MATH2269 Semester 2, 2020 – Applied Bayesian Analysis

Gibbs sampling for normal-gamma model on Melbourne properties sales price

Assignment 1

Millie Woo (s3806940)

Contents

1. Introduction 4
2. Background of the modelling, distributions and apps 4
   i. Mathematical modelling 4
   ii. Summary statistics of the normal likelihood 4
   iii. Normal prior on μ 6
   iv. Gamma prior on τ 6
   iv. Apps to model the posterior distribution 8
3. Steps to generate results 10
   i. Non informative prior 10
      [a] Simulation 10
      [b] Diagnostic Check in representativeness and accuracy 14
   ii. Informative prior 15
      [a] Simulation 15
      [b] Diagnostic Check in representativeness and accuracy 17
4. Analysis 20
   i. Hypothesis test for non-Informative prior 20
   ii. Hypothesis test for Informative prior 21
5. Discussion on Efficiency 22
6. Conclusion 23
Appendix 23
   [A1] – Import packages and data preparation 23
   [A2] – Generate descriptive statistics 24
   [A3] – A histogram for the likelihood data 24

[A4] – Showing the Effect on variance on normal distribution 24
[A5] – Showing the Effect on standard deviation on gamma distribution 25
[A6] – A histogram for the likelihood data with comparison on the μ of likelihood and prior info on μ 25

Table of Figures

Figure 1 - Distribution of likelihood 5
Figure 2 - Normal prior on μ with variance = 1 or 7 6
Figure 3 - Gamma prior on τ with S.D. = 15, 10 or 5 7
Figure 4 - Gamma prior on τ with S.D. = 30, 20 or 15 7
Figure 5 - The Gamma Distribution Specified by Mean and Standard Deviation app 8
Figure 6 - The input parameters for the Gibbs sampling for the normal-gamma model app 9
Figure 7 - Comparing the mean from likelihood with the mean from prior knowledge 10
Figure 8 - posterior distribution of non-informative prior (first trial) 12
Figure 9 - summary on posterior distribution of non-informative prior (first trial) 12
Figure 10 - summary of posterior distribution of non-informative prior (first trial) compare with prior information 13
Figure 11 - posterior distribution of non-informative prior (first trial) compare with prior information 13
Figure 12 - Diagnostics for the gamma prior on τ for non-informative prior (first trial) 13
Figure 13 - Diagnostics for normal prior on μ for non-informative prior (first trial) 13
Figure 14 - Diagnostics for normal prior on μ for non-informative prior (first trial) with increased steps 14
Figure 15 - Diagnostics for the gamma prior on τ for non-informative prior (first trial) with increased steps 14
Figure 16 - Posterior distribution of Informative prior (trial 24) 16
Figure 17 - Summary of Posterior distribution of Informative prior (trial 24) 16
Figure 18 - Diagnostics for normal prior on μ for informative prior (trial 24) 17
Figure 19 - Diagnostics for the gamma prior on τ for informative prior (trial 24) 17
Figure 20 - Diagnostics for the gamma prior on τ for informative prior (trial 24) with not enough iterations 18
Figure 21 - Diagnostics for normal prior on μ for informative prior (trial 24) with not enough iterations 18
Figure 22 - summary of posterior distribution of informative prior (trial 24) with not enough iterations 18
Figure 23 - Posterior distribution of Informative prior (trial 24) with not enough iterations 18
Figure 24 - Diagnostics for the gamma prior on τ for trial 21 19
Figure 25 - Diagnostics for normal prior on μ for trial 21 19
Figure 26 - Diagnostics for the gamma prior on τ for trial 22 20
Figure 27 - Diagnostics for normal prior on μ for trial 22 20

Figure 28 - Posterior distribution of non-Informative prior (trial 2) – compare to values of hypothesis tests 20
Figure 29 - summary of posterior distribution of non-informative prior (trial 2) – compare to values of hypothesis test 20
Figure 30 - Posterior distribution of Informative prior (trial 24) – compare to values of hypothesis tests 21
Figure 31 - summary of posterior distribution of informative prior (trial 24) – compare to values of hypothesis test 21
Figure 32 - Summary of posterior distribution which shows the runtime of trial 21 22
Figure 33 - Summary of posterior distribution which shows the run time of trial 22 23

Table of Tables

Table 1 - Summary statistics on Melbourne properties sales price (In AUD$ 100,000) 5
Table 2 - Input parameters for non-informative prior 11
Table 3 - parameters search for informative prior 15
Table 4 - Runtime for different trials 22

1. Introduction

In this assignment, we are using Bayesian analysis to draw statistical inference on the mean and variance of Melbourne properties sales price. A datafile of sale price of properties in Melbourne (in AUD $100,000) is given to us as the population data of this analysis. We would be making Bayesian inferences with/without any informative prior knowledge and model the distribution of sales price using the provided apps <Gibbs sampling for the normal-gamma model> and document our findings in this report.

2. Background of the modelling, distributions and apps

i. Mathematical modelling

The following mathematical distributions are used to model the data we collected and the parameters of interest.

1. X ~ Normal(μ, τ), collected data of the current Melbourne properties sales price, the likelihood for us to build our posterior model. We are interested in 2 parameters, mean (μ) and variance(τ), from the dataset, since there are 2 parameters of interest, we would not have any conjugate prior settings, but we can use Gibbs sampling to model these parameters jointly to obtain posterior distributions.

2. μ ~ Normal(μ_0, τ_0), mean of the sales price is a continuous parameter, since we have both prior knowledge (AUD$ 750K) and degree of belief (high) for the mean, we need to use a 2-parameter distributions to model it. The domain of μ is [0, +∞], since the variance of μ is small and finite, we could still use the normal distribution [-∞, +∞] for modelling.

3. τ ~ Gamma(α, β), variance of the sales price is a continuous parameter, since we have both prior knowledge (standard deviation of AUD$ 600K) and degree of belief (high) in the variance, we need to use a 2-parameter distributions to model it. The domain of τ is [0, +∞], gamma distribution is perfect for modelling.

The posterior with normal likelihood, normal prior on μ and gamma prior on τ is not mathematically tractable, thus we need to use Gibbs sampling to conduct the posterior model. For the normal prior and gamma prior, we have to model the prior knowledge as the location (mean) and degree of belief as the dispersion (variance or standard deviation) in the distribution.

ii. Summary statistics of the normal likelihood

Please note, for easy interpretation, from this point onwards, all the figures being mentioned would be in the unit of AUD $100,000 (e.g. 6.9 is equivalent to AUD$609,000)

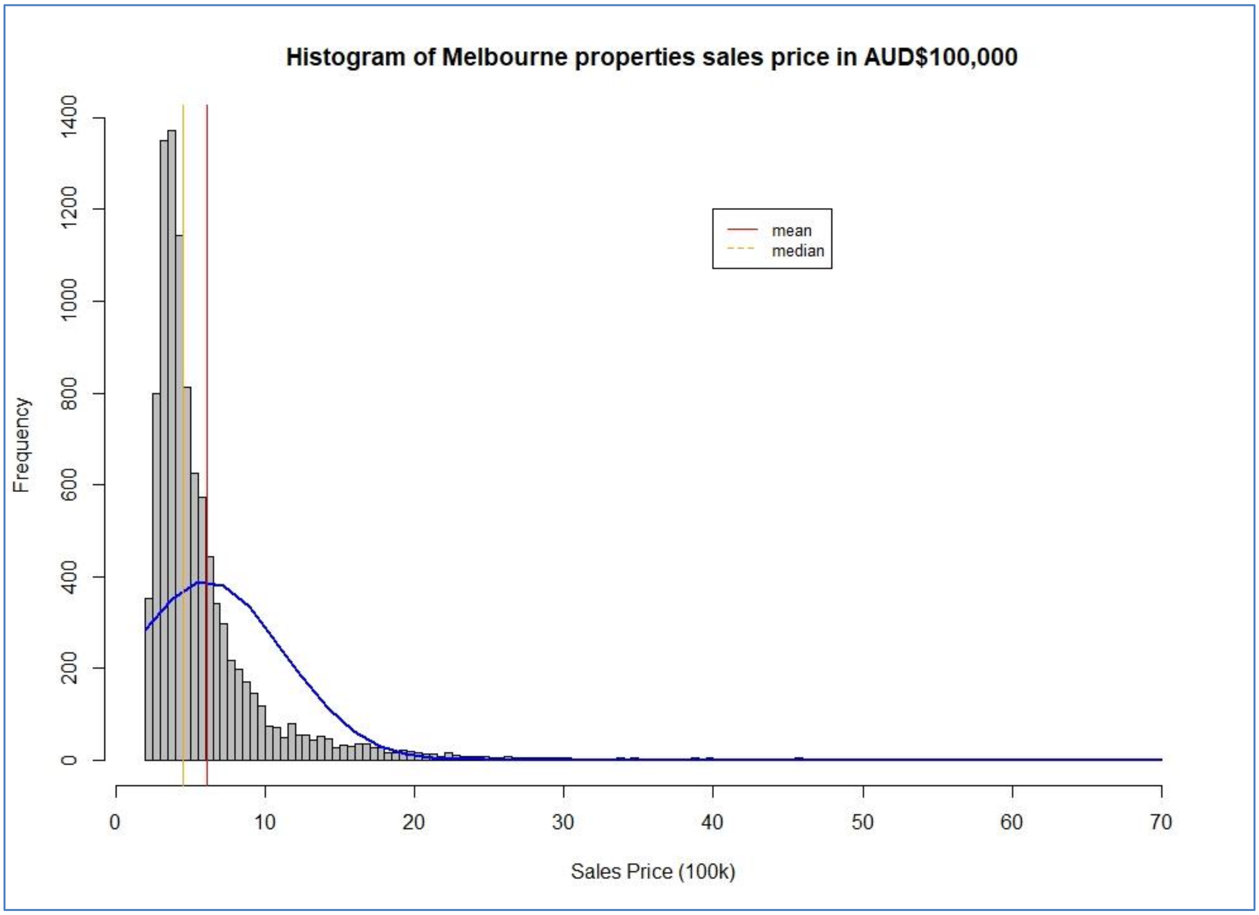
Refer to Table 1, it shows the mean of properties price is 6.09, with a standard deviation of 5.12, highest price at 70.0 and lowest price at 2.0. The median (4.50) is less than the mean, and Q1 at 3.50 and Q3 at 6.55, which both indicate the collected data is right skewed. There are only 10,000 observations in the dataset. (Refer to [A1] and [A2] on the R codes.)

