UCLouvain

Faculté des sciences

Université catholique de Louvain
Louvain School of Statistics

LSTAT2130 - Bayesian Statistics

Project - Group Q

Lionel Lamy - 1294-1700
Adrien Kinart
Simon Lengendre

May 22, 2021

Contents

Introduction

In this work, the Net Monthly Income (HNI) of household older than 30 years is studied across the two Belgian regions. These regions are denoted by $k = \{1, 2\}$ with respect to Flanders and Wallonia, respectively. The 1228 households were listed with respect to 10 income intervals. The detailed frequency table is given below:

	Net monthly household income (in euros)										
Region	<1200	[1200, 1500)	[1500,1800)	[1800,2300)	[2300,2700)	[2700,3300)	[3300,4000)	[4000,4900)	[4900,6000)	≥ 6000	Total
Flanders	25	69	65	106	80	106	136	94	76	46	803
Wallonia	17	36	47	58	47	53	59	54	33	21	425

Table 1: Frequency table of the survey

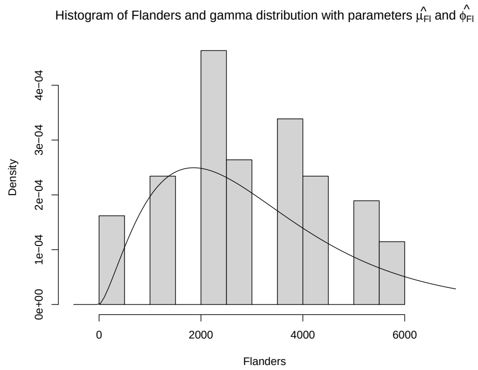
Let X be the HNI regardless the 2 regions , it is assumed it follows a Gamma distribution with parameters κ and λ. It can be reparametrised in terms of its mean μ and dispersion parameter ϕ with the following trick:

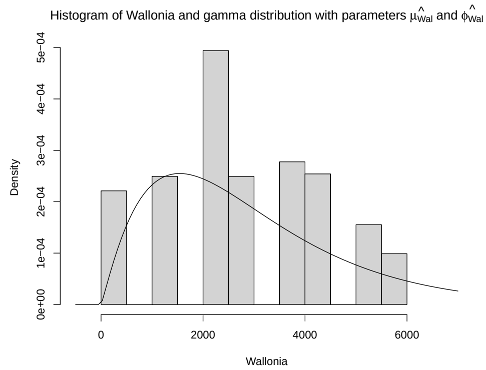
$$\text{shape: } \kappa = \frac{1}{\phi}$$
$$\text{rate: } \lambda = \frac{1}{\phi\,\mu}$$

By assuming the gamma distribution, one can plot their experimental histogram and plot the theoretical density plot by estimating the shape and rate parameters with Maximum of Likelihood. This allows to get a first sight on the data behaviour at hand.

Table 2: Estimated μ and ϕ for Flanders and Wallonia

	Flanders	Wallonia
$\hat{\mu}$	3089.944	2914.882
$\hat{\phi}$	0.399687	0.471615





1 Question 1

Let $\theta_k := (\mu_k, \phi_k)$ be the set of parameters for a HNI with respect to region k.

1.1 (a) Theoretical probability

As expressed earlier, the distribution of monthly net income is described by a distribution, in terms of μ and ϕ, for both region $k = \{1, 2\}$: this gives

$$f(x_k) = \frac{(\phi_k \mu_k)^{-1/\phi_k}}{\Gamma(\phi_k^{-1})} x_k^{1/\phi_k - 1} \exp\left(\frac{-x_k}{\phi_k \mu_k}\right) \tag{1}$$

The cumulative distribution function is defined as :

$$F(x) = \int_0^x f(u)\ \mathrm{d}u = \frac{\gamma(\phi^{-1}, \frac{x}{\phi\mu})}{\Gamma(\phi^{-1})}$$

where $\Gamma(a)$ and $\gamma(a, b)$ are the complete gamma and lower incomplete gamma functions, defined as:

$$\gamma(a, b) = \int_0^b t^{a-1} \exp(-t) \mathrm{d}t$$
$$\Gamma(a) = \gamma(a, \infty)$$

Then, the probability to fall into a certain HNI interval I, for each region, is:

$$P(x \in I_j) = \int_{I_j} \frac{(\phi\mu)^{\frac{-1}{\phi}}}{\Gamma(\phi^{-1})} x^{\frac{1}{\phi} - 1} \exp\left(\frac{-x}{\phi\mu}\right)\ \mathrm{d}x$$

Using CDF writing , the probability can be proposed as a difference of CDF :

$$p_j = \begin{cases} F(x_j) & \text{if} \quad j = 1 \\ F(x_{j+1}) - F(x_j) & \text{if} \quad j \in \{2, .., 9\} \\ 1 - F(x_j) & \text{if} \quad j = 10 \end{cases}$$

$$p_j = \begin{cases} \frac{\gamma(\phi^{-1}, \frac{x_j}{\phi\mu})}{\Gamma(\phi^{-1})} & \text{if} \quad j = 1 \\ \frac{1}{\Gamma(\phi^{-1})} \Big(\gamma(\phi^{-1}, \frac{x_{j+1}}{\phi\mu}) - \gamma(\phi^{-1}, \frac{x_j}{\phi\mu})\Big) & \text{if} \quad j \in \{2, .., 9\} \\ 1 - \frac{\gamma(\phi^{-1}, \frac{x_j}{\phi\mu})}{\Gamma(\phi^{-1})} & \text{if} \quad j = 10 \end{cases}$$

1.2 (b) Theoretical expression for the likelihood

Since the frequency distribution in a given region is assumed multinomial,

We have, writing $P := (p1, .., p_{10})$ and $Y := (Y_1, ...Y_{10})$:

$$
\begin{aligned}
Y|P \sim \mathrm{Mul}(Y, P) &= \frac{(\sum y_i)!}{y_1! \dots y_{10}!}\, p_1^{y_1} \times ... \times p_{10}^{y_{10}} \text{ when} \sum_{j=1}^{10} p_j = 1 \\
&= 0 \text{ otherwise}
\end{aligned}
$$

For such a distribution, the likelihood L is well known. Indeed, up to a multiplicative constant, for a region k, it is given by:

$$L(P_k, Y_k) = P(Y_k|P_k) \propto \prod_{j=1}^{10} p_{k,j}^{y_{k,j}}$$

To translate this likelihood in terms of μ_k and ϕ_k, one can use the definition of the probability referring to the difference of CDF (see equation 1.1). By substituting $p_{k,j}$ with it and writing $[0; x_1], [x_1; x_2], ..., [x_9; x_{10}], [x_{10}, +\infty]$ as $I_1, I_2, ..., I_{10}$, the likelihood function becomes, up to a multiplicative constant, as follows:

$$L(\mu_k, \phi_k, Y_k) \propto \prod_{j=1}^{10} \left(\frac{1}{\Gamma(\phi^{-1})} \int\limits_{\frac{I_j}{\phi_k \mu_k}} x_j^{(\phi_k^{-1}-1)} e^{-x} \mathrm{d}x \right)^{y_{k,j}}$$

and then taking the logarithm, we obtain the log-likelihood ℓ for a certain region:

$$\ell(\mu_k, \phi_k, Y_k) = \sum_{j=1}^{10} y_{k,j} \times \ln(p_{k,j})$$

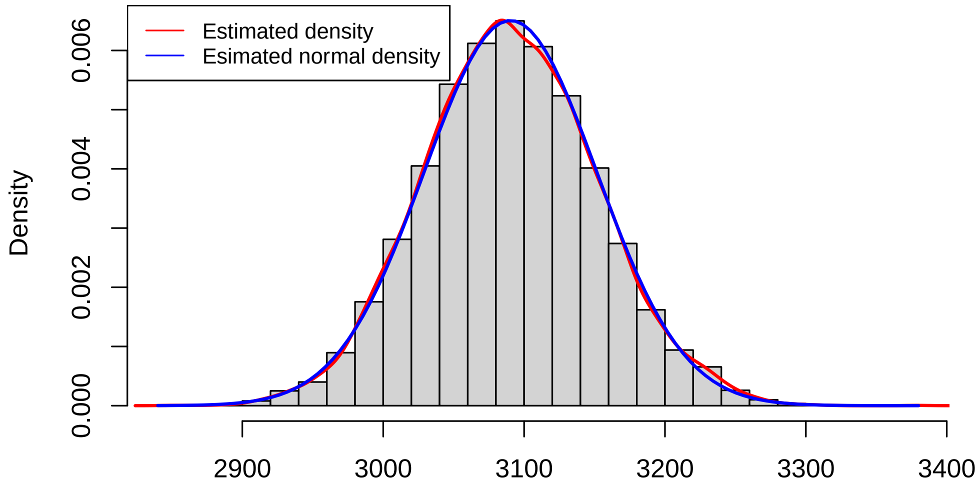
where $y_{k,j}$ is the frequency of households in the interval j in the region k, and we recall that $p_{k,j}$ corresponds to the probability, or the area of the j^{th} interval in the region k. Again, by substituting with the difference of CDF, the log-likelihood becomes:

$$\ell(\mu_k, \phi_k, Y_k) = \sum_{j=1}^{10} y_{k,j} \times \ln \left(\frac{1}{\Gamma(\phi_k^{-1})} \int\limits_{\frac{I_j}{\phi_k \mu_k}} x^{(\phi_k^{-1}-1)} e^{-x} \mathrm{d}x \right) \tag{2}$$

