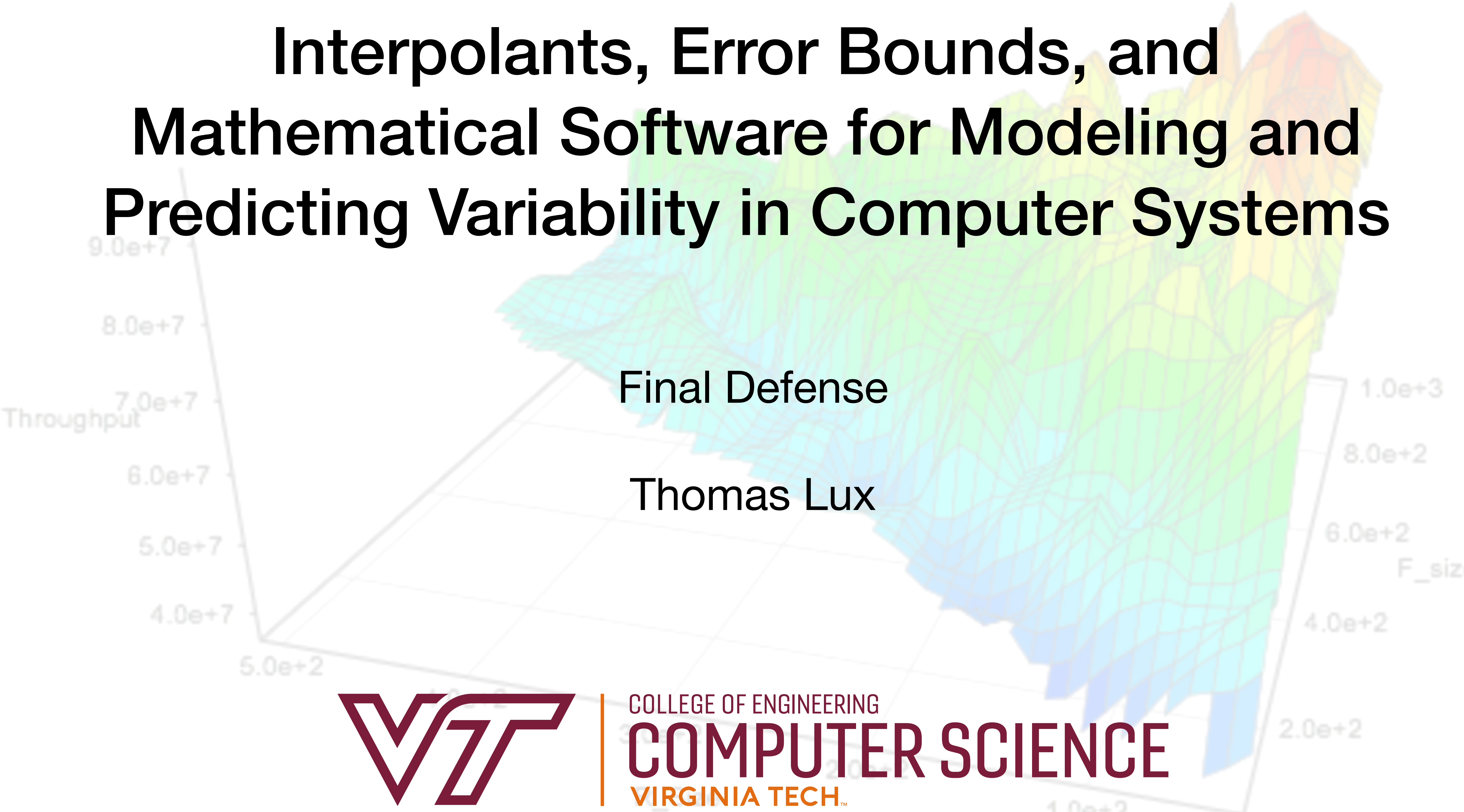
Interpolants, Error Bounds, and Mathematical Software for Modeling and Predicting Variability in Computer Systems

Final Defense

Thomas Lux

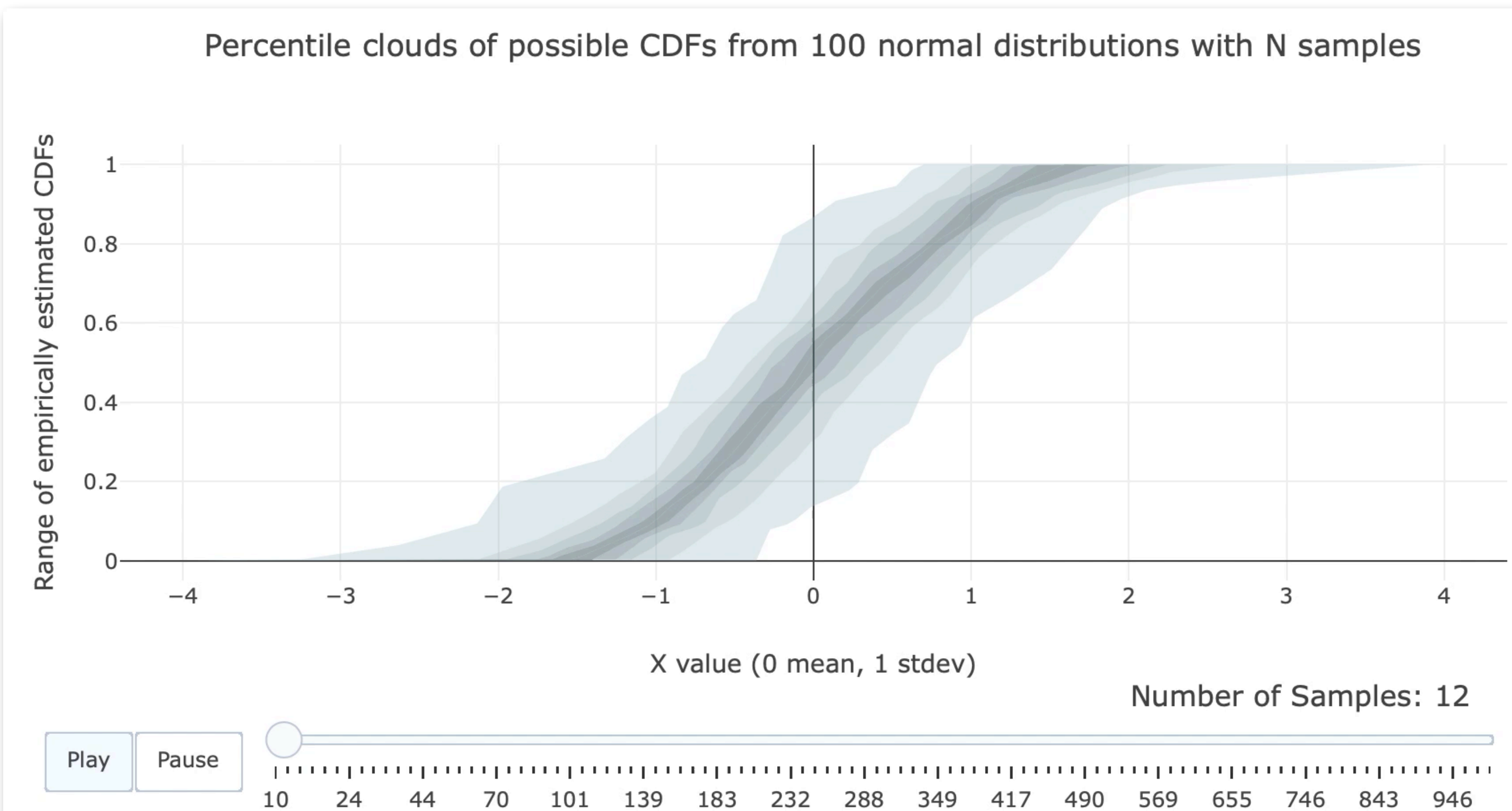


COLLEGE OF ENGINEERING
COMPUTER SCIENCE
VIRGINIA TECH

Chapters

What is “Variability”?

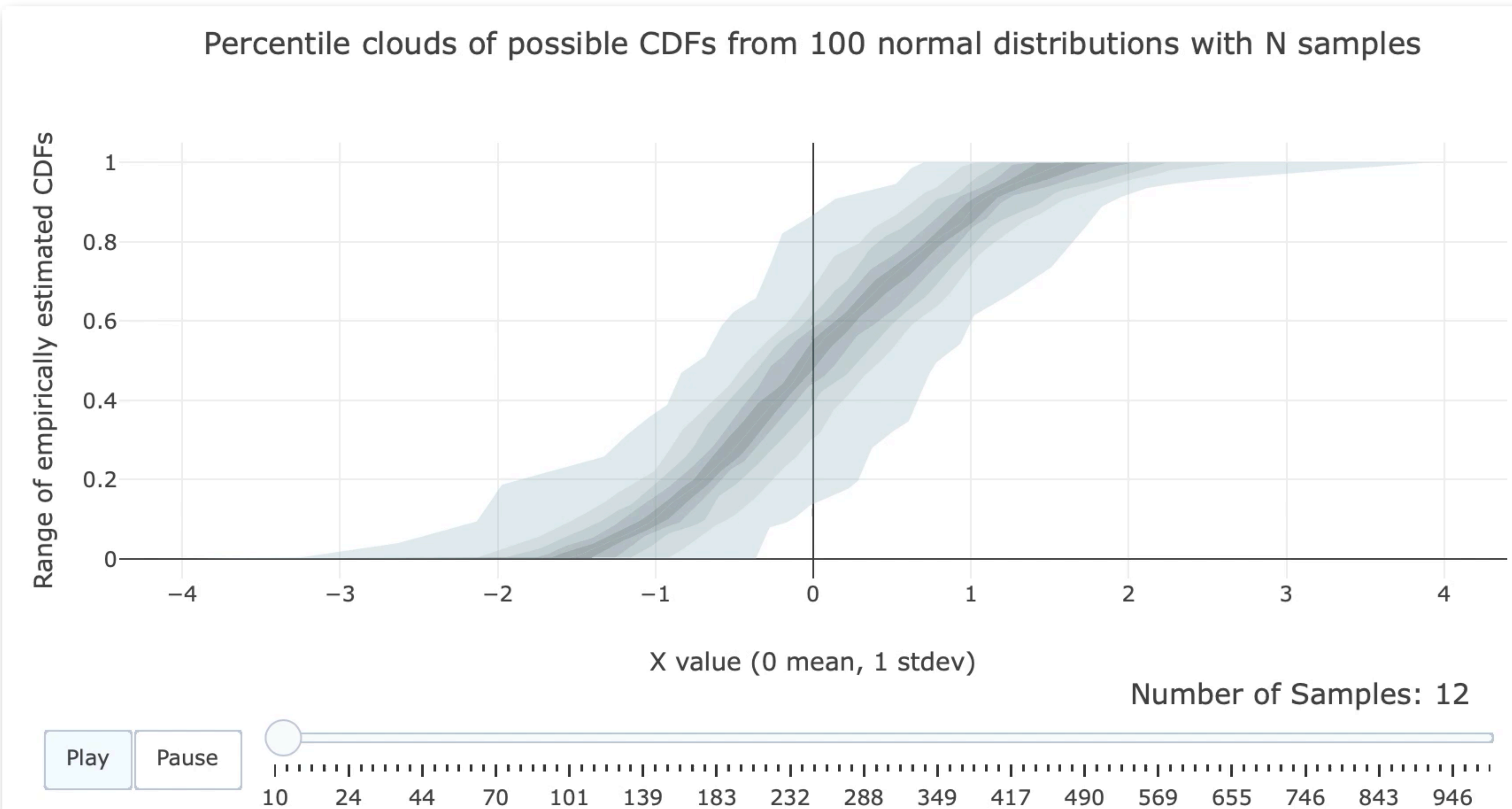
A *random variable* uniquely defined by its Cumulative Distribution Function (CDF) / Probability Density Function (PDF)



COMPUTER SCIENCE
VIRGINIA TECH.

What is “Variability”?

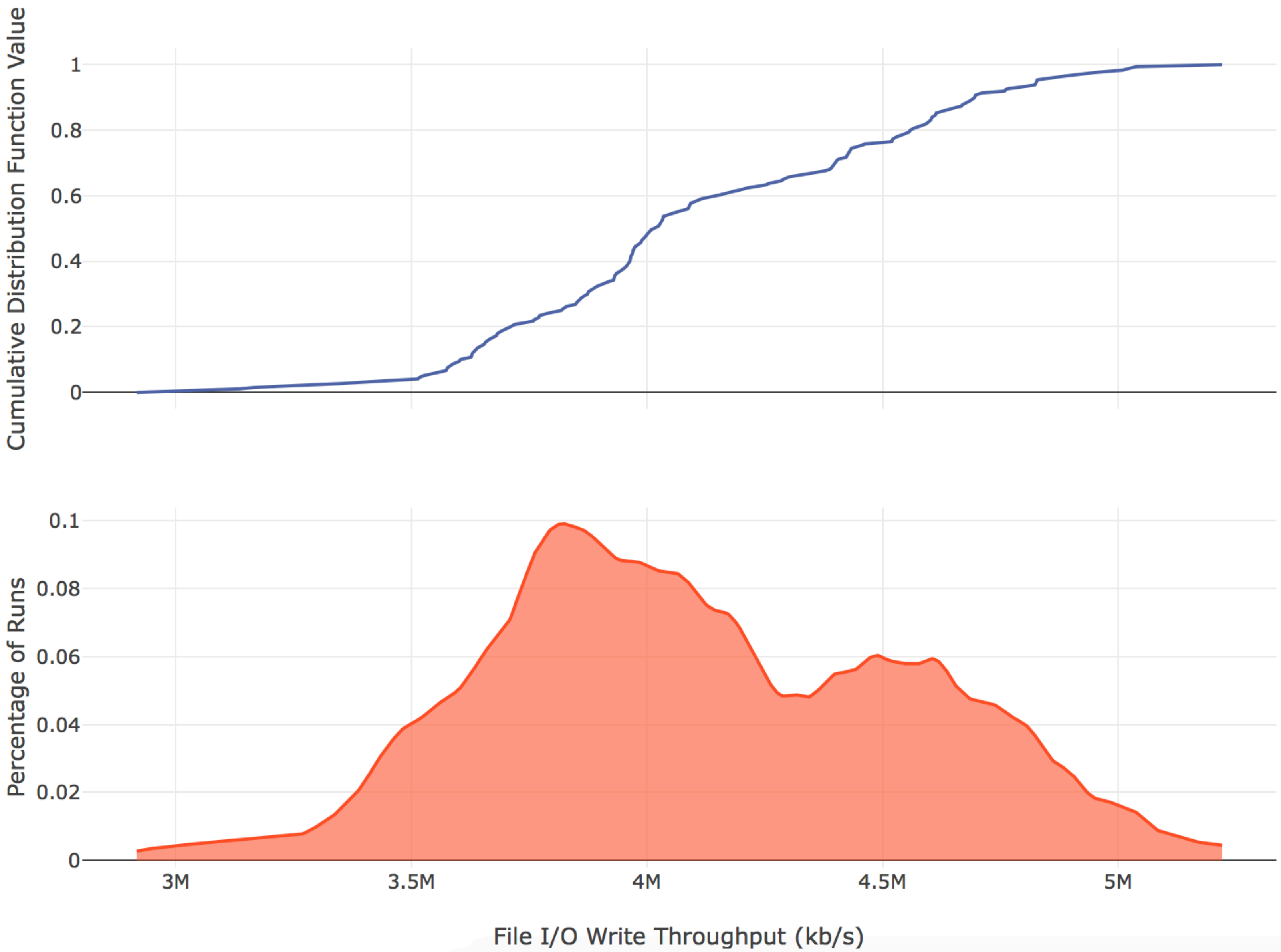
A *random variable* uniquely defined by its Cumulative Distribution Function (CDF) / Probability Density Function (PDF)



COMPUTER SCIENCE
VIRGINIA TECH.

Variability in Systems

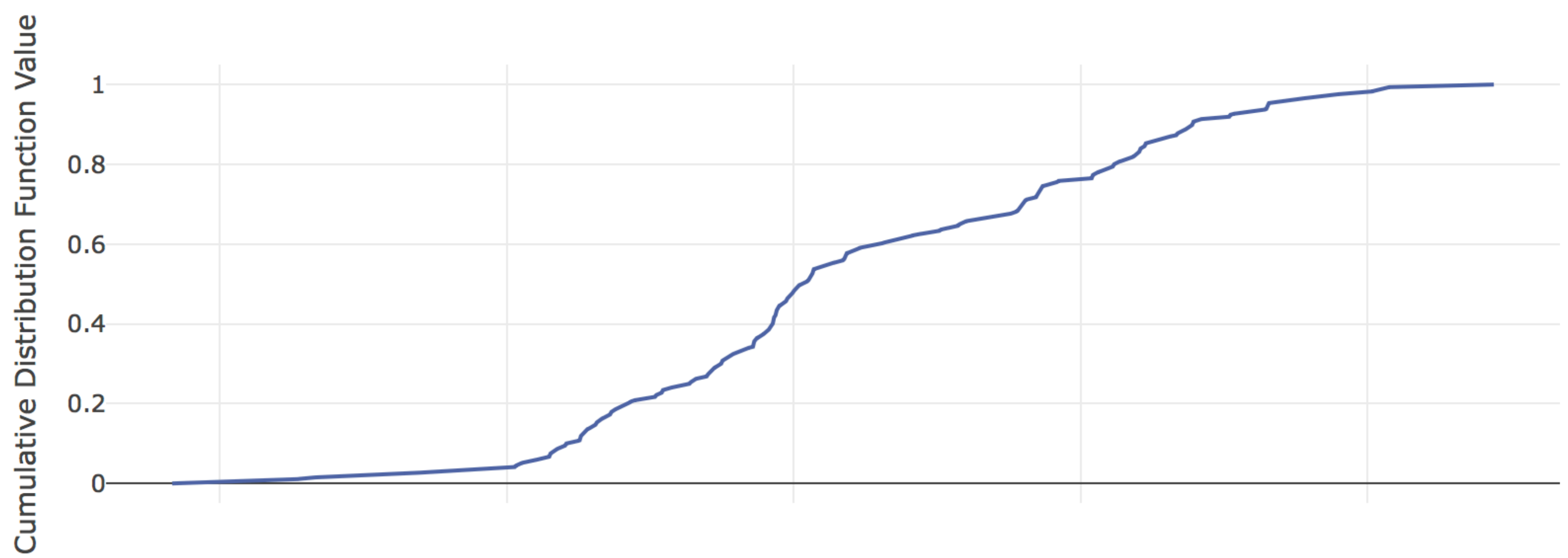
When the same computer is used to execute the same program repeatedly, most measurable performance characteristics will vary.



COMPUTER SCIENCE
VIRGINIA TECH.

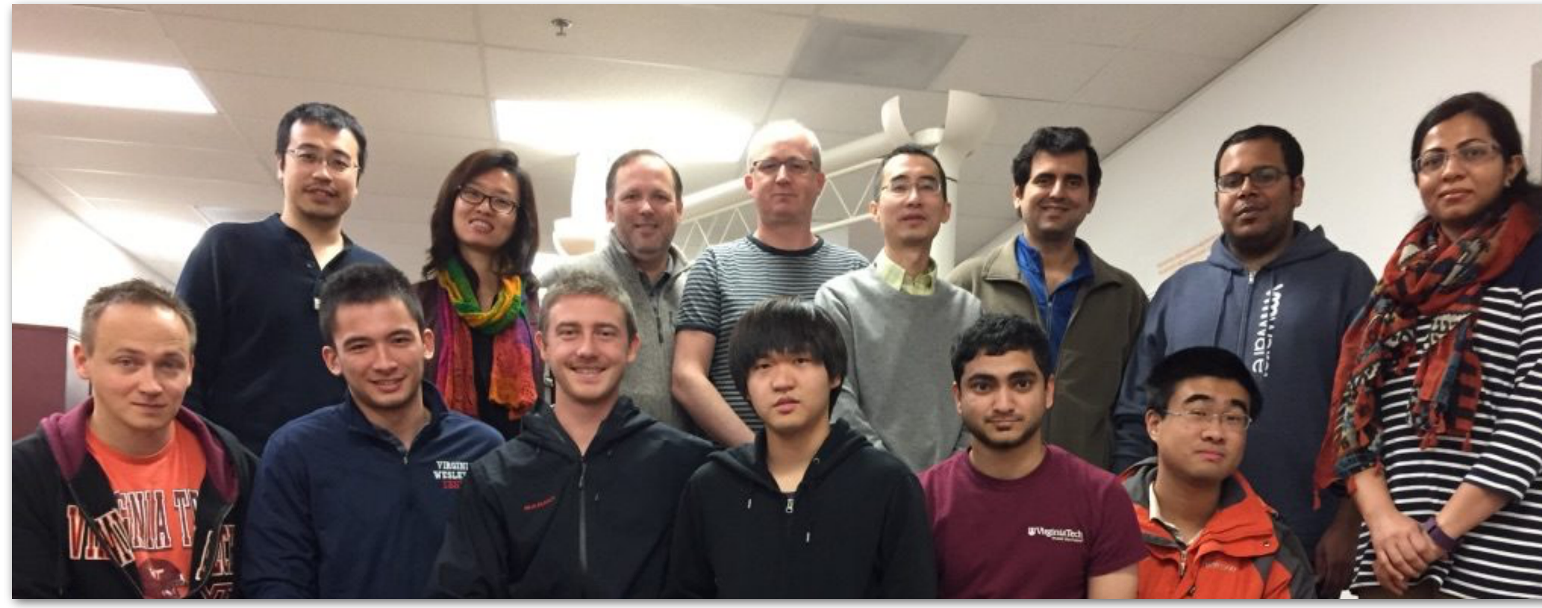
Variability in Systems

When the same computer is used to execute the same program repeatedly, most measurable performance characteristics will vary.



Managing this performance variability can be very important in applications where we want to meet some set requirements.

VarSys: Modeling and Managing Variability



High Performance Computing

HPC systems consume a lot of energy and time, both are functions of how the system was built and configured. Models can be used to optimize a configuration.

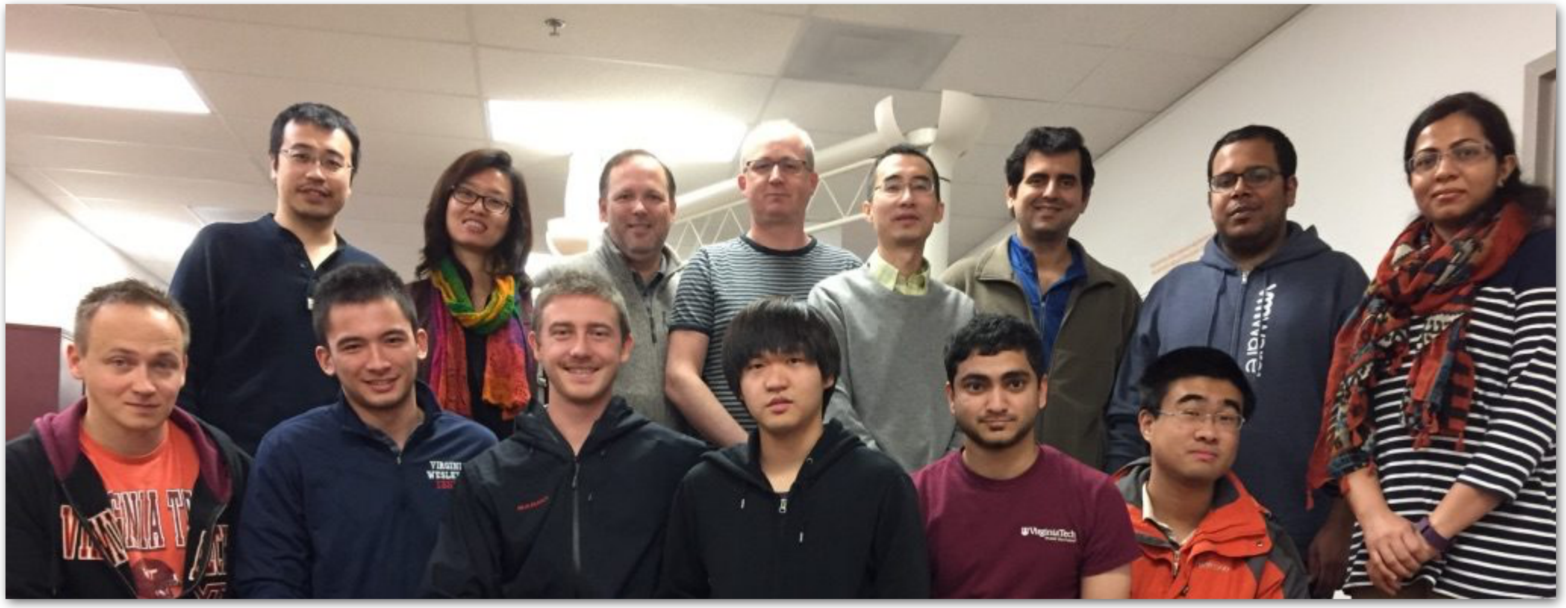
Cloud Computing

Small savings in compute time and performance magnify greatly when 1000's of machines are involved. Service Level Agreements (SLAs) can be tightened.

Computer Security

A strong understanding of variability can improve defenses against malicious users by demonstrating new vulnerabilities, and helping prevent side channel attacks.

VarSys: Modeling and Managing Variability



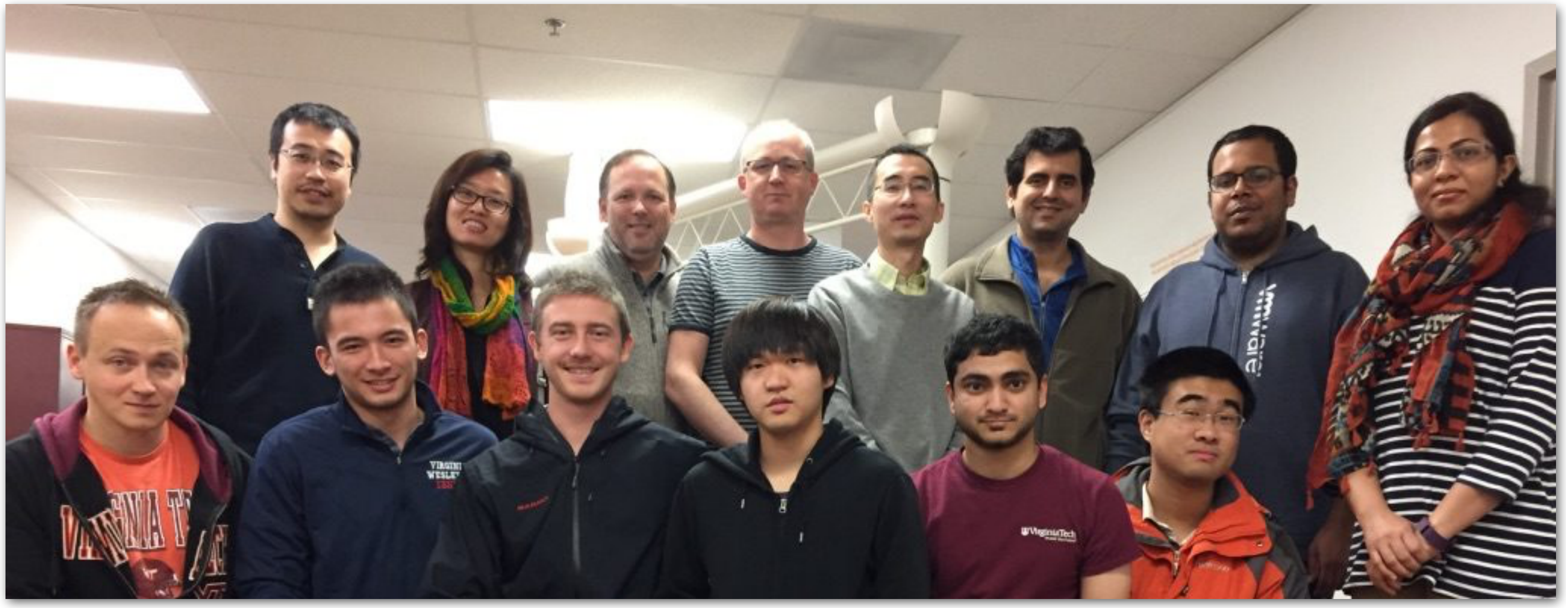
High Performance Computing
HPC systems consume a lot of energy and time, both are functions of how the system was built and configured. Models can be used to optimize a configuration.

Cloud Computing
Small savings in compute time and performance magnify greatly when 1000's of machines are involved. Service Level Agreements (SLAs) can be tightened.

Computer Security
A strong understanding of variability can improve defenses against malicious users by demonstrating new vulnerabilities, and helping prevent side channel attacks.

Variability is important in many aspects of computation.

VarSys: Modeling and Managing Variability



High Performance Computing
HPC systems consume a lot of energy and time, both are functions of how the system was built and configured. Models can be used to optimize a configuration.

Cloud Computing
Small savings in compute time and performance magnify greatly when 1000's of machines are involved. Service Level Agreements (SLAs) can be tightened.

Computer Security
A strong understanding of variability can improve defenses against malicious users by demonstrating new vulnerabilities, and helping prevent side channel attacks.

Variability is important in many aspects of computation.

Quantifying variability and constructing models of it may lead to improvements in all of these aspects of computation.