Table 1 - Summary statistics on Melbourne properties sales price (In AUD$ 100,000)

```
## Descriptive Statistics
## house$price
## N: 10000
##
##                Mean   Median   Std.Dev     Q1     Q3    IQR    Min     Max
## ----------- ------ -------- --------- ------ ------ ------ ------ -------
##       price   6.09     4.50      5.12   3.50   6.55   3.05   2.00   70.00
```

We can use a histogram to visualise this dataset and fit into a normal distribution curve, the distribution is highly right-skewed as shown in Figure 1. (Refer to [A3] on the R codes.)

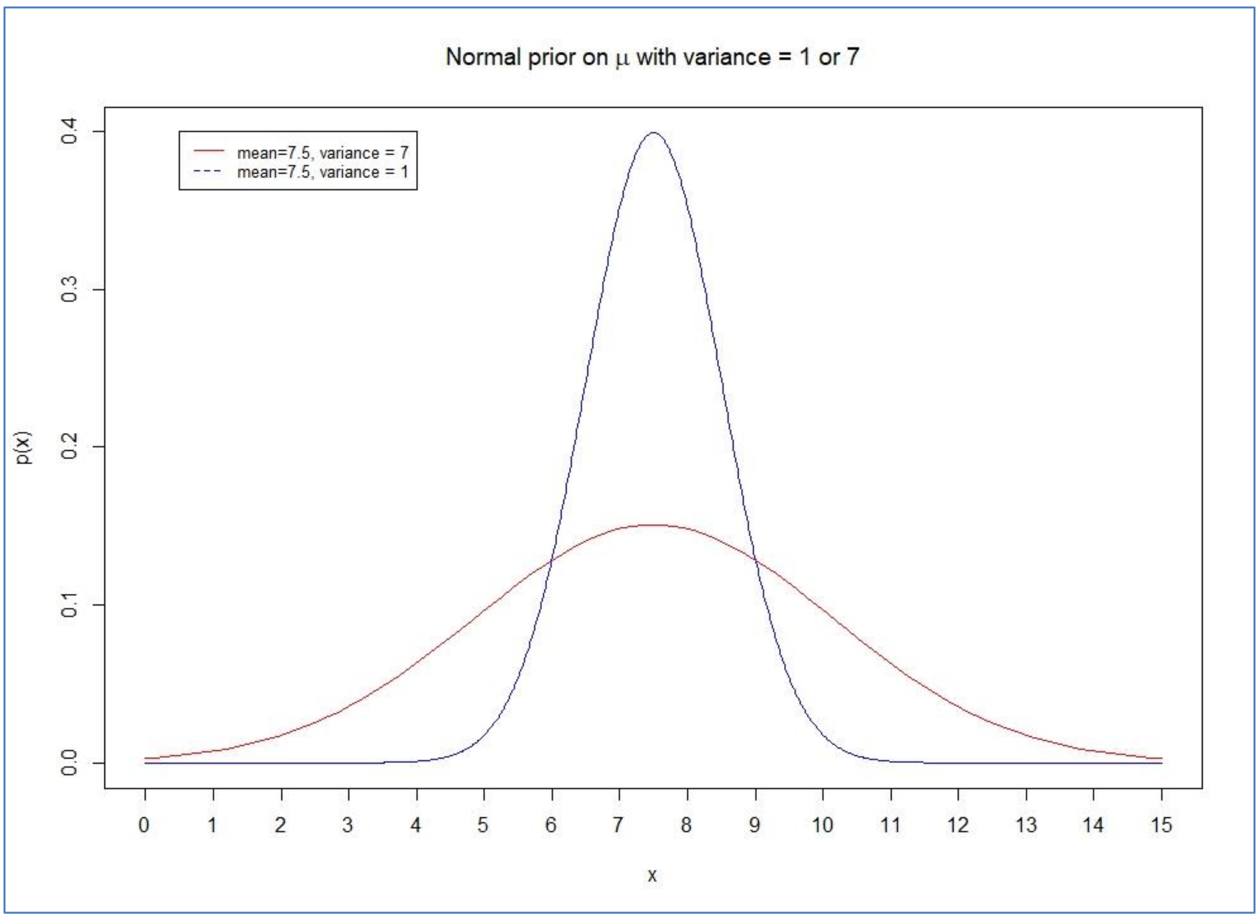
Figure 1 - Distribution of likelihood



iii. Normal prior on µ

For all distribution, decrease in variance (or standard deviation <SD>, which is the square root of variance) will increase concentration which will generate a more informative prior. This is more illustrative for normal distribution, as we have the mean (location of the prior) as the point in x-axis of the graph and variance (dispersion of the prior) is showing directly as the width of the curve, from Figure 2, the curve is more flatten (thus, less concentrated and less informative) when variance is 7 (in red) compare to variance is 1 (in blue). (Refer to [A4] on the R codes.)

Figure 2 - Normal prior on µ with variance = 1 or 7



iv. Gamma prior on τ

For gamma distribution, it is not as simply illustrated as normal distribution, as mean and variance are not shown directly from the plots. For example in Figure 4, when SD=15 (in green), the curve is more flatten, but actually it has higher concentration compares to

SD=30 (in red), we are just mislead by the visual effect on the scale of the plot. If we plot SD=15 (also in green) with SD=10 or SD=5 on the same graph as shown in Figure 3, we can see SD=15 has high concentration when the y-axis drops from 12 to 1, and x-axis extends from 2 to 120. Same applies to SD=5 (in blue), if we move the distribution into another plot and shrink the scale in y-axis and increase the scale in x-axis, it would display SD=5 has the highest concentration among all the other SDs being shown. (Refer to [A5] on the R codes.)

Thus,

Variance ↓, concentration ↑, degree of belief ↑

Figure 4 - Gamma prior on τ with S.D. = 30, 20 or 15

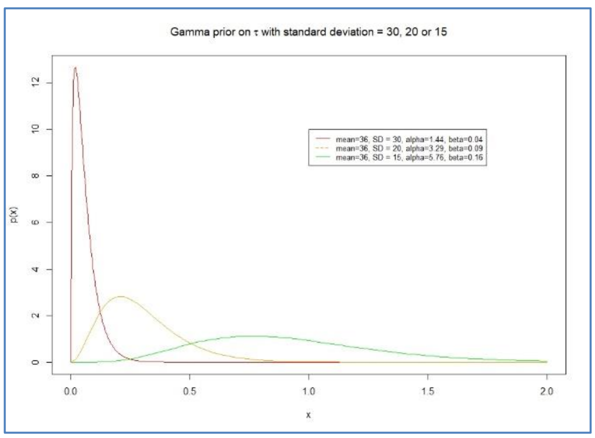
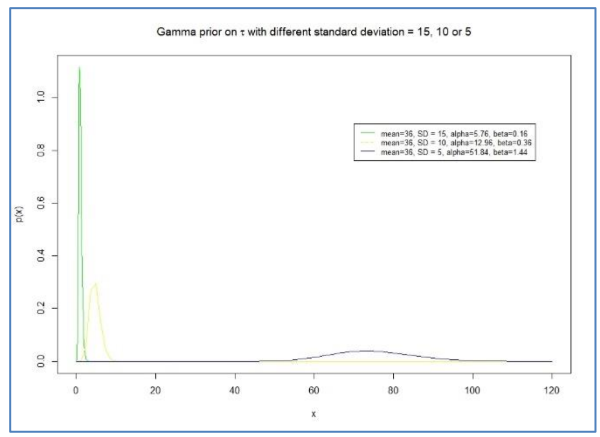


