Dynamic Subclass Instantiation Distribution

Amy Wheeler
David Binkley
Loyola University Maryland

1 floor
3 full time faculty
0 PhD students
What

Dynamic subclass instantiation tracking
Example Distributions

Test Suite Versus "Typical use"

Subclasses from most to least frequent
How: JBoss

Part 1: Join Point Description

```xml
<bind pointcut = "construction(*->new(..))"
    and !construction(Tracer*->new(..))" >
    <before aspect="Tracer"
        name="ConstructorInterceptor" />
</bind>
```
Part 2: Aspect

public Object Tracer::ConstructorInterceptor (ConstructionInvocation ci)
{
    Class c = ci.getConstructor().getDeclaringClass();
    Chain invocation = new Chain();
    while (c != null)
    {
        invocation.addClass(c);
        c = c.getSuperclass();
    }
    return ci.invokeNext();
}
## Who – Subjects Programs

<table>
<thead>
<tr>
<th>Program</th>
<th>LoC</th>
<th>Classes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GanttProject 2.0.9</td>
<td>69K</td>
<td>564</td>
<td>Project Scheduling</td>
</tr>
<tr>
<td>jasmin 2.3</td>
<td>40K</td>
<td>216</td>
<td>Java Assembler</td>
</tr>
<tr>
<td>jess</td>
<td>no src</td>
<td>460</td>
<td>Sandia Rule Engine</td>
</tr>
<tr>
<td>jmeter 2.3.4</td>
<td>147K</td>
<td>792</td>
<td>Testing Tool</td>
</tr>
<tr>
<td>jolden</td>
<td>6215</td>
<td>20</td>
<td>Olden Bench Mark</td>
</tr>
<tr>
<td>jtopas</td>
<td>24K</td>
<td>65</td>
<td>Java Tokenizer Lib</td>
</tr>
<tr>
<td>nanoxml 1.2</td>
<td>95K</td>
<td>611</td>
<td>Java Parser</td>
</tr>
<tr>
<td>siena 0.9</td>
<td>98K</td>
<td>34</td>
<td>Event Services</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>479K</td>
<td>2762</td>
<td></td>
</tr>
</tbody>
</table>
Research Questions

1. More than 1?
   Do classes with instantiated subclasses have more than one instantiated subclass?

2. Uniform or Skewed?
   If multiple subclasses are instantiated, is the distribution of subclass instantiation uniform or skewed?

3. Dominated? Correlated?
   If the distribution is skewed, then does a
   a) single or do a few subclasses dominate?
   b) higher instantiation count correlate to greater skew?
Hypothesis 1

H0: only one of a class's subclass is instantiated

Ha: more than one of a class's subclasses are instantiated
Dynamic Instantiation Count

Ha: more than one of a class's subclasses are instantiated

The mean number of subclasses instantiations is different from 1.0

(p = 0.00044)
Hypothesis 2

H0: the distribution of subclass instantiations is uniform

Ha: the distribution of subclass instantiations is skewed
Aggregate Distribution

Ha: the distribution of subclass instantiations is skewed

There is significant skew $X^2$ test $p < 0.0001$

[X^2 H0 – all outcomes are equally likely]
Data per Class

By class 53% significant skew

4/5 of these $X^2$ $p < 0.0001$

By count 99.998% significant
Hypothesis 3

Hypothesis 3a
Ha: one or a small number of subclasses dominate.

Hypothesis 3b
Ha: higher instantiation count correlates to greater skew.
Next?

Questions?

• Investigate applications

• Forecast what is typical / atypical
The Chicken came first!

• One is
  – First validate assumptions

• One is
  – First explore uses
Hypothesis 3a