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

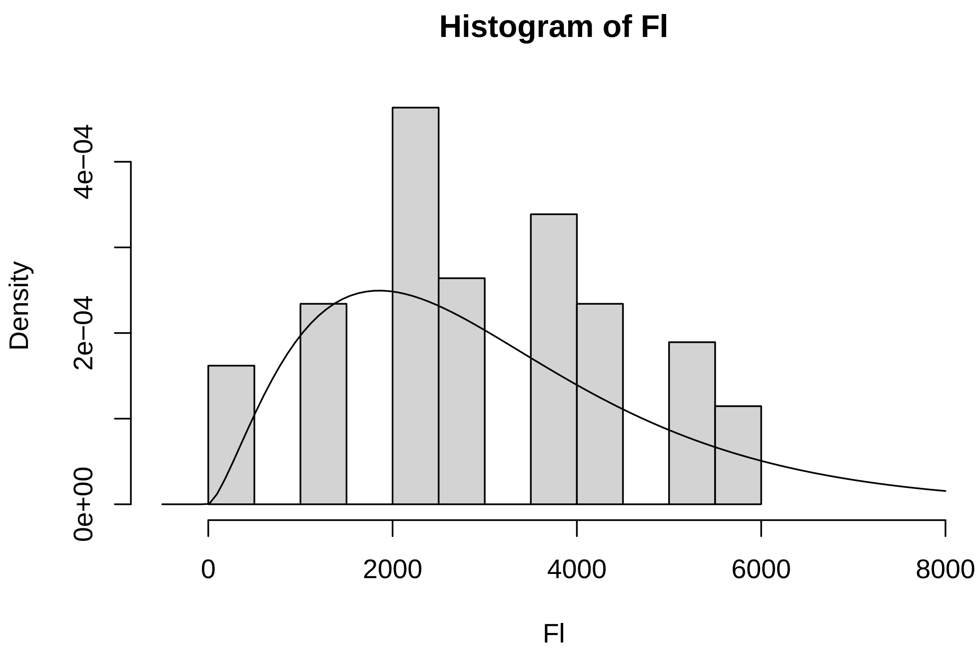
May 14, 2021

Contents

1 Introduction

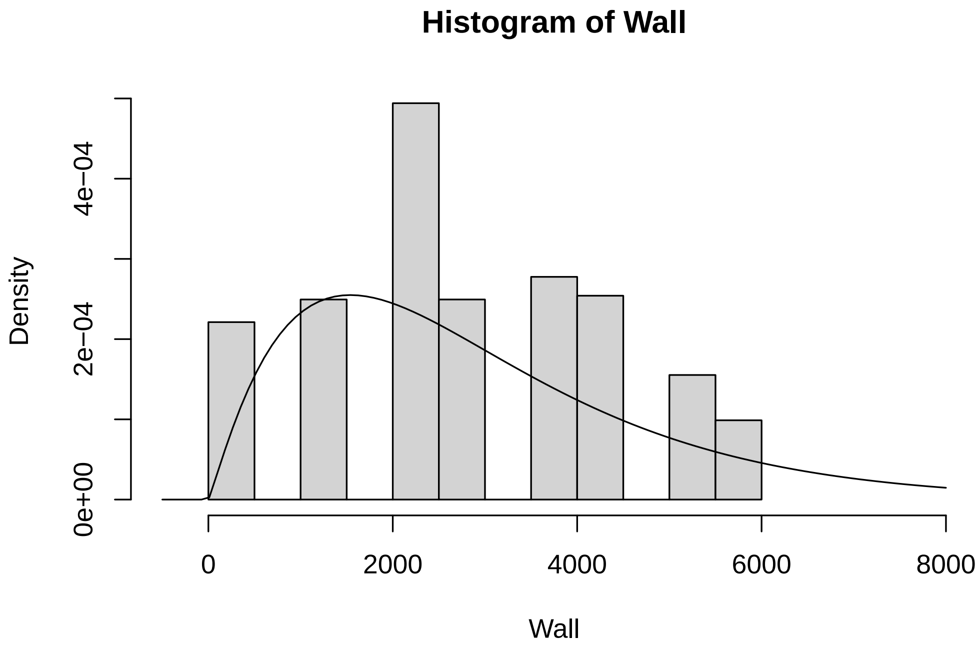
```
##    shape
## 3089.944
```

```
##    shape
## 0.399687
```