Figure 3 - Gamma prior on τ with S.D. = 15, 10 or 5

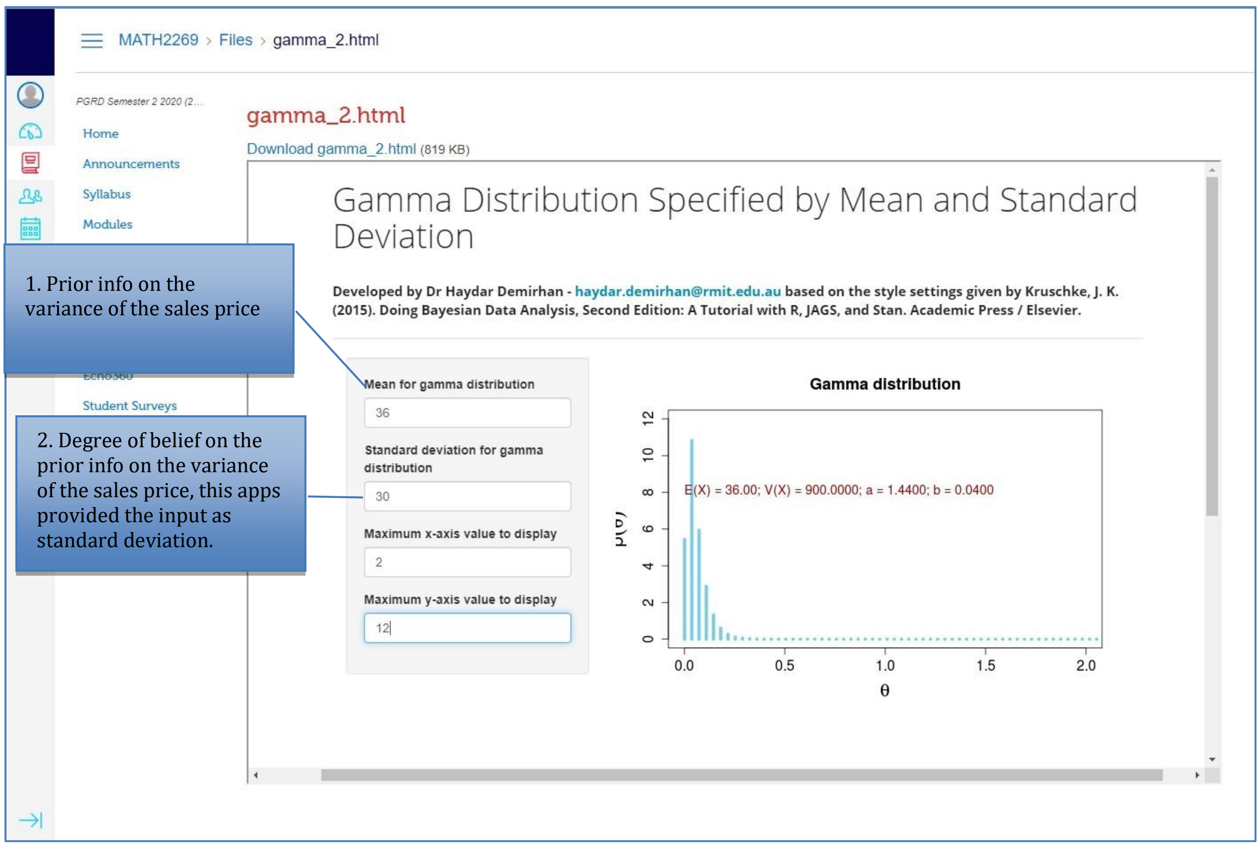


iv. Apps to model the posterior distribution

We need to use 2 apps to model the posterior distribution:

First, the Gamma Distribution Specified by Mean and Standard Deviation app is used to work out the α and β of the gamma prior on τ. When we input the prior knowledge of the variance (we need to input variance instead of standard deviation as the mean of the gamma distribution, because the second app is written to show population variance, refer to Figure 6 point 4 & 5) of sales price (36, square of the standard deviation, which is given as AUD$ 600,000 and is equivalent to 6 in the unit of AUD$100,000) as the location (mean) of gamma prior on τ, and the dispersion (standard deviation) of this gamma prior on τ, this apps will generate the respective α and β as shown in Figure 5 .

Figure 5 - The Gamma Distribution Specified by Mean and Standard Deviation app



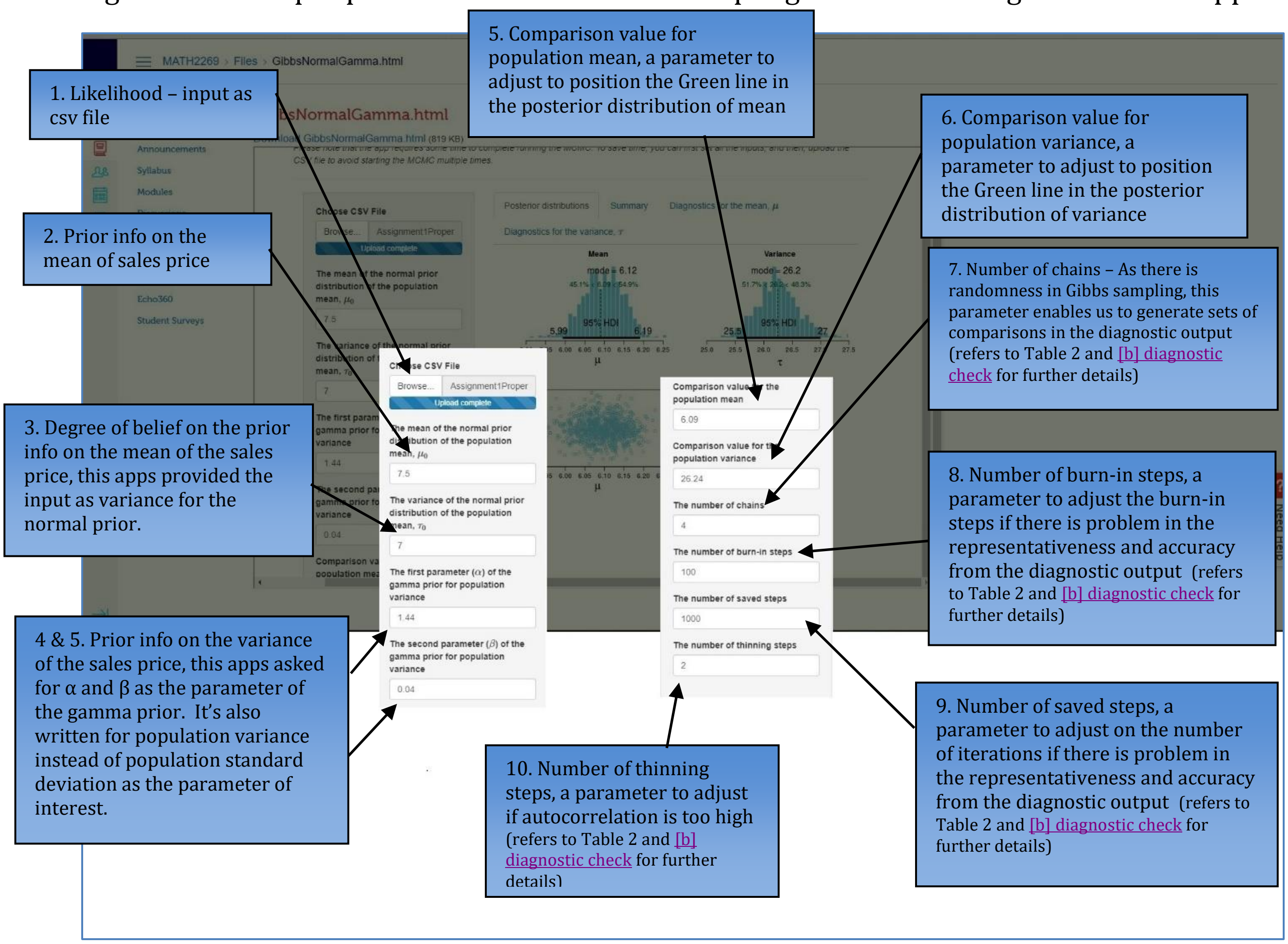
Second, the Gibbs sampling for the normal-gamma model app, we need to input the followings in order to generate the posterior distribution:

1. Likelihood as the given csv file.
2. μ_0, The location (mean) of normal prior on μ, would just be the prior knowledge of the mean of sales price (7.5, which is AUD$ 750,000), it is the μ_0 of the Normal(μ_0, τ_0) prior distribution.

3. τ_0, The dispersion (variance) of normal prior on μ, it is denoted as τ_0 of the Normal(μ_0, τ_0) prior distribution.
4. α, worked out from the first app.
5. β, also worked out from the first app.

And other comparison and tuning parameters as described below.

Figure 6 - The input parameters for the Gibbs sampling for the normal-gamma model app

