

# Source Code Volatility (SCV) to Spot Dead Code

Yegor Bugayenko

April 15, 2020

## 1 Introduction

Volatility of source code is an experimental metric that shows how big is the difference between actively and rarely changed (possibly dead) code. It is assumed that a big percentage of dead code is an indicator of maintainability problems in the project.

## 2 Details

First, by looking at Git history, it is observed how many times every source code file out of  $N$  was touched during the lifetime of the repository (excluding the files that don't exist in the repository anymore):

$$T = [t_1, t_2, \dots, t_N] \quad (1)$$

Then, the entire interval between  $\check{T}$  (the maximum value) and  $\hat{T}$  (the minimum value) is divided to  $Z$  equivalent groups:

$$G = [g_1, g_2, \dots, g_Z] \quad (2)$$

$$\delta = (\check{T} - \hat{T})/Z \quad (3)$$

$$g_j = \sum_{i=1}^N [j(\delta - 1) < t_i < j\delta] \quad (4)$$

Then, the mean  $\mu$  is calculated as:

$$\mu = \frac{1}{Z} \sum_{j=1}^Z g_j \quad (5)$$

Finally, the variance is calculated as:

$$Var(g) = \frac{1}{Z} \sum_{j=1}^Z |g_j - \mu|^2 \quad (6)$$

The variance  $Var(g)$  is the volatility of the source code. The smaller the volatility the more cohesive is the repository and the smaller the amount of the dead code inside it.