Assignment: Statistics and Probability

African Institute for Mathematical Sciences

Quantum Leap Africa

Foundations of Machine Learning

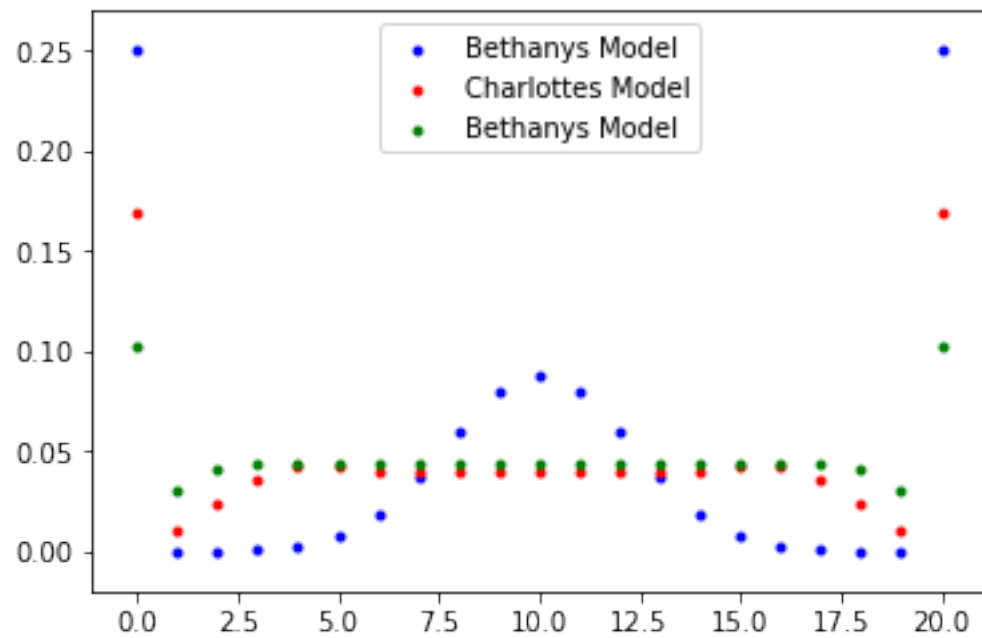*Author: Lionel Ngoupeyou Tondji*

Date: October 16, 2018

1 Discrete Models

c)

The following graph is showing the evidence for each model

d)

1)

Table over the 11 different possibilities under each model

Bethany-model	Charlotte-model	Davina-model
0	0.00000000e+00	0.00000000e+00
0	0.00000000e+00	7.78633801e-09
0	5.66695902e-05	2.79676219e-05
0	0.00000000e+00	2.13746139e-03
0	6.19681968e-02	3.05825945e-02
1	0.00000000e+00	1.55251489e-01
0	7.05856492e-01	3.48354866e-01
0	0.00000000e+00	3.44952256e-01
0	2.32118641e-01	1.14555379e-01
0	0.00000000e+00	4.13797926e-03
0	0.00000000e+00	0.00000000e+00

2)

