UCLouvain

Faculté des sciences

Université catholique de Louvain
Louvain School of Statistics

LSTAT2130 - Bayesian Statistics

Project - Group Q

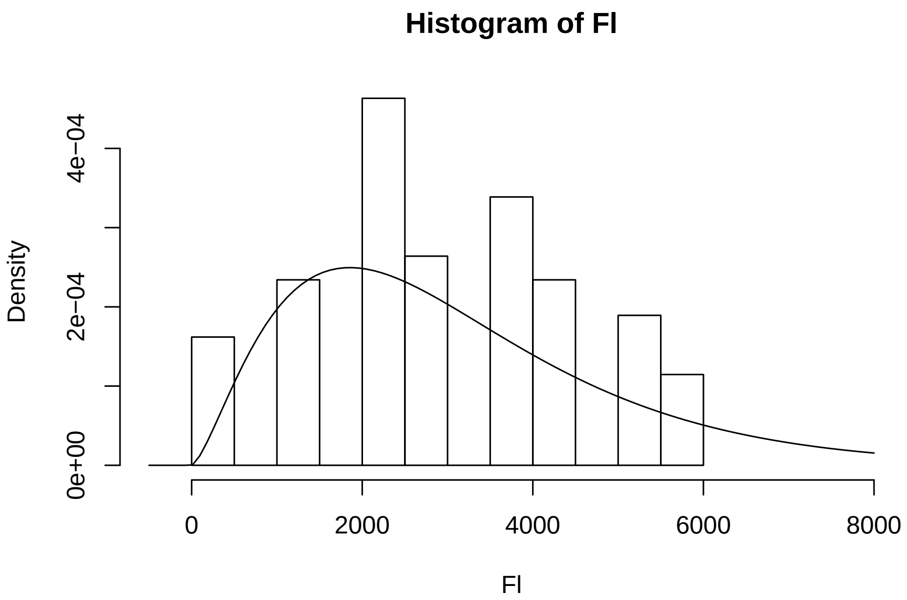
Lionel Lamy - 1294-1700
Adrien Kinart
Simon Lengendre

May 9, 2021

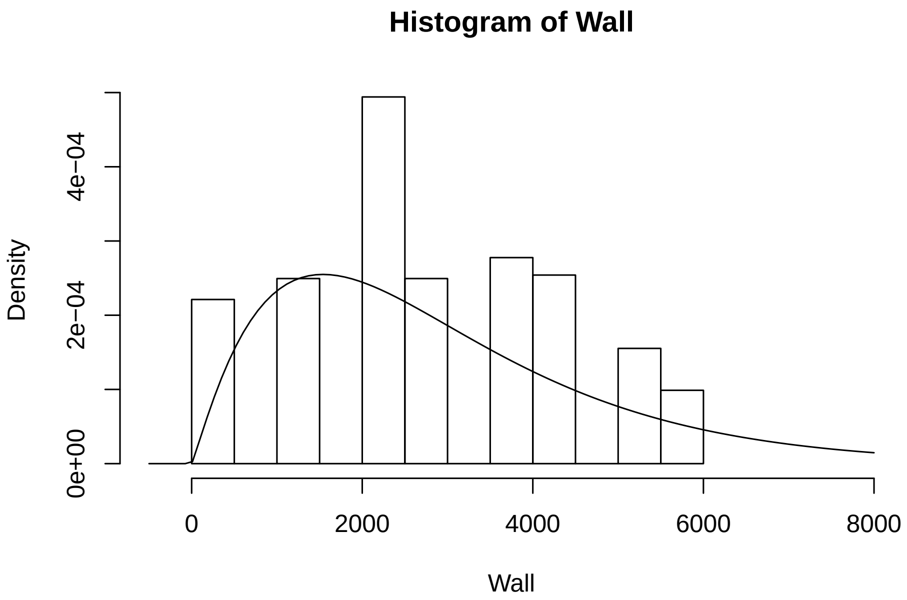
Contents

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

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

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

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

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 $X := (X_1, ...X_{10})$:

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

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.

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

This gives for the likelihood:

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

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 $\bar{x} = 3000$. Then, to get the standard deviation:

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

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

So we have

$$
\bar{x}_k \sim N(\mu = 3000, \sigma = 306) \propto \sigma^{-1/2} \exp\big(-\frac{1}{2\sigma}(x_k - \mu)^2\big)
$$

- 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:

$$\text{prior:} P(\mu_k, \phi_k) = P(\mu_k)(\phi_k) \propto$$

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.