Bayesian Statistics

Fabrizio Ruggeri

Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Alfonso Corti 12, I-20133, Milano, Italy, European Union
fabrizio@mi.imati.cnr.it
www.mi.imati.cnr.it/fabrizio/

REGRESSION

- We now consider linear regression (LR), providing a linear relation between a dependent variable (Y) and an independent one (X), sometimes called *covariate*
- We can distinguish 4 cases based on the dimensions of Y and X
  - Simple LR vs. Multiple LR: just one X or multiple X's
  - Univariate LR vs. Multivariate LR: just one-dimensional Y or multiple dimensional Y
- We consider only the simplest case: Univariate Simple Linear Regression
- $Y = \beta_1 + \beta_2 X + \varepsilon$
- β_1, β_2 univariate unknown parameters
- ε error term with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$ unknown
- We consider $\varepsilon \sim \mathcal{N}(0, \sigma^2)$

REGRESSION

- Observations: $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$, $i = 1, \ldots, n$
- X_i's are supposed known here but they could be r.v.'s as well
- We assume that $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. $\mathcal{N}(0, \sigma^2)$
- Notation:$\underline{Y} = (Y_1, \ldots, Y_n)$ and $\underline{X} = (X_1, \ldots, X_n)$
- Likelihood function $L(\beta_1, \beta_2, \sigma^2 | \underline{Y}, \underline{X})$ given by

$$\begin{aligned}\prod_{i=1}^{n} f(Y_i | X_i, \beta_1, \beta_2, \sigma^2) &= \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{2\sigma^2}\} \right\} \\ &\propto \frac{1}{(\sigma^2)^{n/2}} \exp\{\frac{-\sum_{i=1}^{n} (Y_i - \beta_1 - \beta_2 X_i)^2}{2\sigma^2}\}\end{aligned}$$

- Independent priors with known hyperparameters:
  $\beta_1 \sim \mathcal{N}(0, \tau_1^2)$, $\beta_2 \sim \mathcal{N}(0, \tau_2^2)$ and $\sigma^2 \sim \mathcal{IG}(a, b)$

REGRESSION

- Posterior distribution

$$\pi(\beta_1, \beta_2, \sigma^2 | \underline{Y}, \underline{X}) \propto \frac{1}{(\sigma^2)^{n/2}} \exp\{\frac{-\sum_{i=1}^n (Y_i - \beta_1 - \beta_2 X_i)^2}{2\sigma^2}\} \cdot$$

$$\cdot \quad \exp\{-\beta_1^2/(2\tau_1^2)\} \exp\{-\beta_2^2/(2\tau_2^2)\} \frac{1}{(\sigma^2)^{a+1}} \exp\{-b/\sigma^2\}$$

- Conditional on β_1: $\beta_1 | \beta_2, \sigma^2, \underline{Y}, \underline{X} \sim \mathcal{N}\left(\frac{\sum_{i=1}^n (Y_i - \beta_2 X_1)}{n + \sigma^2/\tau_1^2}, \frac{1}{n/\sigma^2 + 1/\tau_1^2}\right)$

$$\pi(\beta_1 | \beta_2, \sigma^2, \underline{Y}, \underline{X}) \propto \exp\{\frac{-(n\beta_1^2 - 2\beta_1 \sum_{i=1}^n (Y_i - \beta_2 X_i)}{2\sigma^2}\} \exp\{-\beta_1^2/(2\tau_1^2)\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{n}{\sigma^2} + \frac{1}{\tau_1^2}\right)\beta_1^2 - 2\frac{\beta_1}{\sigma^2}\sum_{i=1}^n (Y_i - \beta_2 X_i)\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2(n/\sigma^2 + 1/\tau_1^2)^{-1}}\left[\beta_1^2 - 2\frac{\beta_1}{\sigma^2}\frac{\sum_{i=1}^n (Y_i - \beta_2 X_i)}{n/\sigma^2 + 1/\tau_1^2}\right]\right\}$$

