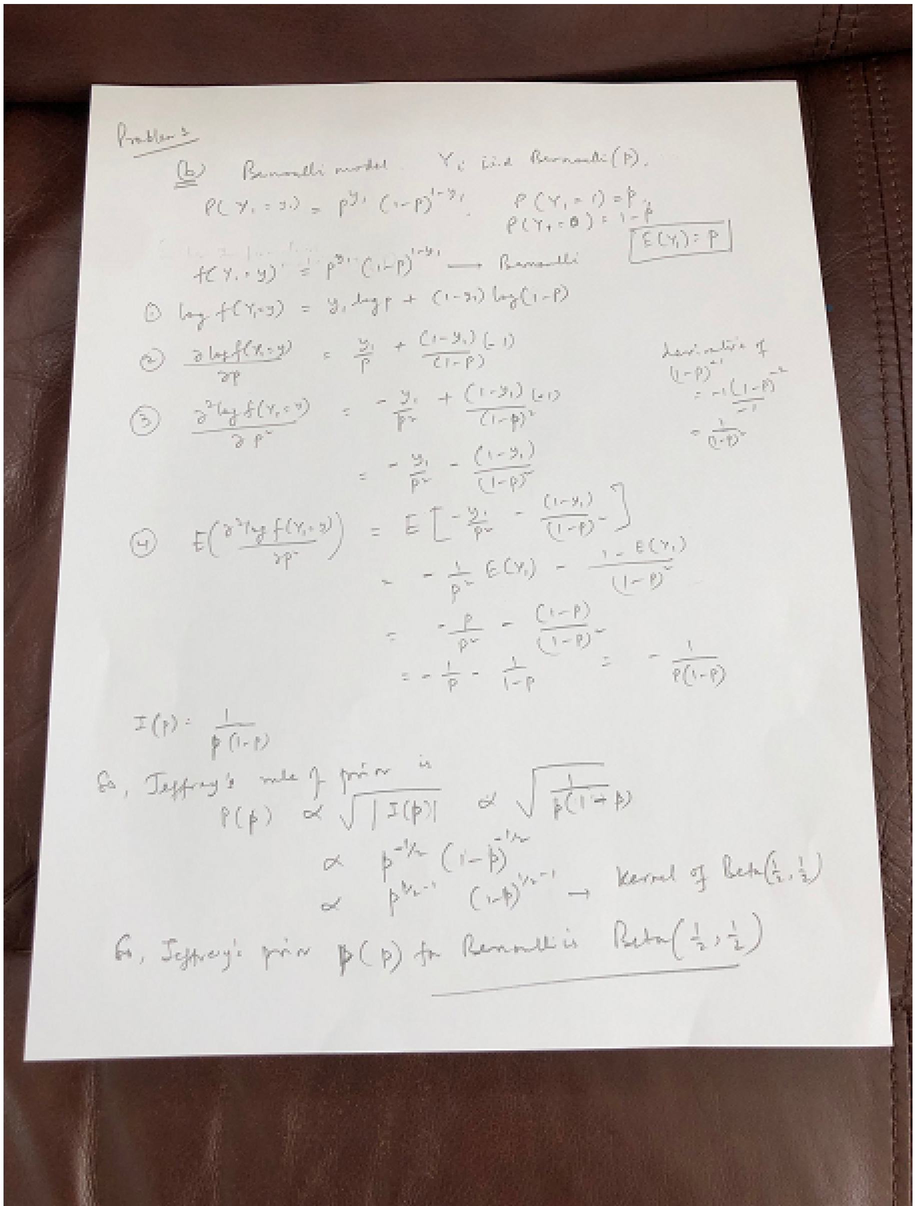
Jeffrey's prior $p(\psi) = \sqrt{\det(I^*(\psi))}$ is invariant and so people like Jeffrey's prior

Drawback → (1) It requires the fisher Information to exist

(2) It fails badly for higher dimensional parameters

Answer-1-a

Answer 1 b



Problem 3

(b) Bernoulli model Y_i iid Bernoulli(p).