Predictive distribution

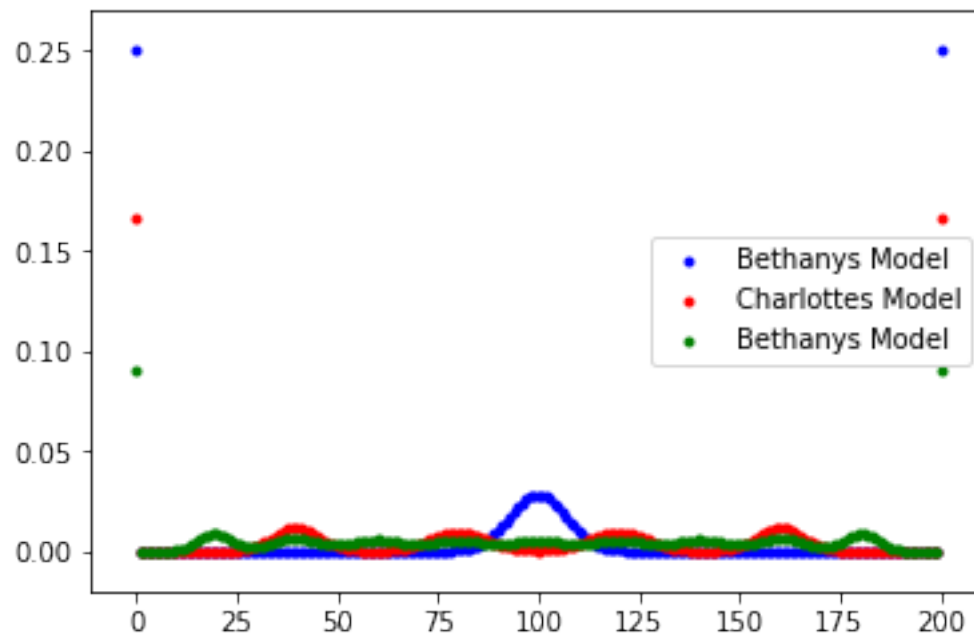| Bethany-model	Charlotte-model	Davina-model
0.5	0.63400742105992502	0.63635359808933689

e)

1)

According to the question $c)$ we can say that all of the three girls are right because the graph is in accord with what they have said at the beginning.

2)

What happen If Andrew had drawn 130 white balls out of 200?

As we can see on the graph, when $nTotal$ is very big the Charlotte's model and the Bethany's model tends to be the same.

2   Continuous Models

a)

Derive Freds maximum likelihood solution for the parameters of his model. The parameters for the Fred's model is $\theta = (\mu, \sigma^2)$.
Likelihood: Probability of the observed data x given the parameters θ

The Likelihood is given by :

$$p(x|\theta) = p(x_1, x_2, ..., x_N \ \theta) = \prod_{n=1}^{N} p(x_n|\theta)$$

Our goal is to Find the parameters that maximize the likelihood $p(x|\theta)$. The step of finding the maximum are :

- Compute gradient with respect to

- Set gradient to 0

- Solve for

So we have :

$$p(x|\theta) = \prod_{n=1}^{N} p(x_n|\theta) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \sigma^2)$$

But instead of maximizing the Likelihood, we will maximize the log-Likelihood because log is a strictly monotonically function increasing function.

Why are we allow to this transformation? because

- We will obtain The same maximum

- Gradients easy

- Fewer numerical problems

Thus, the Log Likelihood is given by :

$$log\, p(x|\theta) = log \prod_{n=1}^{N} p(x_n|\theta)$$

$$= \sum_{n=1}^{N} log \left(\frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2\sigma^2}(x_n - \mu)^2 \right) \right)$$

$$= -\frac{N}{2} log(2\pi\sigma^2) + \sum_{n=1}^{N} \left(-\frac{1}{2\sigma^2}(x_n - \mu)^2 \right)$$

The gradient of the Log Likelihood with respect to μ is given by :

$$\frac{\partial}{\partial\mu} log\, p(x|\theta) = \frac{\partial}{\partial\mu} \left(-\frac{N}{2} log(2\pi\sigma^2) + \sum_{n=1}^{N} \left(-\frac{1}{2\sigma^2}(x_n - \mu)^2 \right) \right)$$

$$= \sum_{n=1}^{N} \frac{\partial}{\partial\mu} \left(-\frac{1}{2\sigma^2}(x_n - \mu)^2 \right)$$

$$= \sum_{n=1}^{N} \left(\frac{1}{\sigma^2}(x_n - \mu) \right)$$

Setting the gradient that we compute before to zero, we obtain:

$$\sum_{n=1}^{N}\left(\frac{1}{\sigma^2}(x_n-\mu)\right)=0 \implies \sum_{n=1}^{N} x_n - N\mu = 0$$

