NMath Case Studies and Use Cases

Technical Report No. 5

CenterSpace™
Software

Introduction

The **NMath** .NET math and statistics libraries from CenterSpace Software provide building blocks for financial, engineering, and scientific applications on the Microsoft .NET platform. The **NMath** libraries are general-purpose numerical toolkits with wide application in multiple industry sectors. This document describes some example case studies and common use cases of **NMath** features, but is by no means exhaustive.

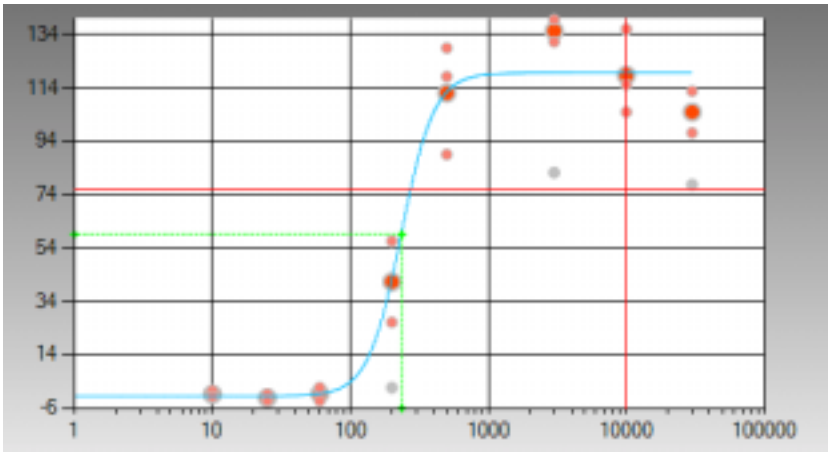
Case Studies

The following case studies illustrate a few of the ways that **NMath** mathematical and statistical functionality can be used to solve real-world problems.

Curve Fitting

SIGA (www.siga.com) is a world leader in designing and developing novel countermeasures to prevent and treat serious infectious diseases, with an emphasis on biological warfare defense. Using **NMath's** curve fitting functionality, SIGA scientists modeled a dose-response system with a logistic curve.

Figure 1 – Dose-Response Curve

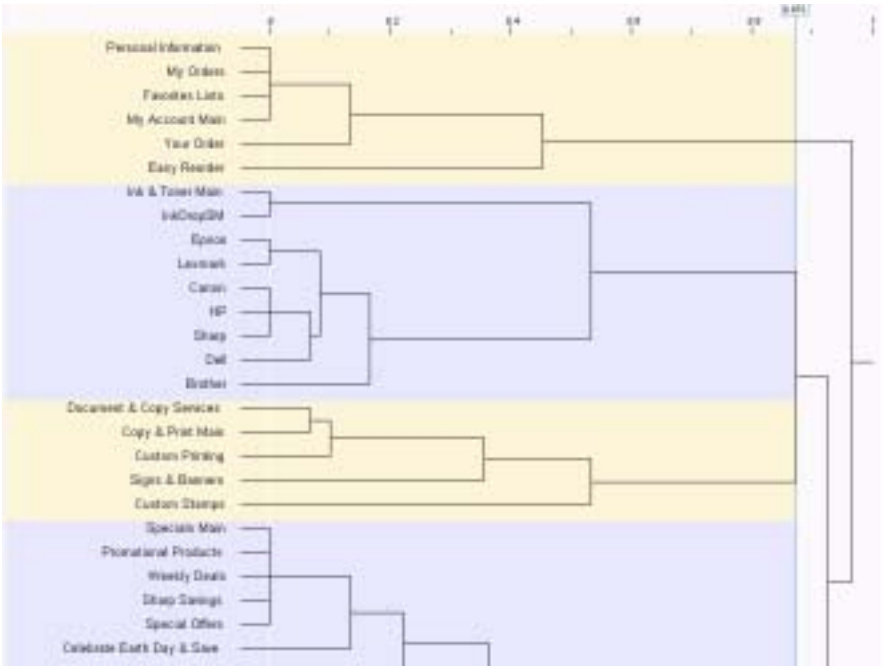


The *x*-axis is the dose, and the *y*-axis is the drug's response to the pathogen. The logistic model is a fundamental non-linear model for many systems, and is widely used in the life sciences, medicine, and environmental toxicology.