$P(Y_i = y_i) = p^{y_i}(1-p)^{1-y_i}$; $P(Y_i = 1) = p$, $P(Y_i = 0) = 1-p$

$\boxed{E(Y_i) = p}$

$f(Y_i, y) = p^{y_i}(1-p)^{1-y_i} \longrightarrow$ Bernoulli

① $\log f(Y_i, y) = y_i \log p + (1-y_i)\log(1-p)$

② $\dfrac{\partial \log f(Y_i, y)}{\partial p} = \dfrac{y_i}{p} + \dfrac{(1-y_i)}{(1-p)}(-1)$

derivative of $(1-p)^{-1} = -1(1-p)^{-2} \cdot (-1) = \dfrac{1}{(1-p)^2}$

③ $\dfrac{\partial^2 \log f(Y_i, y)}{\partial p^2} = -\dfrac{y_i}{p^2} + \dfrac{(1-y_i)}{(1-p)^2}(-1)$

$= -\dfrac{y_i}{p^2} - \dfrac{(1-y_i)}{(1-p)^2}$

④ $E\left(\dfrac{\partial^2 \log f(Y_i, y)}{\partial p^2}\right) = E\left[-\dfrac{y_i}{p^2} - \dfrac{(1-y_i)}{(1-p)^2}\right]$

$= -\dfrac{1}{p^2}E(Y_i) - \dfrac{1-E(Y_i)}{(1-p)^2}$

$= -\dfrac{p}{p^2} - \dfrac{(1-p)}{(1-p)^2}$

$= -\dfrac{1}{p} - \dfrac{1}{1-p} = -\dfrac{1}{p(1-p)}$

$I(p) = \dfrac{1}{p(1-p)}$

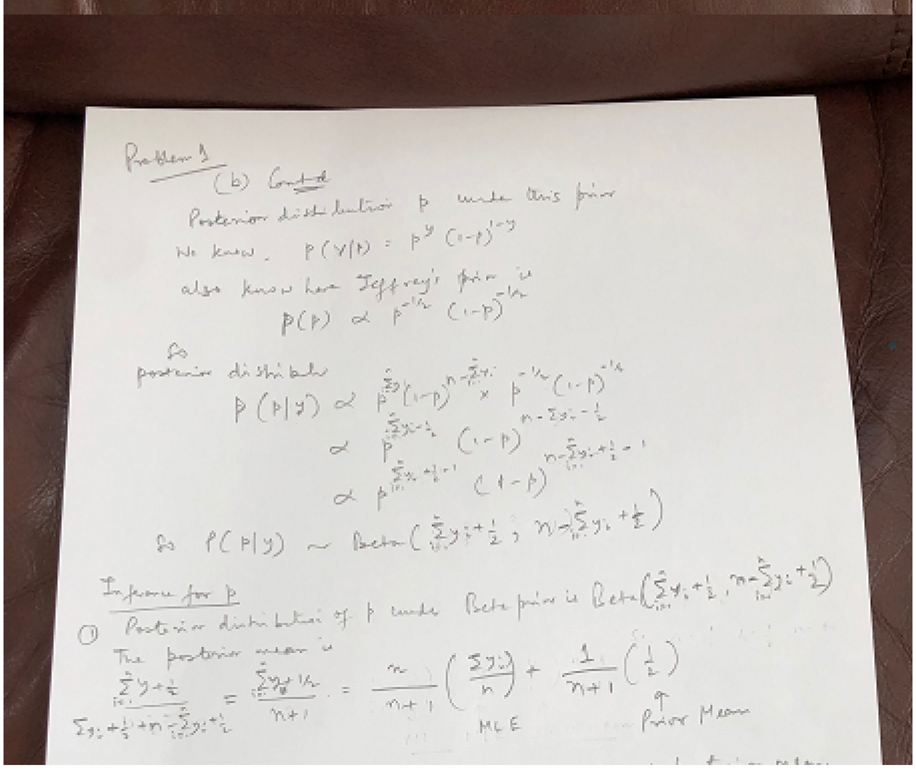
So, Jeffrey's rule of prior is

$P(p) \propto \sqrt{|I(p)|} \propto \sqrt{\dfrac{1}{p(1-p)}}$

$\propto p^{-1/2}(1-p)^{-1/2}$

$\propto p^{1/2-1}(1-p)^{1/2-1} \rightarrow$ Kernel of Beta$(\frac{1}{2}, \frac{1}{2})$

So, Jeffrey's prior $p(p)$ for Bernoulli is Beta$\left(\frac{1}{2}, \frac{1}{2}\right)$



Problem 3

(b) Contd

Posterior distribution p under this prior

We know, $P(Y|p) = p^y(1-p)^{1-y}$

also know here Jeffrey's prior is

$p(p) \propto p^{-1/2}(1-p)^{-1/2}$

So

posterior distribution

$p(p|y) \propto p^{\sum_{i=1}^n y_i}(1-p)^{n-\sum_{i=1}^n y_i} \times p^{-1/2}(1-p)^{-1/2}$

$\propto p^{\sum_{i=1}^n y_i - \frac{1}{2}}(1-p)^{n-\sum y_i - \frac{1}{2}}$

$\propto p^{\sum_{i=1}^n y_i + \frac{1}{2} - 1}(1-p)^{n-\sum_{i=1}^n y_i + \frac{1}{2} - 1}$

