# Computer Hardware Estimation using Hierarchical Poisson Regression

Capstone project report
Bayesian Statistics: Techniques and Models

La Ngoc Thuy An

6 May 2018

**Abstract**

Computer Hardware Estimation is a task that approximates CPU performance based on system configuration and design. It can be explained as a regression analysis: indicating the relationship between computer's attributes (vendor and model name, memory size, number of input/output channels) and performance measurement metric called Published Relative Performance (PRP). To solve this problem, I conducted experiments with linear and poisson regression and considered using log2 transformation for variables. Poisson model seems to be more complex but lower in residual. Finally, I constructed a hierachical model to improve the DIC. According to the result, hierachical structure extending from poisson regression with log2 transformation of some attributes is the best solution.

## 1 Introduction

CPU performance is one of the most important factors to evaluate computers, reflects how fast computers running programs. Executing time is depend on system configuration and design: Main/cache memory, number of input/output channel. To measure and compare this characteristic, previous researches use a measurement metric called Published Relative Performance (PRP).

Qi Zhou et al handled this long-standing challenge by statistical analyses. They examined carefully each dependent variable and found that the best solution when using linear regression was applying log transformation to some independent variables and PRP. Their result of estimating PRP was saved at ERP variable in the dataset.

In this work, I tried many ways to estimate PRP. My first approach was basically poisson regression as a "traditional" solution of predicting counting number. Taking inspiration from Qi Zhou et al, I also checked out linear regression with log transformation. After that, I compared poisson regression and linear regression. Finally, I constructed a hierarchical graph, based on assumption that computers from same vendor are more likely to be similar than those from different vendors.

## 2 Data

### 2.1 Data Description

The dataset was collected from 209 computers on the market from 1981 - 1984, with machine names and 7 attributes:

- Vendor name (30 unique names)
- Model Name
- MYCT: machine cycle time in nanoseconds (integer $[17, 1500]$).
- MMIN: minimum main memory in KB (integer $[64, 32000]$).
- MMAX: maximum main memory in KB (integer $[64, 64000]$ )
- CACHE: cache memory in KB (integer $[0, 256]$).
- CHMIN: minimum channels in units (integer $[0, 52]$)
- CHMAX: maximum channels in units (integer $[0, 176]$)
- PRP: published relative performance (integer)
- ERP: estimated relative performance from Qi Zhou et al (integer)

MYCT means the actual time computers take complete all of operations to do something. The larger MYCT, the lower performance. MMIN, MMAX, CACHE indicates size of memory device. More memory, less time running. CHMIN, CHMAX are related to transfer data rate, so that faster speed, higher CPU performance.

Qi Zhou et al found that applying log transformation to MYCT, MMIN, MMAX and PRP was the best solution when using linear regression.

For full dataset, more details of data description and Qi Zhou, see this link:
https://archive.ics.uci.edu/ml/datasets/Computer+Hardware.
https://www.slideshare.net/QiGilbertZhou/evaluating-cpu-performance

## 2.2 Explore data

PRP has poisson data shape. As a "traditional" solution of predicting counting number, I immediately think of poisson regression. Beside that, poisson distribution can be seen as right-skew normal distribution, then skewness can be removed by using log transformation. Therefore, using log of PRP, we can apply linear regression.
Inspired by work of Qi Zhou et al, I not only took log of PRP, but also some independent variables. The advantage of this pre-processing step is removing skewness, but it also reduces value ranges. So that variable with highly-right-skew and small range are not suitable. Just MYCT, MMIN, MMAX are satisfied. (See figure 1, 2).

