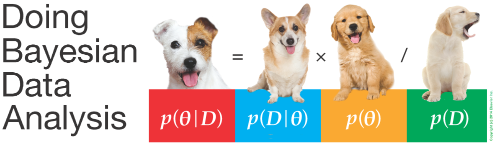
Bayesian Methods for Data Science (DATS 6450 - 11)
Metric Predicted Variable with One Metric Predictor

Yuxiao Huang

Data Science, Columbian College of Arts & Sciences
George Washington University
yuxiaohuang@gwu.edu

November 6, 2019

Reference



Picture courtesy of the book website

- This set of slices is an excerpt of the book by Professor John K. Kruschke, with some trivial changes by the creator of the slides
- Please find the reference to and website of the book below:
  - Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier
  - `https://sites.google.com/site/doingbayesiandataanalysis/`

Overview

1. Generalized Linear Model
2. Simple Linear Regression
3. Robust Linear Regression

Formal expression of the GLM

- The formal definition of the GLM is

$$\mu = f\Big(\text{lin}(x), [\text{parameters}]\Big)$$
$$y \sim \text{pdf}\Big(\mu, [\text{parameters}]\Big)$$

- Here:
    - x are the predictors
    - lin is the linear function of x
    - f is the inverse linkage function (a.k.a., activation function)
    - μ is the central tendency
    - y is the data
- This is illustrated in Figure 15.9 (see next page)

Figure 15.9

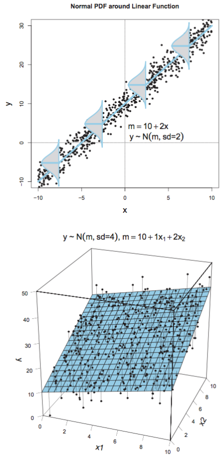


Figure 15.9: Examples of points normally distributed around a linear function. The upper panel shows normal distributions superimposed on the linear function to emphasize that the random variability is vertical, along the *y* axis, and centered on the line. The lower panel shows each datum connected to the plane by a dotted line, to again emphasize the vertical displacement from the plane. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

The data generation process

- In simple linear regression, there is only one predictor
- **Q:** Can you explain how data are generated based on this model?

The data generation process

- In simple linear regression, there is only one predictor
- **Q:** Can you explain how data are generated based on this model?
- **A:** There are two steps:
  1. generating μ based on the liner model determined by the single predictor x and its weights β_0 and β_1

     $$\mu = \beta_0 + \beta_1 x$$

  2. generating y based on a normal distribution determined by μ and σ

     $$y \sim \mathrm{norm}(\mu, \sigma)$$

- This is illustrated in Figure 17.1 (see next page)

Figure 17.1

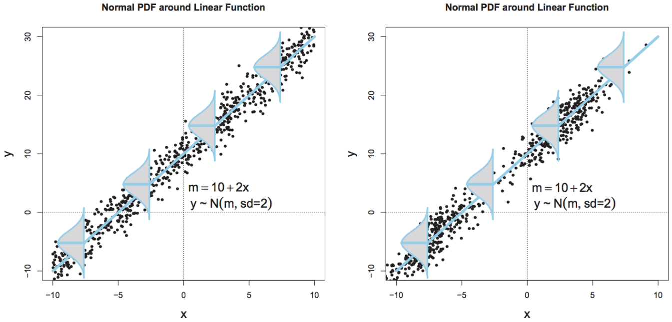


Figure 17.1: Examples of points normally distributed around a linear function. (The left panel repeats Figure 15.9, p. 406.) The model assumes that the data y are normally distributed vertically around the line, as shown. Moreover, the variance of y is the same at all values of x. The model puts no constraints on the distribution of x. The right panel shows a case in which x are distributed bimodally, whereas in the left panel the x are distributed uniformly. In both panels, there is homogeneity of variance. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

Outliers and robust estimation: the t distribution

- In reality, there could be outliers in the data
- Normal distribution may not be able to address the outliers due to its thin tails
- A more robust distribution is (student) t distribution, which has three parameters:
    - μ: the central tendency
    - σ: the standard deviation
    - ν (where $\nu \geq 1$): the normality parameter
- Figure 16.4 to 16.6 (see next three pages) show some t distributions with different value of ν, and the difference between t and normal distribution