So $P(p|y) \sim$ Beta$\left(\sum_{i=1}^n y_i + \frac{1}{2}, n - \sum_{i=1}^n y_i + \frac{1}{2}\right)$

Inference for p

① Posterior distribution of p under Beta prior is Beta$\left(\sum_{i=1}^n y_i + \frac{1}{2}, n - \sum_{i=1}^n y_i + \frac{1}{2}\right)$

The posterior mean is

$\dfrac{\sum_{i=1}^n y + \frac{1}{2}}{\sum y_i + \frac{1}{2} + n - \sum_{i=1}^n y_i + \frac{1}{2}} = \dfrac{\sum_{i=1}^n y_i + 1/2}{n+1} = \dfrac{n}{n+1}\left(\dfrac{\sum y_i}{n}\right) + \dfrac{1}{n+1}\left(\dfrac{1}{2}\right)$

MLE

Prior Mean

So, posterior mean is weighted average of MLE and posterior [illegible]

(2) also in this case Jeffrey's prior is conjugate prior.

(3) We can also find expected value of p

$$E(p|Y) = \frac{\sum y_i + \frac{1}{2}}{n+1}.$$

Answer 1 c

Problem 1 (c) Poisson distribution, y_i, iid Poisson λ.

Jeffrey's prior would be $\sqrt{|I(\theta)|} = \sqrt{-E\left[\frac{\partial^2}{\partial\theta^2} \log f(x|\theta)|\theta\right]}$

$$f(y_n|\lambda) = e^{-n\lambda} \lambda^{\sum y_i} \times \prod_{i=1}^{n} \frac{1}{(y_i)!}$$

$$\log f(x_n|\lambda) = \sum y_i \log \lambda - n\lambda - \log \sum_{i=1}^{n} (y_i!)$$

(1) $\frac{\partial \log f(x_n|\lambda)}{\partial \lambda} = \frac{\sum y_i}{\lambda} - n$

(2) $\frac{\partial^2 \log f(x_n|\lambda)}{\partial \lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2}$

(3) $E\left[\frac{\partial^2 \log f(x|\lambda)}{\partial \lambda^2}\right] = \frac{-n\lambda}{\lambda^2} = -\frac{n}{\lambda}.$

Now $\sqrt{I(\lambda)} = \sqrt{-E\left[\frac{\partial^2 \log f(x_n|\lambda)}{\partial\lambda^2}\Big|\lambda\right]} = \sqrt{\frac{n}{\lambda}}$

$\propto \frac{1}{\sqrt{\lambda}}$

$\propto \lambda^{-1/2}$ → not a proper prior as distribution is not recognizable

→ But Jeffrey's prior can viewed as $\lim_{k \to 0}$ Gamma $(\frac{1}{2}, k)$

Likelihood function: $p(y|\lambda) \propto e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}$

So posterior distribution

$$p(\lambda|y) \propto p(y|\lambda)\, \lambda^{-1/2}$$
$$\propto e^{-n\lambda} \lambda^{\sum y_i} \lambda^{-1/2}$$
$$\propto \lambda^{\sum y_i + \frac{1}{2} - 1} e^{-n\lambda}$$

which is the kernel of Gamma funct.

So posterior distribution of $p(\lambda|y) \propto$ Gamma $\left(\sum_{i=1}^{n} y_i + \frac{1}{2}, n\right)$

for inference, we can say $E(\lambda|Y) = \frac{\sum_{i=1}^{n} y_i + \frac{1}{2}}{n}.$

Answer-1-c

Answer 1 d

Problem 1

(d) Exponential Distribution. y_i iid Exponential(λ)

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

$$\log f(y_n|\lambda) = n \log \lambda - \lambda \sum y_i$$

① $$\frac{\partial \log f(y_n|\lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum y_i$$

② $$\frac{\partial^2 \log f(y_n|\lambda)}{\partial \lambda^2} = -\frac{n}{\lambda^2}$$

③ $$E\left[\frac{\partial^2 \log f(y_n|\lambda)}{\partial \lambda^2}\right] = -\frac{n}{\lambda^2}$$

Now, $$\sqrt{|I(\lambda)|} = \sqrt{-E\left[\frac{\partial^2 \log f(y_n|\lambda)}{\partial \lambda^2} \Big| \lambda\right]} = \sqrt{\frac{n}{\lambda^2}}$$

$$\propto \frac{1}{\lambda}$$

$$\propto \lambda^{-1} \longrightarrow \text{unrecognizable and not proper prior}$$

But, we can say distribution of Jeffrey's prior is $\lim_{k \to 0} \text{Gamma}(k, k)$

