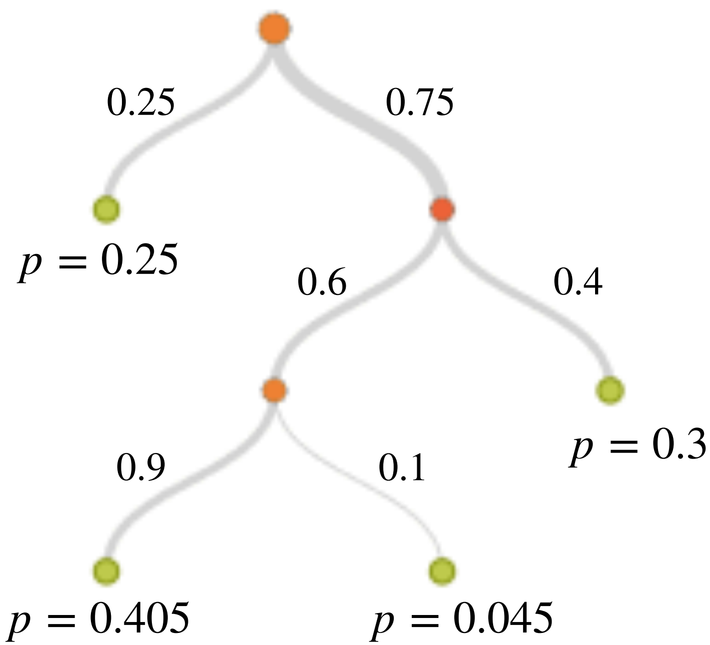
Random Variable Arithmetic

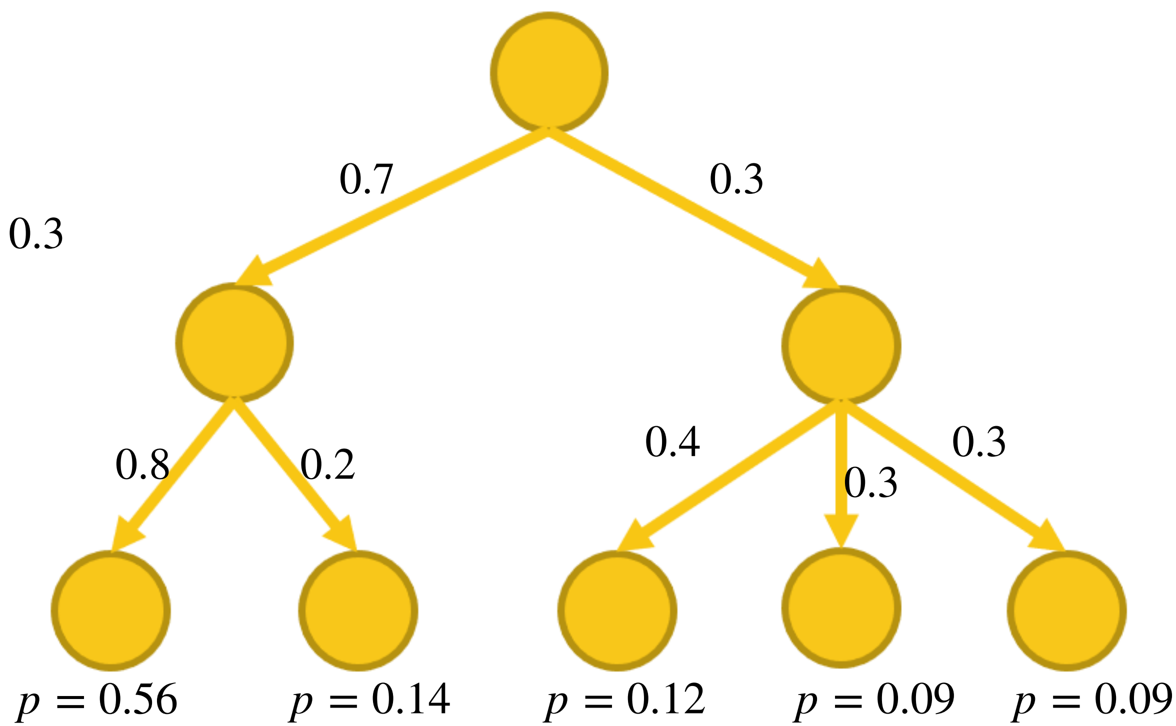
Presenter: Jess Meyer
Collaborators: John Daly, Lauren Huckaba,
Shelby Coppolino, Alex Poulos, Casey Orndoff

Nondeterministic Computing

Nondeterministic computing refers to a model of computation that can have stochastic (or random) outputs.

Probabilistic Turing machines map inputs to one of several outputs.





Background: RELIABLEGUIDE

- (U//FOUO) RELIABLEGUIDE was a three-year exploration of the potential applications of probabilistic computing techniques.
- (U//FOUO) Nondeterministic hardware and programming methods have the potential to deliver multiple orders of magnitude improvement in computational efficiency (operations per Joule) compared to their fully deterministic counterparts.

A Result From RELIABLEGUIDE

In order for a computation to be correct, every step along the way has to be correct.

The Question behind this Project

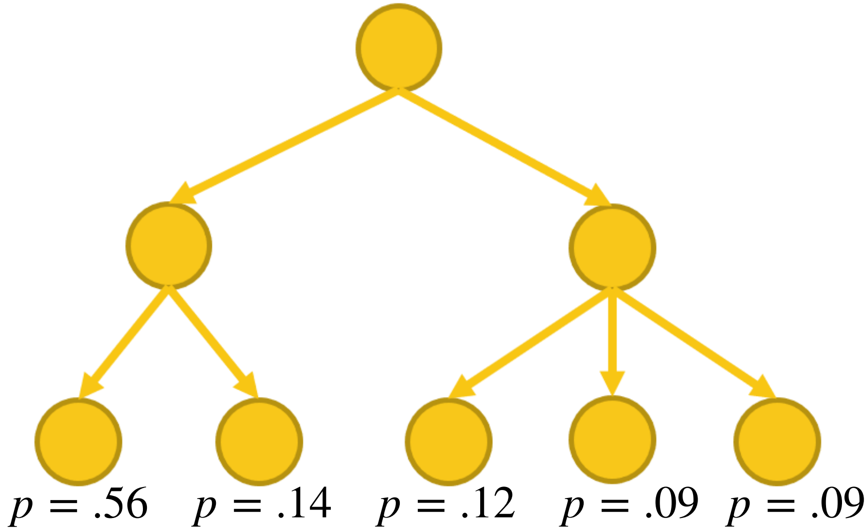
Can we conceive a mechanism whereby we can recover the correct solution to a deterministic problem with some fixed probability even if there are incorrect intermediate results along the way?

Random Variables

Random variables are variables whose values depend on the outcome of random phenomena.

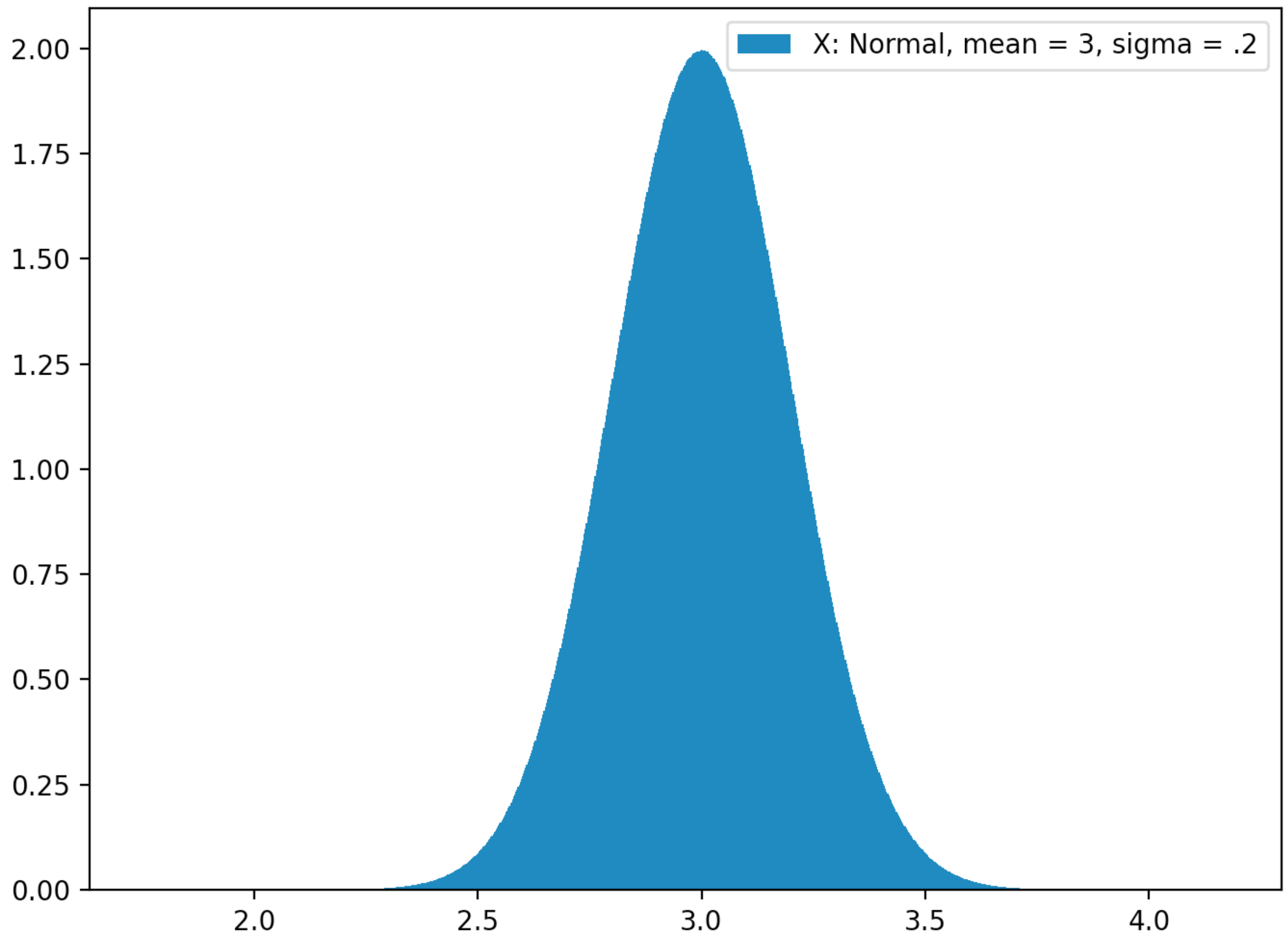
Every random variable has a corresponding **probability distribution**, which describes the probability that the random variable is in any particular range of values.

Probability density functions specify the probability distribution of continuous random variables.



Example: Normal Random Variable

If you randomly generate values according to the distribution of this random variable, 95% of the values will be in the interval (2.6, 3.4).

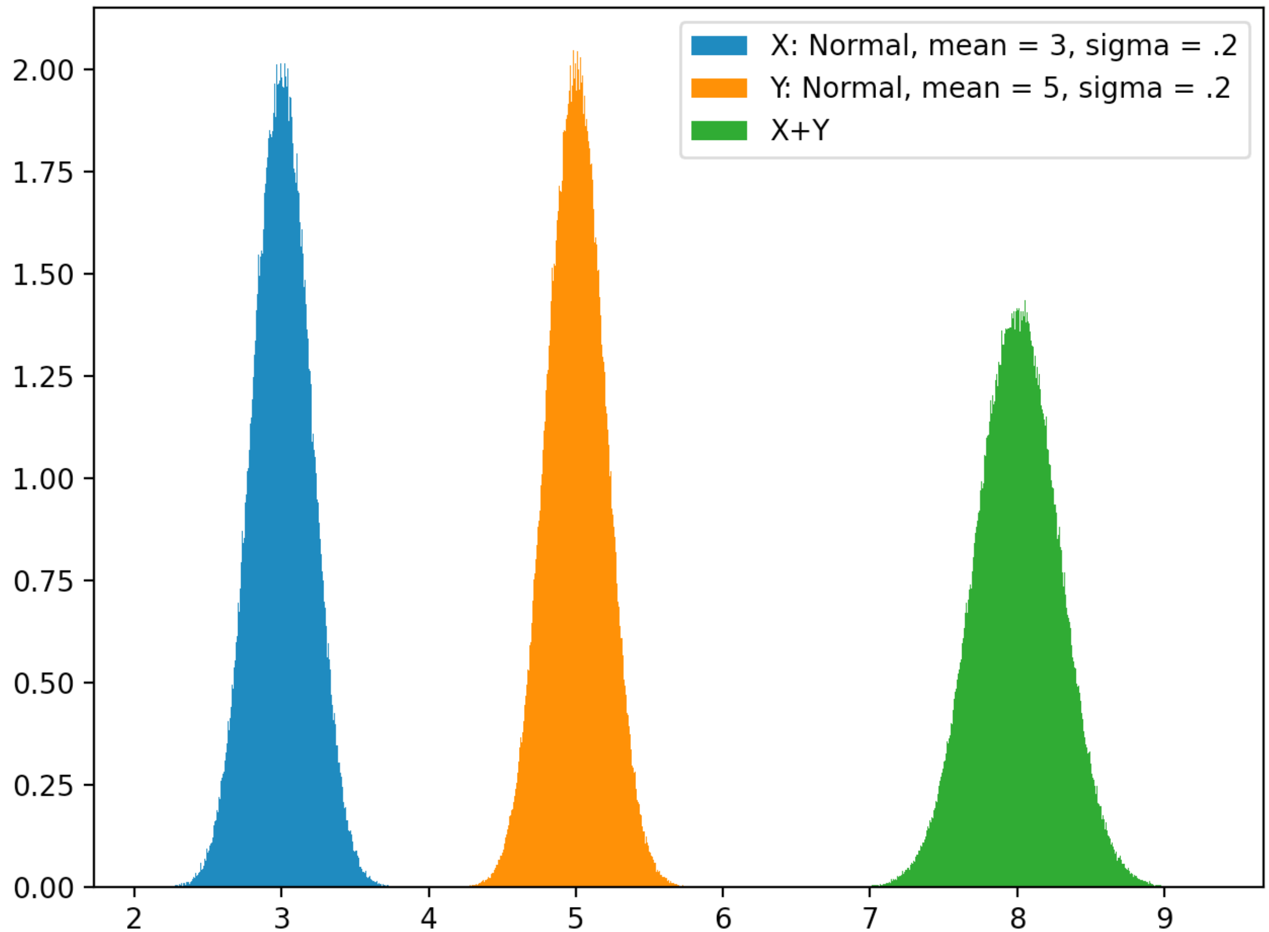


Random Variable Arithmetic

Random Variable Arithmetic is the study of properties and manipulation of random variables.