Hierarchical Clustering

UserZoom (www.userzoom.com) is an international software company specializing in web customer experience and usability testing. UserZoom software includes a tool to run online card sorting exercises with real users. Card Sorting is a technique that information architects and user experience practitioners use to explore how "real people" group items and content.

Figure 2 – Dendrogram



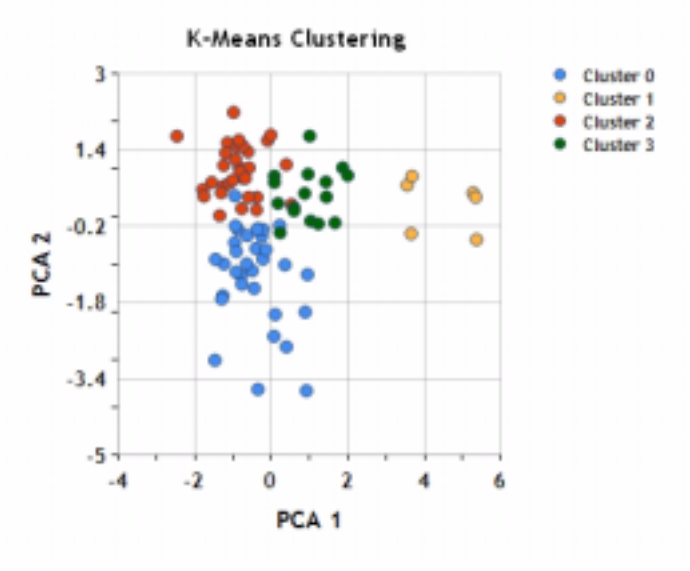
UserZoom developers used **NMath's** hierarchical clustering functionality to analyze card sorting results, and to generate dendrograms to help researchers understand participant's grouping criteria.

Principal Component Analysis (PCA)

A financial company uses principal component analysis (PCA) to reduce the dimensionality of a data set. **NMath's** PCA functionality finds a smaller set of synthetic variables that capture the maximum variance in an original data set. The first principal component accounts for as much of the variability in the data as possible, and each succeeding orthogonal component accounts for as much of the remaining variability as possible.

The original data was 89 stocks rated on 12 separate measures. Using PCA, they plotted the first two principal components against each other, reducing the dimensionality from 12 dimensions to 2. In this case, the first two principal components account for over 50% of the variance.

Figure 3 – The First Two Principal Components

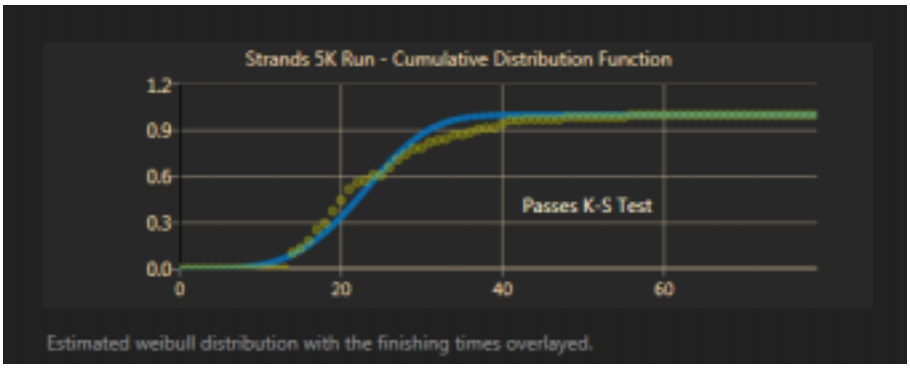


The researchers further used **NMath's** *k*-means clustering algorithm to partition the stocks into four natural groupings.

