SOFTWARE RELIABILITY GROWTH MODELING

Advanced Statistics, Dynamic Programming and Stochastic Control course project

Artur Samigullin
2nd year Big Data Science Master Student
Innopolis University

PART 1. INTRODUCTION

Modelling software reliability used to properly manage product quality and release dates in software development process. Several classes of models have been proposed, including Software Reliability Growth Models, which model reliability estimating the errors that still remain in the software using the history of the detection of errors.

Project focused on measuring performance of Weibull distribution based Software Reliability Growth Model by applying it to open data of development Tizen OS.

We will measure performance of model with following metrics:

1. Goodness of fit (GoF) shows how well the model fits the original data. GoF is calculated as a sum of squared residuals divided by the number of model's freedom degrees
2. The accuracy of final point (AcFP) corresponds to the effectiveness of the model in determining the final number of defects observed in the dataset
3. Predictive ability (PA) shows how early in the testing model is able to predict the final number of defects with maximum 10% error

Data consists of 1327 points. Dependency of cumulative number of registered defects from number of days for original data shown on Figure 1.

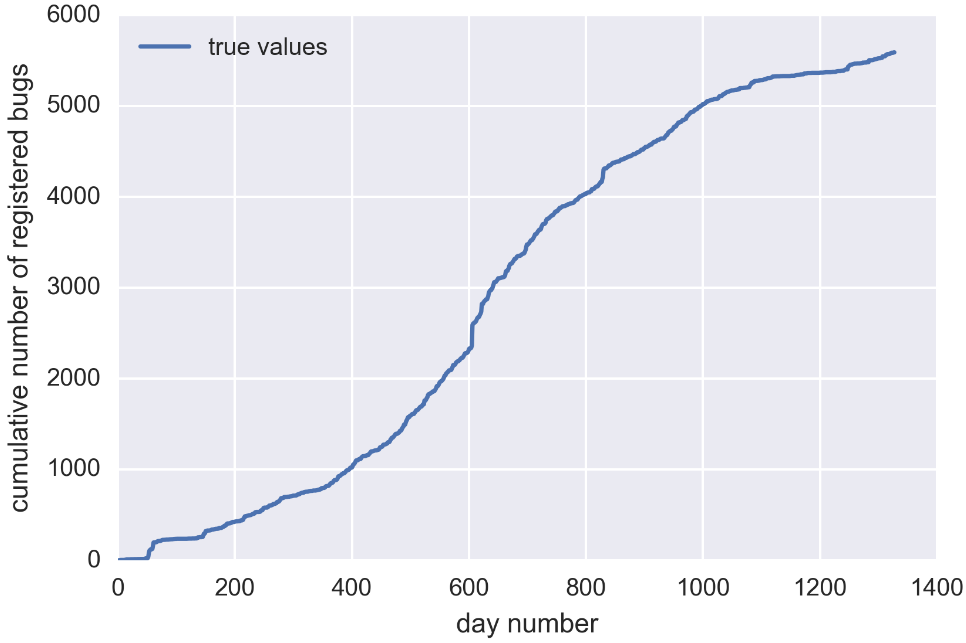

FIGURE 1 TIZEN DEVELOPMENT DATA

By looking on quantiles, 3-sigma test and visualising number of defects registered in each day, found extremely high values (outliers).

