UCLouvain

Faculté des sciences

Université catholique de Louvain

Louvain School of Statistics

LSTAT2130 - Bayesian Statistics

Project - Group Q

Lionel Lamy - 1294-1700

Adrien Kinart

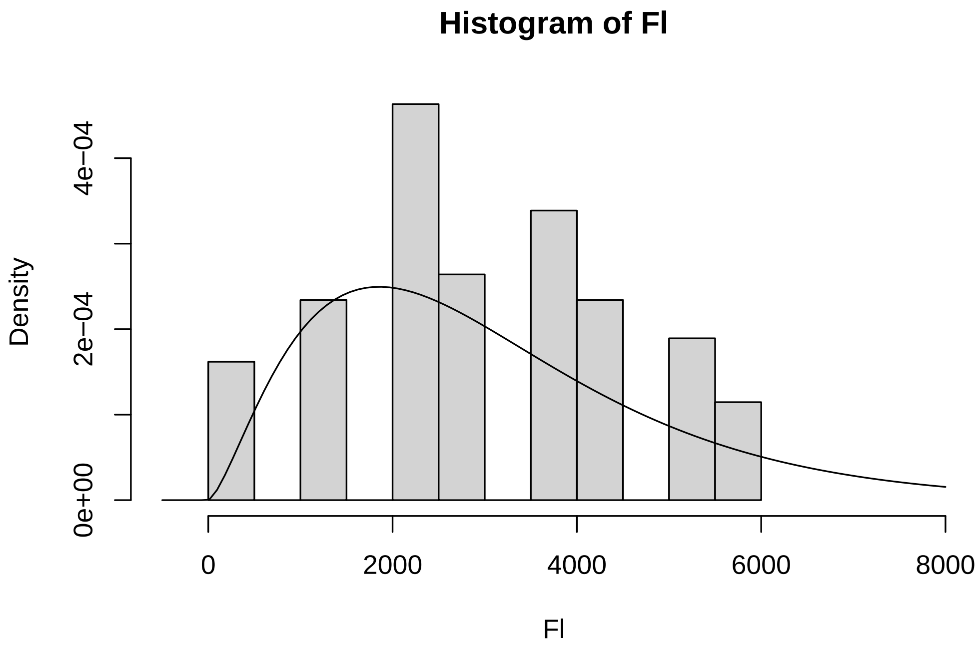
Simon Lengendre

May 12, 2021

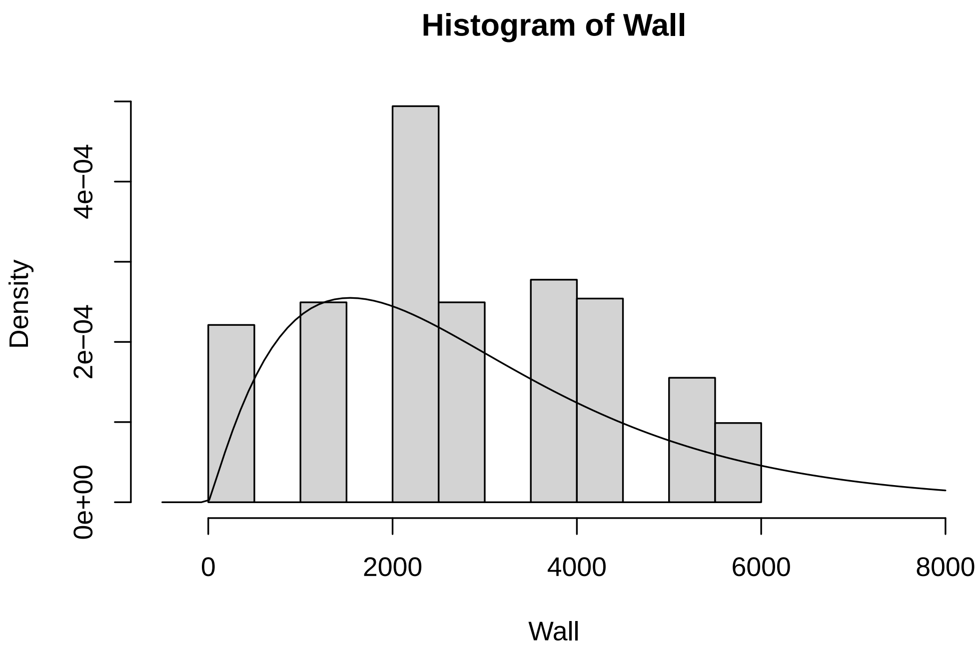
Contents

1 Introduction

```
##    shape
## 3089.944
```




```
##    shape
## 2914.882
```




```
## [1] 0.05892352
```

```
## [1] 0.0001358382
```

2 Question 1

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

2.1 (a) Theoretical probability

Let X be the monthly net income of 1123 Belgian households net income (HNI) older than 30 years. Regardless the 2 regions ($k = \{1, 2\}$ wrt Flanders and Wallonnia, respectively), is assumed it follows a Gamma distribution. It can be reparametrised in terms of its mean μ and dispersion parameter ϕ with the following trick:

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

For both regions $k = \{1, 2\}$: This gives

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

Then, the probability to fall into a certain HNI interval is:

$$P(x_{k,j} < x_k < x_{k,j+1}) = \int_{x_j}^{x_{j+1}} \frac{(\phi_k\mu_k)^{-1/\phi_k}}{\Gamma(\phi_k^{-1})} x_k^{1/\phi_k - 1}\exp\left(\frac{-x_k}{\phi_k\mu_k}\right)\,\mathrm{d}x$$