REGRESSION

- Conditional on β_2: $\beta_2|\beta_1, \sigma^2, \underline{Y}, \underline{X} \sim \mathcal{N}\left(\frac{\sum_{i=1}^n X_i(Y_i - \beta_1)}{\sum_{i=1}^n X_i^2 + \sigma^2/\tau_1^2}, \frac{1}{\sum_{i=1}^n X_i^2/\sigma^2 + 1/\tau_1^2}\right)$

$$\begin{aligned}
\pi(\beta_2|\beta_1, \sigma^2, \underline{Y}, \underline{X}) &\propto \exp\{-\frac{\beta_2^2 \sum_{i=1}^n X_i^2 - 2\beta_2 \sum_{i=1}^n X_i(Y_i - \beta_1)}{2\sigma^2}\} \exp\{-\beta_2^2/(2\tau_2^2)\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\frac{\sum_{i=1}^n X_i^2}{\sigma^2} + \frac{1}{\tau_2^2}\right)\beta_2^2 - 2\frac{\beta_2}{\sigma^2}\sum_{i=1}^n X_i\left(Y_i - \beta_1\right)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2(\sum_{i=1}^n \frac{X_i^2}{\sigma^2} + \frac{1}{\tau_1^2})^{-1}}\left[\beta_2^2 - 2\frac{\beta_2}{\sigma^2}\frac{\sum_{i=1}^n X_i\left(Y_i - \beta_1\right)}{\sum_{i=1}^n X_i^2/\sigma^2 + 1/\tau_1^2}\right]\right\}
\end{aligned}$$

- Conditional on σ^2: $\sigma^2|\beta_1, \beta_2, \underline{Y}, \underline{X} \sim \mathcal{IG}\left(a + n/2, b + \frac{\sum_{i=1}^n (Y_i - \beta_1 - \beta_2 X_i)^2}{2}\right)$

$$\pi(\sigma^2|\beta_1, \beta_2, \underline{Y}, \underline{X}) \propto \frac{1}{(\sigma^2)^{n/2}} \exp\{-\frac{\sum_{i=1}^n (Y_i - \beta_1 - \beta_2 X_i)^2}{2\sigma^2}\} \frac{1}{(\sigma^2)^{a+1}} \exp\{-b/\sigma^2\}$$

- $\Rightarrow$ Gibbs sampling

REGRESSION

- Different approach, with slight change, from Press, J. (2002), *Subjective and Objective Bayesian Statistics*, Wiley
- $\pi(\beta_1, \beta_2, \sigma^2) \propto \sigma^2$
- Joint posterior

$$\pi(\beta_1, \beta_2, \sigma^2 | \underline{Y}, \underline{X}) \quad \propto \quad \frac{1}{(\sigma^2)^{n/2}} \text{exp}\{-\frac{\sum_{i=1}^{n}(Y_i - \beta_1 - \beta_2 X_i)^2}{2\sigma^2}\}\frac{1}{\sigma^2}$$

- Integrating out $\sigma^2 \Rightarrow (\beta_1, \beta_2)|\underline{Y}, \underline{X}$ bivariate Student distribution

$$\pi(\beta_1, \beta_2 | \underline{Y}, \underline{X}) \propto \frac{1}{\left[\sum_{i=1}^{n}(Y_i - \beta_1 - \beta_2 X_i)^2\right]^{n/2}}$$

- Bivariate Student's $t \Rightarrow$ Marginal Student's t
- $\Rightarrow \beta_1|\underline{Y}, \underline{X} \sim t$, $\beta_2|\underline{Y}, \underline{X} \sim t$ and $\sigma^2|\beta_1, \beta_2, \underline{Y}, \underline{X} \sim \mathcal{IG}$ as before
- $\sigma_2^{(j)}$ from inverse gamma with values $\beta_1^{(j)}$ and $\beta_2^{(j)}$ generated from the t-distributions