Chapters

Approximation: The Problem Description

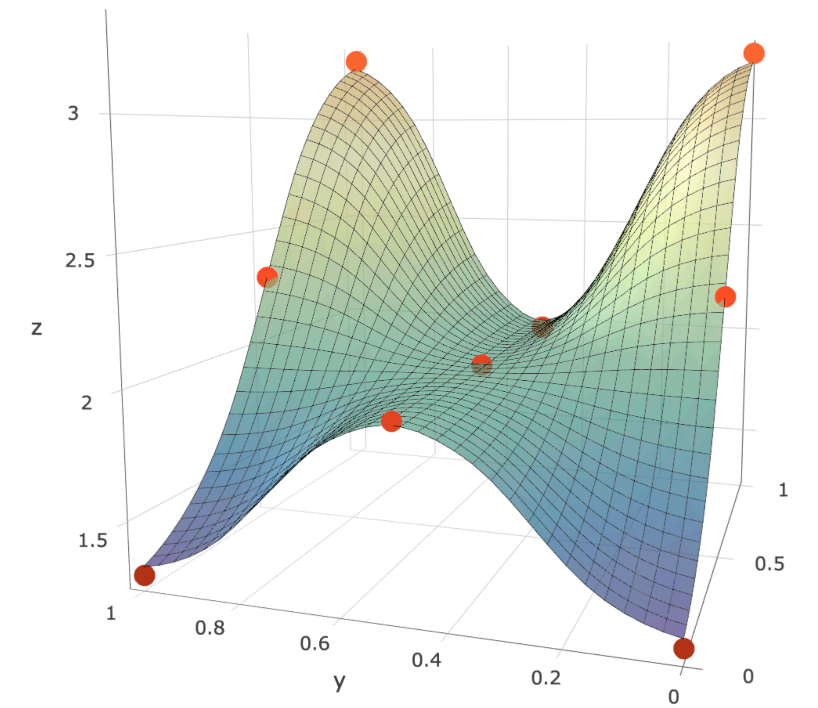
Given

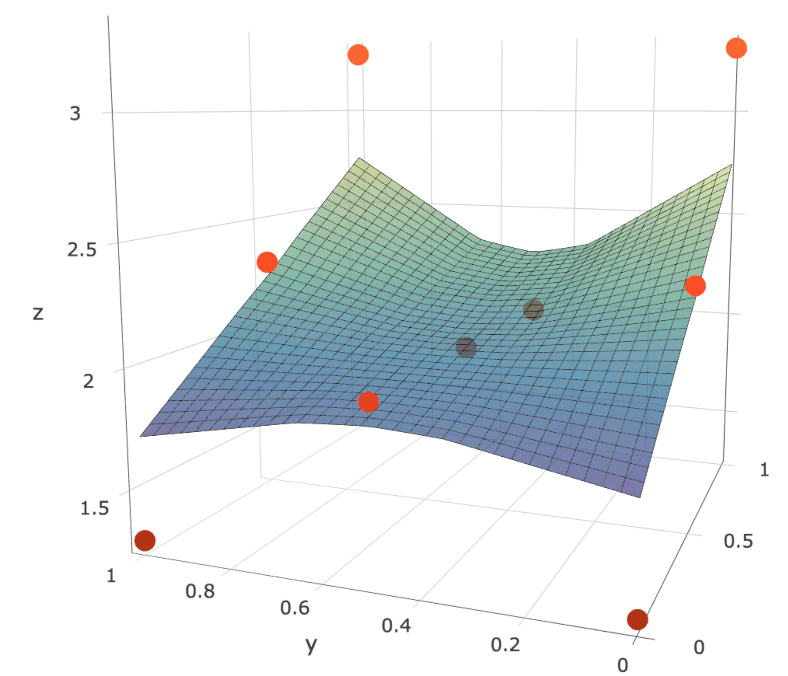
underlying function $f : \mathbb{R}^d \to \mathbb{R}$

data matrix $X^{d \times n}$ with column vectors $x^{(i)} \in \mathbb{R}^d$

function values $f(x^{(i)})$ for all $x^{(i)}$

vector $f(X)$ has elements $f(x^{(i)})$





Approximation: The Problem Description

Given

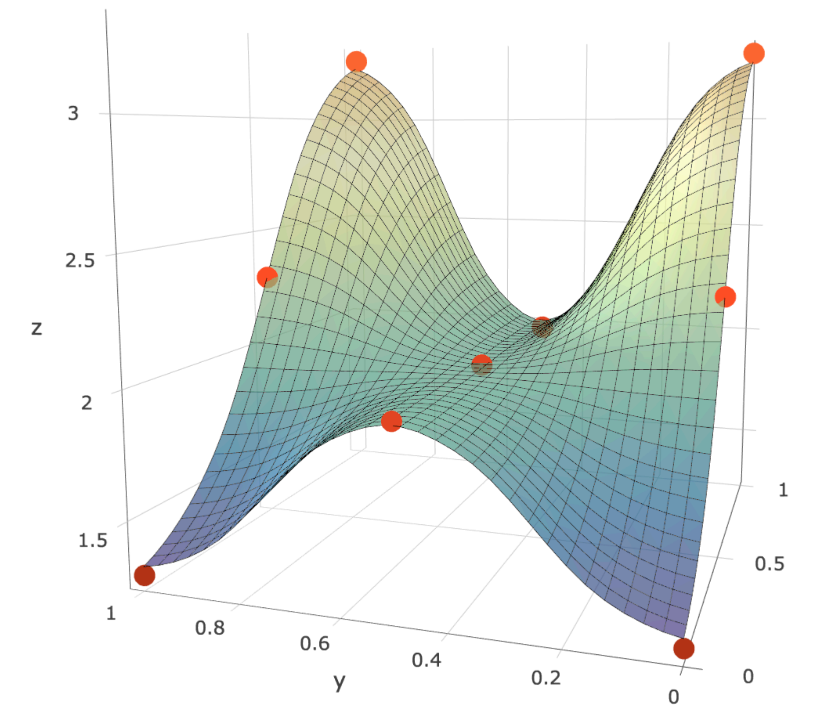
underlying function $f: \mathbb{R}^d \to \mathbb{R}$

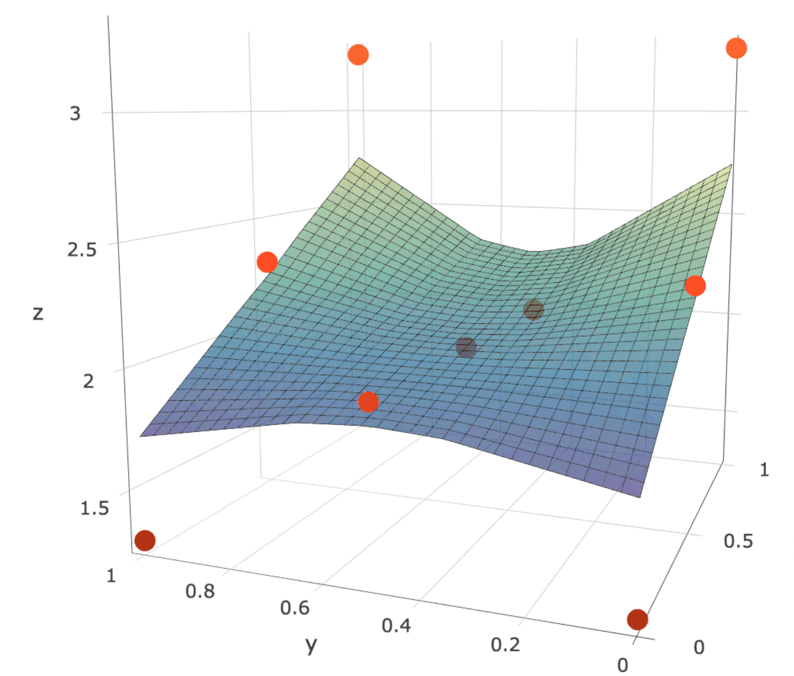
data matrix $X^{d \times n}$ with column vectors $x^{(i)} \in \mathbb{R}^d$

function values $f(x^{(i)})$ for all $x^{(i)}$

vector $f(X)$ has elements $f(x^{(i)})$

Generate a function $\hat{f}: \mathbb{R}^d \to \mathbb{R}$ such that:





Approximation: The Problem Description

Given

underlying function $f : \mathbb{R}^d \to \mathbb{R}$

data matrix $X^{d \times n}$ with column vectors $x^{(i)} \in \mathbb{R}^d$

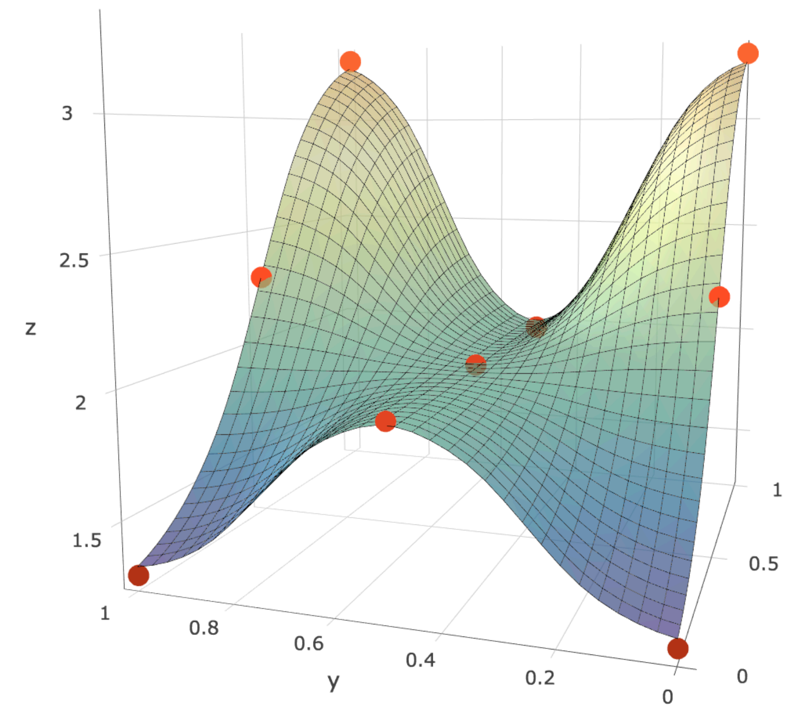
function values $f(x^{(i)})$ for all $x^{(i)}$

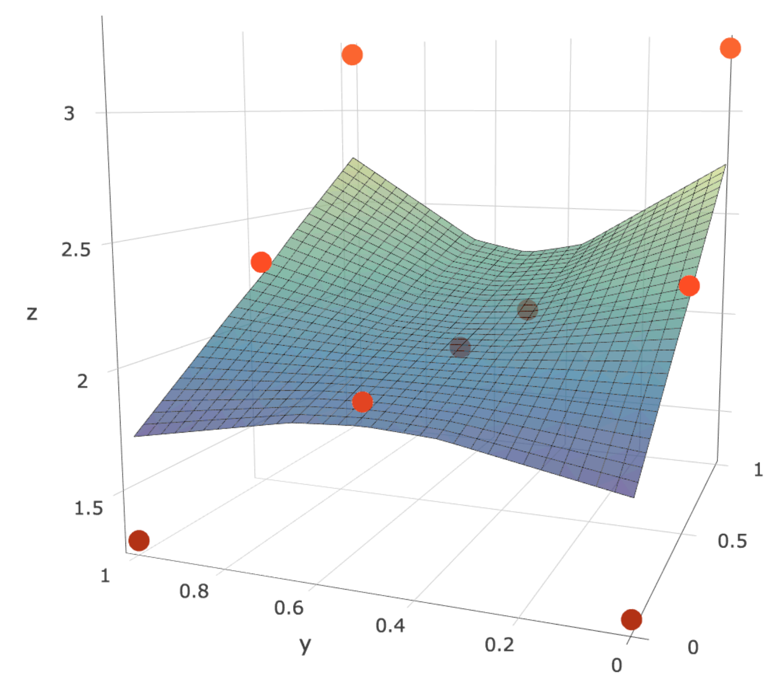
vector $f(X)$ has elements $f(x^{(i)})$

Generate a function $\hat{f} : \mathbb{R}^d \to \mathbb{R}$ such that:

Interpolation

$\hat{f}(x^{(i)})$ equals $f(x^{(i)})$ for all $x^{(i)}$





Approximation: The Problem Description

Given

underlying function $f : \mathbb{R}^d \to \mathbb{R}$

data matrix $X^{d \times n}$ with column vectors $x^{(i)} \in \mathbb{R}^d$

function values $f(x^{(i)})$ for all $x^{(i)}$

vector $f(X)$ has elements $f(x^{(i)})$

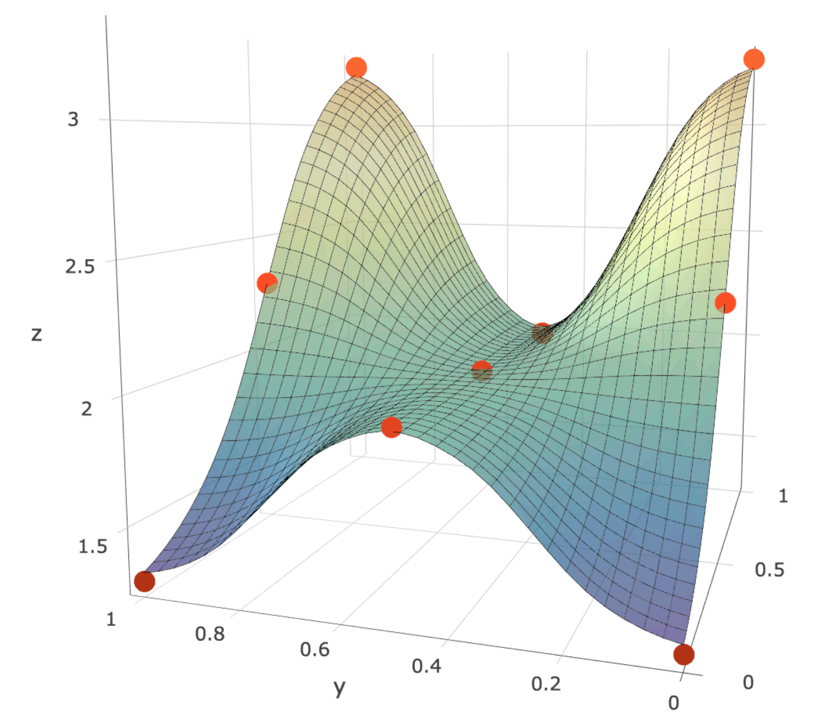
Generate a function $\hat{f} : \mathbb{R}^d \to \mathbb{R}$ such that:

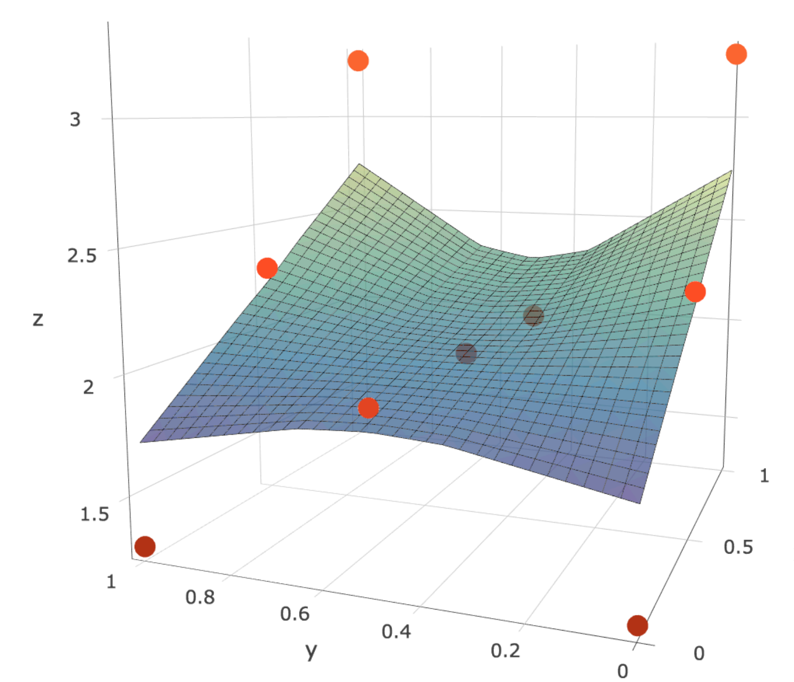
Interpolation

$\hat{f}(x^{(i)})$ equals $f(x^{(i)})$ for all $x^{(i)}$

Regression

$\hat{f}_p$ has parameters p and is the solution to $\min_p \| \hat{f}(X) - f(X) \|$



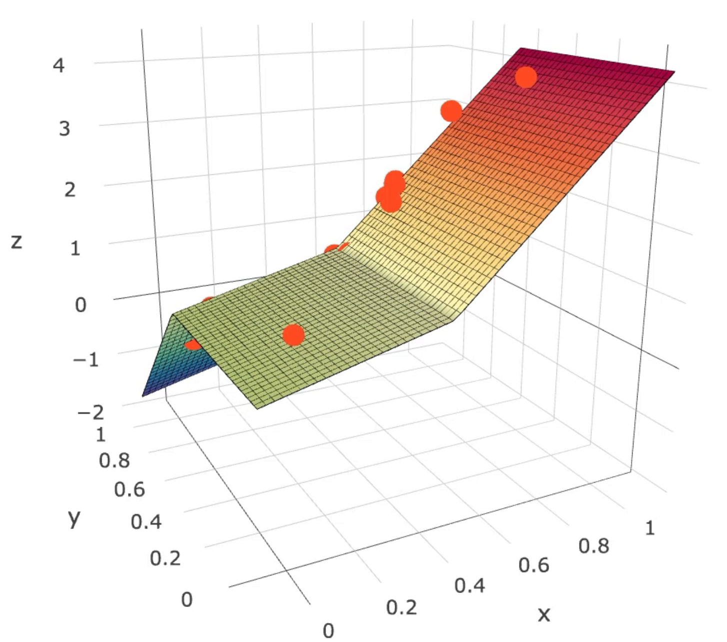


COMPUTER SCIENCE
VIRGINIA TECH.

Regression Techniques

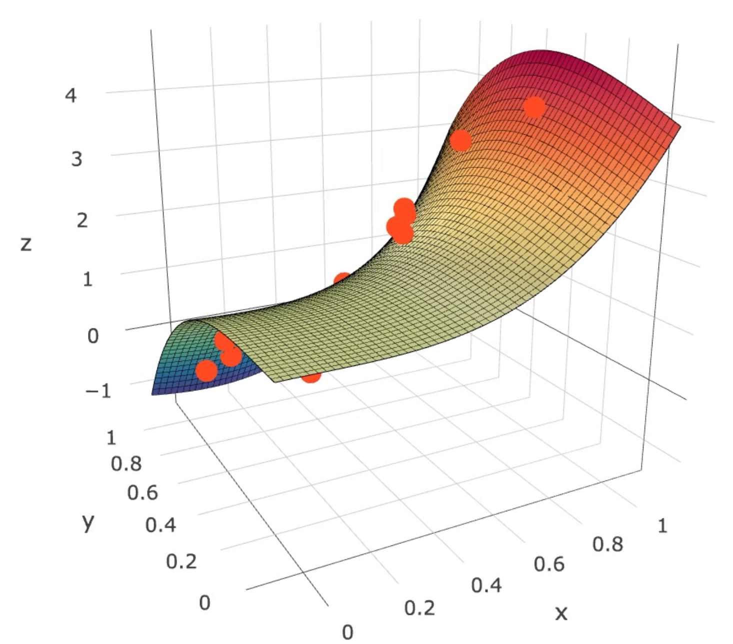
Multivariate Adaptive Regression Splines

$$B_{2j-1}(x) = B_l(x)\left(x_i - x_i^{(p)}\right)_+$$
$$B_{2j}(x) = B_k(x)\left(x_i - x_i^{(p)}\right)_-$$



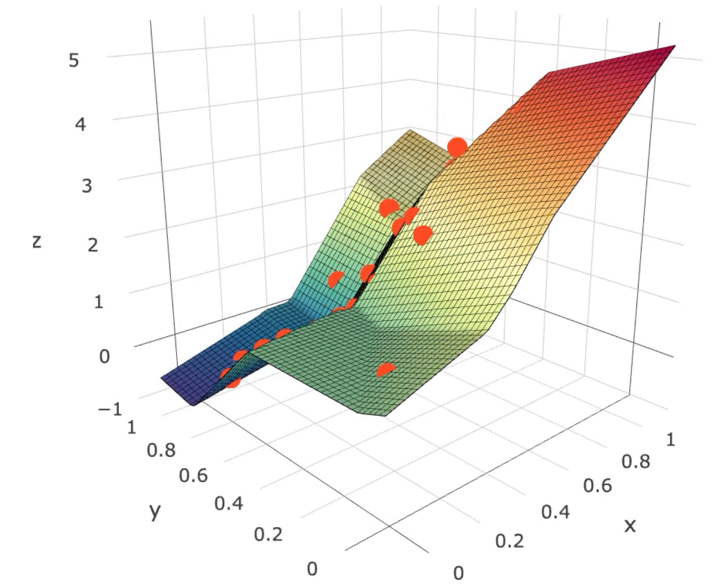
Support Vector Regressor

$$p(x) = \sum_{i=1}^{n} a_i K\left(x, x^{(i)}\right) + b$$



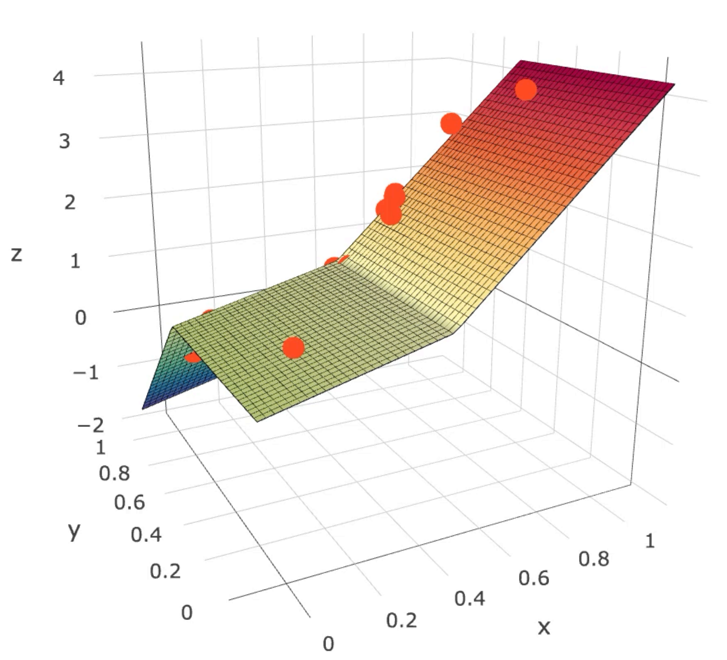
Multilayer Perceptron Regressor

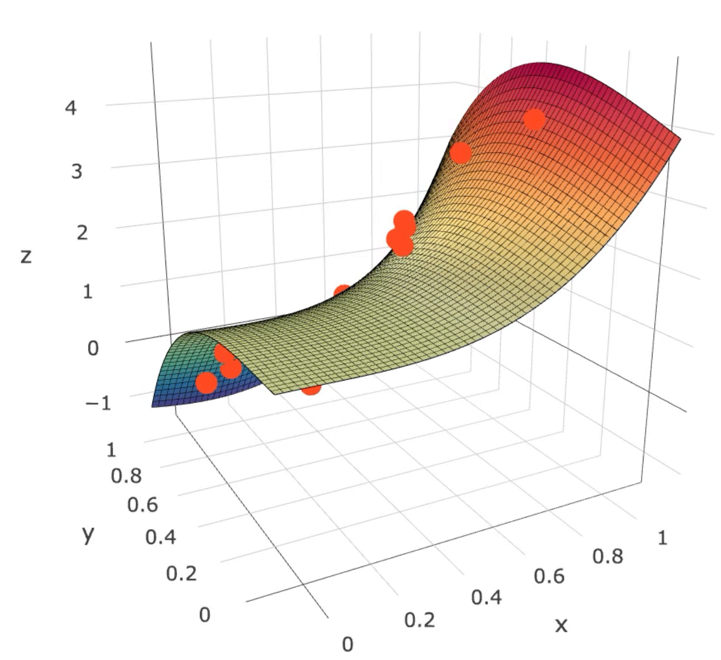
$$l(u) = \left(u^t W_l + c_l\right)_+$$

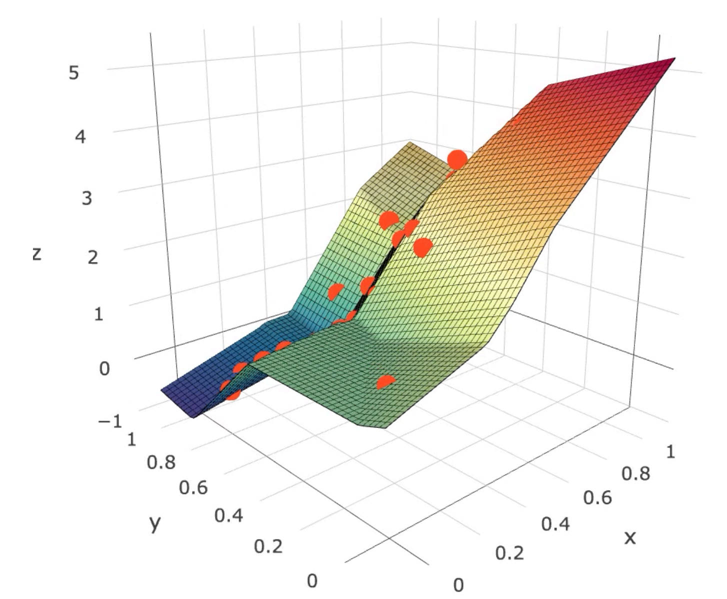


Regression Techniques

Multivariate Adaptive Regression Splines	Support Vector Regressor	Multilayer Perceptron Regressor
$B_{2j-1}(x) = B_l(x)\left(x_i - x_i^{(p)}\right)_+$ $B_{2j}(x) = B_k(x)\left(x_i - x_i^{(p)}\right)_-$	$p(x) = \sum_{i=1}^{n} a_i K\left(x, x^{(i)}\right) + b$	$l(u) = \left(u^t W_l + c_l\right)_+$



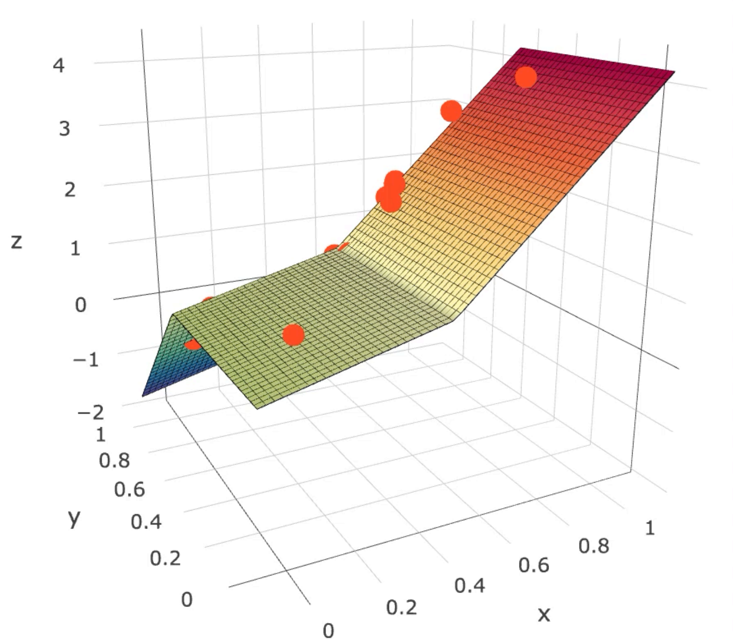




Regression Techniques

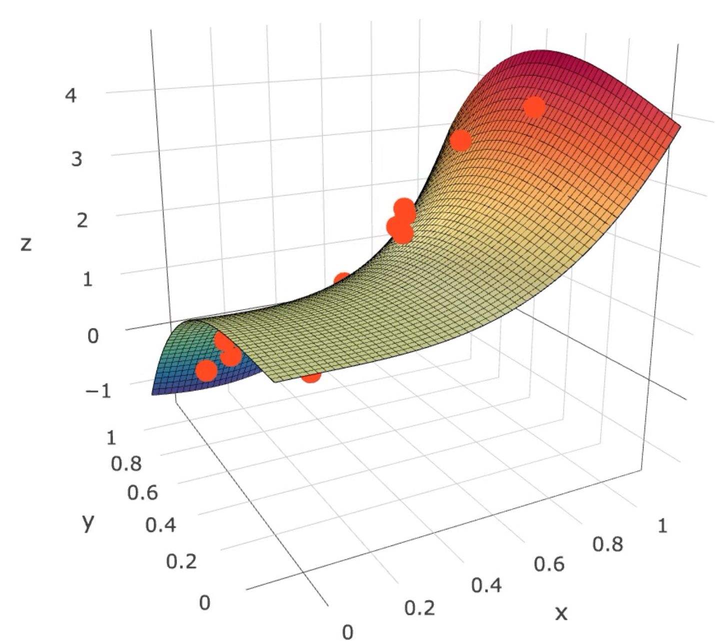
Multivariate Adaptive Regression Splines

$$B_{2j-1}(x) = B_l(x)\left(x_i - x_i^{(p)}\right)_+$$
$$B_{2j}(x) = B_k(x)\left(x_i - x_i^{(p)}\right)_-$$



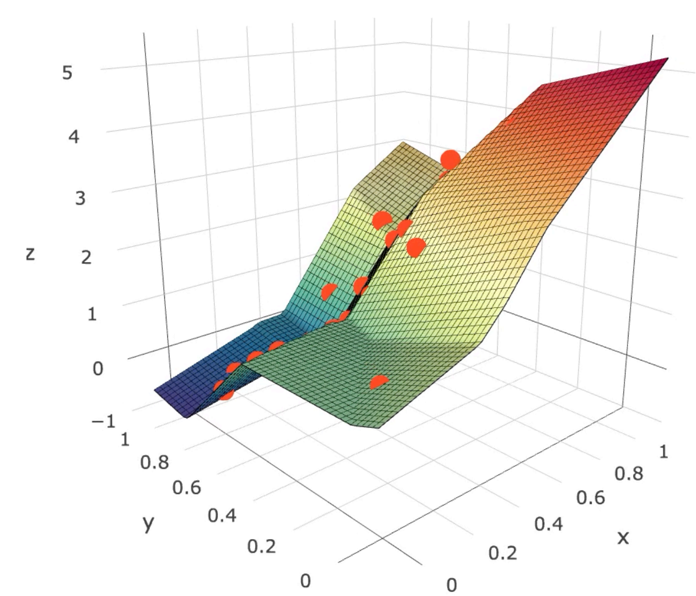
Support Vector Regressor

$$p(x) = \sum_{i=1}^{n} a_i K\left(x, x^{(i)}\right) + b$$



Multilayer Perceptron Regressor

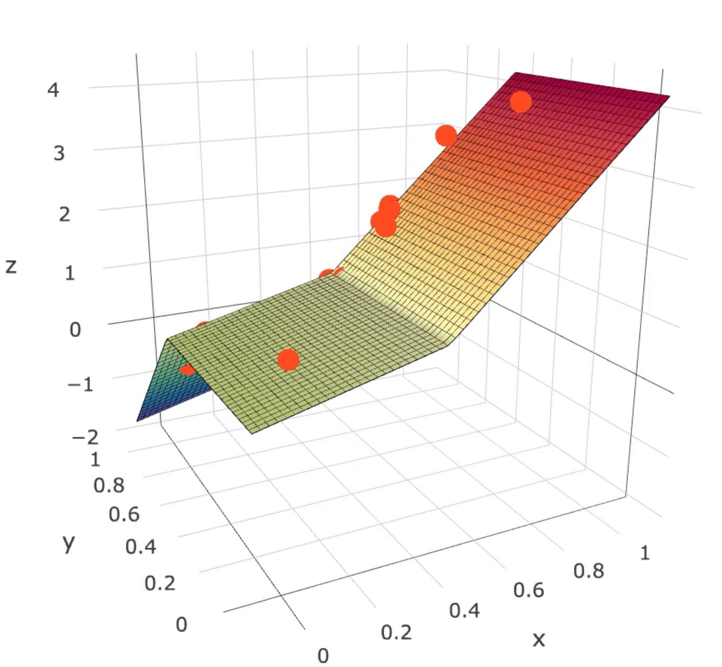
$$l(u) = \left(u^t W_l + c_l\right)_+$$



Regression Techniques

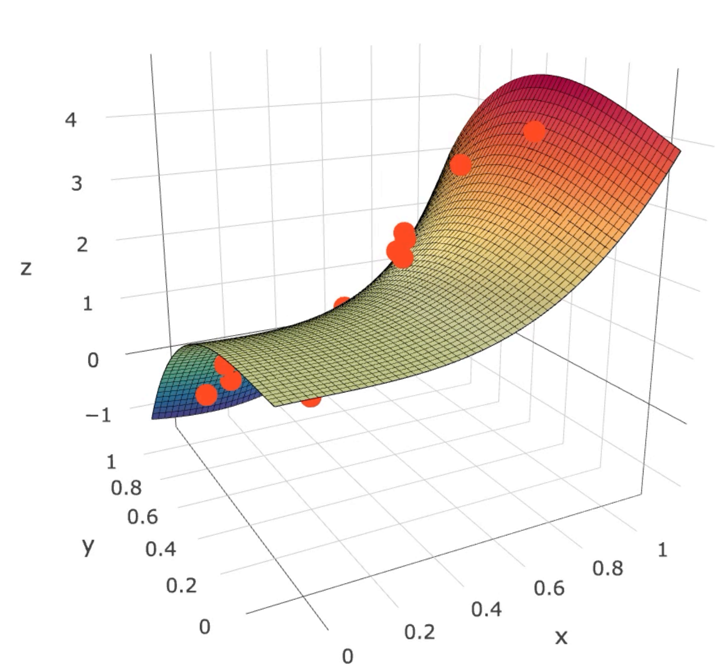
Multivariate Adaptive Regression Splines

$$B_{2j-1}(x) = B_l(x)\left(x_i - x_i^{(p)}\right)_+$$
$$B_{2j}(x) = B_k(x)\left(x_i - x_i^{(p)}\right)_-$$



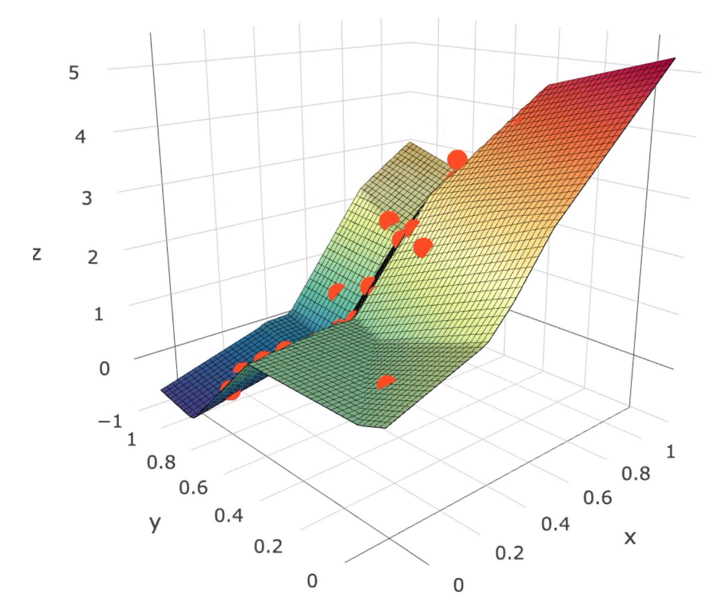
Support Vector Regressor

$$p(x) = \sum_{i=1}^{n} a_i K\left(x, x^{(i)}\right) + b$$



Multilayer Perceptron Regressor

$$l(u) = \left(u^t W_l + c_l\right)_+$$



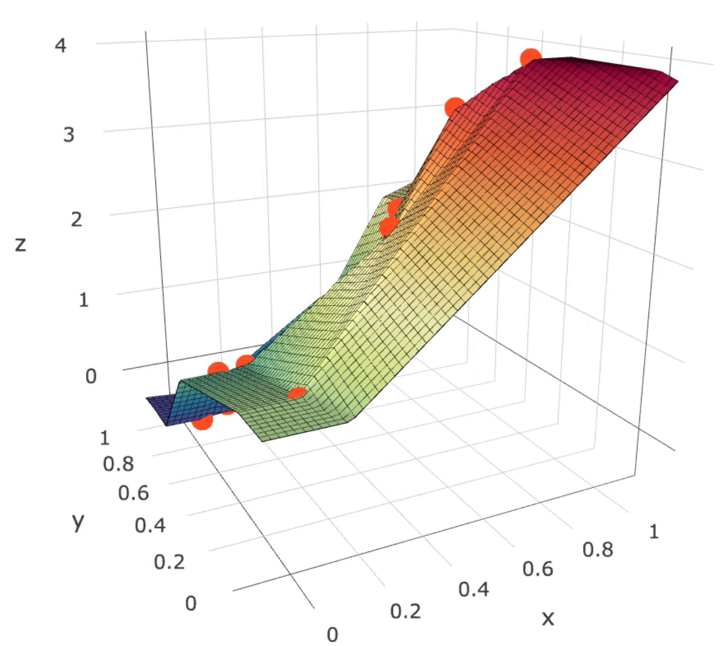
Interpolation Techniques

Delaunay

simplicial mesh

$$y = \sum_{i=0}^{d} w_i x^{(i)}, \quad \sum_{i=0}^{d} w_i = 1, \quad w_i \geq 0, \quad i = 0, \ldots, d$$