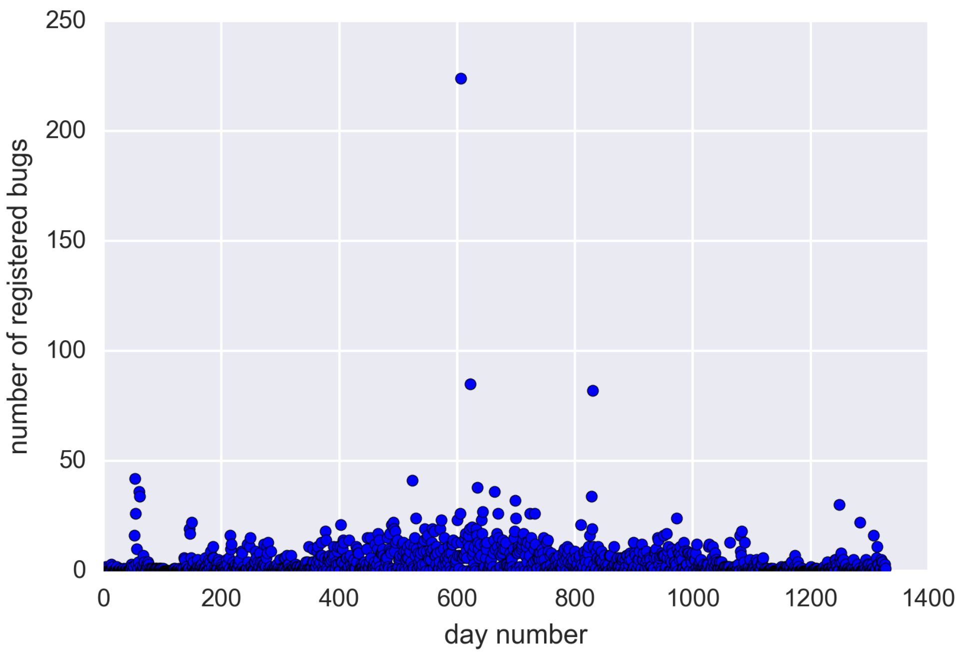

FIGURE 2 NUMBER OF REGISTERED BUGS IN EACH DAY

PART 3. MODEL

WEIBULL DISTRIBUTION

Weibull distribution is a continuous probability distribution. It is named after mathematician Wallodi Weibull, who described it in detail in 1951. Weibull distribution has plenty of applications, including failure analysis and reliability engineering.

The probability density function of a Weibull random variable is:

$$f(x;\lambda,k) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0, \\ 0 & x < 0, \end{cases}$$

where $\boldsymbol{k} > 0$ is shape parameter and $\boldsymbol{\lambda} > 0$ is scale parameter of the distribution.

WEIBULL MORE S-SHAPED MODEL

In our study we use More S-Shaped version of Weibull model:

$$\boldsymbol{a}\,(1 - \,(1 + \boldsymbol{b} \times t^c)e^{-b \times t^c}$$

where

- $a > 0$ is an expected cumulative total number of defects,
- $b > 0$ is a defect detection rate,
- $c > 0$ is a defect detection rate booster.

By fitting model to Tizen dataset we get:

Metric	Value
GoF	20761,527
AcFP	0,4986 %
PA	73,3032 %

TABLE 1 FITTED MODEL MEASUREMENTS

Visualisation of predicted data vs original data shown on Figure 4.