Posterior distribution

$$p(\lambda|y_n) \propto p(y_n|\lambda) \cdot \lambda^{-1}$$

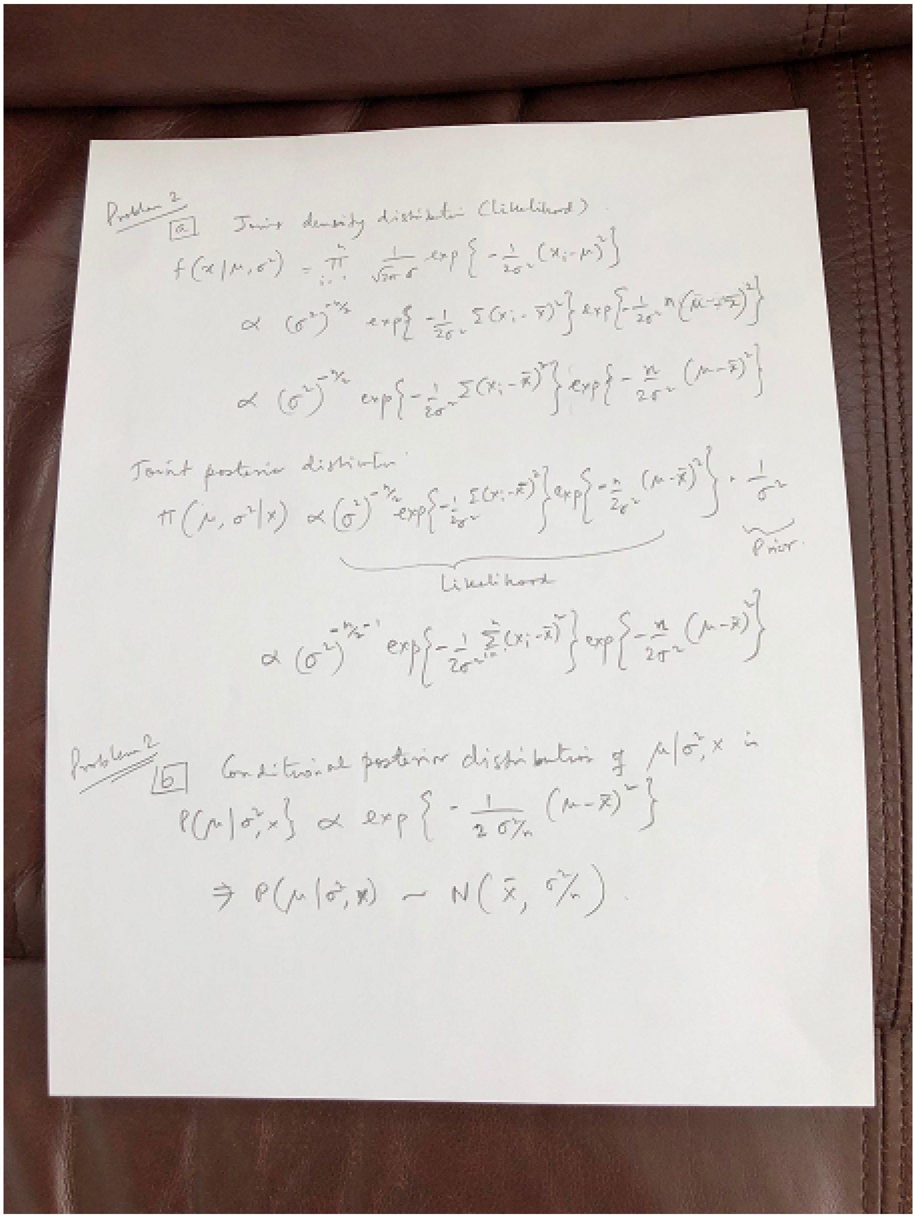
$$\propto \lambda^{n-1} e^{-\lambda \sum y_i}$$

which is the kernel of Gamma distribution $\text{Gamma}(n, \sum_{i=1}^{n} y_i)$

for inference, we can say $E(\lambda|y) = \frac{n}{\sum_{i=1}^{n} y_i} = \frac{1}{\bar{y}}$.