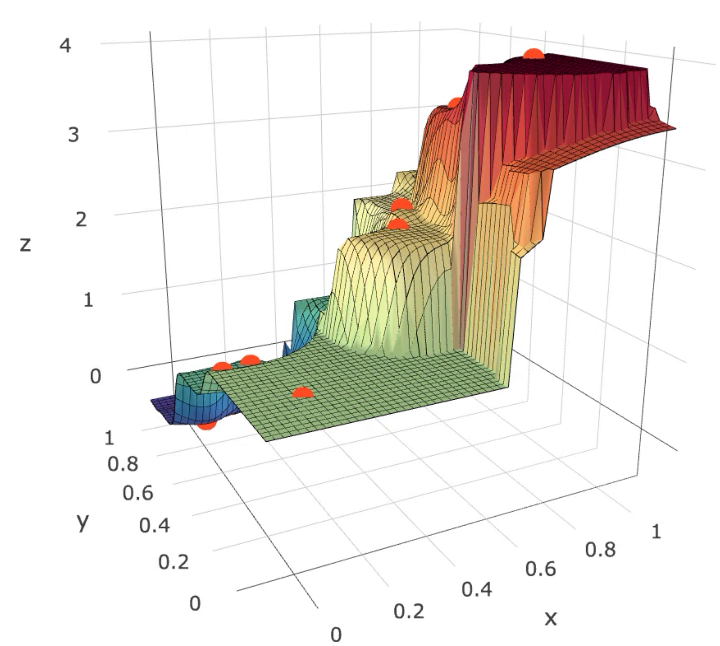
$$\hat{f}(y) = \sum_{i=0}^{d} w_i f\left(x^{(i)}\right)$$



Modified Shepard

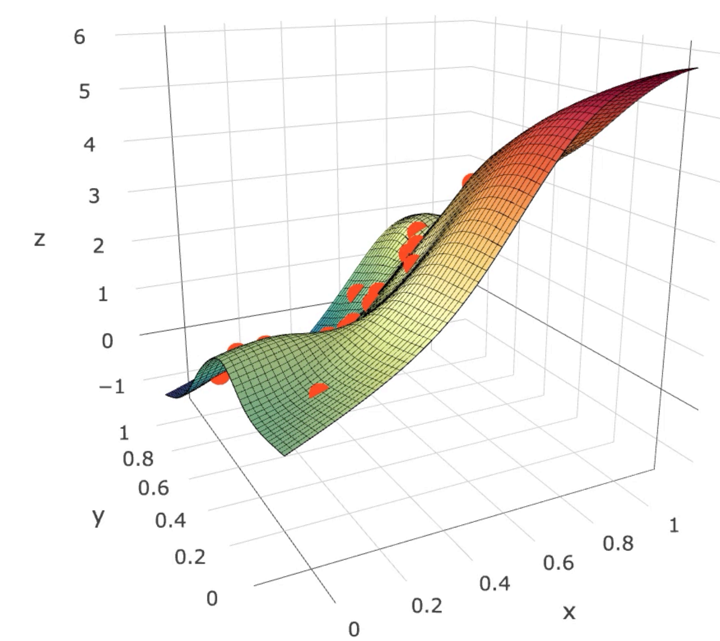
squared inverse distance

$$p(x) = \frac{\sum_{k=1}^{n} W_k(x) f\left(x^{(k)}\right)}{\sum_{k=1}^{n} W_k(x)}$$



Linear Shepard

local linear fit



Interpolation Techniques

Delaunay

simplicial mesh

$$y = \sum_{i=0}^{d} w_i x^{(i)}, \quad \sum_{i=0}^{d} w_i = 1, \quad w_i \geq 0, \quad i = 0, \ldots, d$$

$$\hat{f}(y) = \sum_{i=0}^{d} w_i f\left(x^{(i)}\right)$$

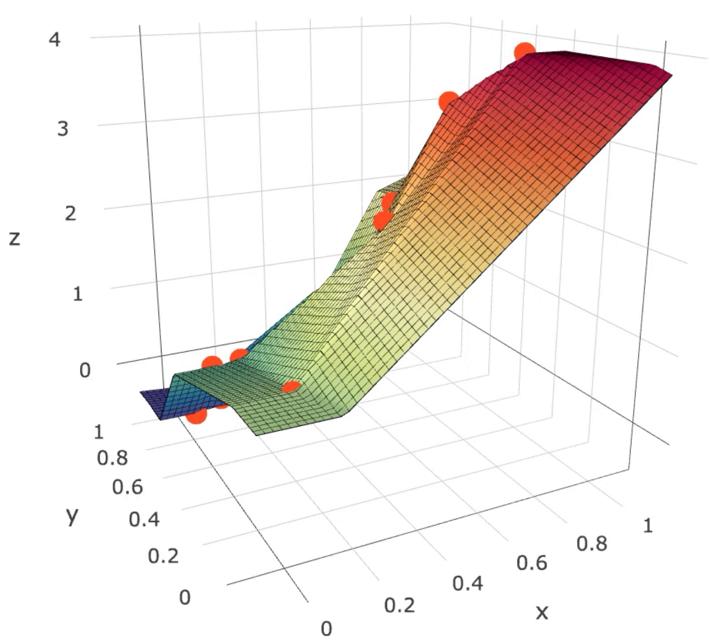
Modified Shepard

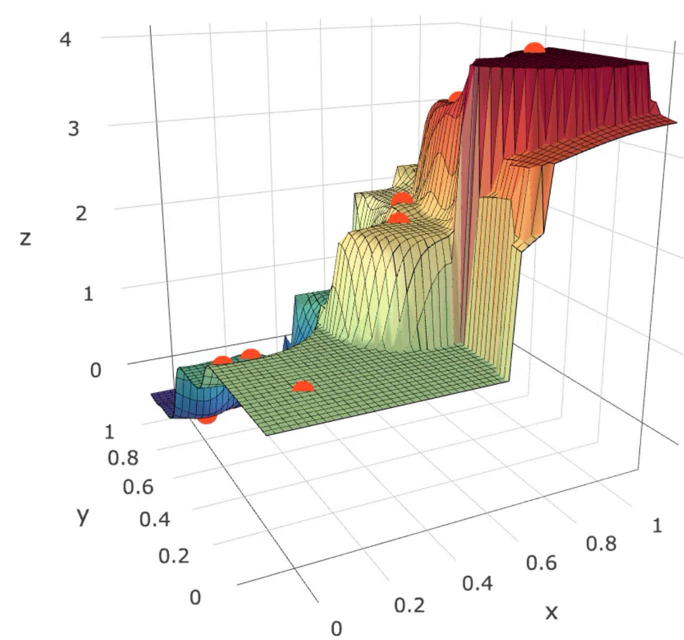
squared inverse distance

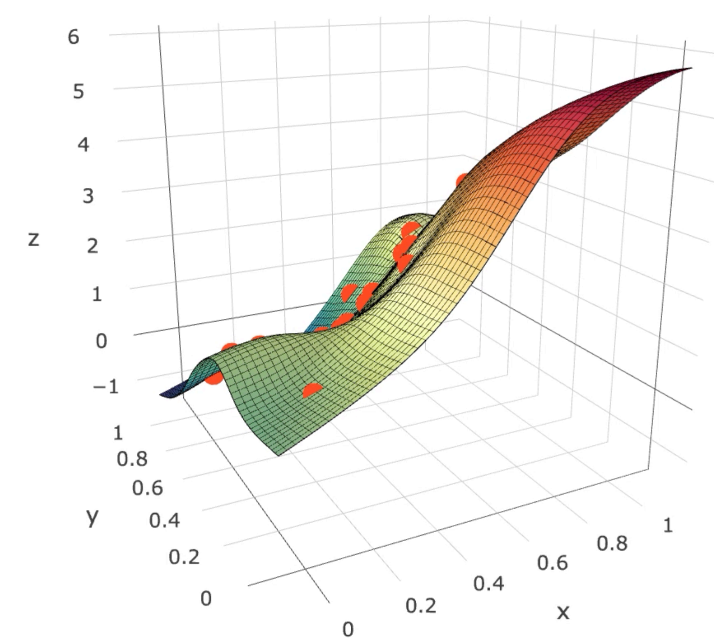
$$p(x) = \frac{\sum_{k=1}^{n} W_k(x) f\left(x^{(k)}\right)}{\sum_{k=1}^{n} W_k(x)}$$

Linear Shepard

local linear fit







Interpolation Techniques

Delaunay

simplicial mesh

$$y = \sum_{i=0}^{d} w_i x^{(i)}, \quad \sum_{i=0}^{d} w_i = 1, \quad w_i \geq 0, \quad i = 0, \ldots, d$$

$$\hat{f}(y) = \sum_{i=0}^{d} w_i f\left(x^{(i)}\right)$$

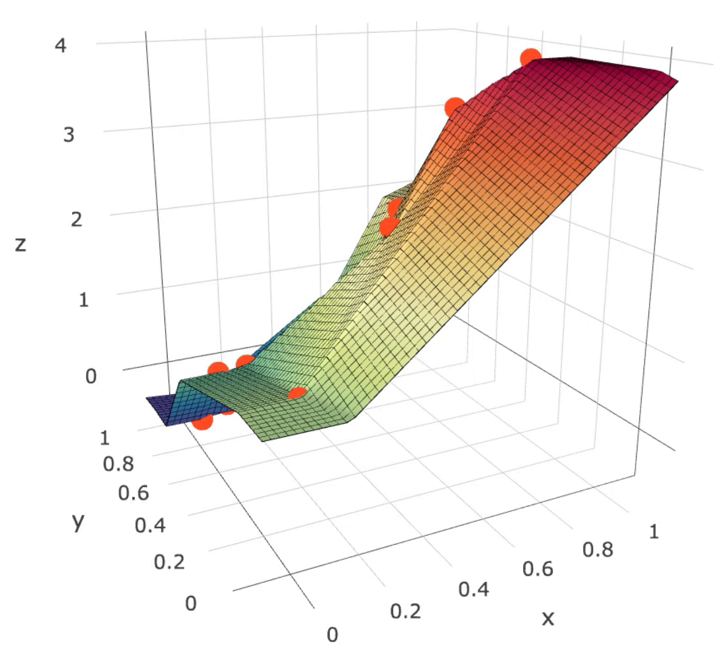
Modified Shepard

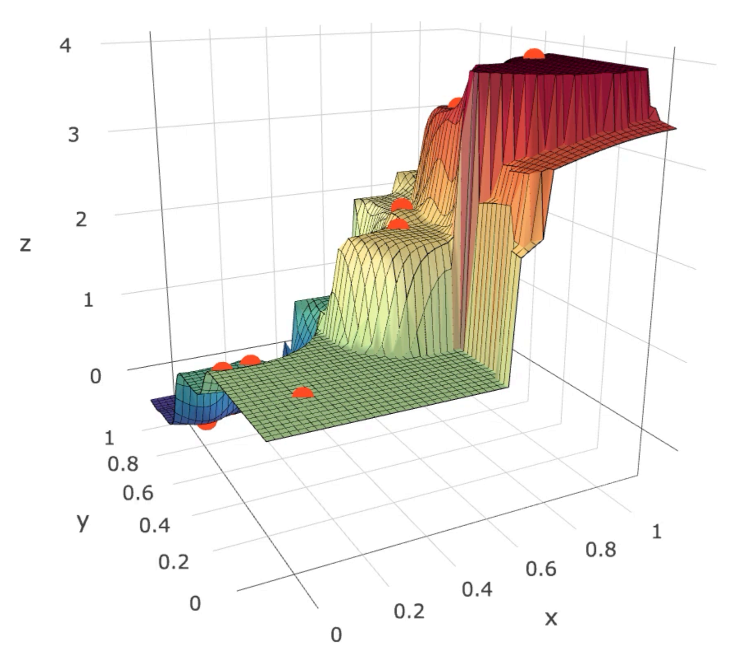
squared inverse distance

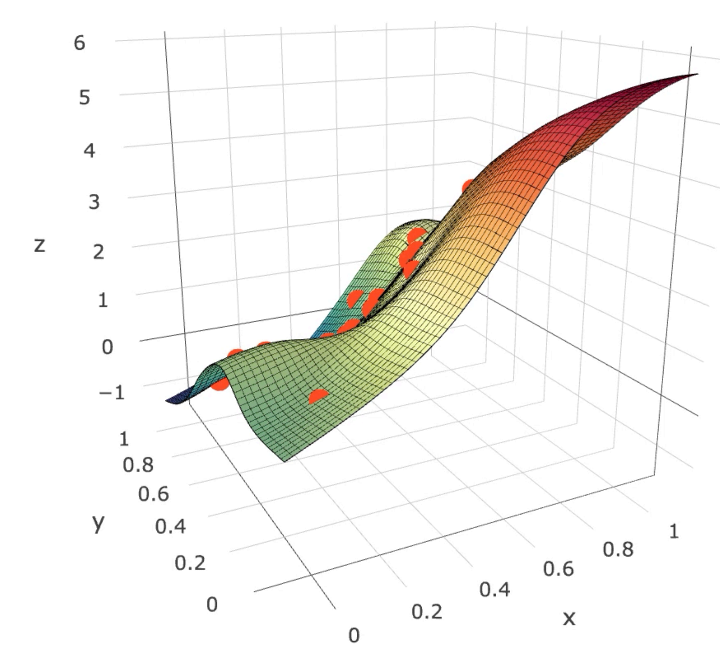
$$p(x) = \frac{\sum_{k=1}^{n} W_k(x) f\left(x^{(k)}\right)}{\sum_{k=1}^{n} W_k(x)}$$

Linear Shepard

local linear fit







Interpolation Techniques

Delaunay

simplicial mesh

$$y = \sum_{i=0}^{d} w_i x^{(i)}, \quad \sum_{i=0}^{d} w_i = 1, \quad w_i \geq 0, \quad i = 0, \ldots, d$$

$$\hat{f}(y) = \sum_{i=0}^{d} w_i f\left(x^{(i)}\right)$$

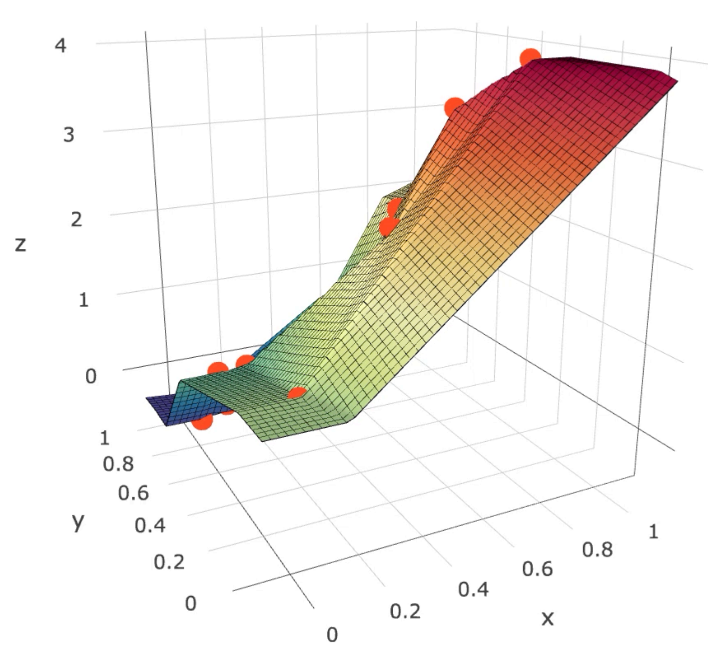
Modified Shepard

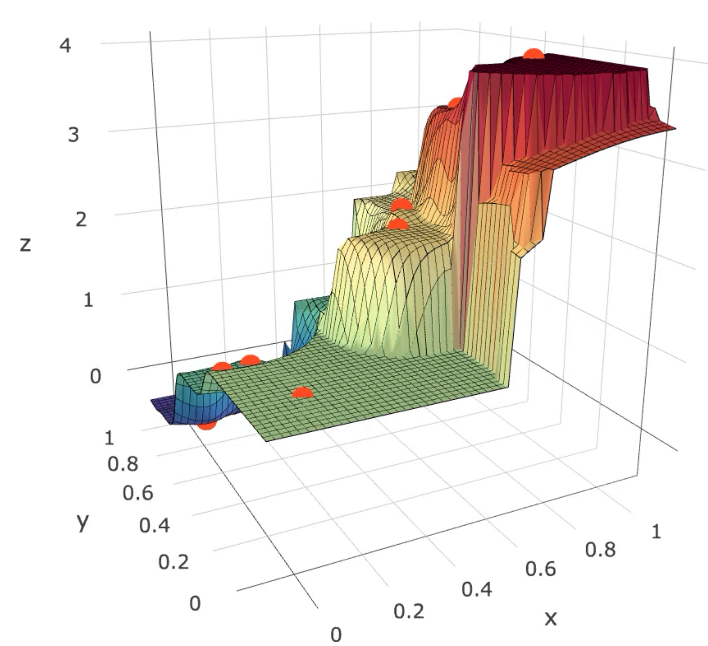
squared inverse distance

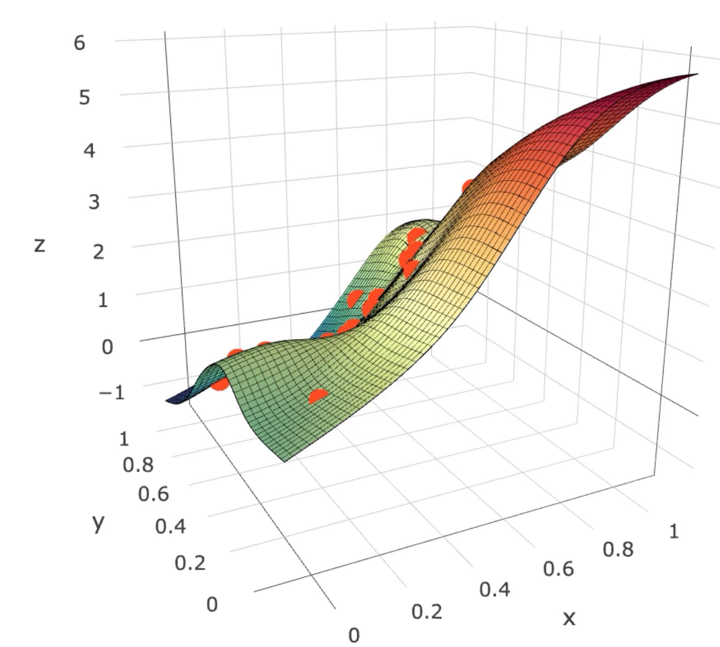
$$p(x) = \frac{\sum_{k=1}^{n} W_k(x) f\left(x^{(k)}\right)}{\sum_{k=1}^{n} W_k(x)}$$

Linear Shepard

local linear fit







Chapters

IOzone — A System Benchmark

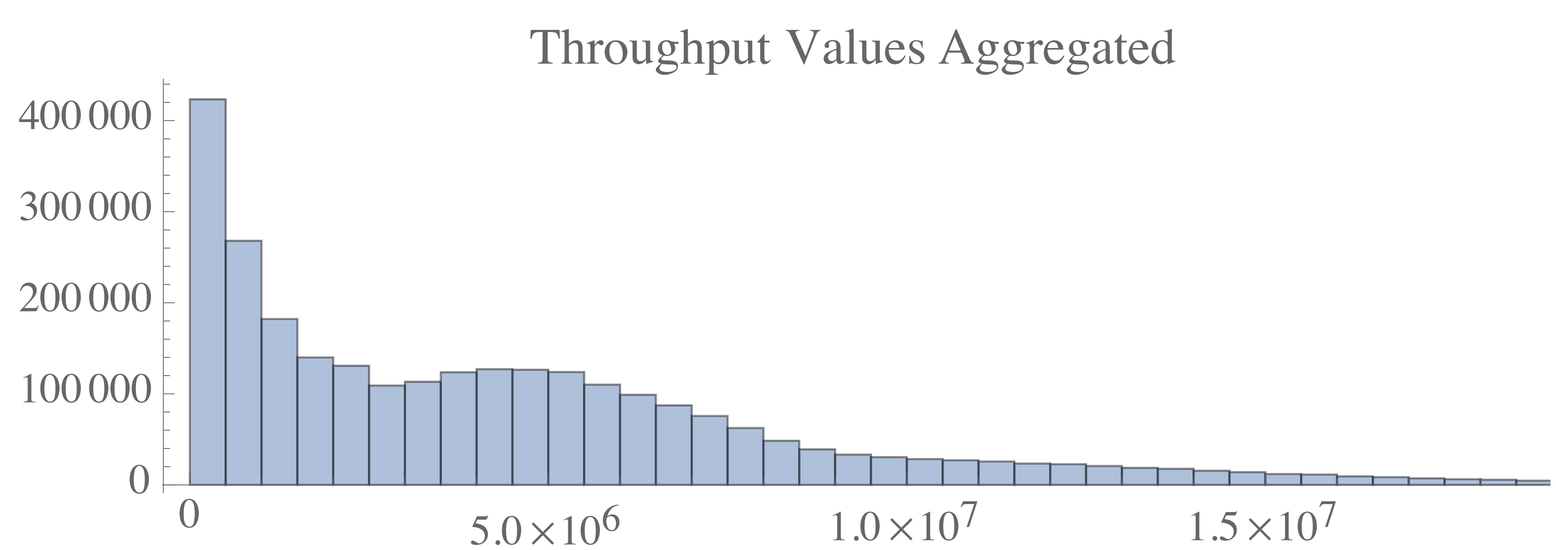
System Specs:

Two Intel Xeon E5-2637 CPUs with total 16 CPU cores and 16GB DRAM per node, at 12 nodes.

Ext4 filesystem above an Intel SSDSC2BA20 SSD drive.

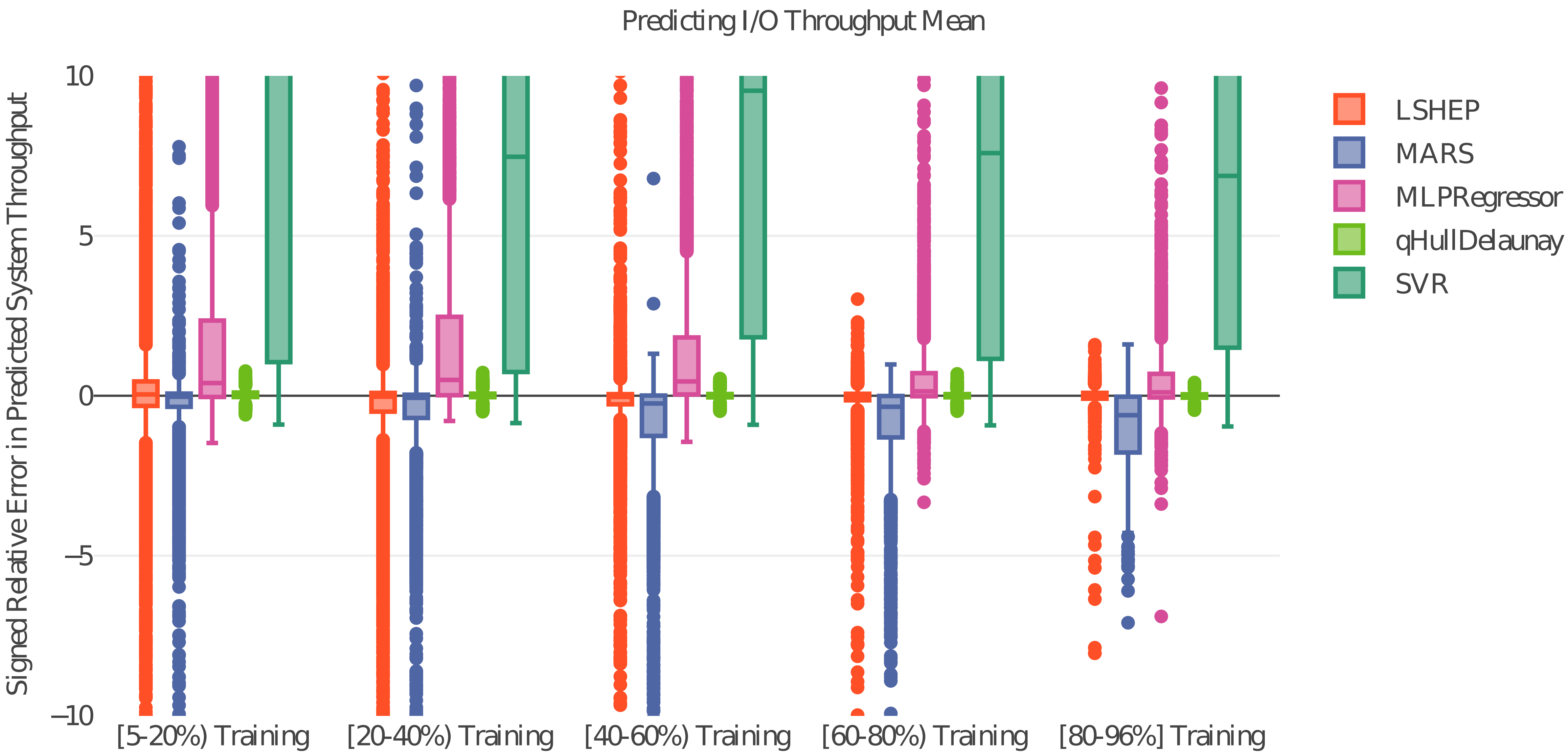
Each of ~20K unique system configurations were run 150 times.

System Parameter	**Values**
File Size (KB)	4, 16, 64, 256, 1024, 4096, 8192, 16384
Record Size (KB)	4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384
Thread Count	1, 8, 16, 24, 32, 40, 48, 56, 64
Frequency (GHz)	1.2, 1.6, 2, 2.3, 2.8, 3.2, 3.5
Test Type	Readers, Rereaders, Random Readers, Initial Writers, Rewriters, Random Writers



COMPUTER SCIENCE
VIRGINIA TECH.

Mean Prediction Results



COMPUTER SCIENCE
VIRGINIA TECH.

Variance Prediction Results

Chapter Takeaways

Multivariate models of HPC system performance can predict I/O throughput mean and variance.

The Delaunay method produces considerably better results for mean and variance prediction.

Throughput variance is harder to predict than mean throughput.

Chapters

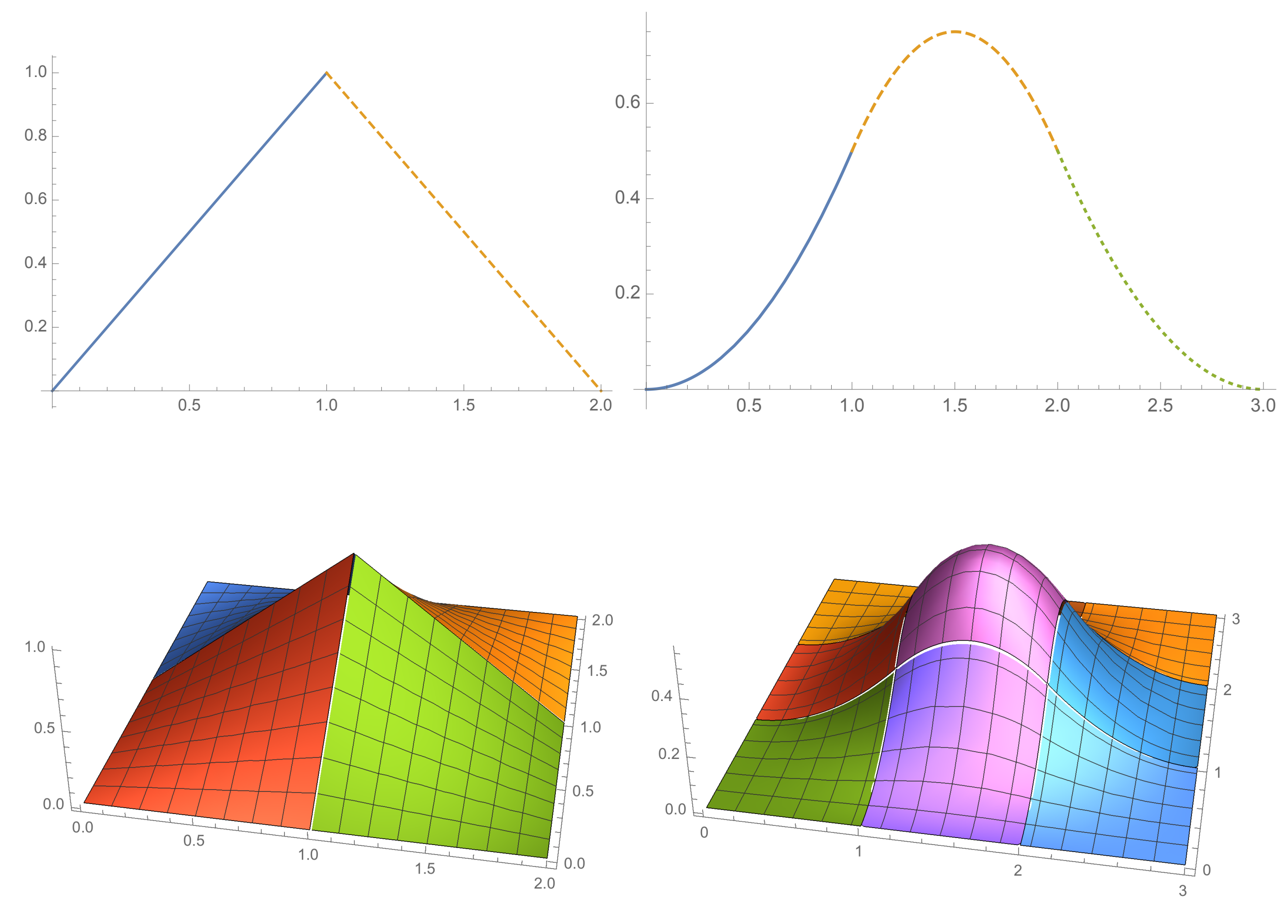
1. The Importance and Applications of Variability
2. Algorithms for Constructing Approximations
3. Naive Approximations of Variability
4. Box-Splines: Uses, Constructions, and Applications
5. Stronger Approximations of Variability
6. An Error Bound for Piecewise Linear Interpolation
7. A Package for Monotone Quintic Spline Interpolation

Box Splines

Proposed by C. de Boor as an extension of B-Splines into multiple dimensions (without using tensor products).

Can be shifted and scaled without losing smoothness.

Computationally expensive.

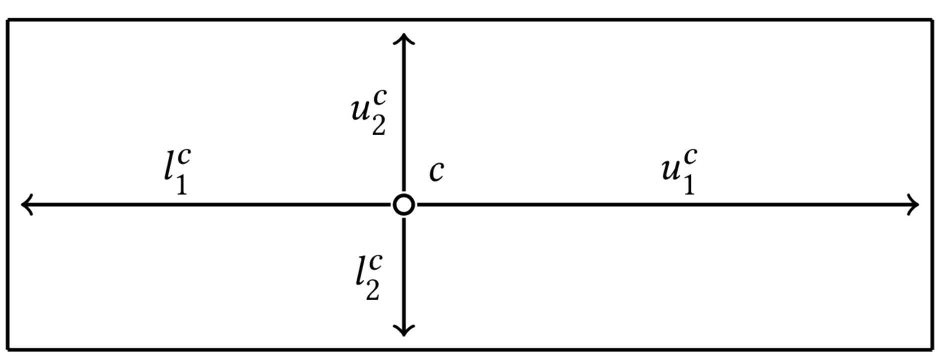


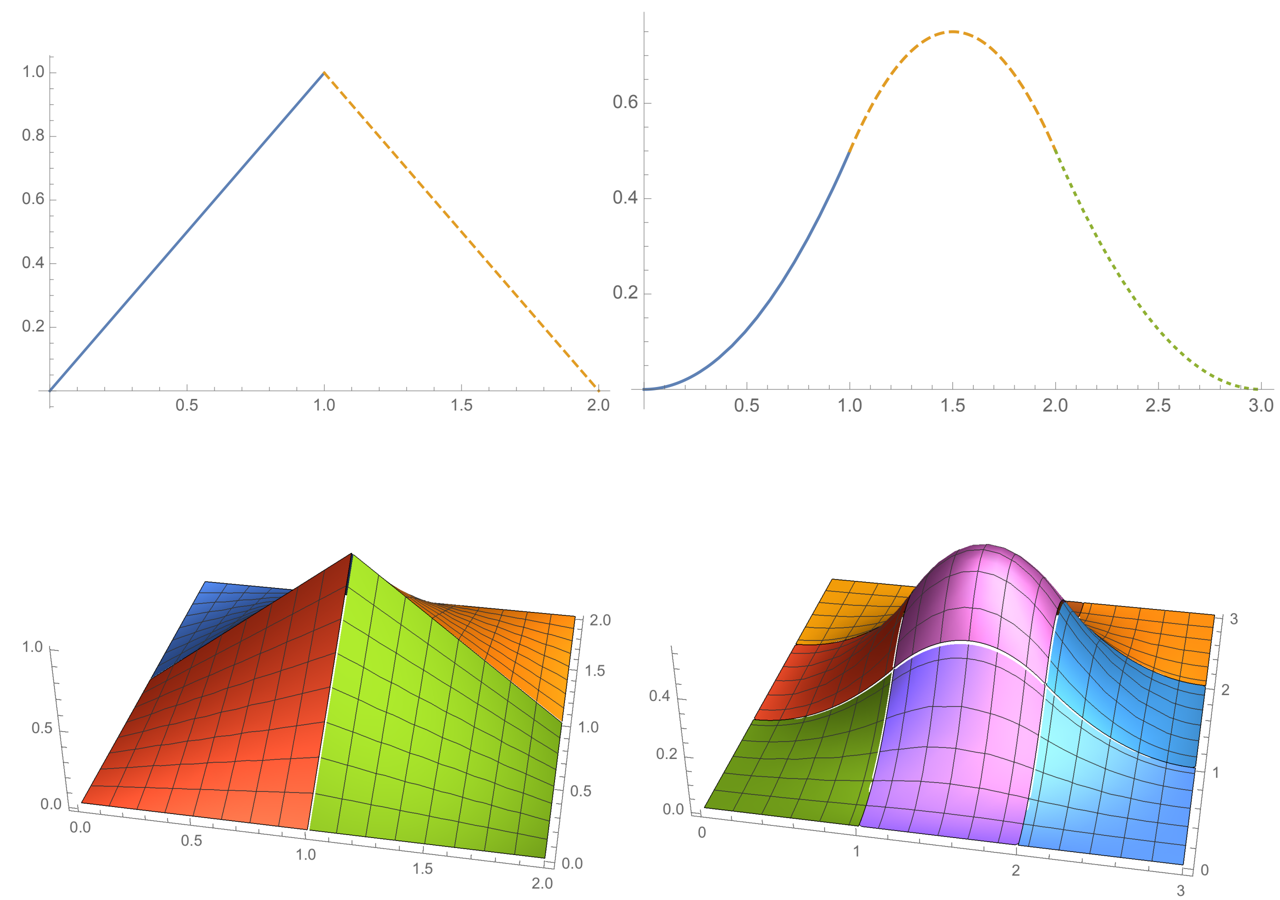
Box Splines

Proposed by C. de Boor as an extension of B-Splines into multiple dimensions (without using tensor products).

