Simulation Input Data Modeling

OSMAN BALCI
Professor

Department of Computer Science
Virginia Polytechnic Institute and State University (Virginia Tech)
Blacksburg, VA 24061, USA

https://manta.cs.vt.edu/balci

Review of Probability & Statistics

What is a Random Variable?

- "A random variable is a mathematical function that maps outcomes of random experiments to numbers.
- It can be thought of as the numeric result of operating a non-deterministic mechanism or performing a non-deterministic experiment to generate a random result.
- For example, a random variable can be used to describe the process of rolling a fair dice and the possible outcomes { 1, 2, 3, 4, 5, 6 }.
- Another random variable might describe the possible outcomes of picking a random person and measuring his or her height."
- See http://en.wikipedia.org/wiki/Random_variable for more information.

Review of Probability & Statistics

What is a Probability Distribution?

- Every random variable gives rise to a probability distribution.
- If `X` is a random variable, the corresponding probability distribution assigns to the interval `[a, b]` the probability `Pr [a ≤ X ≤ b]`, i.e. the probability that the variable `X` will take a value in the interval `[a, b]`.
- The probability distribution of the random variable `X` can be uniquely described by its cumulative distribution function `F(x)`, which is defined as

```
F(x) = Pr [X <= x]
```

  where `x` is a particular value (variate) of the random variable `X`.

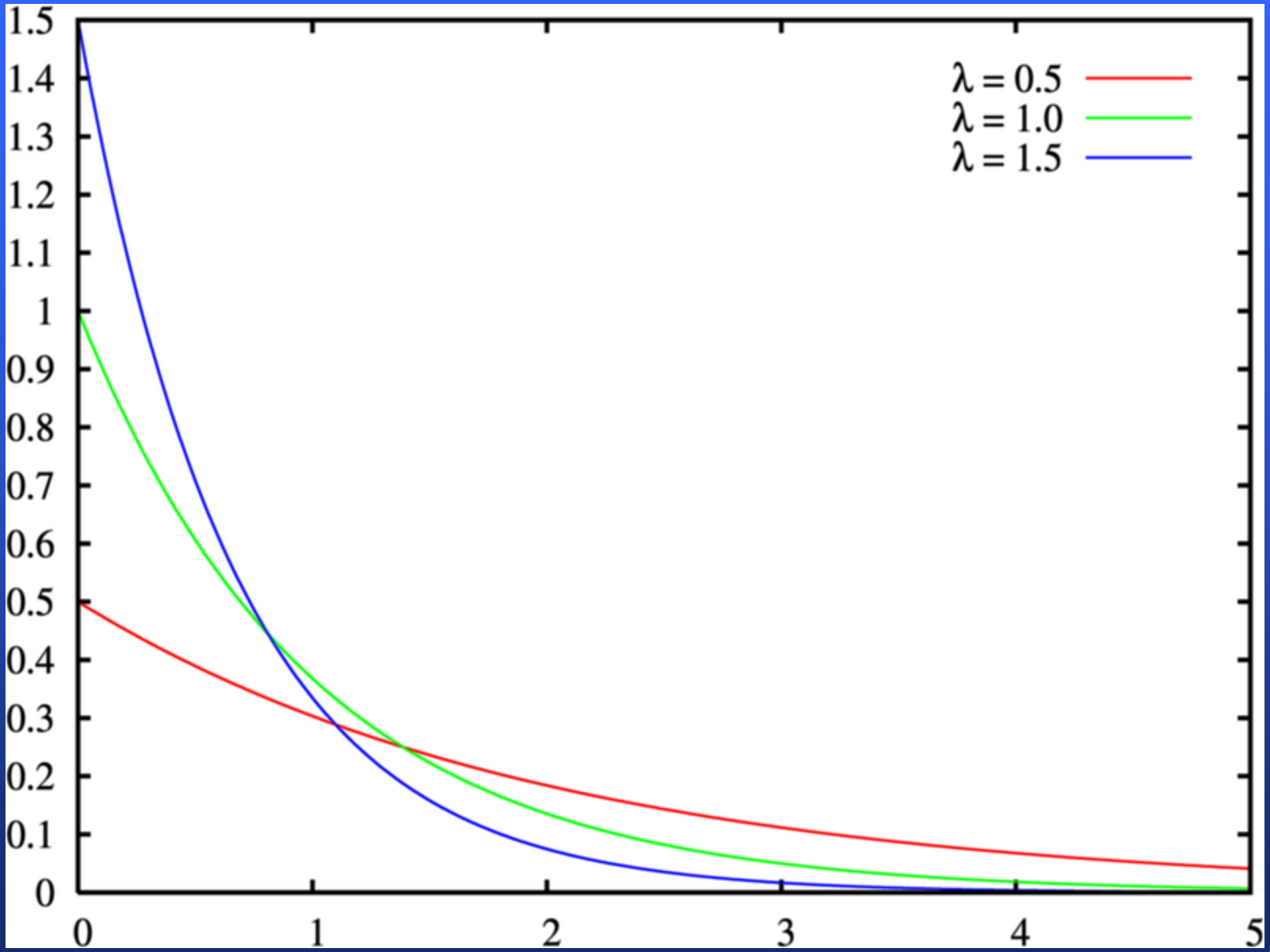
- See http://en.wikipedia.org/wiki/Probability_distribution for more information.

