Roman Shaikh

Student number: 18300989

CS7DS3-Applied Statistical Modelling

Short-Assignment-1

Question 1:

Given:

Let C -> Cohen brothers, E -> Eve

P (Coen brothers) $=$ 0.4 can also be written as P(C) $= 0.4$

P (Coen brothers | Eve) $=$ 0.2 can also be written as P(C|E) $= 0.2$

P(Coen brothers | Not Eve) $=$ 0.7 can also be written as P (C|Ē) $= 0.7$

a) To show that 60% of the time I watch movies with Eve i.e. P (Eve) = 0.6

$$P\ (C)\ =\ P\ (C, E)\ +\ P\ (C, \bar{E})$$

$$P(E)\ P(C|E)\ +\ P(\bar{E})\ P\ (C\ |\ \bar{E})$$

Assuming P (E) to be X which gives us $P(\bar{E})\ =\ 1 - X$

Therefore, $0.4\ =\ X * 0.2\ +\ (1 - X)\ *\ 0.7$

$$= 0.2X\ + 0.7 - 0.7X$$

$$X = 0.3/0.5\ =\ 0.6$$

Therefor $P(E)\ =\ 0.6$ i.e. I watch movies with Eve 60% of the time.

b) Probability of Eve watching Coen Brother movie with you i.e. P (Eve | Coen brothers)?

By Bays theorem we know that,

$$P(E|C)\ = P\ (E, C)/\ P(C)$$

$$= P(C|E)\ P(E)/\ P\ (C)$$

$$= (0.2\ *\ 0.6)\ /\ 0.4$$

$$= 0.3$$

Therefore, the probability of Eve watching Coen Brother movie with you is 30%

Question 2:

Given:

$T \sim Exp(\theta)$

a) Let $y1, y2 \ldots, yn.$ be some random samples from exp(θ),

Then we can give the likelihood function by $\mathrm{L}(\bar{\mathrm{y}}, \theta) = \theta^{\mathrm{n}} \mathrm{e}^{-\sum_{i=1}^{n} \theta \mathrm{Y_i}}$

$$\mathrm{Log\,L}(\bar{\mathrm{y}}, \theta) = \mathrm{Log}\,\theta^{\mathrm{n}} - \theta \sum_{i=1}^{n} \mathrm{Y_i}$$

$$\frac{\partial \mathrm{LogL}}{\partial \theta} = \frac{n}{\theta} - \sum Y_i$$

$$\frac{\partial \mathrm{LogL}}{\partial \theta} = 0 \;\;=>\;\; \frac{n}{\theta} = \sum Y_i$$

$$\Rightarrow \;\; \theta = \frac{n}{\sum Y_i}$$

$$\Rightarrow \;\; = \frac{\frac{1}{\sum Y_i}}{n}$$

$$\Rightarrow \;\; = \frac{1}{\bar{\mathrm{y}}}$$

i.e. $\frac{1}{\bar{\mathrm{y}}}$ is the MLE for θ

b) We are given that $\theta \sim G(\,a_0, b_0)$

i.e. $f(\theta) = \frac{1}{\Gamma a_0 b_0^{a_0}} \cdot e^{-\frac{\theta}{b_0}} \cdot \theta^{a_0 - 1}$

to find the posterior distribution we have,

$$f(y|\theta) = \frac{\theta^n . e^{-\sum_{i=1}^{n} \theta \mathrm{Y_i}} . \frac{1}{\Gamma a_0 b_0^{a_0}} . e^{-\frac{\theta}{b_0}} . \theta^{a_0-1}}{\int_0^\infty \frac{1}{\Gamma a_0 b_0^{a_0}} . e^{-\frac{\theta}{b_0}} . \theta^{a_0-1} . \theta^n . e^{-\sum_{i=1}^{n} \theta \mathrm{Y_i}}}$$

$$= \frac{\theta^{n+a_0-1} . e^{-\theta \sum \mathrm{Y_i} - \frac{\theta}{b_0}}}{\int_0^\infty \theta^{n+a_0-1} . e^{-\theta(\sum \mathrm{y_i} + \frac{1}{b_0})}}$$

$$= \frac{\frac{1}{\Gamma n + a_0}}{\left(\sum \mathrm{y_i} + \frac{1}{b_0}\right)^{n+a_0}} \; . \theta^{n+a_0-1} \; . \, e^{-\theta\left[\sum y_i + \frac{1}{b_0}\right]}$$