Can be shifted and scaled without losing smoothness.

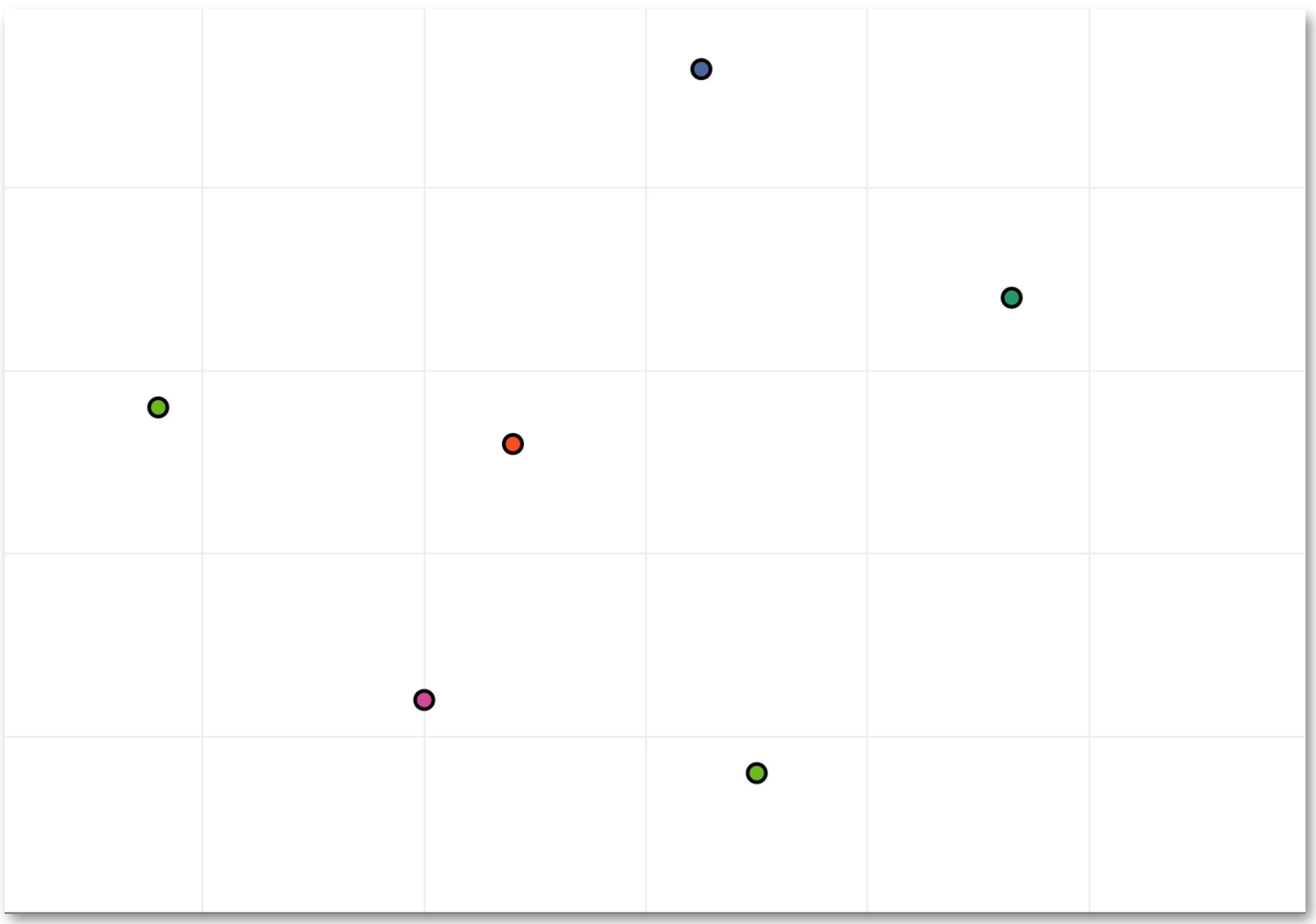
Computationally expensive.





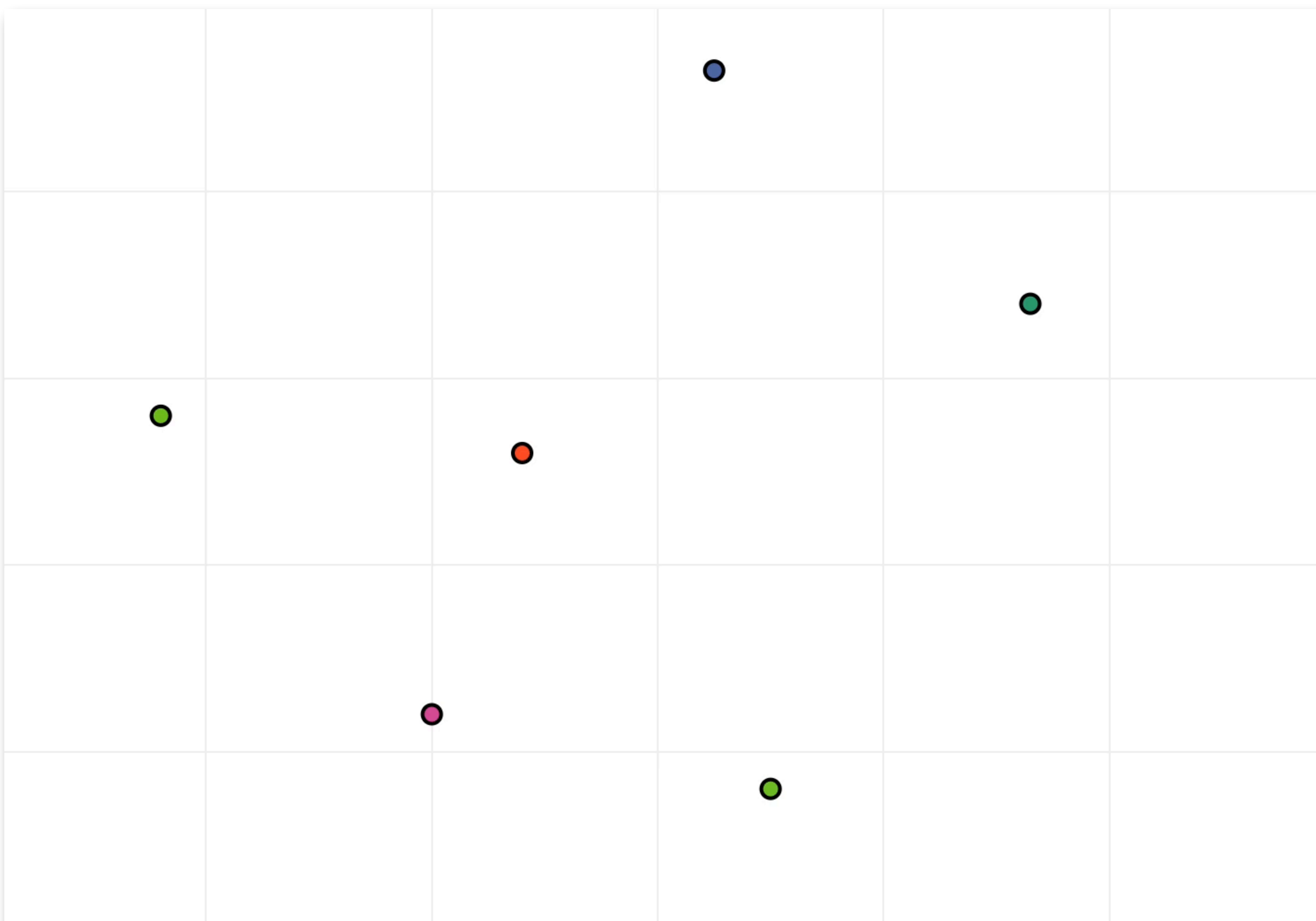
Construction of a *Mesh*

Max Boxes	Iterative Boxes	Voronoi Cells
$\mathcal{O}(n^2 d \log n)$	$\mathcal{O}(n^2 d)$	$\mathcal{O}(n^2 d)$



Construction of a *Mesh*

Max Boxes	Iterative Boxes	Voronoi Cells
$\mathcal{O}(n^2 d \log n)$	$\mathcal{O}(n^2 d)$	$\mathcal{O}(n^2 d)$

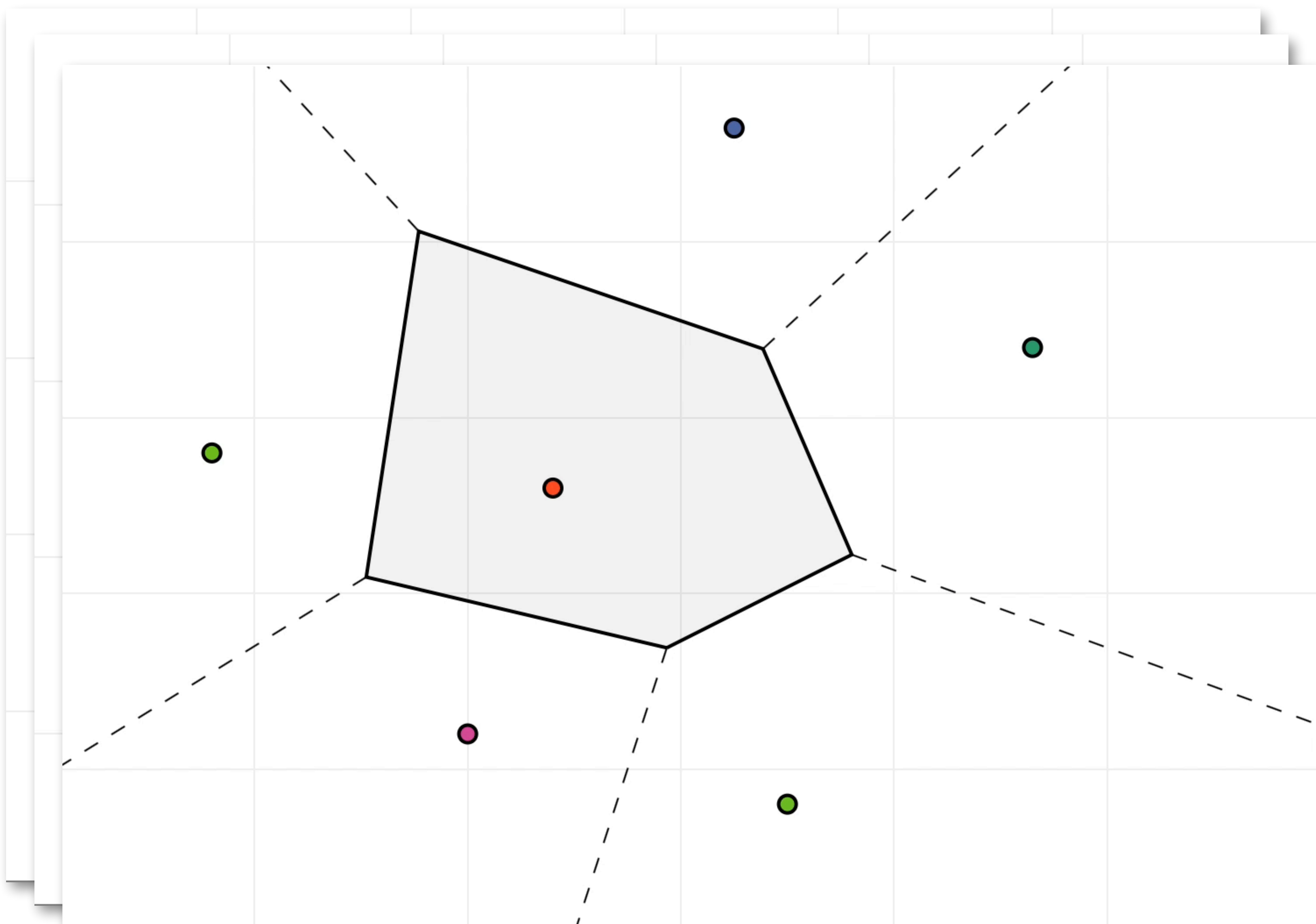


Construction of a *Mesh*

Max Boxes	Iterative Boxes	Voronoi Cells
$\mathcal{O}(n^2 d \log n)$	$\mathcal{O}(n^2 d)$	$\mathcal{O}(n^2 d)$

Construction of a *Mesh*

Max Boxes	Iterative Boxes	Voronoi Cells
$\mathcal{O}(n^2 d \log n)$	$\mathcal{O}(n^2 d)$	$\mathcal{O}(n^2 d)$



COMPUTER SCIENCE
VIRGINIA TECH

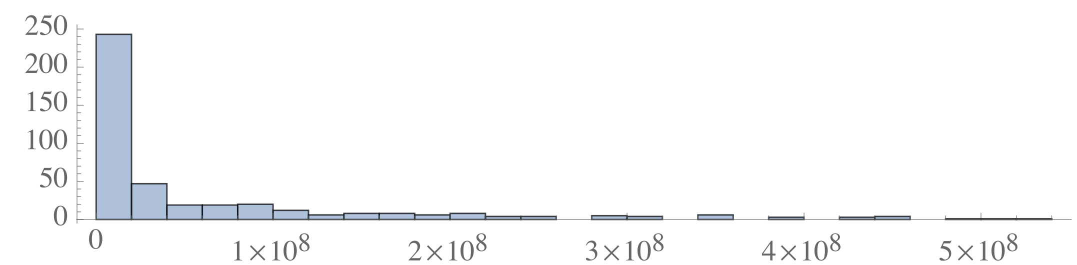
Testing and Evaluation: Data

Testing and Evaluation: Data

High Performance Computing File I/O

$n = 532, d = 4$

predicting *file I/O throughput*

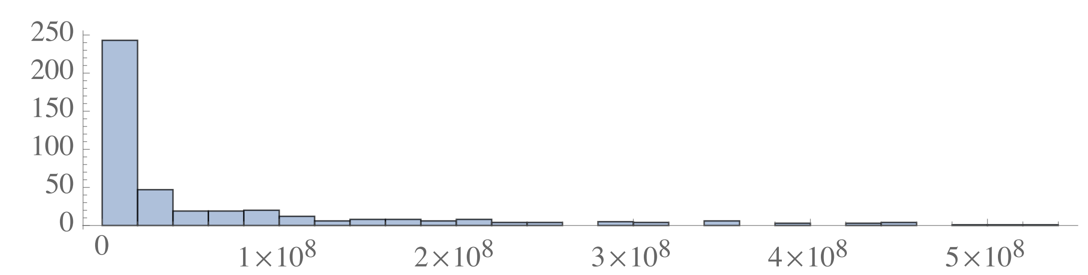


Testing and Evaluation: Data

High Performance Computing File I/O

$n = 532, d = 4$

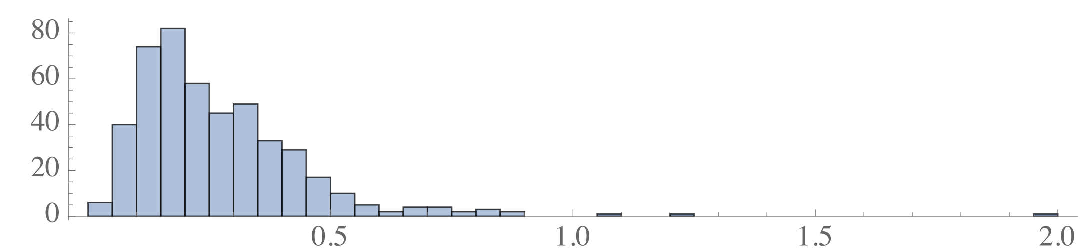
predicting *file I/O throughput*



Forest Fire

$n = 517, d = 12$

predicting *area burned*

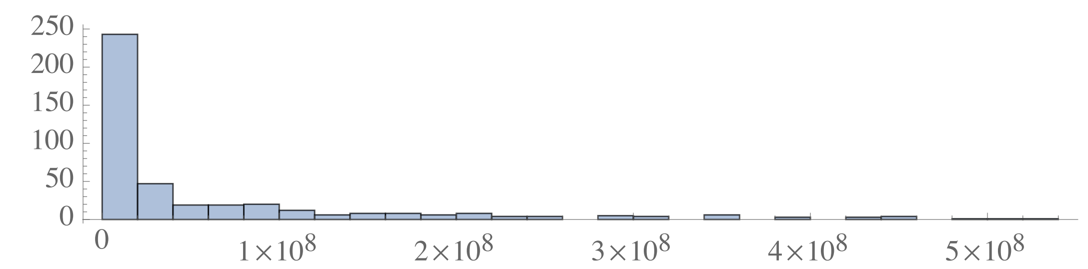


Testing and Evaluation: Data

High Performance Computing File I/O

$n = 532, d = 4$

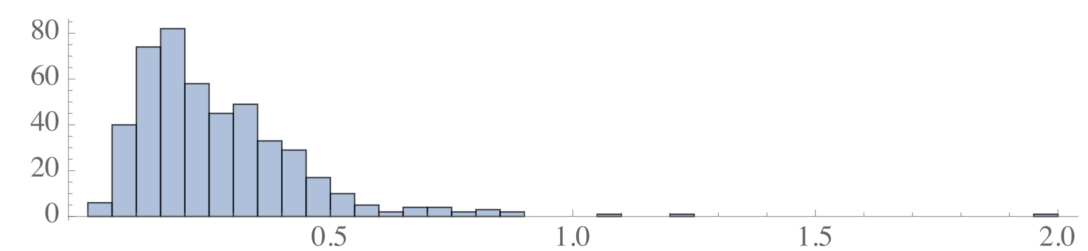
predicting *file I/O throughput*



Forest Fire

$n = 517, d = 12$

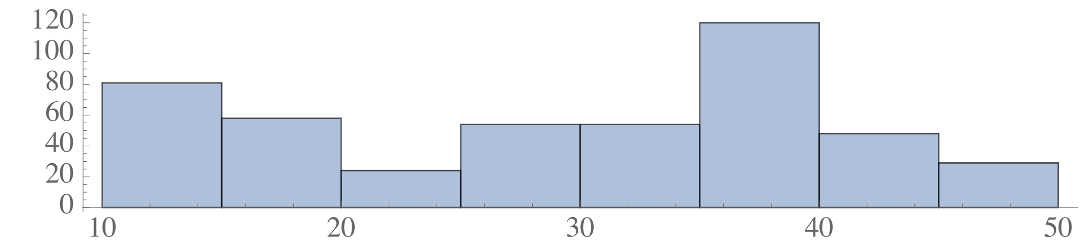
predicting *area burned*



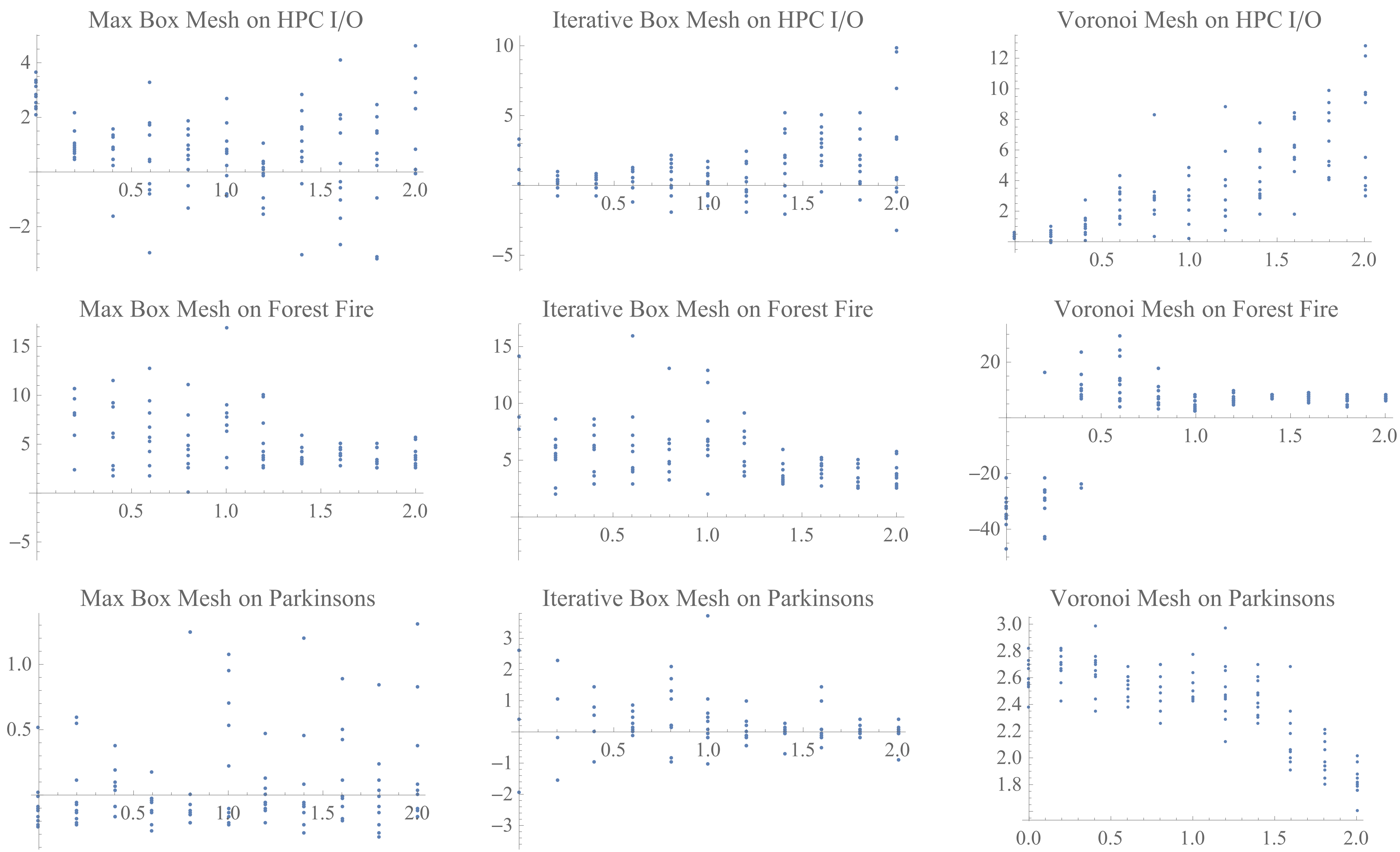
Parkinson's Clinical Evaluation

$n = 468, d = 16$

predicting *total clinical "UPDRS" score*



Average Relative Testing Error (y-axis) versus Relative Error Tolerance (x-axis)



Chapter Takeaways

The “Max” method appears to produce better results than the “Iterative” method. The Voronoi Cell method is best for I/O, but worst for all other tests.

The bootstrapping combined with the least squares computation incurs a lot of computational expense. This methodology **cannot be scaled** to more than 100’s of points.

Chapters

1. The Importance and Applications of Variability
2. Algorithms for Constructing Approximations
3. Naive Approximations of Variability
4. Box-Splines: Uses, Constructions, and Applications
5. Stronger Approximations of Variability
6. An Error Bound for Piecewise Linear Interpolation
7. A Package for Monotone Quintic Spline Interpolation

Interpolating Distributions

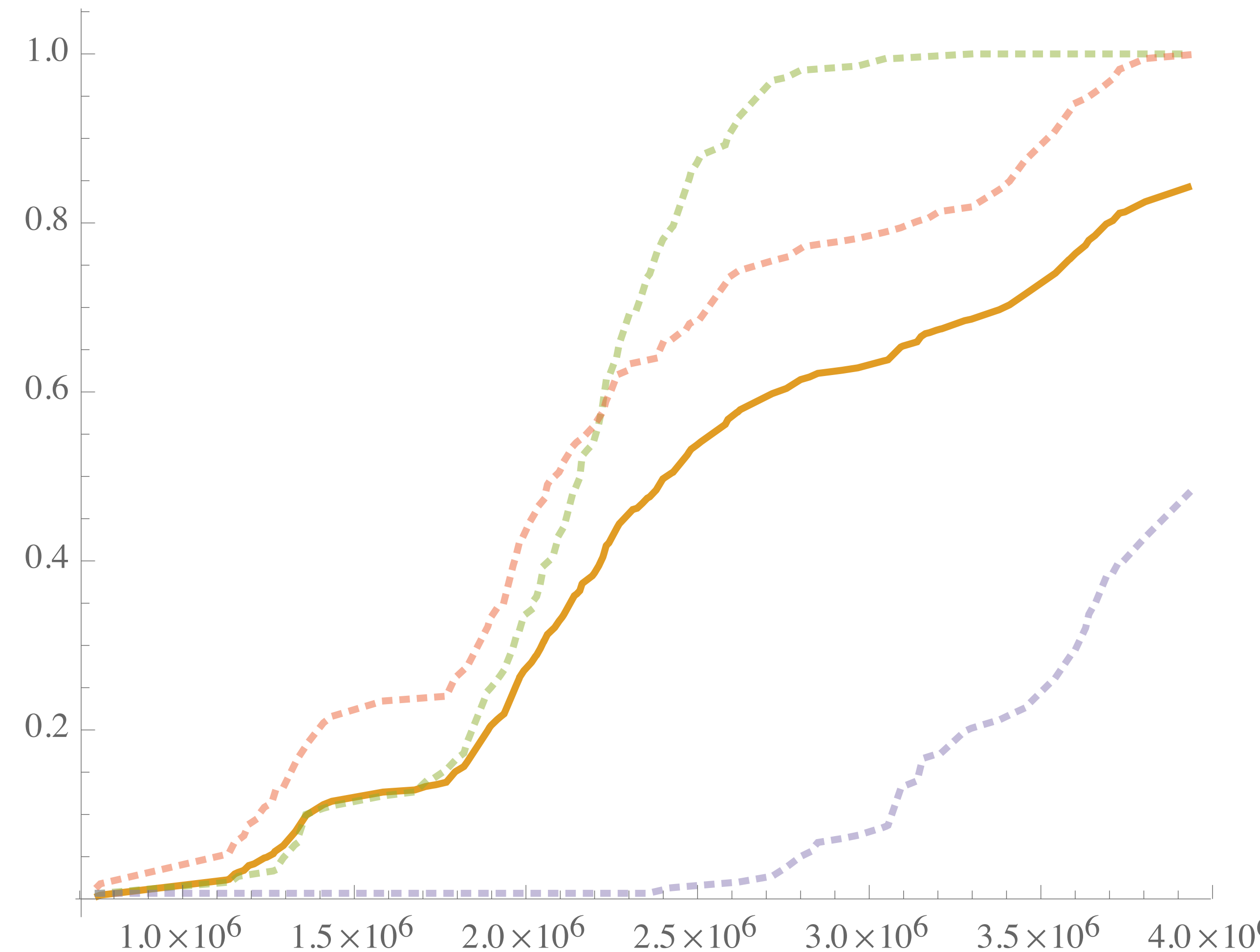
A Cumulative Distribution Function (CDF) $F : \mathbb{R} \rightarrow \mathbb{R}$ must maintain the properties:

$$F(x) \in [0, 1]$$

$F(x)$ is absolutely continuous and nondecreasing.

A convex combination of CDFs results in a valid CDF. Consider this example, solid line is the weighted sum:

{.3 Red, .4 Green, .3 Blue}



Measuring Error in a Prediction

Kolmogorov Smirnov (KS) statistic, max-norm difference.

Null hypothesis (of distributions being same) is rejected at confidence level p according to

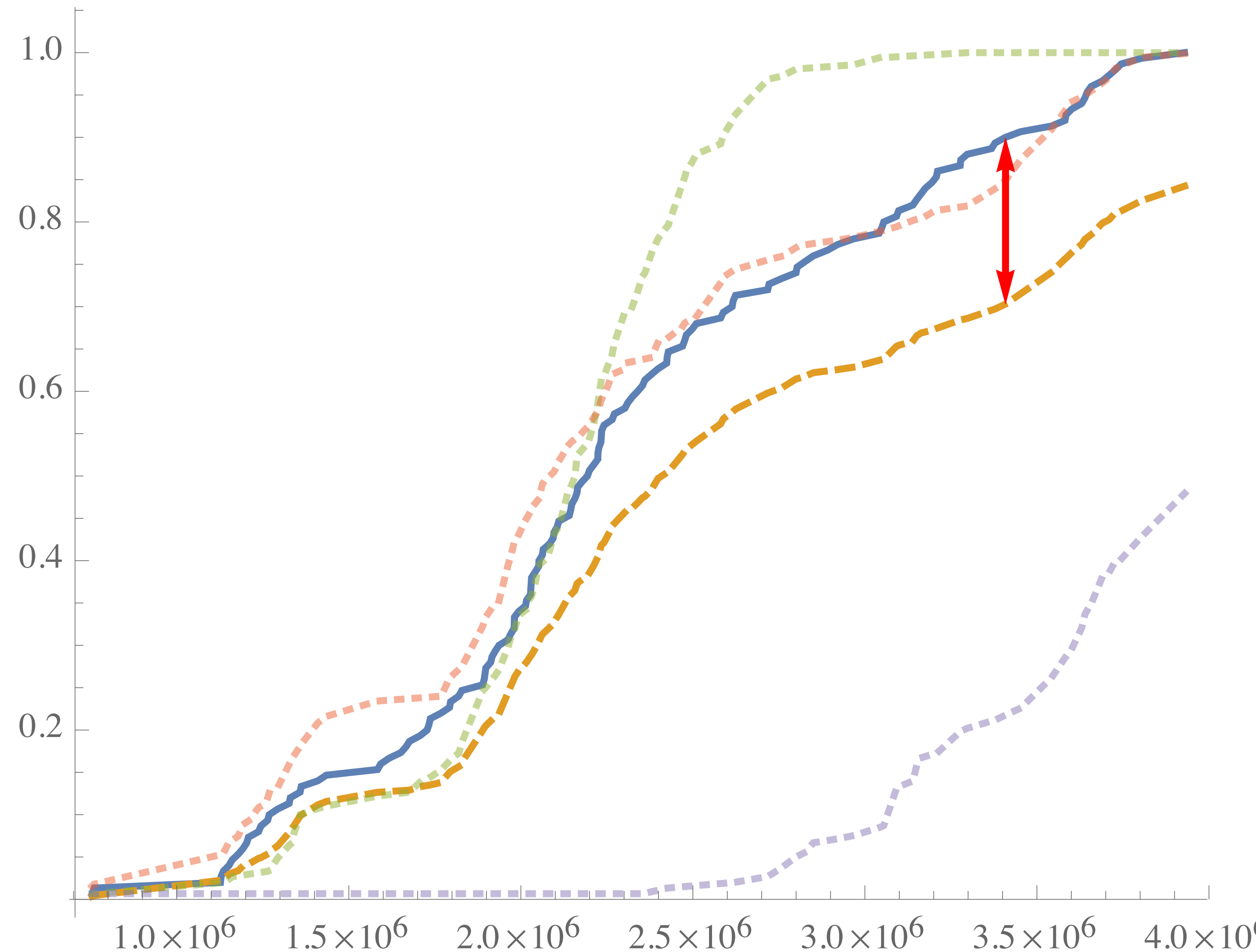
$$KS > \sqrt{-\frac{1}{2}\ln\left(\frac{p}{2}\right)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Dotted lines –
source CDFs

Dashed line –
predicted CDF (Delaunay)

Solid line –
true CDF

Red arrow –
KS statistic between predicted and true (.2)