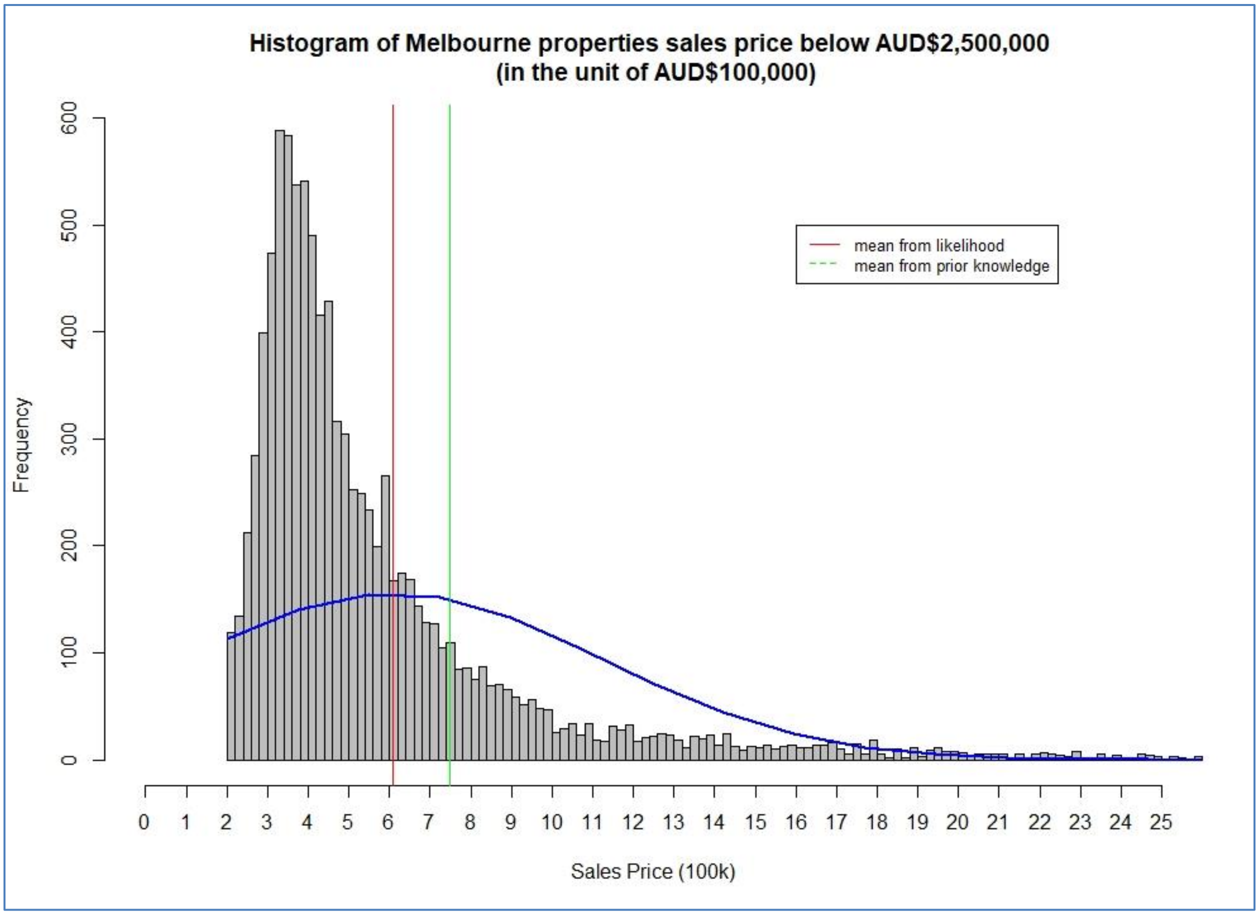


As we know that,

$$\text{Posterior} \propto \text{Prior} * \text{Likelihood}$$

If prior is less informative (less dominant), posterior will have the tendency to be shaped towards likelihood, if prior is more informative (more dominant), posterior will tend to follow prior.

Figure 7 - Comparing the mean from likelihood with the mean from prior knowledge



Thus, if our posterior is not affected by the prior info, the mean would be very close to 6.09 (without rounding is 6.093602), standard deviation would be very close to 5.12 (without rounding is 5.123238, variance is close to $5.123238^2 = 26.24756$).

We need to find out what is the degree of belief (in terms of variance) of our normal prior on μ to make the posterior mean jump from 6.09 towards 7.5 (as shown in Figure 7, refer to [A6] on the R codes) and the degree of belief (in terms of standard deviation, then work out α and β) of our gamma prior on τ to make the posterior variance jump from 26.24 to 36.

3. Steps to generate results

i. Non informative prior

[a] Simulation

I have chosen the following values for my non-informative normal prior on μ and gamma on τ as follows, parameters 1 – 10 are the input for the Gibbs sampling for the normal-gamma model app , parameters A1 and A2 are the input for the Gamma Distribution Specified by Mean and Standard Deviation app:

Table 2 - Input parameters for non-informative prior

	Parameters	Values	Reasons
1	μ_0, The location (mean) of normal prior on μ	7.5	This is given prior info on mean
2	τ_0, The dispersion (variance) of normal prior on μ	7	Simply rounding down 7.5 as an integer value to make a non-informative prior as this is a relative big variance compare to μ_0 (explained in Figure 2), pending to adjust if the posterior distribution is too far away from parameter 5
A1	The location (mean) of gamma prior on τ	36	The given prior info on standard deviation is 6, which makes the variance = 36, use this to generate parameter 3 and 4 as described in Figure 5
A2	The dispersion (standard deviation) of gamma prior on τ	30	Simply rounding down 36 to the closest tens value place to make a non-informative prior as this is a relative big S.D. compare to 36, (explained in Figure 4 and Figure 3), use this to generate parameter 3 and 4 as described in Figure 5, pending to adjust if this is too far away from parameter 6.
3	α of the gamma prior	1.44	Derived from step A1 and A2
4	β of the gamma prior	0.04	Derived from step A1 and A2
5	Comparison value for population mean	6.09	This is the mean value from likelihood
6	Comparison value for population variance	26.24	This is the variance from likelihood
7	Number of chains	4	Default value, pending to adjust after diagnostic check
8	Number of burn-in steps	100	My guess on the number of samples that I should disregard before getting stable into the high density region for the posterior simulation, pending to adjust after diagnostic check
9	Number of saved steps	1000	My guess on the number of samples that should be enough for getting good results for

			representativeness and accuracy on the posterior simulation, pending to adjust after diagnostic check
10	Number of thinning steps	2	Default value, pending to adjust if there is high autocorrelation in the diagnostic check

We would get 4 tabs after the simulation is done:

Tab 1. The posterior distributions (Figure 8)

Tab 2. The summary on the posterior distributions (Figure 9)

Figure 8 - posterior distribution of non-informative prior (first trial)

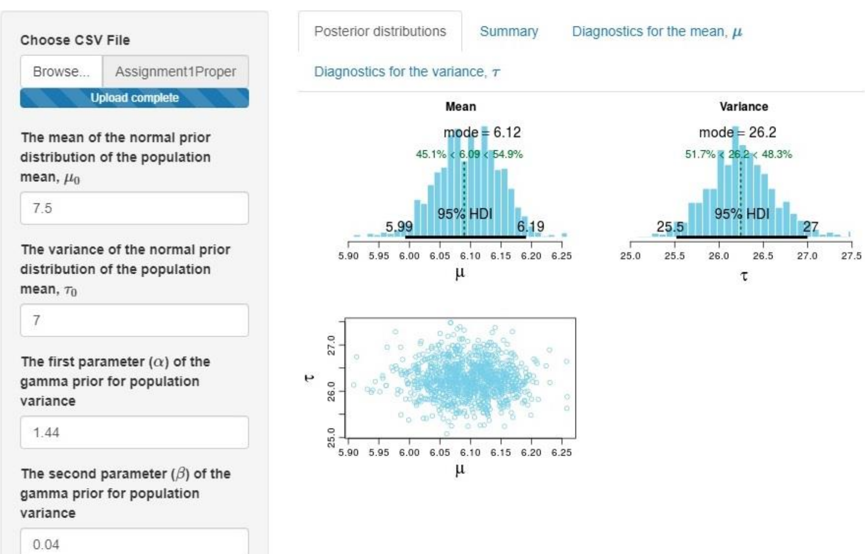


Figure 9 - summary on posterior distribution of non-informative prior (first trial)

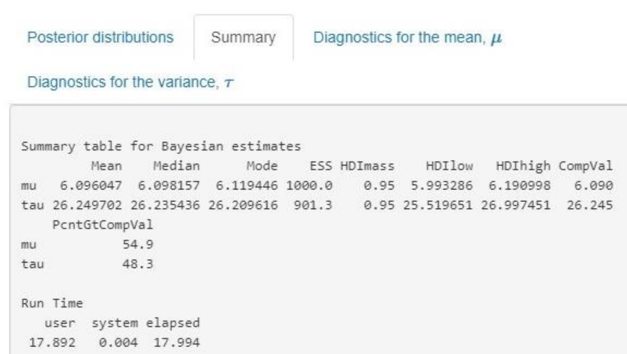


Figure 8 and Figure 9 show us that the mode and mean of the posterior distribution for μ is 6.119 and 6.096047, mode and mean of the posterior distribution for τ is 26.209 and 26.249702, probability is 45.1% for the posterior distribution for μ to be less than 6.09, and 54.9% to be higher than 6.09. Probability is 51.7% for the posterior distribution for τ to be less than 26.24, and 48.3% to be higher than 26.24.

95% of HDI interval shows that the probability of the posterior distribution for μ will be in between 5.9932 and 6.1909 is 0.95, the probability of the posterior distribution for τ will be in between 25.519 and 26.997 is 0.95. If we change parameter 5 and 6 (refers to Table 2) to the value of the prior info. Refer to Figure 11 and Figure 10, probability is 0% for the posterior distribution on μ and τ to be greater 7.5 and 36 respectively.

Figure 11 - posterior distribution of non-informative prior (first trial) compare with prior information

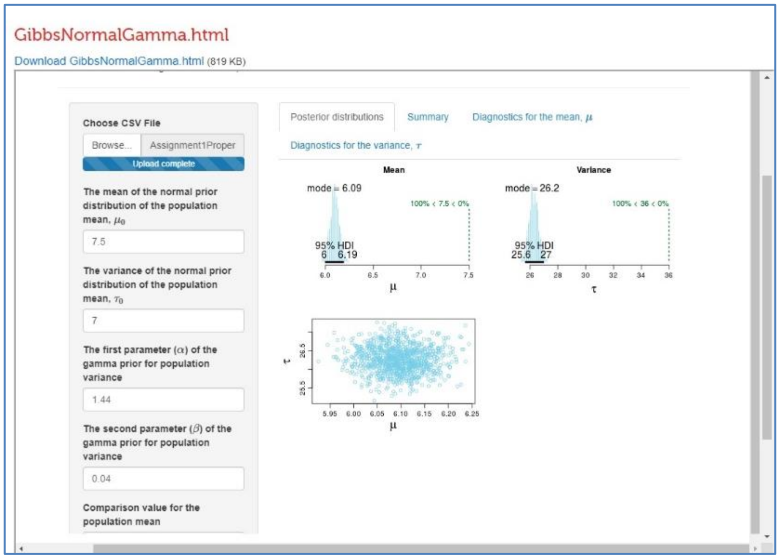


Figure 10 - summary of posterior distribution of non-informative prior (first trial) compare with prior information

With this parameter settings, the mean and variance of posterior distribution is much closer to the likelihood, and is far away from the prior knowledge. We could consider our parameter 2 and A2 is larger enough to set the prior as non-informative.