REGRESSION

- Another different approach from Cowles, M.K., (2013), *Applied Bayesian Statistics*, Springer
- Center the X_i's around their mean $\overline{X}$:$\Rightarrow X_i \to X_i - \overline{X}$
- $\Rightarrow Y_i|X_i, \beta_1, \beta_2, \sigma^2 \sim \mathcal{N}(\beta_1 + \beta_2(X_i - \overline{X}), \sigma^2),\ i = 1, \ldots, n$
- $\pi(\beta_1, \beta_2, \sigma^2) \propto \sigma^2$
- We will consider three sufficient statistics: $\widehat{\beta}_1, \widehat{\beta}_2, SSR$ (sum of squared residuals)
- Given a r.v. X with density $f(X|\theta)$, a statistic $t = T(X)$ is said sufficient for θ if $f(X|t = T(X)$ does not depend on θ
- In words, a sufficient statistic contains all the information provided by the data about the model parameters
- Those statistics are estimators from a frequentist viewpoint

REGRESSION

- Likelihood: $\prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2\sigma^2} \left[Y_i - \beta_1 - \beta_2(X_i - \overline{X})\right]^2\} \right\}$ i.e.
$$\frac{1}{\left[\sqrt{2\pi}\sigma\right]^n} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left[Y_i - \beta_1 - \beta_2(X_i - \overline{X})\right]^2\}$$

- Loglikelihood $\propto l(\beta_1, \beta_2, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left[Y_i - \beta_1 - \beta_2(X_i - \overline{X})\right]^2$

- $\frac{\partial l}{\partial \beta_1} = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left[Y_i - \beta_1 - \beta_2(X_i - \overline{X})\right] = 0 \Leftrightarrow \sum_{i=1}^{n} Y_i - n\beta_1 - \beta_2 \sum_{i=1}^{n} \left[X_i - \overline{X}\right] = 0$

- $\Rightarrow \widehat{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i}{n} = \overline{Y}$ since $\sum_{i=1}^{n} \left[X_i - \overline{X}\right] = 0$

- Note: if we had not centered the X_i around $\overline{X}$, we would have got $\widehat{\beta}_1 = \overline{Y} - \widehat{\beta}_2 \overline{X}$

REGRESSION

$$\begin{aligned}
\frac{\partial l}{\partial \beta_2} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})\left[Y_i - \beta_1 - \beta_2(X_i - \overline{X})\right] = 0 \\
&\Leftrightarrow \sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \widehat{\beta}_1) - \widehat{\beta}_2 \sum_{i=1}^{n}(X_i - \overline{X})^2 = 0 \\
&\Leftrightarrow \widehat{\beta}_2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}
\end{aligned}$$

- We have plugged-in $\widehat{\beta}_1$ and $\widehat{\beta}_2$ since the equality conditions are to be satisfied by them

- Sum of squared residuals $SSR = \sum_{i=1}^{n}\left[Y_i - \widehat{\beta}_1 - \widehat{\beta}_2(X_i - \overline{X})\right]^2$

- Remember: in simple linear regression $\Rightarrow$ Sample variance $S = \dfrac{SSR}{n-2}$

- We now go back to the Bayesian computations, using $\widehat{\beta}_1, \widehat{\beta}_2$ and SSR

REGRESSION

- Priors: $\pi(\beta_1) \propto c_1, \pi(\beta_2) \propto c_2, \sigma^2 \propto 1/\sigma^2$
- $\Rightarrow \pi(\beta_1, \beta_2, \sigma^2) \propto 1/\sigma^2$
- Posterior $\pi(\beta_1, \beta_2, \sigma^2 | \underline{X}, \underline{Y})$

$$\propto \quad \frac{1}{(\sigma^2)^{n/2}} \exp\{-\frac{\sum_{i=1}^n \left[Y_i - \beta_1 - \beta_2(X_i - \overline{X})\right]^2}{2\sigma^2}\} \cdot \frac{1}{\sigma^2}$$

$$\propto \quad \frac{1}{(\sigma^2)^{n/2+1}} \exp\{-\frac{\sum_{i=1}^n \left\{\left[Y_i - \widehat{\beta}_1 - \widehat{\beta}_2(X_i - \overline{X})\right] - (\beta_1 - \widehat{\beta}_1) - (\beta_2 - \widehat{\beta}_2)(X_i - \overline{X})\right\}^2}{2\sigma^2}\}$$