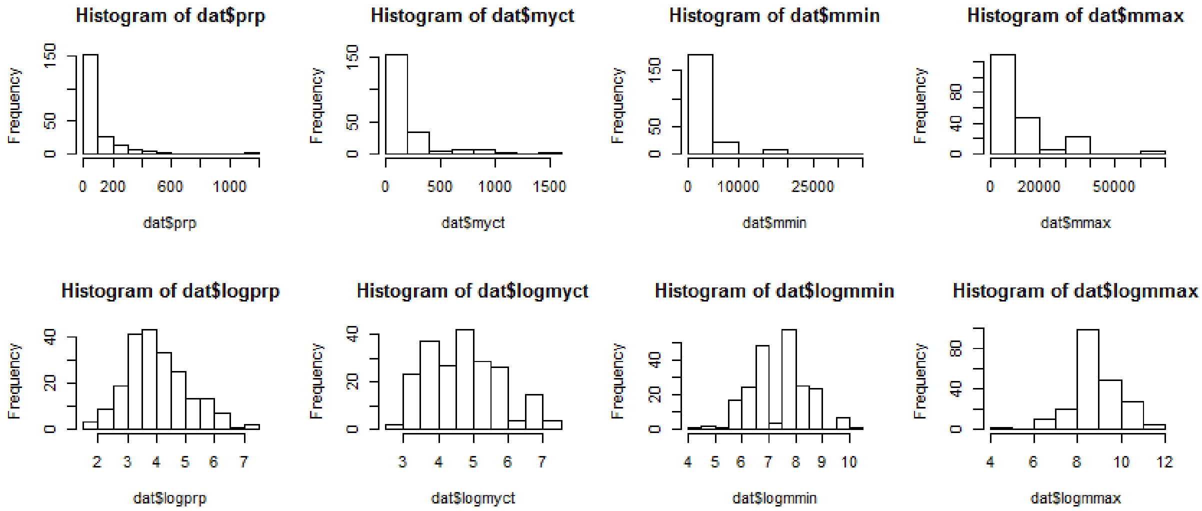


Figure 1: Histogram of variables and their log transfomation

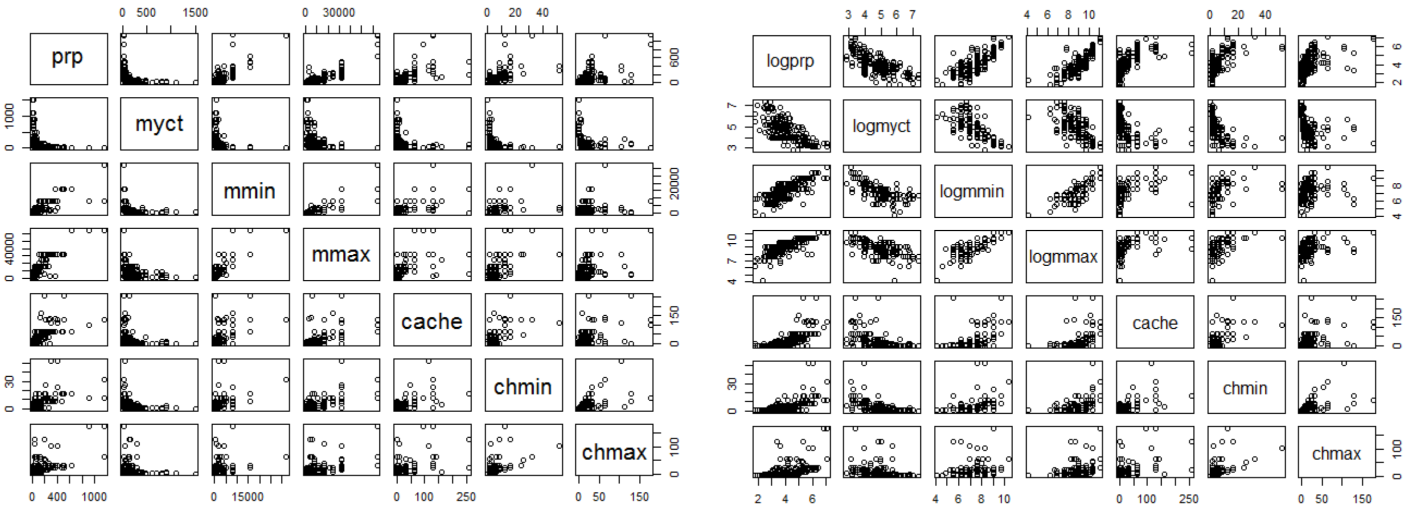


Figure 2: Relationship between PRP and independent variables without (left) and with (right) log transfomation

Plotting relationship between vendor and log(prp), it can be seen that computers from same vendor are more likely to be similar than those from different vendors. (See figure 3).

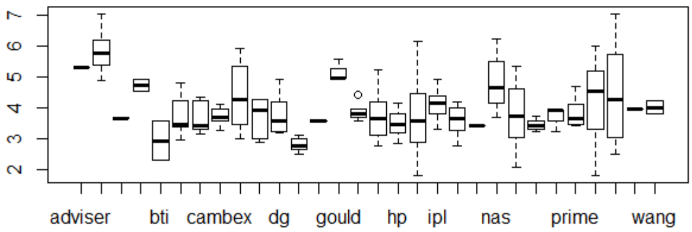


Figure 3: Relationship between vendor and log(PRP)

After this exploring data, I decided to:

- Compare poisson regression and linear regression with log(PRP)
- Compare using those model with and without log transformation of MYCT, MMIN, MMAX.
- Construct hierachical model.