Example: Random Variable Arithmetic

- Deterministic problem: Add 3 and 5
- Represent 3 and 5 as normal random variables, centered at 3 and 5
- The output is a normal random variable, centered at 8



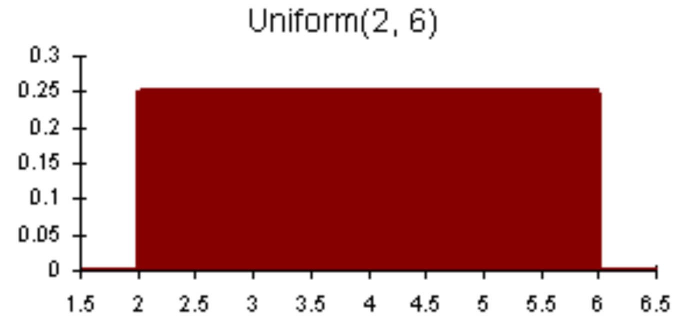
The Question behind this Project

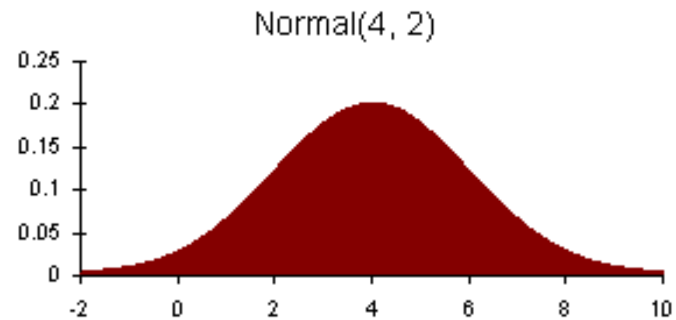
Can we conceive a mechanism whereby we can recover the correct solution to a deterministic problem with some fixed probability even if there are incorrect intermediate results along the way?

We are using random variable arithmetic to model nondeterministic computation, i.e., random variable arithmetic is the proposed "mechanism".

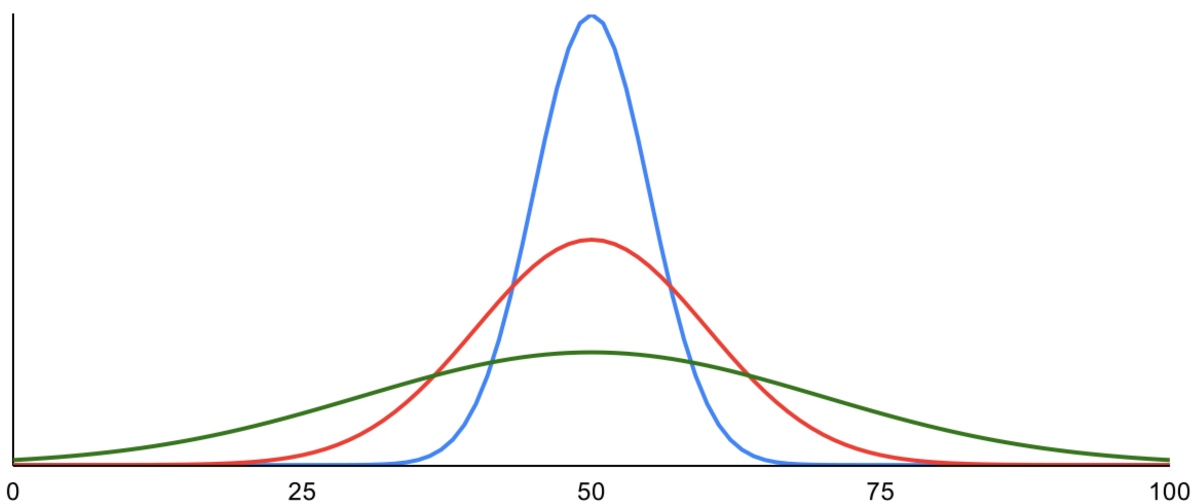
Dispersion

Dispersion is the extent to which a probability distribution is “spread out”.





Probability density functions of normal random variables all centered at 50.



Probability density functions of random variables that are

- ***uniform*** on the interval [2,6]
- ***normal*** and centered at 4.

Variance is a Measure of Dispersion

In order for any model of nondeterministic computing to be practical:

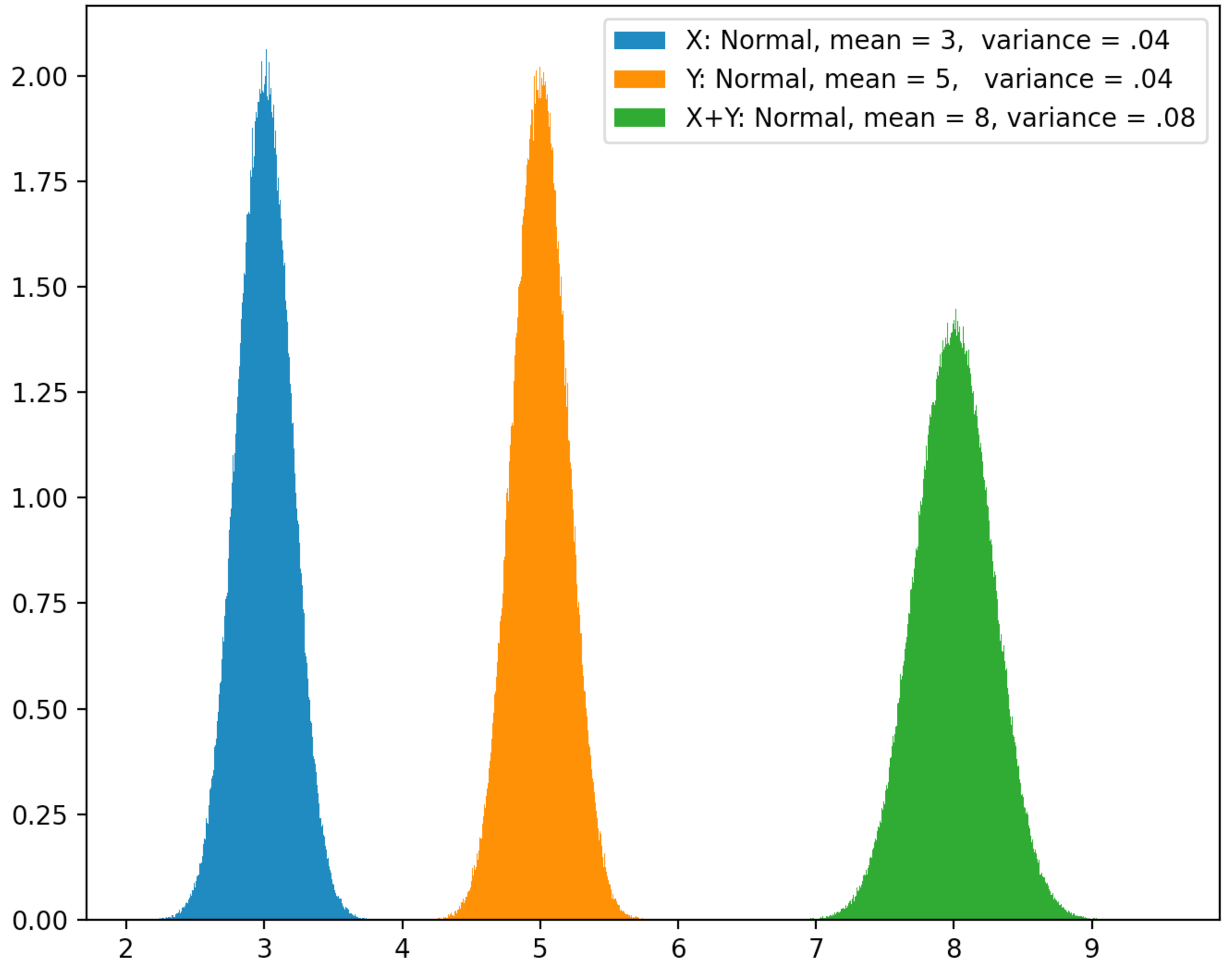
- We need to be able to predict and control the **dispersion** of the output.

Parameters that measure the dispersion are called **Measures of Dispersion.**

❖ **Variance** is the most common measure of dispersion.

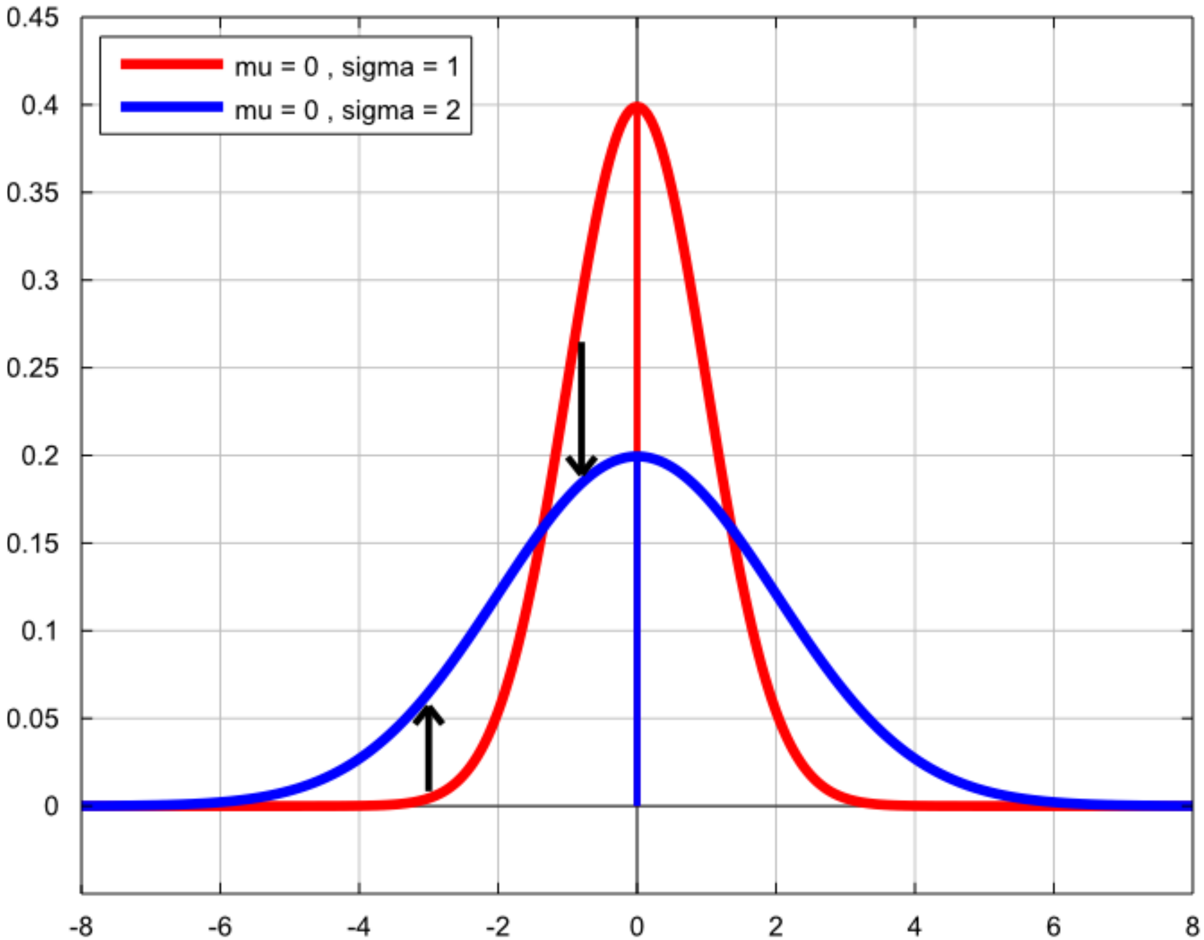
Example: Adding Two Normal Random Variables

- Deterministic problem: Add 3 and 5
- Represent 3 and 5 as normal random variables, with variance .04
- The output is a normal random variable, with variance .08



Statistical Dependence

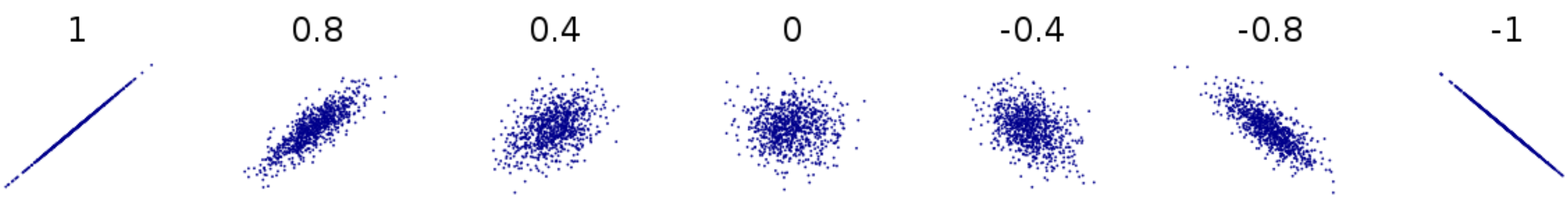
Random variables are **dependent** if they are “related" in some way.



Correlation: Linear Dependence

Correlation is a measure of linear dependence.

- **Pearson's correlation coefficient**, ρ, is the most common measure of linear dependence.
- $-1 \leq \rho \leq 1$,
- The magnitude of ρ corresponds to the strength of the relationship.



