Group 8: Robust Regression

Aimee, Roger and Hari

November 29, 2015

1. The Laplace distribution

1.1 Write down the likelihood for μ, Λ

$\prod_{i=1}^{N} \frac{1}{2}\Lambda e^{-\Lambda|t_i-\mu|} = \frac{1}{2}\Lambda e^{\sum_{i=1}^{N} -\Lambda|t_i-\mu|}$

1.2. Show that

$\mu_{mle} = median(t_1, ..., t_N)$

The likelihood function expands to:

$f(\mu, \Lambda) = \frac{1}{2}\Lambda e^{\sum_{i=1}^{N} -\Lambda|t_i-\mu|}$

Maximizing the likelihood function is equivalent to maximizing the log of the likelihood function.

$log(f(\mu, \Lambda)) = Nlog(\frac{1}{2}) + Nlog(\Lambda) + \sum_n -\Lambda|t_n - \mu|$

Taking the derivate with respect to μ and set equal to 0:

$-\Lambda \sum_n sgn(t_n - \mu) = 0$

If N is odd there are $\frac{N-1}{2}$ cases where $t_i < \mu$ and for the other $\frac{N-1}{2}$ cases $t_i > \mu$ and thus we have an equality with zero and the median of t_i is μ_{mle}. If N is even, there is no way to satisfy this equality but we can minimize it by taking either $\frac{t_i}{N}$ or $\frac{t_i}{N-1}$

1.3. Show that

$\Lambda_{mle} = (\frac{1}{N}\sum_{i=1}^{N} |t_i - \mu_{mle}|)^{-1}$

Taking the derivative of the likelihood function and setting it equal to zero results in:

$\frac{N}{\Lambda} = \sum_{i=1}^{N} |t_i - \mu|$

which implies

$\Lambda = (\frac{1}{N}\sum_{i=1}^{N} |t_i - \mu|)^{-1}$

And replacing μ with μ_{mle}, we get

$\Lambda_{MLE} = (\frac{1}{N}\sum_{i=1}^{N} |t_i - \mu_{mle}|)^{-1}$

1.4. Show that

$\Sigma_{mle} = 2\frac{1}{N}\sum_n |t_n - \mu_{mle}|$

Where $\Sigma := var[t]$ and $var[t] = \frac{2}{\Lambda^2}$

Using the equivariance property

$\Sigma_{mle} = \frac{2}{\Lambda_{mle}^2}$

We know that $\Lambda_{MLE} = (\frac{1}{N}\sum_{i=1}^{N} |t_i - \mu_{mle}|)^{-1}$. Therefore

$\Sigma_{mle} = 2(\frac{1}{N}\sum_n |t_n - \mu_{mle}|)^2$

2. EM algorithm for robust regression

2.1. Show that if

$t_n|x_n, w, q, \eta_n \sim N(\phi(x_n^T w, (\eta_n q)^{-1} I))$ and $\eta_n \sim Gam(\frac{\nu}{2}, \frac{\nu}{2} - 1)$ then

$\eta_n|t_n, x_n, w, q \sim Gam(\frac{(\nu+1)}{2}, \frac{(\nu+qe_n^2)}{2} - 1)$

with $e_n := t_n - \phi(x_n)^T w$

By Bayes, we know that $p(\eta_n|t_n, x_n, w, q) \propto p(\eta)p(t_n|x_n, w, q, \eta_n)$

η follows the known gamma distribution: $p(\eta) = \frac{(\frac{\nu}{2}-1)^{\frac{\nu}{2}}}{(\frac{\nu}{2}-1)!}(\nu)^{\frac{\nu}{2}-1} exp^{-(\frac{\nu}{2}-1)\eta_n}$

The conditional distribution of t follows the known normal distribution: $p(t_n|x_n, w, q, \eta_n) = \frac{1}{(2\pi)^{\frac{1}{2}}}(q\eta_n)^{\frac{1}{2}} exp^{-\frac{1}{2}e_n^2 q}$

Which implies that the conditional distribution of η is as follows:

$p(\eta_n|t_n, x_n, w, q) \propto e^{-(\frac{\nu}{2}-1)\eta_n} e^{-\frac{1}{2}e_n^2 q} * (\nu)^{\frac{\nu}{2}-1}\eta_n^{\frac{1}{2}}$

$p(\eta_n|t_n, x_n, w, q) \propto exp^{-(\frac{\nu+qe_n^2}{2}-1)\eta_n}\eta_n^{(\frac{\nu-1}{2})}$