This equation again follows gamma distribution with parameters $G(\,n + a_0, \sum y_i + \frac{1}{b_0})$. Here we can note that the posterior parameters $n + a_0 \; and \sum y_i + \frac{1}{b_0}$ depend upon the prior once $a_0 \; and \; b_0$

c) Given:

$\bar{y} = 4.71$ min ; A is viewed 30 times

$\bar{y} = 4.87$ min ; B is viewed 20 times

for worst layout minimum engagement time is 45 mins i.e. $E[T]_{MIN} = 0.75$

for optimistic layout maximum engagement time is more than 10 mins i.e. $E[T]_{MAX} = 10$

We know that, $E[\theta] = \frac{a}{b}$ and $Var[\theta] = \frac{a}{b^2}$ (because $\theta \sim G(a, b)$)……………………………………………………{1}

Now, Given $E[T] = \frac{1}{\theta}$ and $Var[T] = \frac{1}{\theta^2}$

Therefore, $E[T]_{MIN} = \frac{1}{\theta_{MAX}}$ and $E[T]_{MAX} = \frac{1}{\theta_{MIN}}$

$$\theta_{MAX} = \frac{1}{E[T]_{MIN}} = \frac{4}{3}$$

$$\theta_{MIN} = \frac{1}{E[T]_{MAX}} = \frac{1}{10}$$

$$E[\theta] = \frac{\theta_{MIN} + \theta_{MAX}}{2} = \frac{\frac{4}{3} + \frac{1}{10}}{2} = \frac{43}{60} \text{……………………………………………………………………………………………………}\{2\}$$

From {1} and {2} we can say that,

$$a = \frac{43}{60} b$$

Further,

$$\theta_{MAX} = \mu + 2\sigma$$

$$\sigma = \frac{\theta_{MAX} - E[\theta]}{2} * \frac{9}{10}$$

$$= \frac{\frac{4}{3} - \frac{43}{6}}{2} * 0.9$$

$\sigma = 0.2775$

$\sigma = \frac{a}{b^2} = 0.2775$

Therefore,

$a = 0.2775 * b^2$

$0.2775 * b^2 = \frac{43}{60}\ b$

Which gives

$b = 0.7167 * 3.6036$

$b = 2.58$

And there for

$a = 0.7167 * 2.58$

$a = 1.85$

Finally, we get,

$a_0 = 1.85\ and\ b_0 = 2.58$

It is possible to specify a non-informative prior in this case non-informative is nothing but uniform distribution the prior given has gamma distribution. Thus, gamma distribution is converted to uniform distribution with $a_0, b_0 = 0$. The information distribution will be of the form

$$f(x) = \begin{Bmatrix} \frac{1}{b-a} & if\ \ a \le x \le b \\ 0 & Otherwise \end{Bmatrix}$$

d) Using Monte Carlo methods in the following R code we estimate the probability that the new layout B has increased user engagement compared to the existing Layout A i.e. $\mathrm{P}(\theta_A > \theta_B)$

i. 95% interval for the difference in engagement rates
   - 95% interval is [-0.03833525, 0.0614802]

ii. The probability P ($\theta_A > \theta_B$) = 0.5658, which is approximately 50 %. This suggests us that the we are unsure of the probability that θ_B is greater than θ_A.
   Also, the density plot suggests that the difference in engagement rate is concentrated near zero, which supports our uncertainty about the probability.
   So, we conclude that more information is required to support the proposed profess of Layout B being superior to Layout A.

```
a0 = 5.98614
b0 = 1.11370
y_mean.A = 4.71
y_mean.B = 4.87
A_n = 30
B_n = 20

an = a0 + 30
bn = (30*4.71)+b0

theta.sim_A <- rgamma(5000, shape = an, rate = bn)

alpha_m = alpha_0 + 20
beta_m = (20*4.87)+beta_0

theta.sim_B <- rgamma(5000, shape = alpha_m, rate = beta_m)

sum(theta.sim_B > theta.sim_A)/5000
#Result 0.5658

diffrence_between_theta = theta.sim_B - theta.sim_A

#Finding 95 % internal of the difference in engagement rates
95percent_interval_diffrence =
trim(sort(diffrence_between_theta),trim=0.025)

min(95percent_interval_diffrence)
#Result -0.03833525

max(95percent_interval_diffrence)
#Result 0.0614802

df_theta <- data.frame(diffrence_between_theta)

#Plotting the density plot
ggplot(df_theta, aes(diffrence_between_theta, ..density..)) +
geom_histogram(bins = 5, aes(alpha = 0.5, colour = "blue"),
show.legend = FALSE) + geom_density(fill = "pink", aes(alpha = 0.5),
show.legend = FALSE) + xlab("Difference in theta") + ylab("Density") +
ggtitle("Difference in engagement rates")
```

e) Given:

y – time a user spends on a webpage

g - denotes the value (in euros) of this time to the company

Then $g(y) = 2(\log y + 5)$ is value of user engagement.