Values drawn from pairs of correlated random variables, for various values of ρ

Addition of Normal Random Variables

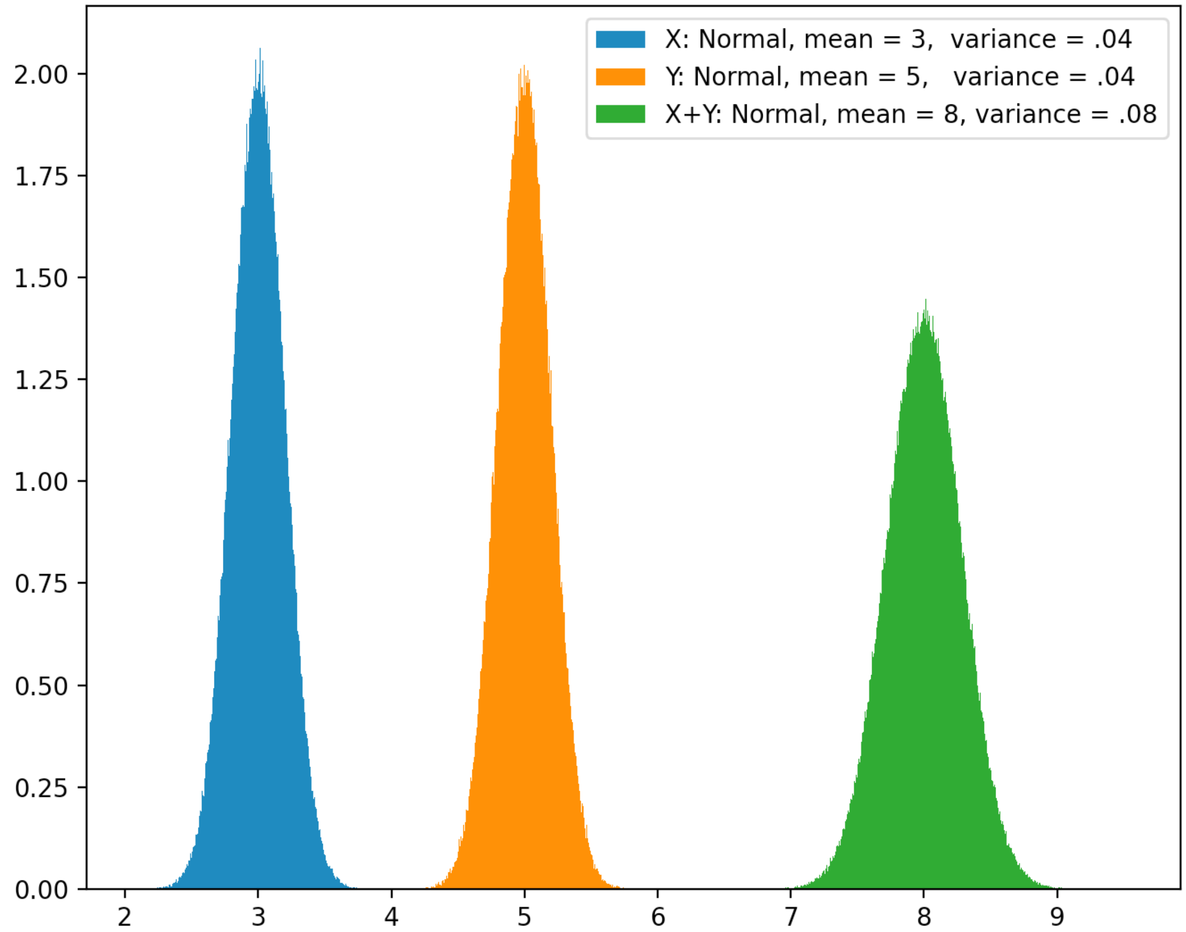
The ***variance*** of the output random variable depends on three values:

- the ***number*** of input random variables,
- the ***variances*** of the input random variables, and
- the ***correlation*** between the input random variables.

Example: Addition of normal random variables

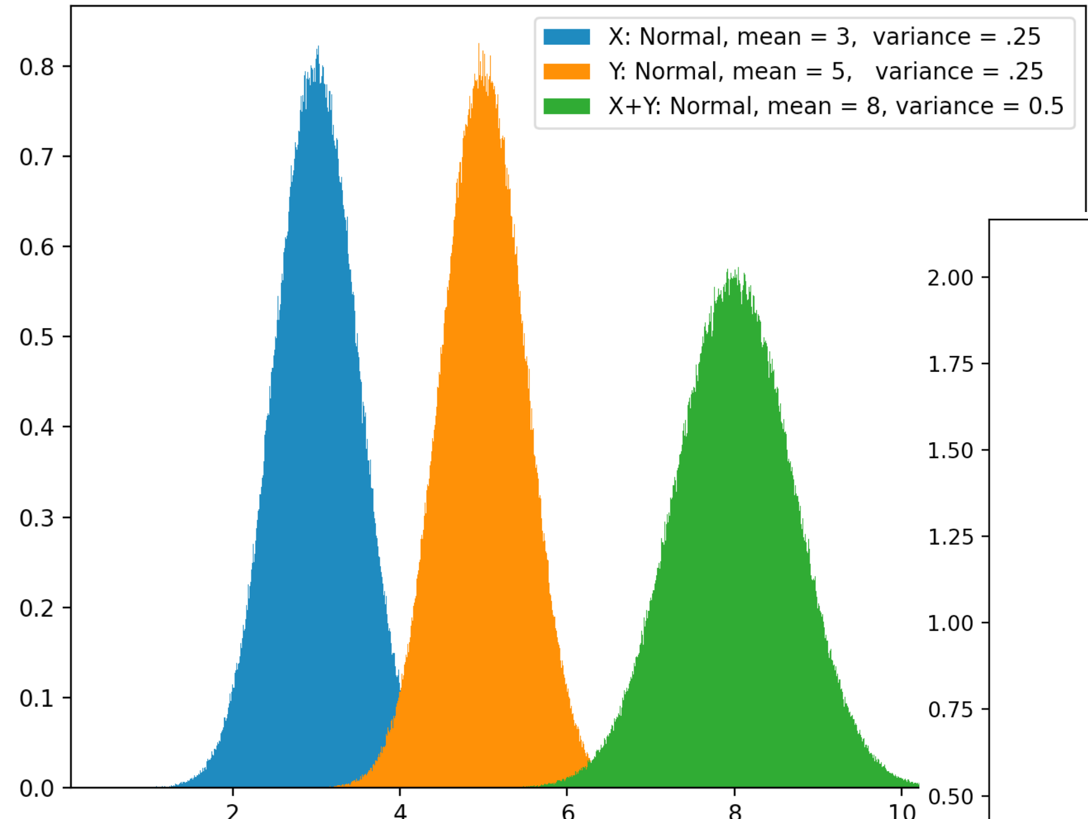
- Number of Inputs: 2
- Variance of input random variables: .04
- Correlation between input random variables: $\rho = 0$

Variance of output random variable: .08



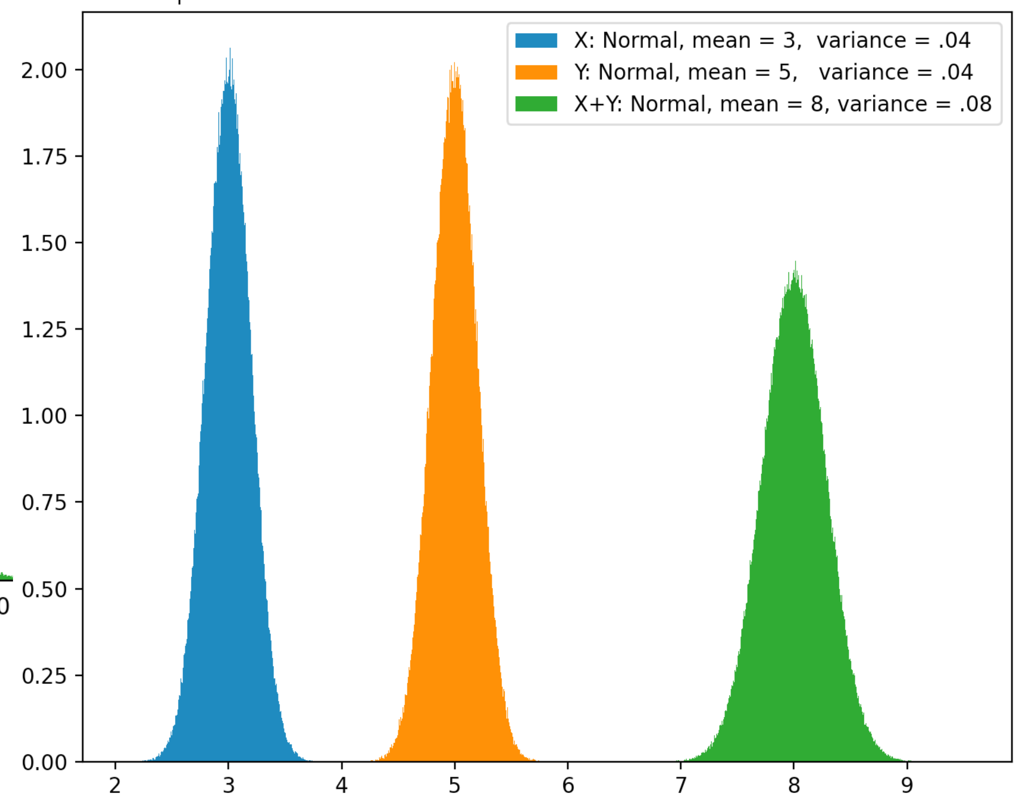
Example: Addition of normal random variables

- Number of Inputs: 2
- Variance of input random variables: .5
- Correlation between input random variables: $\rho = 0$



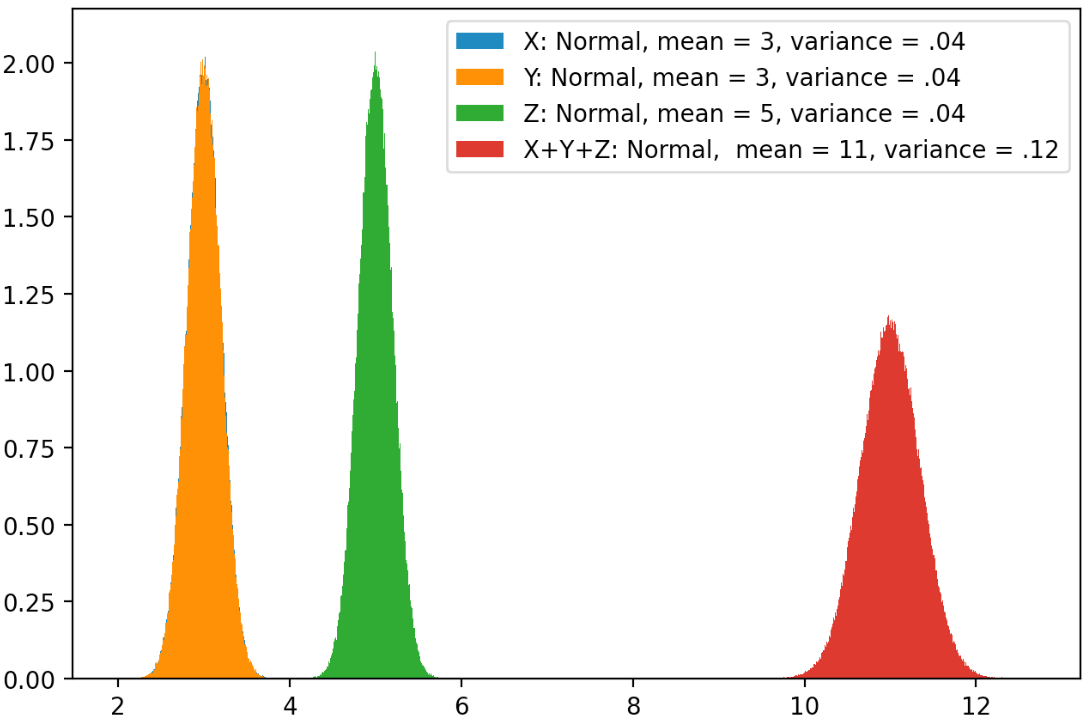
Variance of output random variable: .25

Variance of output random variable: .08



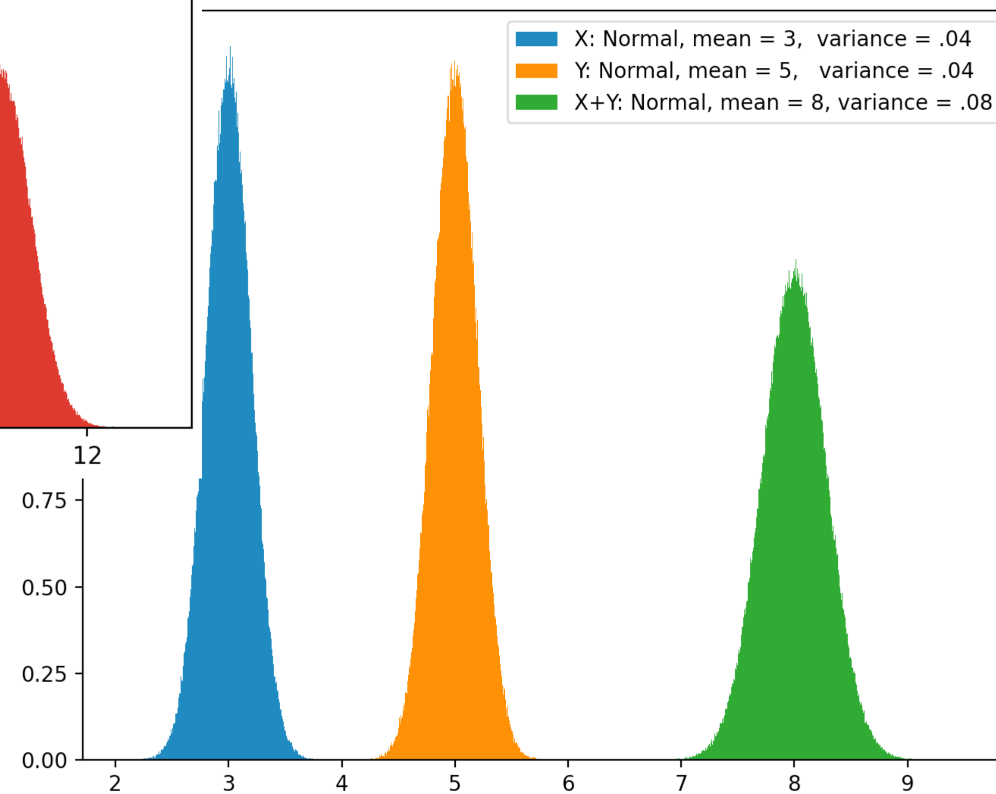
Example: Addition of normal random variables

- Number of inputs: 3
- Variance of input random variables: .04
- Correlation between input random variables: $\rho = 0$



