Representing Uncertainty in SysML 2.0

Chris Paredis
Director
Model-Based Systems Engineering Center
Georgia Tech
chris.paredis@me.gatech.edu

Overview

- Motivation
- Objectives
- What needs to be represented?
- Thoughts on how to model uncertainty in SysML
- Some background on probability theory

Motivation

- Decision making under uncertainty is ubiquitous in SE—risk, safety/reliability, but...
  - Current approaches are often ad-hoc and qualitative
    - » E.g., fever charts
  - Engineers and other decision makers often lack a deep understanding of the underlying theory
  - The cost of inference with uncertainty is high
    - » Both in cost of modeling and cost of computation
- To move toward more quantitative, theoretically rigorous approaches for decision making under uncertainty, we need to express/model uncertainty explicitly
- Aim to make SysML 2.0 sufficiently expressive and precise to support rigorous inference under uncertainty

Objectives

- Enable representation of uncertain information in SysML
  - Rigorous, precise
  - Explicit
  - Useful (the most common constructs but not necessary all)
  - Extensible
- Primary focus today:
  - Make some strategic decisions about scope, approach, transformation path… which will then inform technical implementation
- Warning: more questions than answers…

Rigorous — Some Background on Uncertainty

Based on Weatherford: “Phil. Found. of Prob. Theory”

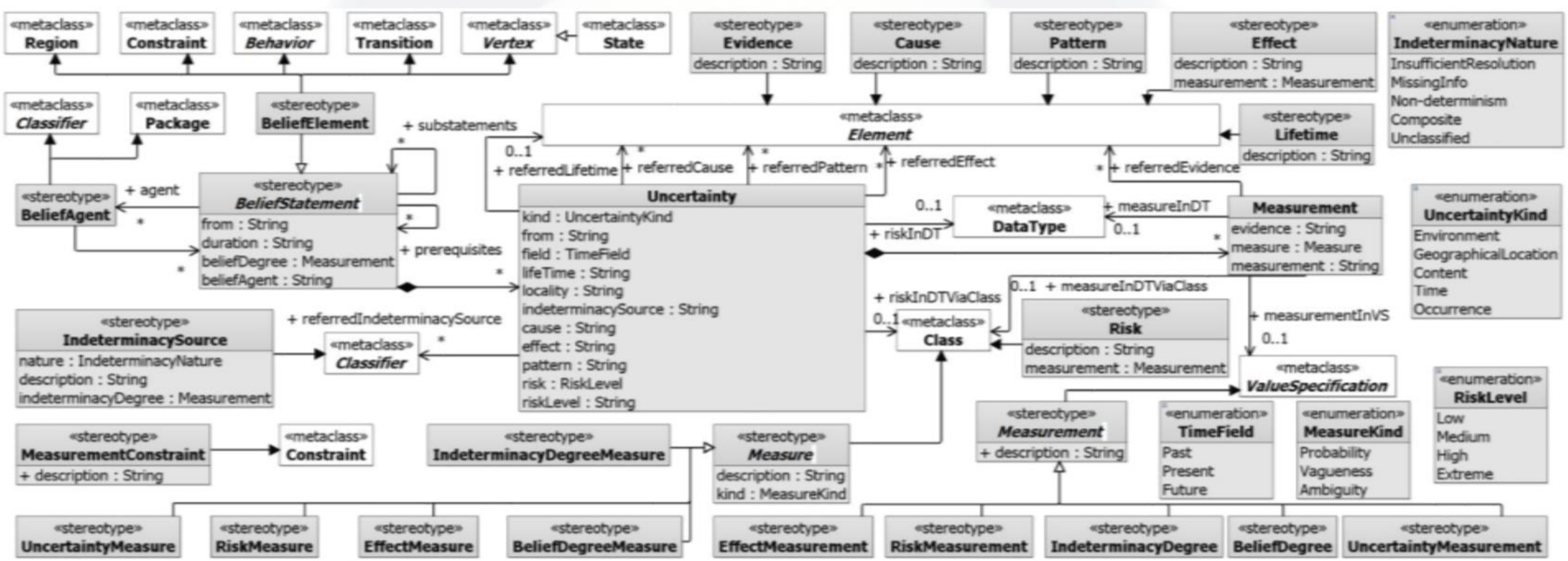
- Uncertainty relates to “asserting with less than certainty”
  - Predicting a future event
  - Supporting a conclusion based on available (limited) evidence
- Many formalisms have been proposed for characterizing and reasoning with uncertainty:
  - Probability theory
  - Fuzzy set theory, fuzzy number, fuzzy random numbers, random fuzzy numbers…
  - Possibility theory, Dempster-Shafer theory, imprecise probability theory,…
  - …
- Strong consensus in the philosophical/scientific community that reasoning under uncertainty should (always) be based on Probability Theory
  - e.g. all NASA uncertainty and risk handbooks & best practices specify a probabilistic approach
- Other approaches have been shown to lead to inconsistencies
  - e.g. Dutch book argument in support of Kolmogorov axioms, etc.
- Supporting other formalisms besides probability theory is a bad idea: we lose (all) rigor