Review of Probability & Statistics

Click here to see a comprehensive list of probability distributions.

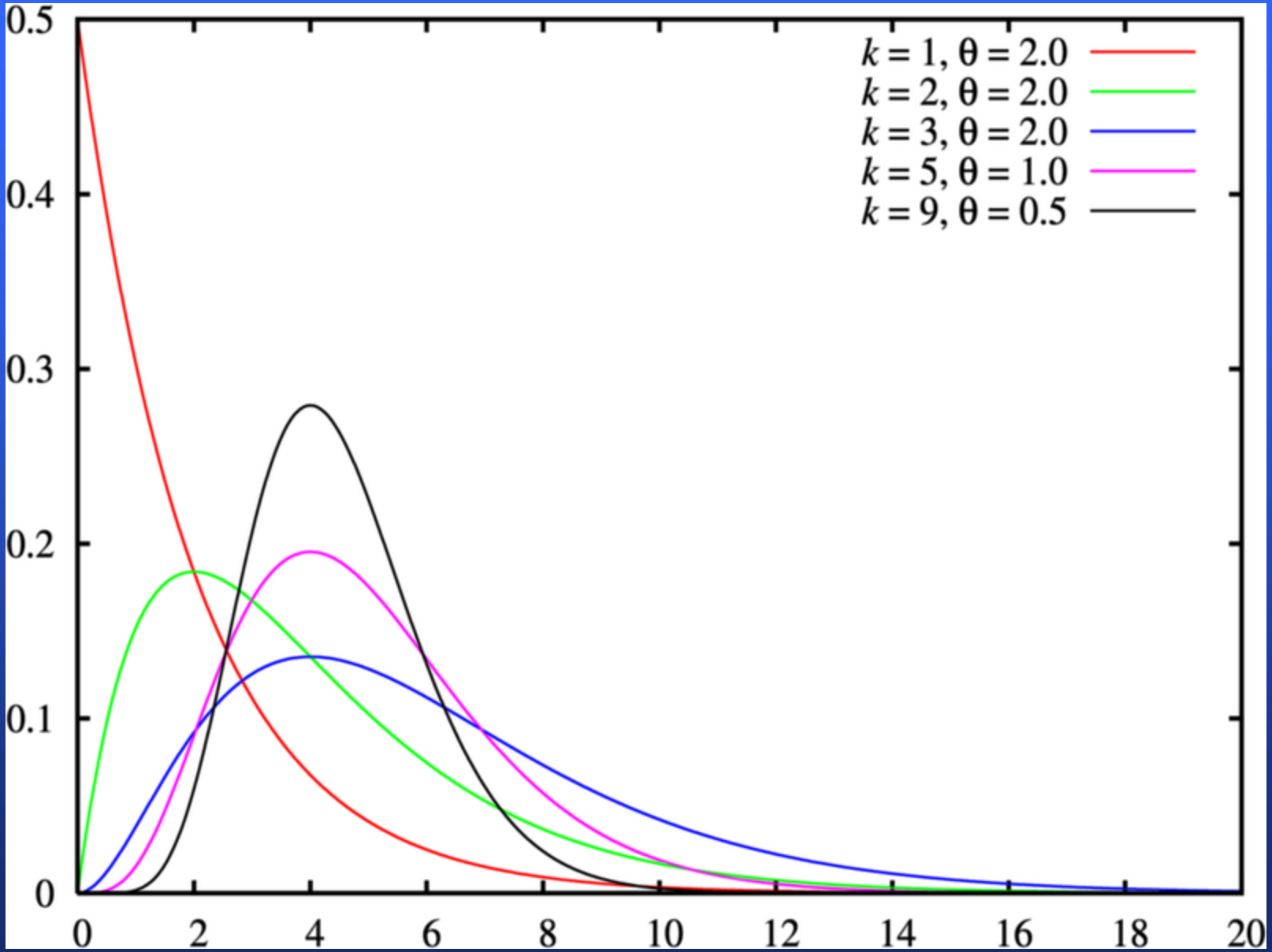
Category	Probability Distribution
Non-negative Continuous	Exponential
	Gamma
	Inverse Gaussian
	Inverted Weibull
	Log-Laplace
	Log-Logistic
	Lognormal
	Pareto
	Pearson Type V
	Pearson Type VI
	Random Walk
	Weibull
Bounded Continuous	Beta
	Johnson SB
	Triangular
	Uniform
Unbounded Continuous	Extreme Value Type A
	Extreme Value Type B
	Johnson SU
	Laplace
	Logistic
	Normal
Discrete	Binomial
	Discrete Uniform
	Geometric
	Negative Binomial
	Poisson