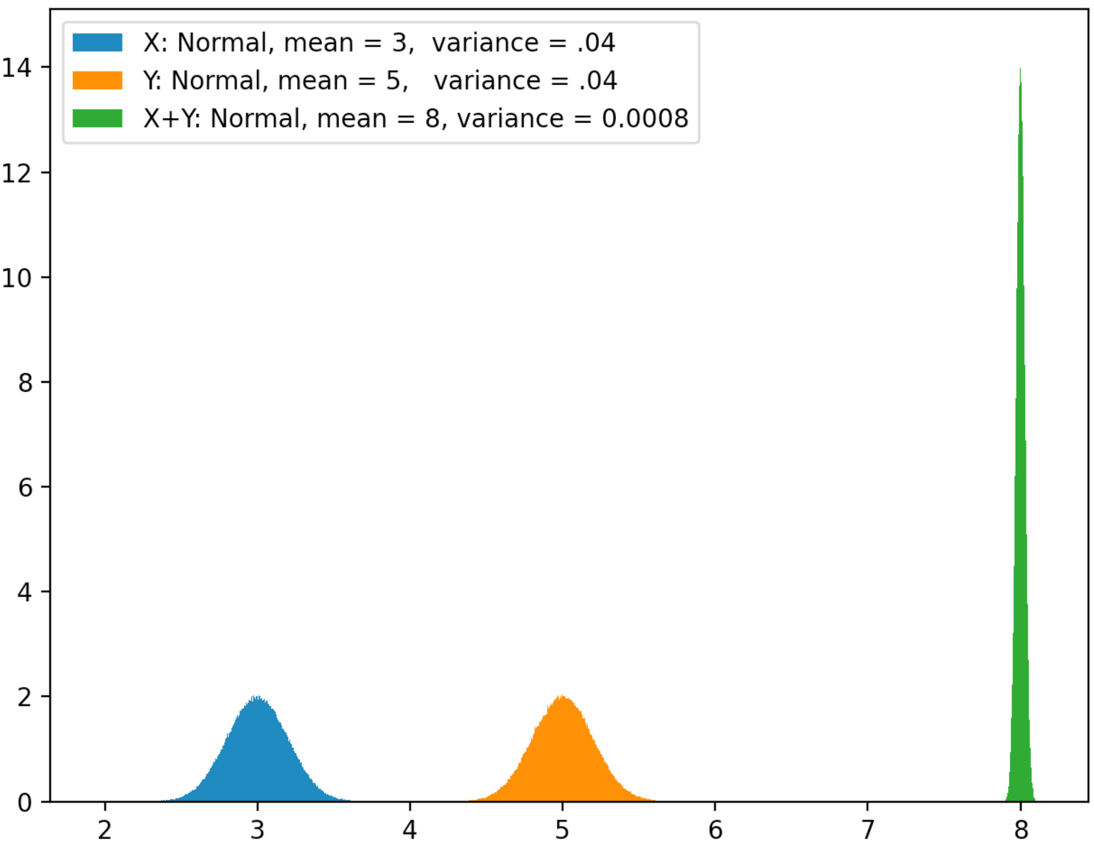
Variance of output random variable: .12

Variance of output random variable: .08



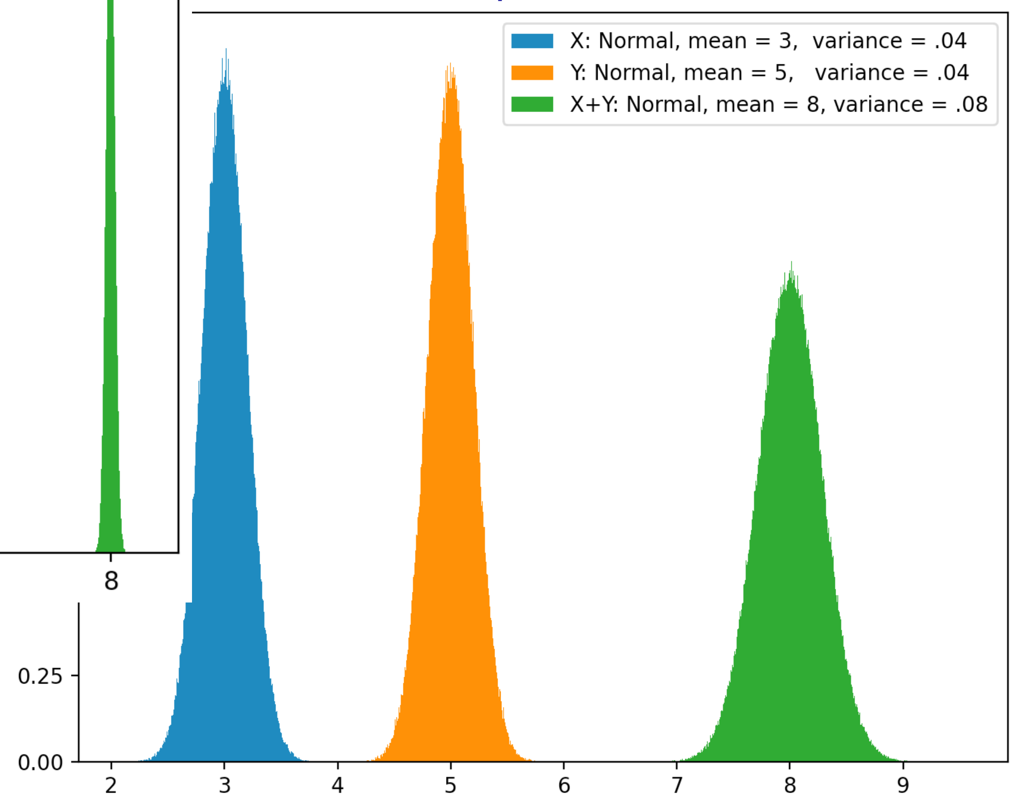
Example: Addition of normal random variables

- Number of inputs: 2
- Variance of input random variables: .04
- Correlation between the input random variables: $\rho = -0.99$



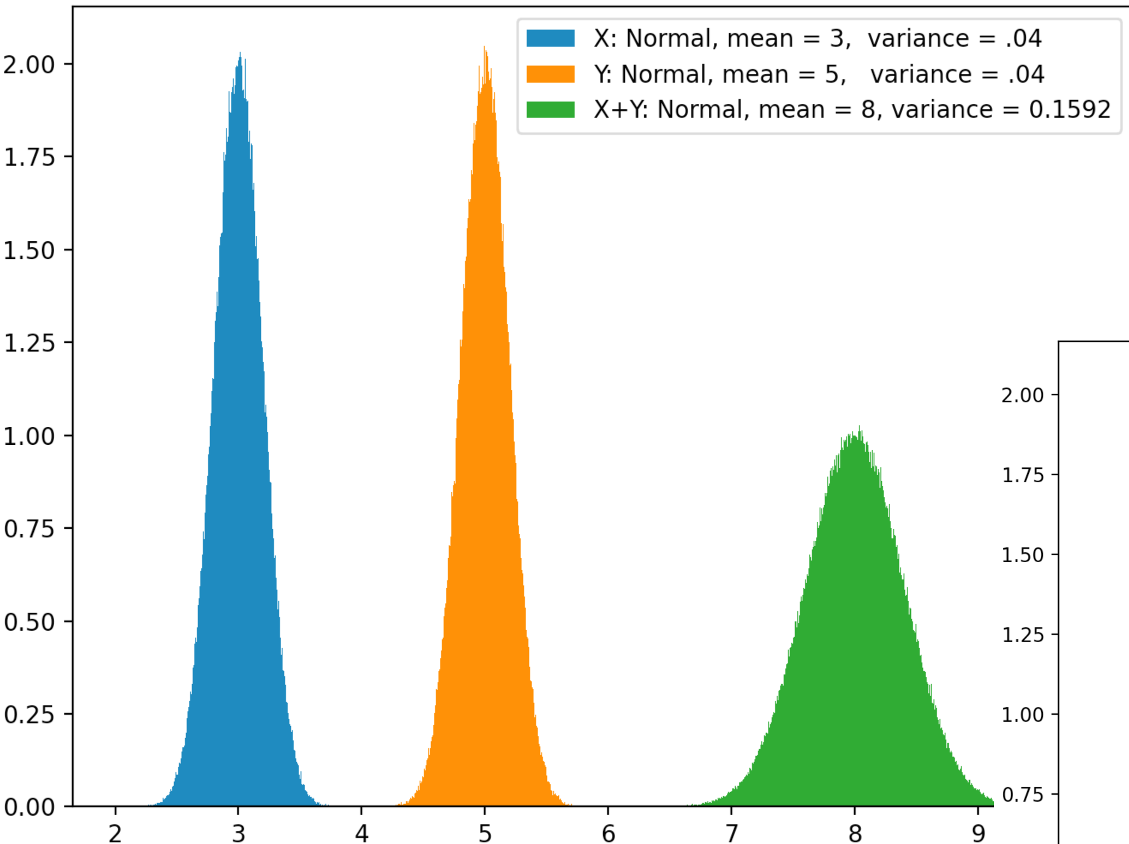
Variance of output random variable: .0008

Variance of output random variable: .08



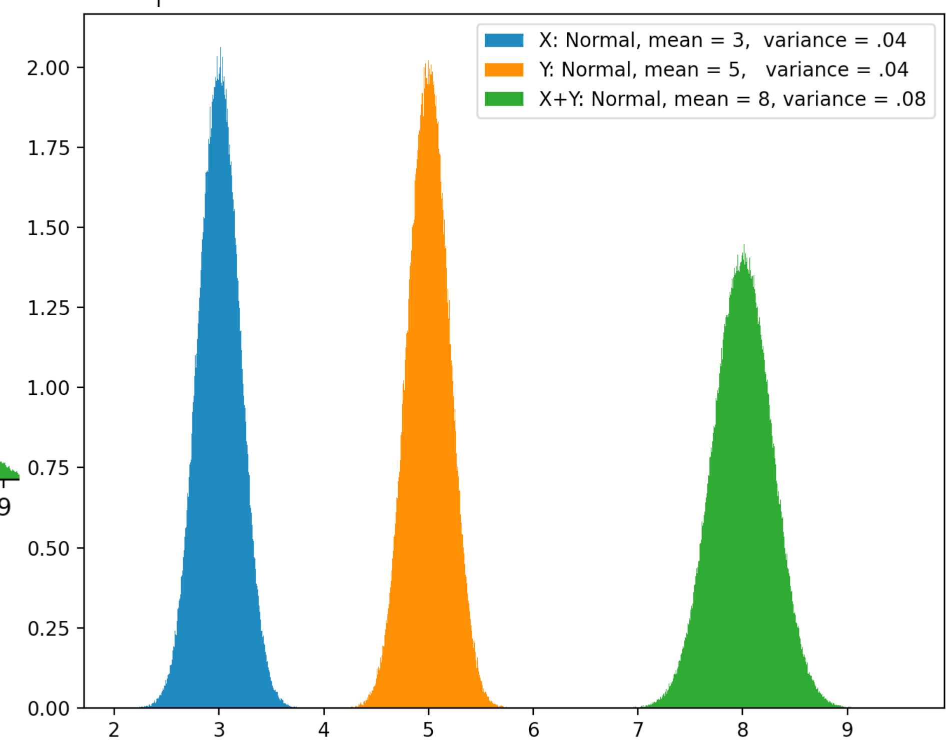
Example: Addition of normal random variables

- Number of inputs: 2
- Variance of input random variables: .04
- Correlation between the input random variables: $\rho = 0.99$



Variance of output random variable: .1592

Variance of output random variable: .08



Variance of a Sum of Normal Random Variables

Suppose that there are k ***normal*** input random variables, $X_1, \ldots, X_k$

- all with equal variance v,
- and such that the linear correlation between each pair is equal to ρ.

Then the variance of the output random variable $Y = X_1 + \cdots + X_k$ is

$$kv + k(k-1)\rho v\,.$$

- If $\rho = -\frac{1}{k}$ the variance of the output random variable is equal to v.

Variance of a Sum of Two Normal Random Variables

Suppose that there are 2 normal random variables:

- with equal variance v
- and with a correlation of ρ.

Then the variance of the output random variable $Y = X_1 + X_2$ is

$$2v\,(1+\rho)\,.$$

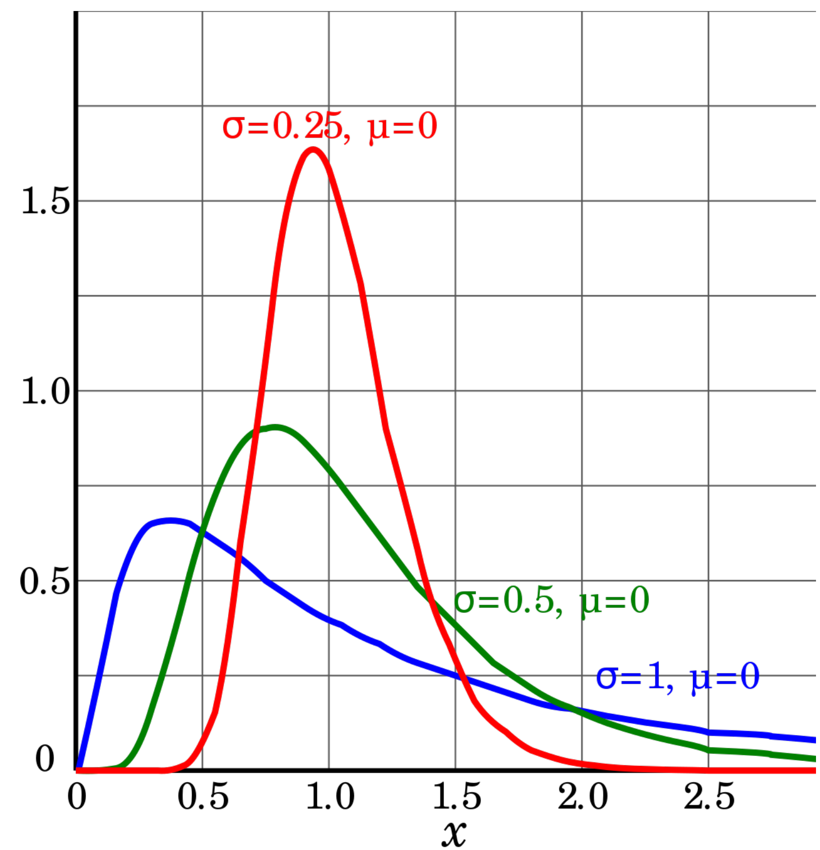
- If $\rho = -\frac{1}{2}$ the variance of the output random variable is equal to v.
- As ρ approaches -1, the variance of the output random variable approaches 0.
- As ρ approaches 1, the variance of the output random variable approaches $4v$.

Variance of a Product of Random Variables

For ***addition*** of ***normal*** random variables, the ***variance*** of the output can be made arbitrarily small.

For ***multiplication***, the same is true if we use the ***lognormal*** random variables and ***geometric variance***.