Rigorous — Some Background on Uncertainty

Based on Weatherford: “Phil. Found. of Prob. Theory”

- This seems to be in stark contrast with the new OMG RFI on Uncertainty Modeling which proposes to include everything and the kitchen sink (http://www.omgwiki.org/uncertainty/doku.php?id=start)

UML UNCERTAINTY PROFILE (UUP): IMPLEMENTATION OF U-TAXONOMY



M. Zhang, S. Ali, T. Yue and P. H. Nguyen, Uncertainty Modeling Framework for the Integration Level V.1, https://www.simula.no/file/uupv1pdf-1/download

Rigorous — Some Background on Uncertainty

Based on Weatherford: “Phil. Found. of Prob. Theory”

- The calculus of probability:
  - A set of rules for manipulating numbers of a certain type in order to produce more numbers of the same type
  - Kolmogorov’s axiomatization (1933)
- A theory of probability:
  - An interpretation to the calculus which leads to probability judgments
  - Subjective probability by Ramsey, de Finetti, Savage, etc.
- Probability:
  - An expression of belief — personal / subjective
  - Belief is measured by willingness to bet
  - Unfortunately, many engineers have been taught that a probability is a relative frequency — which is reasonable in some contexts, but is ultimately too limiting
- Conclusion: to be rigorous, we should limit ourselves to probability theory

What Needs to be Represented?

Recommended Modeling Choices

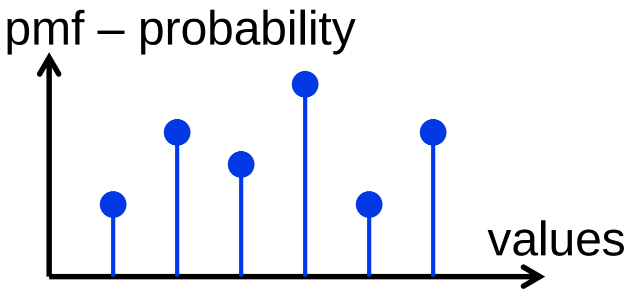
- Which mathematical formalism?
  - Probability theory (and only probability theory)
- Which constructs?
  - Probabilities
  - Distributions: PDF, PMF, CDF, moments,…
  - Joint distributions
  - Random processes
- Meta-information?
  - Author(s), pedigree, history, underlying data/models, etc.

Probability

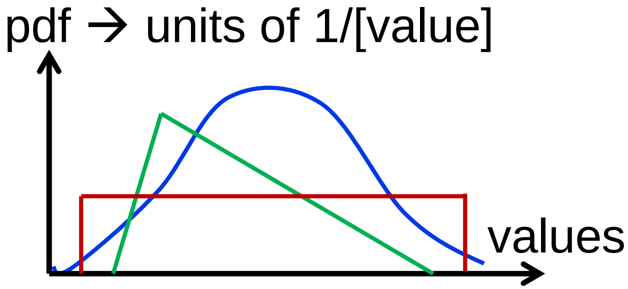
- Motivating example:
  - Probability of Loss of Mission — $P(LoM)$
  - LoM is an event — a subset of the sample space of mission outcomes
- Unitless value type: $P \in [0,1] \subset \mathbb{R}$
- This will fit directly with the proposed ValueType definition
- Questions
  - Is it important for the "event" (and maybe even the corresponding sample space) to be modeled explicitly?
  - Or is it sufficient to be informally referenced in the name of the value property?