So we obtain :

$$\boxed{\mu_{ML}=\frac{1}{N}\sum_{n=1}^{N}x_n} \tag{1}$$

Let pose $\alpha=\sigma^2$.
The gradient of the Log Likelihood with respect to σ^2 is given by :

$$\begin{aligned}
\frac{\partial}{\partial\alpha}log\,p(x|\theta) &= \frac{\partial}{\partial\alpha}\left(-\frac{N}{2}log(2\pi\alpha)+\sum_{n=1}^{N}\left(-\frac{1}{2\alpha}(x_n-\mu)^2\right)\right)\\
&= \frac{\partial}{\partial\alpha}\left(-\frac{N}{2}log(2\pi\alpha)\right)+\frac{\partial}{\partial\alpha}\left(\sum_{n=1}^{N}\left(-\frac{1}{2\alpha}(x_n-\mu)^2\right)\right)\\
&= -\frac{N}{2\alpha}+\sum_{n=1}^{N}\frac{1}{2\alpha^2}(x_n-\mu)^2
\end{aligned}$$

Setting the gradient that we compute before to zero, we obtain:

$$-\frac{N}{2\alpha}+\sum_{n=1}^{N}\frac{1}{2\alpha^2}(x_n-\mu)^2=0 \implies -N\alpha+\sum_{n=1}^{N}(x_n-\mu)^2=0$$

Replacing α by σ^2

$$-N\sigma^2+\sum_{n=1}^{N}(x_n-\mu)^2=0$$

So we obtain :

$$\boxed{\sigma_{ML}^2=\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2} \tag{2}$$

6

b)

Using only the properties of expectation and variance of linear combinations of iid random variables, show that the maximum likehood estimator σ^2_{ML} is biased. (that is, taking the expectation of the maximum likelihood estimator with respect $\mathcal{N}(\mu, \sigma^2)$ does not return σ^2).

From the previous question, we already have the value of σ^2_{ML}

$$
\begin{aligned}
\sigma^2_{ML} &= \frac{1}{N} \sum_{n=1}^{N} \left(x_n - \mu \right)^2 \\
&= \frac{1}{N} \sum_{n=1}^{N} \left(x_n - \mu_{ML} \right)^2 \\
&= \frac{1}{N} \sum_{n=1}^{N} x_n^2 - \frac{2}{N} \sum_{n=1}^{N} x_n \mu_{ML} + \mu_{ML}^2 \\
&= \frac{1}{N} \sum_{n=1}^{N} x_n^2 - 2\mu_{ML} \left(\frac{1}{N} \sum_{n=1}^{N} x_n \right) + \mu_{ML}^2 \\
&= \frac{1}{N} \sum_{n=1}^{N} x_n^2 - 2\mu_{ML}^2 + \mu_{ML}^2 \\
&= \frac{1}{N} \sum_{n=1}^{N} x_n^2 - \mu_{ML}^2
\end{aligned}
$$

Now we can compute the expectation σ_{ML}^2:

$$E(\sigma_{ML}^2) = E\left(\frac{1}{N}\sum_{n=1}^{N} x_n^2 - \mu_{ML}^2\right)$$

$$= E\left(\frac{1}{N}\sum_{n=1}^{N} x_n^2\right) - E(\mu_{ML}^2)$$

$$= \frac{1}{N}\sum_{n=1}^{N} E(x_n^2) - E(\mu_{ML}^2)$$

$$\overset{\text{iid}}{=} E(x_N^2) - E(\mu_{ML}^2)$$

According to the alternative definition of variance, $\sigma_x = E(x^2) - E(x)^2$ and similarly, $\sigma_{\mu_{ML}} = E(x_{\mu_{ML}}^2) - E(x_{\mu_{ML}})^2$ where the random variable is μ_{ML}. We can notice that $E(x) = E(x_{\mu_{ML}}) = \mu$. Plug the 2 equations to the derivation:

$$E(\sigma_{ML}^2) = (\sigma_x^2 + \mu^2) - (\sigma_{\mu_{ML}}^2 + \mu^2)$$

$$= \sigma_x^2 - \sigma_{\mu_{ML}}^2$$

We have :