Answer-1-d

Answer 2 a and 2 b



Problem 2

$\boxed{a}$ Joint density distribution (Likelihood)

$$f(x|\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x_i-\mu)^2\right\}$$

$$\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\Sigma(x_i-\bar{x})^2\right\} \exp\left\{-\frac{1}{2\sigma^2} n(\mu-\bar{x})^2\right\}$$

$$\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\Sigma(x_i-\bar{x})^2\right\} \exp\left\{-\frac{n}{2\sigma^2}(\mu-\bar{x})^2\right\}$$

Joint posterior distribution

$$\pi(\mu,\sigma^2|x) \propto \underbrace{(\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\Sigma(x_i-\bar{x})^2\right\} \exp\left\{-\frac{n}{2\sigma^2}(\mu-\bar{x})^2\right\}}_{\text{Likelihood}} \cdot \underbrace{\frac{1}{\sigma^2}}_{\text{Prior}}$$

$$\propto (\sigma^2)^{-n/2-1} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right\} \exp\left\{-\frac{n}{2\sigma^2}(\mu-\bar{x})^2\right\}$$

Problem 2

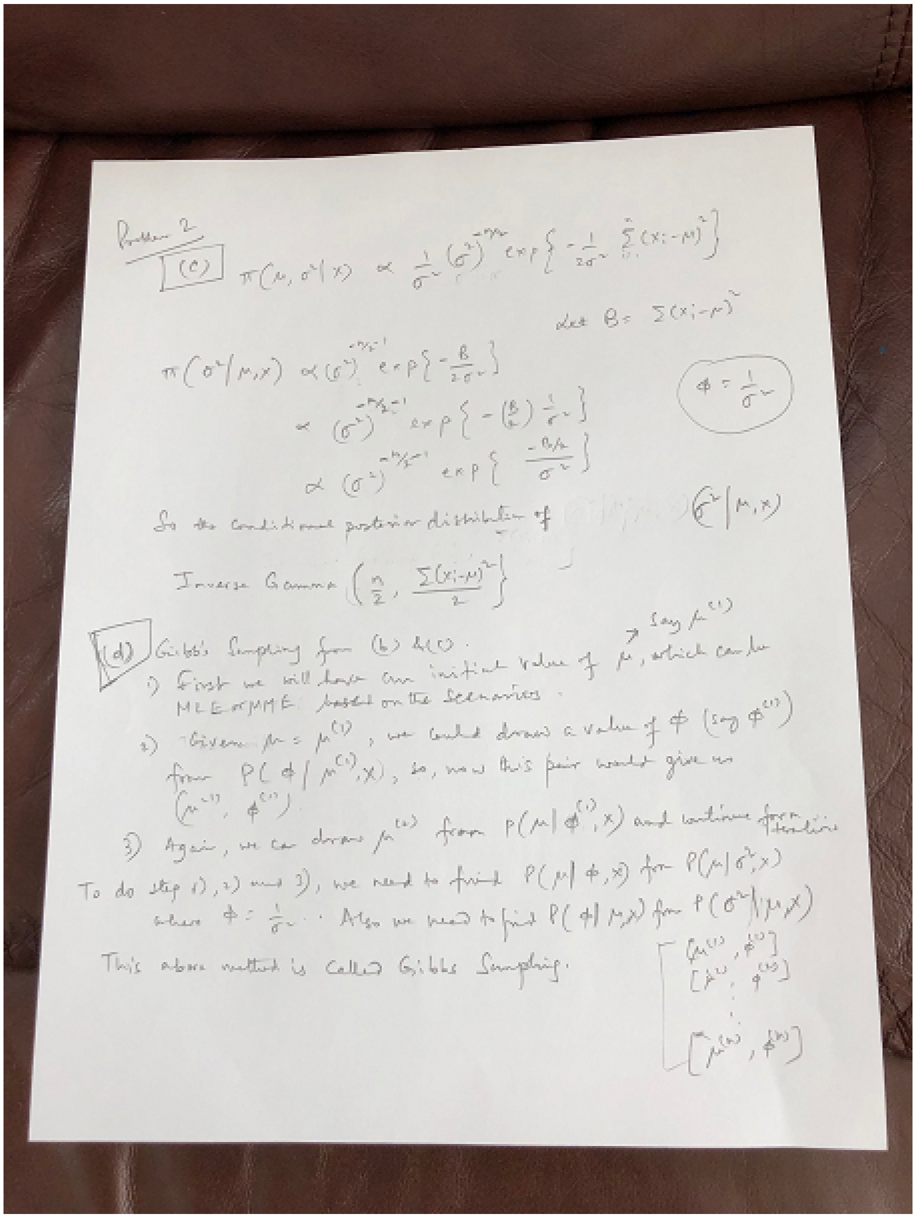
$\boxed{b}$ Conditional posterior distribution of $\mu|\sigma^2,x$ is

$$P(\mu|\sigma^2,x\} \propto \exp\left\{-\frac{1}{2\,\sigma^2/n}(\mu-\bar{x})^2\right\}$$

$$\Rightarrow P(\mu|\sigma^2,x) \sim N(\bar{x}, \sigma^2/n)$$

Answer-2-a-b

Answer 2 c and 2 d



Problem 2

(c) $\pi(\mu, \sigma^2 | x) \propto \frac{1}{\sigma^2} (\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$