Therefore

$\eta_n|t_n, x_n, w, q \sim Gam(\frac{(\nu+1)}{2}, \frac{(\nu+qe_n^2)}{2} - 1)$

2.2. Show that

when $\theta = (w, q)$ and $\theta' = (w', q')$:

$Q(\theta, \theta') = \frac{N}{2}log(q) - \frac{q}{2}(t - \Phi w)^T) diag(\mathbb{E}[\eta|T, x, \theta'])(t - \Phi w) + C$

where

$\mathbb{E}[\eta|T, x, \theta']$ is a vector with elements $\frac{\nu+1}{\nu+q'(t_N-\phi(x_n)^T w')^2-2}$

$Q(\theta, \theta') = \int logp(t|\eta, \theta)p(\eta|t, \theta')d\eta$

$Q(\theta, \theta') = \mathbb{E}(Log(L(\theta|t, \eta)))$

$Q(\theta, \theta') = \mathbb{E}(Log(L(t|\eta, \theta)p(\eta|\theta', t)))$

$Q(\theta, \theta') = \mathbb{E}(\frac{N}{2}log(q) - \frac{1}{2}(t - \Phi w)^T q\eta_n I(t - \Phi w) + C)$

$Q(\theta, \theta') = \frac{N}{2}log(q) - \frac{q}{2}diag(\mathbb{E}[\eta|T, x, \theta']) + C$

and

$\mathbb{E}[\eta|t, x, \theta'] = \frac{(\frac{(\nu+1)}{2})}{(\frac{(\nu+q'(t_n-\phi(x_n)^t w')^2-2)}{2})}$

therefore

$\mathbb{E}[\eta|t, x, \theta'] = \frac{(\nu+1))}{(\nu+q'(t_n-\phi(x_n)^t w')^2-2))}$

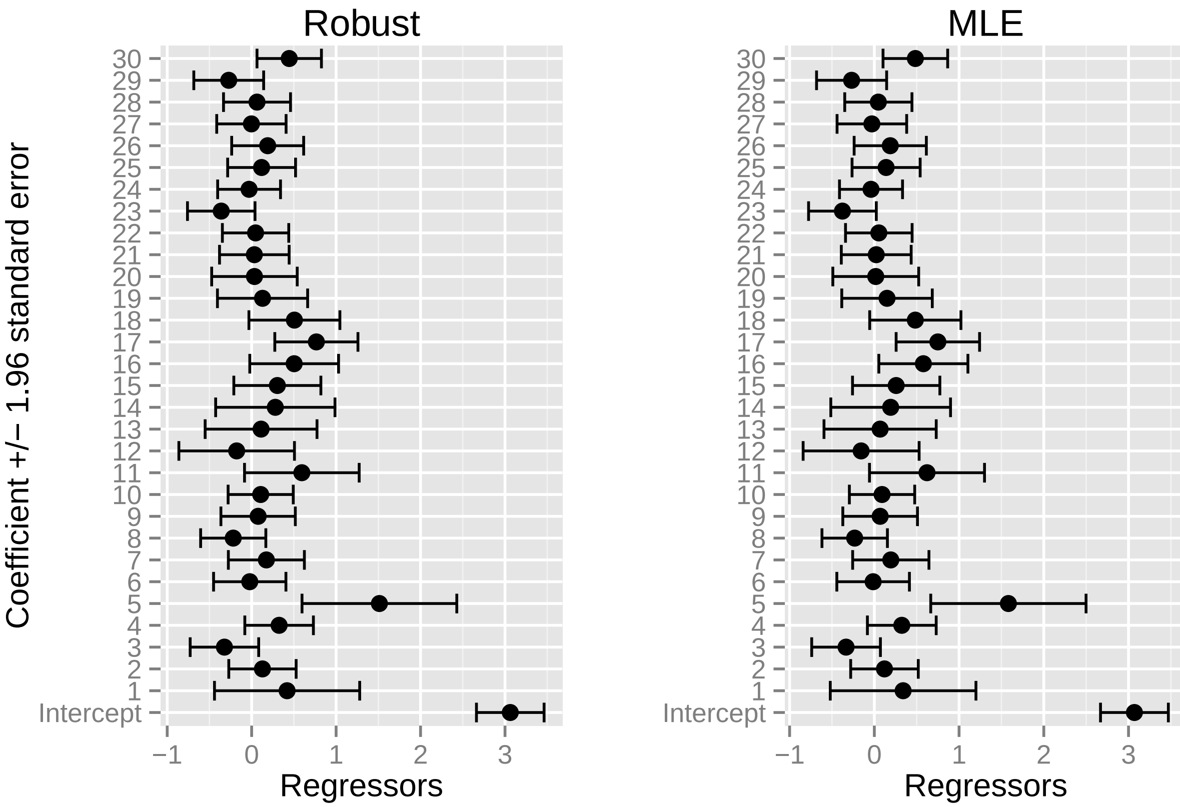
3. R exercise

3.1.