REGRESSION

$$
\begin{aligned}
\propto \quad & \frac{1}{(\sigma^2)^{n/2+1}} \exp\{-\frac{\sum_{i=1}^n \left[Y_i - \widehat{\beta}_1 - \widehat{\beta}_2(X_i - \overline{X})\right]^2 + \sum_{i=1}^n (\beta_1 - \widehat{\beta}_1)^2}{2\sigma^2}\} \\
\cdot \quad & \exp\{-\frac{\sum_{i=1}^n (\beta_2 - \widehat{\beta}_2)^2 (X_i - \overline{X})^2}{2\sigma^2}\} \cdot \\
\cdot \quad & \exp\{-\frac{2\sum_{i=1}^n \left[Y_i - \widehat{\beta}_1 - \widehat{\beta}_2(X_i - \overline{X})\right](\beta_1 - \widehat{\beta}_1)}{2\sigma^2}\} \cdot \\
\cdot \quad & \exp\{-\frac{2\sum_{i=1}^n \left[Y_i - \widehat{\beta}_1 - \widehat{\beta}_2(X_i - \overline{X})\right](\beta_2 - \widehat{\beta}_2)(X_i - \overline{X})}{2\sigma^2}\} \cdot \\
\cdot \quad & \exp\{-\frac{2\sum_{i=1}^n (\beta_1 - \widehat{\beta}_1)(\beta_2 - \widehat{\beta}_2)(X_i - \overline{X})}{2\sigma^2}\}
\end{aligned}
$$

- All the double products will cancel, as we are going to see

REGRESSION

- $\sum_{i=1}^{n} \left[Y_i - \widehat{\beta}_1 - \widehat{\beta}_2(X_i - \overline{X})\right] (\beta_1 - \widehat{\beta}_1) = (\beta_1 - \widehat{\beta}_1) \left\{ \sum_{i=1}^{n} \left[Y_i - \widehat{\beta}_1\right] - \widehat{\beta}_2 \sum_{i=1}^{n} (X_i - \overline{X}) \right\}$

  $\Rightarrow (\beta_1 - \widehat{\beta}_1) \left\{ \sum_{i=1}^{n} \left[Y_i - \overline{Y}\right] - \widehat{\beta}_2 \sum_{i=1}^{n} (X_i - \overline{X}) \right\} = 0$

- $\sum_{i=1}^{n} \left[Y_i - \widehat{\beta}_1 - \widehat{\beta}_2(X_i - \overline{X})\right] (\beta_2 - \widehat{\beta}_2)(X_i - \overline{X}) =$

  $= (\beta_2 - \widehat{\beta}_2) \left\{ \sum_{i=1}^{n} \left[Y_i - \widehat{\beta}_1\right] (X_i - \overline{X}) - \widehat{\beta}_2 \sum_{i=1}^{n} (X_i - \overline{X})^2 \right\} =$

  $= (\beta_2 - \widehat{\beta}_2) \left\{ \sum_{i=1}^{n} \left[Y_i - \overline{Y}\right] (X_i - \overline{X}) - \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 \right\} = 0$

- $\sum_{i=1}^{n} (\beta_1 - \widehat{\beta}_1)(\beta_2 - \widehat{\beta}_2)(X_i - \overline{X}) = (\beta_1 - \widehat{\beta}_1)(\beta_2 - \widehat{\beta}_2) \sum_{i=1}^{n} (X_i - \overline{X}) = 0$