Let $B = \sum (x_i - \mu)^2$

$\pi(\sigma^2 | \mu, x) \propto (\sigma^2)^{-n/2 - 1} \exp\left\{ -\frac{B}{2\sigma^2} \right\}$

$\propto (\sigma^2)^{-n/2 - 1} \exp\left\{ -\left(\frac{B}{2}\right) \frac{1}{\sigma^2} \right\}$

$\propto (\sigma^2)^{-n/2 - 1} \exp\left\{ \frac{-B/2}{\sigma^2} \right\}$

$\phi = \frac{1}{\sigma^2}$

So the conditional posterior distribution of $\sigma^2 | \mu, x$

Inverse Gamma $\left(\frac{n}{2}, \frac{\sum (x_i - \mu)^2}{2} \right)$

(d) Gibb's Sampling from (b) & (c).

1) First we will have an initial value of μ → say $\mu^{(1)}$, which can be MLE or MME based on the scenarios.

2) Given $\mu = \mu^{(1)}$, we could draw a value of ϕ (say $\phi^{(1)}$) from $P(\phi | \mu^{(1)}, x)$, so, now this pair would give us $(\mu^{(1)}, \phi^{(1)})$

3) Again, we can draw $\mu^{(2)}$ from $P(\mu | \phi^{(1)}, x)$ and continue for n iterations

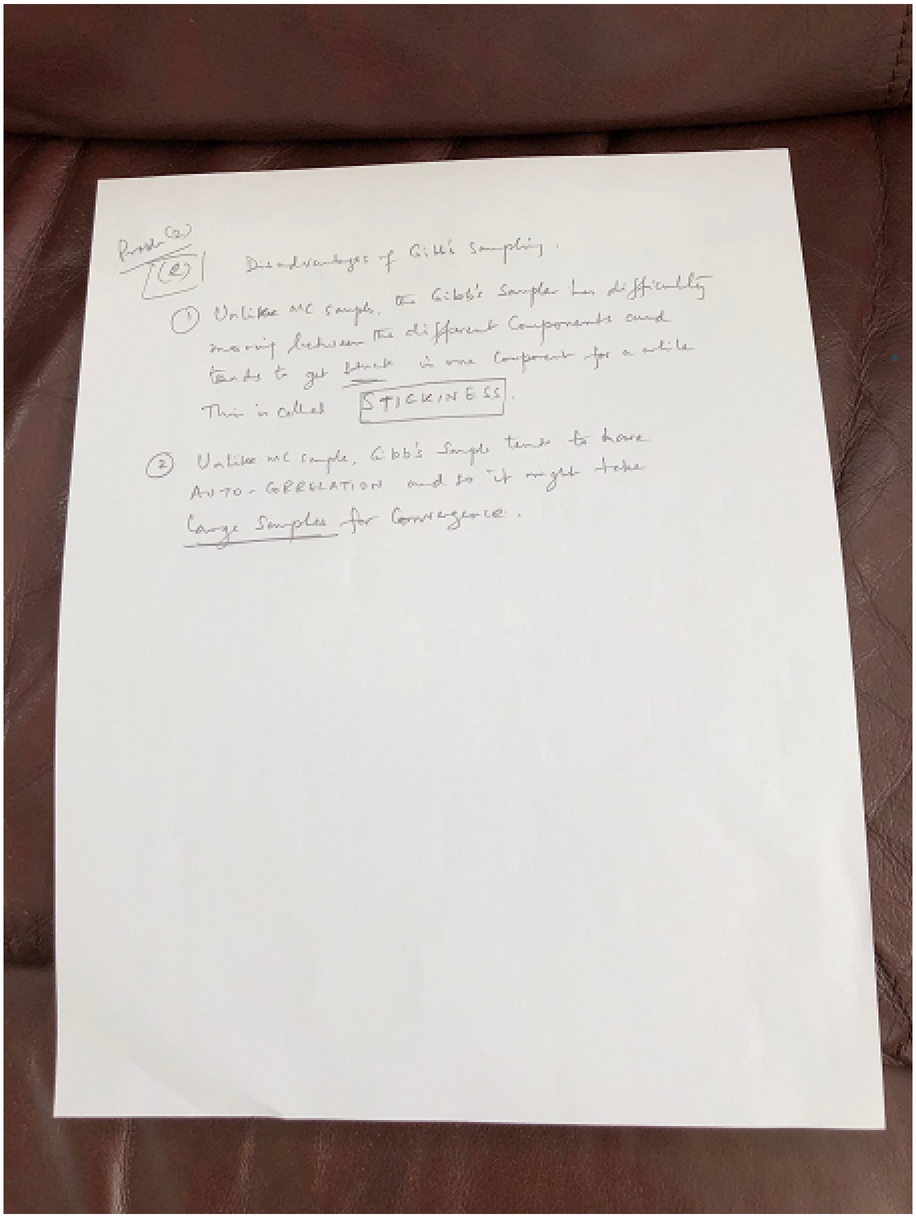
To do step 1), 2) and 3), we need to find $P(\mu | \phi, x)$ for $P(\mu | \sigma^2, x)$ where $\phi = \frac{1}{\sigma^2}$. Also we need to find $P(\phi | \mu, x)$ for $P(\sigma^2 | \mu, x)$