Probability Distribution — Random Variable

- Discrete: can take on only certain separated values
  - – Number of possible values could be finite or infinite
  - – Examples: colors, number of valves of an engine



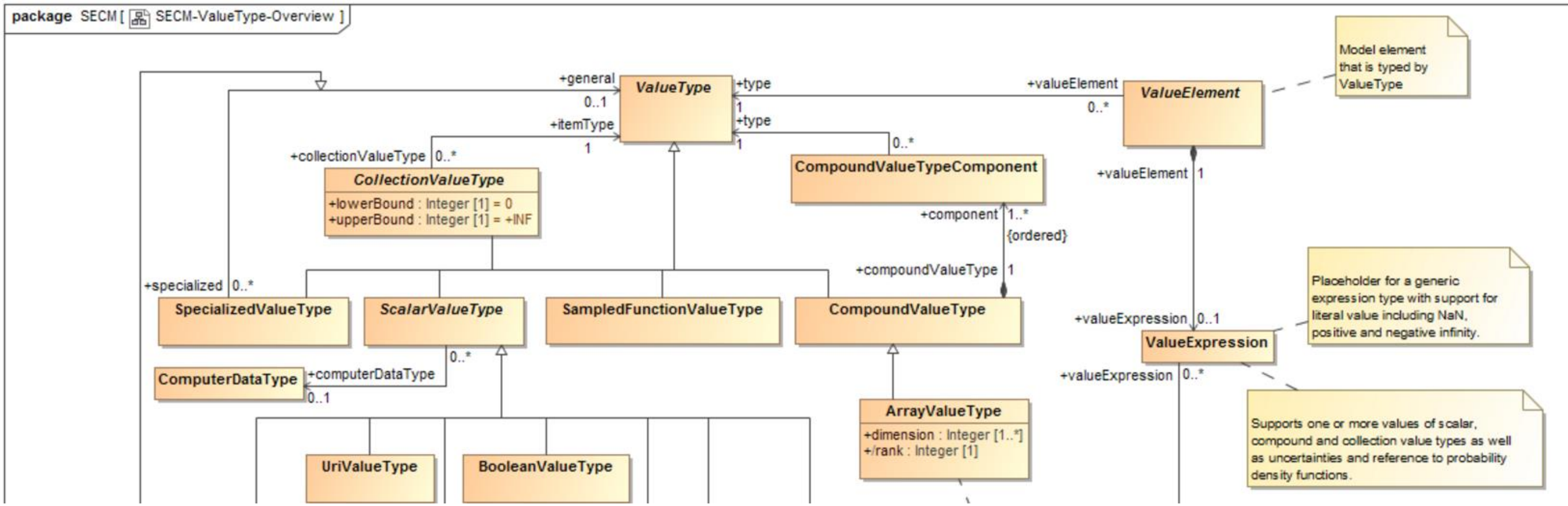
- Continuous: can take on any real value in some range
  - – Number of possible values is always infinite
  - – Range could be bounded on both sides, just one side, or neither
  - – Examples: weight of an object, accuracy of machining



Probability Distribution — Random Variable

- Motivating example:
  - car.mass:kg = 1000
  - car.mass 1000 [kg]
- Any physical quantity is uncertain and should therefore be modeled as a random variable
  - The realized mass for a particular vehicle is unknowable
  - E.g., predicted car mass, measured car mass
  - car.mass Normal(1000,100) [kg] (mean and stdev as params)
  - The value specification is replaced by a distribution specification

Probability Distribution — Random Variable



- Where does Distribution fit in?
  - Not a ValueType but a ValueElement
  - Should not be a stereotype (as in SysML1.x) but a model element that is defined in a library and is thus extensible

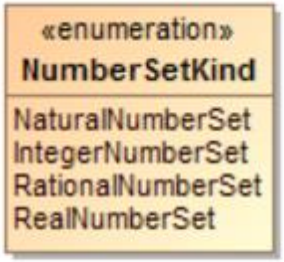
Probability Distribution — Random Variable