Exponential Probability Distribution



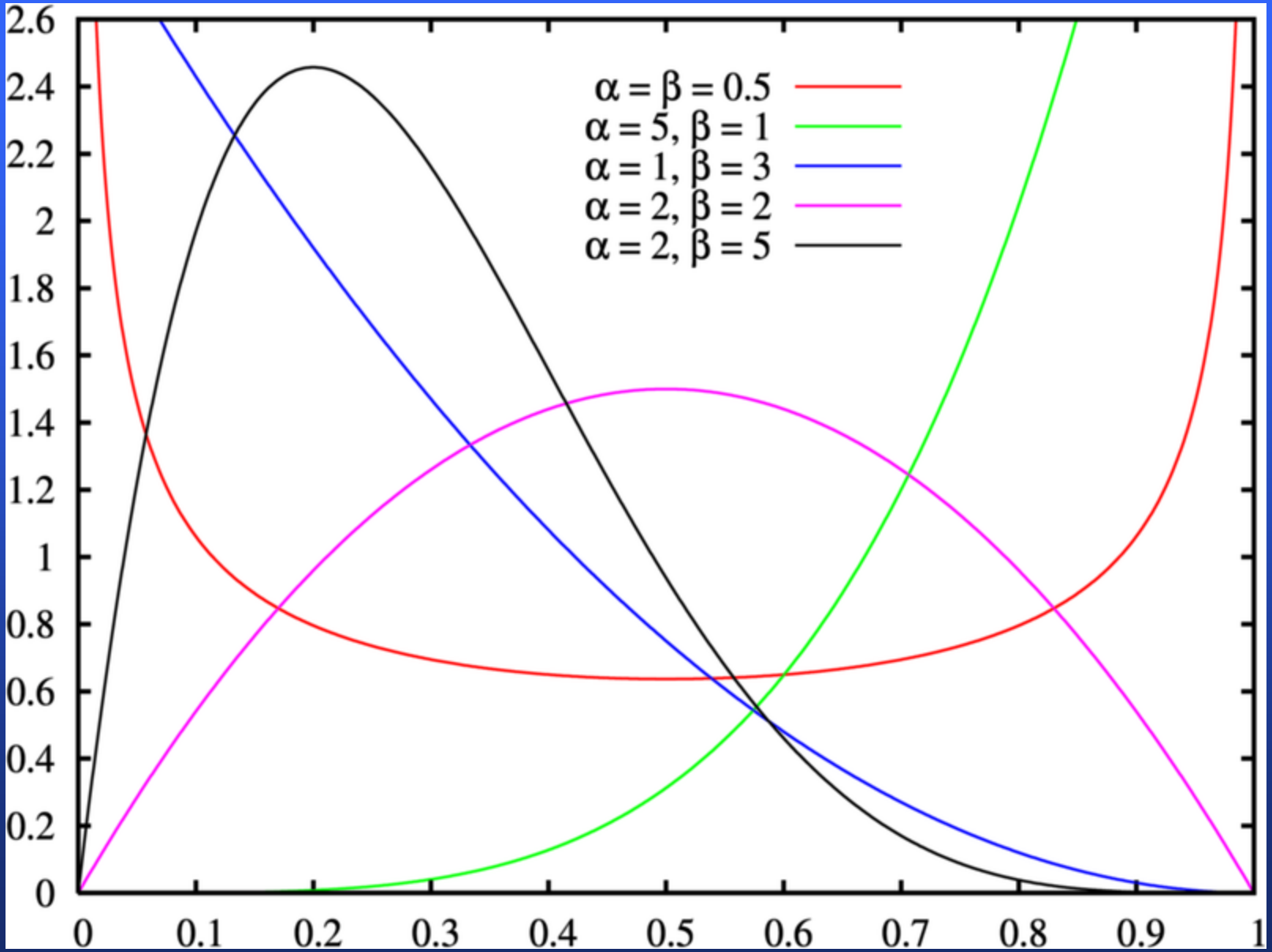
See http://en.wikipedia.org/wiki/Exponential_probability_distribution for more information.