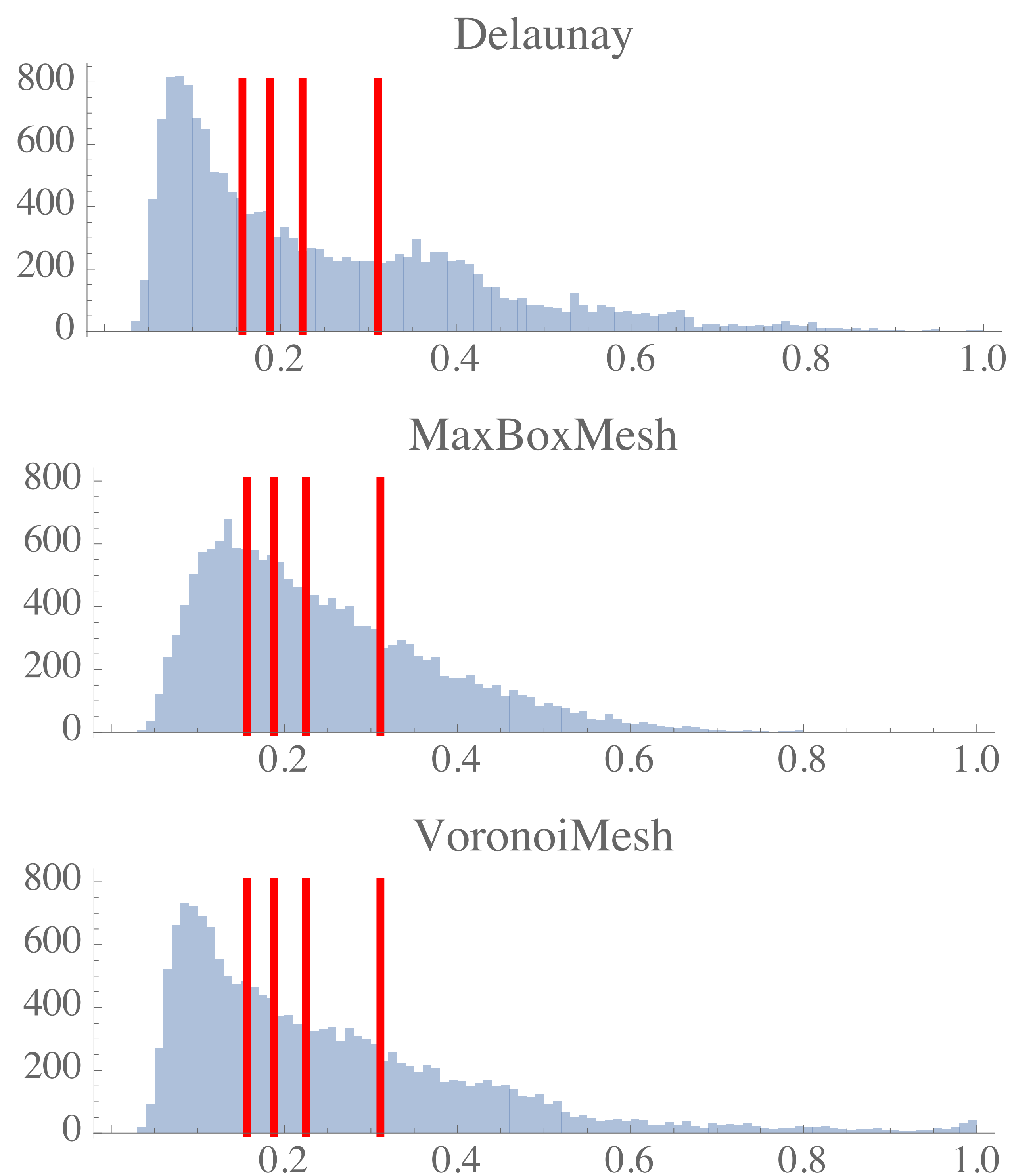


Results Applied to IOzone

x-axis – KS Statistic
y-axis – Number of predictions

Red lines:
KS significance levels at
{.1 .05, .01, .001}

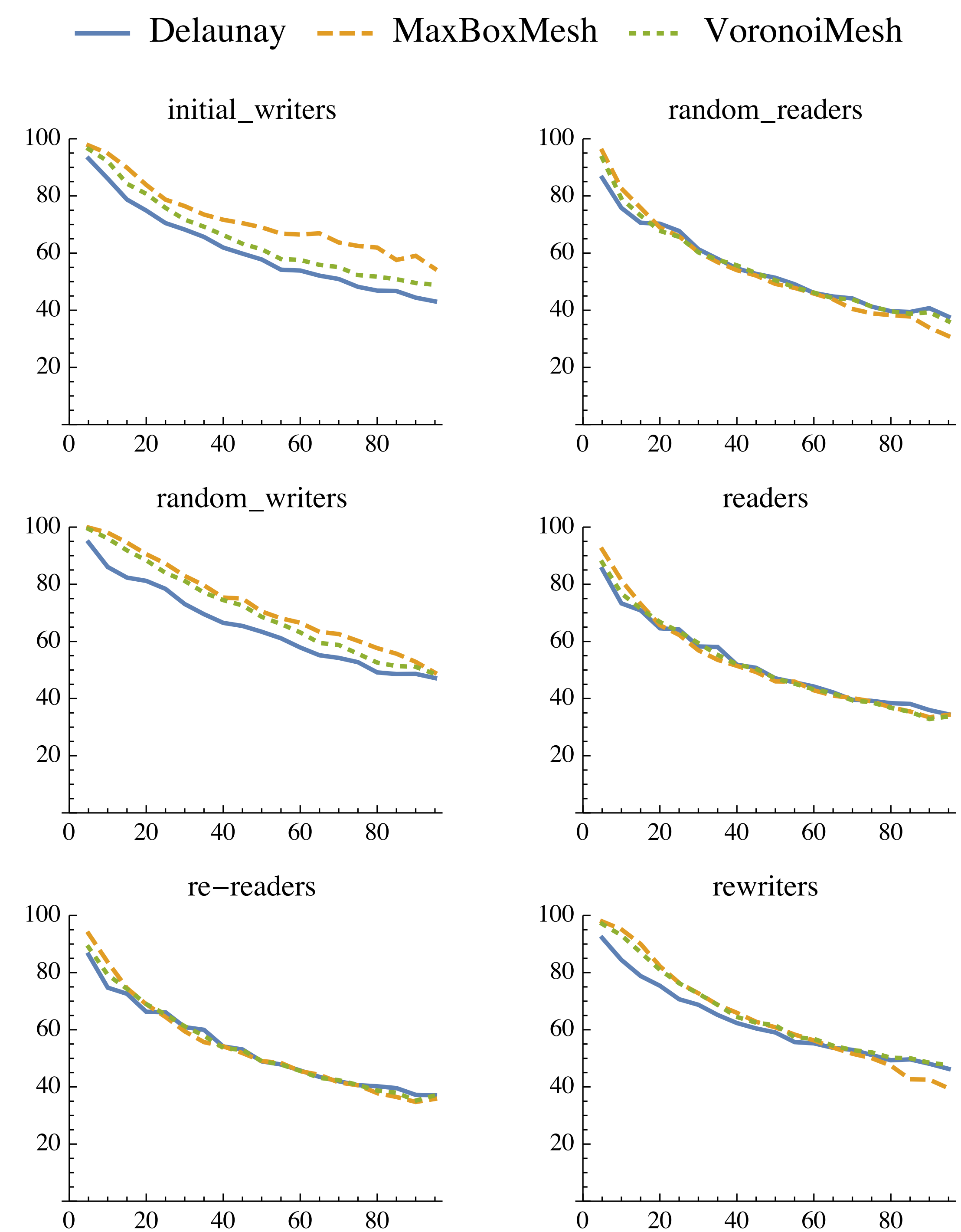
Consider all values to the right of a red line an “incorrect” prediction at that significance.

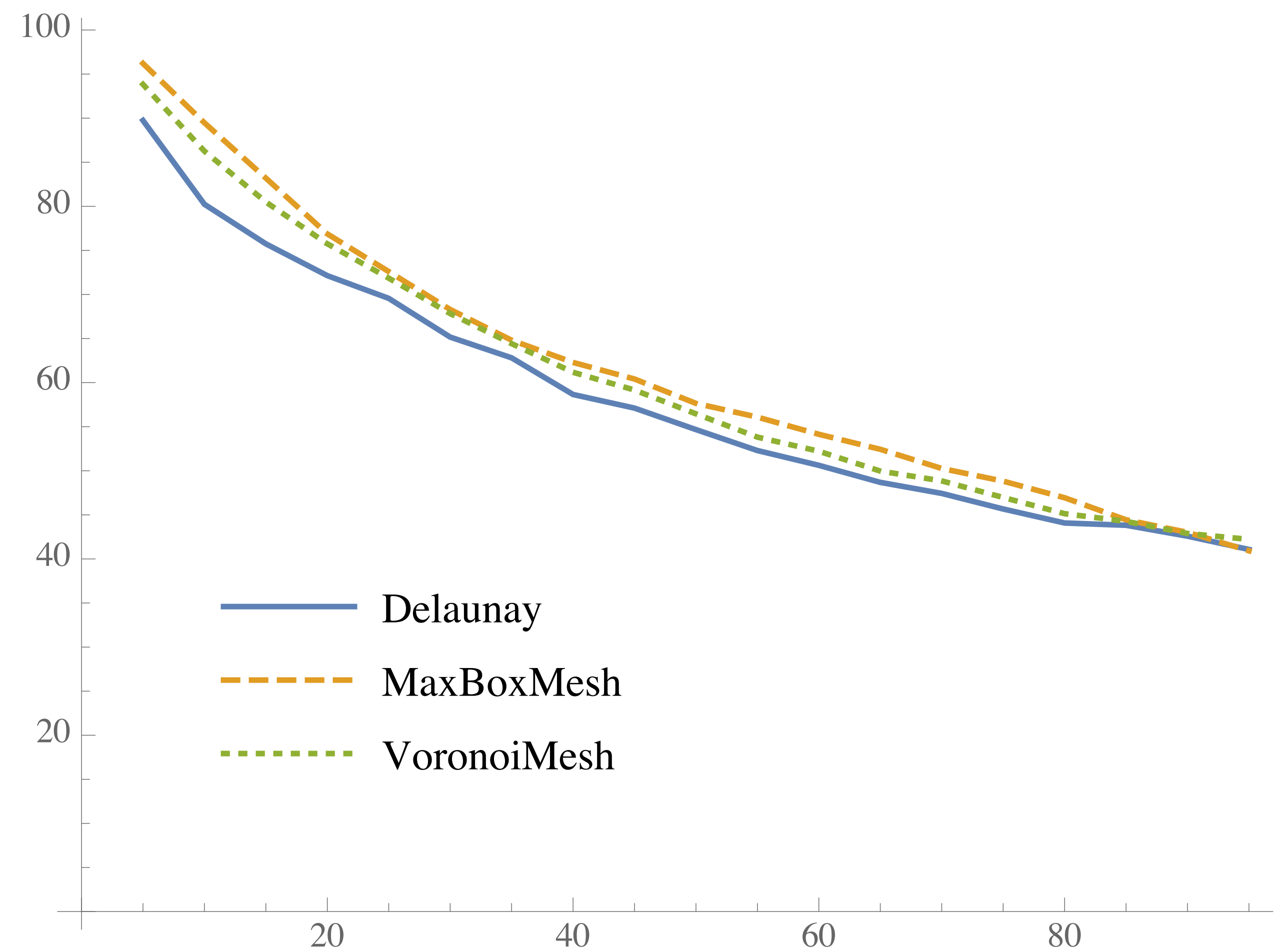


COMPUTER SCIENCE
VIRGINIA TECH.

Increasing Training Data

x-axis – Percentage training data
y-axis – Percentage N.H. rejections
Below: Aggregate
Right: Breakdown by Test





COMPUTER SCIENCE
VIRGINIA TECH.

Improving Performance with *Tuning*

Algorithm	P-Value	Unweighted % N.H. Rejection	Weighted % N.H. Rejection
Delaunay	.05	24.9	30.2
Max Box Mesh		21.3	21.2
Voronoi Mesh		18.7	11.3
Delaunay	.01	21.6	27.4
Max Box Mesh		16.4	16.4
Voronoi Mesh		14.9	7.0
Delaunay	.001	19.7	25.4
Max Box Mesh		13.1	13.1
Voronoi Mesh		12.3	4.6
Delaunay	1.0e-6	17.9	23.4
Max Box Mesh		11.3	11.3
Voronoi Mesh		8.5	2.3

Consensus optimal weighting of (.001, 2, 1.7, 1.5), for frequency, file size, record size, and number of threads. Frequency is unimportant.

Chapter Takeaways

Without any modification, many interpolants can be used to predict distributions! Particularly, those that make predictions with convex combinations of known function values.

Distribution prediction performs well, impressively so with tuning (however the tuning is less provably useful).

20K system configurations appears to approach the limit of distribution prediction accuracy. If we had a better way to approximate distributions, we might reduce error further.

Chapters

1. The Importance and Applications of Variability
2. Algorithms for Constructing Approximations
3. Naive Approximations of Variability
4. Box-Splines: Uses, Constructions, and Applications
5. Stronger Approximations of Variability
6. An Error Bound for Piecewise Linear Interpolation
7. A Package for Monotone Quintic Spline Interpolation

The Theory

$$\left|f(z) - \hat{f}(z)\right| \leq \frac{\gamma \|z - x_0\|_2^2}{2} + \frac{\sqrt{d}\,\gamma k^2}{2\sigma_d} \|z - x_0\|_2$$

The Theory

$$\left|f(z) - \hat{f}(z)\right| \leq \frac{\gamma \|z - x_0\|_2^2}{2} + \frac{\sqrt{d}\,\gamma k^2}{2\sigma_d}\|z - x_0\|_2$$

The absolute error of a linear interpolant is tightly upper bounded by

COMPUTER SCIENCE
VIRGINIA TECH

The Theory

$$\left|f(z) - \hat{f}(z)\right| \leq \frac{\gamma \|z - x_0\|_2^2}{2} + \frac{\sqrt{d}\,\gamma k^2}{2\sigma_d}\|z - x_0\|_2$$

The absolute error of a linear interpolant is tightly upper bounded by

the max change in slope of the function

The Theory

$$\left|f(z) - \hat{f}(z)\right| \leq \frac{\gamma \|z - x_0\|_2^2}{2} + \frac{\sqrt{d}\,\gamma k^2}{2\sigma_d}\|z - x_0\|_2$$

The absolute error of a linear interpolant is tightly upper bounded by

the max change in slope of the function

times the distance to the nearest known point squared

The Theory

$$\left|f(z) - \hat{f}(z)\right| \leq \frac{\gamma\|z - x_0\|_2^2}{2} + \frac{\sqrt{d}\,\gamma k^2}{2\sigma_d}\|z - x_0\|_2$$

The absolute error of a linear interpolant is tightly upper bounded by

the max change in slope of the function

times the distance to the nearest known point squared

plus the square root of the dimension times the max change in slope

The Theory

$$\left|f(z) - \hat{f}(z)\right| \leq \frac{\gamma \|z - x_0\|_2^2}{2} + \frac{\sqrt{d}\,\gamma k^2}{2\sigma_d}\|z - x_0\|_2$$

The absolute error of a linear interpolant is tightly upper bounded by

the max change in slope of the function

times the distance to the nearest known point squared

plus the square root of the dimension times the max change in slope

times the longest edge length between points defining the linear interpolant squared

The Theory

$$\left|f(z) - \hat{f}(z)\right| \leq \frac{\gamma \|z - x_0\|_2^2}{2} + \frac{\sqrt{d}\,\gamma k^2}{2\sigma_d} \|z - x_0\|_2$$

The absolute error of a linear interpolant is tightly upper bounded by

the max change in slope of the function

times the distance to the nearest known point squared

plus the square root of the dimension times the max change in slope

times the longest edge length between points defining the linear interpolant squared

divided by how close the interpolated points are to being planar.

The Importance

The approximation error of a linear (simplicial) interpolant tends quadratically towards zero when approaching observed data only when the diameter of the simplex is also reduced proportionally.

In practice, only linear convergence to the true function can be achieved (because the evaluation points don't move).

Approximation error is largely determined by **data spacing**!

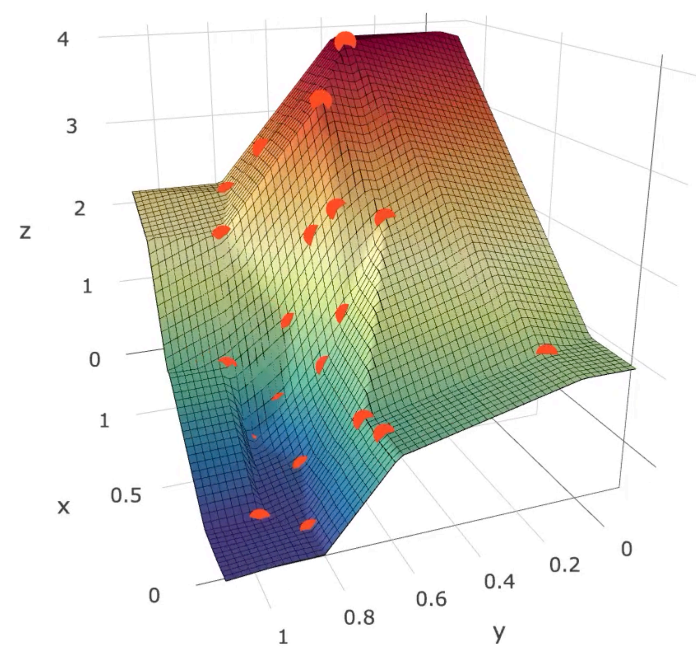
This theory only directly applies to Delaunay, but may give insight into the approximation behavior of other techniques.

Piecewise Linear Approximations

Delaunay
(interpolant)

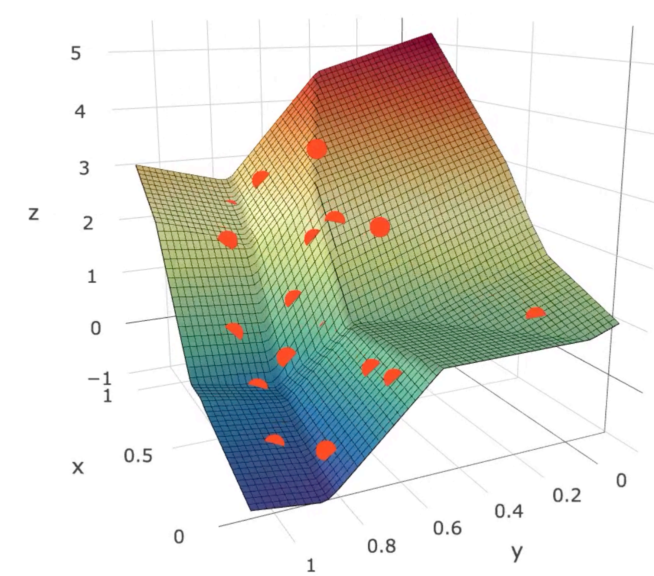
$$y = \sum_{i=0}^{d} w_i x^{(i)}, \quad \sum_{i=0}^{d} w_i = 1, \quad w_i \geq 0, \quad i = 0, \ldots, d$$

$$\hat{f}(y) = \sum_{i=0}^{d} w_i f\left(x^{(i)}\right)$$



Multilayer Perceptron
(regressor)

$$l(u) = \left(u^t W_l + c_l\right)_+$$

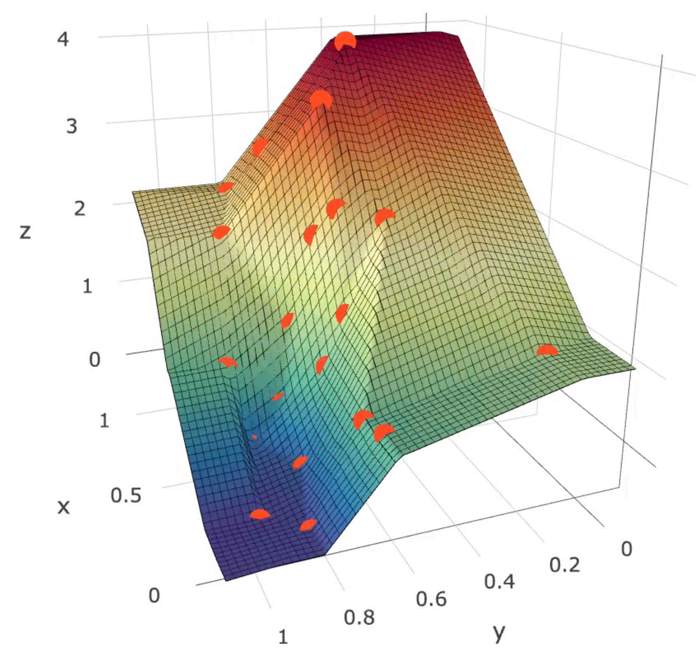


Piecewise Linear Approximations

Delaunay
(interpolant)

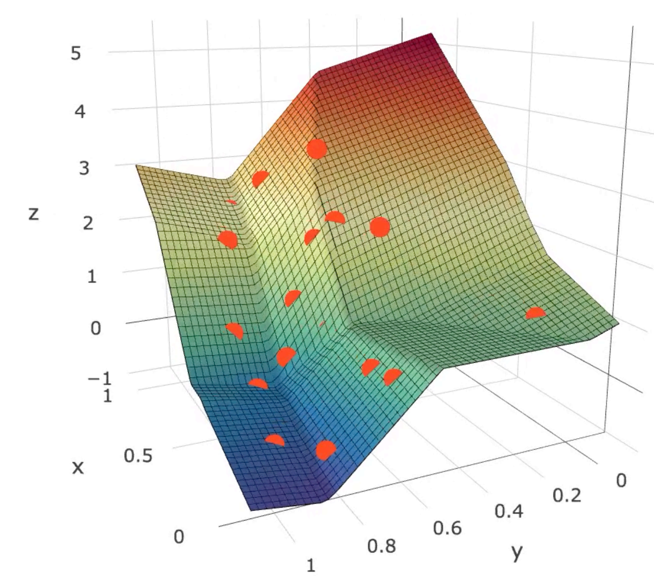
$$y = \sum_{i=0}^{d} w_i x^{(i)}, \quad \sum_{i=0}^{d} w_i = 1, \quad w_i \geq 0, \quad i = 0, \ldots, d$$

$$\hat{f}(y) = \sum_{i=0}^{d} w_i f\left(x^{(i)}\right)$$



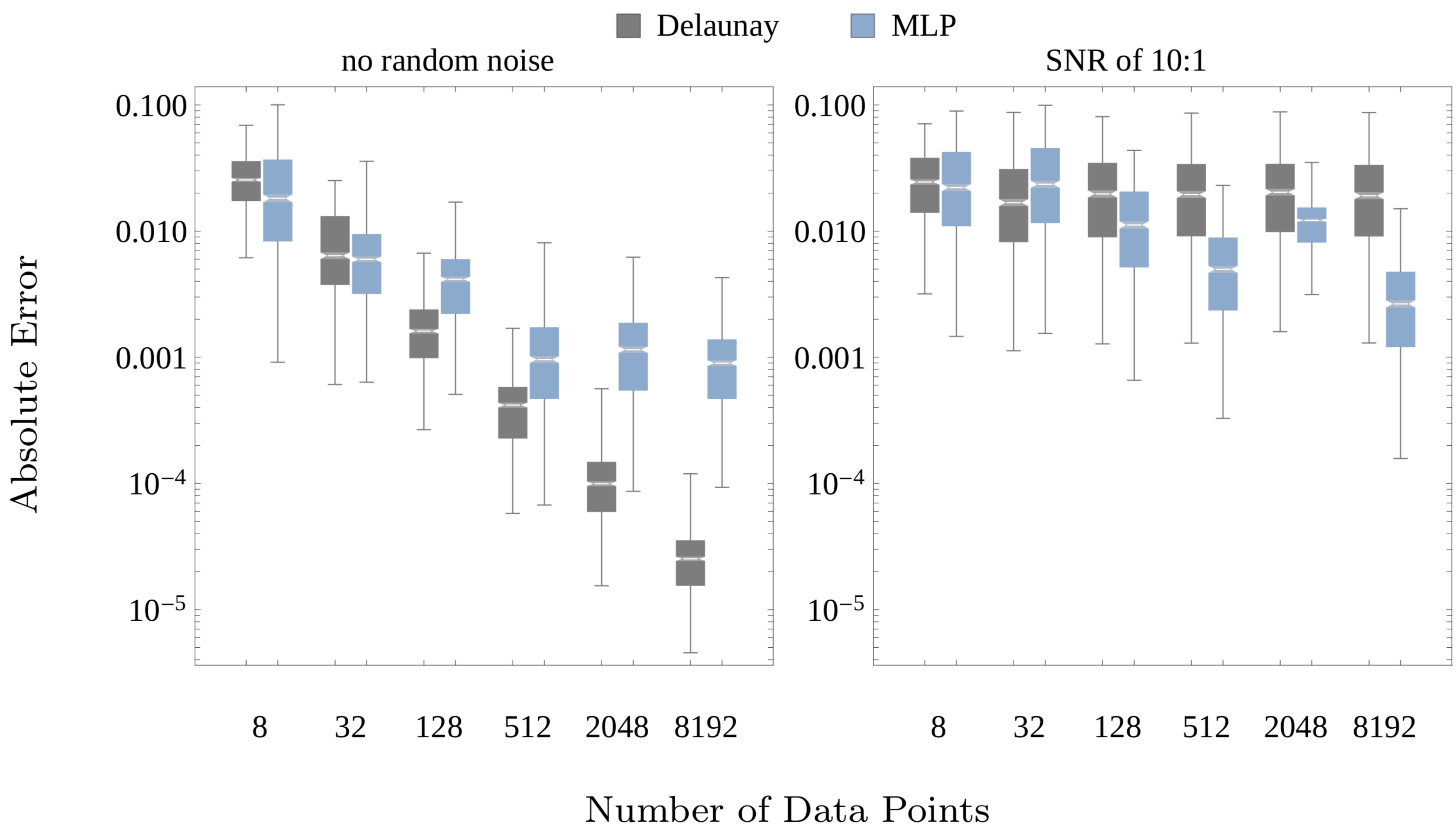
Multilayer Perceptron
(regressor)

$$l(u) = \left(u^t W_l + c_l\right)_+$$



Approximating $f(x) = \cos(\|x\|_2)$

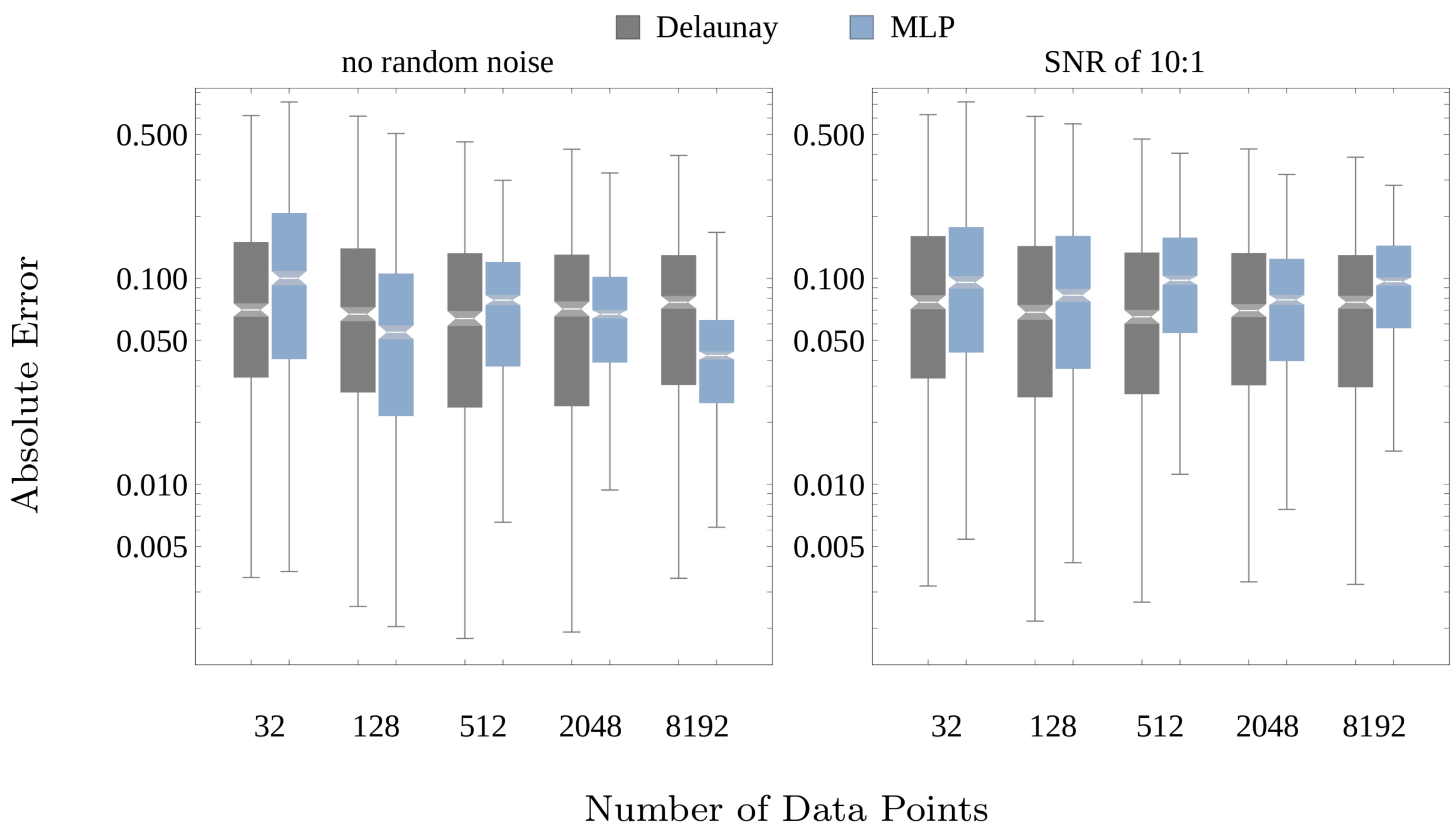
In 2 dimensions, we get expected results.
Delaunay is better at interpolation, MLP better at regression.



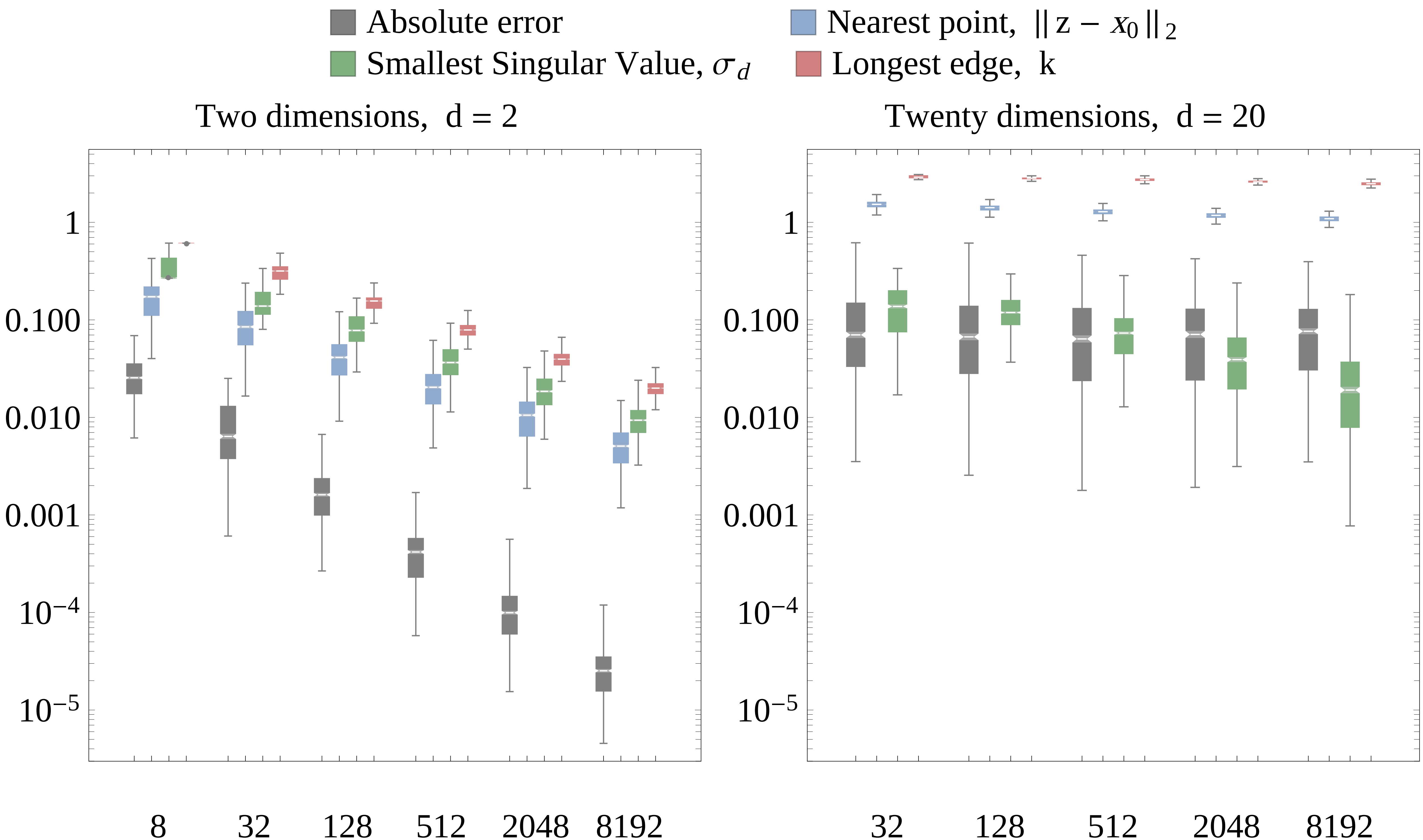
COMPUTER SCIENCE
VIRGINIA TECH

Approximating $f(x) = \cos(\|x\|_2)$

In 20 dimensions, the *intuitive* trend disappears!
Delaunay and MLP look the same.



Explaining the Convergence



COMPUTER SCIENCE
VIRGINIA TECH.

Connecting Back to Theory

$$\left|f(z) - \hat{f}(z)\right| \leq \frac{\gamma \|z - x_0\|_2^2}{2} + \frac{\sqrt{d}\,\gamma k^2}{2\sigma_d}\|z - x_0\|_2$$

Connecting Back to Theory

$$\left|f(z) - \hat{f}(z)\right| \leq \frac{\gamma\|z - x_0\|_2^2}{2} + \frac{\sqrt{d}\,\gamma k^2}{2\sigma_d}\|z - x_0\|_2$$

the average closest
point doesn't get
much closer and …

Connecting Back to Theory

$$\left|f(z) - \hat{f}(z)\right| \leq \frac{\gamma \|z - x_0\|_2^2}{2} + \frac{\sqrt{d}\,\gamma k^2}{2\sigma_d}\|z - x_0\|_2$$

the average closest point doesn't get much closer and …

the longest edge does not meaningfully shrink in 20 dimensions with thousands of points

Data Sets for Empirical Evaluation

Forest Fire ($n = 504, d = 12$)
given meteorological information about a park, predict the amount of land that would be burned in a forest fire.