2 Question 2 : Prior beliefs

2.1 Prior for the mean

Prior to the study, the average net monthly household income, regardless the region, is believed to be within the interval (2400, 3600) on a 95% confidence level. A convenient probability function that translates this belief is the Gaussian distribution with a mean at the center and standard deviation such that its 95% confidence interval corresponds to the mentioned bonds. This belief is confirmed by Monte Carlo simulations. Indeed, if one generate a large amount of samples and repeatedly take its mean, one can determine an estimated density for its mean. It is symmetric without fat tails, which confirm the goodness of this assumption. The density is shown here below:



Hence, by assuming such a distribution for the prior, one can easily estimate its parameters. For both regions $\mu_0 = 3000$ (the center of the confidence interval). Then, to get the standard deviation, one can use the decomposition of a classical confidence interval for the mean.

$$3000 - t_{(n_k-1,1-\alpha/2)}\frac{s_k}{\sqrt{n_k}} = 2400$$
$$\rightarrow \hat{\sigma}_0 = \frac{s_k}{\sqrt{n_k}} = \frac{600}{t_{(n_k-1,1-\alpha/2)}}$$
$$\text{where: } t_{n_1-1,1-\alpha/2} \approx t_{n_2-1,1-\alpha/2} \approx 1.96$$

Therefore $\hat{\sigma}_{Fl} \approx \hat{\sigma}_{Wal} \approx 306.12$, then $\mu_k \sim N(\mu_{0,k} = 3000, \sigma_{0,k} = 306.12)$. So we get as a prior for the parameter μ, up to a multiplicative constant:

$$\pi(\mu_k) \propto \exp\big(-\frac{1}{2\sigma_0^2}(\mu_k - \mu_0)^2\big)\ \forall k \in \{1,2\} \tag{3}$$

2.2 Prior for the dispersion parameter

It is certain that ϕ_k lies in the interval $(0.0, 10.0)$ but there is no knowledge on how the probability mass is broken down. In order to reflect this prior knowledge, one can assume that the dispersion parameter is uniformly distributed with lower and upper bonds of 0 and 10, respectively, i.e. $\phi_k \sim \mathcal{U}(0, 10)$. Hence, up to a multiplicative constant, the prior probability can be represented via the indicator function as below:

$$\pi(\phi_k) \propto \mathbf{1}_{0,10}\ \forall k \in \{1, 2\}$$

2.3 The conjugate prior

Since there is no prior information on the dependence structure of the priors, it is recommended to consider them as independent in order to remain as least-informative as possible(this follows the maximal entropy principle). By doing so, the conjugate prior becomes:

$$\begin{aligned} \pi(\mu_k, \phi_k) &= \pi(\mu_k)\ \pi(\phi_k) \\ &\propto \exp\big(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\big)\mathbf{1}_{0;10} \end{aligned} \tag{4}$$

3 Question 3

3.1 Question 3.a: Posterior for Flanders

Since the likelihood and prior for μ_k and ϕ_k have been defined, a joint posterior can be expressed as their product. The joint posterior for Flanders, i.e. of $\theta_1 = (\mu_1, \phi_1)$ is defined, up to a multiplicative constant, here below:

$$P(\mu_1, \phi_1|Y_1) \propto \left(\prod_{j=1}^{10}\left(\frac{1}{\Gamma(\phi_1^{-1})}\int_{\frac{I_j}{\phi_1\mu_1}} x^{(\phi_1^{-1}-1)}e^{-x}\mathrm{d}x\right)^{y_{1,j}}\right)\exp\big(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\big)1_{0;10}$$

3.2 3.b: Compute log-posterior

Computing the logarithm of this expression, we find the log likelihood posterior, which is, up to an additive constant:

$$h(\mu_1, \phi_1, Y_1) = C^t + \ell(\mu_1, \phi_1, Y_1) + \ln(\pi(\mu_1, \phi_1))$$

Replacing the log-likelihood and log of the prior by their expression (see Equations 2 and 4), the log-posterior can be theoretically written as below:

$$h(\mu_1, \phi_1, Y_1) = C^t + \sum_{j=1}^{10} y_{1,j} \times \ln\left(\frac{1}{\Gamma(\phi^{-1})}\int_{\frac{I_j}{\phi_1\mu_1}} x^{(\phi_1^{-1}-1)}e^{-x}\mathrm{d}x\right) - \frac{1}{2\sigma_0^2}(\mu - \mu_0)^2 + \ln(\mathbf{1}_{0;10}) \tag{5}$$

This expression has been implemented in R by not considering the constant term.

```
# Intervals of the frequency table
intervals = c(0, 1200, 1500, 1800, 2300, 2700, 3300, 4000, 4900, 6000,
    Inf)

# Utility function to compute the prob of being between high and low
pgammadiff = function(low, high, kappa, lambda) {
    pgamma(high, kappa, lambda) - pgamma(low, kappa, lambda)
}

# Utility fonctions (already defined in the setup)
kappa = function(phi) {
    1/phi
}
lambda = function(phi, mu) {
    1/(phi * mu)
}

mu_prior = 3000
sigma_prior = 306.12
lpost = function(theta, freq) {
    # Transform mu and phi -> kappa and lambda
    kappa = kappa(theta[2])
    lambda = lambda(theta[2], theta[1])

    # Likelihood : sum_j^n[x_j * ln(probability_of_being_in_interval_j)]
    LL = sum(sapply(1:length(freq), function(j) {
        freq[j] * log(pgammadiff(low = intervals[j], high = intervals[j +
            1], kappa, lambda))
    }))

    # Log posterior
    lpi = dnorm(theta[1], mu_prior, sigma_prior, log = T) + dunif(theta[2],
        0, 10, log = T)
    lpost = LL + lpi

    names(lpost) = "lpost"
    return(lpost)
}
```

4 4.

4.1 Estimation

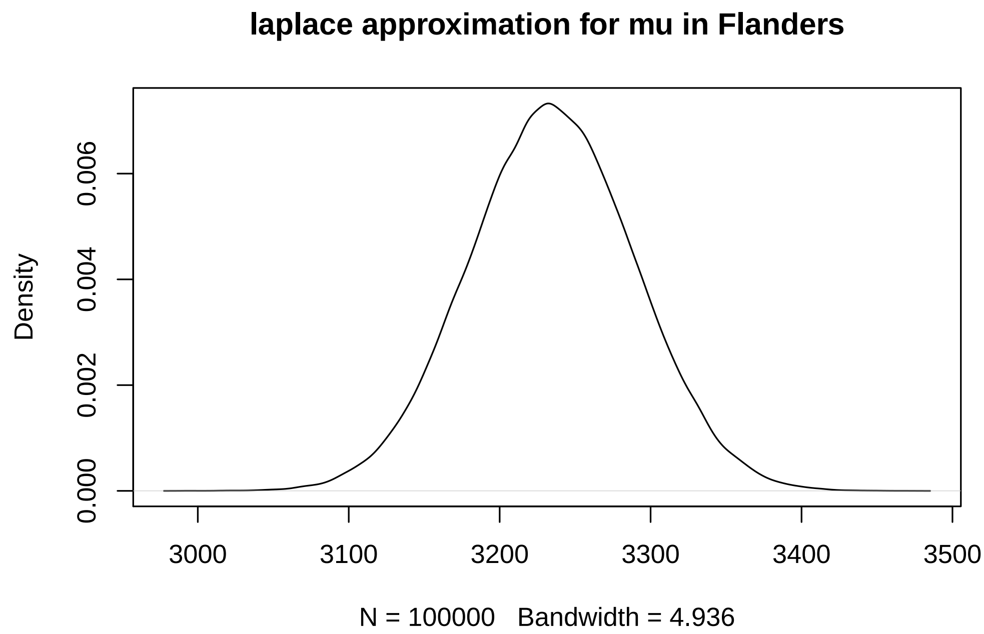
Start from the log-posterior $h(\mu_k, \phi_k)$. One assumes the distribution is unimodale then, if it is normally distributed, then the mode=mean, which means that the mean can be found by optimizing the log-likelihood with respect to its parameters. (aussi assumer μ et ϕ indépendant??), then blablabla (expliquez avec hessienne).