# 3 Model

## 3.1 Postulate a model

1. Orginal attributes
   (a) Linear Regression with log(prp):

$$log(prp_i) = \beta_0 + \sum_{att} \beta_{att} * x_{iatt} + \epsilon_i \sim^{iid} N(0, \sigma^2)$$

$$log(prp_i)|x_i, \beta, \sigma^2 \sim^{ind} N(\beta_0 + \sum_{att} \beta_{att} * x_{iatt}, \sigma^2)$$

   (b) Poisson Regression:

$$prp_i|\lambda_i \sim^{ind} Pois(\lambda_i)$$

$$log(\lambda_i) = \beta_0 + \sum_{att} \beta_{att} * x_{iatt}$$

   with $att$ means attributes, $att \in [myct, mmin, mmax, cache, chmin, chmax]$

   Here, in linear regression, $\mu$ represents expected value of log(PRP), log(prp) is dependent on $att$ and $\beta_0, \beta_{att}$. In poisson regression, $\lambda$ represents expected value of PRP, PRP is dependent on $att$ and $\beta_0, \beta_{att}$

2. Log transformation MYCT, MMIN, MMAX
   Similar to equations above, but $att \in [logmyct, logmmin, logmmax, cache, chmin, chmax]$.

3. Hierarchical Poisson Regression
   Instead of $\beta_0$ for all vendors, use $\beta_{vendor} \sim norm(0.0, 10^6)$

## 3.2 Fit the model

1. Linear Regression with log(prp):

$$log(prp) \sim norm(\mu, \sigma^2)$$

$$\mu = \beta_0 + \sum_{att} \beta_{att} * att$$

$$\beta_0 \sim norm(0.0, 10^6), \beta_{att} \sim norm(0.0, 10^6), \sigma^2 \sim norm(0.0, 10^6)$$

2. Poisson Regression

$$prp \sim poisson(\lambda)$$

$$log(\lambda) = \beta_0 + \sum_{att} \beta_{att} * att$$

$$\beta_0 \sim norm(0.0, 10^6), \beta_{att} \sim norm(0.0, 10^6)$$

with $att$ means attributes, $att \in [myct, mmin, mmax, cache, chmin, chmax]$

Here, in linear regression, $\mu$ represents expected value of log(PRP). log(prp) is dependent on $att$ and $\beta_0, \beta_{att}$. In poisson regression, $\lambda$ represents expected value of PRP.

### 3.3 Check the model

- Convergence: Linear Regression with log(PRP) is easy to reach convergence (lag 50). The second one is Poisson Regression. When using log of MYCT, MMIN, MMAX, coefficients of those variables need larger number of iterators to be converged (lag 2000). In hierachical model, intercepts of vendors have slower convergence speed, too (lag 3000).
- Residual: All models have good residual plots (without any trending or pattern). (See figure 5).
- Precision and Complexity of Model: use DIC.
- Predictive performance: I use Root Mean Square Error (RMSE) to calculate the difference between PRP and prediction. RMSE between PRP and ERP (estimating result of Qi Zho et al) is also be calculated to comapre with my model.

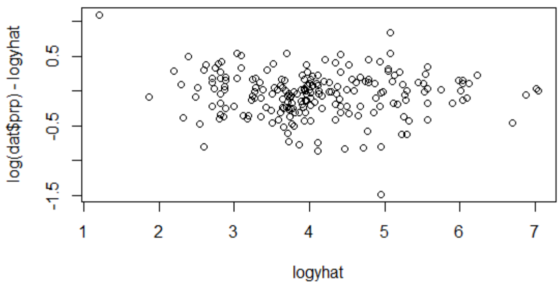


Figure 4: Residual of hierachical poisson regression model

| Experiments | | RMSE | Mean deviance | Penalty | Penalized deviance |
|---|---|---|---|---|---|
| Without log transformation of attributes | Linear Regression with log(PRP) | 198.2 | 269.8 | 8.114 | 277.9 |
| | Poisson Regression | 58.2 | 4781.0 | 7.157 | 4788.0 |
| Log transformation of MYCT, MMIN and MMAX | Linear Regression with log(PRP) | 73.1 | 245.8 | 7.699 | 253.5 |
| | Poisson Regression | 46.0 | 4265.0 | 8.841 | 4274 |
| | Hierachical Model | 39.3 | 2952.0 | 34.55 | 2987 |
| Qi Zhou et al | | 41.68 | | | |

## 4 Result and Conclusions

According to the result table above:

- Whether using log transformation of attributes, poison regression has lower error but higher complexity than Linear Regression. Linear Regression just fits with log(PRP), which is much smaller than PRP, so that it is easy to be converged but reduces sensitivity of meaning.
- Using log transformation for MYCT, MMIN, MMAX improved the performance in both RMSE and DIC:
  - Linear Regression: RMSE: 198 down to 73.1, Mean deviance: 269.8 down to 245.8.
  - Poisson Regression: RMSE: 58.2 down to 46.0, Mean Deviance: 4781.0 down to 4265.0.

  .
- Hierachical is the best solution. It's better than ERP - estimating result of Qi Zhou et al.

Comparing my Linear Regression to Qi Zhou et al, despite of the same pre-processing step (apply log transformation to MYCT, MMIN, MMAX), my result has higher RMSE. Maybe, Qi Zhou et al has some other sophisticated pre-processing or sensitive tunning prior distribution.