This above method is called Gibbs Sampling.

$[\mu^{(1)}, \phi^{(1)}]$
$[\mu^{(2)}, \phi^{(2)}]$
$\vdots$
$[\mu^{(n)}, \phi^{(n)}]$

Answer-2-c-d

Answer 2 e



Prob (2)

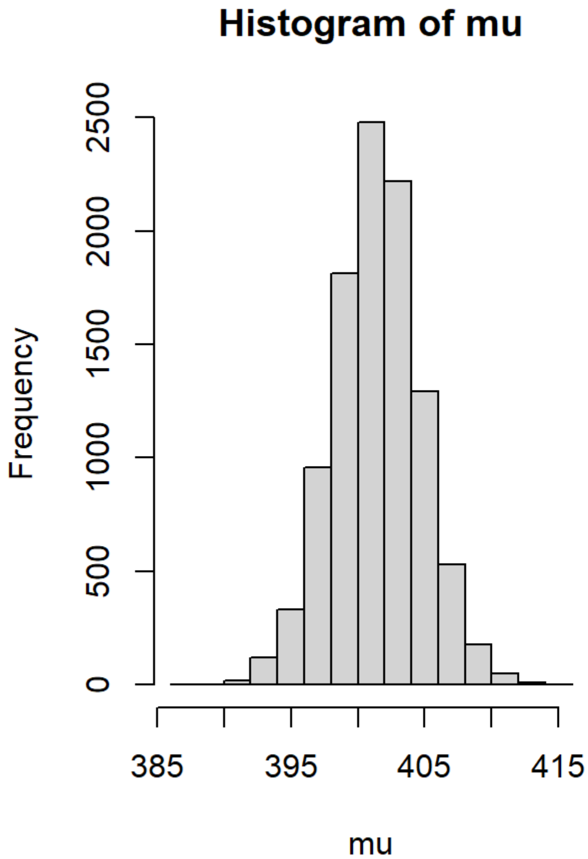
(e) Disadvantages of Gibb's Sampling.

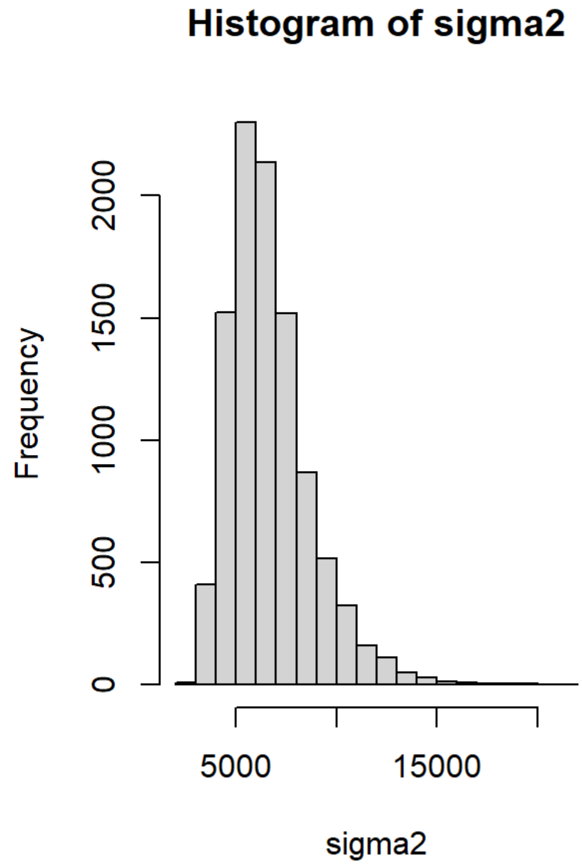
(1) Unlike MC sample, the Gibb's Sampler has difficulty moving between the different Components and tends to get <u>stuck</u> in one Component for a while. This is called $\boxed{\text{STICKINESS}}$.