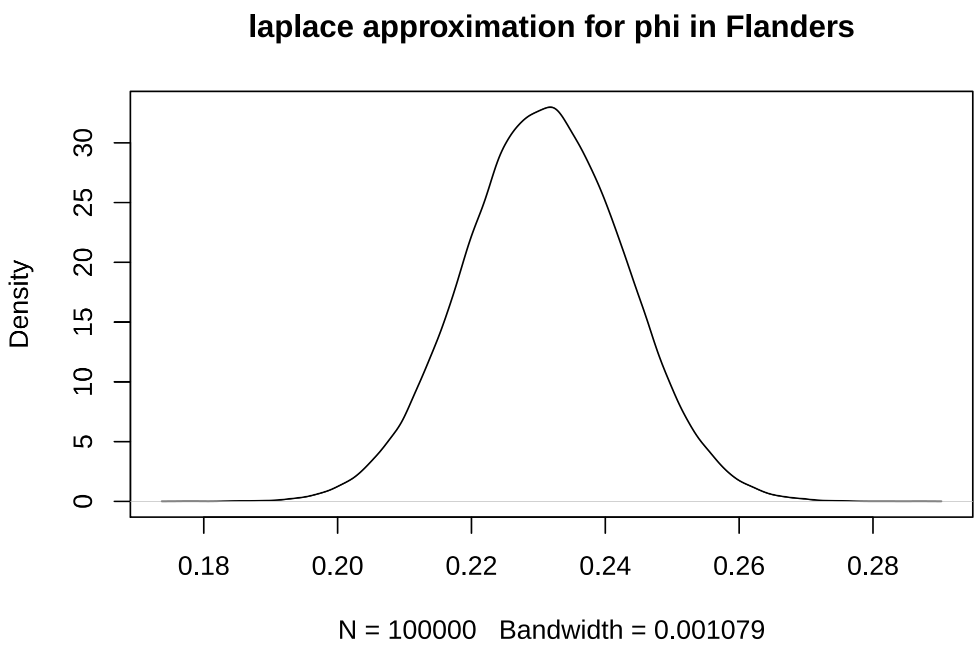
Based on http://www.sumsar.net/blog/2013/11/easy-laplace-approximation/#:~:text=Laplace%20Ap

4.1.1 Flanders

```
##  mu.shape phi.shape
##   3234.50      0.23
```

```
##           mu.shape phi.shape
## mu.shape  3023.270     0.025
## phi.shape    0.025     0.000
```

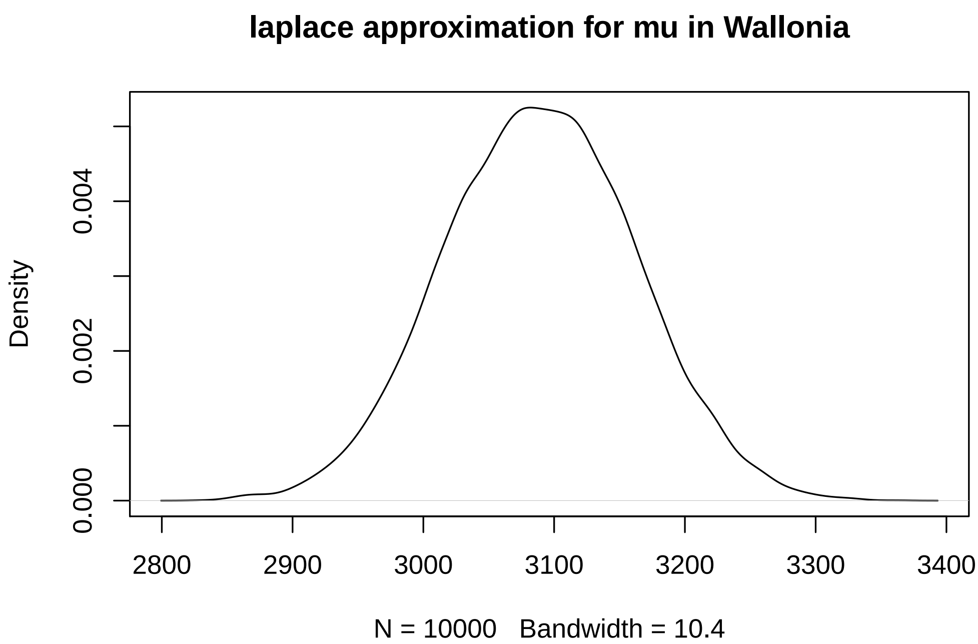


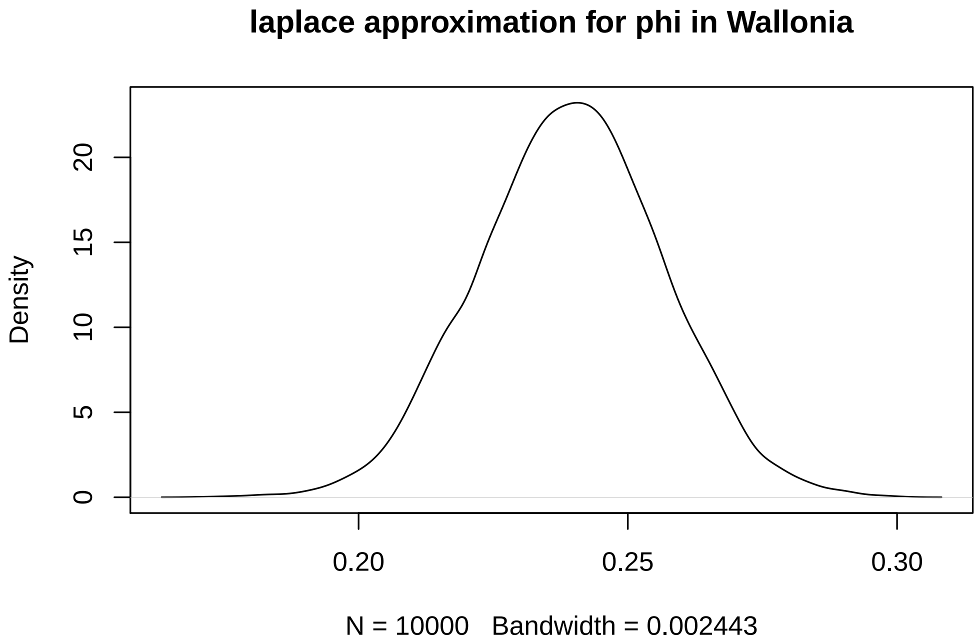


4.1.2 Wallonia

```
##  mu.shape phi.shape
##   3089.32      0.24
```

```
##           mu.shape phi.shape
## mu.shape  5252.793     0.039
## phi.shape    0.039     0.000
```