Tab 3. Diagnostics for the normal prior on μ (Figure 13)

Tab 4. Diagnostics for the gamma prior on τ (Figure 12)

Figure 13 - Diagnostics for normal prior on μ for non-informative prior (first trial)

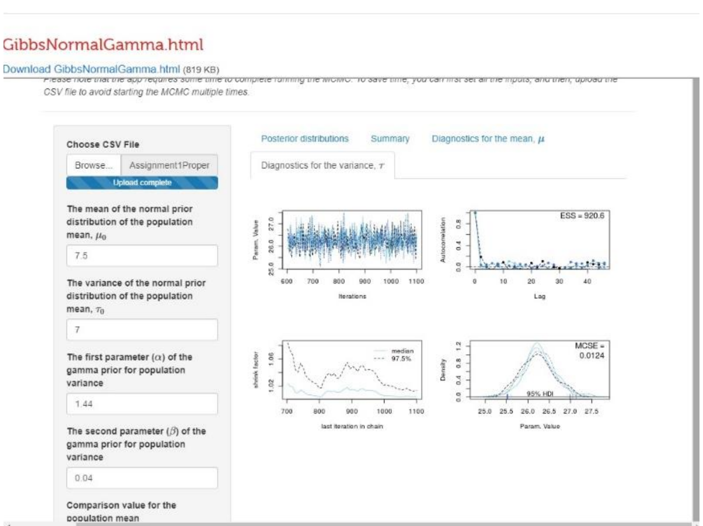
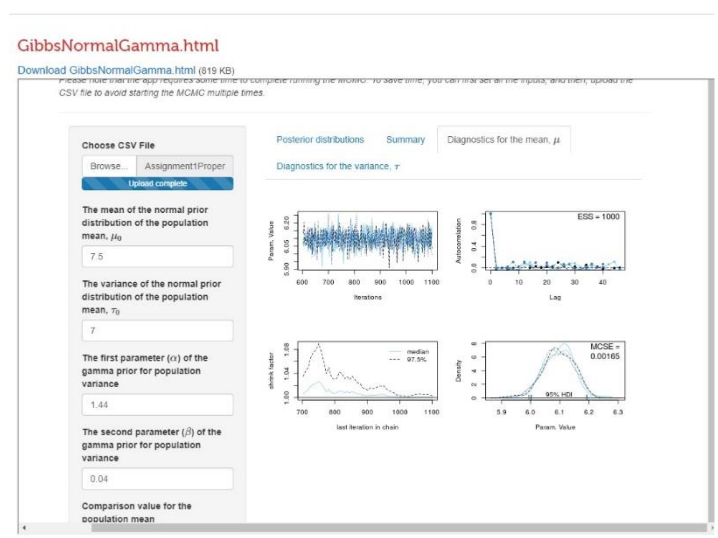


Figure 12 - Diagnostics for the gamma prior on τ for non-informative prior (first trial)



[b] Diagnostic Check in representativeness and accuracy

Figure 12 and Figure 13, tell us that:

1) for both normal prior on μ and gamma prior on τ, 4 chains are all overlapping for the param values in the later iterations (top-left plot),
2) there are almost no autocorrelations in the chains (from the top-right plot), ESS is close to number of saved steps, which implies we do not have to change the number of thinning steps).
3) shrink factor is always less than 1.2 (lower-left plot), which implies there are no orphaned or stuck chains
4) However, in the density plot (lower-right plot), the shape and HDI interval coverage of different chains seem do not overlap very well. I have tried to adjust Number of burn-in steps and Number of saved steps to 500 and 5000, we get some significant improvement as shown in Figure 14 and Figure 15.

Figure 14 - Diagnostics for normal prior on μ for non-informative prior (first trial) with increased steps

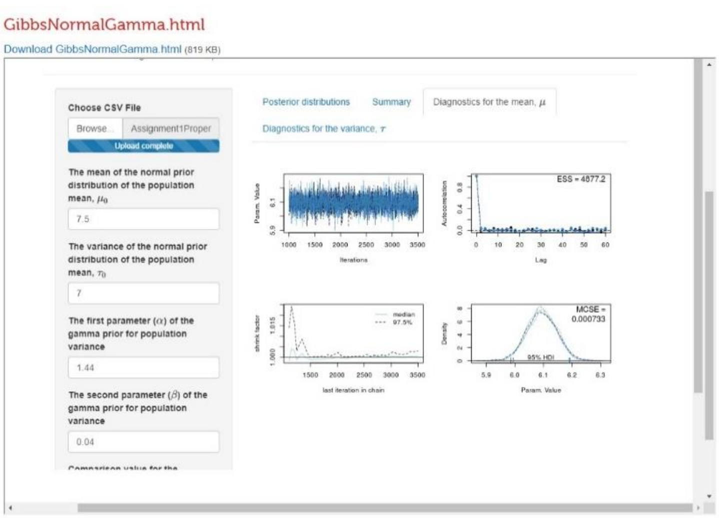
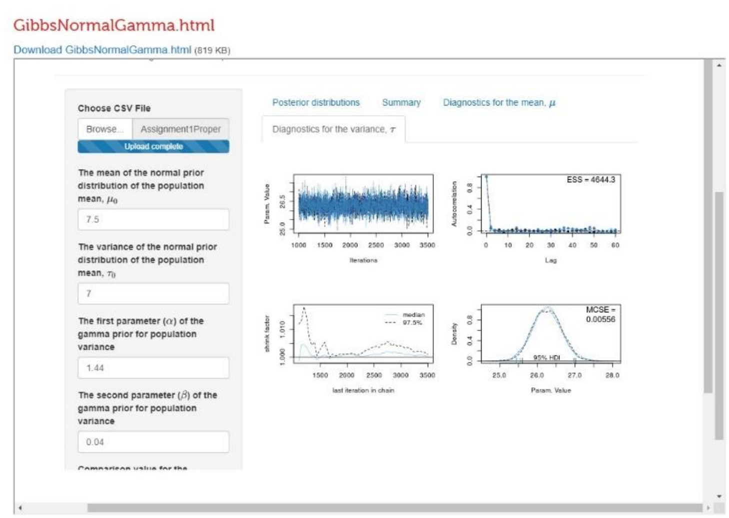


Figure 15 - Diagnostics for the gamma prior on τ for non-informative prior (first trial) with increased steps



ii. Informative prior

[a] Simulation

I have run extensive trials in the apps with the following parameters to search for informative priors:

Table 3 - parameters search for informative prior

	Input values								Posterior Results					
Trials	Normal prior μ_0	Normal prior τ_0	Gamma prior Mean	Gamma prior S.D	Gamma prior α	Gamma prior β	Burin-in Steps	Saved Steps	Posterior Mean for μ	Lower Bound of HDI for μ	Upper Bound of HDI for μ	Posterior Mean for τ	Lower Bound of HDI for τ	Upper Bound of HDI for τ
1	7.5	7	36	30	1.44	0.04	100	1000	6.096	5.993	6.191	26.25	25.52	26.997
2	7.5	7	36	30	1.44	0.04	500	5000	6.095	5.993	6.191	26.255	25.503	26.98
3	7.5	7	36	20	3.24	0.09	500	5000	6.094	5.995	6.195	26.277	25.569	26.993
4	7.5	7	36	15	5.76	0.16	500	5000	6.093	5.992	6.191	26.252	25.565	26.974
5	7.5	7	36	10	12.96	0.36	500	5000	6.095	5.995	6.192	26.272	25.549	26.976
6	7.5	7	36	5	51.84	1.44	500	5000	6.093	5.986	6.186	26.323	25.584	27.086
7	7.5	7	36	4.5	64	1.778	500	5000	6.094	5.995	6.194	26.359	25.678	27.102
8	7.5	7	36	4	81	2.25	500	5000	6.094	5.997	6.192	26.374	25.655	27.118
9	7.5	7	36	3.5	105.796	2.939	500	5000	6.093	5.991	6.189	26.389	25.641	27.133
10	7.5	7	36	3	144	4	500	5000	6.095	5.993	6.191	26.457	25.764	27.205
11	7.5	7	36	2.5	207.36	5.76	500	5000	6.096	5.99	6.193	26.541	25.839	27.263
12	7.5	7	36	2	324	9	500	5000	6.095	5.99	6.188	26.705	25.973	27.447
13	7.5	7	36	1.5	576	16	500	5000	6.096	5.982	6.191	27.045	26.367	27.791
14	7.5	7	36	1	1296	36	500	5000	6.093	5.988	6.195	27.88	27.172	28.587
15	7.5	7	36	0.5	5184	144	500	5000	6.093	5.98	6.196	30.827	30.17	31.458
16	7.5	1	36	0.5	5184	144	500	5000	6.098	5.996	6.206	30.837	30.19	31.456
17	7.5	0.1	36	0.5	5184	144	500	5000	6.136	6.033	6.248	30.829	30.166	31.441
18	7.5	0.01	36	0.5	5184	144	500	5000	6.425	6.332	6.521	30.898	30.252	31.56
19	7.5	0.001	36	0.5	5184	144	500	5000	7.161	7.106	7.215	31.485	30.811	32.13
20	7.5	0.005	36	0.5	5184	144	500	5000	7.308	7.267	7.348	31.672	31.034	32.33
21	7.5	0.0005	36	0.1	129600	3600	500	5000	7.328	7.286	7.367	35.690	35.506	35.880
22	7.5	0.0005	36	0.1	129600	3600	2000	20000	7.327	7.287	7.370	35.692	35.504	35.886
23	7.5	0.0001	36	0.1	129600	3600	2000	20000	7.462	7.442	7.480	35.705	35.523	35.905
24	7.5	0.00005	36	0.1	129600	3600	3000	30000	7.480	7.466	7.494	35.707	35.516	35.903

From Table 3, it shows that the posterior mean for μ only go over 7 when the normal prior $\tau_0 \leq 0.001$ and τ will go over 30 when gamma prior S.D. ≤ 0.5. There is one interesting fact that the HDI interval decreases when the prior become more informative. (e.g. HDI interval for μ is around 0.2 (6.191-5.993) at trial 1, and go down to 0.03 (7.494-7.466) at trial 24, HDI interval for τ is around 1.48 (26.997-25.52) at trial 1, and go down to 0.39 (35.903-35.516) at trial 24), All the diagnostic plots satisfy the requirements as stated on page 13 until trial 21, we would discuss in details at case 2 of the diagnostic section.