Using Monte Carlo methods in the following R code, we find the expected value of the company website when visiting layout A.

- The expected value comes out to be E2.532182
- The short comings of this projections can be
- Alternative approach.

```
mle_for_theta = 1 / mean_y.A

#Generating 5000 samples of y.A
y.A.1 = rexp(5000,mle_for_theta)

#calculating the value for user engament i.e.  g(y)
g_of_y_for_A.1  = 2*log(y.A.1 + 0.5)

#Calculating the expected value to the company
mean(g_of_y_for_A.1 )
#Result 2.532182
```

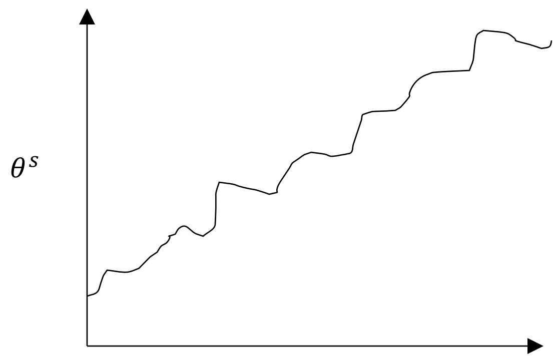
Question 3:

Given:

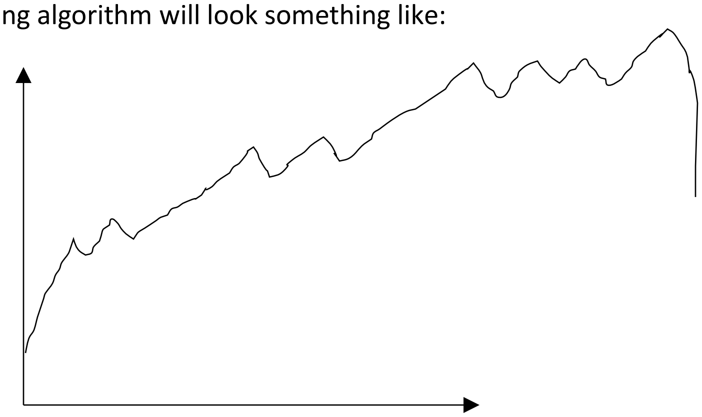
$$\theta^* = \theta^s + u^*$$

where $u^* \sim U(0,1)$ and U denotes a uniform distribution

a) since we know that u* ranges from 0 to 1 can only be positive, the H-M algorithm will resemble an increasing graph.
   Like $\theta^* = \theta^{s+1} \geq \theta^s$



Optimal performing algorithm will look something like:

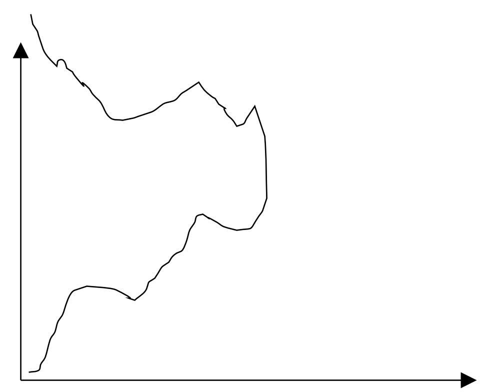


In this case of optimally performing algorithm, the kernel holds the following properties:

- Irreducible: cannot be reduced to separate smaller states
- Aperiodic: The system does not return to same state after.
- Recurrent: The expected number of steps for returning to the same state is finite.

b) By the property of Reversibility, we can deduce that if θ^{s-1} is taken from $P(\theta\ |y)$ then θ^{s} can also be drawn from $P(\theta\ |y)$

- Therefor the condition will not be true in that case.
- Because the M-H algorithm can only be positive values, in this case and hence it won't be able to trace back the path.
- Therefore θ^{s-1} cannot be concluded from $P(\theta\ |y)$ even if we can draw θ^{s} from $P(\theta\ |y)$
- Visualizing the trace path of the θ^{s-1} from $P(\theta\ |y)$, it looks like below(somewhat)



c) A good alternative of proposal distribution for this would be $U(-1,1)$

- This will help in controlling the ever-increasing value of the M-H algorithm to run optimally.
- Also, a normal distribution $N(-1,1)$ can also help with the problem.