REGRESSION

- Posterior $\pi(\beta_1, \beta_2, \sigma^2 | \underline{X}, \underline{Y})$

$$\propto \quad \frac{1}{(\sigma^2)^{n/2+1}} \text{exp}\{-\frac{\sum_{i=1}^n \left[Y_i - \widehat{\beta}_1 - \widehat{\beta}_2(X_i - \overline{X})\right]^2 + \sum_{i=1}^n (\beta_1 - \widehat{\beta}_1)^2}{2\sigma^2}\} \cdot$$

$$\cdot \quad \text{exp}\{-\frac{\sum_{i=1}^n (\beta_2 - \widehat{\beta}_2)^2 (X_i - \overline{X})^2}{2\sigma^2}\} \cdot$$

$$\propto \quad \frac{1}{(\sigma^2)^{n/2+1}} \text{exp}\{-\frac{SSR + n(\beta_1 - \widehat{\beta}_1)^2 + (\beta_2 - \widehat{\beta}_2)^2 \sum_{i=1}^n (X_i - \overline{X})^2}{2\sigma^2}\}$$

$$\propto \quad \frac{1}{(\sigma^2)^{(n+1)/2}} \text{exp}\{-\frac{SSR + n(\beta_1 - \widehat{\beta}_1)^2}{2\sigma^2}\} \frac{1}{(\sigma^2)^{1/2}} \text{exp}\{-\frac{(\beta_2 - \widehat{\beta}_2)^2 \sum_{i=1}^n (X_i - \overline{X})^2}{2\sigma^2}$$

- We integrate out $\beta_2 \sim \mathcal{N}(\widehat{\beta}_2, \frac{\sigma^2}{\sum_{i=1}^n (X_i - \overline{X})^2})$

- $\Rightarrow \pi(\beta_1, \sigma^2 | \underline{X}, \underline{Y}) \propto \frac{1}{(\sigma^2)^{(n+1)/2}} \text{exp}\{-\frac{SSR + n(\beta_1 - \widehat{\beta}_1)^2}{2\sigma^2}\}$

REGRESSION

- We integrate out $\sigma^2 \sim \mathcal{IG}((n-1)/2, (SSR + n(\beta_1 - \widehat{\beta}_1)^2)/2)$
- $\Rightarrow \beta_1 | \underline{X}, \underline{Y}$ will have a Student's t-distribution
- Student's t density $f(t)$ with mean 0, scale 1 and degrees of freedom (d.f.) ν

$\Rightarrow f(t) \propto \dfrac{1}{\left[1 + \frac{t^2}{\nu}\right]^{(\nu+1)/2}}$

$$
\begin{aligned}
\pi(\beta_1 | \underline{X}, \underline{Y}) &\propto \frac{1}{\left[SSR + n(\beta_1 - \widehat{\beta}_1)^2\right]^{(n-1)/2}} \\
&\propto \frac{1}{\left[1 + \dfrac{(\beta_1 - \widehat{\beta}_1)^2}{SSR/n}\right]^{[(n-2)+1]/2}} \\
&\propto \frac{1}{\left[1 + \dfrac{\frac{(\beta_1 - \widehat{\beta}_1)^2}{S^2/n}}{n-2}\right]^{[(n-2)+1]/2}}
\end{aligned}
$$

REGRESSION

- I used $\dfrac{SSR}{n} = \dfrac{SSR}{n-2}\dfrac{n-2}{n} = S^2\dfrac{n-2}{n}$ where S^2 is an unbiased estimator of σ^2

- $\Rightarrow \beta_1|\underline{X}, \underline{Y} \sim t_{n-2}(\widehat{\beta}_1, S^2/n)$, with mean $\widehat{\beta}_1$, scale S^2/n and $(n-2)$ d.f.