$$\sigma_{\mu_{ML}}^2 = Var(\frac{1}{N}\sum_{n=1}^{N} x_n)$$

$$= \frac{1}{N^2}Var(\sum_{n=1}^{N} x_n)$$

$$\overset{\text{iid}}{=} \frac{1}{N^2}\sum_{n=1}^{N} Var(x_n)$$

$$= \frac{1}{N^2}\sum_{n=1}^{N} \sigma_x^2$$

$$= \frac{1}{N}\sigma_x^2$$

Plug back to the $E(\sigma^2_{ML})$ derivation,
So we have :

$$E(\sigma^2_{ML}) = \sigma^2_x - \sigma^2_{\mu_{ML}}$$
$$= \sigma^2_x - \frac{1}{N}\sigma^2_x$$
$$= \frac{N-1}{N}\sigma^2_x$$

So we get :

$$E(\sigma^2_{ML}) \neq \sigma^2_{ML} \tag{3}$$

Then we conclude that the estimator σ^2_{ML} is biased.

c)

Derive the MAP solution for in Georges model. Do your analysis replacing 10 with μ_0 and 25 with σ^2_0 (this makes it easier to read). Write your answer in terms of Freds maximum likelihood solution μ_{ML}.

The Posterior is given by :

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

Where $x = (x_1, x_2, ..., x_n)$, $x \sim \mathcal{N}(\mu, \sigma^2)$ and $\mu \sim \mathcal{N}(\mu_0, \sigma^2_o)$
Our goal is to Find the parameters that maximize the log-Posterior $p(\theta|x)$. The step of finding the maximum are :

- Compute gradient with respect to

- Set gradient to 0

- Solve for

9

Thus, the Log Posterior is given by :

$$log\, p(\theta|x) = log\, p(x|\theta) + log\, p(\theta) + cste$$

$$= -\frac{N}{2}log\left(2\pi\sigma^2\right) - \sum_{n=1}^{N}\left(\frac{1}{2\sigma^2}\left(x_n - \mu\right)^2\right) + log\,\mathcal{N}(\mu_0, \sigma_0^2)$$

$$= -\frac{N}{2}log\left(2\pi\sigma^2\right) - \sum_{n=1}^{N}\left(\frac{1}{2\sigma^2}\left(x_n - \mu\right)^2\right) - \frac{1}{2}log\left(2\pi\sigma_0^2\right) - \left(\frac{1}{2\sigma_0^2}\left(\mu - \mu_0\right)^2\right)$$

The gradient of the Log Posterior with respect to μ is given by :

$$\frac{\partial}{\partial\mu}log\, p(\theta|x) = \frac{\partial}{\partial\mu}\left(-\frac{N}{2}log\left(2\pi\sigma^2\right) - \sum_{n=1}^{N}\left(\frac{1}{2\sigma^2}\left(x_n - \mu\right)^2\right) - \frac{1}{2}log\left(2\pi\sigma_0^2\right) - \left(\frac{1}{2\sigma_0^2}\left(\mu - \mu_0\right)^2\right)\right)$$

$$= \sum_{n=1}^{N}\left(\frac{1}{\sigma^2}\left(x_n - \mu\right)\right) - \frac{1}{\sigma_0^2}\left(\mu - \mu_0\right)$$

Setting this gradient that we compute before to zero, we obtain:

$$\frac{\partial}{\partial\mu}log\, p(\theta|x) = 0 \implies \sum_{n=1}^{N}\left(\frac{1}{\sigma^2}\left(x_n - \mu\right)\right) - \frac{1}{\sigma_0^2}\left(\mu - \mu_0\right) = 0$$

$$\implies \sum_{n=1}^{N}\frac{1}{\sigma^2}x_n - \frac{N}{\sigma^2}\mu - \frac{1}{\sigma_0^2}\mu + \frac{1}{\sigma_0^2}\mu_0 = 0$$

$$\implies \frac{N}{\sigma^2}\mu_{ML} - \frac{N}{\sigma^2}\mu - \frac{1}{\sigma_0^2}\mu + \frac{1}{\sigma_0^2}\mu_0 = 0$$

$$\implies \mu\left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right) = \frac{N}{\sigma^2}\mu_{ML} + \frac{1}{\sigma_0^2}\mu_0$$

So we obtain :

$$\boxed{\mu_{MAP} = \frac{1}{\left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right)}\left(\frac{N}{\sigma^2}\mu_{ML} + \frac{1}{\sigma_0^2}\mu_0\right)} \qquad (4)$$

10

d)

Explaining your reasoning, calculate the posterior for Georges model. Show that the MAP point you calculated in the previous exercise is also the mean, and give a reason why this is true in this example but not true in general. Again use μ_0 and σ_0^2 rather than use the actual numbers.