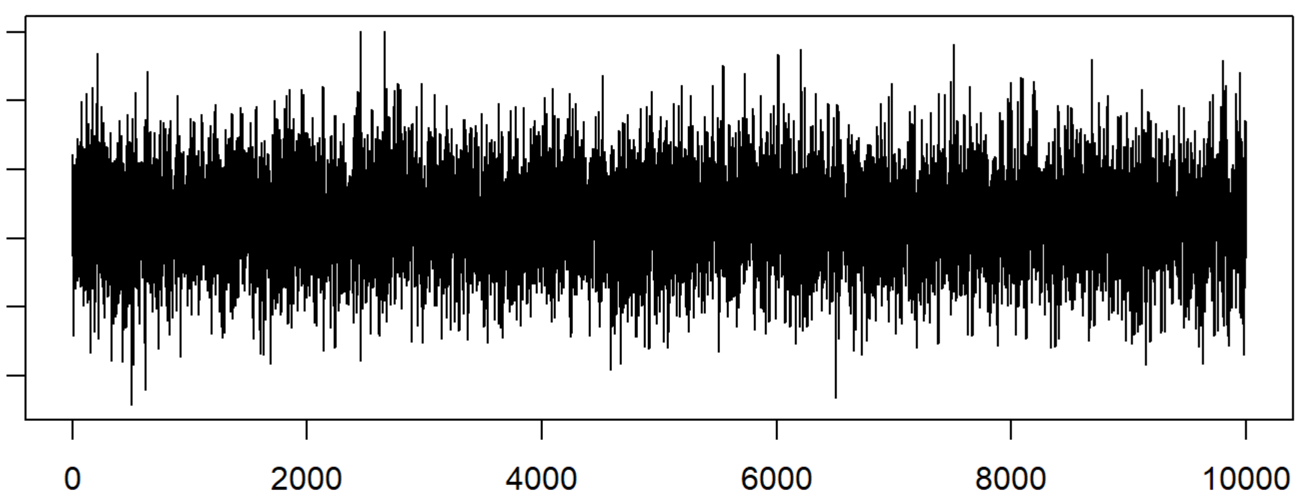
(2) Unlike MC sample, Gibb's Sample tends to have AUTO-CORRELATION and so it might take <u>Large Samples</u> for Convergence.

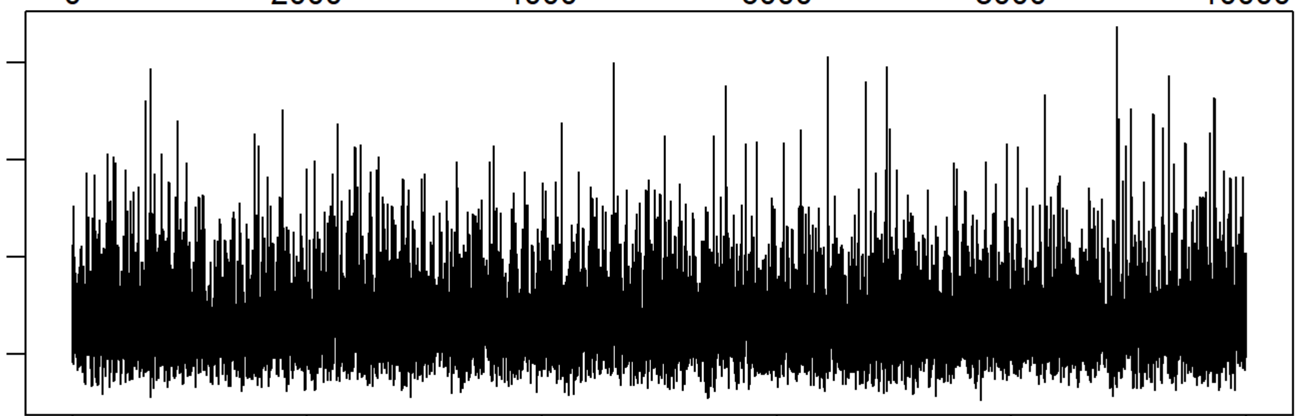
Answer-2-e

Answer 2f


Histogram of mu
Frequency
0
500
1000
1500
2000
2500
385
395
405
415
mu


Histogram of sigma2
Frequency
0
500
1000
1500
2000
5000
15000
sigma2


0
2000
4000
6000
8000
10000



The 95% credible interval for μ is displayed below :

```
##     2.5%    97.5% 
## 394.9349 407.9236
```

The 95% credible interval for σ^2 is displayed below :

```
##      2.5%     97.5% 
##  3770.402 11868.201
```