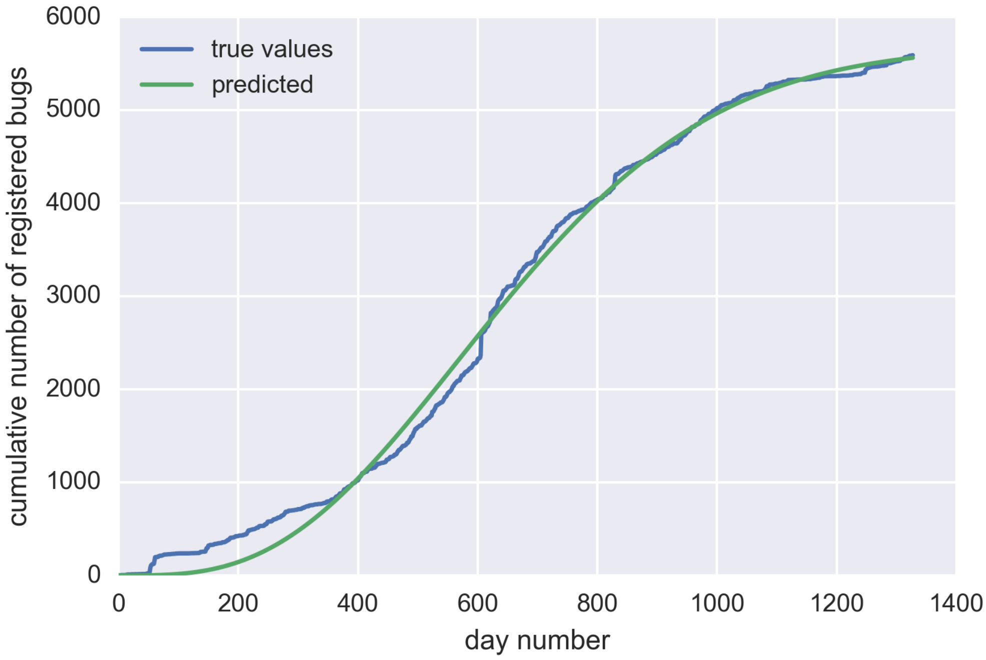

FIGURE 3 ORIGINAL DATA VS PREDICTION

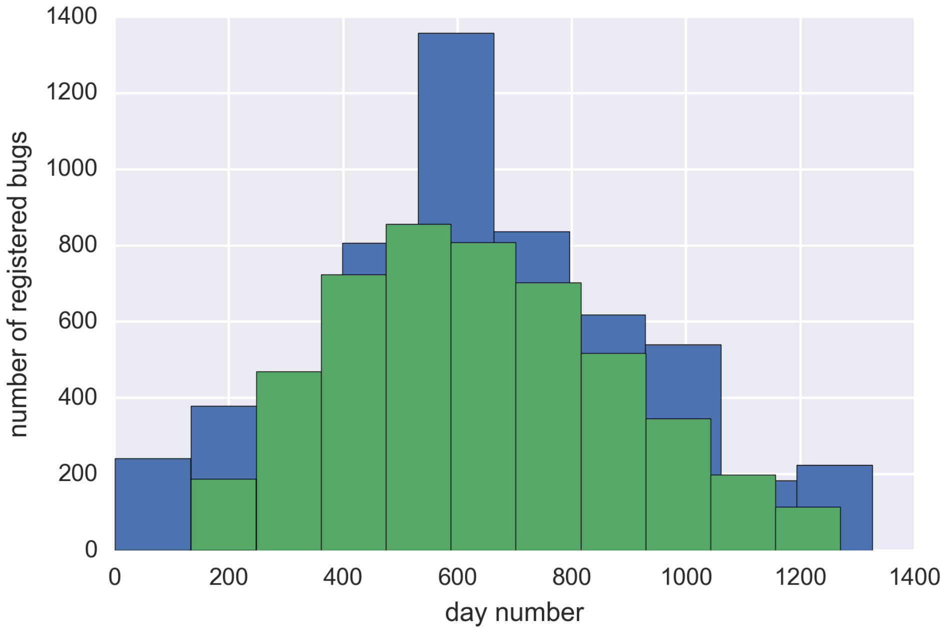

FIGURE 4 DEFECT REGISTRATION EVENTS DISTRIBUTION

Sometimes discovered bugs can be registered in bugtracker not in discovering day. Because of that we need to test our model with changed dataset. New dataset generated as follow:

1. Get number of bugs registered at each day
2. For every discovered bug generate $Shift \sim \mathcal{N}(0, \sigma)$ that shows possible shift for registration. Negative shift shows that bug registered with delay, and possitive shift shows possible delay for registering this bug.
3. Recalculate number of registered bugs for each day

Bugs that will move out interesting time period will be dropped down, because they will not change parameter ***a*** too much with adequate σ. Data that generated with described algorithm tries to preserve one of the model characteristics - expected cumulative total number of defects.