- Similarly, it possible to prove that $\beta_2|\underline{X}, \underline{Y} \sim t_{n-2}(\widehat{\beta}_2, \dfrac{S^2}{\sum_{i=1}^n (X_i - \overline{X})^2})$

- We go back to $\pi(\beta_1, \sigma^2|\underline{X}, \underline{Y}) \propto \dfrac{1}{(\sigma^2)^{n/2+1/2}} \exp\{-\dfrac{SSR + n(\beta_1 - \widehat{\beta}_1)^2}{2\sigma^2}\}$

- We integrate out $\beta_1 \sim \mathcal{N}(\widehat{\beta}_1, \sigma^2/n)$

- $\Rightarrow \pi(\sigma^2|\underline{X}, \underline{Y}) \propto \dfrac{1}{(\sigma^2)^{n/2}} \exp\{-\dfrac{SSR}{2\sigma^2}\}$

- $\Rightarrow \sigma^2|\underline{X}, \underline{Y} \sim \mathcal{IG}(n/2 - 1, SSR/2)$

REGRESSION

- Summarising, we have been able to get the three posteriors in closed form:
  - $\beta_1|\underline{X}, \underline{Y} \sim t_{n-2}(\widehat{\beta}_1, S^2/n)$
  - $\beta_2|\underline{X}, \underline{Y} \sim t_{n-2}(\widehat{\beta}_2, \dfrac{S^2}{\sum_{i=1}^n (X_i - \overline{X})^2})$
  - $\sigma^2|\underline{X}, \underline{Y} \sim \mathcal{IG}(n/2 - 1, SSR/2)$
- Note that the posterior means for β_1 and β_2 coincide with the MLEs: this is not uncommon when considering improper priors
- Warning: the posteriors are proper if and only if $n > 2$ and not all X_i's are equal!
- The posterior mean of σ^2 exists if and only if $n > 4$ since the mean of the inverse gamma $\mathcal{IG}(a, b)$ is $\dfrac{b}{a-1} = \dfrac{SSR/2}{(n/2-1)-1} = \dfrac{SSR}{n-4}$

REGRESSION

- We have found the posterior distributions of the parameters, either in closed form or in a suitable one to apply MCMC $\Rightarrow$ now we can estimate them, e.g., considering the posterior mean, and build credible intervals in a way similar to what we saw earlier (and I will not repeat it)
- When considering more than one covariate, i.e., $X_1, \ldots, X_n$, still Gaussian priors should be considered for each of them
- Similarly to the frequentist approach, there is an interest for the covariates which are significant
  - Instead of considering p-values, Bayesians look for a credible interval and check if 0 belongs to it
  - If the credible interval contains 0 then the covariate is not significant; otherwise, it is
  - We will see an example next
- If Y is multivariate, then multivariate Gaussian distributions are chosen to model the observations and as a prior for the mean, while an Inverse Wishart distribution is chosen for the covariance matrix

REGRESSION

- Both frequentist and Bayesian methods will be applied in the next example
- 713 observations corresponding to the days where the prices of the Bitcoins in 8 different exchange markets were recorded together with the prices of the classical assets and the exchange rates
- We will use the package `rstanarm` and the function `stan_glm`, whose usage is similar to `lm`
- Use of improper priors leading to results close to frequentist ones
- You could try other priors, using the `R` tutorials, like `?stan_glm`
- For this example, I tried `stan_lm`, the very equivalent of `lm` (both about linear models) but it did not work, so that I used the one for generalised linear models
- I first present the commands for the frequentist analysis

REGRESSION

```
rm(list=ls()) # Clear the environment
install.packages("ggplot2",dependencies=TRUE)
install.packages("readxl",dependencies=TRUE)
install.packages("corrplot",dependencies=TRUE)
library(ggplot2);library(readxl);library(corrplot)
exchanges<-read_excel("exchanges.xlsx") # Read in working directory
data<-exchanges
data1<-data[-1] # Remove the first column from data
# New dataset with returns instead of prices: (log(x)-log(x-1))
data2<-as.data.frame(sapply(data1,function(x)diff(log(x),lag=1)))
attach(data2) # Bring the names of the variables directly into memory
```

REGRESSION

```
# Multiple linear regression [btc_coinbase on all other variables]
model_3<-lm(btc_coinbase~.,data=data2)
summary(model_3)
# Get and plot residuals
res<-model_3$residuals
plot(res,type='l')
install.packages("rstanarm",dependencies=TRUE)
library(rstanarm)
model_b<-stan_glm(btc_coinbase~.,data=data2)
summary(model_b, digits=3)
# Get and plot residuals
resb<-model_b$residuals
plot(resb,type='l')
```