- Distribution
  - ValueType – must equal the type of the ValueElement
  - DistributionKind → continuous, discrete, mixed,… TBD
  - 1..* parameters
    - » The number of parameters depends on the distribution type
    - » The types of the parameters depends on the ValueType (but is not necessarily the same – depends on distribution type)
  - Some derived properties for all distributions…
    - » \mean:ValueType
    - » \median:ValueType
    - » \mode:ValueType
    - » \stdDev:ValueType
    - » \skewness:Real
    - » \excKurtosis:Real

Not sure how important all these are and whether they should be required. Not easily computed for some distribution types…

Probability Distribution — Random Variable

- Should every value property be probabilistic?
  - No. – e.g., a specified upper bound on car.mass is deterministic
  - car.maxMass Deterministic(1000) [kg]
- Should the default be a Deterministic distribution or a null/unspecified distribution?
- We need to distinguish between integer-valued and real-valued properties and constrain the allowed DistributionKind accordingly



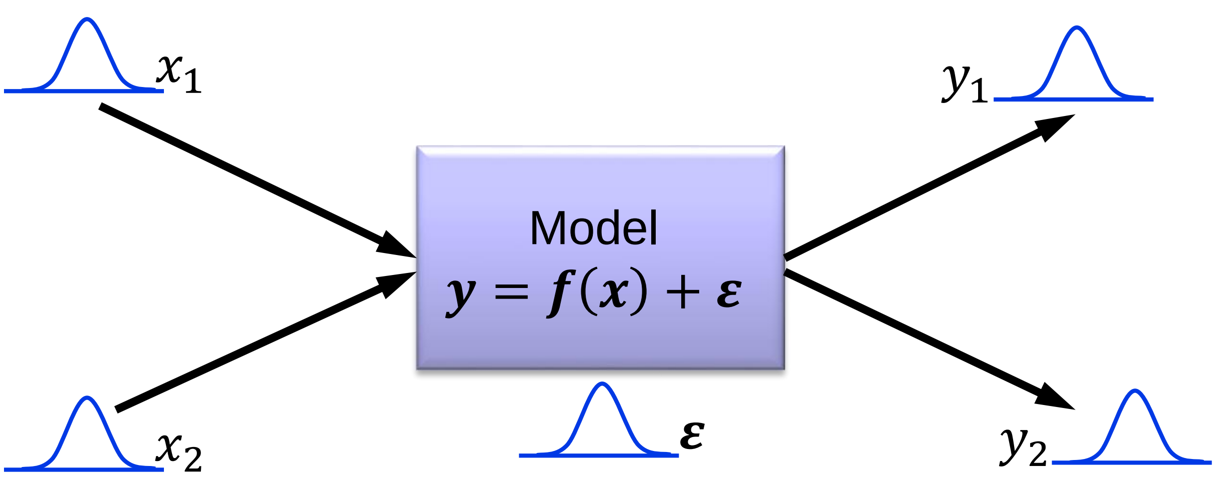
Probability Distribution — Random Variable

- What does an instance specification look like?
  - Is it a sample of a distribution or an instance of the distribution with specific distribution parameters?

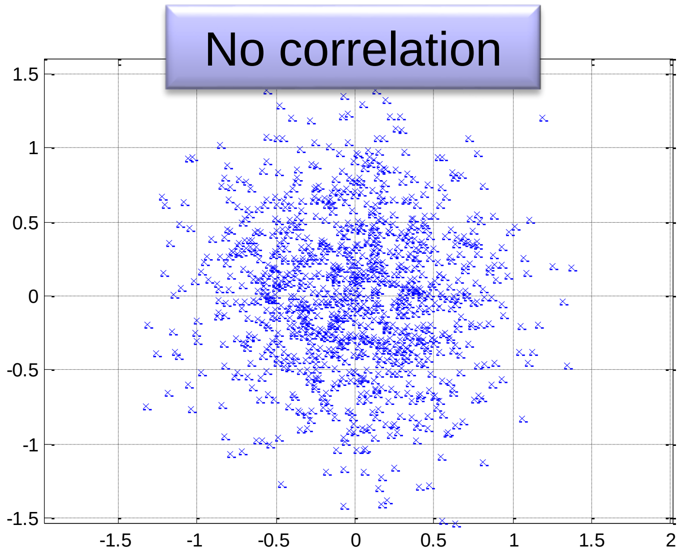
- Some parametrizations of distributions can be quite complicated – e.g., a PMF is defined as a set of (value, probability) pairs. What is the right balance between expressivity, extensibility and simplicity in the definition of Distribution?