Figure 16 - Posterior distribution of Informative prior (trial 24)

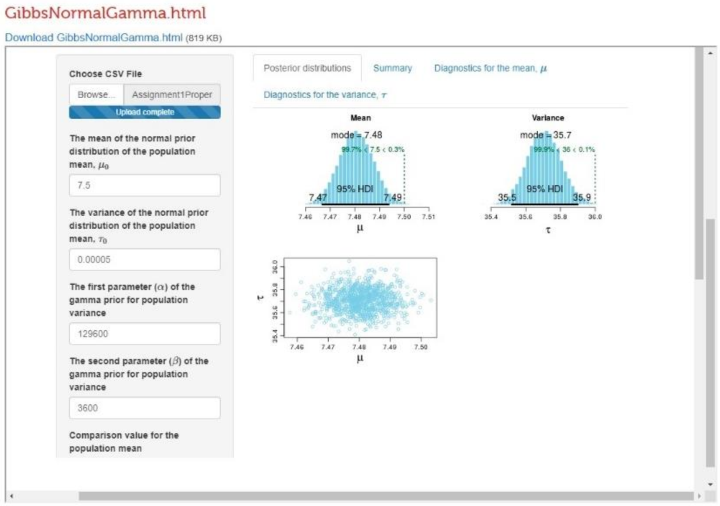


Figure 17 - Summary of Posterior distribution of Informative prior (trial 24)

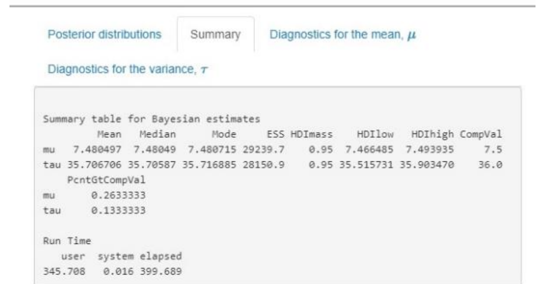


Figure 16 and Figure 17, show us that the mode and mean of the posterior distribution for μ is 7.480715 and 7.48097, mode and mean of the posterior distribution for τ is 35.716885 and 35.706706, probability is 99.7% for the posterior distribution for μ to be less than 7.5, and 0.3% to be higher than 7.5. Probability is 99.9% for the posterior distribution for τ to be less than 36, and 0.1% to be higher than 36.

95% of HDI interval shows that the probability of the posterior distribution for μ will be in between 7.466485 and 7.493935 is 0.95 (the difference between 7.5 and lower bound is 0.034, upper bound is 0.006), the probability of the posterior distribution for τ will be in between 35.5157 and 35.903470 (the difference between 36 and the lower bound is 0.484, upper bound is 0.097) is 0.95.

Although there is a slight chance for the posterior distribution to go over 7.5 (for μ) and 36 (for τ), the 95% HDI interval still do not capture these values. I believe if we proceed further with normal prior on μ_0 = 0.00001 and τ_0 = 0.05 (α=518400 and β=14400), we could have a better capture for 95% HDI on μ=7.5 and τ=36. With limitation on resources in running the online apps, and the upper bound is just 0.006 away from the prior info for μ and 0.097 away from prior info for τ, we would now take trial 24 as our best posterior model for informative prior.

[b] Diagnostic Check in representativeness and accuracy

For Informative prior, we would be discussing 3 diagnostic plots for the trials highlighted in orange in Table 3.

Case 1

For trial 24, referring to Figure 18 and Figure 19, the diagnostic plots tell us that:

1) for both normal prior on μ and gamma prior on τ, 4 chains are all overlapping for the param values in the later iterations (top-left plot),
2) there are almost no autocorrelations in the chains (from the top-right plot), ESS is close to number of saved steps, which implies we do not have to change the number of thinning steps.
3) shrink factor is always less than 1.2 (lower-left plot), which implies there are no orphaned or stuck chains
4) The density plot (lower-right plot) shows the shape and HDI interval coverage of different chains overlap very well and converge.

Figure 18 - Diagnostics for normal prior on μ for informative prior (trial 24)

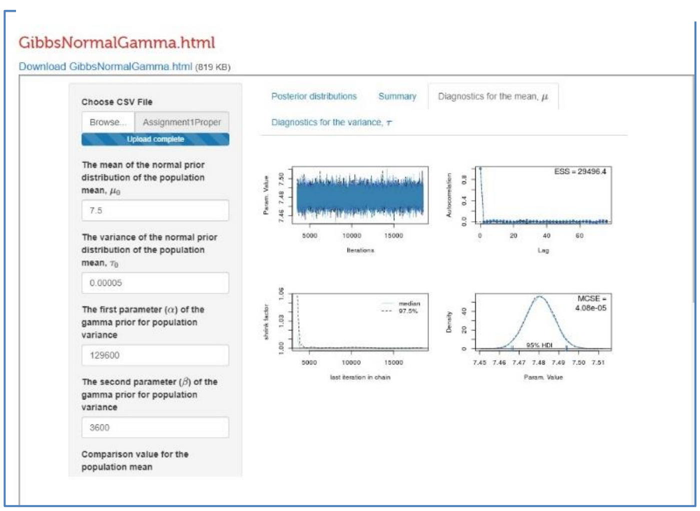
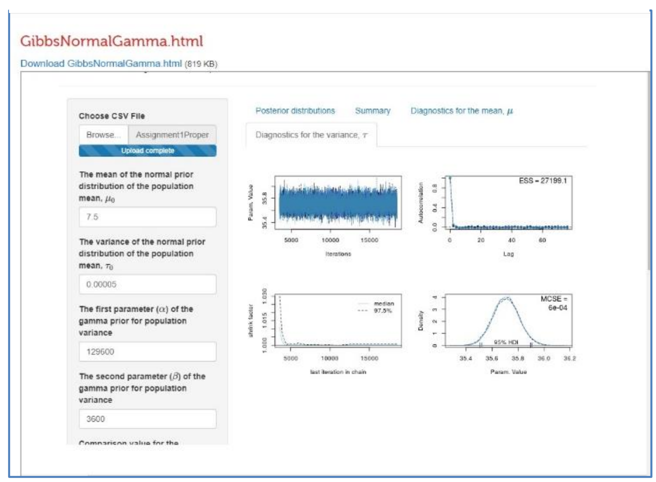


Figure 19 - Diagnostics for the gamma prior on τ for informative prior (trial 24)



But before we got these good diagnostic plots, we were using 2000 and 20,000 as the number of burn-in steps and number of saved steps for trial 24 (We call this trial 24a). It seems the chains are overlapping in the top-left and right bottom plot (refers to Figure 21 and Figure 20). However, the ESS samples is 9414.6 for μ and 8798.3 for τ, which are only around half of the number of saved steps (20,000).

Figure 21 - Diagnostics for normal prior on μ for informative prior (trial 24) with not enough iterations

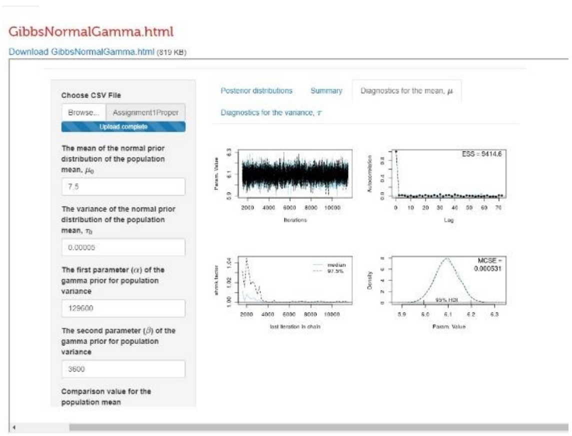
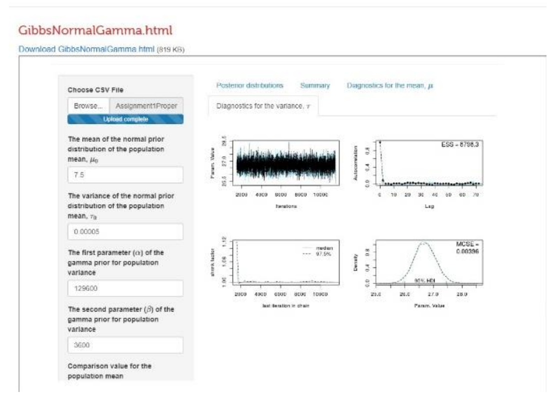


Figure 20 - Diagnostics for the gamma prior on τ for informative prior (trial 24) with not enough iterations



Also, the figures in the summary do not match with the numbers shown on the distributions tab. (refers to Figure 23 and Figure 22) This misalignment is solved if we increase the number of burn-in steps and number of saved steps to 3000 and 30,000 respectively.