Gamma Probability Distribution



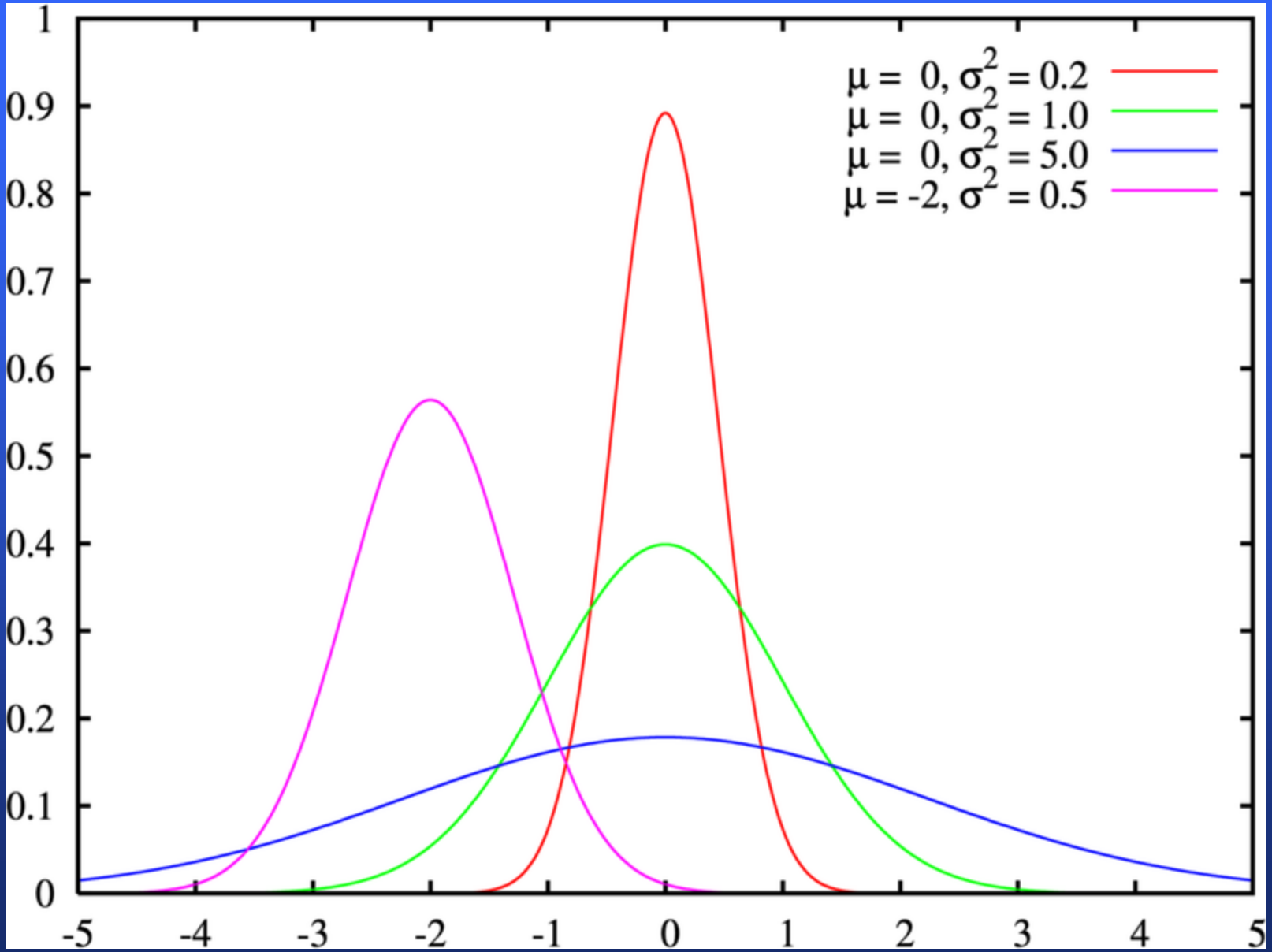
See http://en.wikipedia.org/wiki/Gamma_distribution for more information.