Parkinson's Telemonitoring ($n = 5875, d = 19$)
given features of audio recorded in the home of someone with Parkinson's, predict their next clinical evaluation score.

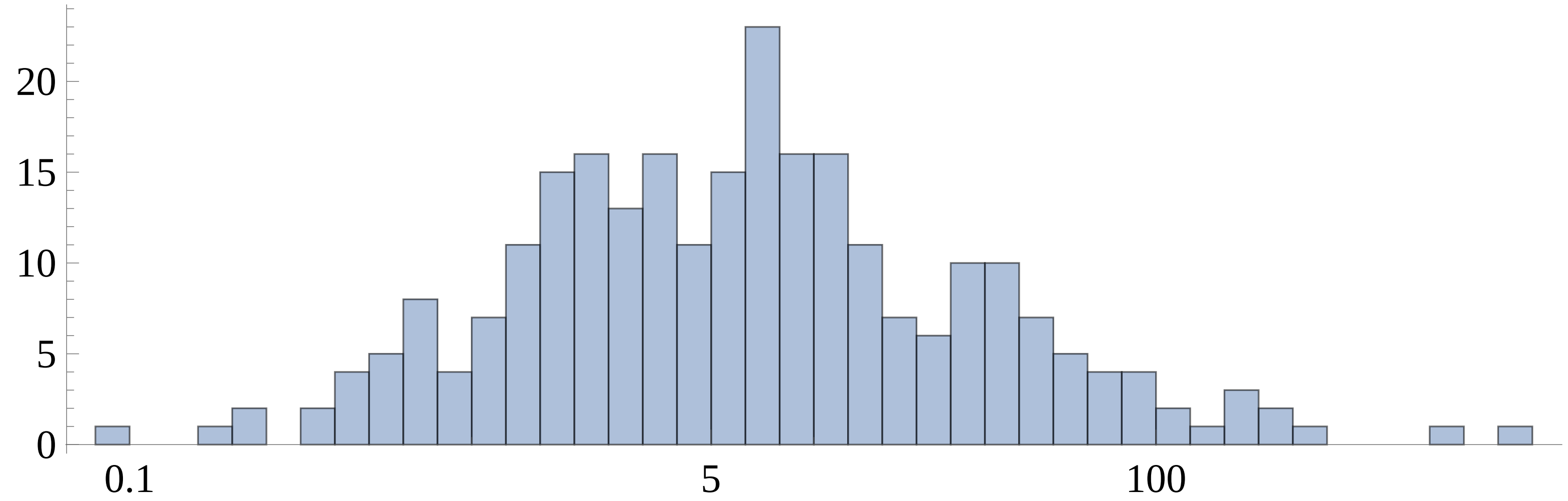
Australian Daily Rainfall Volume ($n = 2609, d = 23$)
given meteorological data around Sydney, Australia, predict the amount of rainfall that will occur on the next day.

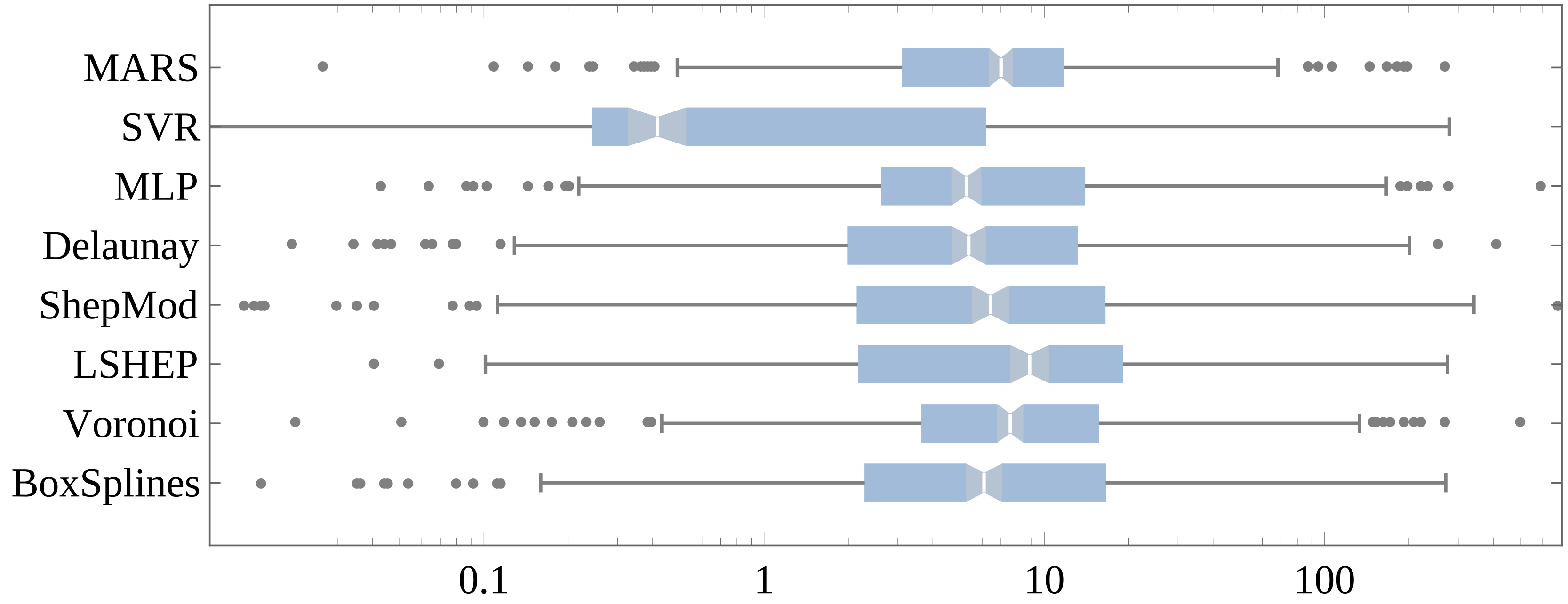
Credit Card Transaction Amount ($n = 5562, d = 28$)
given anonymized electronic transaction features (output of PCA) predict the amount of money that the transaction will process.

High Performance Computing I/O ($n = 3016, d = 4$)
given system configuration information, predict the distribution of I/O throughput that will be seen at a new configuration.

Empirical Results for Each Data Set

Forest fire data

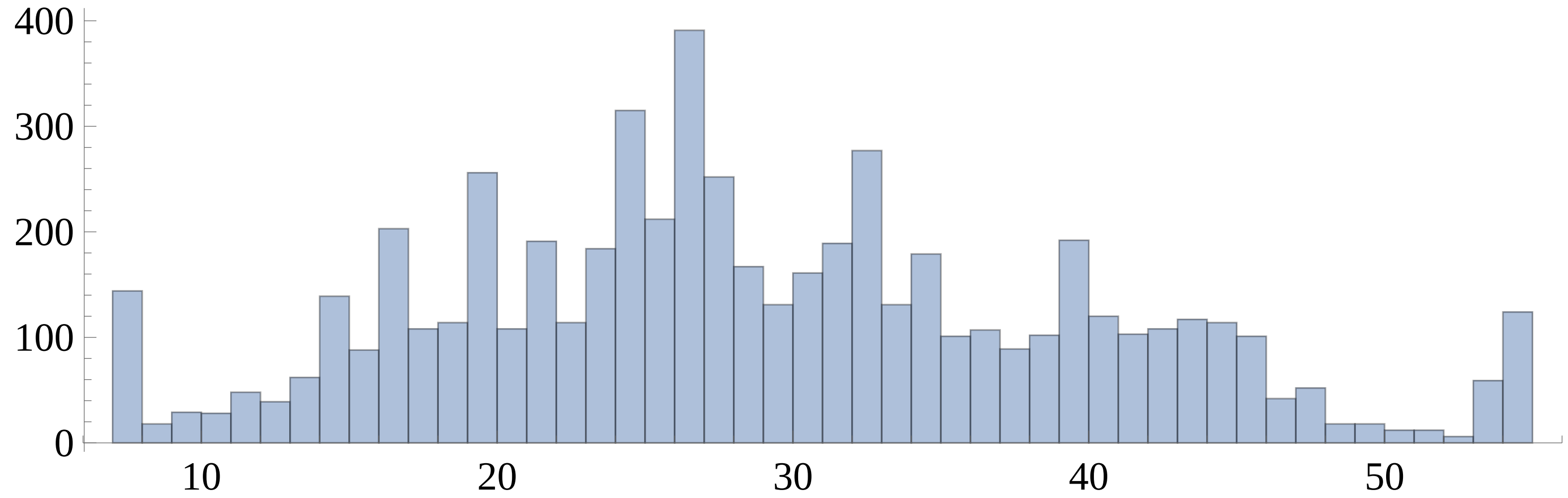


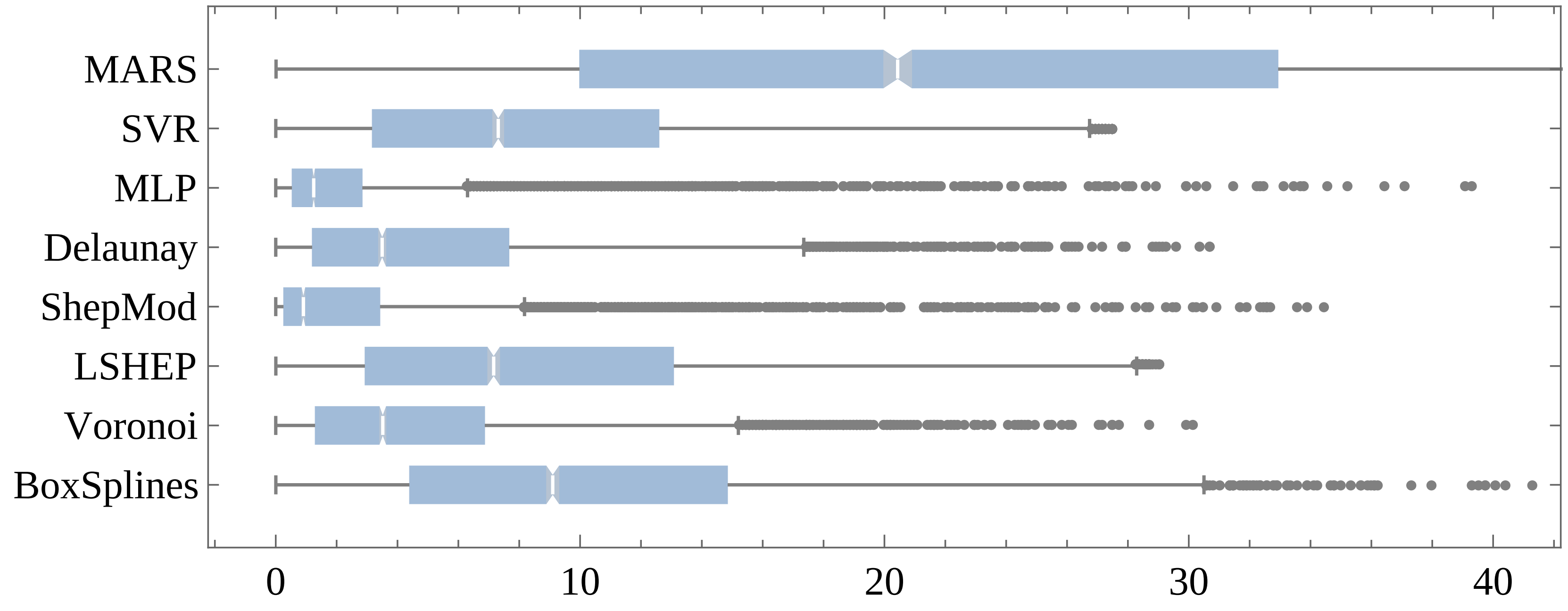


COMPUTER SCIENCE
VIRGINIA TECH.

Empirical Results for Each Data Set

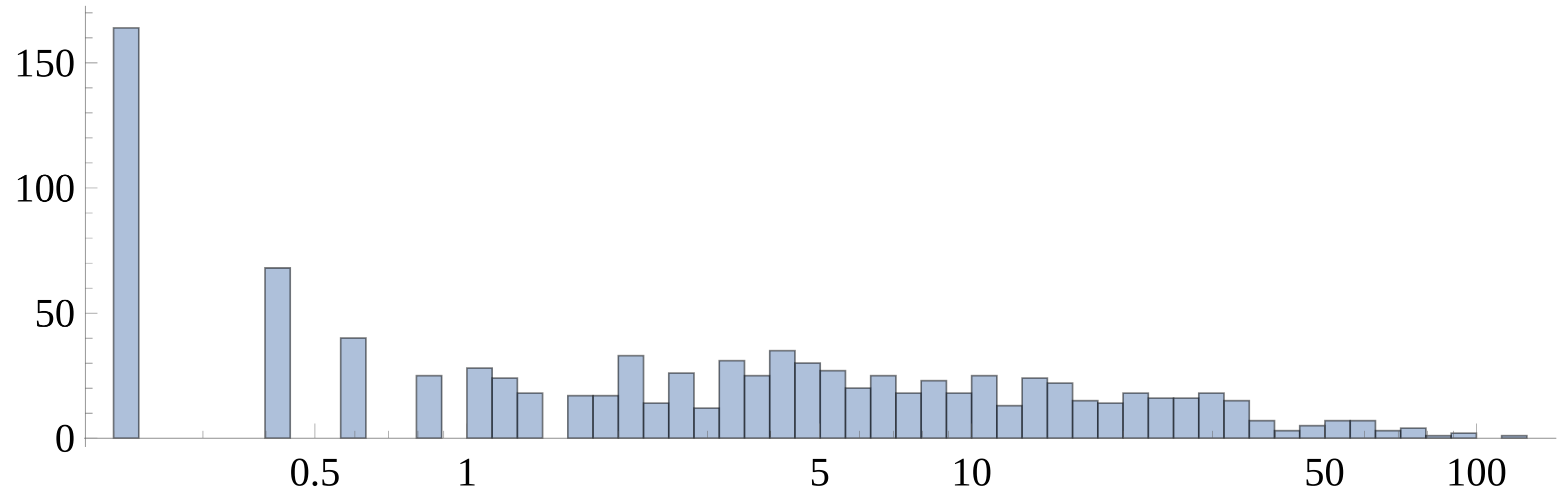
Parkinson's Telimonitoring

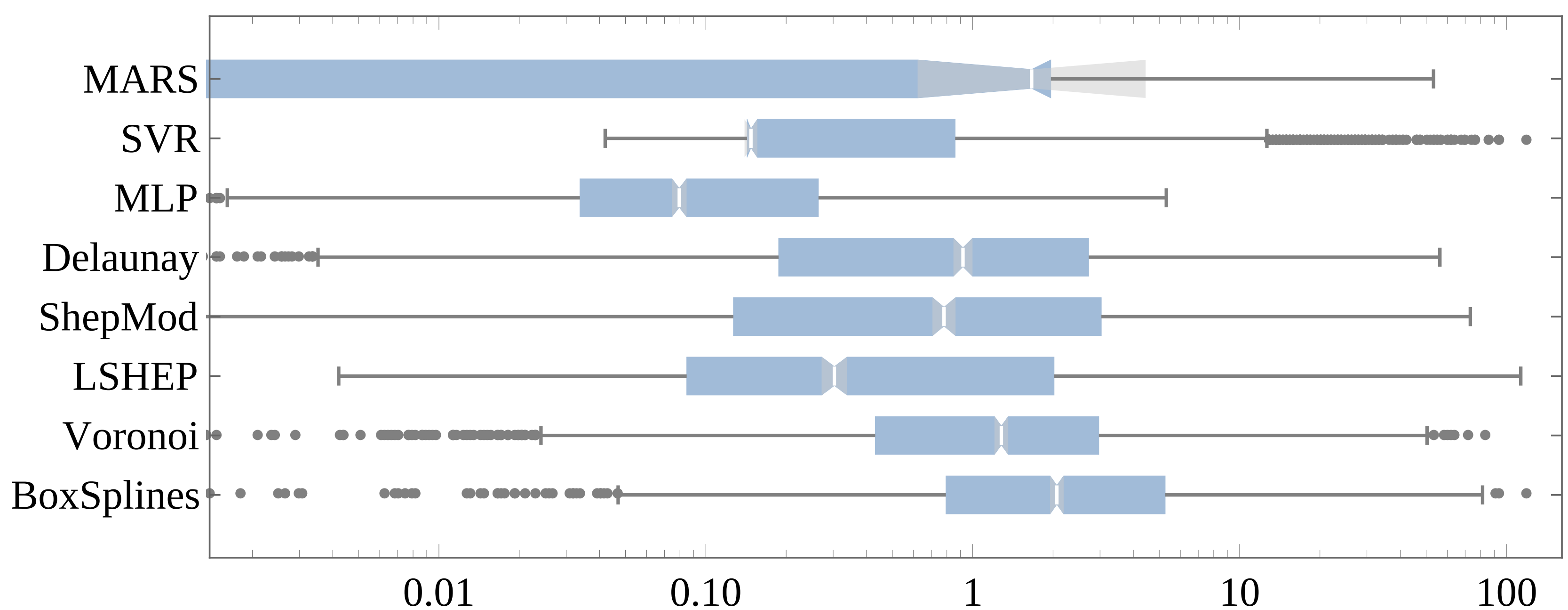




Empirical Results for Each Data Set

Australian Rainfall

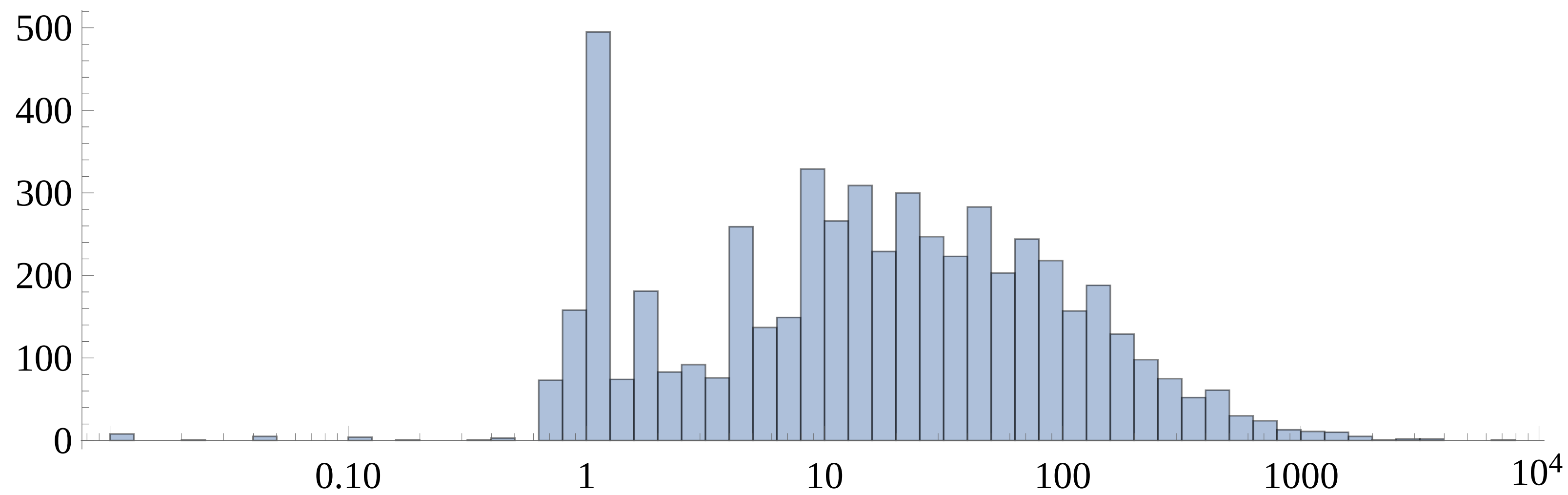


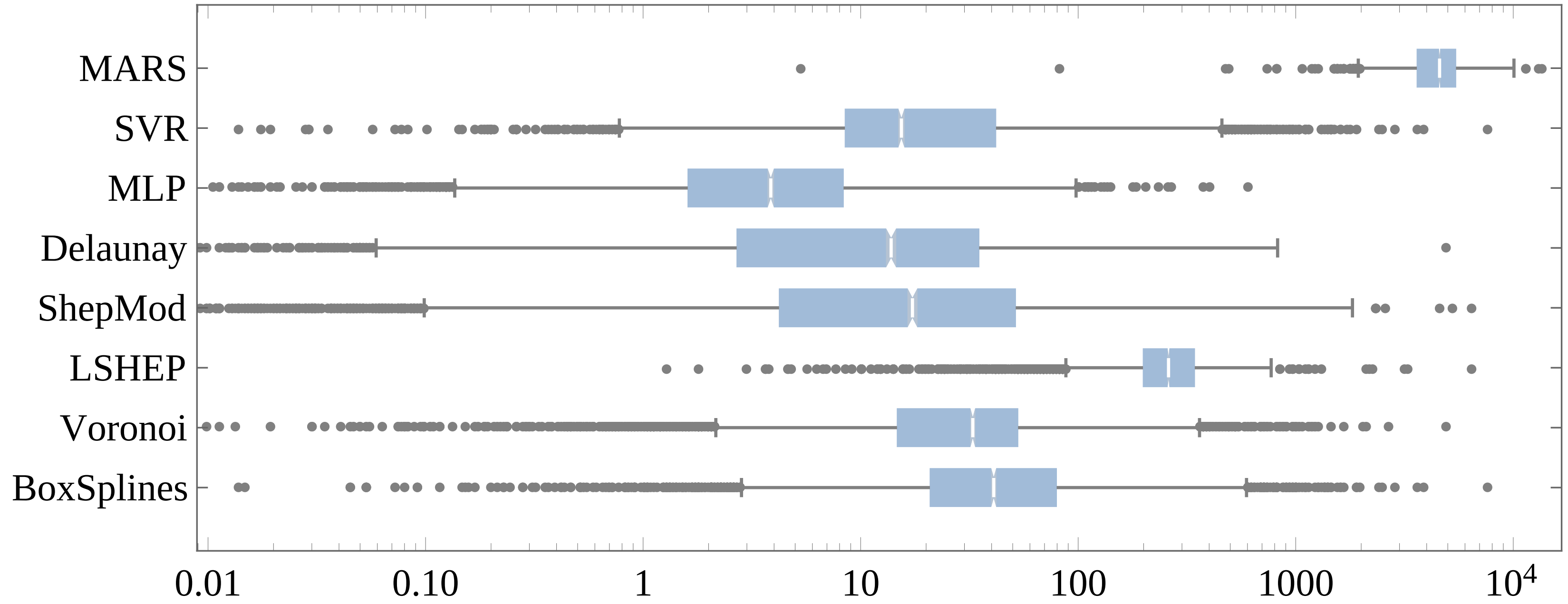


COMPUTER SCIENCE
VIRGINIA TECH

Empirical Results for Each Data Set

Credit Card Transactions

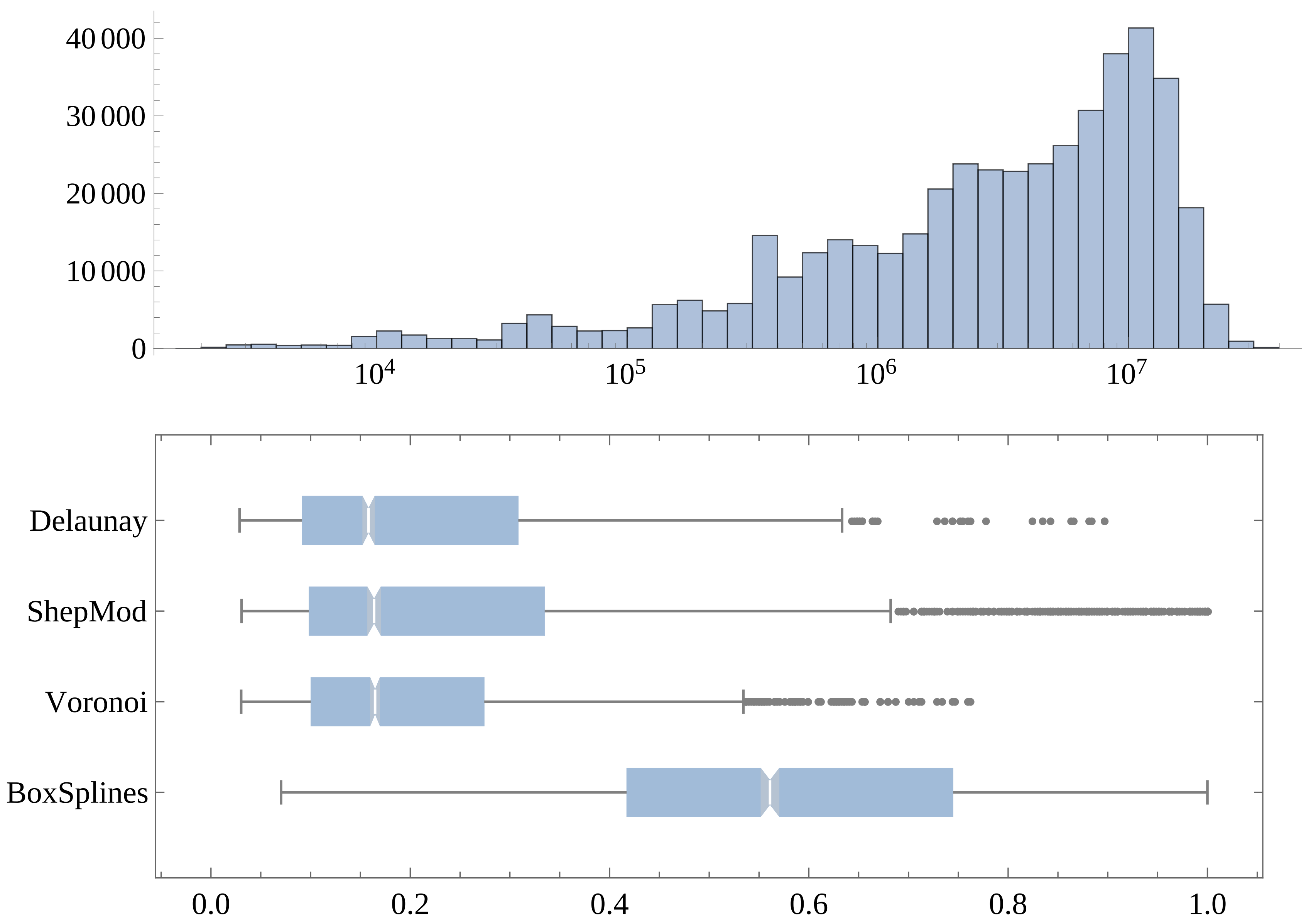




COMPUTER SCIENCE
VIRGINIA TECH.

Empirical Results for Each Data Set

IOzone Distribution Models



Chapter Takeaways

Interpolants have provable convergence properties.

Interpolants produce competitive approximations in medium dimension, while requiring very little "fit" time.

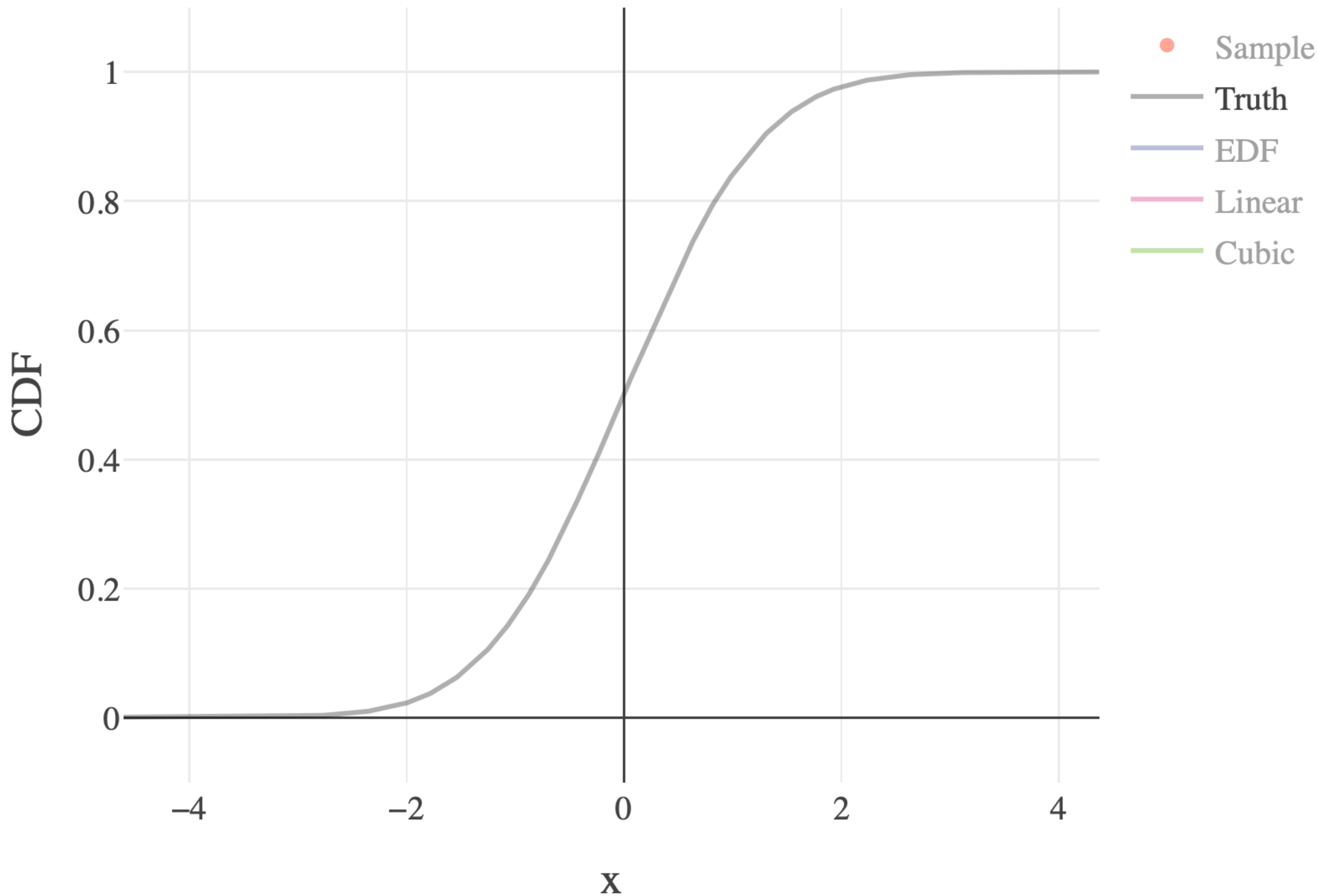
Some interpolants easily generalize to predicting functions.