The following graphs compare the coefficients for both plain vanilla MLE estimators and the ones obtained by a robust regression. As can be observed, the differences are very small, indicating that there are no outliers in the data leveraging the standard MLE estimators.

```
## number of iterations 10

## Loading required package: grid
```

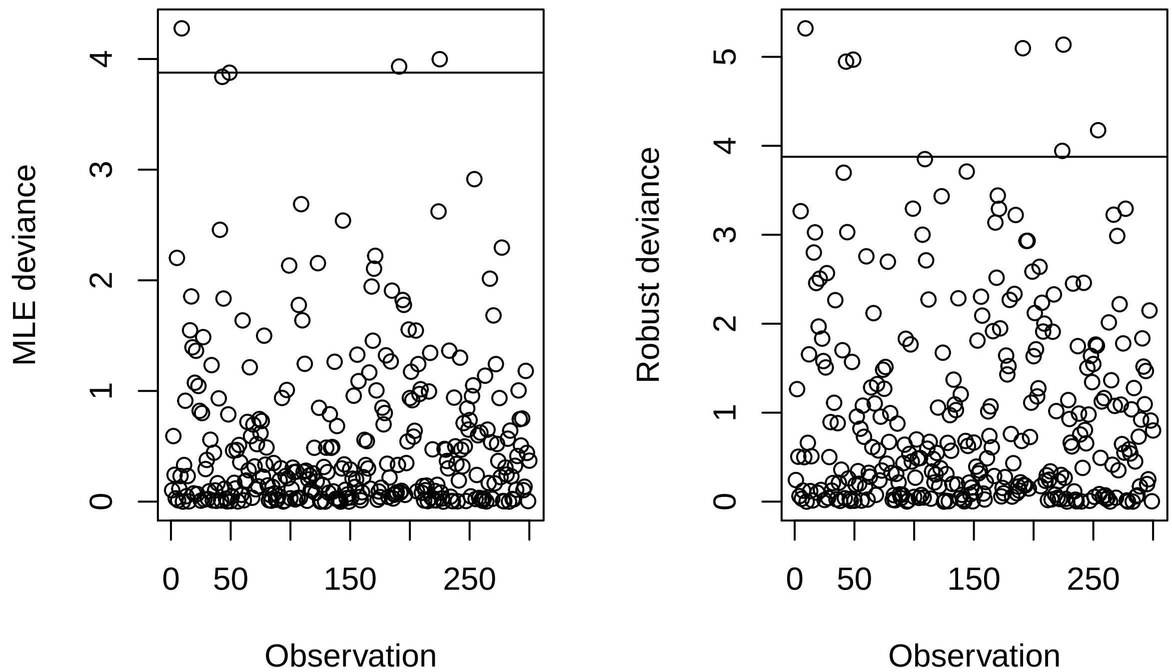



```
## [1] 2
```

3.2

The following graph plots the deviance residuals for both the standard Gaussian model and the robust regression.

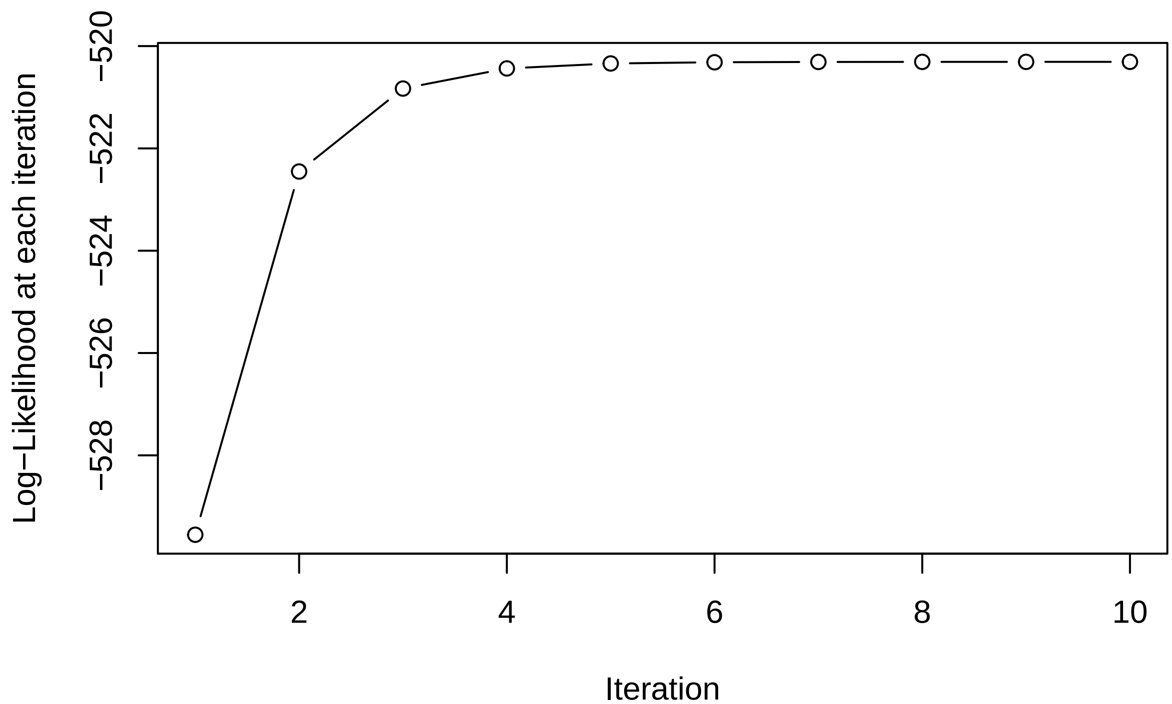
```
## numeric(0)
```




```
## numeric(0)
```