Beta Probability Distribution



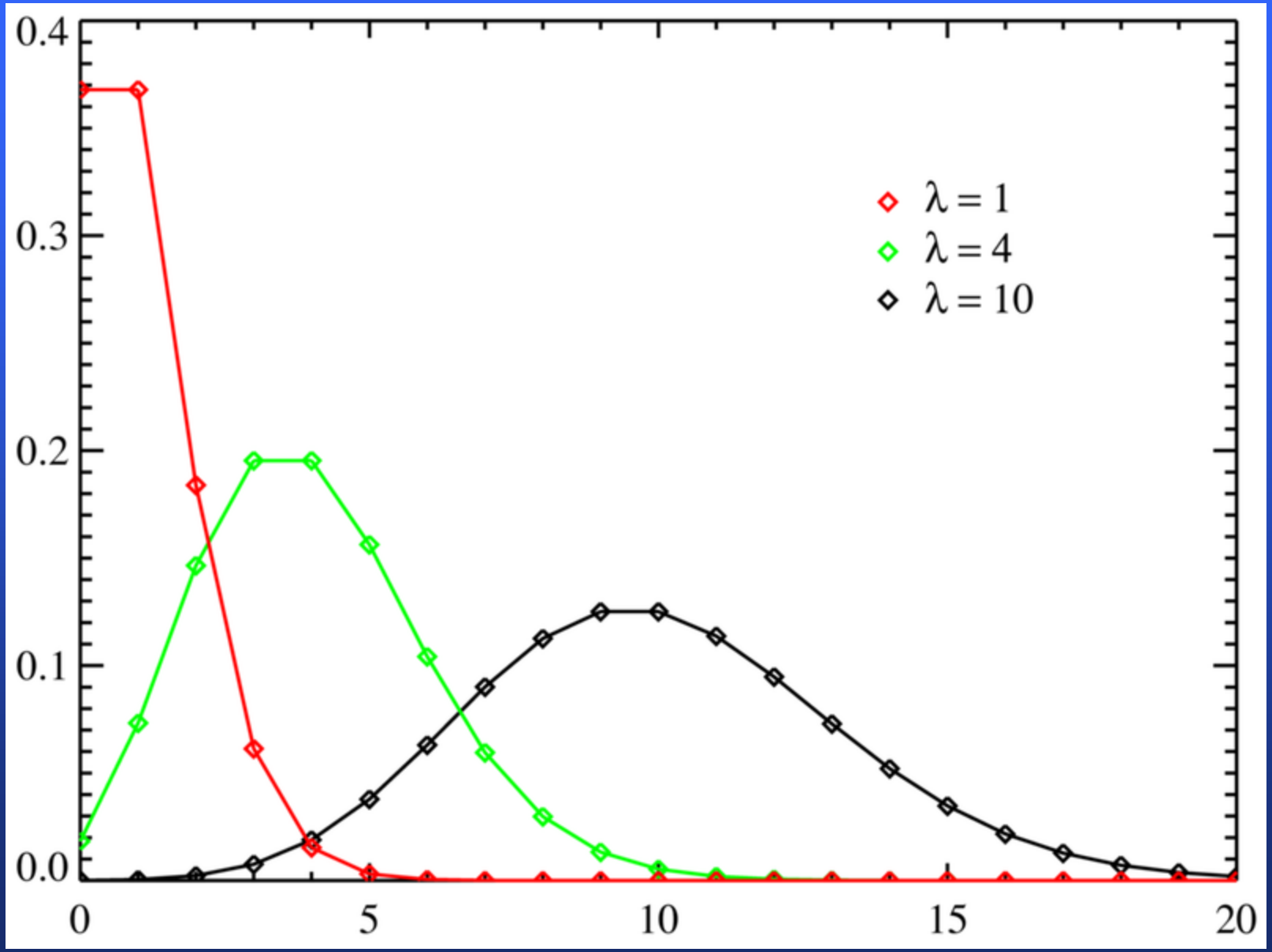
See http://en.wikipedia.org/wiki/Beta_distribution for more information.

Normal Probability Distribution