Algorithm	Avg. % Best	Avg. Fit or Prep. Time (s)	Avg. App. Time (s)
MARS	4.5	20.0s	0.001s
SVR	*19.5*	**0.5**s	**0.0001**s
MLP	**43.1**	200.0s	0.001s
Delaunay	5.2	1.0s	1.0s
ShepMod	18.0	*0.7*s	**0.0001**s
LSHEP	8.4	2.0s	**0.0001**s
Voronoi	0.5	1.0s	0.04s
BoxSplines	3.5	0.8s	*0.0005*s

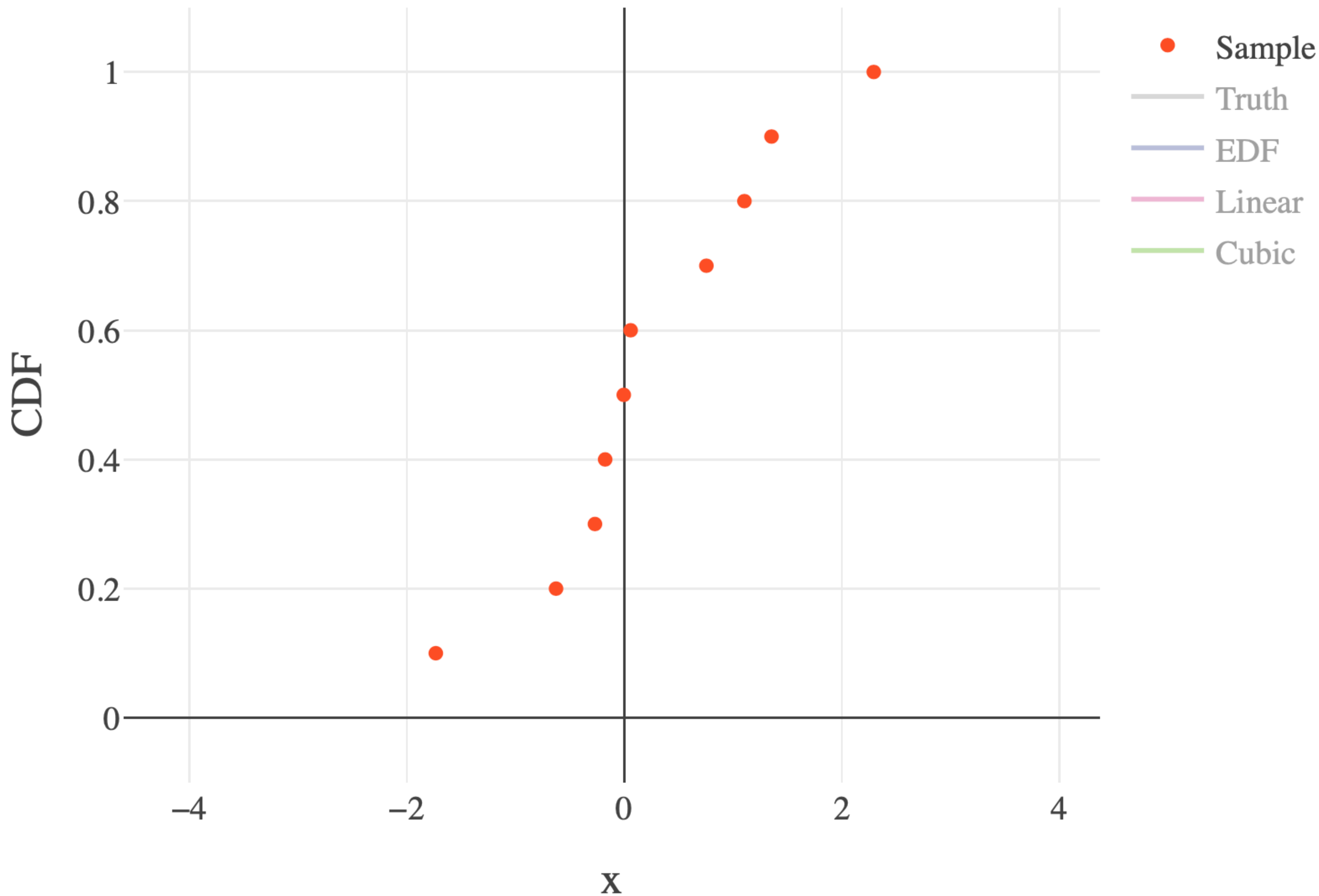
Chapters

1. The Importance and Applications of Variability
2. Algorithms for Constructing Approximations
3. Naive Approximations of Variability
4. Box-Splines: Uses, Constructions, and Applications
5. Stronger Approximations of Variability
6. An Error Bound for Piecewise Linear Interpolation
7. A Package for Monotone Quintic Spline Interpolation

Empirical Distribution Approximations

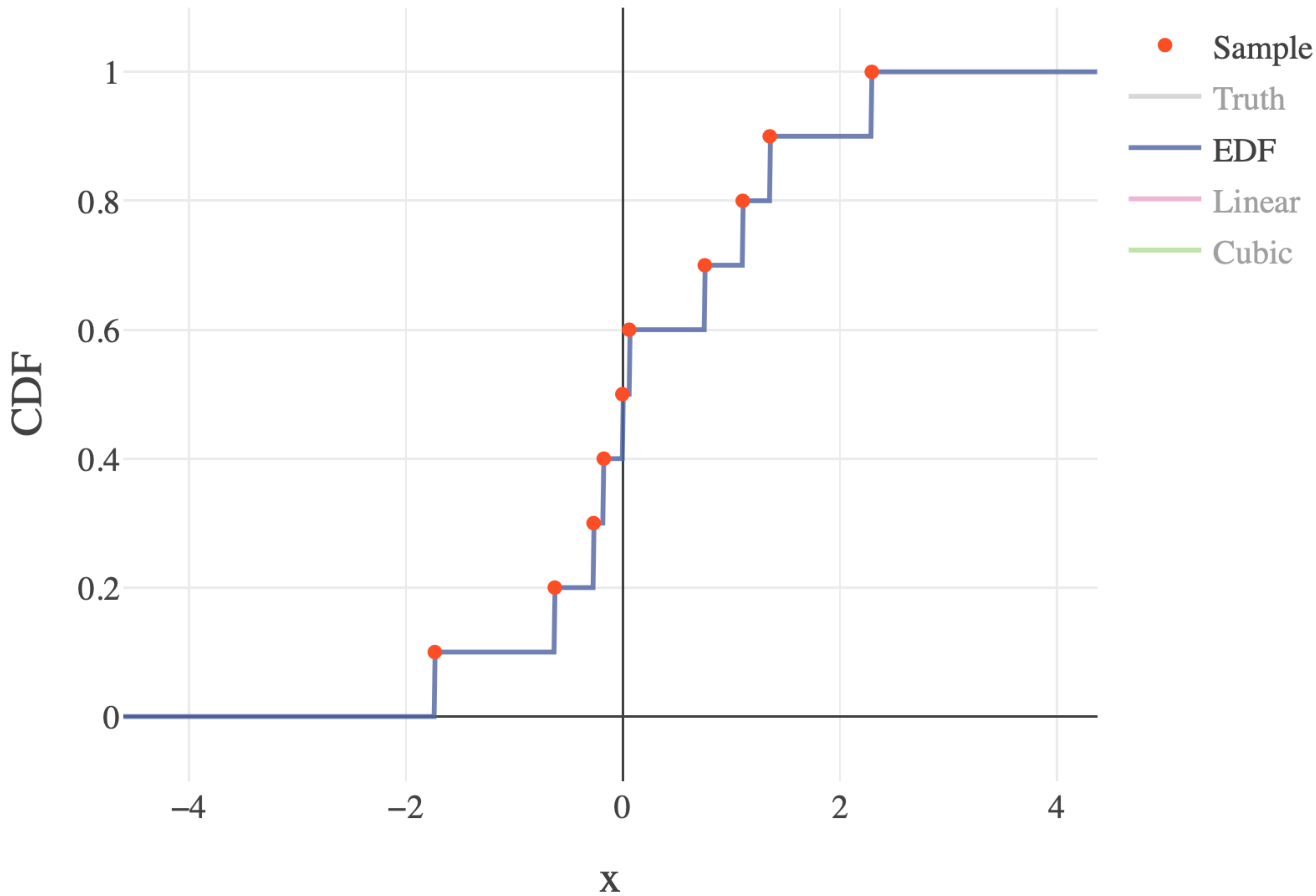


Empirical Distribution Approximations

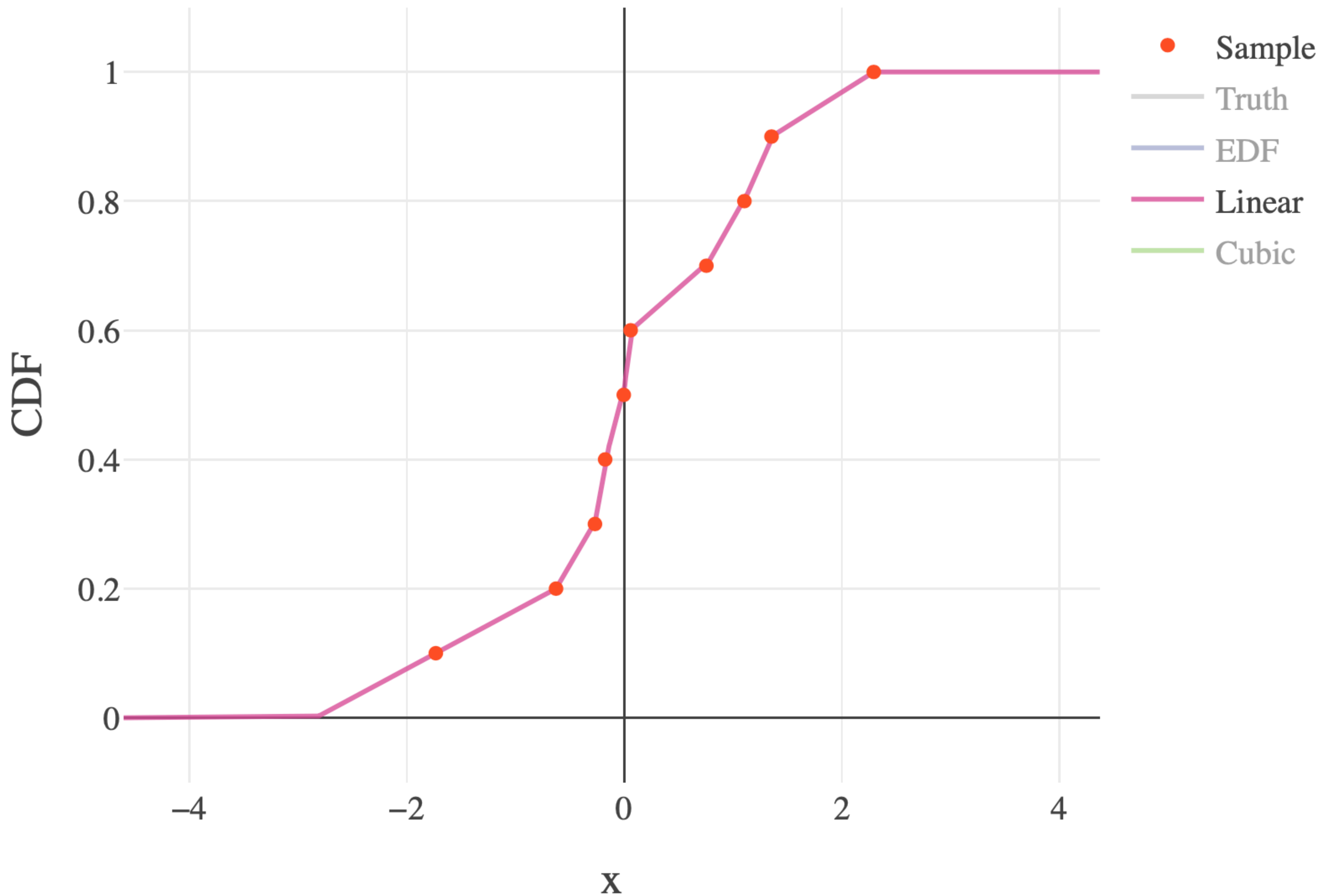


COMPUTER SCIENCE
VIRGINIA TECH.

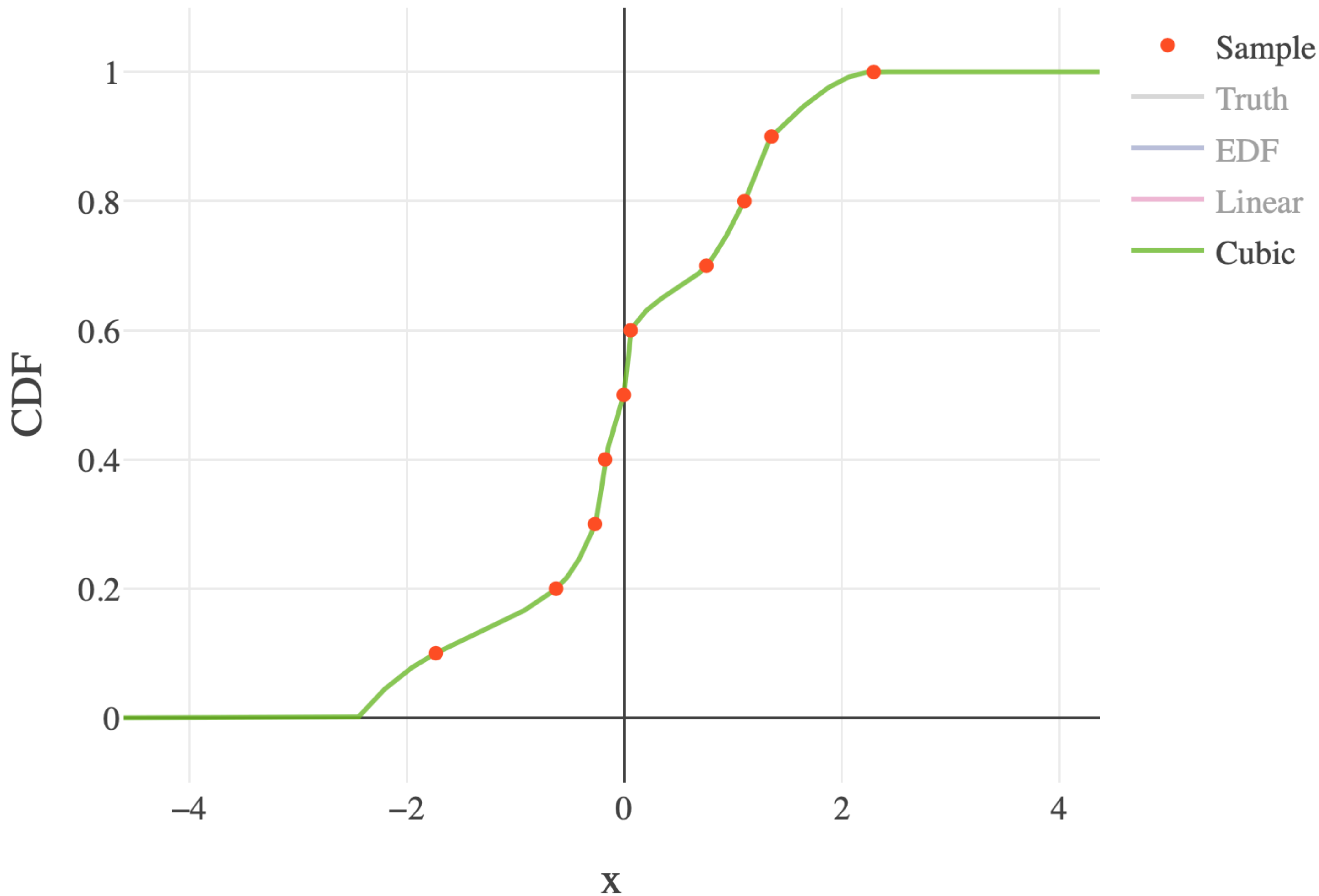
Empirical Distribution Approximations



Empirical Distribution Approximations

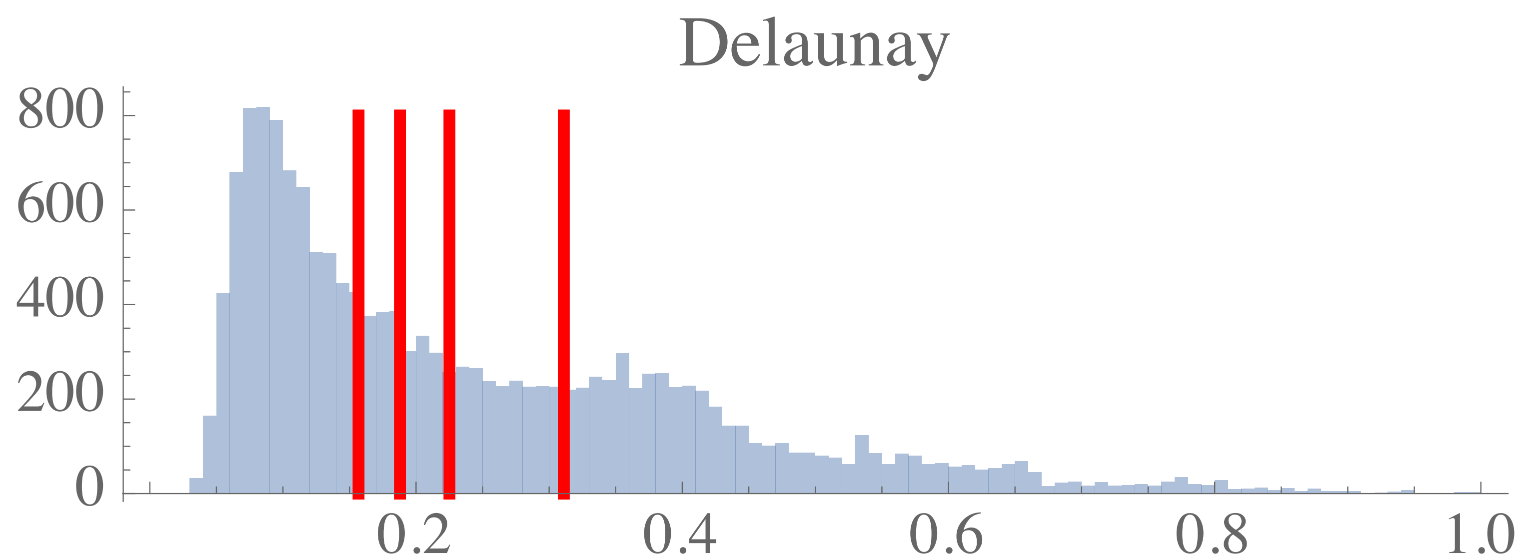


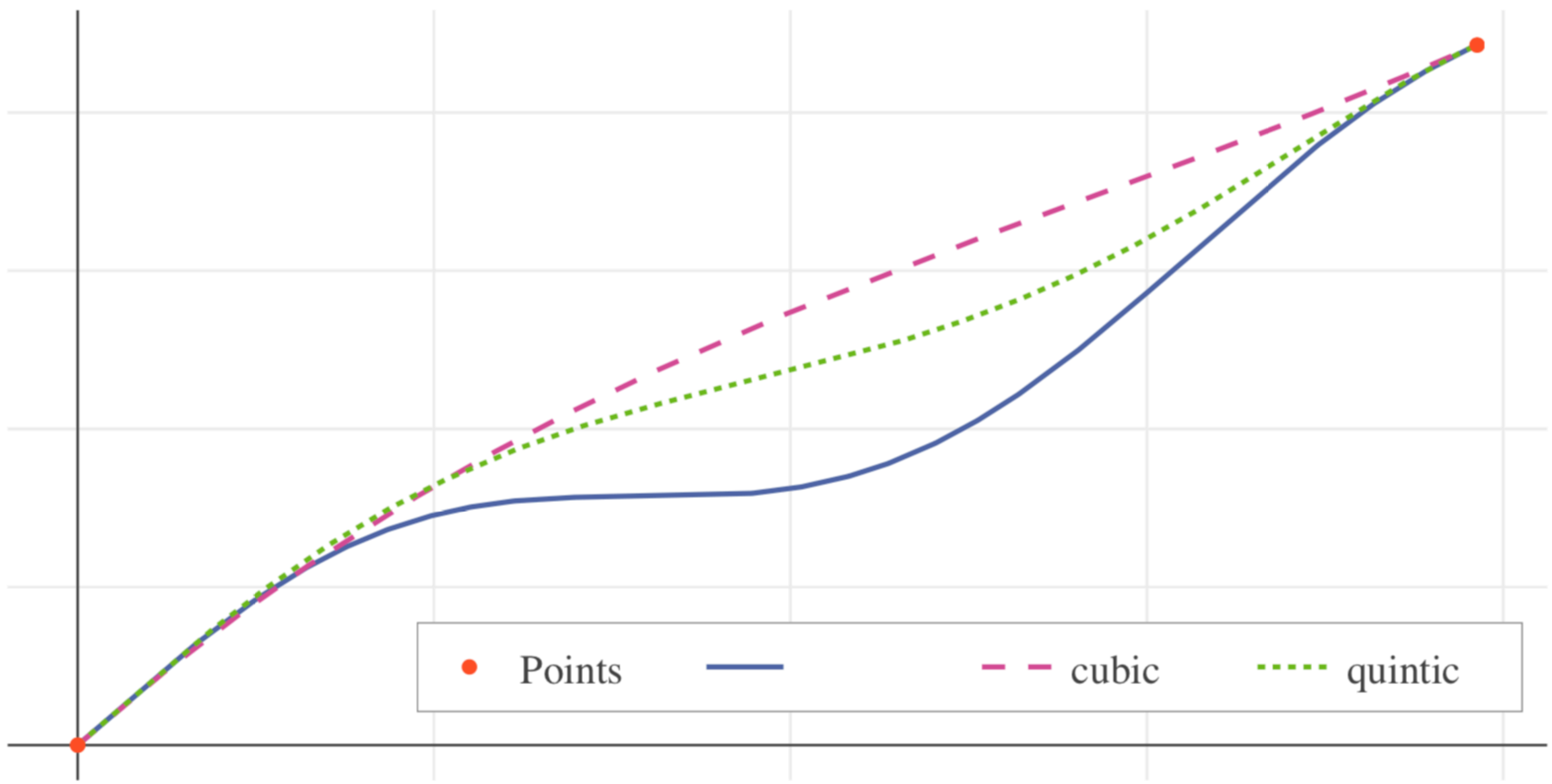
Empirical Distribution Approximations



COMPUTER SCIENCE
VIRGINIA TECH

Connection to Variability





Monotone Piecewise Quintic Splines

Theory existed to create splines that are composed of quintics, but no mathematical software had been produced!

G. Ulrich and L. Watson. Positivity conditions for quartic polynomials. SIAM Journal on Scientific Computing, 15(3):528–544, 1994. doi: 10.1137/0915035. URL https://doi. org/10.1137/0915035.

Walter Hess and Jochen W Schmidt. Positive quartic, monotone quintic c2-spline interpolation in one and two dimensions. Journal of Computational and Applied Mathematics, 55(1): 51–67, 1994. doi: 10.1016/0377-0427(94)90184-8.

Dougherty, Randall L., Alan S. Edelman, and James M. Hyman. Nonnegativity-, monotonicity-, or convexity-preserving cubic and quintic Hermite interpolation. Mathematics of Computation 52.186 (1989): 471-494.

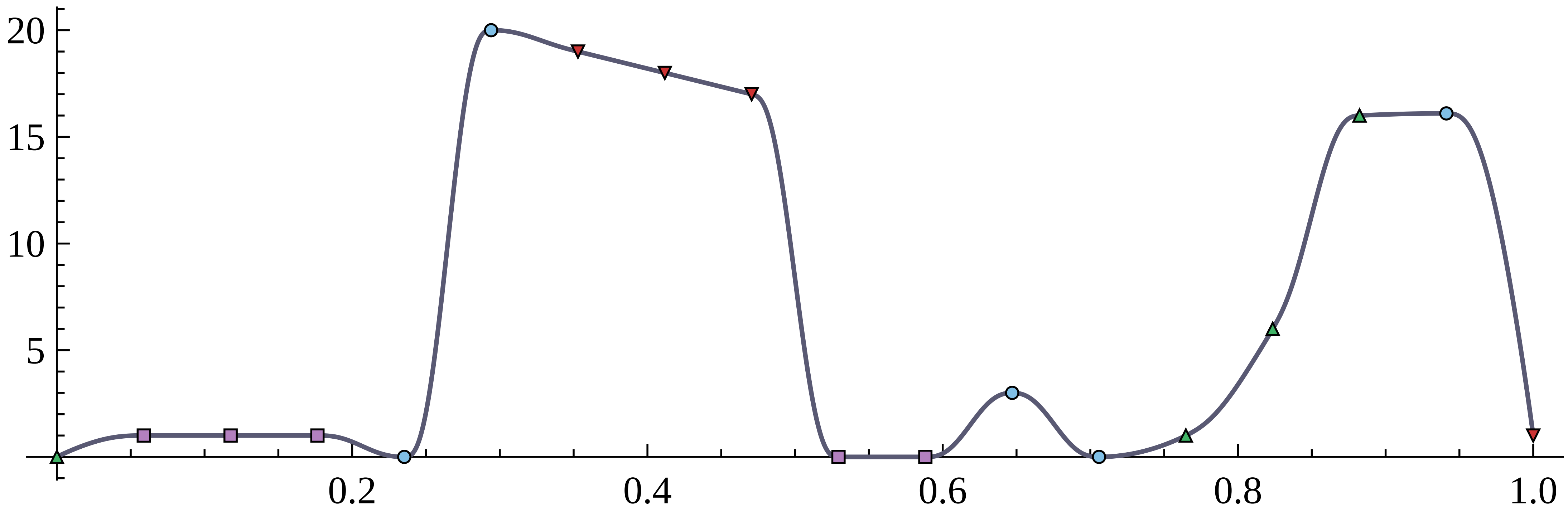
Proposed Software Package, MQSI

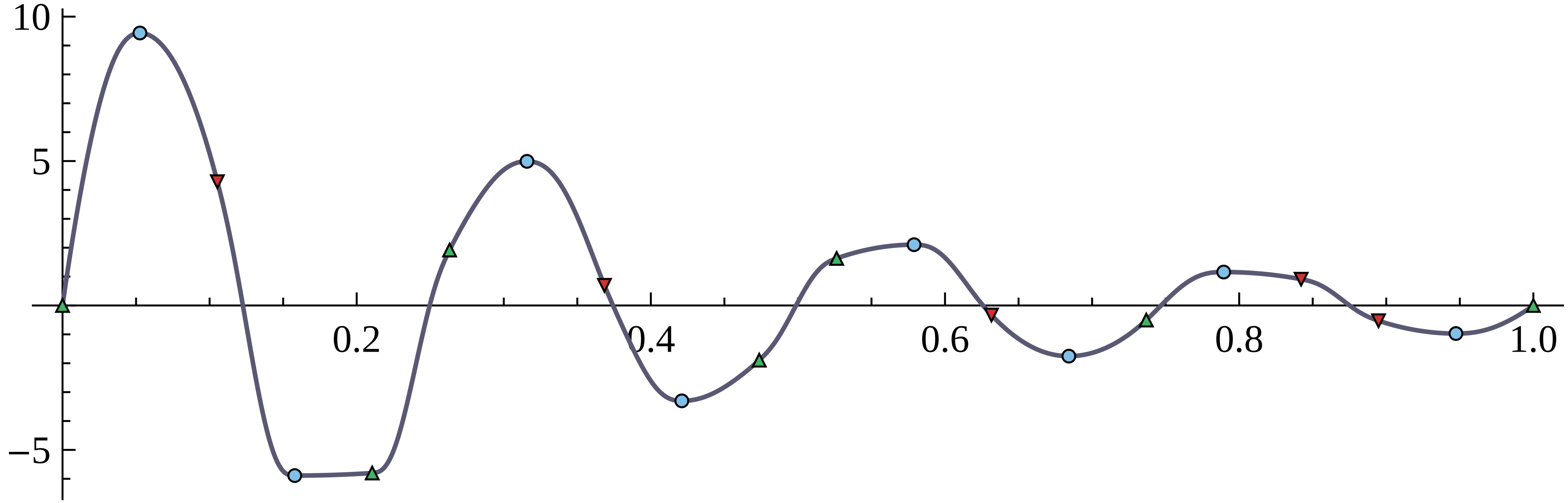
Algorithm 1: `QUADRATIC_FACET`$\left(X(1{:}n),\ Y(1{:}n),\ i\right)$
where $X_j,\ Y_j \in \mathbb{R}$ for $j = 1, \ldots, n$, $1 \leq i \leq n$, and $n \geq 3$. Returns the slope and curvature at X_i of the local quadratic interpolant with minimum magnitude curvature.

Algorithm 2:
`IS_MONOTONE`$\left(x_0,\ x_1,\ f(x_0),\ Df(x_0),\ D^2f(x_0),\ f(x_1),\ Df(x_1),\ D^2f(x_1)\right)$
where $x_0,\ x_1 \in \mathbb{R}$, $x_0 < x_1$, and f is an order six polynomial defined by $f(x_0)$, $Df(x_0)$, $D^2f(x_0)$, $f(x_1)$, $Df(x_1)$, $D^2f(x_1)$. Returns `TRUE` if f is monotone increasing on $[x_0, x_1]$.

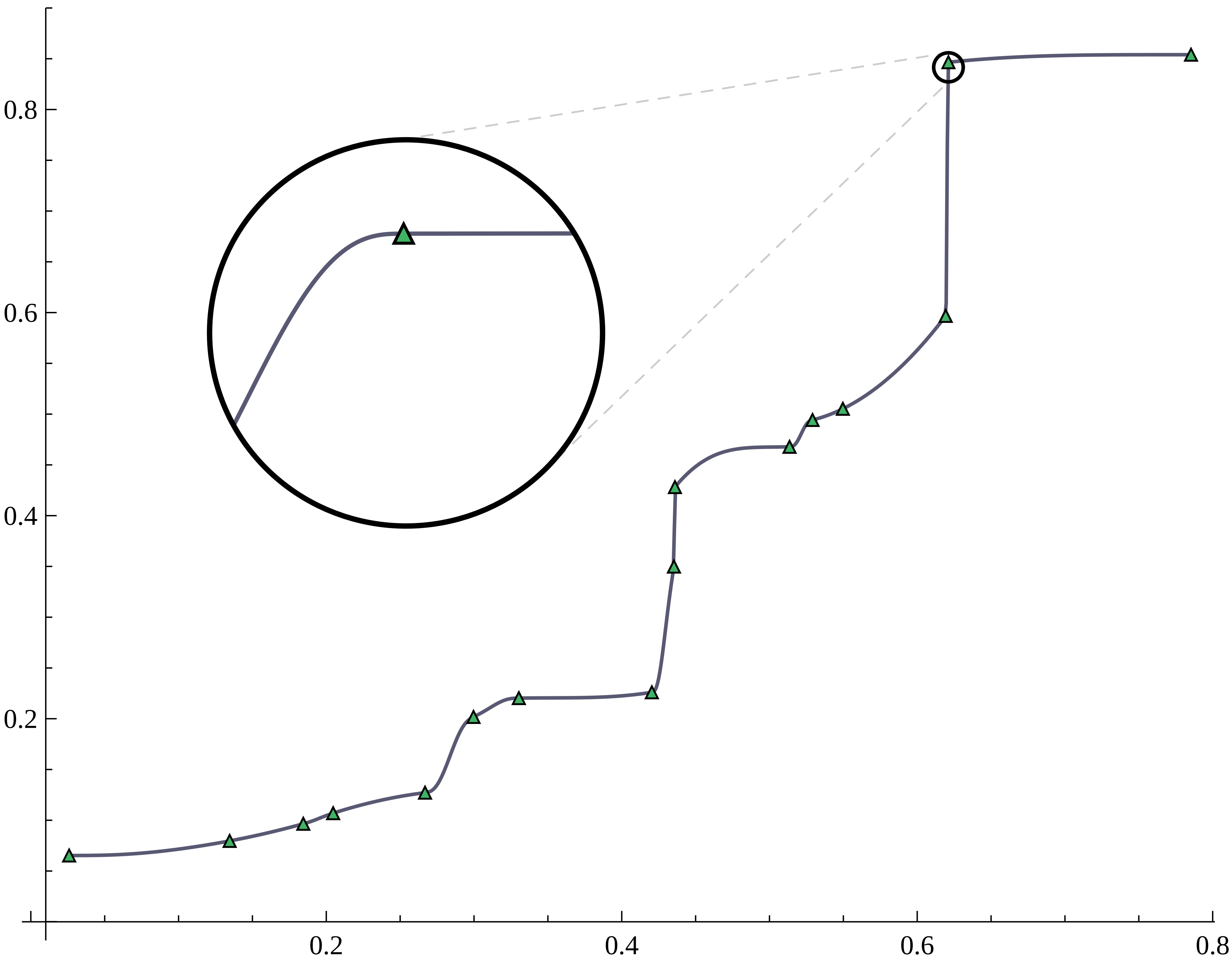
Algorithm 3: `MQSI`$\left(X(1{:}n), Y(1{:}n)\right)$
where $(X_i, Y_i) \in \mathbb{R} \times \mathbb{R}$, $i = 1, \ldots, n$ are data points. Returns monotone quintic spline interpolant $Q(x)$ such that $Q(X_i) = Y_i$ and is monotone increasing (decreasing) on all intervals that Y_i is monotone increasing (decreasing).

