

# Mathematical Models for Regression Testing: Would they help?

(in making Regression Testing Cost Effective)

Ranjit Shringarpure

September 2014



# Premise

- Regression Testing (or Testing *for* Regression?)
  - is an *Essential* activity that
  - may (or may not) be *Effective* and is
  - potentially *Expensive*

I'd like to concentrate on the last two aspects

# Why is Regression Testing Expensive?

- Two examples from very recent projects are good indicators
  - Contractual Obligation
  - Email interaction with Automation Test Manager

# What is 'Expensive'?

Not merely attributed to

*Cost incurred in terms of time and resource(s)*

But also about *Effectiveness*

*(ability to help reduce residual bugs / improve the bug detection rate)*

# Cost-Effective

- We all want More; for Less! We want Regression Testing to be cost-effective.
- We get more by
  - Automating stuff (checks)
  - Not Automating stuff (run relevant, old tests)
  - New tests (design new tests – Things could be worse video by Michael Bolton)
  - Use Heuristics (e.g. Karen Johnson's RCRCRC heuristics)
  - More available online

But we want it all for Less

# Topic of interest: Regression Testing

- Academics, researchers of Software Engineering discipline
- Empirical studies, research papers, algorithms, discussions and debates since late 70s!
- Two main areas of research of cost effectiveness of Regression Testing
  - Choosing the right (number of) tests (addresses Cost aspect)
  - Improving fault detection rates (address Effectiveness aspect)

# Regression Test Selection

- Most commonly employed technique
  - testers perform tests
  - They identify some tests as regression candidates
  - They (pass it onto members who) automate them to build a repository of checks.
  - At the next release, all items within this 'repository' is run as 'regression test'
  - We'll include genuine 'tests' that will also be run
- 'Retest-all' technique. (Caveat – all non-obsolete / non-redundant tests)
- There are two other major techniques

# Regression Test Selection

- Regression Test Selection
  - Only tests that are necessary to test the modified/impacted part of the software under test are chosen
  - A flavour of this technique also includes Test case reduction and test case minimisation too.
- Test Case Prioritization
  - Tests are ordered based on a set criteria (e.g. biggest coverage, quick wins etc) and reiterate until set number of tests are completed.
  - Helps early feedback and allows further prioritization



# Costs

- attributed to time and resource (performing various testing related activity). Mainly includes costs associated with
  - *Analysis*: of the techniques to be employed; either by themselves or even being a mentor to other testers
  - *Obsolete/Redundant tests*: cost of identification and possible fixing of obsolete and non-obsolete tests;
  - *Test Setup and Execution*: Costs associated with procuring hardware, software, configuration and setting up environment,
  - *Test execution and interpretation* – manual or automated runs, interpretation of result at the end of test run etc
  - *Miscellaneous*: includes test management, reporting, databases and many more.

# Models

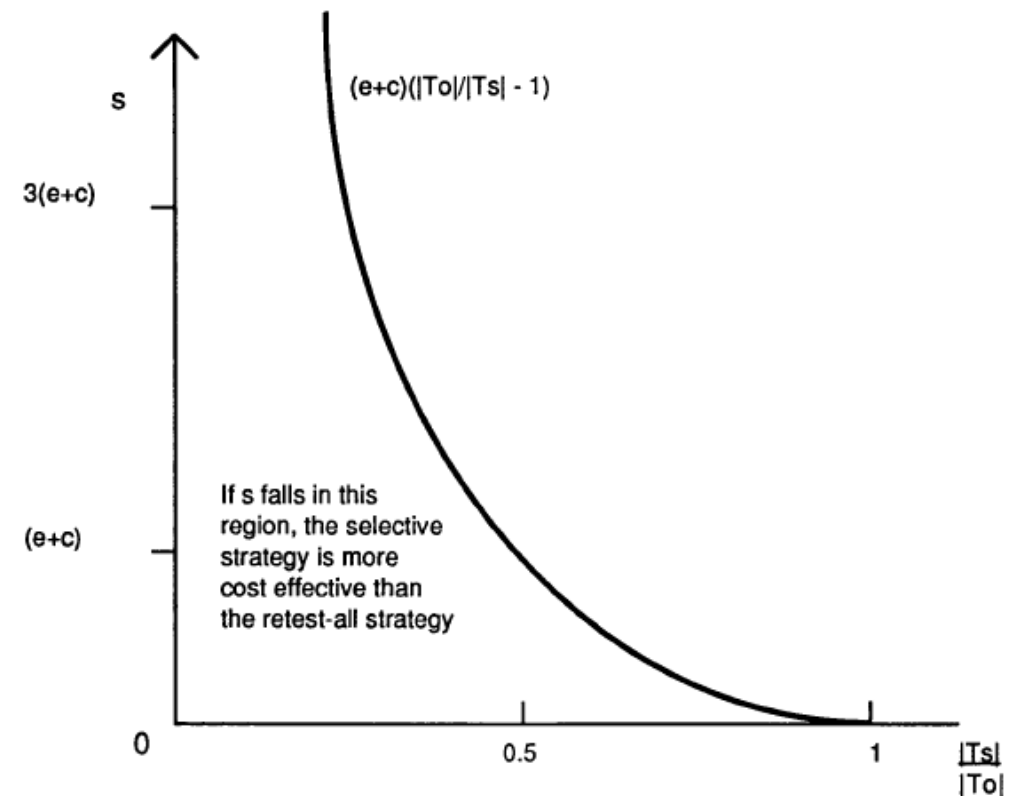
- Mathematicians and Statisticians have used models for generations to describe systems
- Examples
  - Growth based on food intake
  - Education vs Experience vs wages relationships
  - etc
- Modelling techniques have been used in fields of natural sciences, engineering, social sciences and even Regression Testing