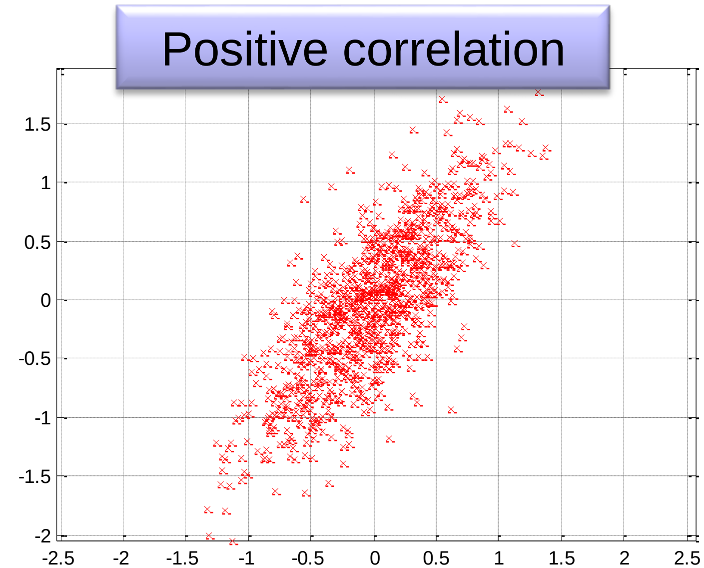
Joint Probability Distributions

- Multiple random variables may be dependent
- This dependence is a relationship between random variables
- Motivating example:
  - Modeling the uncertainty of properties constrained by a model will require joint distributions

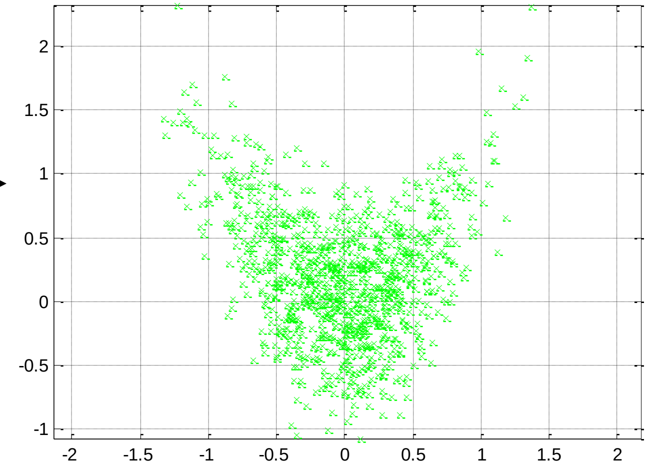


Joint Probability Distributions





What about this one?