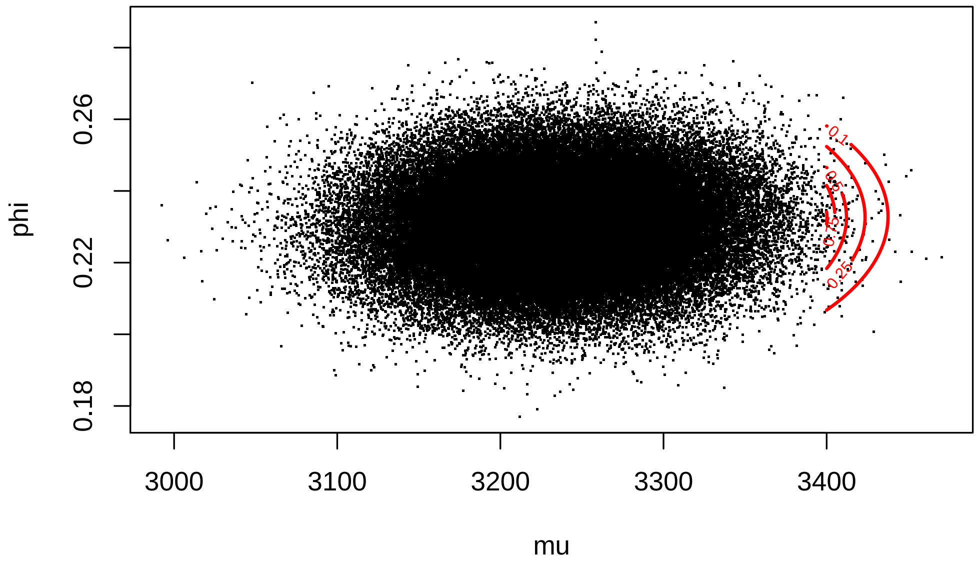




4.2 Credible intervals Laplce approximation

4.2.1 Flanders

The credible region is given here below:



We needs to get the credible interval for μ. This means that we needs the marginal posterior distribution:

$$P(\mu|D) \propto \int p(\mu, \phi|D) \mathrm{d}\phi$$

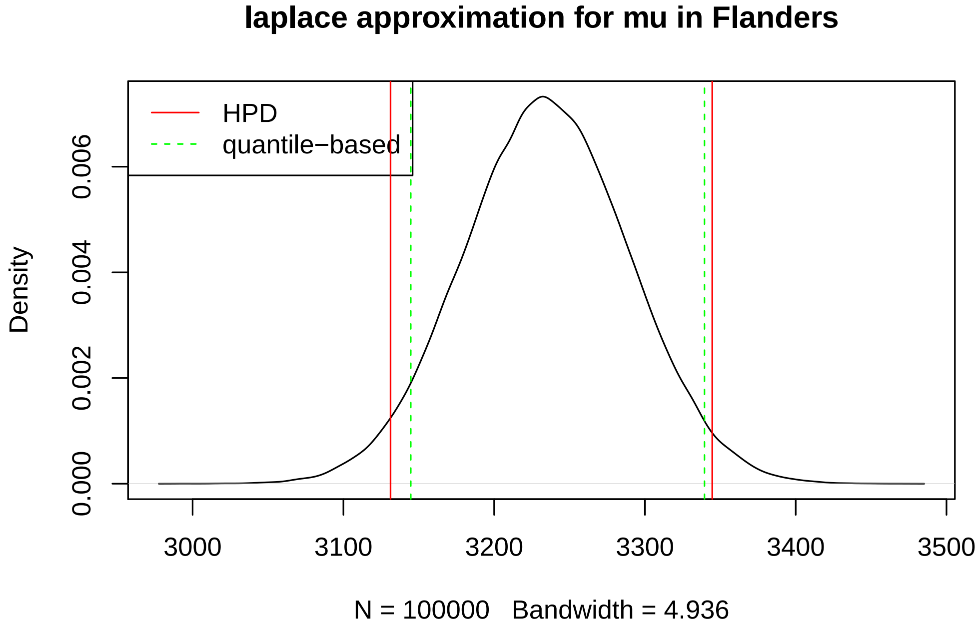
It can be shown that the marginal (univariate) distribution of the bivariate Gaussian distribution $N(\mu_\theta, \Sigma)$ with $\mu_\theta = (E(\mu), E(\phi))$ and

$$\Sigma = \begin{pmatrix} \sigma_\mu & \sigma_{\mu,\phi} \\ \sigma_{\phi,\mu} & \sigma_\phi \end{pmatrix}$$

also follows a normal distribution. See here. Hence, for μ, one has:

$$\mu|D{=}N(E(\mu), \sigma_\mu)$$

We needs that 95% of the marginal posterior to fall into the interval for Flanders:



5 Question 5

5.1 (a) Random walk component-wise Metropololis algorithm

One first needs starting values for $\theta_t := (\mu_t, \phi_t)$. One can take advantage of the MLE estimations of κ and λ and compute μ_0 and ϕ_0 using the relations described earlier. At each iteration, a new candidate will be proposed. To build it, it is a good idea to use the variance-covariance matrix computed into the Laplace approximation section since it should, more or less, describe the standard deviations of the parameters of interest $\hat{\sigma}_\mu$ and$\hat{\sigma}_\phi$. A factor, f_1 and f_2 should multiply their standard deviation in order to optimize the acceptance rate (goal is 40%). Moreover, since the posterior probability is on a log-scale, the comparison of probabilities should be made taking the exponential of the difference of the candidate with the current state. In this work, 10% of the generated data is considered as burn-in.

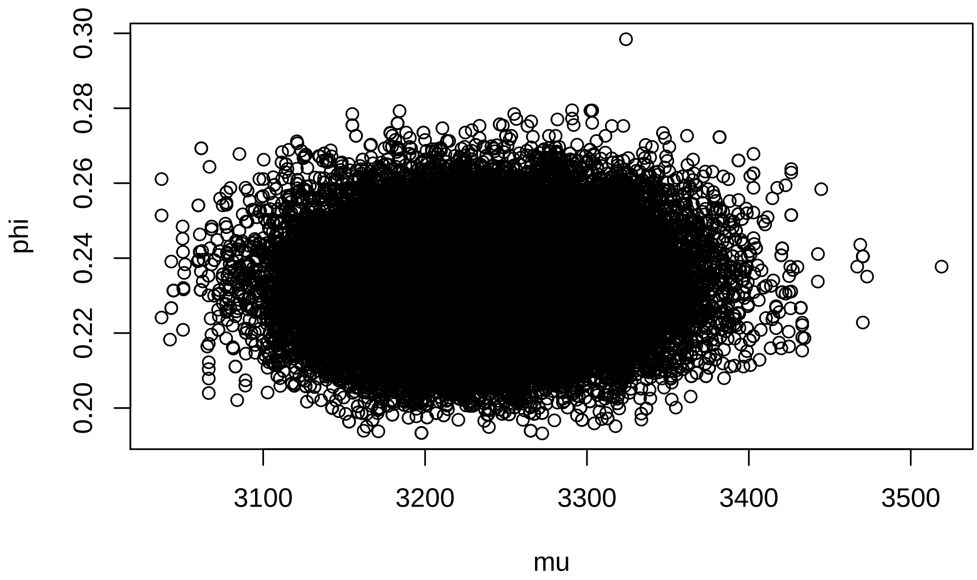
That being, this gives the following algorithm:

1. $t = 0$: Set starting values of the parameters of interest to their MLE: $\theta_0 := (\hat{\mu}_{MLE}, \hat{\phi}_{MLE})$
2. for $t = 2, 3, ..55000$
   - Draw μ_{prop} from a normal distribution centered on value derived at $t-1$, i.e. $\mu_{\text{prop}} \sim N(\mu_{t-1}, f_1\hat{\sigma}_\mu)$. The candidate is then $\theta^\mu_{prop} := (\mu_{\text{prop}}, \phi_{t-1})$
   - Compute $prob_\mu = \min\Big(1, \exp\big(h(\theta^\mu_{prop}) - h(\theta_{t-1})\big)\Big)$
   - Set $\mu_t = \mu_{\text{prop}}$ with probability $prob_\mu$, $\mu_t = \mu_{t-1}$ otherwise.
   - Draw ϕ_{prop} from a normal distribution centered on value derived at $t-1$, i.e. $\phi_{\text{prop}} \sim N(\phi_{t-1}, f_2\hat{\sigma}_\phi)$. The candidate is then $\theta^\phi_{prop} := (\mu_{t-1}, \phi_{\text{prop}})$

- Compute $prob = \min\left(1, \exp\left(h(\theta^{\phi}_{prop}) - h(\theta_{t-1})\right)\right)$
- Set $\phi_t = \phi_{\text{prop}}$ with probability $prob_{\phi}$, $\phi_t = \phi_{t-1}$ otherwise.
- Set $\theta_t = (\mu_t, \phi_t)$

```
## Acceptance rate for mu:  40.22833 %
```