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

```
##    shape
## 0.471615
```




```
## [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 I is:

$$P(x_k \in I_j) = \int_{I_j} \frac{(\phi_k \mu_k)^{-1/\phi_k}}{\Gamma(\phi_k^{-1})} x_k^{1/\phi_k - 1} \exp\left(\frac{-x_k}{\phi_k \mu_k}\right) \,\mathrm{d}x$$

Using CDF writing, for a region k ,

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

$$p_j = \begin{cases} \frac{1}{\Gamma(\kappa)} \gamma(\kappa, \lambda x_j) \text{ if j } = 1 \\ \frac{1}{\Gamma(\kappa)} \Big(\gamma(\kappa, \lambda x_{j+1}) - \gamma(\kappa, \lambda x_j) \Big) \text{ if} j \in \{2, ..9\} \\ 1 - \frac{1}{\Gamma(\kappa)} \gamma(\kappa, \lambda x_j) \text{ if j } = 10 \end{cases} \tag{2}$$

$$p_j = \begin{cases} \frac{1}{\Gamma(\kappa)} \int_0^{\lambda x_j} t^{\kappa-1} e^{-t} \mathrm{d}t \text{ if j } = 1 \\ \frac{1}{\Gamma(\kappa)} \int_{\lambda x_j}^{\lambda x_{j+1}} t^{\kappa-1} e^{-t} \mathrm{d}t \text{ if} j \in \{2, ..9\} \\ \frac{1}{\Gamma(\kappa)} \int_{\lambda x_j}^{+\infty} t^{\kappa-1} e^{-t} \mathrm{d}t \text{ if j } = 10 \end{cases} \tag{3}$$

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

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

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

$$L(\theta_k, D_k) = P(D_k | \mu_k, \phi_k) \propto \prod_{j=1}^{10} p_{k,j}^{x_{k,j}}$$

2.3 Taking approximation

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.

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

This gives for the likelihood:

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

2.4 Not taking approximation but the CDF differences

$$
\begin{aligned}
P(D|\kappa,\lambda) &\propto \gamma(x_1,\kappa,\lambda)^{x_1}\left[\prod_{j=2}^{9}\Big(\gamma(x_j,\kappa,\lambda)-\gamma(x_{j-1},\kappa,\lambda)\Big)^{x_j}\right]\big(1-\gamma(x_{10},\kappa,\lambda)\big)^{x_{10}} \\
&\propto \left(\int_0^{\lambda x_1} x^{\kappa-1}e^{-x}\mathrm{d}x\right)^{x_1}\left[\prod_{j=2}^{9}\Big(\int_{\lambda x_j-1}^{\lambda x_j} x^{\kappa-1}e^{-x}\mathrm{d}x\Big)^{x_j}\right]\left(\int_{\lambda x_{10}}^{+\infty} x^{\kappa-1}e^{-x}\mathrm{d}x\right)^{x_{10}}
\end{aligned}
$$

If we write $[0;x_1],[x_1;x_2],...,[x_9;x_{10}],[x_{10},+\infty]$ as $I_1,I_2,...,I_9$ and I_{10}. The notation can be lightened. With respect to the region k, this gives:

$$
P(D_k|\kappa_k,\lambda_k) \propto \prod_{j=1}^{10}\left(\int_{\lambda_k I_j} x^{\kappa_k-1}e^{-x}\mathrm{d}x\right)^{x_{k,j}}
$$

Writing in terms of μ_k and ϕ_k:

$$
P(D_k|\mu_k,\phi_k) \propto \prod_{j=1}^{10}\left(\int_{\frac{I_j}{\phi_k\mu_k}} x^{(\phi_k^{-1}-1)}e^{-x}\mathrm{d}x\right)^{x_{k,j}}
$$

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

$$\mu \sim N(\mu_0 = 3000, \sigma_0 = 306)$$

$$\pi(\mu) \propto \sigma_0^{-1/2} \exp\big(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\big)$$

- 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:

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

4 Question 3

4.1 Question 3a: posterior

4.1.1 With approximation

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

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

4.1.2 Without approximation

$$P(\mu_k,\phi_k|D) \propto \left(\prod_{j=1}^{10}\left(\int_{\frac{I_j}{\phi_k\mu_k}} x^{(\phi_k^{-1}-1)}e^{-x}\mathrm{d}x\right)^{x_{k,j}}\right)\sigma_0^{-1/2}\exp\big(-\frac{1}{2\sigma_0^2}(\mu-\mu_0)^2\big)1_{0;10}$$

4.2 3.b

the log likelihood is:

$$LL \propto \sum_{j=1}^{10} x_j \ln \left(\int_{\frac{I_j}{\phi_k \mu_k}} x^{(\phi_k^{-1}-1)} e^{-x} \mathrm{d}x \right)$$

```
## [1] -1755.954
```

```
## [1] -956.3159
```

5 4.

$$h(\mu, \phi) \propto l(\mu, \phi) + \ln(\pi(\mu, \phi))$$

$$h(\mu, \phi) = Ct + \sum_j x_j \ln \big(F(\mu, \phi, x_{j+1}) - F(\mu, \phi, x_j) \big) - \frac{1}{2\sigma_0^2}(\mu - \mu_0)^2$$

... pour le moment... (voir photo envoyée sur messenger)

$$\mu^* = \mu_0 + \sigma_0^2 (\sum^{10} x_j \frac{f_{j+1} - f_j}{F_{j+1} - Fj})$$
$$\sigma^{2*} = \sigma_0^2$$

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.