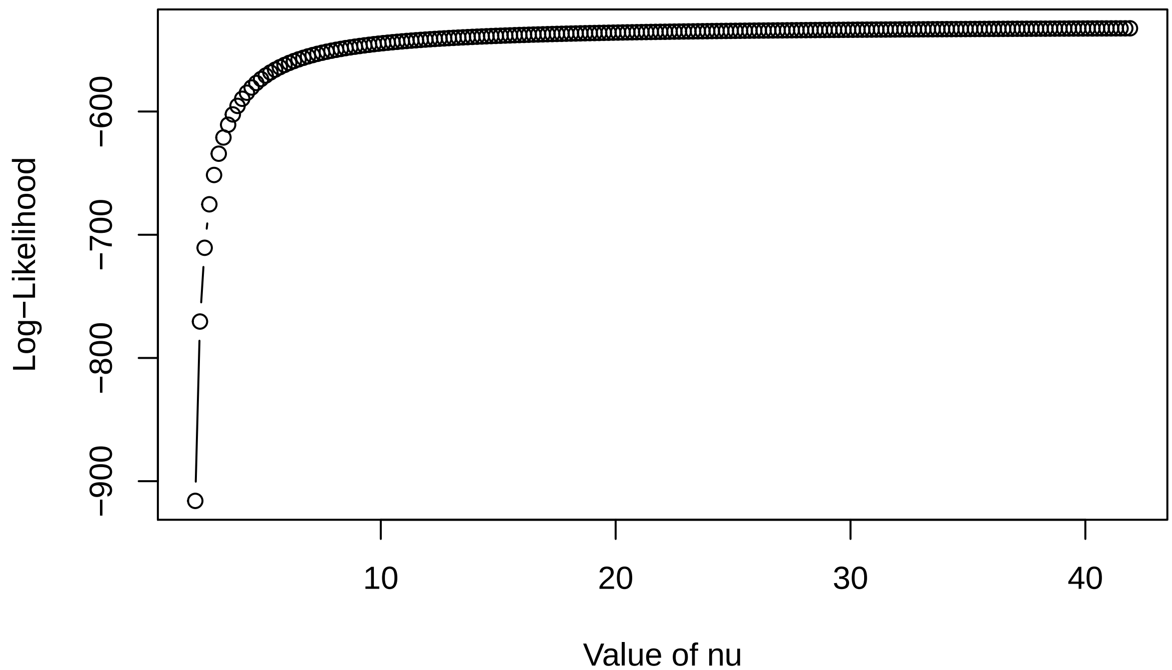
3.3.

The following graph plots the log-likelihood at each iteration of the EM algorithm. To build it, we have included a counter in the EM algorithm recording the value of the log-likelihood at each iteration, along with a rule to indicate the algorithm that enough convergence was reached. Specifically, we have indicated that the algorithm stops when the increments in the log-likelihood are smaller than 0.0001. We can see that convergence is reached at about the tenth iteration.



3.4.

To find an estimate of the degrees of freedom for the latent variable η, we have applied the EM algorithm for a range of values for nu, looking for the moment at which increases in nu cease to increase significantly the log-likelihood. We can see in the graph produced below that for values of nu greater than around 20, there is barely any improvement in the likelihood.


Log−Likelihood
−600
−700
−800
−900
10
20
30
40
Value of nu