See http://en.wikipedia.org/wiki/Normal_distribution for more information.

Poisson Probability Distribution



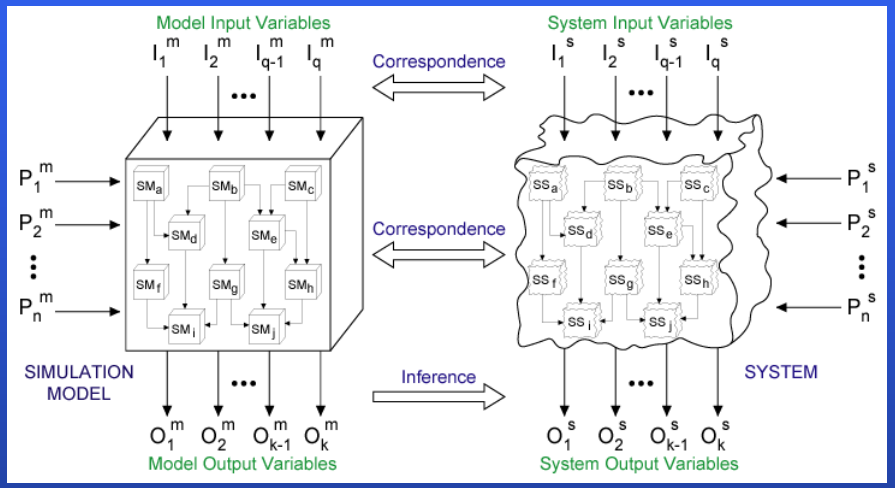
See http://en.wikipedia.org/wiki/Poisson_distribution for more information.