The Posterior is given by :

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

Where $x = (x_1, x_2, ..., x_n)$, $x \sim \mathcal{N}(\theta, \sigma^2)$ and $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$.
Given the formula for the posterior we can say that :

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

We get :

$$p(x|\theta)p(\theta) = e^{-\frac{1}{2}M}$$

Where :

$$M = \sum_{n=1}^{N} \frac{1}{\sigma^2}(x_n - \theta)^2 + \frac{1}{\sigma_0^2}(\theta - \mu_0)^2$$

$$= \frac{1}{\sigma^2}\sum_{n=1}^{N} x_n^2 + \frac{N}{\sigma^2}\theta^2 - \frac{2\theta}{\sigma^2}\sum_{n=1}^{N} x_n + \frac{\theta^2}{\sigma_0^2} + \frac{\mu_0^2}{\sigma_0^2} - \frac{2\theta\mu_0}{\sigma_0^2}$$

$$= \underbrace{\left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right)}_{a}\theta^2 - 2\underbrace{\left(\frac{1}{\sigma^2}\sum_{n=1}^{N} x_n + \frac{\mu_0}{\sigma_0^2}\right)}_{b}\theta + \underbrace{\left(\frac{1}{\sigma^2}\sum_{n=1}^{N} x_n^2 + \frac{\mu_0^2}{\sigma_0^2}\right)}_{c}$$

So :

$$M = a\theta^2 - 2b\theta + c$$

$$= a\left(\theta^2 - \frac{2b}{a} + \frac{c}{a}\right)$$

$$= a\left[\left(\theta - \frac{b}{a}\right)^2 - \frac{b^2}{a^2} + \frac{c}{a}\right]$$

$$= a\left(\theta - \frac{b}{a}\right)^2 - \frac{b^2}{a} + c$$

Then :

$$p(x|\theta)p(\theta) = exp\left[-\frac{1}{2}\left(a(\theta - \frac{b}{a})^2 - \frac{b^2}{a} + c\right)\right]$$

$$= Cste \times exp\left(-\frac{1}{2}a(\theta - \frac{b}{a})^2\right)$$

We can conclude that $p(\theta|x) \sim \mathcal{N}(\frac{b}{a}, \frac{1}{a})$, where :

$$a = \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right)$$

$$b = \left(\frac{1}{\sigma^2}\sum_{n=1}^{N} x_n + \frac{\mu_0}{\sigma_0^2}\right)$$

$$c = \left(\frac{1}{\sigma^2}\sum_{n=1}^{N} x_n^2 + \frac{\mu_0^2}{\sigma_0^2}\right)$$

1)

We can see that the mean of the posterior is given by :

$$\frac{b}{a} = \frac{1}{\left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right)}\left(\frac{1}{\sigma^2}\sum_{n=1}^{N} x_n + \frac{\mu_0}{\sigma_0^2}\right)$$

$$= \frac{1}{\left(\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}\right)}\left(\frac{N}{\sigma^2}\mu_{ML} + \frac{\mu_0}{\sigma_0^2}\right)$$

$$= \mu_{MAP}$$

So finally we end-up with :

$$\boxed{\frac{b}{a} = \mu_{MAP}} \tag{5}$$

2)

Reason why this is true in this example but not true in general
The MAP point that we calculated previously is also the mean because the
distribution used is symmetric But in general if the distribution is not sym-
metric we will not have the equality.

e)

Derive the MAP estimate for Harrys model. Use a and b for the shape and scale 2 respectively instead of the numbers. Write the your answer for σ^2_{MAP} in terms of Freds maximum likelihood result σ^2_{ML} . NB you might find it easier to work with (and differentiate with respect to) σ^2 rather than σ

The Posterior is given by :

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

Where $x = (x_1, x_2, ..., x_n)$, $x \sim \mathcal{N}(\mu, \theta)$ and $\theta \sim \mathcal{IG}(a, b)$, where $\mathcal{IG}$ stand for the inverse Gamma distribution.