# Mathematical Models for testing

- Hareton K. N. Leung & Lee White proposed a cost model to compare cost advantage of selective regression testing against retest-all strategy
- Y Chen, D Rosenblum and K Vo at AT & T Bell laboratories proposed a 'TestTube' system which claimed reduction of number test cases required by 50% or more
- Gregg Rothermel, Alexey Malishevsky and Sebastian Elbaum improved upon previously published work and demonstrated prioritization techniques significantly improved the fault detection rate

# Examples of models for Regression Testing

- Hareton K. N. Leung & Lee White's Cost Model to compare Regression Test Strategies
- Depends on
  - Selection cost ( $s$ )
  - Execution cost ( $e$ )
  - Checking cost ( $c$ )
  - Ratio of selected previous test ( $T_s$ ) to total previous test ( $T_o$ )



# Examples of models for Regression Testing

- Gregg Rothermel & Hyunsook Do's Cost-Benefit Model
- Includes
  - Context factors (by considering costs for many essential testing related activities) and
  - lifetime factors (time constraints, incremental resource availability)

$$Cost = PS * \sum_{i=2}^n (CS(i) + CO_i(i) + CO_r(i) + b(i) * CV_d(i) + c(i) * CF(i)) \quad (4)$$

$$Benefit = REV * \sum_{i=2}^n (ED(i) - (CS(i) + CO_i(i) + CO_r(i) + a_{in}(i-1) * CA_{in}(i-1) + a_{tr}(i-1) * CA_{tr}(i-1) + CR(i) + b(i) * (CE(i) + CV_i(i) + CV_d(i)) + CD(i))) \quad (5)$$

$$Benefit_A - Cost_A \quad (6)$$

$$(Benefit_A - Cost_A) - (Benefit_B - Cost_B) \quad (7)$$

# ...and many more

- Including Hyunsook Do & Panduka Nagahawatte's study of Regression Testing Prioritization technique has indeed been effective in reducing the occurrence of residual defects
- And Gregg Rothermel & Mary Jean Harrold's algorithms for selecting regression tests for Object Oriented Software And many more
- But nearly all studies, research and studies point to avenues for future work including
  - More controlled experiments
  - Development of new regression techniques
  - Development of new algorithms

# Finally...

- There's tons of work done by researchers out there that can help testers make regression testing cost effective
- Just one of the research paper quote around 150 publications between 1977 and 2009
- Nothing available in workplaces that encourage testers to tap into this minefield of information (yes, following Standards is one of the reason for it!)

# References

Some of the references are:

- <http://cse.unl.edu/~grother/>
- [http://web.cs.du.edu/~sazghand/background\\_chap\\_papers/A%20cost%20model%20to%20compare%20regression%20test%20strategies.pdf](http://web.cs.du.edu/~sazghand/background_chap_papers/A%20cost%20model%20to%20compare%20regression%20test%20strategies.pdf)
- <http://cse.unl.edu/~grother/papers/icsm02b.pdf>
- <http://cs.txstate.edu/~rp31/papersSQ/ChenRosenblumVo.pdf>
- <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1031&context=csetechreports>
- And many, many more



# About me

- Independent Tester
- @ranjitsh | ranjitsh@gmail.com
- [www.iveritas.co.uk](http://www.iveritas.co.uk) | [uk.linkedin.com/in/ranjitsh/](https://uk.linkedin.com/in/ranjitsh/)