Simulation of Random Phenomenon

- What is a random phenomenon?
  - Arrivals of vehicles to a traffic intersection
  - Arrivals of passengers to an airport
  - Arrivals of jobs to a computer system
  - Repair times of a machine
  - Cashier service times
  - Number of e-mail packets received per unit time
  - Flight time of an airplane between two cities
  - Time between computer system failures
  - Response time for an e-commerce system
- How do we characterize / model random phenomenon?
  - Using Simulation Input Data Modeling
- How do we represent the random phenomenon in our simulation model?
  - Using Random Variate Generation

Simulation Input Data Modeling

- **Trace-Driven Simulation:** Requires no modeling. Input data traced from the real operation of the system are used directly.
- **Self-Driven Simulation:** Probabilistic modeling of simulation input



- Collect data on an input variable
- Fit the collected data to a probability distribution and estimate its parameters (probabilistic modeling) (To do this, use a software product such as ExpertFit.)
- Sample from the fitted probability distribution using Random Variate Generation to drive the simulation model.

Example

Play the Video

- **Random Phenomenon:**
  - Arrivals of vehicles to a lane
- **Random Variable of Interest:**
  - Inter-arrival times of vehicles to a lane
- **Data Collection and Modeling:**
  1. Record the arrival times of vehicles to a lane (in seconds):

     0, 12, 20, 21, 23, 34, 42, 50, 51, 55, 60, 62, 76, 82, 90, 101, …

     12 8 1 2 11 8 8 1 4 5 2 14 6 8 11
  2. Compute the inter-arrival times:

     12, 8, 1, 2, 11, 8, 8, 1, 4, 5, 2, 14, 6, 8, 11, …
  3. Fit the inter-arrival times to a probability distribution and estimate its parameters, e.g., Exponential with mean = 8
  4. Use the Exponential (8) Random Variate Generation to simulate the random phenomenon of vehicle arrivals

Input Data Modeling Approaches

Case 1: Data can be collected

a. Collected data fit to one of the known probability distributions. Use the Random Variate Generation (RVG) algorithm for that probability distribution.

b. Collected data do not fit to one of the known probability distributions.

   i. If more than 100 independent observations are available, then use the empirical approach for modeling the data. Use a table lookup algorithm for the collected data.

   ii. If less than 100 independent observations are available, then use the uniform, triangular or beta approach.

Case 2: Data cannot be collected

a. Use the uniform, triangular or beta approach for input data modeling in the absence of data.

Input Data Modeling in the Absence of Data: Uniform

Weighted Average

Weighted Average

		weight	estimate	estimate
Subject Matter Expert (SME)		w_1	a_1	b_1
		w_2	a_2	b_2
		w_3	a_3	b_3
		w_4	a_4	b_4

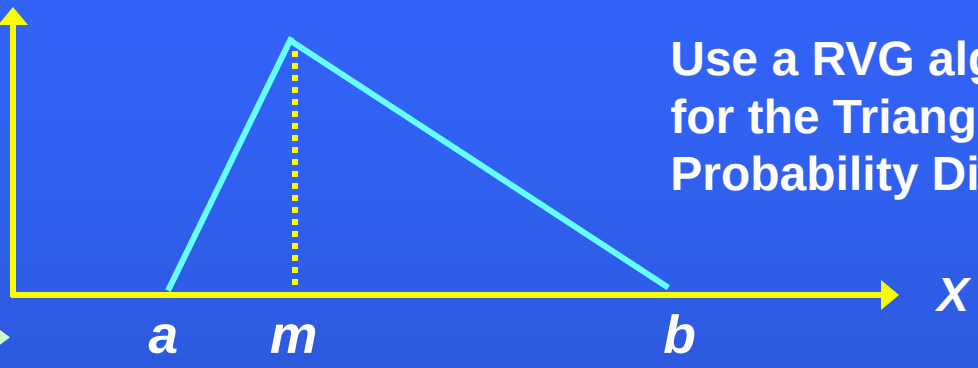
Ask each SME to estimate a and b

Weight each SME since one SME can be more knowledgeable than another.