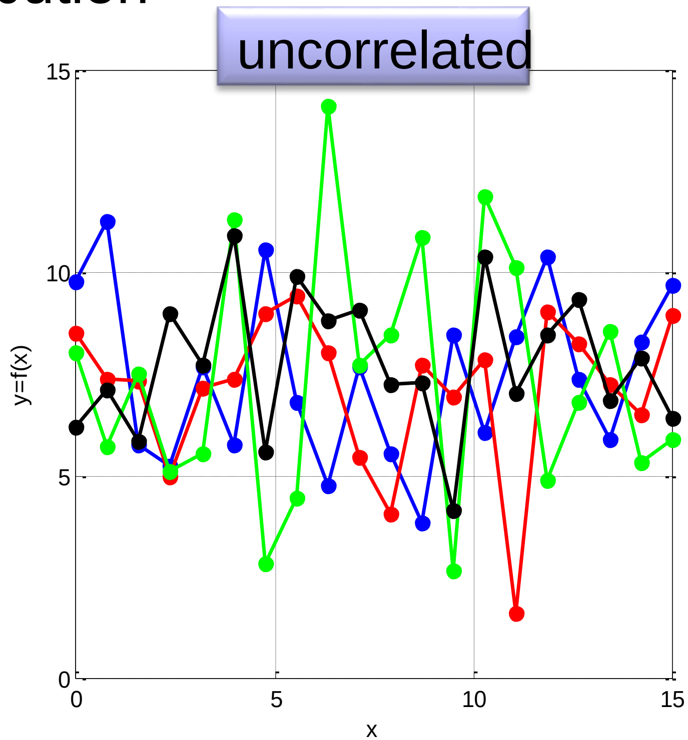
Joint Probability Distributions

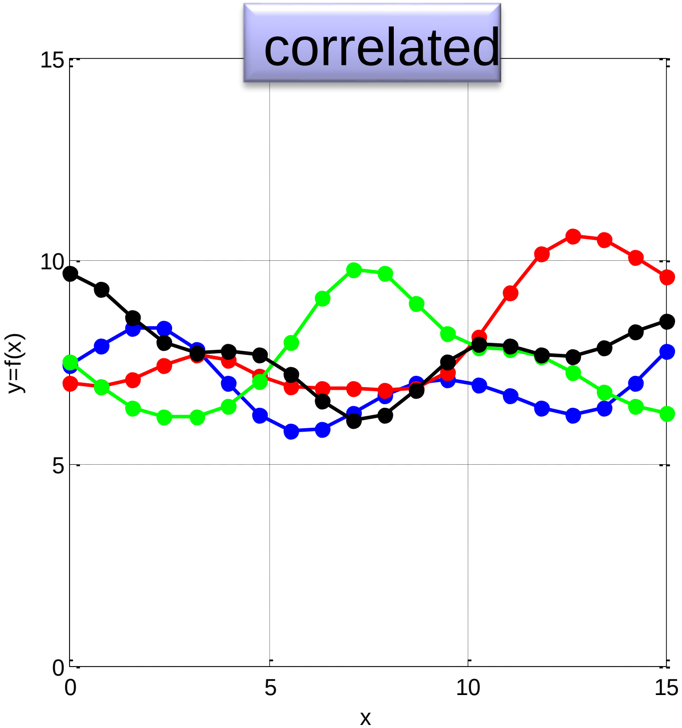
- …to be completed

- Some ideas for representation:
  - Parametric joint distributions: e.g., multivariate Gaussian is characterized by a mean vector and a covariance matrix
  - Copulas
  - Joint sample sets

Random Process

- A stochastic process or random process is a collection of (dependent) random variables — one random variable, y, for each value of x (or t for time series)
- E.g., for a zero-mean Gaussian process, $y \sim \mathcal{N}(0, \sigma^2)$ for each x_i
- In addition, the Gaussian process is characterized by its covariance function — $y(\mathbf{x})$, is characterized by a multivariate Gaussian distribution





Random Process

- Motivating example:
  - Environment temperature as a function of time

- Representation
  - Parametric process models: e.g., Gaussian Process model is characterized by a covariance kernel and the corresponding parameters

Some Final Questions

- How about legacy representations?
  1. Representations that are mapable to probabilities
  2. Representations that are not rigorous
- How about meta-information?
  - To what extent should this be supported in the metamodel?

Backup Slides

Brief Introduction to Probability Theory

- Probability
- Random variables
- Distribution functions: PMF, PDF, CDF
- Some common distributions: uniform, triangular, …
- Covariance
- Frequentist vs Bayesian Subjective

Basic Definitions

- Experiment: Activity or process with an uncertain outcome (e.g. coin flip, rolling dice, ...)
- Simple (or elementary) event: An individual, undecomposable outcome e_i of an experiment
- Sample Space: Set $S = \{e_1, e_2, \ldots\}$ containing all possible outcomes (simple events) of an experiment
  - Could be easy to characterize (e.g. rolling two dice) or hard (e.g. the voting mechanisms used for elections in 2020)
  - May not always be characterized explicitly

Basic Definitions (2)