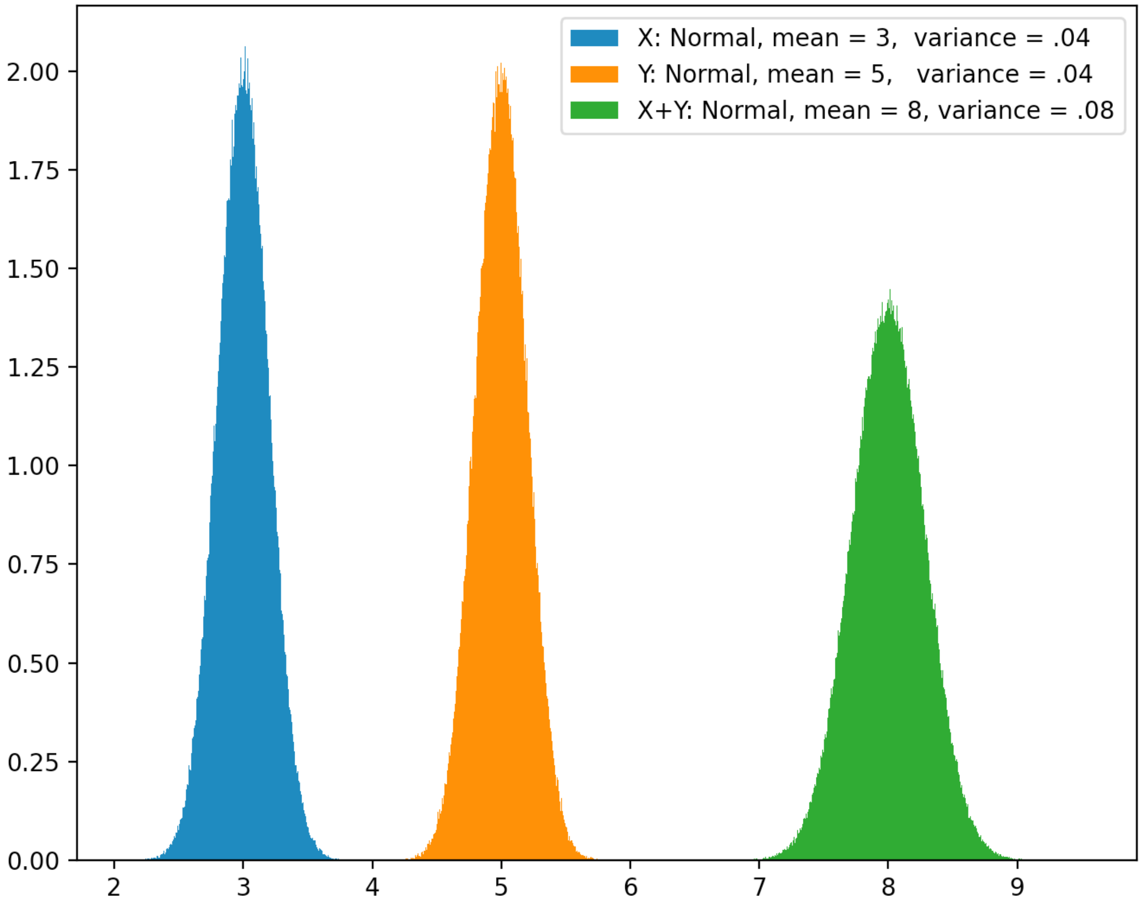
Probability density functions of various *lognormal* random variables.



Accuracy and Measures of Center

In order for any model of nondeterministic computing to be practical:

- We need to be able to predict and control the **accuracy** of the output.



Addition of normal random variables

Mean as a Measure of Center

In order for any model of nondeterministic computing to be practical:

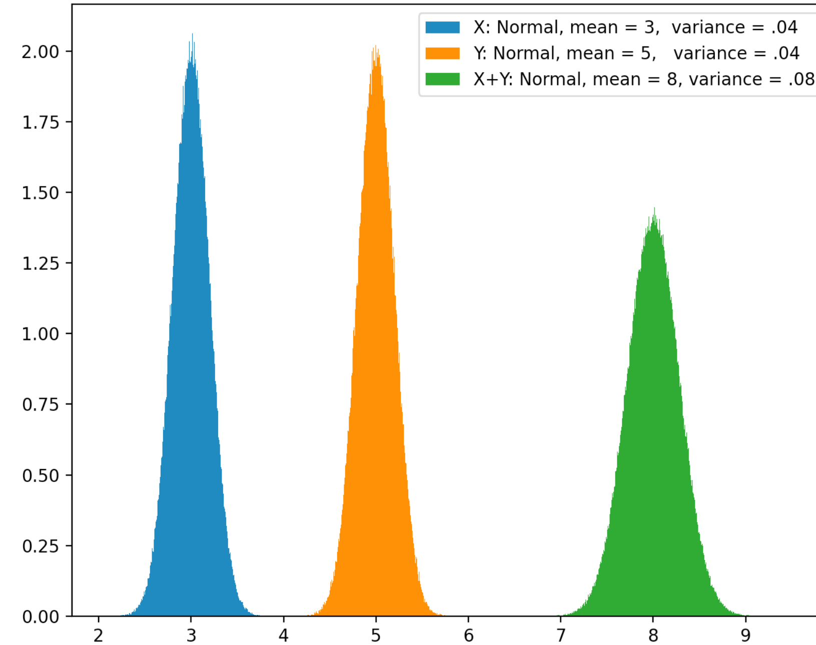
- We need to be able to predict and control the **accuracy** of the output.

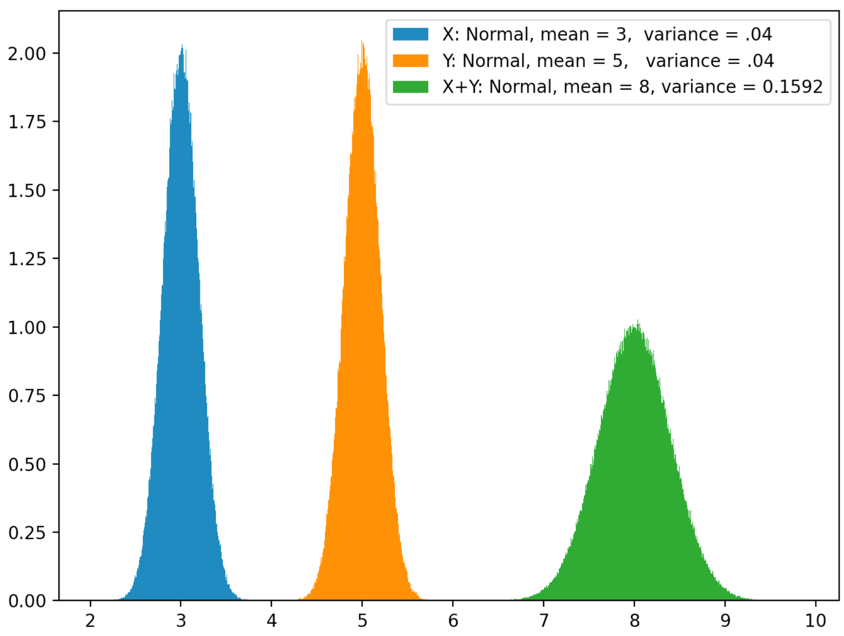
Parameters that measure the accuracy are called **Measures of Center.**

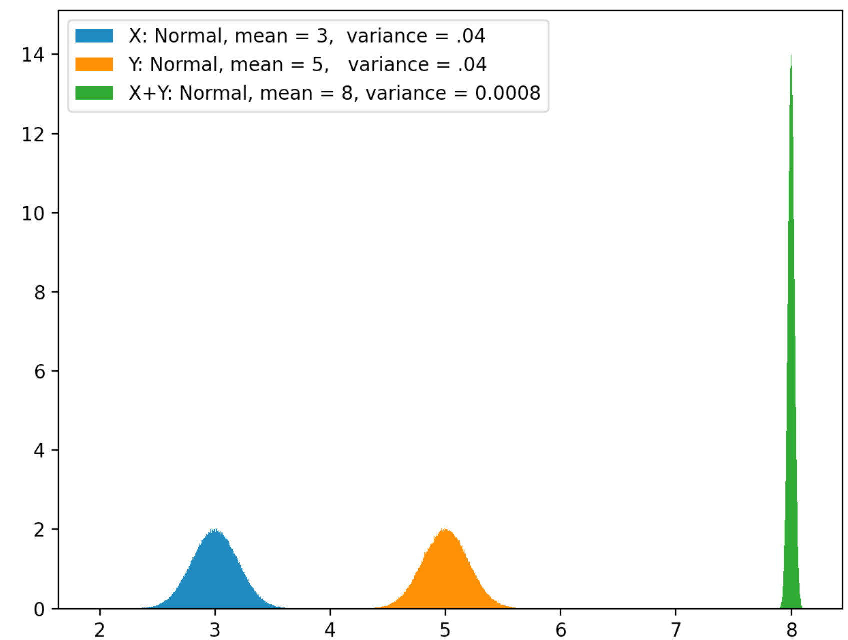
- The **mean** is the most common measure of center.

A Measure of Center that Respects Addition

For ***addition*** of ***normal*** random variables, the ***mean*** is always accurate.







Mean and Multiplication of Normal Random Variables

Suppose X and Y are **normal** random variables

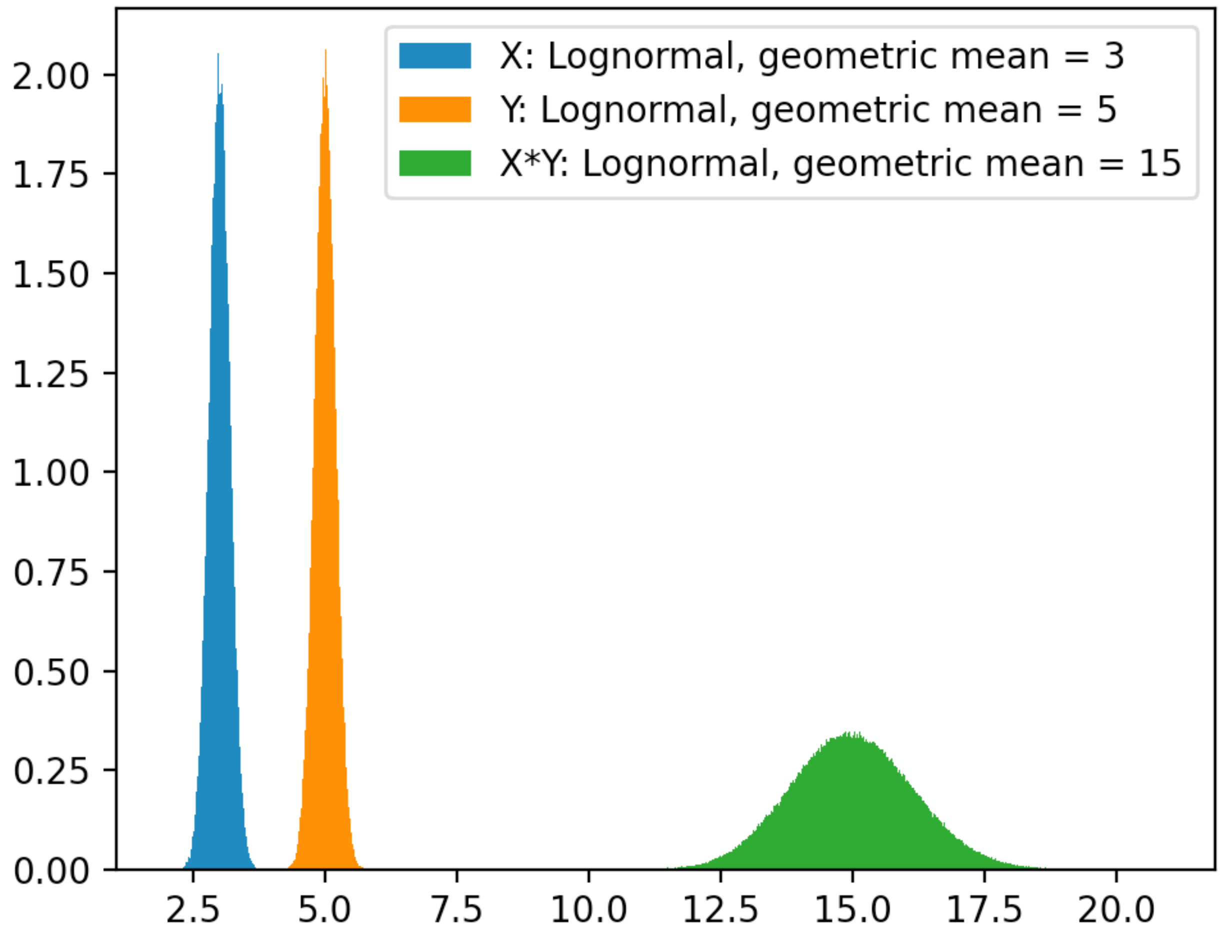
- with means μ_x and μ_y , and
- with variances v_x and v_y .

Then the mean of the product $X * Y$ is

$$1 + \frac{\mu_x \mu_y}{\sqrt{v_x v_y}} .$$

A Measure of Center that Respects Multiplication

For ***multiplication*** of ***lognormal*** random variables, the ***geometric mean*** is always accurate.



Defining the Problem

In order for any model of nondeterministic computing to be practical:

- We need to be able to predict and control the **accuracy** of the output.
- We need to be able to predict and control the **dispersion** of the output.

Goal:

- Perform *arbitrary* operations on input random variables,
- predict and control the *dispersion* and *accuracy* of the output random variable.

Families of Distributions

The distribution of the output random variable might not be the same as the distributions of the input random variables.

- The distribution of the ***product*** of ***normal*** random variables is not normal.
- The distribution of the ***sum*** of ***lognormal*** random variable is not known.