Given the formula for the posterior we can say that :

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

The Log Prior is given by :

$$log\, p(\theta) = log\, \left(\frac{b^a}{\Gamma(a)} (\frac{1}{\theta})^{a+1} exp(-\frac{b}{\theta}) \right)$$
$$= cste - (a+1)log\,\theta - \frac{b}{\theta}$$

Thus, the Log Posterior is given by :

$$log\, p(\theta|x) = log\, p(x|\theta) + log\, p(\theta) + cste$$
$$= -\frac{N}{2}log\,(2\pi\theta) - \sum_{n=1}^{N} \frac{1}{2\theta}(x_n - \mu)^2 + log\,\mathcal{IG}(a,b)$$
$$= -\frac{N}{2}log\,(2\pi\theta) - \sum_{n=1}^{N} \frac{1}{2\theta}(x_n - \mu)^2 - (a+1)log\,\theta - \frac{b}{\theta}$$

The gradient of the Log Posterior with respect to θ is given by :

$$\frac{\partial}{\partial \theta} log\, p(\theta|x) = \frac{\partial}{\partial \theta} \left(-\frac{N}{2} log(2\pi\theta) - \sum_{n=1}^{N} \left(\frac{1}{2\theta} (x_n - \mu)^2 \right) - (a+1) log\, \theta - \frac{b}{\theta} \right)$$

$$= -\frac{N}{2\theta} + \sum_{n=1}^{N} \frac{1}{2\theta^2} (x_n - \mu)^2 - \frac{(a+1)}{\theta} + \frac{b}{\theta^2}$$

Setting the previous gradient to zero, we get:

$$-\frac{N}{2\theta} + \sum_{n=1}^{N} \frac{1}{2\theta^2} (x_n - \mu)^2 - \frac{(a+1)}{\theta} + \frac{b}{\theta^2} = 0$$

$$\implies \frac{\theta^2}{N} \left(-\frac{N}{2\theta} + \sum_{n=1}^{N} \frac{1}{2\theta^2} (x_n - \mu)^2 - \frac{(a+1)}{\theta} + \frac{b}{\theta^2} \right) = 0$$

$$\implies -\frac{\theta}{2} + \frac{1}{2N} \sum_{n=1}^{N} (x_n - \mu)^2 - \frac{\theta(a+1)}{N} + \frac{b}{N} = 0$$

$$\implies \theta \left(-\frac{1}{2} - \frac{(a+1)}{N} \right) + \frac{1}{2N} \sum_{n=1}^{N} (x_n - \mu)^2 + \frac{b}{N} = 0$$

$$\implies \theta \left(\frac{1}{2} + \frac{(a+1)}{N} \right) = \frac{1}{2N} \sum_{n=1}^{N} (x_n - \mu)^2 + \frac{b}{N}$$

$$\implies \theta \left(\frac{1}{2} + \frac{(a+1)}{N} \right) = \frac{1}{2} \sigma_{ML}^2 + \frac{b}{N}$$

Then we have :

$$\boxed{\theta_{MAP} = \frac{1}{\left(\frac{1}{2} + \frac{(a+1)}{N} \right)} \left(\frac{1}{2} \sigma_{ML}^2 + \frac{b}{N} \right)} \tag{6}$$

14

f)

Derive Harry's posterior distribution. You may reuse some of your working from your previous answer. State also the posterior mean and explain why it is not equal to the MAP estimate you found in the previous part. You may use standard results for the mean of the Inverse Gamma distribution.