Input Data Modeling in the Absence of Data: Uniform

- **Step 1:** Identify the random variable of interest, X.
- **Step 2:** Identify N Subject Matter Experts (SMEs) with expertise and experience in the problem domain.
- **Step 3:** Assign relative criticality weights (weight is a fractional value between 0 and 1; All SME weights must sum to 1) to the SMEs, w_j, j=1,2,…N
- **Step 4:** Ask each SME to subjectively estimate the lowest value, a, for X. a_j, j=1,2,…,N
- **Step 5:** Ask each SME to subjectively estimate the highest value, b, for X. b_j, j=1,2,…,N
- **Step 6:** Use a Uniform probability distribution for X over a and b, where a and b are weighted averages computed as
  - $a = \sum a_j \times w_j$ for j = 1,2,…,N
  - $b = \sum b_j \times w_j$ for j = 1,2,…,N

Input Data Modeling in the Absence of Data: Triangular

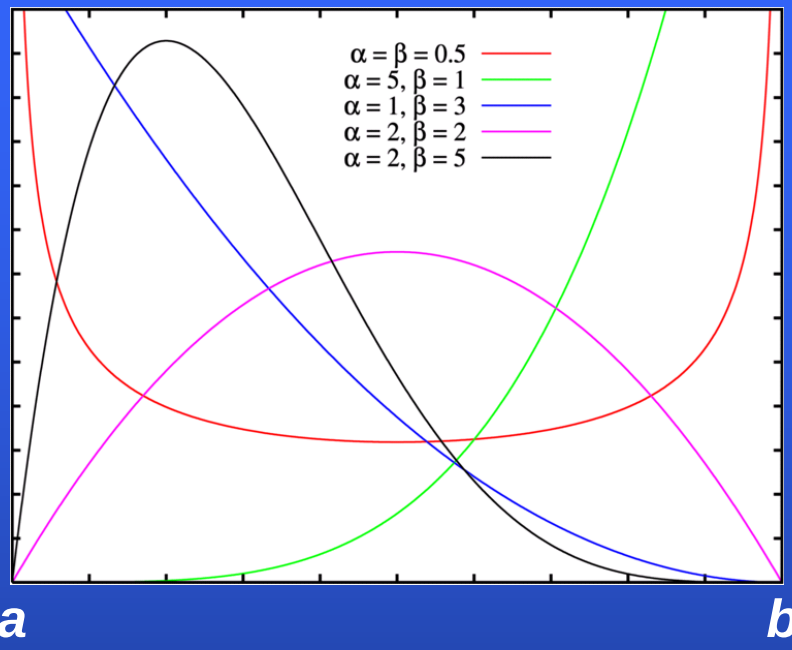


Use a RVG algorithm for the Triangular Probability Distribution

Weighted Averages

SME	weight	estimate	estimate	estimate
	w_1	a_1	m_1	b_1
	w_2	a_2	m_2	b_2
	w_3	a_3	m_3	b_3
	w_4	a_4	m_4	b_4

Input Data Modeling in the Absence of Data: Beta



- **Step 1:** Use the approach for the Uniform distribution to estimate lowest and highest values, *a* and *b*.
- **Step 2:** Ask each SME to estimate the shape parameters α and β to suggest a distribution of values over the range *a* to *b*. Note that $\alpha = 1$ and $\beta = 1$ create a Uniform distribution.
- **Step 3:** Use a RVG algorithm for the Beta probability distribution with SME weighted averages for α and β.

Simulation of Random Phenomenon

- The Probability Distribution identified as the best fit to the collected data becomes the Probabilistic Model of the random phenomenon.
- The random phenomenon is simulated in the simulation model by using an algorithm, called Random Variate Generator (RVG).
- Later, we will learn how to develop an RVG to generate random values so as to form the fitted probability distribution with the estimated parameters.