Product of Two Normal Random Variables

Theorem ([Cui+16], pp.1662-1663)

Let X and Y two real Gaussian random variables $X \sim N(\mu_x, \sigma_x)$ and $Y \sim N(\mu_y, \sigma_y)$ with ρ the correlation coefficient. Then the exact PDF $f_Z(z)$ of the product $Z = XY$ is given by:

$$\exp\left\{-\frac{1}{2(1-\rho^2}\left(\frac{\mu_x^2}{\sigma_x^2}+\frac{\mu_y^2}{\sigma_y^2}-\frac{2\rho(x+\mu_x\mu_y)}{\sigma_x\sigma_y}\right)\right\}$$

$$\times\sum_{n=0}^{\infty}\sum_{m=0}^{2n}\frac{x^{2n-m}|x|^{m-n}\sigma_x^{m-n-1}}{\pi(2n)!(1-\rho^2)^{2n+1/2}\sigma_y^{m-n+1}}\left(\frac{\mu_x}{\sigma_x^2}-\frac{\rho\mu_y}{\sigma_x\sigma_y}\right)^m$$

$$\binom{2n}{m}\times\left(\frac{\mu_y}{\sigma_y^2}-\frac{\rho\mu_x}{\sigma_x\sigma_y}\right)^{2n-m}K_{m-n}\left(\frac{|x|}{(1-\rho^2)\sigma_x\sigma_y}\right)$$

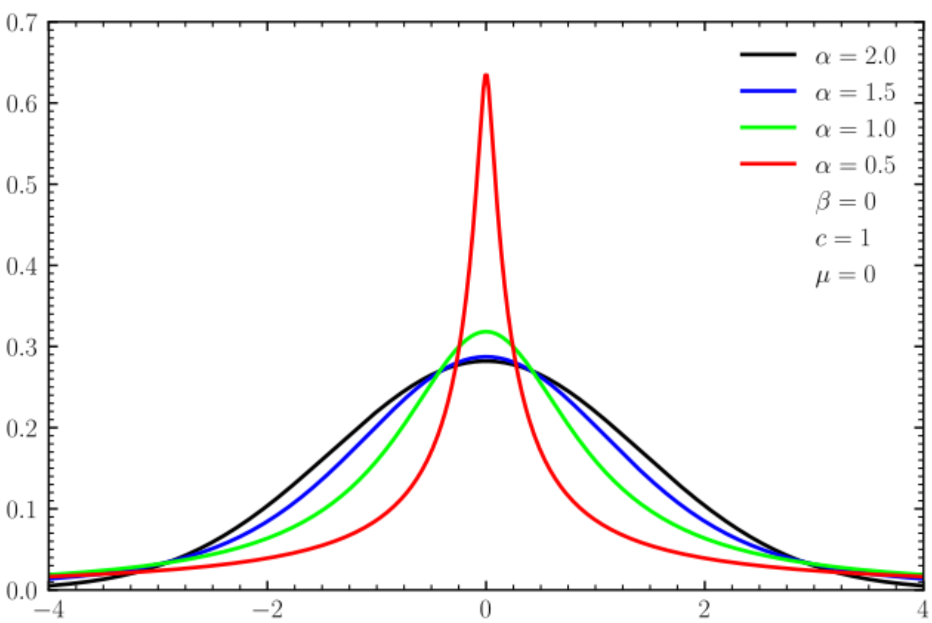
where $K_v(\cdot)$ denotes the modified Bessel function of the second kind and order v.

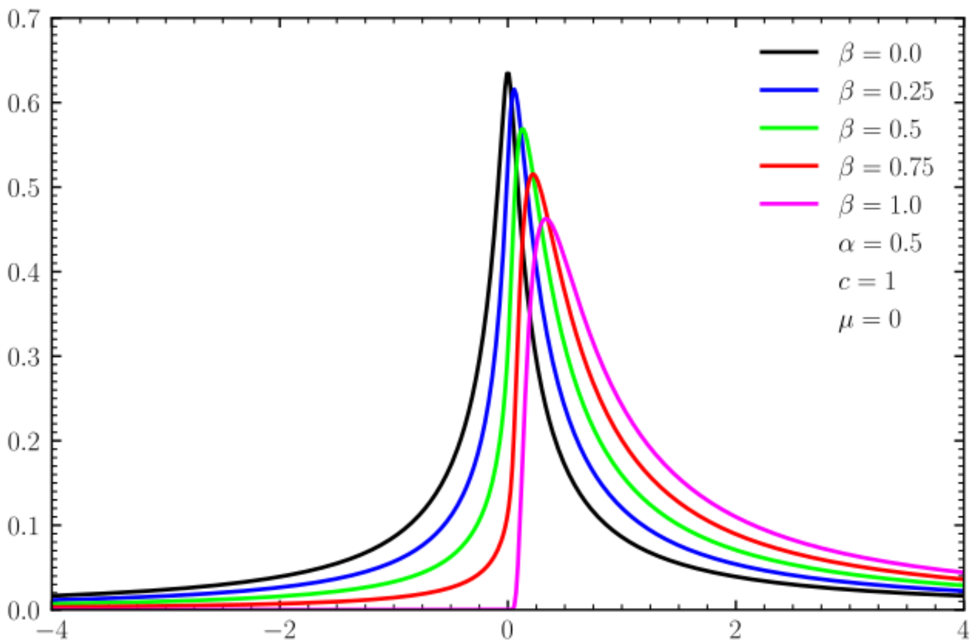
A Question Concerning Distributions

Is there a family of distributions that is closed with respect to addition and multiplication?

Stable Distributions

- **Stable distributions** are a family of distributions that is *closed* with respect to *addition* (of independent random variables).
- Generally, random variables with stable distributions have undefined expectation and infinite variance.





Generalized Gamma Convolutions

The class of **generalized gamma convolutions (GGCs)** is closed with respect to addition and multiplication of *independent* random variables.

A Question Concerning Parameters

Which parameters should we use?

Median

Current Approach:

Measure of center: median

- The ***mean*** of ***normal*** random variables is also the ***median***.
- The ***geometric mean*** of ***lognormal*** random variables is also the ***median***.

Median: Intuitive Definition

- The mean of normal random variables and the geometric mean of lognormal random variables are both also the median of those distributions.

- The median, $\mathscr{M}$, of a random variable is the value such that half of the area of the distribution is above $\mathscr{M}$, and half is below.

Median: Mathematical Definition

- The median, $\mathscr{M}$, of a random variable is the value such that half of the area of the distribution is above $\mathscr{M}$, and half is below.

Suppose X is a continuous random variable, with probability density function $f_X(x)$.

Then the *median* of X is the value $\mathscr{M}$ such that

$$\int_{\mathscr{M}}^{\infty} f_X(x)\,dx = \int_{-\infty}^{\mathscr{M}} f_X(x)\,dx = \frac{1}{2}$$

Median Absolute Deviation: Intuitive Definition

Current Approach:

Measure of dispersion: median absolute deviation (MAD)

- The median absolute deviation, λ, of a random variable with median $\mathscr{M}$ is the value such that half of the distribution lies in the interval

$$(\mathscr{M} - \lambda, \mathscr{M} + \lambda)\,.$$

Median Absolute Deviation: Mathematical Definition

- The median absolute deviation, λ, of a random variable is the value such that half of the distribution lies in the interval $(\mathscr{M} - \lambda, \mathscr{M} + \lambda)$.

Suppose X is a continuous random variable, with probability density function $f_X(x)$ and median $\mathscr{M}$.

Then the *median absolute deviation* of X is the value λ such that

$$\lambda = \text{median}\left(\,|X - \mathscr{M}|\,\right).$$

In other words,

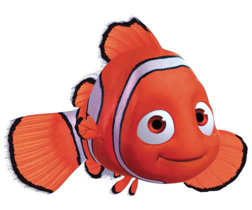
$$\int_{\mathscr{M}-\lambda}^{\mathscr{M}+\lambda} f_X(x)\,dx = \frac{1}{2}\,.$$

Current Choice of Parameters

Measure of center: median

Measure of dispersion: median absolute deviation (MAD)

- Every random variable has a median and a median absolute deviation.



The Question of Parameters

Which parameters should we use?

Current Approach:

- ***Measure of center:*** *median*
- ***Measure of dispersion:*** *median absolute deviation (MAD)*

Questions Concerning Dependence

How do we model dependence?

Are there limits on the possible dependency structures?

How does the dependence structure of the input random variables affect the median and median absolute deviation of the output random variable?

Modeling Dependence

How do we model dependence?

A **joint probability distribution**

- is the probability distribution of multiple random variables,
- it contains the dependence structure of the random variables.

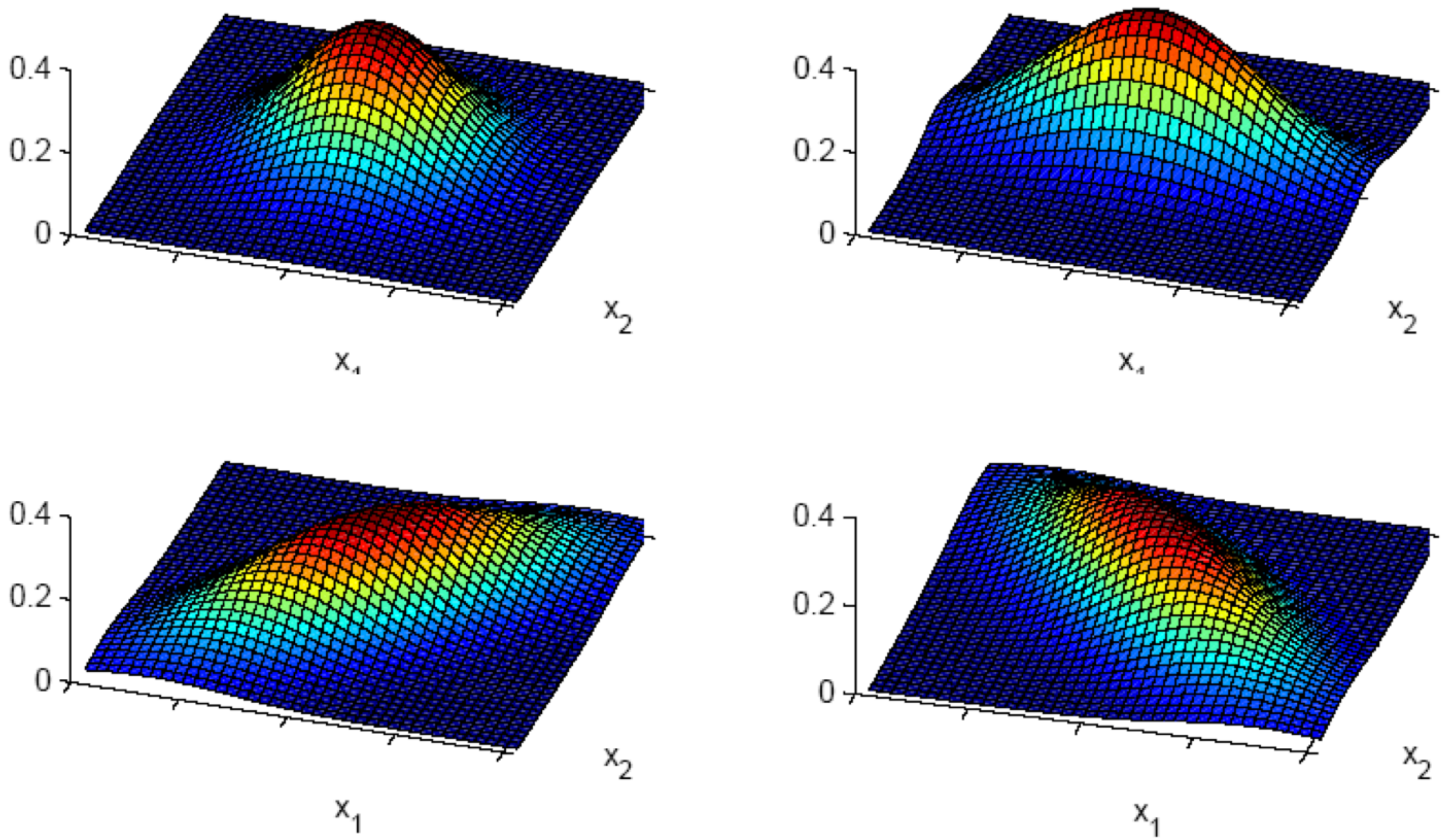
The **marginals** are the individual random variables in the joint distribution.

Examples: Linear Dependence

Two Dimensional Joint Probability Density Functions with Normal Marginals

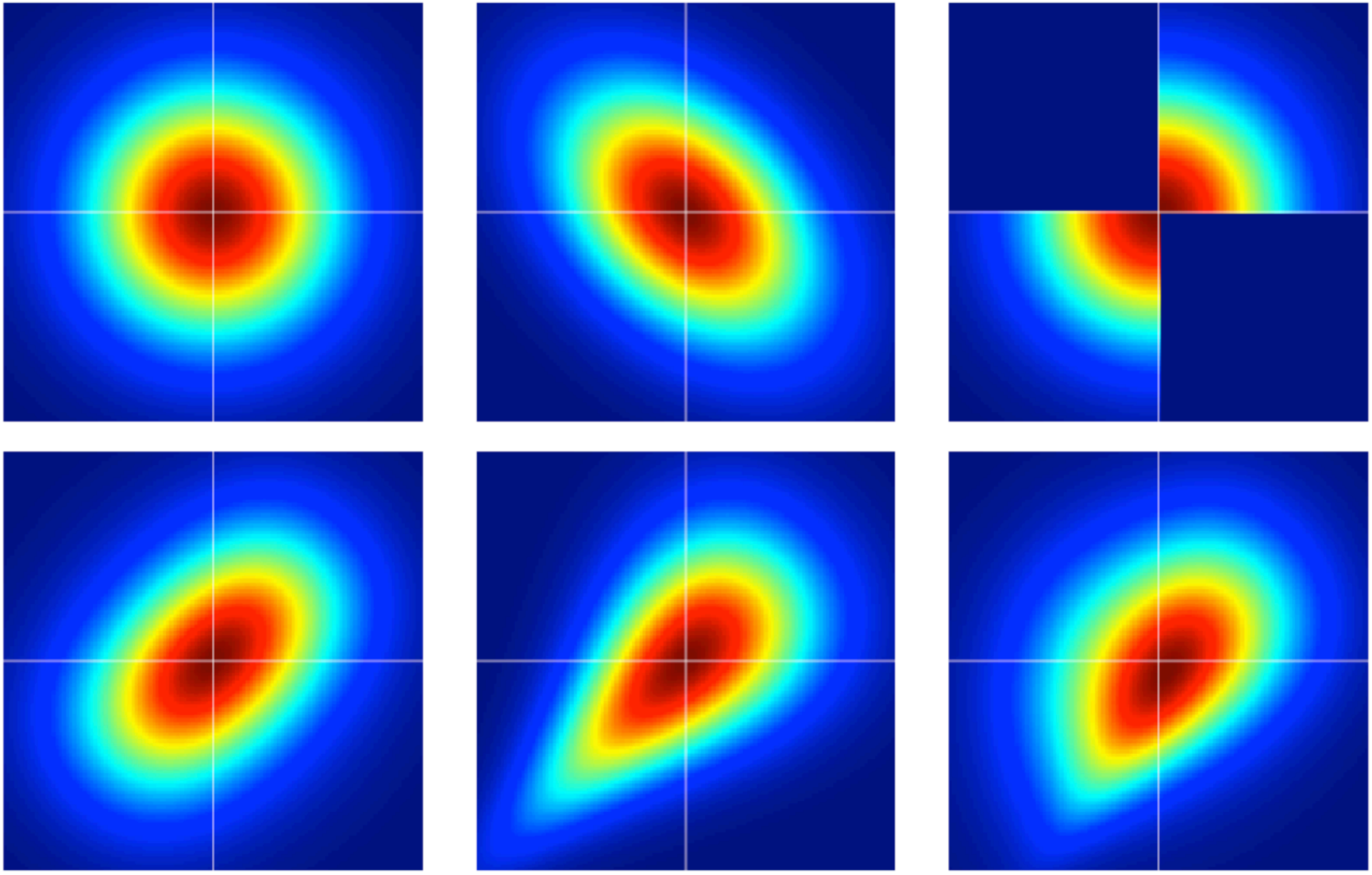
Top row: independent, equal variance;
independent, unequal variance