Using CDF writing:

$$\begin{aligned}
p_j &= F(x_{j+1}, \kappa, \lambda) - F(x_j, \kappa, \lambda) \\
&= \frac{1}{\Gamma(\kappa)}\Big(\gamma(\kappa, \lambda x_{j+1}) - \gamma(\kappa, \lambda x_j)\Big) \\
&= \sum_{i=0}^{\kappa-1} \frac{\lambda x_j}{i!}\exp(-\lambda x_j) - \sum_{i=0}^{\kappa-1}\frac{\lambda x_{j+1}}{i!}\exp(-\lambda x_{j+1}) \\
&= \lambda\exp(-\lambda)\sum_{i=0}^{\kappa-1}\left(\frac{x_j}{i!}\exp(-x_j) + \frac{x_{j+1}}{i!}\exp(-x_{j+1})\right)
\end{aligned} \tag{1}$$

2.2 (b) Theoretical expression for the likelihood

On behalf of writing simplicity, the region index is removed. Since the frequency distribution in a given region is multinomial, i.e.

We have, writing $P := (p1, .., p_{10})$ and $X := (X_1, ...X_{10})$:

$$X|P \sim \text{Mul}(N,P) = \frac{x!}{x_1! \ldots x_{10}!} p_1^{x_1} \times \ldots \times p_{10}^{x_{10}} \text{ when} \sum_{j=1}^{10} p_j = 1$$
$$= 0 \text{ otherwise}$$

Up to a multiplicative constant, the likelihood can be written as follow:

$$L(\theta, D) = P(D|\mu,\phi) \propto \prod_{j=1}^{10} p_j$$

p_j corresponds to the area in the j^{th} interval. One can take the approximation mean the mean,e.g. $x_{Flanders,3} = (1500+1800)/2 = 1650$? On can approximate that with $f(x_i)\Delta$ where Δ is the unit of measurement.

$$\begin{aligned} p_j = P(x_j - \frac{\Delta_j}{2} < x_j < x_j + \frac{\Delta_j}{2}) &\approx f(x_j)\Delta_j \\ &\approx \frac{1}{\Gamma(\phi_k^{-1})} \left(\phi_k \mu\right)^{-1/\phi_k} x_j^{\frac{1}{\phi_k}-1} \exp\left(\frac{-x_j}{\phi_k \mu}\right)\Delta_j \end{aligned}$$

This gives for the likelihood:

$$\begin{aligned} P(D|\mu,\phi) &\propto \prod_{j=1}^{10} x_j^{\frac{1}{\phi_k}-1} \exp\left(\frac{-x_j}{\phi_k\mu}\right)\Delta_j \\ &\propto \exp\left(\frac{-\sum x_j}{\phi_k \mu}\right) \prod_{j=1}^{10} x_j^{\frac{1}{\phi_k}-1} \Delta_j \end{aligned}$$

3 Question 2 : Priors

- Statement 1: we are at 95% convinced that the mean net monthly household income in a given region is in the interval (2400, 3600). If one assumes a normal distribution for the mean μ_k (à justifier), then it is possible to get a prior of the distriubtion for both regions:

For both regions $\mu_0 = 3000$. Then, to get the standard deviation:

$$\begin{aligned} 3000 - t_{(n_k-1,1-\alpha/2)} \frac{s_k}{\sqrt{n_k}} &= 2400 \\ \rightarrow \hat{\sigma}_0 = \frac{s_k}{\sqrt{n_k}} &= \frac{600}{t_{(n_k-1,1-\alpha/2)}} \end{aligned}$$

$\hat{\sigma}_{Fl} = 306$

So we have

$$\mu \sim N(\mu_0 = 3000, \sigma_0 = 306) \propto \sigma_0^{-1/2} \exp\big(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\big)$$

- dispersion parameter:

If one considers that the parameter can be in any point within the interval (0.0, 10.0) with the same probability, then one could say that it follows a uniform distribution between those to interval

$$\phi_k \sim U(a = 0, b = 10) \propto 1_{0,10}$$

Using the entropy theorem, the conjugate prior would be the product of the two last quantity:

$$\begin{aligned}\text{prior:} P(\mu_k, \phi_k) &= P(\mu_k)\ P(\phi_k) \\ &\propto \sigma_0^{-1/2} \exp\big(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\big) 1_{0;10}\end{aligned} \tag{2}$$

4 Question 3

4.1 Question 3a: posterior

$\sigma^2 = \phi\mu^2$

$$\begin{aligned}P(\mu, \phi|D) &\propto \exp\big(\frac{-\sum x_j}{\phi_k \mu}\big) \prod_{j=1}^{10} x_j^{\frac{1}{\phi_k}-1} \Delta_j \sigma^{-1/2} \exp\big(-\frac{1}{2\sigma^2}(\bar{x}_k - \mu)^2\big) 1_{0,10} \\ &\propto \exp\big(\frac{-\sum x_j}{\phi_k \mu}\big) \prod_{j=1}^{10} x_j^{\frac{1}{\phi_k}-1} \Delta_j \big(\frac{1}{\sqrt{\phi}\mu}\big) \exp\big(-\frac{1}{2\phi\mu^2}(\bar{x}_k - \mu)^2\big) 1_{0,10}\end{aligned} \tag{3}$$

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.