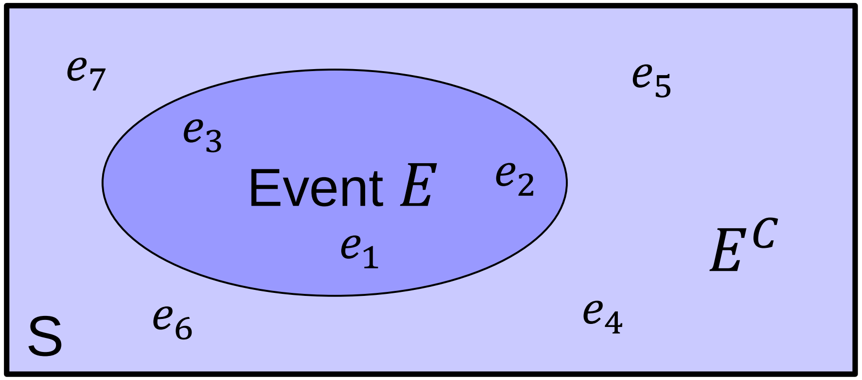
- Event: A subset $E \subseteq S$ of the sample space, usually denoted by E, F, E_1, E_2, etc.
  - Set of simple events
  - E.g. "Sum of faces after rolling two dice is 4" is a set containing 3 simple events: $\{(1,3),(2,2),(3,1)\}$
  - If E and F are events, then so are $E \cap F$, $E \cup F$ and $E \setminus F$
- Probability: Relative likelihood (probability) $p(e_i)$ of an (simple) event e_i occuring when doing an experiment
  - For any $e_i \in S: 0 \leq p(e_i) \leq 1$
  - Furthermore: $\sum_{e \in S} p(e) = 1$
  - And with the definition for events: $\sum_{e \in E} p(e) = p(E)$

Some More Important Properties…

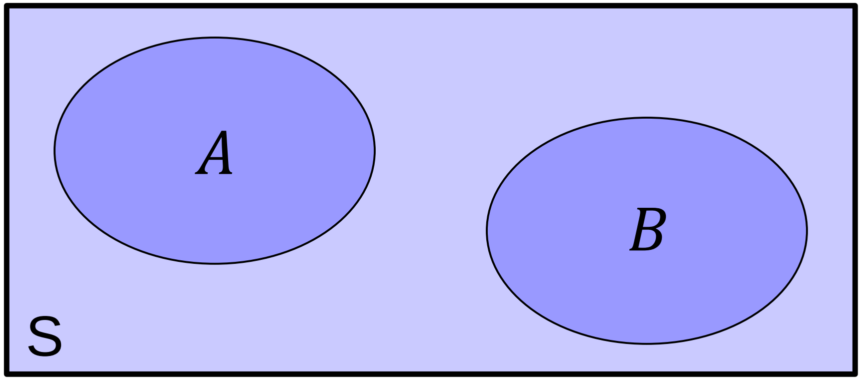
- $p(S) = 1$ must always be true
- Events $E \neq S$ with $p(E) = 1$ can exist
- If ∅ is the empty event (empty set), then $p(\emptyset) = 0$
- There may be events $E \neq \emptyset$ with $p(E) = 0$
- If E^C is the complement of E, then $p(E^C) = 1 - p(E)$
- $p(E \cup F) = p(E) + p(F) - p(E \cap F)$
- If E and F are mutually exclusive (i.e., $E \cap F = \emptyset$), then $p(E \cup F) = p(E) + p(F)$

Venn Diagrams

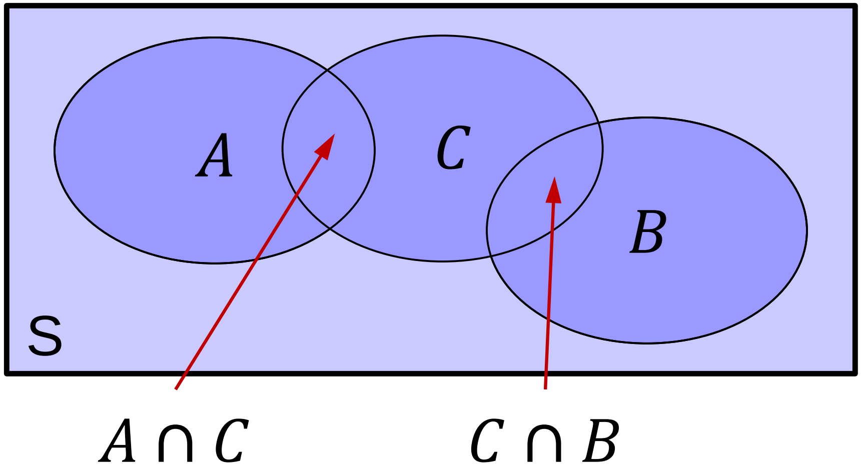
- Useful for visualizing basic properties
- Sets denoted as bounded, 2-dimensional structures