Some examples



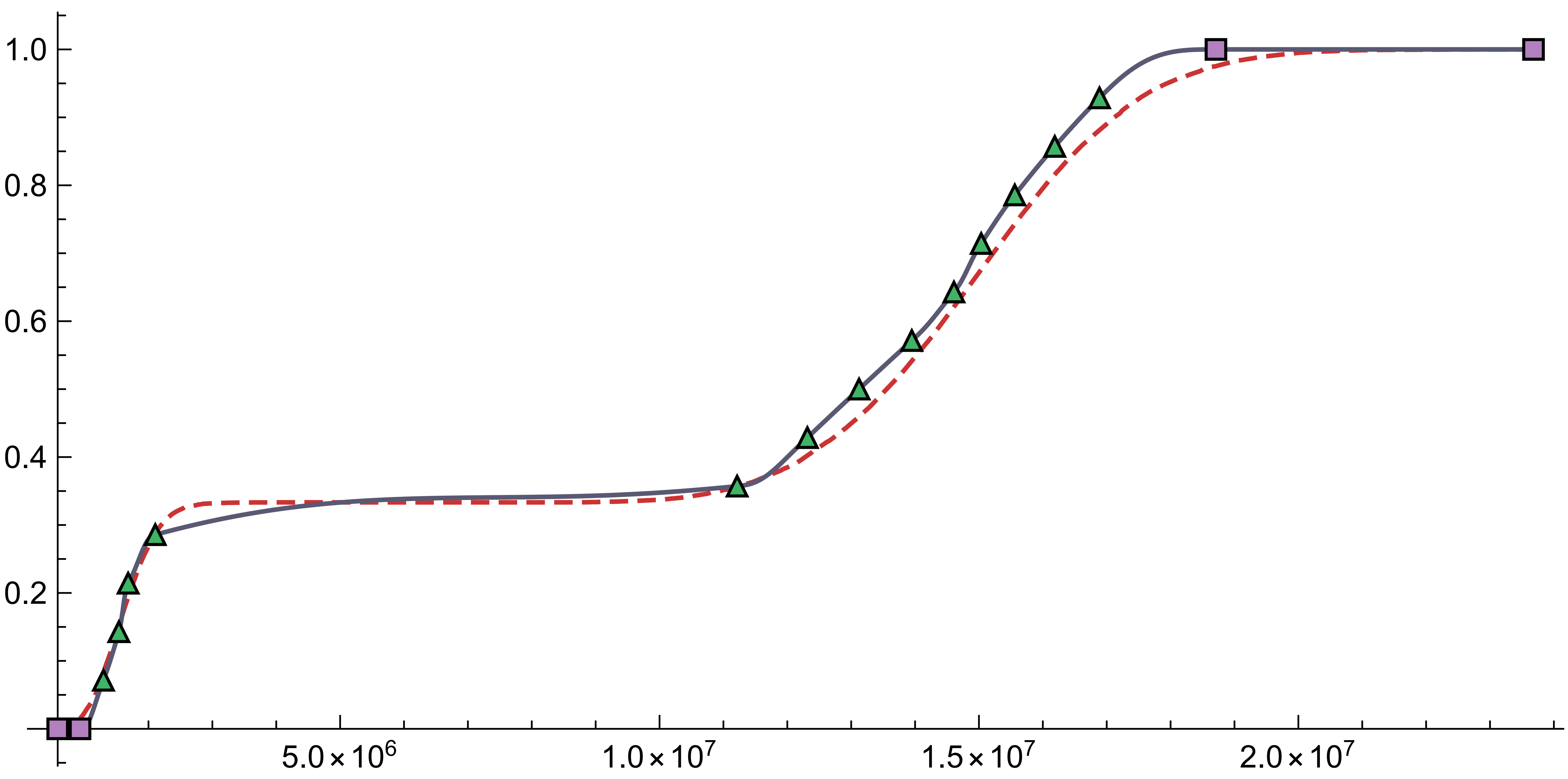


MQSI given difficult data

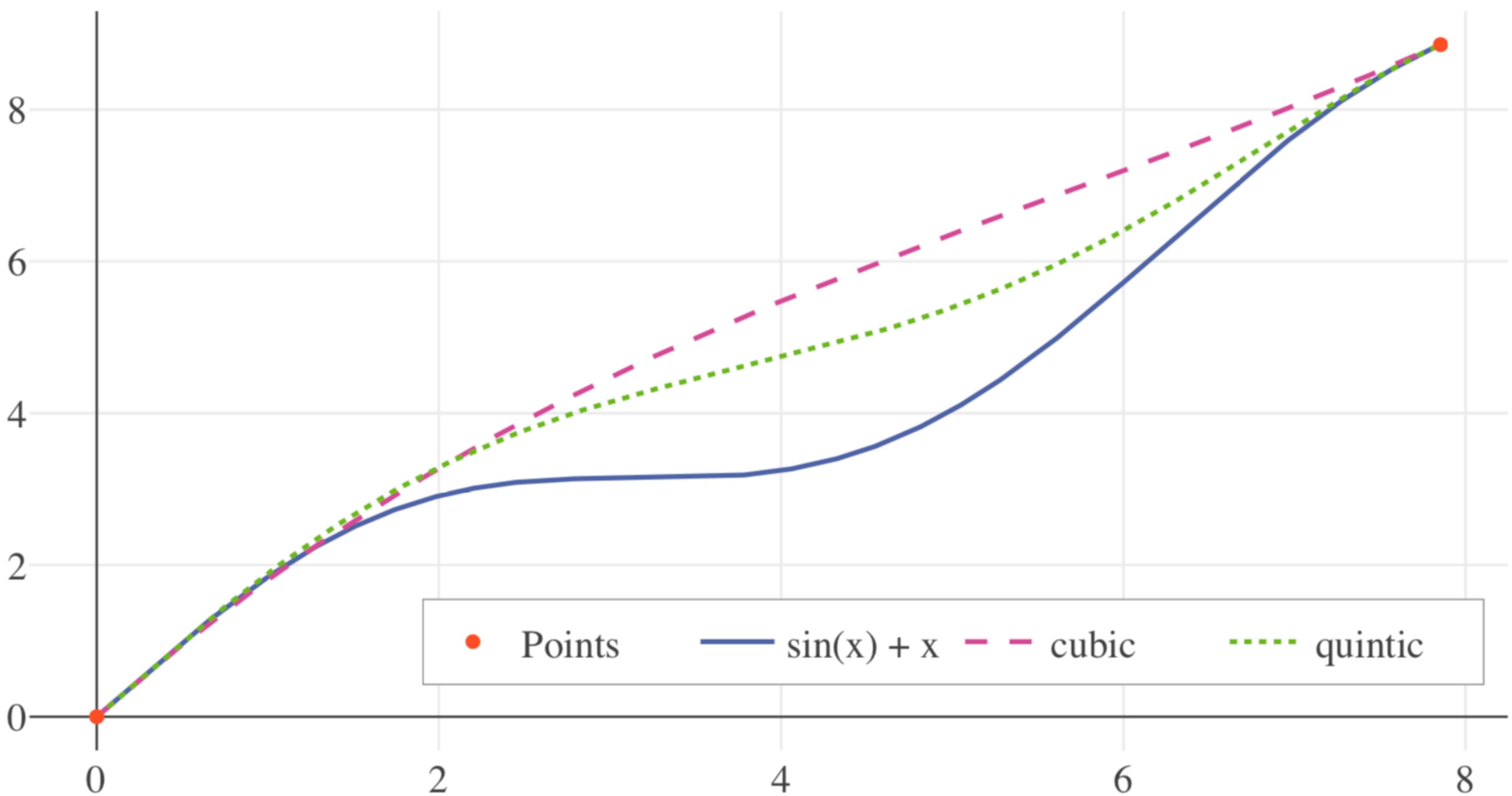


COMPUTER SCIENCE
VIRGINIA TECH.

Approximating VarSys Data

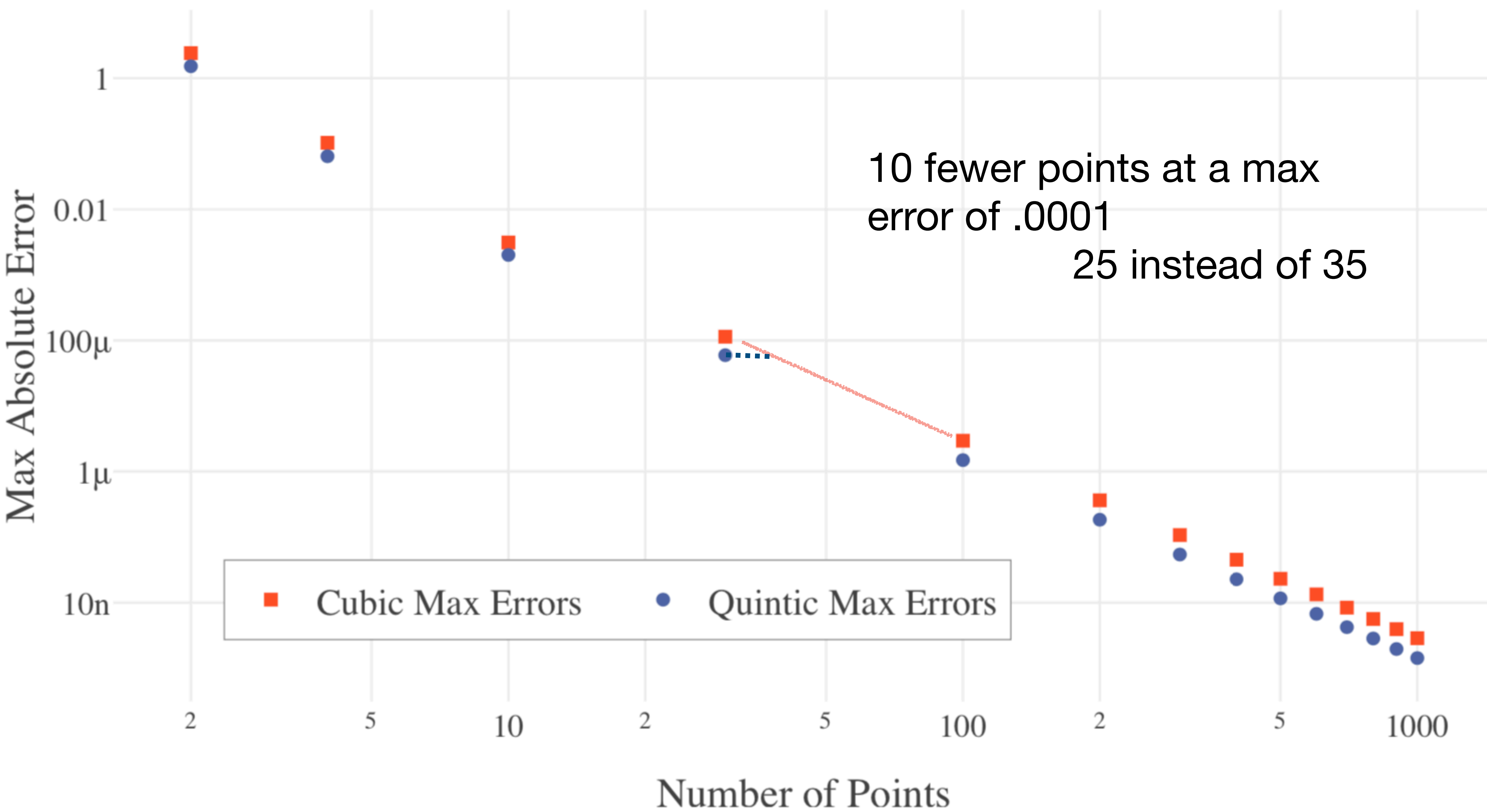


Measuring Accuracy, a Test



The Maximum Error with More Data

The Maximum Error with More Data

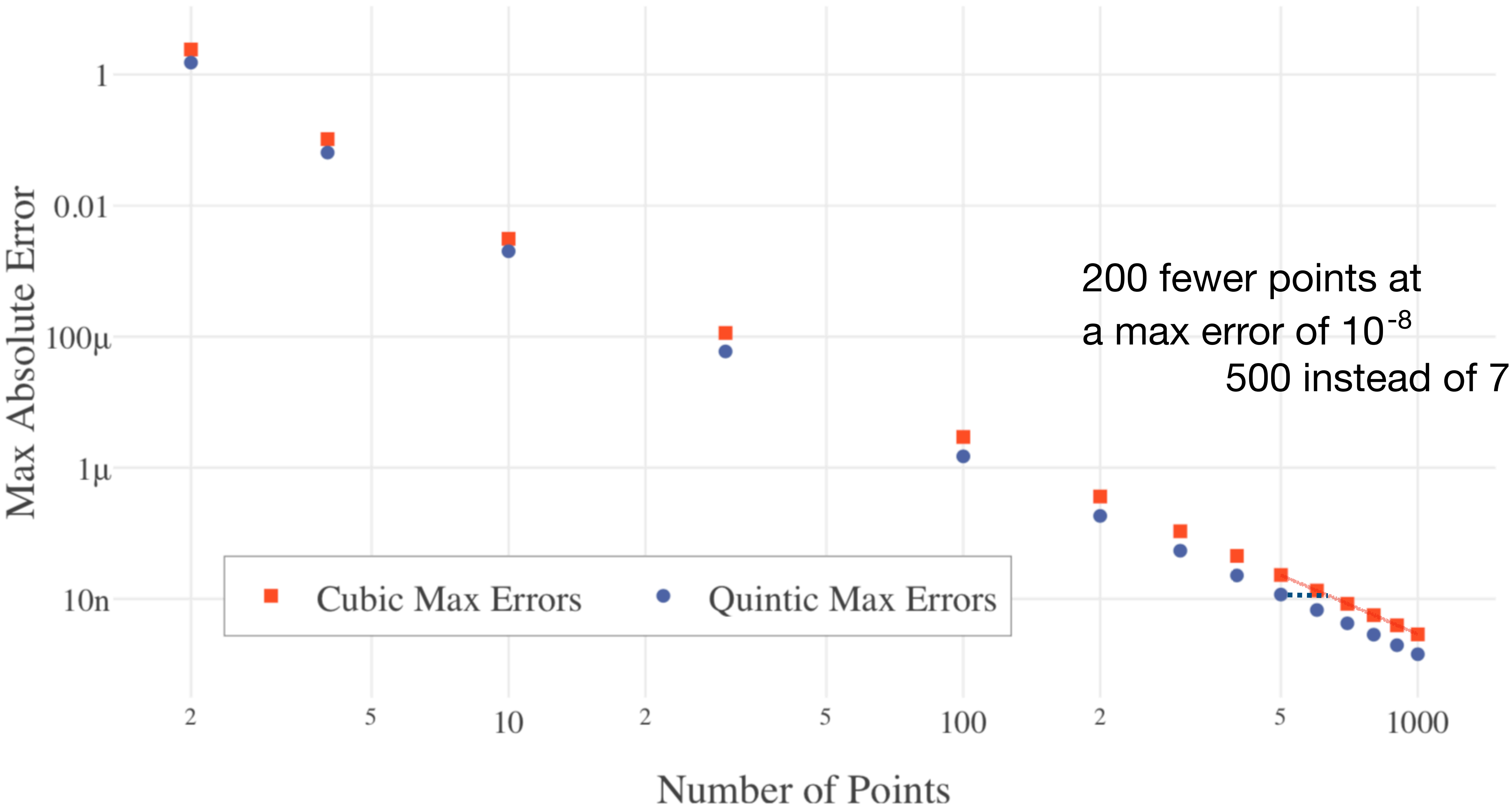


COMPUTER SCIENCE
VIRGINIA TECH.

The Maximum Error with More Data

COMPUTER SCIENCE
VIRGINIA TECH.

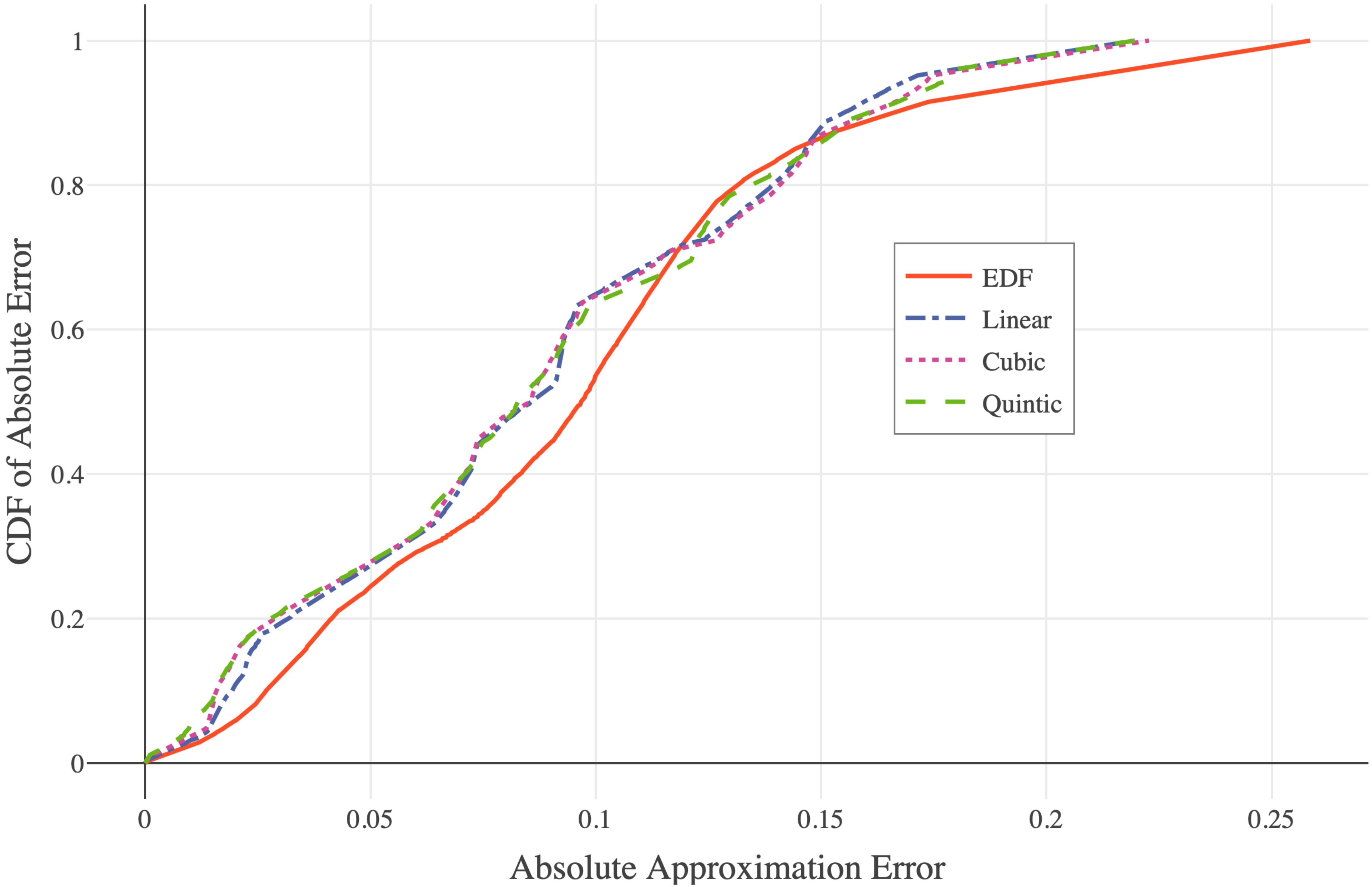
The Maximum Error with More Data



200 fewer points at a max error of 10^{-8}

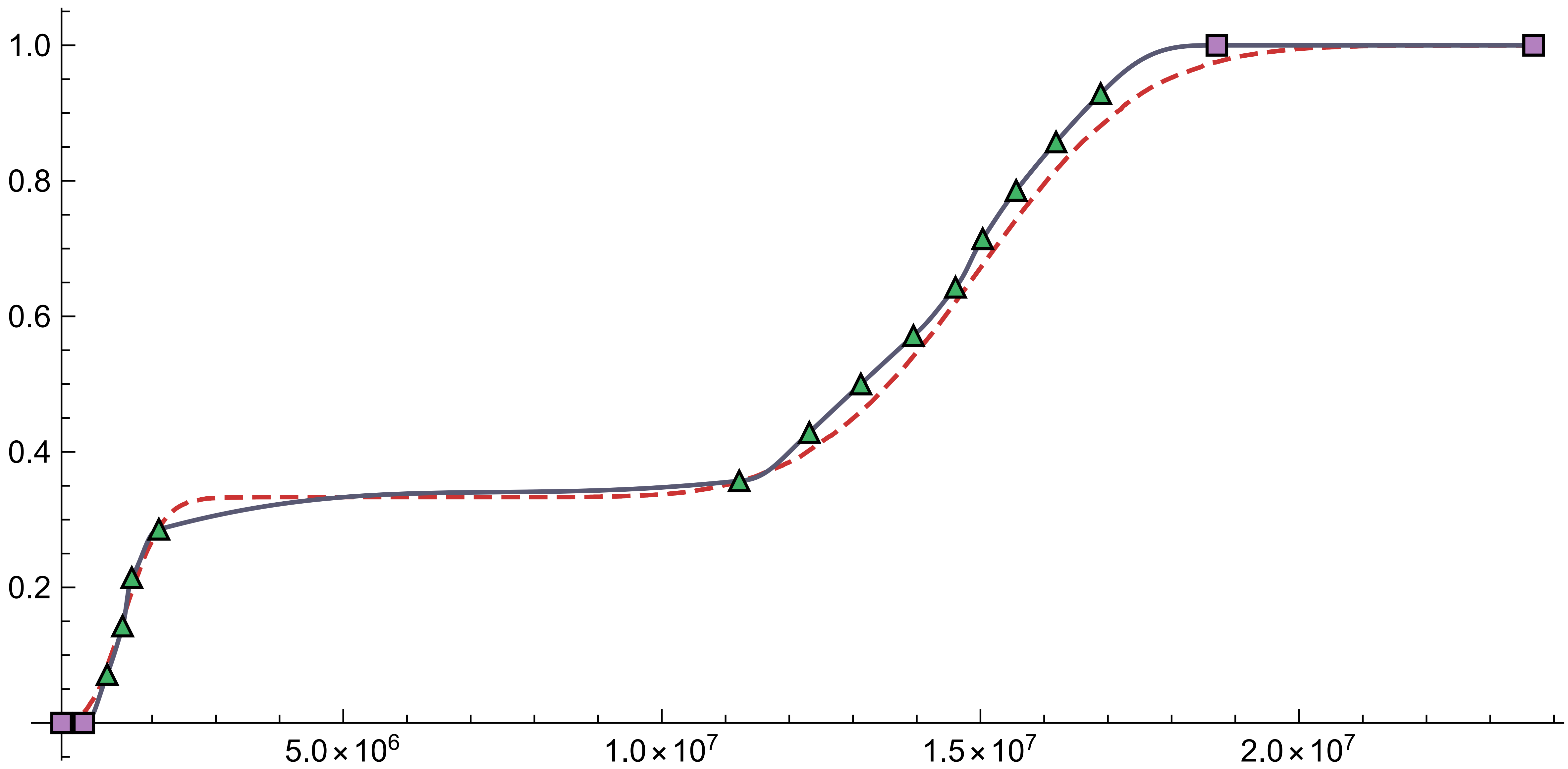
500 instead of 700

The Quintic is Often Better for Tested VarSys Data



COMPUTER SCIENCE
VIRGINIA TECH.

Approximating VarSys Data



The Big Picture

1. The Importance and Applications of Variability
2. Algorithms for Constructing Approximations
3. Naive Approximations of Variability
4. Box-Splines: Uses, Constructions, and Applications
5. Stronger Approximations of Variability
6. An Error Bound for Piecewise Linear Interpolation
7. A Package for Monotone Quintic Spline Interpolation

Publications along the way

T. C. H. Lux, T. H. Chang, L. T. Watson, J. Bernard, B. Li, L. Xu, G. Back, A. R. Butt, K. W. Cameron, and Y. Hong, “Predictive modeling of I/O characteristics in high performance computing systems”, in Proc. 2018 Spring Simulation Multiconference, 26th High Performance Computing Symp., L. T. Watson, W. I. Thacker, M. Sosonkina, J. Weinbub, and K. Rupp (eds.), Soc. for Modelling and Simulation Internat., Vista, CA, 2018, 434–445.

Thomas C. H. Lux, Layne T. Watson, Tyler H. Chang, Jon Bernard, Bo Li, Xiaodong Yu, Li Xu, Godmar Back, Ali R. Butt, Kirk W. Cameron, Danfeng Yao, and Yili Hong. 2018. Novel meshes for multivariate interpolation and approximation. In Proceedings of the ACMSE 2018 Conference (ACMSE '18). ACM, New York, NY, USA, Article 13, 7 pages. DOI: https://doi.org/10.1145/3190645.3190687

Thomas C. H. Lux, Layne T. Watson, Tyler H. Chang, Jon Bernard, Bo Li, Xiaodong Yu, Li Xu, Godmar Back, Ali R. Butt, Kirk W. Cameron, Danfeng Yao, and Yili Hong. 2018. Nonparametric Distribution Models for Predicting and Managing Computational Performance Variability. In Proceedings of the IEEE SoutheastCon 2018 Conference (IEEESE '18). IEEE, Tampa, FL, USA, 7 pages.

T. C. H. Lux, L. T. Watson, T. H. Chang, Y. Hong, K. C. Cameron (2019) Interpolation of Sparse High-Dimensional Data. Springer Numerical Algorithms.

Lux, T.C.H., Chang, T.H. & Tipirneni, S.S. Least-squares solutions to polynomial systems of equations with quantum annealing. Quantum Inf Process 18, 374 (2019). https://doi.org/10.1007/s11128-019-2489-x

T. C. H. Lux, S. Nagy, M. Almanaa, S. Yao and R. Bixler, "A Case Study on a Sustainable Framework for Ethically Aware Predictive Modeling," 2019 IEEE International Symposium on Technology and Society (ISTAS), Medford, MA, USA, 2019, pp. 1-7.

Thomas C. H. Lux, Layne T. Watson, Tyler H. Chang, Li Xu, Yueyao Wang, Yili Hong. An Algorithm for Constructing Monotone Quintic Interpolating Splines. Spring Simulation Multiconference, High Performance Computing Symposium. February, 2020.

Thomas C. H. Lux, Layne T. Watson, Tyler H. Chang, Li Xu, Yueyao Wang, Jon Bernard, Yili Hong, Kirk W. Cameron. Effective Nonparametric Distribution Modeling for Distribution Approximation Applications. Institute of Electrical and Electronics Engineers Southeastcon. January, 2020.

Thomas C. H. Lux, Tyler H. Chang. Analytic Test Functions for Generalizable Evaluation of Convex Optimization Techniques. Institute of Electrical and Electronics Engineers Southeastcon. January, 2020.

Thomas C. H. Lux, Layne T. Watson, Tyler H. Chang, William I. Thacker. Algorithm XXXX: MQSI—Monotone Quintic Spline Interpolation. ACM Transactions on Mathematical Software. Planning to submit August, 2020.

Publications along the way

Contributed Chapters	
1,3	T. C. H. Lux, T. H. Chang, L. T. Watson, J. Bernard, B. Li, L. Xu, G. Back, A. R. Butt, K. W. Cameron, and Y. Hong, “Predictive modeling of I/O characteristics in high performance computing systems”, in Proc. 2018 Spring Simulation Multiconference, 26th High Performance Computing Symp., L. T. Watson, W. I. Thacker, M. Sosonkina, J. Weinbub, and K. Rupp (eds.), Soc. for Modelling and Simulation Internat., Vista, CA, 2018, 434–445.
4	Thomas C. H. Lux, Layne T. Watson, Tyler H. Chang, Jon Bernard, Bo Li, Xiaodong Yu, Li Xu, Godmar Back, Ali R. Butt, Kirk W. Cameron, Danfeng Yao, and Yili Hong. 2018. Novel meshes for multivariate interpolation and approximation. In Proceedings of the ACMSE 2018 Conference (ACMSE '18). ACM, New York, NY, USA, Article 13, 7 pages. DOI: https://doi.org/10.1145/3190645.3190687
1,5	Thomas C. H. Lux, Layne T. Watson, Tyler H. Chang, Jon Bernard, Bo Li, Xiaodong Yu, Li Xu, Godmar Back, Ali R. Butt, Kirk W. Cameron, Danfeng Yao, and Yili Hong. 2018. Nonparametric Distribution Models for Predicting and Managing Computational Performance Variability. In Proceedings of the IEEE SoutheastCon 2018 Conference (IEEESE '18). IEEE, Tampa, FL, USA, 7 pages.
1,2,6	T. C. H. Lux, L. T. Watson, T. H. Chang, Y. Hong, K. C. Cameron (2019) Interpolation of Sparse High-Dimensional Data. Springer Numerical Algorithms.
-	Lux, T.C.H., Chang, T.H. & Tipirneni, S.S. Least-squares solutions to polynomial systems of equations with quantum annealing. Quantum Inf Process 18, 374 (2019). https://doi.org/10.1007/s11128-019-2489-x
-	T. C. H. Lux, S. Nagy, M. Almanaa, S. Yao and R. Bixler, "A Case Study on a Sustainable Framework for Ethically Aware Predictive Modeling," 2019 IEEE International Symposium on Technology and Society (ISTAS), Medford, MA, USA, 2019, pp. 1-7.
1,7	Thomas C. H. Lux, Layne T. Watson, Tyler H. Chang, Li Xu, Yueyao Wang, Yili Hong. An Algorithm for Constructing Monotone Quintic Interpolating Splines. Spring Simulation Multiconference, High Performance Computing Symposium. February, 2020.
7	Thomas C. H. Lux, Layne T. Watson, Tyler H. Chang, Li Xu, Yueyao Wang, Jon Bernard, Yili Hong, Kirk W. Cameron. Effective Nonparametric Distribution Modeling for Distribution Approximation Applications. Institute of Electrical and Electronics Engineers Southeastcon. January, 2020.
-	Thomas C. H. Lux, Tyler H. Chang. Analytic Test Functions for Generalizable Evaluation of Convex Optimization Techniques. Institute of Electrical and Electronics Engineers Southeastcon. January, 2020.
1,7	Thomas C. H. Lux, Layne T. Watson, Tyler H. Chang, William I. Thacker. Algorithm XXXX: MQSI—Monotone Quintic Spline Interpolation. ACM Transactions on Mathematical Software. Planning to submit August, 2020.

VarSys Publications coauthored

Tyler H. Chang, Layne T. Watson, Thomas C. H. Lux, Bo Li, Li Xu, Ali R. Butt, Kirk W. Cameron, and Yili Hong. 2018. A polynomial time algorithm for multivariate interpolation in arbitrary dimension via the Delaunay triangulation. In Proceedings of the ACMSE 2018 Conference (ACMSE '18). ACM, New York, NY, USA, Article 12, 8 pages. DOI: https://doi.org/10.1145/3190645.3190680

T. H. Chang, L. T. Watson, T. C. H. Lux, J. Bernard, B. Li, L. Xu, G. Back, A. R. Butt, K. W. Cameron, and Y. Hong, ``Predicting system performance by interpolation using a high-dimensional Delaunay triangulation'', in Proc. 2018 Spring Simulation Multiconference, 26th High Performance Computing Symp., L. T. Watson, W. I. Thacker, M. Sosonkina, J. Weinbub, and K. Rupp (eds.), Soc. for Modelling and Simulation Internat., Vista, CA, 2018, 363--374.

Tyler H. Chang, Layne T. Watson, Thomas C. H. Lux, Sharath Raghvendra, Bo Li, Li Xu, Ali R. Butt, Kirk W. Cameron, and Yili Hong. 2018. Computing the Umbrella Neighborhood of a Vertex in the Delaunay Triangulation and a Single Voronoi Cell in Arbitrary Dimension. In Proceedings of the IEEE SoutheastCon 2018 Conference (IEEESE '18). IEEE, Tampa, FL, USA, 7 pages.

Cameron, K.W., Anwar, A., Cheng, Y., Xu, L., Li, B., Ananth, U., Bernard, J., Jearls, C., Lux, T., Hong, Y., Watson, L.T. and Butt, A. R. (2019). MOANA: Modeling and Analyzing I/O Variability in Parallel System Experimental Design. IEEE Transactions on Parallel and Distributed Systems, 1–1. doi:10.1109/tpds.2019.2892129

T. H. Chang, L. T. Watson, T. C. H. Lux, A. R. Butt, K. C. Cameron, Y. Hong (2019) Algorithm XXX: DELAUNAYSPARSE: Interpolation via a Sparse Subset of the Delaunay Triangulation in Medium to High Dimensions. ACM Transaction on Math Software.

Li Xu, Thomas Lux, Tyler Chang, Bo Li, Yili Hong, Layne Watson, Kirk Cameron, and Jon Bernard. 2018. Prediction for Distributional Outcomes in the Management of High-Performance Computing Input/output (I/O) Variability. Joint Statistical Meetings - American Statistical Association.

Li Xu, Yueyao Wang, Thomas Lux, Tyler Chang, Jon Bernard, Bo Li, Yili Hong, Kirk Cameron, Layne Watson, Modeling I/O performance variability in high-performance computing systems using mixture distributions, Journal of Parallel and Distributed Computing, Volume 139, 2020, Pages 87-98.

Tyler H. Chang, Jeffrey Larson, Layne T. Watson, Thomas C. H. Lux. Managing Computationally Expensive Blackbox Multiobjective Optimization Problems with libEnsemble. Spring Simulation Multiconference, High Performance Computing Symposium. accepted February, 2020.

Tyler H. Chang, Layne T. Watson, Jeffery Larson, Thomas C. H. Lux. Algorithm XXX: VTMOP: Multi-objective optimization for black box problems. ACM Transactions on Mathematical Software. Planning to submit May, 2020.

Many thanks to

Advisors

Dr. Layne Watson

Dr. Yili Hong

Committee Members

Dr. Kirk Cameron

Dr. Bert Huang

Dr. Gang Wang

Dr. Yang Cao

Many thanks to

Advisors

Dr. Layne Watson

Dr. Yili Hong

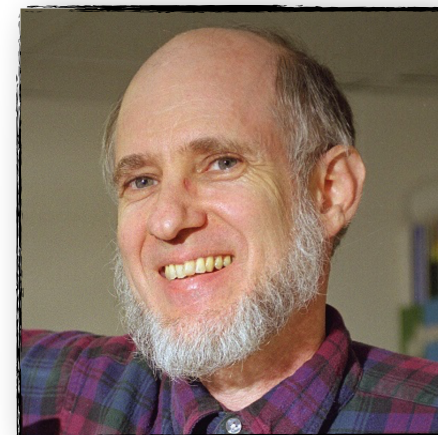
Committee Members

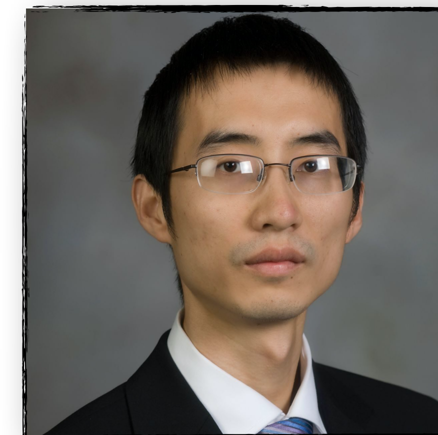
Dr. Kirk Cameron

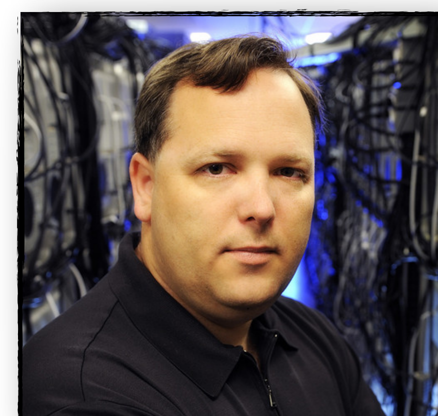
Dr. Bert Huang

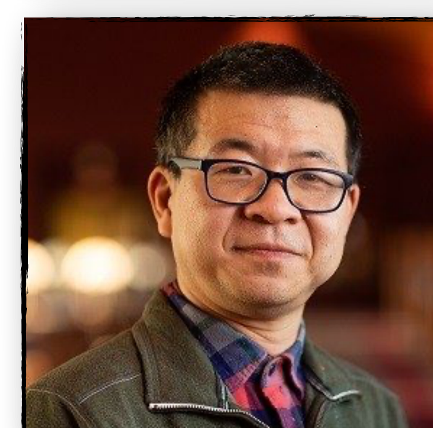
Dr. Gang Wang

Dr. Yang Cao









Overview

1. The Importance and Applications of Variability
   define variability, why is it important?
2. Algorithms for Constructing Approximations
   approximation, regression and interpolation techniques
3. Naive Approximations of Variability
   mean, variance, and standard deviation prediction with IOzone
4. Box-Splines: Uses, Constructions, and Applications
   spline overview, box splines, meshes, fitting, and data sets
5. Stronger Approximations of Variability
   predicting distributions, measuring error, and tuning
6. An Error Bound for Piecewise Linear Interpolation
   theoretical bound, synthetic demo, and empirical results
7. A Package for Monotone Quintic Spline Interpolation
   MQSI algorithms, example pictures, performance study