2'nd row: positive linear dependence, equal variance;
negative linear dependence, equal variance



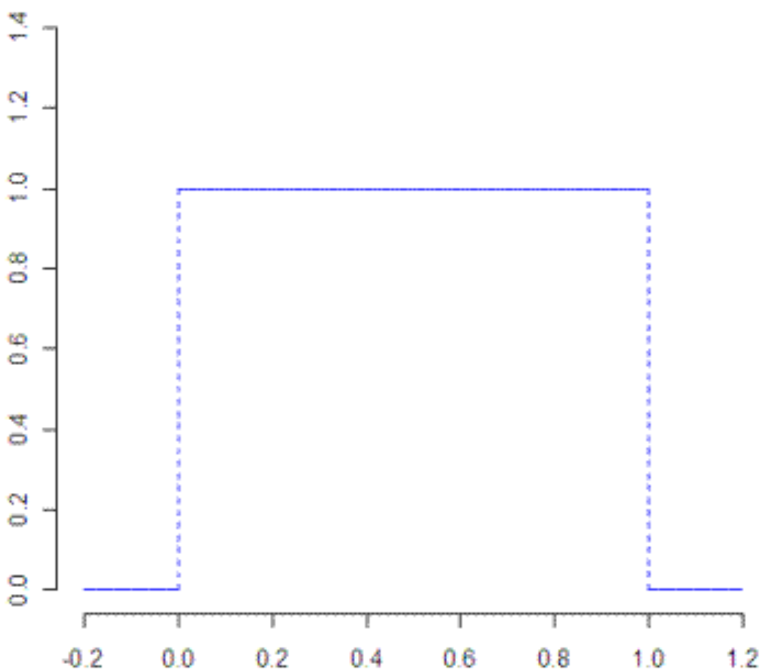
Examples: Nonlinear Dependence

Two Dimensional Joint Probability Density Functions with Normal Marginals

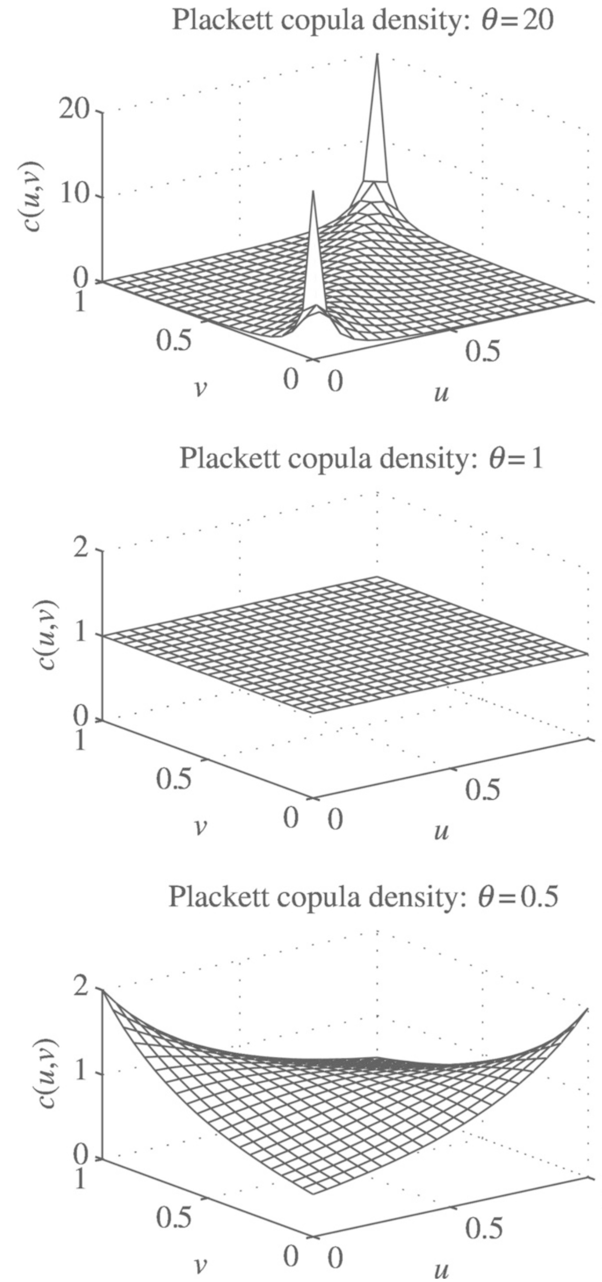


Copulas: What?

Copulas are joint probability distributions of marginals that are uniform on the interval [0,1].



The probability density function of a random variable that is uniform on the interval [0,1]



Copulas: How?

Sklar's theorem:

Every joint probability distribution can be decomposed into a copula and a set of marginals.

Copulas: How?

If $X_1, \ldots, X_d$ are random variables with a joint probability distribution

$$h(x_1, \ldots, x_d),$$

then

$$h(x_1, \ldots, x_d) = c(F_1(x_1), \ldots, F_d(x_d)) \cdot f_1(x_1) \, \cdots \, f_d(x_d),$$

where

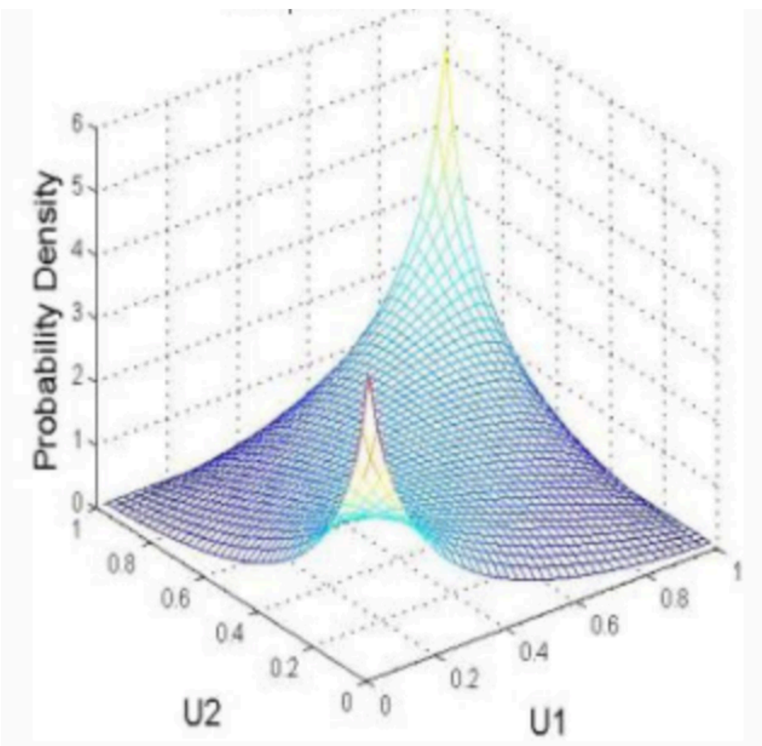
- c is the probability density function of a copula,
- $f_1, \ldots, f_d$ are the probability density function of the marginals,
- and $F_1, \ldots, F_d$ are the **cumulative probability distributions** of the marginals, defined by

$$F_i(x) = \int_{-\infty}^{x} f_i(t)\, dt, \quad 1 \leq i \leq d\,.$$

Gaussian Copula: What?

Probability density function of the Gaussian copula:

$$c(t_1, \ldots, t_d) = \frac{1}{\sqrt{\det R}} \exp\left(-\frac{1}{2} \begin{bmatrix} \varphi^{-1}(t_1) \\ \vdots \\ \varphi^{-1}(t_d) \end{bmatrix}^T (R^{-1} - I) \begin{bmatrix} \varphi^{-1}(t_1) \\ \vdots \\ \varphi^{-1}(t_d) \end{bmatrix} \right)$$



Probability density function of 2 dimensional Gaussian copula.

Gaussian Copula: How?

Multivariate distribution with Gaussian Copula:

$$h(x_1, \ldots, x_d) = \frac{1}{\sqrt{\det R}} f_1(x) \cdots f_d(x) \times \exp\left(-\frac{1}{2} \begin{bmatrix} \varphi^{-1}(F_1^{-1}(x_1)) \\ \vdots \\ \varphi^{-1}(F_d^{-1}(x_d)) \end{bmatrix}^T (R^{-1} - I) \begin{bmatrix} \varphi^{-1}(F_1^{-1}(x_1)) \\ \vdots \\ \varphi^{-1}(F_d^{-1}(x_d)) \end{bmatrix} \right)$$

- *Specifying the copula:* φ is the cumulative distribution of the normal distribution with mean 0 and variance 1.
- *Specifying the marginals:* The $F_i's$ are the cumulative distributions of the marginals.
- *Specifying the dependency structure:* R is a positive definite matrix.

Copulas: Why?

Copulas give us the ability to create joint probability distributions with *any* set of marginals.

The Gaussian copula gives us a “domain” of dependency structures (i.e., the set of positive definite matrices).

Questions Concerning Dependence

How do we model dependence?

Copulas!

Current Approach: *Gaussian copulas.*

Are there limits on the possible dependency structures?

How does the dependence structure of the inputs affect the median and MAD of the output random variable?

Are there limits on the possible dependencies?

Gaussian Copulas:

$$c(t_1, \ldots, t_d) = \frac{1}{\sqrt{\det R}} \exp\left(-\frac{1}{2} \begin{bmatrix} \varphi^{-1}(t_1) \\ \vdots \\ \varphi^{-1}(t_d) \end{bmatrix}^T (R^{-1} - I) \begin{bmatrix} \varphi^{-1}(t_1) \\ \vdots \\ \varphi^{-1}(t_d) \end{bmatrix} \right)$$

Specifying the dependency structure: R is a positive definite matrix with unit diagonal.

$$R = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1d} \\ \rho_{12} & 1 & \cdots & \rho_{2d} \\ \vdots & & & \vdots \\ \rho_{1d} & \rho_{2d} & \cdots & 1 \end{bmatrix}$$

Linear Dependence Constraints in 2 Dimensions

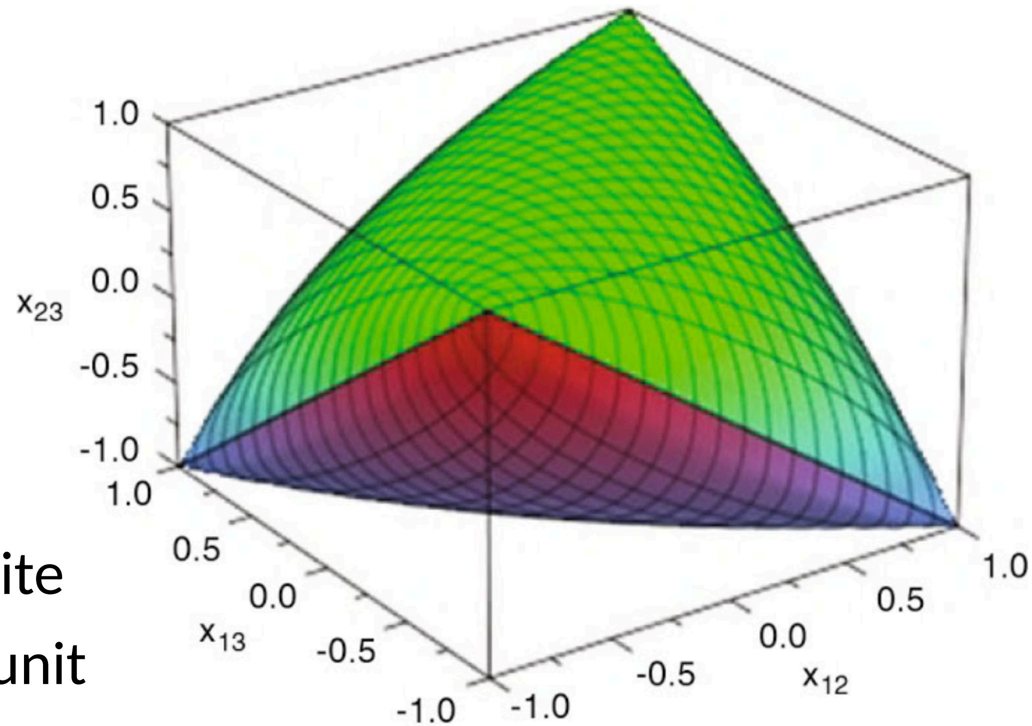
$$R = \begin{bmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{bmatrix}$$