Hypothesis Testing

Strands (www.strands.com) creates a workout analyzer and digital training tools for athletes. They host an annual 5K run in Corvallis, OR. The central limit theorem tells us to expect the running times for a foot race (of enough participants) to be normally distributed. But what happens when Strands offers a $10,000 prize as an inducement to participants? Strands hypothesized that the finishing times would not be normally distributed, because the big prize would attract many fast runners, and many average runners would enjoy the race from the sidelines. This would group the finishing times around the winner (many close finishers), so the finishing times would no longer normally distributed.

Figure 4 – Estimated Weibull Distribution



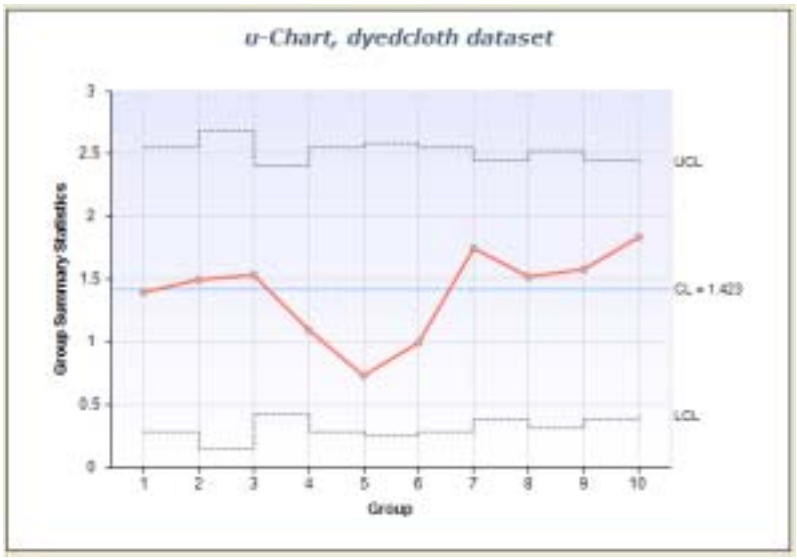
NMath's Kolmogorov-Smirnov test is a hypothesis test to determine whether a set of data points are drawn from a reference distribution. Using this tool, Strands found that the running times did not in fact follow a normal distribution, but rather a Weibull distribution.

Statistical Process Control

Statistical quality control charts, or Shewart charts, are used across nearly all sectors of industry to maintain and improve product quality. These process control charts are independent of any engineering decision-making about the particular process at hand, but are instead based on the statistical nature of the process itself. For example, a u-chart shows non-conformances per unit within some subgroup.

A clothing company uses quality control charts to detect when their dyeing exceeds its historic process variation and needs analysis and/or intervention to remedy the out-of-control process (known as special cause variation). Using **NMath's** descriptive statistics and probability distributions, Acme creates a u-chart for their process control engineers.

Figure 5 – u-Chart



Data Smoothing

An engineering company needed to filter experimental data without losing important peaks. **NMath's** Savitzky-Golay smoothing functionality effectively removes local signal noise while preserving the shape of the signal. Commonly, it's used as a preprocessing step with experimental data, especially spectrometry data, because of it's effectiveness at removing random variation while minimally degrading the signal's information content.

The following three images show some real experimental data and a comparison of two filtering algorithms. The first image shows the raw data, the second image shows the effect of a standard averaging filter, and the last image demonstrates a Savitzky-Golay smoothing filter.