Default priors: standard Gaussian for intercept and coefficients and exponential of parameter 1 for σ^2

REGRESSION

Results based on MLE

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.0001059  0.0003123   0.339  0.73454
btc_kraken    0.0210321  0.0123977   1.696  0.09025 .
btc_bitstamp  0.0384385  0.0359272   1.070  0.28503
btc_itbit     0.0130343  0.0256007   0.509  0.61082
btc_bitfinex  0.2297741  0.0315236   7.289 8.47e-13 ***
btc_hitbtc    0.0821093  0.0184755   4.444 1.03e-05 ***
btc_gemini    0.5981632  0.0308680  19.378  < 2e-16 ***
btc_bittrex   0.0056419  0.0145595   0.388  0.69850
usdyuan      -0.1045943  0.2066436  -0.506  0.61291
usdeur        0.2060414  0.0986501   2.089  0.03710 *
gold          0.0712161  0.0575053   1.238  0.21597
oil          -0.0595675  0.0192726  -3.091  0.00208 **
sp500        -0.0952889  0.0569865  -1.672  0.09495 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

REGRESSION

- `Warmup` is better known as *burn-in*, i.e. the first values are discarded because affected by the starting values
- We now consider different priors, like Student t for each coefficient, Cauchy for the intercept and exponential for σ^2
- We consider also 1 chains, setting a seed and the number of iterations

```
model_b<-stan_glm(btc_coinbase~.,chains=1,seed=12345,iter=250,
prior=student_t(df=4,0,2.5),prior_intercept=cauchy(0,10),prior_aux =
exponential(1/2),data=data2)
summary(model_b, digits=3)
print(model_b)
prior_summary(model_b) # To see the chosen priors
library(bayesplot)
mcmc_dens(model_b)
library(bayestestR)
hdi(model_b)
```

LOGISTIC REGRESSION

- The previous example dealt with continuous variables but what about a response (*dependent variable*) taking only a finite number of integer values?
- Consider people applying for mortgages (or subject to surgery): are they able to pay the mortgage back (or will they survive)?
- The observations are 1's (pays back/survives) and 0's (does not pay back/dies)
- We are still interested in studying the effect of covariates (*independent variables*), like age and gender, on the final result
- We cannot use $Y = \beta_1 + \beta_2 X + \epsilon$ with $Y = 0, 1$ since it is almost impossible to choose r.h.s. terms such that there is always either 0 or 1 in the l.h.s.
- We consider $\pi = P(Y = 1)$ but we cannot use $\pi = \beta_1 + \beta_2 X + \epsilon$ since it is almost impossible to choose r.h.s. terms such that the l.h.s. will be always between 0 and 1
- (`logit`) transformation: $\log\left(\frac{\pi}{1-\pi}\right) = X^{'}\beta$, with $X^{'}, \beta$ vectors of size k
- Earlier: $X^{'} = (1, X), \beta^{'} = (\beta_1, \beta_2)$

LOGISTIC REGRESSION

- $\log\left(\frac{\pi}{1-\pi}\right) = X^{'}\beta \Rightarrow \pi = \frac{e^{X'\beta}}{1+e^{X'\beta}}$

- For each $i = 1, \ldots, n$, consider n_i observations (y_i, x_i) and the related probability π_i (e.g. y_i, out of n_i, persons with features x_i, paid the mortgage back)

- $\underline{y} = (y_1, \ldots, y_n)$, $\underline{x} = (x_1, \ldots, x_n)$, $\underline{\pi} = (\pi_1, \ldots, \pi_n)$ and $\underline{n} = (n_1, \ldots, n_n)$

- We consider a Binomial model (Bernoulli if $n_i = 1$)

$$P(Y_i = y_i|\pi_i, n_i, x_i) = \binom{n_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i} = \binom{n_i}{y_i}\left(\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}\right)^{y_i}\left(\frac{1}{1+e^{x_i'\beta}}\right)^{n_i-y_i}$$