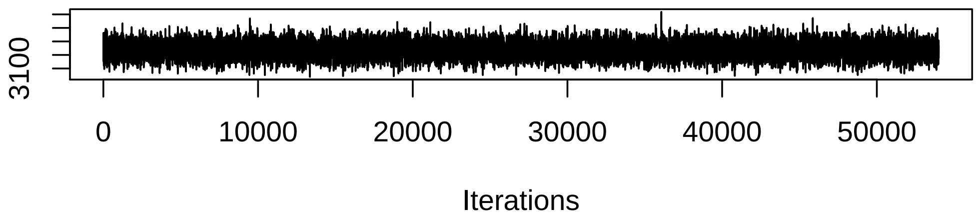
```
## Acceptance rate for phi:  40.20667 %
```

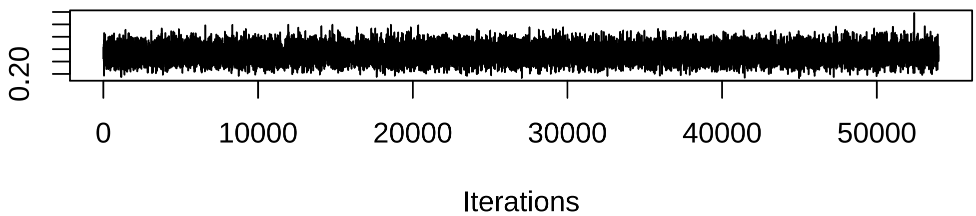


5.2 (b) diagnostic for convergence

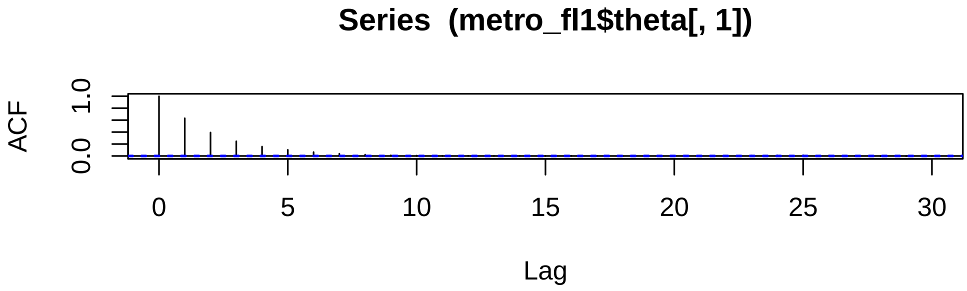
5.2.1 Graphs analysis

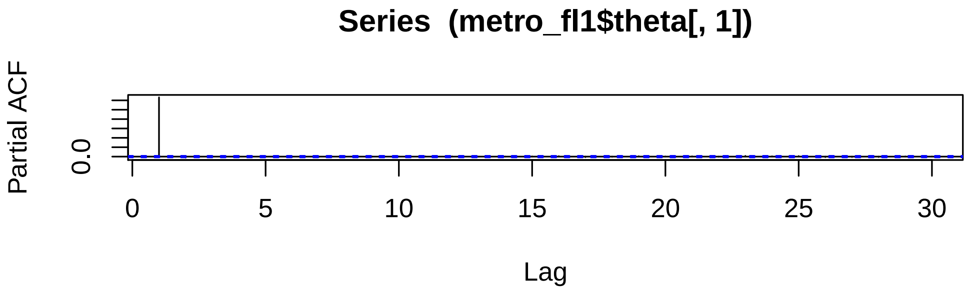
The first, trivial, way of checking if the convergence occurred is looking at the plot of the generated variables:

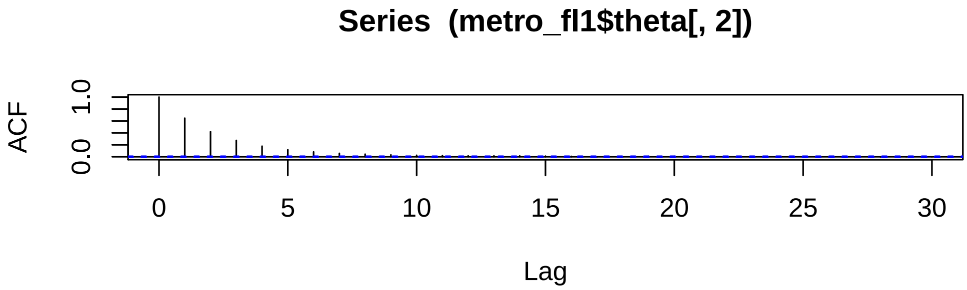


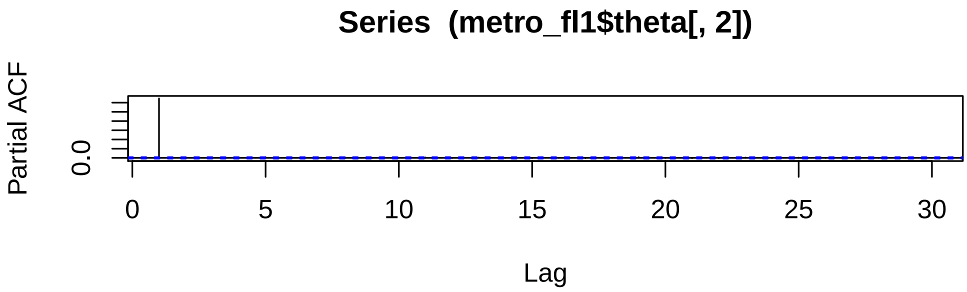


By doing so, one sees that the mixing seems pretty good. For example, the autocorrelation is significative up to not a too high order. As such, the parameter space of *theta* does not seem to be visited to slowly.





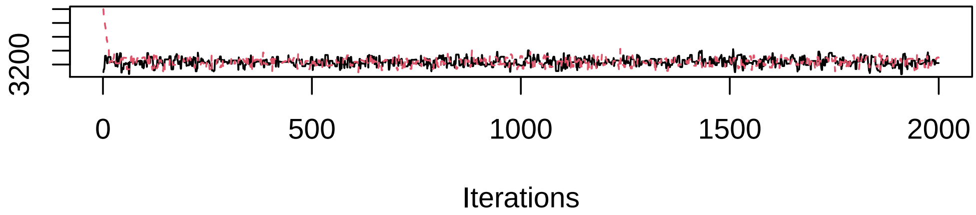


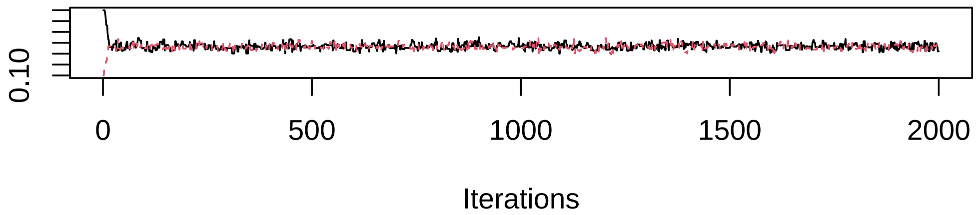


Since the chains clearly seem to be generated with respect to an Autoregressive process (see ACF/PACF plots), one can compute effectively the effective sample size. It represents how many samples would be generated if there were no autocorrelation.

```
##     var1     var2
## 12484.70 11879.18
```

Gelman-Rubin This method requires multiple chains. In this work, the diagnostic will be done with 2 generated chains with two different starting values. As a second chain, more exotic values (but still possible , a priori) are proposed for initialization: 4000 for μ_1 and 0.1 for ϕ_1. After having run the second chain, the traceplot of both chain is plotted with respect to the first 1000 generations.