Figure 6 – Raw Data

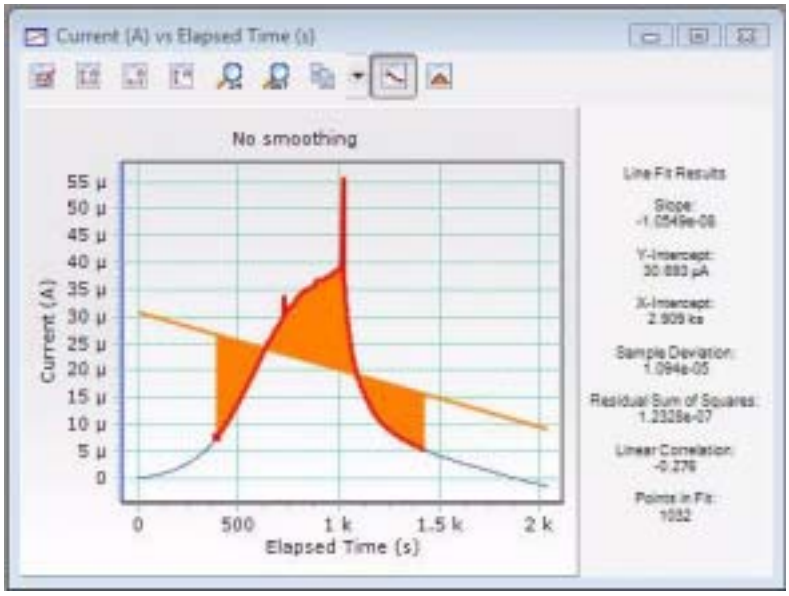


Figure 7 – Averaging Filter

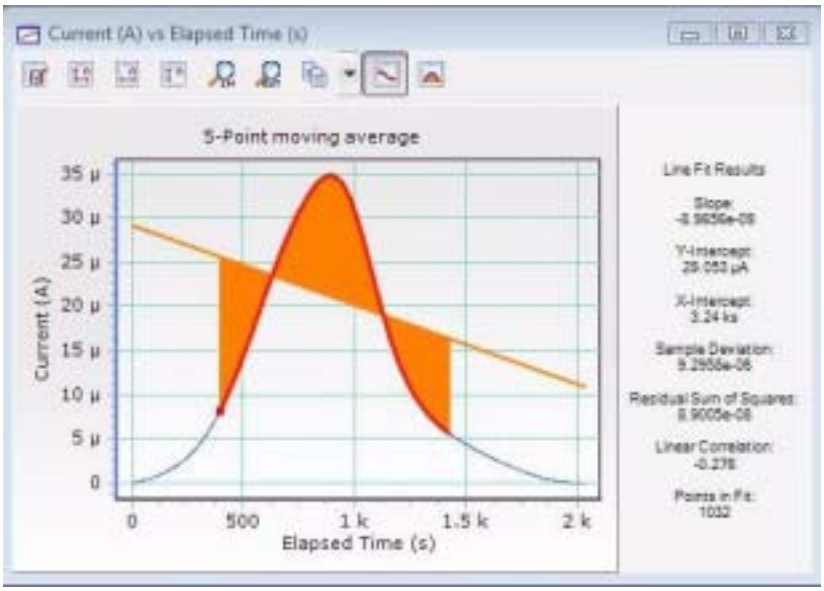
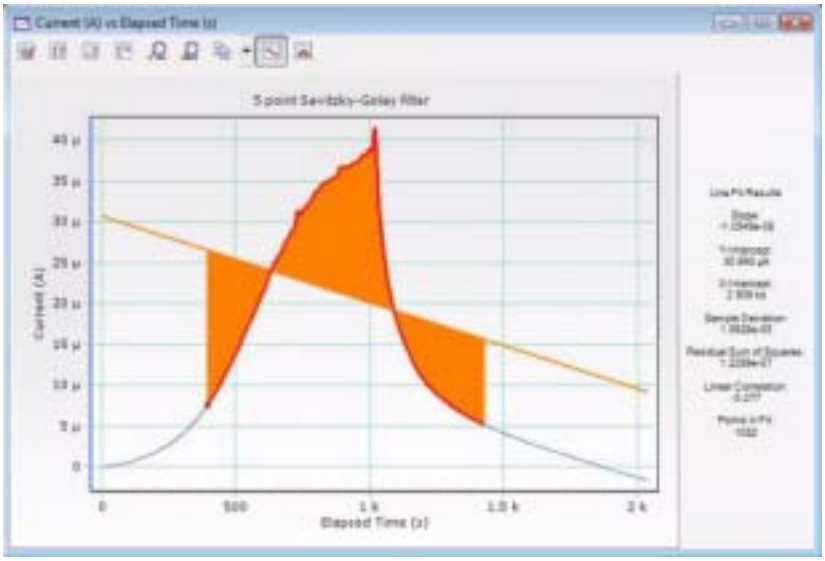


Figure 8 – Savitsky-Golay Filter



Use Cases

The following tables show some common applications of **NMath** features.