- Likelihood: $\prod_{i=1}^{n}\binom{n_i}{y_i}\frac{e^{y_i x_i'\beta}}{\left(1+e^{x_i'\beta}\right)^{n_i}}$

- Prior distribution on β: e.g. Multivariate Gaussian (simplest: product of independent univariate Gaussian distributions)

LOGISTIC REGRESSION

- Survey of 3200 residents in a small area of Bangladesh suffering from arsenic contamination of groundwater*

- Respondents with elevated arsenic levels in their wells were encouraged to switch their water source to a safe well in the nearby area and the survey was conducted several years later to learn which of the affected residents had switched wells

- The goal of the analysis is to learn about the factors associated with switching wells

- To start, we will use `dist` (the distance from the respondent's house to the nearest well with safe drinking water) as the only predictor of `switch` (1 if switched, 0 if not).

- Then we will expand the model by adding the arsenic level of the water in the resident's own well as a predictor and then we will add all variables

- After loading the `wells` data, we first rescale the `dist` variable (measured in meters) so that it is measured in units of 100 meters

*Example due to Gabry and Goodrich (website), based on Gelman and Hill's book

LOGISTIC REGRESSION

```
library(rstanarm)
data(wells)
wells$dist100 <- wells$dist / 100
head(wells)
library(ggplot2)
ggplot(wells,aes(x=dist100,y=after_stat(density),fill=switch==1)) +
geom_histogram() + scale_fill_manual(values=c("gray30", "skyblue"))
```

- Distribution of dist100: 1737 residents who switched (blue bars) and 1283 who did not (dark grey bars)
- It is just one density (not two!) which describes also the proportion of switch/no switch at various distances
- For the residents who switched wells, the distribution of `dist100` is more concentrated at smaller distances
- We use a Student t prior with coefficients close to 0 but with chances of being large (less likely under Gaussian)

LOGISTIC REGRESSION

```
t_prior <- student_t(df = 7, location = 0, scale = 2.5)
fit1<-stan_glm(switch ~ dist100,data=wells,seed = 12345,
  family = binomial(link = "logit"),
  prior = t_prior, prior_intercept = t_prior)
summary(fit1,digits=3)
round(posterior_interval(fit1, prob = 0.5), 3) # digits=3
fit2 <- update(fit1, formula = switch ~ dist100 + arsenic)
round(coef(fit2), 3)
summary(fit2,digits=3)
fit3<-stan_glm(switch ~ arsenic+assoc+educ+dist100,data=wells,
family = binomial(link = "logit"),seed = 12345,
  prior = t_prior, prior_intercept = t_prior)
  summary(fit3,digits=3)
```

LOGISTIC REGRESSION

- `switch` – binary/dummy (0 or 1) for well-switching
- **0.468**: `arsenic` – arsenic level in respondent's well
- **-0.897**: `dist100` – distance (100 meters) from the respondent's house to the nearest well with safe drinking water
- **-0.125**: `association` – binary/dummy (0 or 1) if member(s) of household participate in community organizations
- **0.043**: `educ` – years of education (head of household)
- Interpretation of those numbers (posterior means)?

LOGISTIC REGRESSION

- Using the coefficient estimates from the first model, we can plot the predicted probability of `switch = 1` (as a function of `dist100`)
- `plogis` is the cdf of a logistic distribution

```
t_prior <- student_t(df = 7, location = 0, scale = 2.5)
fit1<-stan_glm(switch ~ dist100,data=wells,seed = 12345,
  family = binomial(link = "logit"),
  prior = t_prior, prior_intercept = t_prior)
summary(fit1,digits=3)
pr_switch <- function(x, ests) plogis(ests[1] + ests[2] * x)
coef(fit1)[1]; coef(fit1)[2]
aa=seq(0,12,0.25)
plot(aa,pr_switch(aa,coef(fit1)),type='l')
```