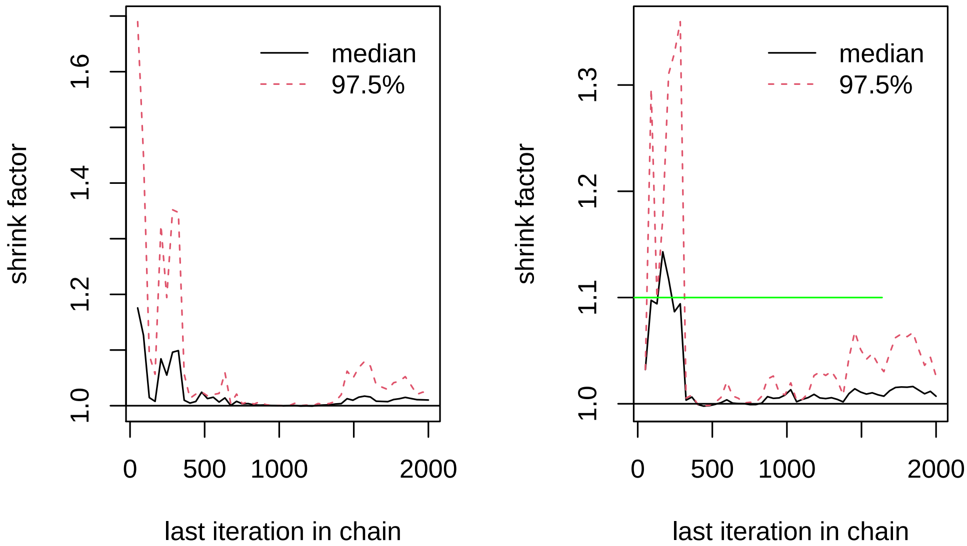




It seems that convergence occurs rather quickly. After a early generations, the chains seem to have similar mean variance. This can be more formally assessed with the Gelman-Rubin statistic that one wishes to be smaller that, say, 1.1. Here below is shown the estimated statistic along with its upper bond.

```
## Potential scale reduction factors:
##
##      Point est. Upper C.I.
## [1,]       1.01       1.02
## [2,]       1.01       1.03
##
## Multivariate psrf
##
## 1.01
```

Those are clearly below, which is a good sign of convergence. To see when convergence may have occurred, one could look at the statistic with respect to the number of generations. Here below are presented the statistics, up to 2000 samples:



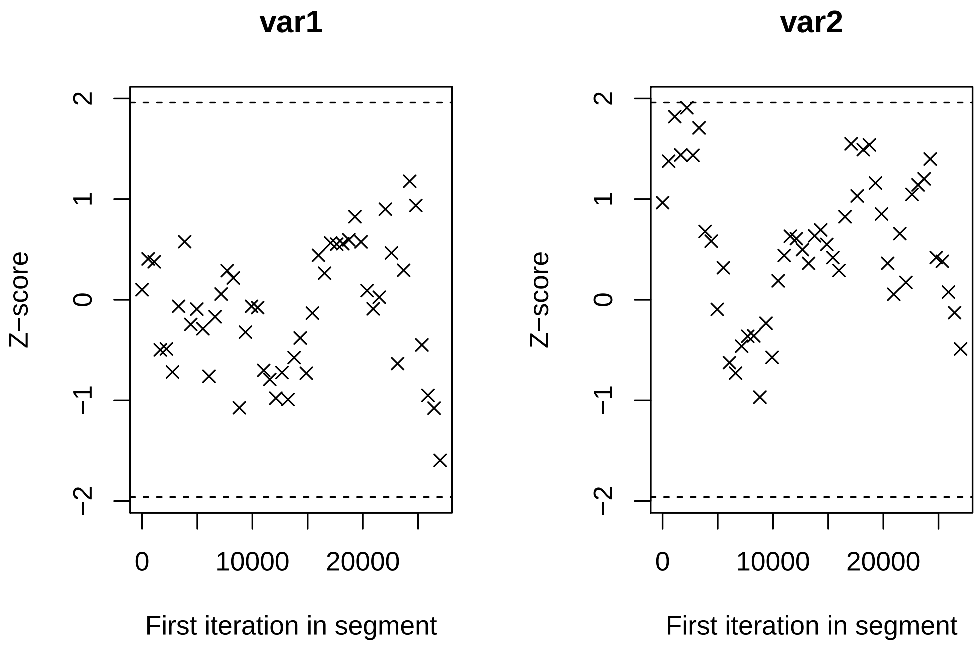
Accordingly, the convergence seems to occurs after roughly 1500 observations for both parameters.

5.2.2 Geweke diagnostic

The Geweke diagnostic is useful in the sense that it allows for only one chain to be generated. The chain will be separated into two part: the first accounts for the first 10% of the data while the other accounts for the last 50% of the chain. Under the hypothesis of convergence, the mean of both subsets should be similar. Hence, one can construct a corresponding two means test. The Z-scores are given given here below for μ and ϕ, respectively:

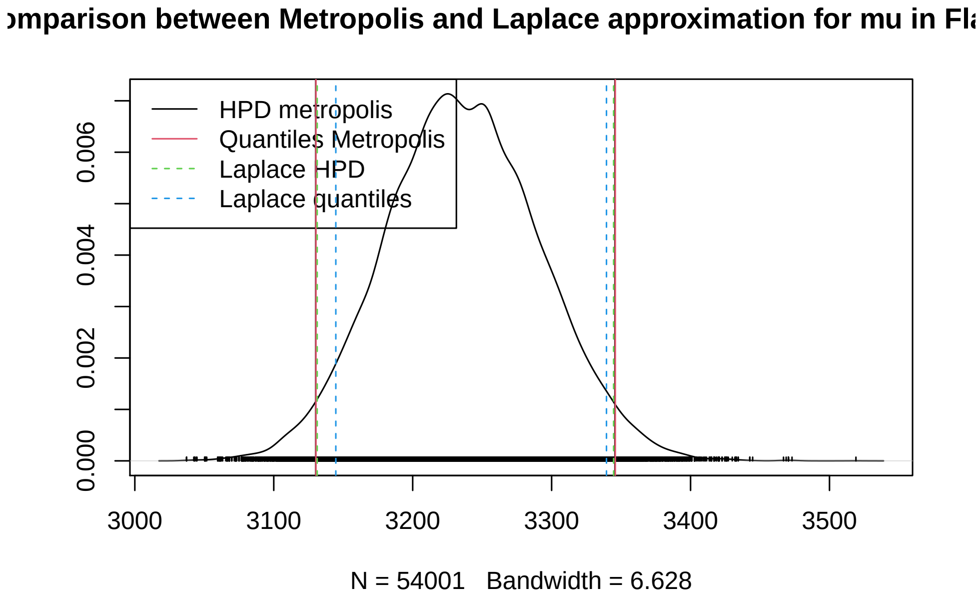
```
##
## Fraction in 1st window = 0.1
## Fraction in 2nd window = 0.5
##
##   var1   var2
## 0.0996 0.9653
```

To strengthen the belief of convergence, one can successively discard larger numbers of iterations from the beginning of the chain. From there, the z-score is successively computed and presented here below:



Except for one in ϕ, the hypothesis of same mean in never rejected. There is still no proof of non convergence of the generated chains.