Figure 16.4

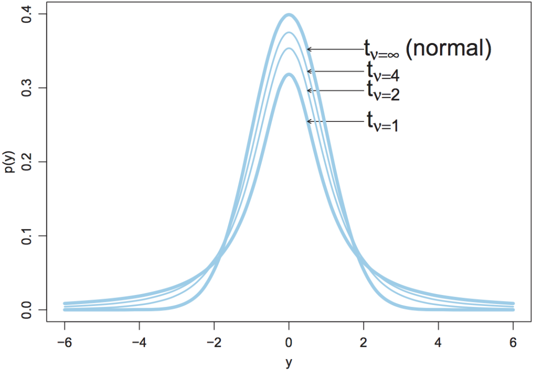


Figure 16.4: Examples of t distributions. In all cases, $\mu = 0$ and $\sigma = 1$. The normality parameter, ν, controls the heaviness of the tails. Curves for different values of ν are superimposed for easy comparison. The abscissa is labelled as y (not x) because the distribution is intended to describe predicted data. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

Figure 16.5

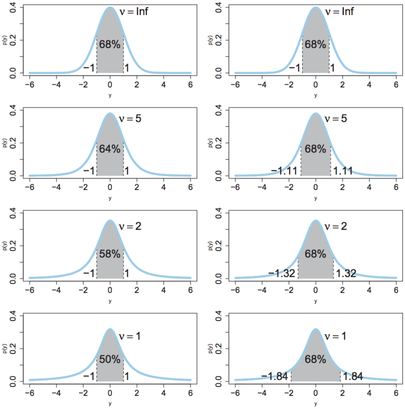


Figure 16.5: Examples of t distributions with areas under the curve. In all cases, $\mu = 0$ and $\sigma = 1$. Rows show different values of the normality parameter, ν. Left column shows area under the t distribution from $y = -1$ to $y = +1$. Right column shows values of $\pm y$ needed for an area of 68.27%, which is the area under a standardized normal curve from $y = -1$ to $y = +1$. The abscissa is labelled as y (not x) because the distribution is intended to describe predicted data. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

Figure 16.6

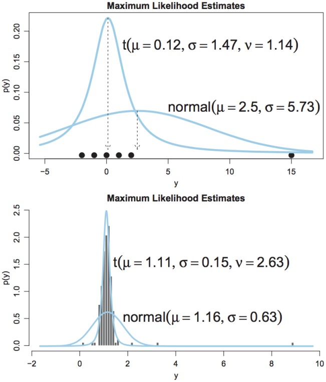


Figure 16.6: The maximum likelihood estimates of normal and *t* distributions fit to the data shown. Upper panel shows "toy" data to illustrate that the normal accommodates an outlier only by enlarging its standard deviation and, in this case, by shifting its mean. Lower panel shows actual data (Holcomb & Spalsbury, 2005) to illustrate realistic effect of outliers on estimates of the normal. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

Using the t distribution

- Since t distribution is more robust than normal distribution, instead of using normal distribution in the linear model, we use the t distribution to do so
- Here, the new parameter ν follows an exponential distribution with mean λ
- Figure 16.7 (see next page) shows the exponential distribution
- Figure 17.2 (see next page) shows the hierarchical structure of the robust model

Figure 16.7

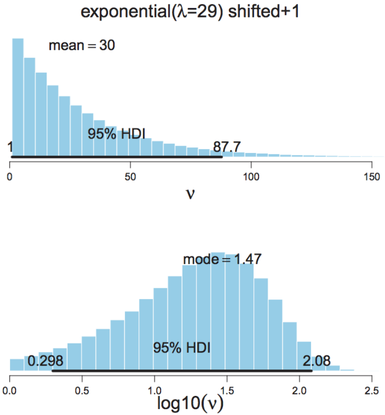


Figure 16.7: The prior on the normality parameter. Upper panel shows the shifted exponential distribution on the original scale of ν. Lower panel shows the same distribution on a logarithmic scale. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

Figure 17.2

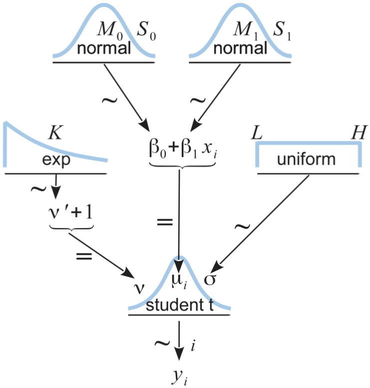


Figure 17.2: A model of dependencies for robust linear regression. The datum, y_i at the bottom of the diagram, is distributed around the central tendency μ_i, which is a linear function of x_i. Compare with Figure 16.11 on p. 437. Copyright © Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan. 2nd Edition.* Academic Press / Elsevier.

Robust linear regression in JAGS

- The posteriors of parameters in robust linear regression can be estimated using JAGS
- There are five major steps
  1. standardize the data
  2. specify the model
  3. generate the MCMC chain
  4. transform the standardized data back to the original form
  5. plot the posteriors
- See `Jags-Ymet-Xmet-Mrobust-Example.R` for details