Figure 23 - Posterior distribution of Informative prior (trial 24) with not enough iterations

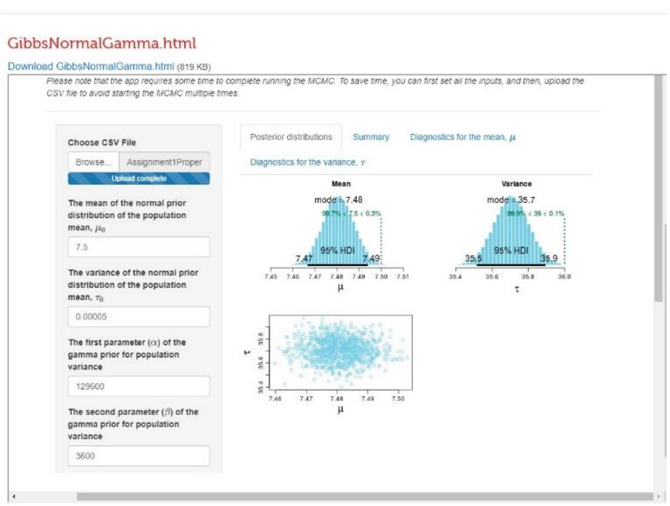
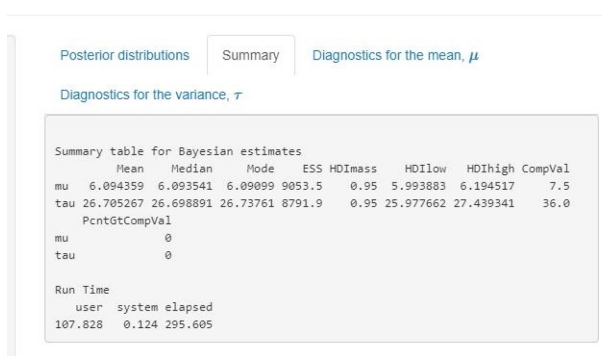


Figure 22 - summary of posterior distribution of informative prior (trial 24) with not enough iterations



Case 2

At trial 21, when we set the normal prior τ_0 to 0.0005 and gamma prior S.D. = 0.1 (α= 12960 and β=3600), referring to Figure 25 and Figure 24, the diagnostic plots tell us that:

1) for both normal prior on μ and gamma prior on τ, 4 chains do not overlap or converge for the param values in the later iterations (top-left plot),
2) there are significant autocorrelations in the chains (from the top-right plot), ESS is only 50, which is so low compare to the number of saved steps (5,000)
3) shrink factor plot (lower-left plot) did not come up, which implies the shrink factor is too large to be displayed on the app.
4) The density plot (lower-right plot) shows the shape and HDI interval coverage of different chains do not overlap.

Figure 25 - Diagnostics for normal prior on μ for trial 21

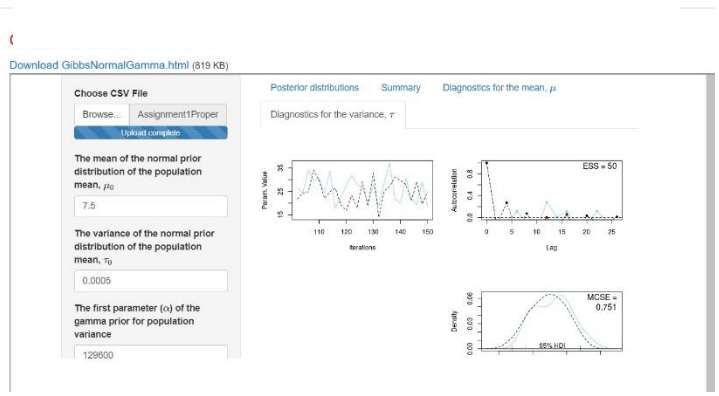
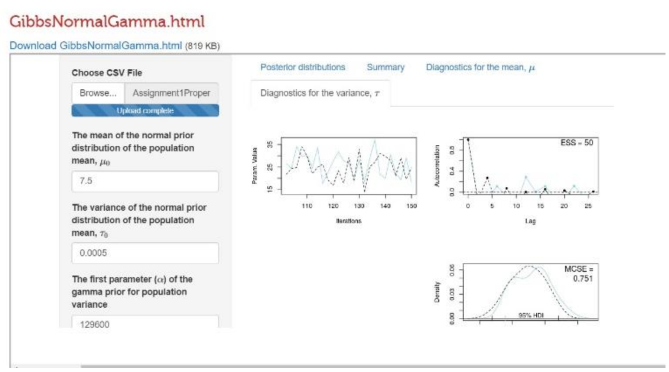


Figure 24 - Diagnostics for the gamma prior on τ for trial 21



Thus we increased the Number of burn-in steps and Number of saved steps to 2000 and 20,000, and we get some significant improvement for trial 22 as shown in Figure 26 and Figure 27.

Figure 27 - Diagnostics for normal prior on μ for trial 22

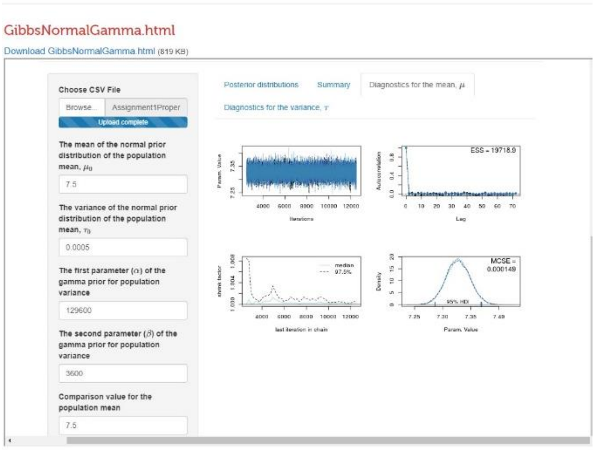
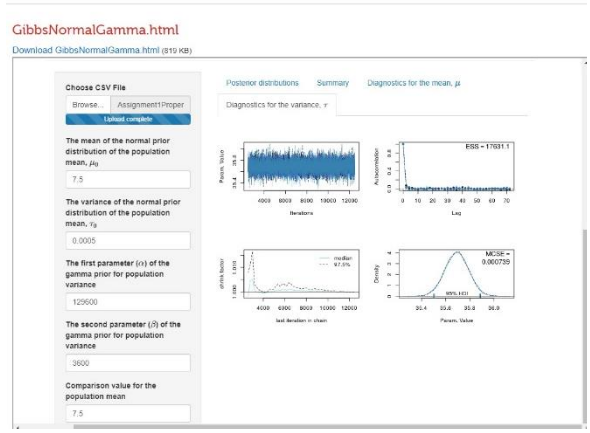


Figure 26 - Diagnostics for the gamma prior on τ for trial 22



4. Analysis

i. Hypothesis test for non-Informative prior

Refer to Figure 28 and Figure 29, 95% HDI do not capture 8.5 as the μ and 9 as the τ (standard deviation = 3) for non-informative prior, thus we reject the null hypothesis that the mean sales price is 850,000 AUD and the standard deviation is 300,000 AUD.

Figure 28 - Posterior distribution of non-Informative prior (trial 2) – compare to values of hypothesis tests

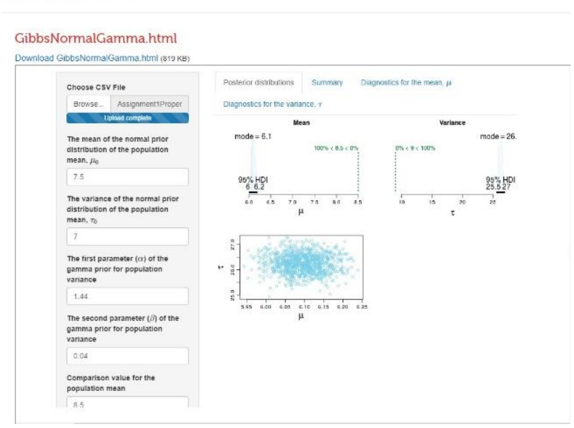


Figure 29 - summary of posterior distribution of non-informative prior (trial 2) – compare to values of hypothesis test

```
Summary table for Bayesian estimates
         Mean    Median      Mode    ESS HDImass    HDIlow  HDIhigh CompVal
mu   6.094496  6.094109  6.099443 2202.0    0.95  5.995421  6.19541     8.5
tau 26.255831 26.262544 26.281211 1696.1    0.95 25.548141 26.97417     9.0
    PcntGtCompVal
mu              0
tau           100

Run Time
   user  system elapsed
 28.416   0.000  28.496
```

ii. Hypothesis test for Informative prior

Refer to Figure 30 and Figure 31, 95% HDI do not capture 8.5 as the μ and 9 as the τ (standard deviation = 3) for informative prior, thus we reject the null hypothesis that the mean sales price is 850,000 AUD and the standard deviation is 300,000 AUD.

Figure 30 - Posterior distribution of Informative prior (trial 24) – compare to values of hypothesis tests

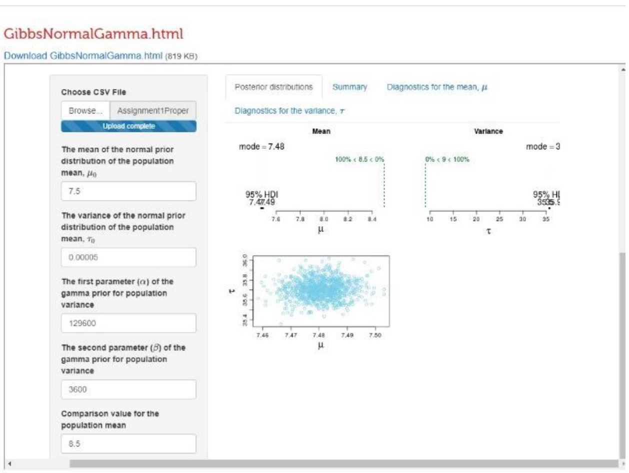


Figure 31 - summary of posterior distribution of informative prior (trial 24) – compare to values of hypothesis test

```
Summary table for Bayesian estimates
        Mean    Median      Mode     ESS HDImass    HDIlow   HDIhigh CompVal
mu   7.48067  7.480666  7.480482 30000.0    0.95  7.466995  7.494433     8.5
tau 35.70659 35.706690 35.706598 26961.4    0.95 35.510135 35.895333     9.0
    PcntGtCompVal
mu              0
tau           100

Run Time
   user  system elapsed
351.184   0.028 878.046
```

5. Discussion on Efficiency

We could decrease the number of saved steps to obtain a shorter run time, for example, in trial 21, we observed poor diagnostic plots for saved steps =5000, and after trying to set the saved steps=20,000, we could obtain excellent plots. Theoretically, we could fine tune the saved steps by slowly decreasing 20,000 to get minimum run time with acceptable diagnostic results, however, as the apps we are using is online, run-time might not be consistent (refers to the cells highlighted in orange in Table 4, we noticed there are less saved steps in trial 24a compare to trial 24, but the system run time is exceptionally high, this also applies to the elapsed time for trial 22) for parameter tuning. I believe we would encounter a more stable platform by installing JAGS locally, which makes the platform more reliable and consistent on tuning the efficiency of the posterior generation.