The constraint imposed by positive definiteness of R:

$$-1 < \rho_{12} < 1.$$

Linear Dependence Constraints in 3 Dimensions

$$R = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$$

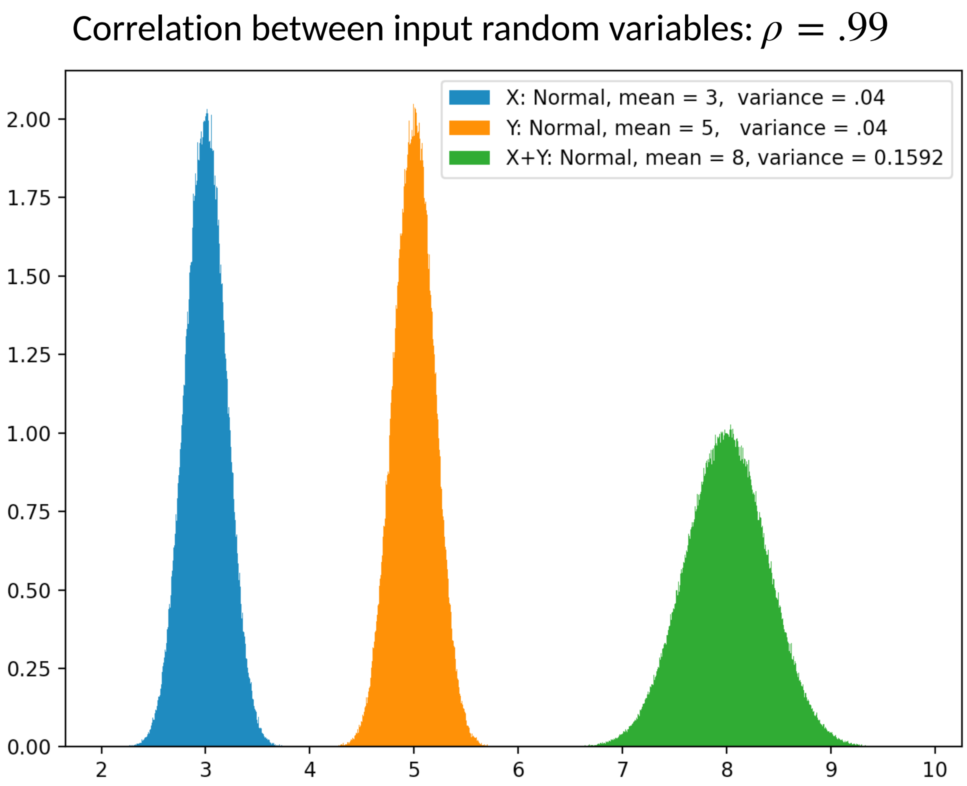


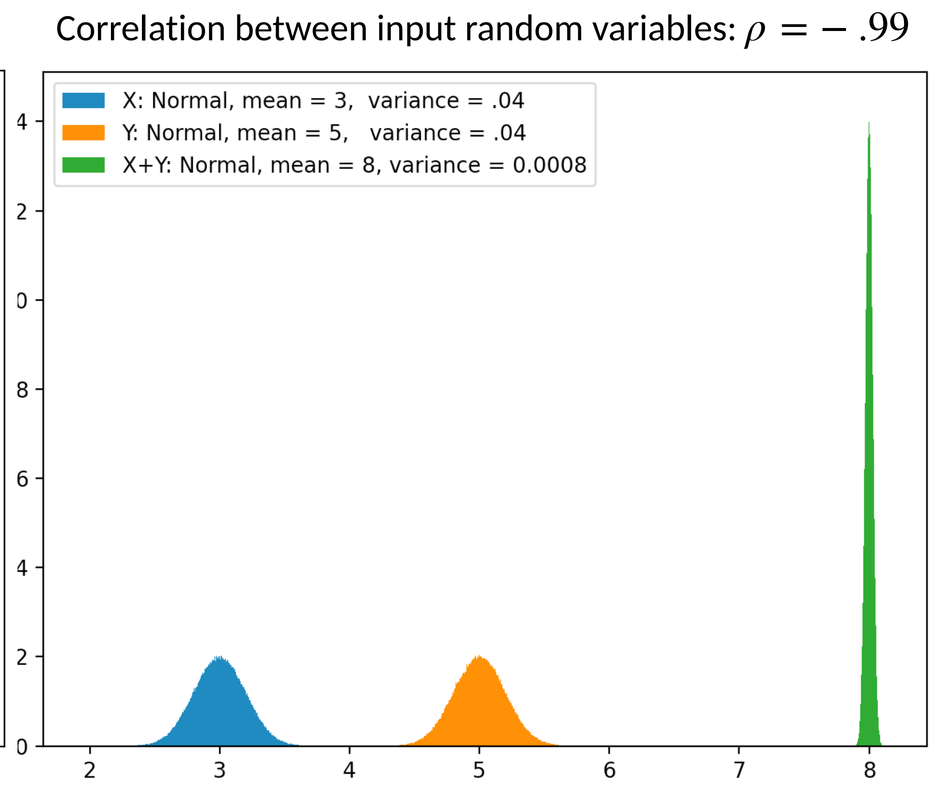
The space of positive (semi)definite matrices defines a subset of the unit cube called the **elliptope.**

Examples of correlation matrices that are not positive definite:

$$\begin{bmatrix} 1 & -.5 & -.5 \\ -.5 & 1 & -.5 \\ -.5 & -.5 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & -.2 & -.5 \\ -.2 & 1 & -.8 \\ -.5 & -.8 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & .2 & -.9 \\ .2 & 1 & -.9 \\ -.9 & -.9 & 1 \end{bmatrix}$$

Addition of normal random variables:





Open Questions Concerning Dependence

Are there limits on the possible dependency structures?

Yes.

Do these constraints on the possible dependency structures effect our ability to control the dispersion and accuracy of the output random variable?

Would focusing on a different type of dependency (e.g., nonlinear) remove these constraints?

Would using a different copula remove these constraints?

Questions Concerning Dependence

How do we model dependence? *Copulas!*

Are there limits on the possible dependency structures?

Yes.

How does the dependence structure of the inputs affect the median and median absolute deviation of the output random variable?

Ideal Effect of Dependence

How does the dependence structure of the inputs affect the median and MAD of the output random variable?

The Ideal:

- The dependence structure of the inputs has no effect on the median (*accuracy*) of the output random variable.
- The dependence structure of the inputs do affect the median absolute deviation (*dispersion*) of the output random variable — specifically, it enables us to minimize that median absolute deviation.

SNAPDRAGON is a simulation framework to explore random variable arithmetic.

SNAPDRAGON: Why?

Goal: Perform arbitrary operations on input random variables and be able to predict and control the dispersion and accuracy of the output random variable.

- In general, median and median absolute deviation don't have simple mathematical expressions.
- Output random variables don't always have known distributions.
- The domain of dependency structures is complicated.

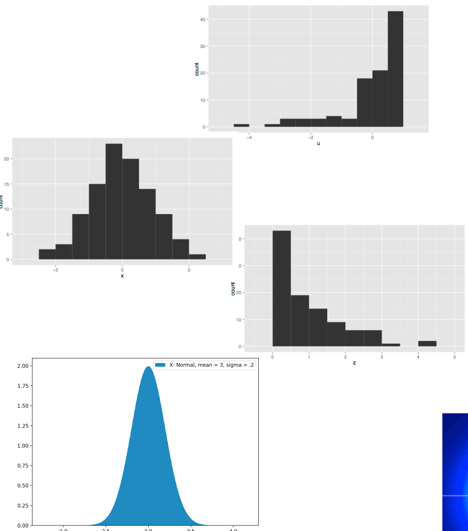
SNAPDRAGON: How?

SNAPDRAGON enables us to

- perform arbitrary operations on input random variables with arbitrary probability distributions,
- measure the median and median absolute deviation of the output random variable, and
- minimize median absolute deviation over the dependency structures (and various other parameters).

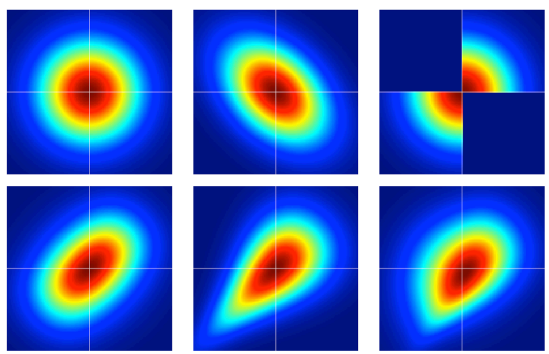
SNAPDRAGON: What?

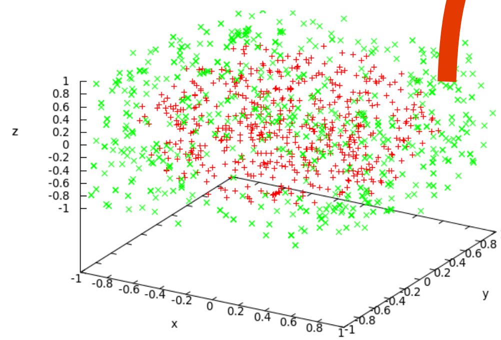
SNAPDRAGON is a flexible framework of components and interfaces



STEP #1: CREATE INPUT RANDOM VARIABLES

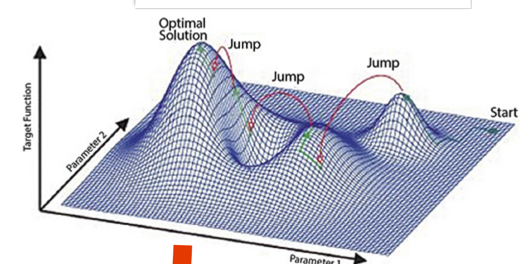
STEP #2: USE COPULAS TO MERGE INPUTS TO CREATE JOINT DISTRIBUTION

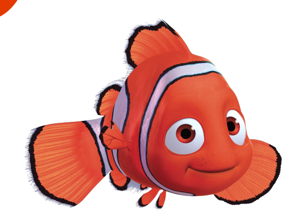




STEP #3: USE NUMERICAL INTEGRATION TO COMPUTE OUTPUT DISTRIBUTION

STEP #4: FIND DEPENDENCY STRUCTURES TO CONTROL AND OPTIMIZE OUTPUT PARAMETERS





Questions Concerning Dependence

How does the dependence structure of the inputs affect the median and median absolute deviation of the output random variable?

The Question of Parameters

Which parameters should we use?

Current Approach:

- *Measure of center*: median
- *Measure of dispersion*: median absolute deviation

The Question Concerning Distributions

Is there a family of distributions that is closed under addition and multiplication?

Open Questions Concerning Dependence

Are there limits on the possible dependency structures?

Yes.

Do these constraints on the possible dependency structures effect our ability to control the dispersion and accuracy of the output random variable?

Would focusing on a different type of dependency remove these constraints?

Would using a different copula remove these constraints?

Revisiting the Dependency Structures Question

- Consider the deterministic problem: $a*b+c*d+e*f$
- Represent every value with a random variable: $A*B+C*D+E*F$

Approach 1:

- Create a 6 dimensional joint distribution,
- perform all 5 operations in one step.

Approach 2:

- Subdivide the problem,
- create various lower dimensional joint distributions,
- perform multiple operations.

Revisiting the Dependency Structures Constraints

Approach 1: Create a single joint distribution and perform all operations in one step.

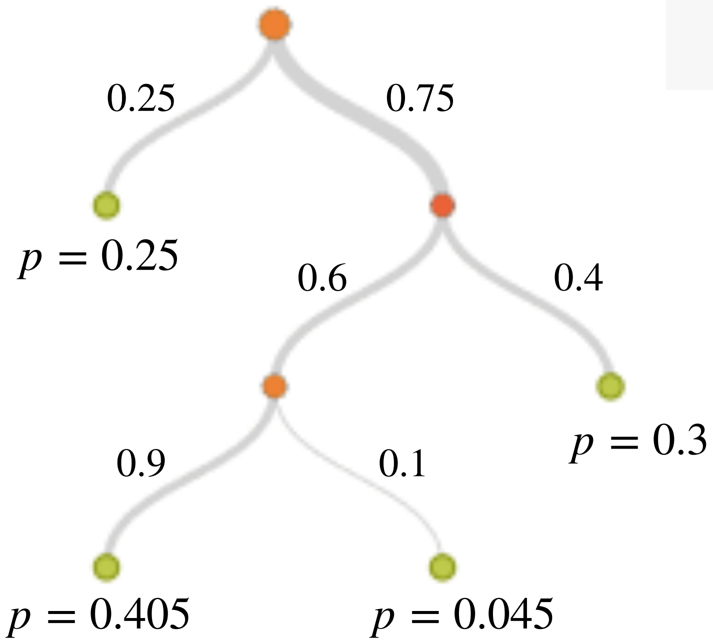
Approach 2: Subdivide the problem, create lower dimensional joint distributions, and perform multiple operations.

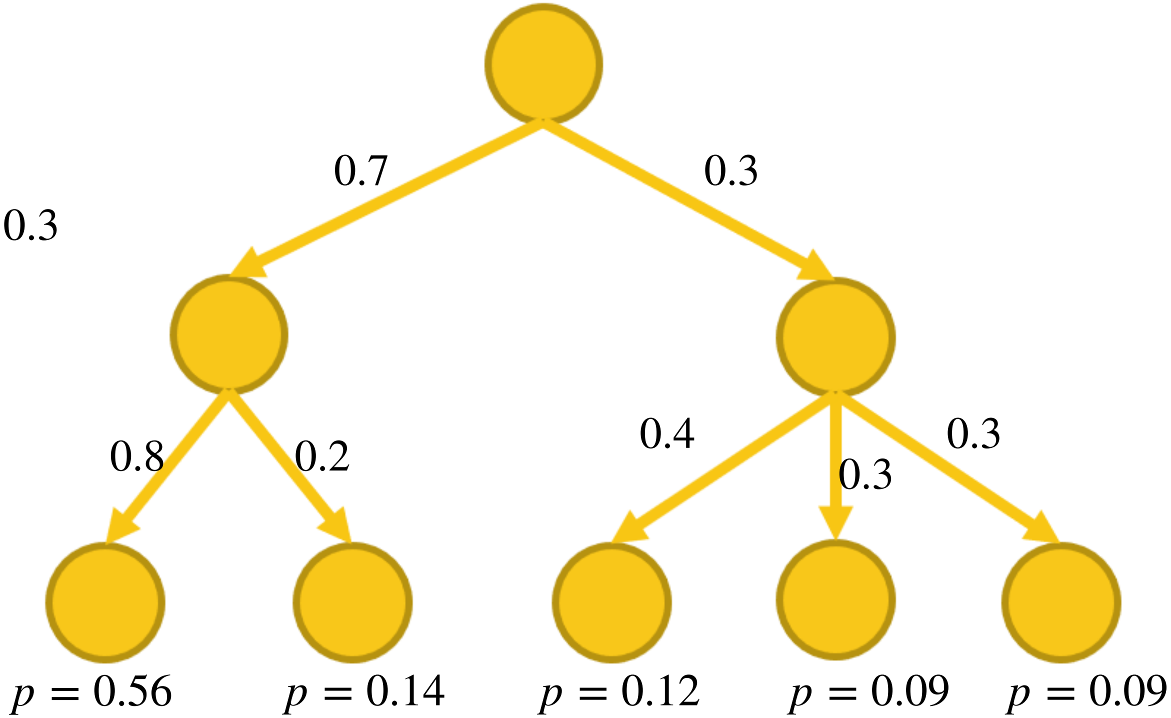
Which approach allows us to better control the dispersion and accuracy of the output random variable?

The Question Behind this Project

Can we conceive a mechanism whereby we can recover the correct solution to a deterministic problem with some fixed probability even if there are incorrect intermediate results along the way?

Is there a formal mapping between random variable arithmetic and another model of computation, e.g., a probabilistic Turing machine?





The Four Stages of a Project

Mathematical Theory

Simulation

Emulation

Computer Hardware

Random Variable Arithmetic

Mathematical Theory

Random Variable Arithmetic
Copula Theory
Probability Theory

Simulation

SNAPDRAGON

Emulation

Computer Hardware

Future Work

How can we design hardware that operates using the random variable arithmetic framework?

That's All Folks!

Thank you!