5.3 (c) Credible intervals for mu1

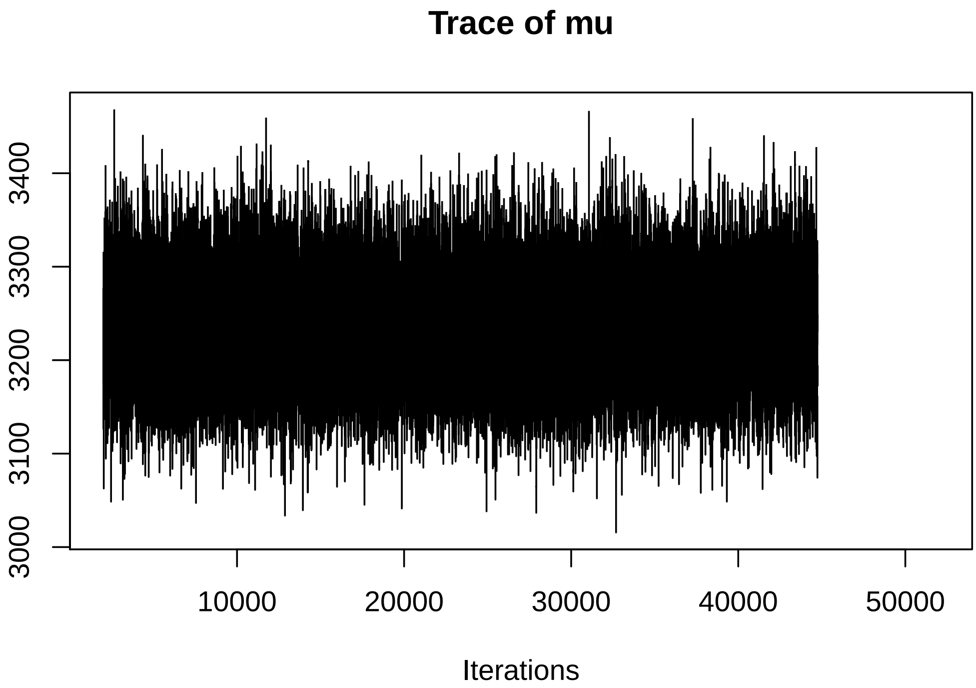


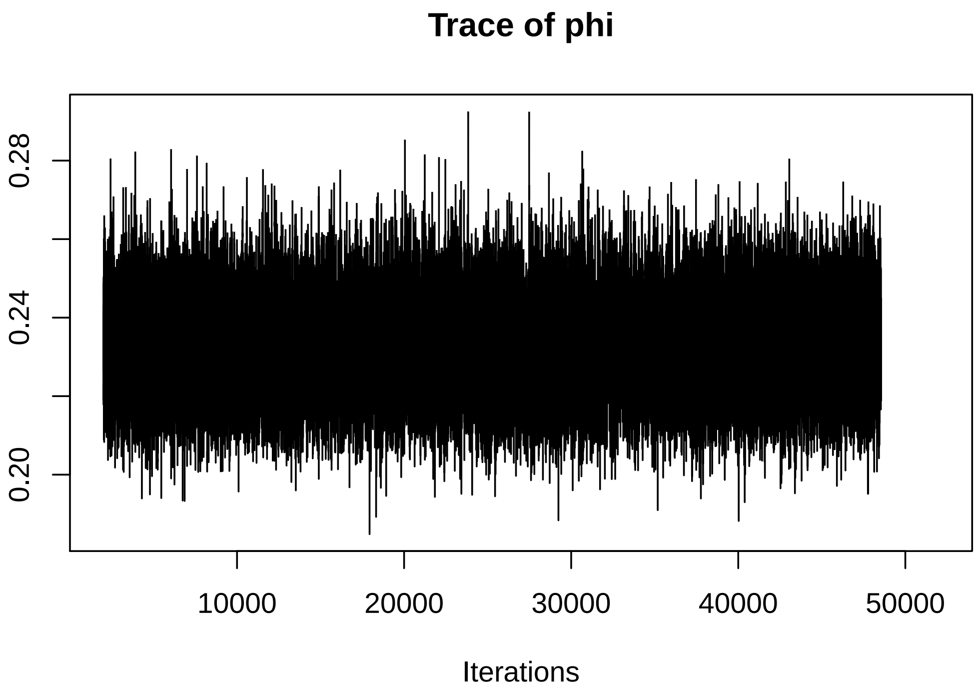
6 Question 6 : same question as 5 but with JAGS

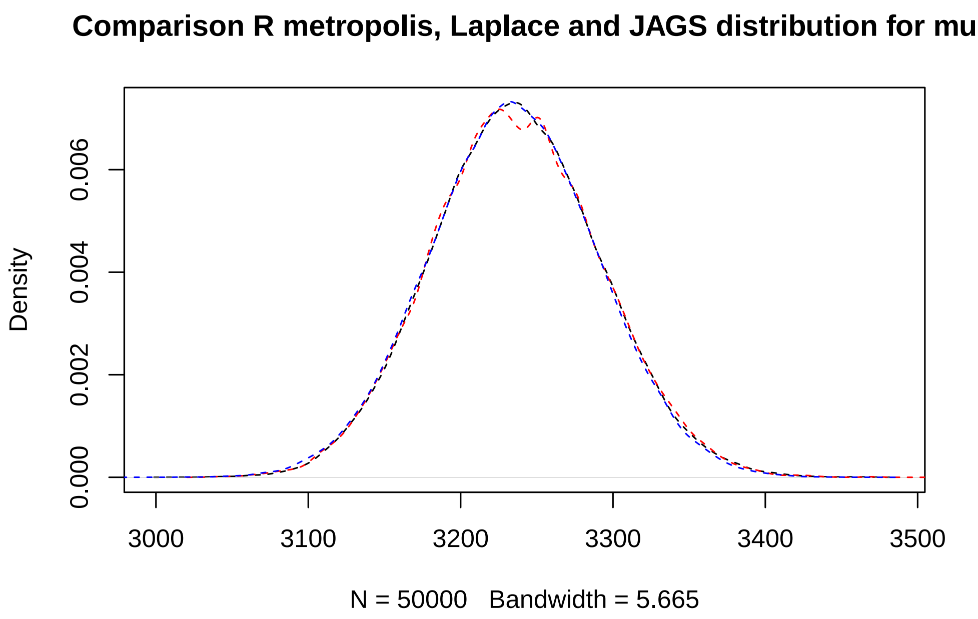
```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 1
##    Unobserved stochastic nodes: 2
##    Total graph size: 48
```

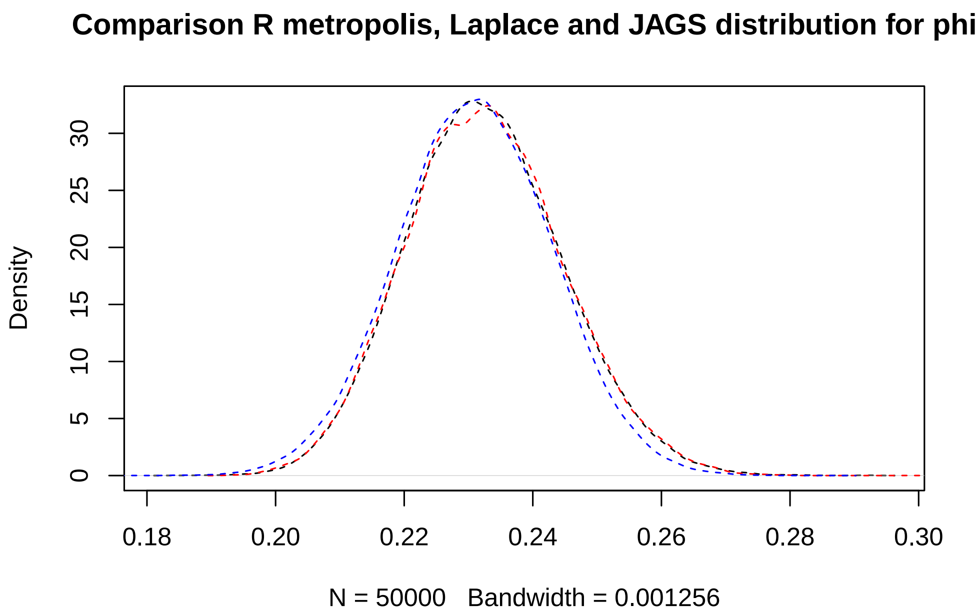
```
##
## Initializing model

##        mu              phi
##  Min.   :3016   Min.   :0.1848
##  1st Qu.:3199   1st Qu.:0.2242
##  Median :3236   Median :0.2321
##  Mean   :3237   Mean   :0.2326
##  3rd Qu.:3273   3rd Qu.:0.2405
##  Max.   :3468   Max.   :0.2925
```