Table 4 - Runtime for different trials

	Input values								Run Time			
Trials	Normal prior μ_0	Normal prior τ_0	Gamma prior Mean	Gamma prior S.D	Gamma prior α	Gamma prior β	Burin-in Steps	Saved Steps	User	System	Elapsed	Refers to
21	7.5	0.0005	36	0.1	129600	3600	500	5000	66.652	0.016	66.751	Fig 32
22	7.5	0.0005	36	0.1	129600	3600	2000	20000	260.752	0.012	634.433	Fig 33
24a	7.5	0.00005	36	0.1	129600	3600	2000	20000	107.828	0.124	295.605	Fig 22
24	7.5	0.00005	36	0.1	129600	3600	3000	30000	345.708	0.016	399.689	Fig 16

Figure 32 - Summary of posterior distribution which shows the runtime of trial 21

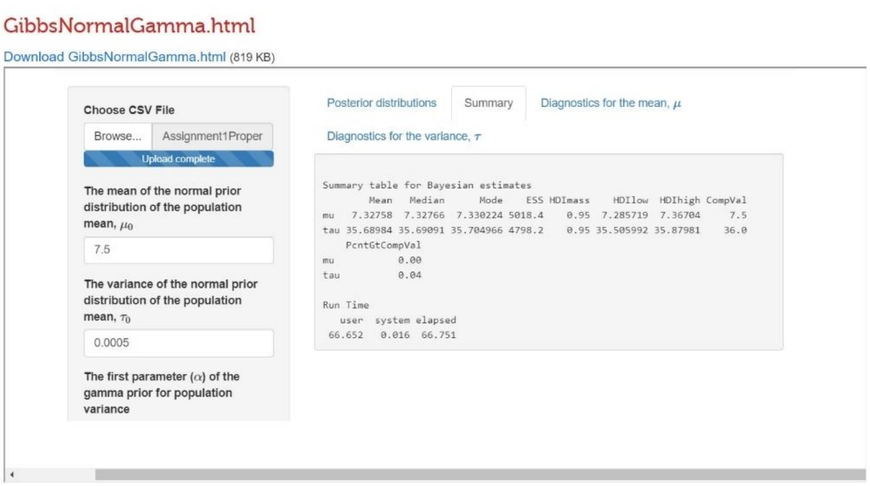
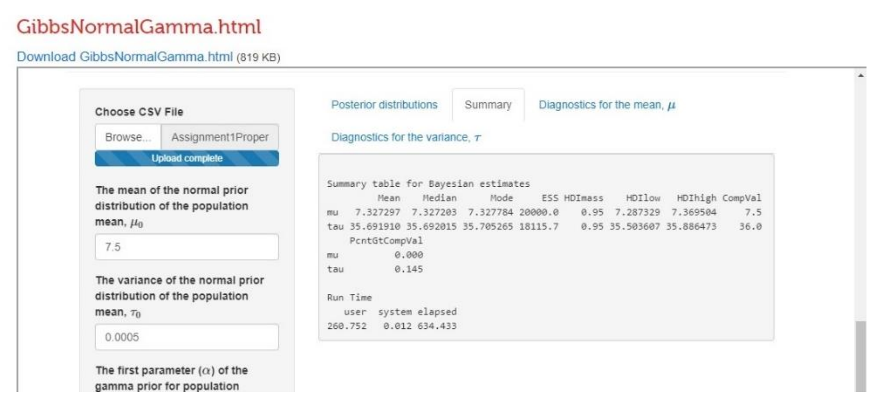


Figure 33 - Summary of posterior distribution which shows the run time of trial 22



6. Conclusion

In this assignment, we have learnt how to use Gibbs sampling to model population data when there are 2 parameters (mean (μ) and variance(τ)) of interest with Bayesian analysis to perceive the theory of "Posterior ∝ Prior * Likelihood".

We first thoroughly investigated the mathematical model which we used to apply on the data and parameter of interest, then we found out the nature of the mathematical model (normal and gamma distribution), and how to apply the location and degree of belief in the domain knowledge to generate informative / non-informative prior with given apps. Finally, we observed how the posterior simulation will change from likelihood dominant to prior dominant when we adjust the variance / standard deviation of the normal and gamma prior distribution.

For both non-informative and informative prior, we reject the null hypothesis that the mean sale price is AUD 850,000 and standard deviation is AUD 300,000.

Appendix

[A1] – Import packages and data preparation

```
#The following packages are needed in this assignment:

library(knitr)

library(summarytools)

library(dplyr)
```

```
house <- read.csv("D:/RMIT Master of Analytics/semester 3/MATH2269 Applied Bayesian Statistics/Assignment 1/Assignment1PropertyPrices.csv")
```

[A2] – Generate descriptive statistics

```
names(house)[1] <- "price"
descr(house, stats = c("mean", "med", "sd", "Q1", "Q3","IQR", "min", "max"), transpose = TRUE)
```

[A3] – A histogram for the likelihood data

```
h <- house$price %>% hist(col="grey",xlab="Sales Price (100k)", main="Histogram of Melbourne properties sales price in AUD$100,000", breaks=100)

xfit<-seq(min(house$price),max(house$price),length=40)
yfit<-dnorm(xfit,mean=mean(house$price),sd=sd(house$price))
yfit <- yfit*diff(h$mids[1:2])*length(house$price)
lines(xfit, yfit, col="blue", lwd=2)

abline(v=mean(house$price),col="red")
abline(v=median(house$price),col="orange")

legend(40, 1200, legend=c("mean", "median"),
       col=c("red", "orange"), lty=1:2, cex=0.8)
```

[A4] – Showing the Effect on variance on normal distribution

```
##Distribution of normal prior on mu
x <- seq(0, 15, length=1000)
y <- dnorm(x, mean=7.5, sd=1)
plot(x, y, type="l", lwd=1, col="blue",
ylab="p(x)",main=expression(paste("Normal prior on ", mu, " with variance = 1 or 7")), xaxp=c(0,15, 15))

curve(dnorm(x, mean=7.5, sd=sqrt(7)), add = TRUE, col = "red")
legend(0.5, 0.4, legend=c("mean=7.5, variance = 7", "mean=7.5, variance = 1"),
       col=c("red", "blue"), lty=1:2, cex=0.8)
```

[A5] – Showing the Effect on standard deviation on gamma distribution

```r
##Distribution of gamma prior on tau

x <- seq(0, 2, length=1000)
y <- dgamma(x = x, shape = 1.44, rate = 1/0.04)
plot(x, y, type="l", lwd=1, col="red", ylab="p(x)",
main=expression(paste("Gamma prior on ", tau, " with standard deviation = 30,
20 or 15")))
curve(dgamma(x = x, shape = 3.29, rate = 1/0.09), add = TRUE, col = "orange")
curve(dgamma(x = x, shape = 5.76, rate = 1/0.16), add = TRUE, col = "green")
legend(1, 10, legend=c("mean=36, SD = 30, alpha=1.44, beta=0.04", "mean=36,
SD = 20, alpha=3.29, beta=0.09",
                         "mean=36, SD = 15, alpha=5.76, beta=0.16"),
col=c("red", "orange", "green"),lty=1:2, cex=0.8)

x <- seq(0, 120, length=1000)
y <- dgamma(x = x, shape = 5.76, rate = 1/0.16)
plot(x, y, type="l", lwd=1, col="green", ylab="p(x)",
main=expression(paste("Gamma prior on ", tau, " with different standard
deviation = 15, 10 or 5")))
curve(dgamma(x = x, shape = 12.96, rate = 1/0.36), add = TRUE, col =
"yellow")
curve(dgamma(x = x, shape = 51.84, rate = 1/1.44), add = TRUE, col = "blue")
legend(70, 0.9, legend=c("mean=36, SD = 15, alpha=5.76, beta=0.16", "mean=36,
SD = 10, alpha=12.96, beta=0.36",
                       "mean=36, SD = 5, alpha=51.84, beta=1.44"),
col=c("green", "yellow", "blue"),lty=1:2, cex=0.8)
```

[A6] – A histogram for the likelihood data with comparison on the µ of likelihood and prior info on µ

```r
##Histogram of likelihood

h <- house$price %>% hist(col="grey",xlab="Sales Price (100k)",
main="Histogram of Melbourne properties sales price below AUD$2,500,000 \n
(in the unit of AUD$100,000)", breaks=300, xaxp=c(0,25,25), xlim=c(0,25))

xfit<-seq(min(house$price),max(house$price),length=40)
yfit<-dnorm(xfit,mean=mean(house$price),sd=sd(house$price))
yfit <- yfit*diff(h$mids[1:2])*length(house$price)
lines(xfit, yfit, col="blue", lwd=2)

abline(v=mean(house$price),col="red")
abline(v=7.5,col="green")
```

```
legend(16, 500, legend=c("mean from likelihood", "mean from prior
knowledge"),
       col=c("red", "green"), lty=1:2, cex=0.8)
```