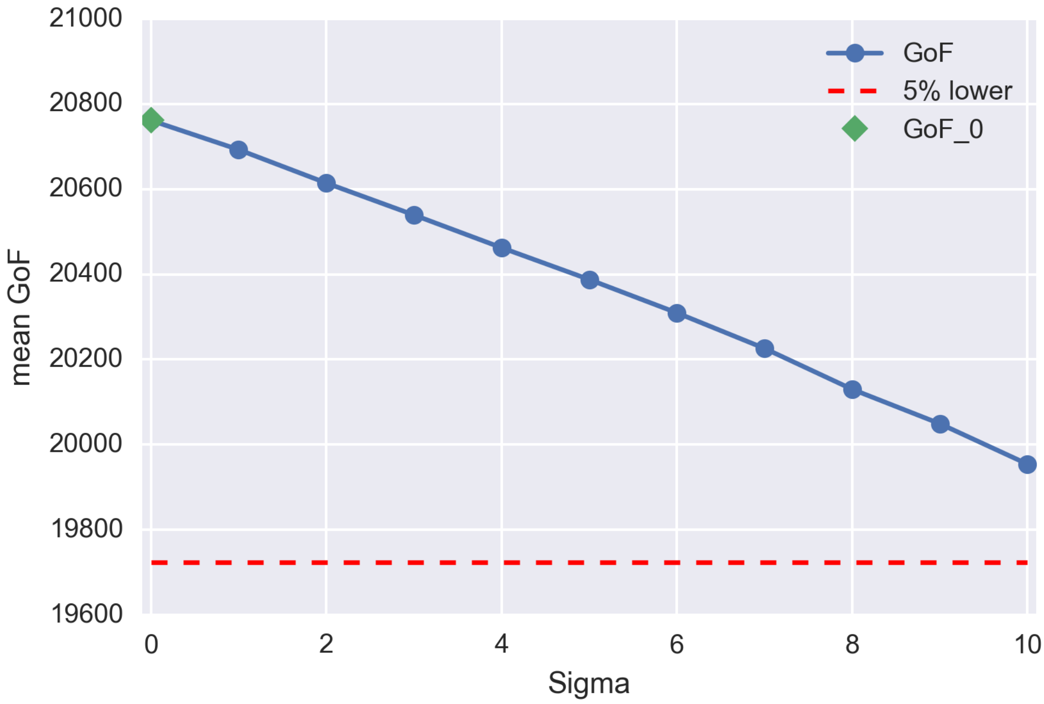

FIGURE 5 MEAN GOF ON NOISE SET

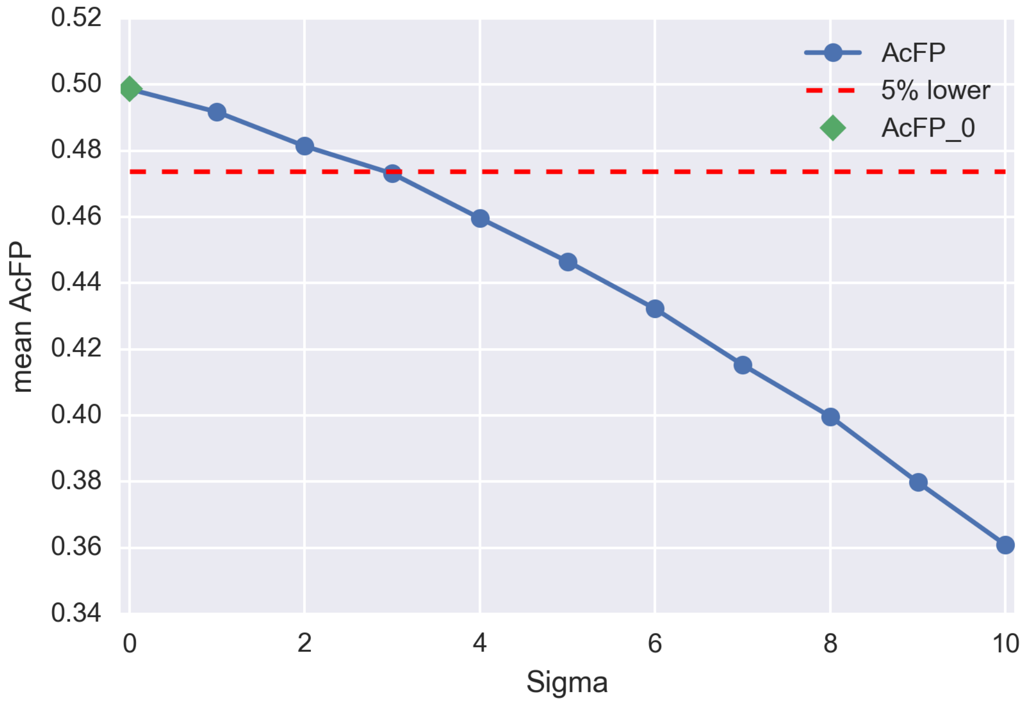

FIGURE 6 MEAN ACFP ON NOISE SET

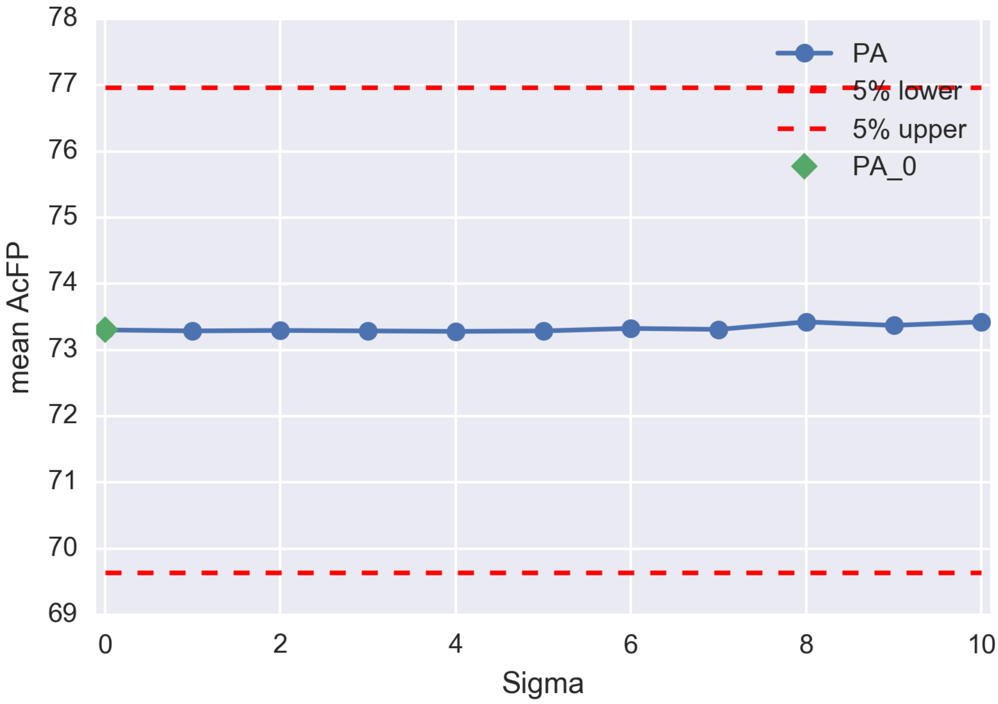

FIGURE 7 MEAN PA ON NOISE SET

RESULTS

Following have been observed on noise test:

1. More S-Shaped Weibull model gives better performance measurement in terms of GoF and AcFP with smoothing data by adding Gaussian noise
2. It is not stable to noise with $\sigma \geq 3$ in terms of AcFP – observed 5% changes
3. In terms of GoF and PA More S-Shaped Weibull model stable for Gaussian noise with σ in [1; 10].

Source code of SRGM package available in GitHub[1]. This package written for project and supports Scikit-learn API.

[1] https://github.com/deusesx/SRMlib