# Towards a Bell-Curve Calculus and its Application to e-Science

*Author Lin Yang*

*Supervisors A. Bundy; D. Berry; S. Huczynska; C. Hughes*

## Background

Quality of Service (QoS) properties, such as accuracy, reliability and runtime, will have a range of possible values. Bell curves (normal distributions) are one way of modelling these ranges. They give more information than using error bounds, as error bounds only give an worst-case analysis, whereas bell curves give an indication of the probability of each value occurring.

## Introduction

We define a bell-curve calculus for reasoning about QoS properties. The advantages of using bell-curve calculus are:

(1) Since there are only two parameters in the expression of a bell curve, it is easy to store and propagate;
(2) Bell curves can deal with complex workflows efficiently;
(3) According to Central Limit Theorem and experimental result from DIGS*, bell curves commonly occur in the real world.

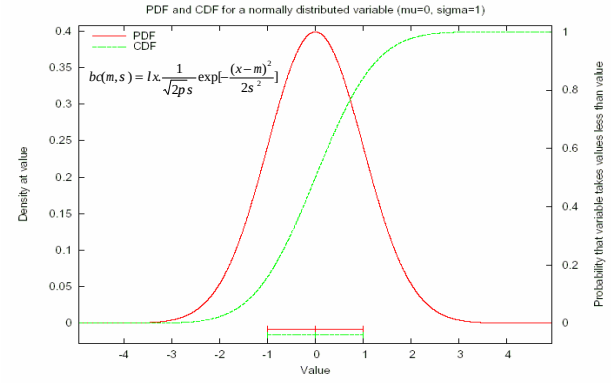
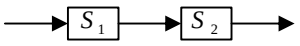



Figure 1. A Standard Bell Curve

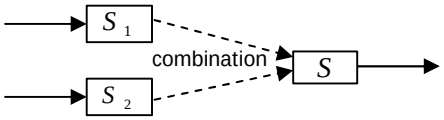
## Definition

Three QoS properties are investigated in our project – accuracy, reliability and runtime, with four ways of combining Grid services:

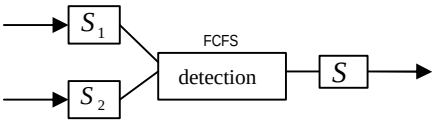
(1) Sequential



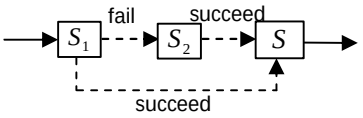

(2) Parallel_All




(3) Parallel_First




(4) Conditional




So there are 3x4=12 simple situations:

| | Seq | Para_All | Para_Fir | Cond |
|---|---|---|---|---|
| run time | sum | max | min | cond1 |
| accuracy | mult | combine1 | varies? | cond2 |
| reliability | mult | combine2 | varies? | cond3 |

Table 1. Operations Of 12 Simple Situations

## Current Work

1. Define $bc(m_m, s_m)$ as the bell-curve approximation of the combined curve $F(bc(m_r, s_r), bc(m_g, s_g))$. For each 12 functions defined above, induce the 24 function for $m_m$ and $s_m$ in terms of $m_r$, $m_g$, $s_r$ and $s_g$.
   e.g. for runtime in sequential structure:

$$m_m = m_r + m_g \qquad s_m = \sqrt{s_r^2 + s_g^2}$$

2. Test these functions in Agrajag** and determine the ranges of acceptable error.

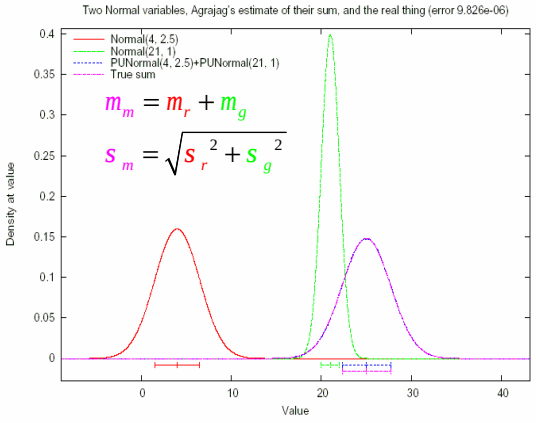



Figure 2. Sum Of Two Runtime Bell Curves

## Future Work

1. Make combinations of different QoS properties;
2. Find better representations of probabilistic behaviour of QoS properties; e.g. log-normal distribution
3. Define more QoS properties to apply the calculus to.

* *Dependability Infrastructure for Grid Services* is an EPSRC project to deal with fault-tolerance and consider wide Quality of Service issues in Service-Oriented Architectures.
** *Agrajag* is a framework developed by DIGS to express some basic distribution functions and define operations on them using Perl and C languages.

School of informatics