The Posterior is given by :

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

Where $x = (x_1, x_2, ..., x_n)$, $x \sim \mathcal{N}(\mu, \theta)$ and $\theta \sim \mathcal{IG}(a, b)$, where $\mathcal{IG}$ stand for the inverse Gamma distribution.

Given the formula for the posterior we can say that :

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

So :

$$p(\theta|x) = \left(\theta^{-\frac{N}{2}}exp\left(-\sum_{n=1}^{N}\frac{(x_n - \mu)^2}{2\theta}\right)\right)\left(\frac{b^a}{\Gamma(a)}\left(\frac{1}{\theta}\right)^{a+1}exp(-\frac{b}{\theta})\right)$$

$$= Cste \times \theta^{-\left(a+\frac{N}{2}+1\right)}exp\left(\frac{-\left(\frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)^2 + b\right)}{\theta}\right)$$

We can conclude that :

$$\boxed{p(\theta|x) \sim \mathcal{IG}\left(a + \frac{N}{2}, b + \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)^2\right)} \qquad (7)$$

Recall : Let a random variable $X \sim \mathcal{IG}(a, b)$, then it probability density function is $f_X(x) = \frac{b^a}{\Gamma(a)}\left(\frac{1}{x}\right)^{a+1}exp(-\frac{b}{x})$ where $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1}e^{-\alpha}dt$.

The mean is given by the expectation of X.

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

$$= \int_{0}^{+\infty} \frac{b^a}{\Gamma(a)} \left(\frac{1}{x}\right)^a \exp\left(-\frac{b}{x}\right) dx$$

$$= \frac{b^a}{\Gamma(a)} \int_{0}^{+\infty} \left(\frac{1}{x}\right)^a \exp\left(-\frac{b}{x}\right) dx$$

$$= \frac{b^a}{\Gamma(a)} \left(\underbrace{\left[\frac{1}{1-a} e^{-\frac{b}{x}} x^{1-a}\right]_{0}^{+\infty}}_{0} - \int_{0}^{+\infty} \frac{b}{1-a} x^{-a-1} e^{-\frac{b}{x}} dx \right)$$

$$= -\frac{b^{a+1}}{(1-a)\Gamma(a)} \left(\int_{0}^{+\infty} x^{-a-1} e^{-\frac{b}{x}} dx \right)$$

Let us pose $X = \frac{b}{x} \implies x = \frac{b}{X} \implies dx = -\frac{b}{X^2} dX$.
Plug this to the previous integral, we have :

$$E(X) = -\frac{b^{a+1}}{(1-a)\Gamma(a)} \left(-\int_{0}^{+\infty} \frac{b^{-a-1}}{X^{-a-1}} e^{-X} \left(-\frac{b}{X^2}\right) dX \right)$$

$$= -\frac{b}{(1-a)\Gamma(a)} \int_{0}^{+\infty} X^{a-1} e^{-X} dX$$

$$= \frac{b}{(a-1)\Gamma(a)} \Gamma(a)$$

We have show that :

$$\boxed{E(X) = \frac{b}{a-1}}$$

By the previous result we can now compute the mean of the posterior following an inverse Gamma distribution $\mathcal{IG}\left(a + \frac{n}{2}, \, b + \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)^2\right)$

$$posterior\ mean = \frac{b + \frac{1}{2}\sum_{n=1}^{N}(x_n - \mu)^2}{a + \frac{N}{2} - 1}$$

$$= \frac{1}{\left(\frac{1}{2} + \frac{(a-1)}{N}\right)}\left(\frac{1}{2}\sigma^2_{ML} + \frac{b}{N}\right)$$

We can notice that the mean of the posterior is not equal to the mean obtain by the MAP method.

1)

Justifications for the reasons why the posterior mean is not equal to the MAP estimate that we found in the previous part.
The MAP point that we calculated previously is not the mean because the distribution used is not symmetric.

g)

Explain what would happen to the result of inference in the three models if Elizabeth was to take a very large sample from the random number generator.

For the three model when $N \to +\infty$,

$$\mu_{MAP} \to \mu_{ML}$$

$$\sigma^2_{MAP} \to \sigma^2_{ML}$$