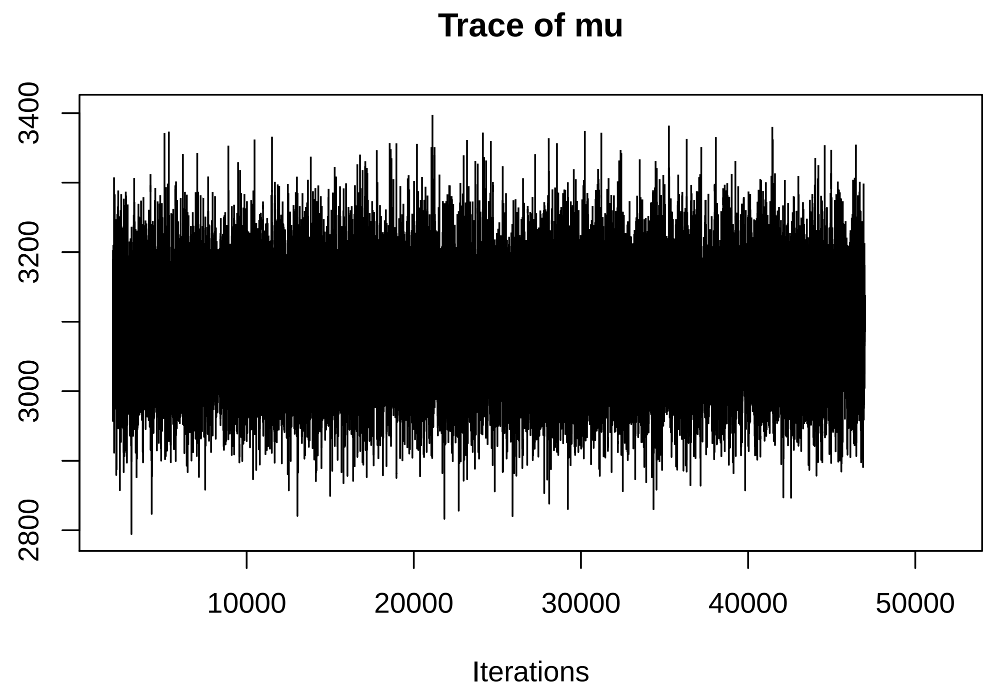


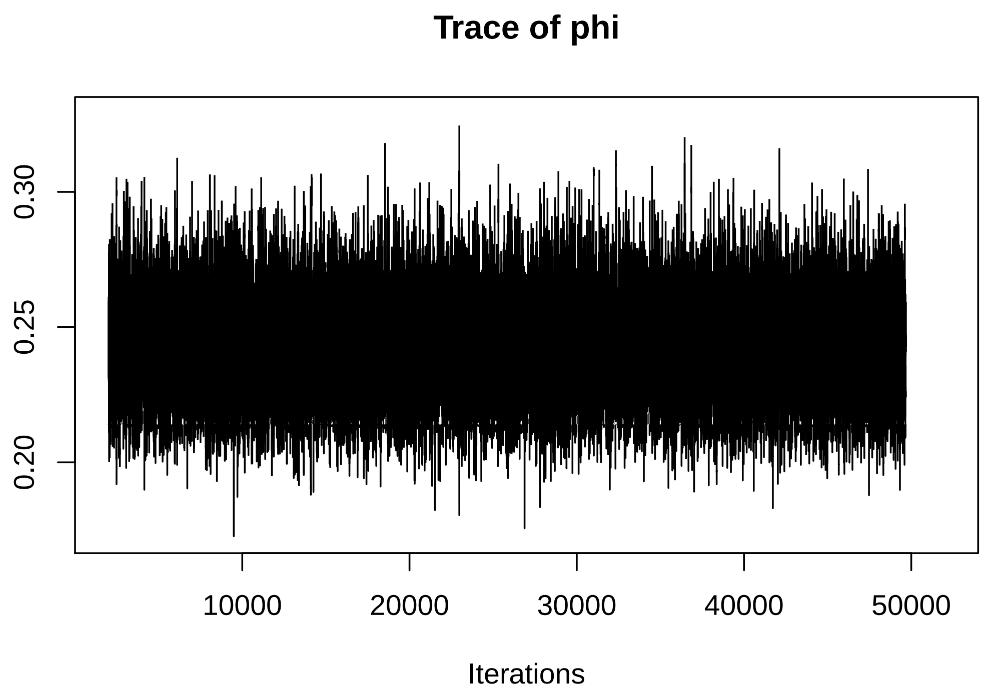
7 Question 7: same but for Wallonia

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 1
##    Unobserved stochastic nodes: 2
##    Total graph size: 48
##
## Initializing model


##
## Iterations = 2001:52000
```

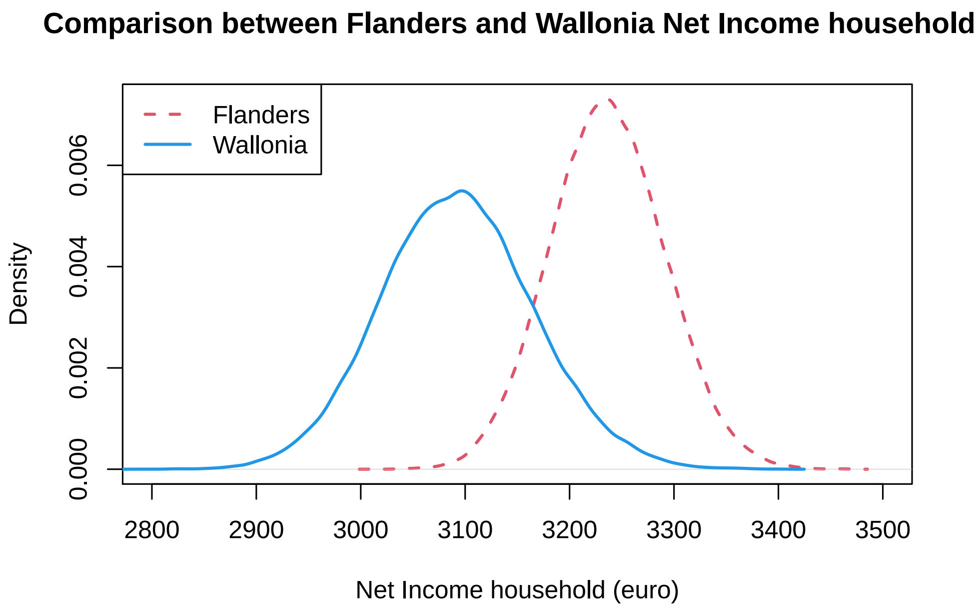
```
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 50000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##          Mean       SD  Naive SE Time-series SE
## mu  3092.7831 73.08517 0.3268468      0.4336739
## phi    0.2421  0.01735 0.0000776      0.0001013
##
## 2. Quantiles for each variable:
##
##          2.5%       25%       50%       75%     97.5%
## mu  2952.3323 3042.9440 3091.8296 3140.6805 3240.3645
## phi    0.2105    0.2301    0.2413    0.2533    0.2781
```



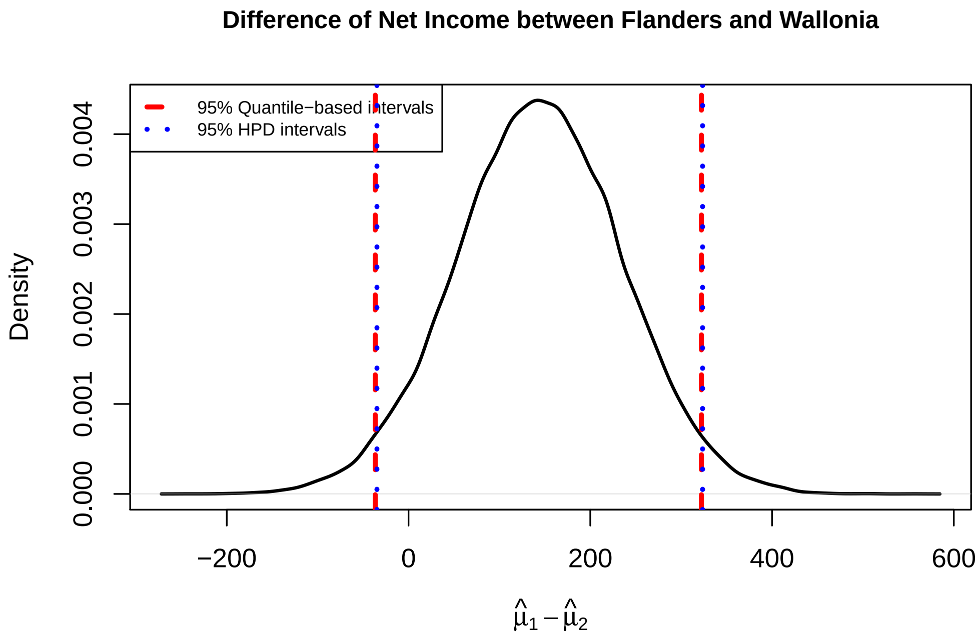


8 Question 8: Credible interval for means difference

The ultimate goal of this work is comparing the households income divergence between Flanders and Wallonia given some a priori and after a experimental study through 1228 Belgian households. To do so, 50000 samples have been generated through the metropolis algorithm for Flanders and its French-talking counterpart. Their estimated density is jointly represented here below:



There is a `143.74` eur difference on average. However, a large part of the density seems to overlap. The density of their difference is shown below along with a 95% qunatile based and HPD credible intervals:



On a 95% confidence level, it is impossible to conclude that the average household income level is significantly different between the regions. According to this Bayesian approach, one can only be `88.17` % sure that the income is different from one another. This is computed by looking from which confidence level one can reject the null hypothesis of same average income.

A Appendix

A.1 Figures

A.2 Code

Note

For reproducibility purposes, the complete R project containing the source code and the results is available on github.com.