NMath

Table 1 – NMath Use Cases

Feature	Target Industries	Common Applications
Random number generators	Biosciences Finance Social Sciences Software/Technology	Statistical sampling Randomized designs Cryptography Computer simulations (Monte Carlo simulations) Financial modeling Portfolio optimization
Linear algebra (matrix/vector types, matrix factorization and decompositions, eigenvalues)	All	Optimization Linear system modeling

Table 1 – NMath Use Cases

Feature	Target Industries	Common Applications
Fast Fourier Transforms (FFTs)	Biosciences Engineering Finance Oil/Gas Social Sciences Software/Technology Telecommunications	Signal analysis Vibration analysis General sound analysis Communication signal analysis (radio, wireless, acoustic) Seismic signal analysis Financial trends analysis Voice signal analysis Noise detection, analysis and removal Signal and electronic intelligence Medical signal analysis (ECG, EEG) Statistical data analysis
Signal processing (convolution, correlation, and filtering)	Biosciences Engineering Finance Oil/Gas Social Sciences Software/Technology Telecommunications	Peak finding Smoothing Time series analysis Financial analysis Exponential weighted moving average Image, audio, and video processing
Calculus (numerical integration, differentiation, minimization, interpolation, and root finding)	All	Structural engineering Physics modeling
Least squares minimization (linear, nonlinear, weighted, and partial)	All	Prediction and modeling Forecasting Inferring relations between variables
Curve and surface fitting	All	Dose-response modeling Experimental science Bond models Yield curves

Table 1 – NMath Use Cases

Feature	Target Industries	Common Applications
Linear Programming (LP), Nonlinear Programming (NLP), and Quadratic Programming (QP)	Transportation Energy Telecommunications Manufacturing/Retail	Planning Routing Scheduling Industrial design
Simulated annealing	Manufacturing/Retail	Optimization Structural design Transportation and distribution design Industrial design
Differential equations	Engineering Finance Health Sciences	Dynamic modeling Population modeling Disease transmission models Drug modeling Derivatives

NMath Stats

Table 2 – NMath Stats Use Cases

Feature	Target Industries	Example Uses
Description statistics	All	Quantitative data analysis Summarization of experimental, clinical, or demographic data
Probability distributions	All	Statistical hypothesis testing Modeling of probabilistic processes or systems
Linear regression	Biosciences Finance Social Sciences Manufacturing/Retail	Trend analysis Risk analysis Predication and forecasting Quality control
Hypothesis tests	Biosciences Social Sciences	Experimental analysis Statistical inference from data

Table 2 – NMath Stats Use Cases

Feature	Target Industries	Example Uses
Analysis of variance (ANOVA, RANOVA)	Biosciences Social Sciences	Experimental analysis Statistical inference from data
Principal Component Analysis (PCA)	All	Dimension reduction and model simplification Data visualization Exploratory data analysis Latent factor discovery
Clustering (k-means, hierarchical, nonnegative matrix factorization)	All	Data visualization Data mining and classification Pattern recognition Information retrieval and document classification Recommender systems Bioinformatics Market research Image segmentation
Partial Least Squares (PLS)	All	Predictive modeling Chemometrics Sensory evaluation Analysis of functional brain imagery

NMath Case Studies and Use Cases

The correct bibliographic reference for this document is:
NMath Case Studies and Use Cases, Technical Report No. 5, CenterSpace Software, Corvallis, OR.

Printed in the United States.
Printing Date: August, 2011

CenterSpace Software

Address:	230 SW 3rd St., Suite 311, Corvallis, OR 97333 USA
Phone:	(541) 896-1301
Web:	http://www.centerspace.net
Sales:	sales@centerspace.net
Technical Support:	support@centerspace.net