- Mutual exclusivity:



- Intersecting events:



Conditional Probability

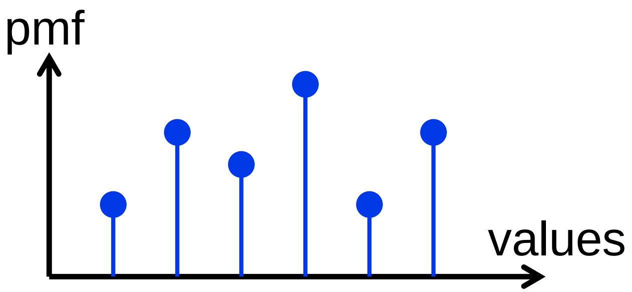
- Two or more events may be related
  - Knowing that an event F occurred might affect the probability that another event E also occurred
  - Reduce the effective sample space from S to F, then measure "size" of E relative to its overlap (if any) in F, rather than relative to S
  - Definition (assuming $p(F) \neq 0$): $$p(E|F) = \frac{p(E \cap F)}{p(F)}$$
- E and F are independent if $p(E \cap F) = p(E)p(F)$
  - Knowing that one event occurs tells you nothing about other
  - Implies $p(E|F) = p(E)$ and $p(F|E) = p(F)$
  - Not to be confused with mutual exclusivity! ($E \cap F = \emptyset$)

Random Variables

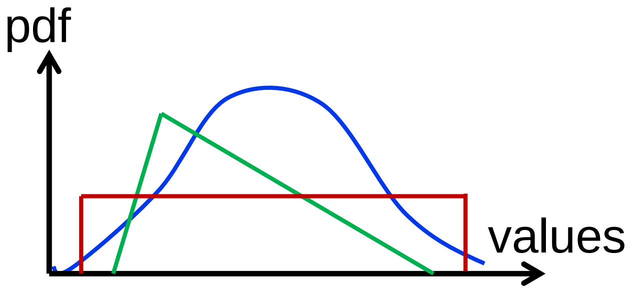
- A random variable (RV) is a numerically valued function representing the set of events that can result from an experiment
  - RV is a number whose value we don't know for sure but we'll usually know something about what it can be or what it is likely to be
  - Usually denoted using capital letters: X, Y, W_1, W_2, etc.
  - A RV is an assigned number to an event, *not* the event itself
  - Examples:
    - » Coin toss: assign 0 to event "Heads" and 1 to "Tails"
    - » Measuring weight: $X = 2.305\ kg$
- Probabilistic behavior described by a distribution

Discrete vs. Continuous RVs

- Discrete: can take on only certain separated values
  - Number of possible values could be finite or infinite
  - Examples: colors, number of valves of an engine



- Continuous: can take on any real value in some range
  - Number of possible values is always infinite
  - Range could be bounded on both sides, just one side, or neither
  - Examples: weight of an object, accuracy of machining



Discrete Distribution Functions

- Let X be a discrete RV with possible values (range) $x_1, x_2, \ldots$ (finite or infinite list)

- Probability mass function (PMF):

$$p(x_i) = p(X = x_i) \quad \text{for } i = 1, 2 \ldots$$

  – The statement "$X = x_i$" is an event that may or may not happen, so it has a probability of happening, as measured by the PMF
  – Since X must be equal to *some* x_i, and since the x_i's are all distinct: $\sum_{all\ i} p(x_i) = 1$

Discrete Distribution Functions (2)

- Cumulative distribution function (CDF)

$$F(x) = p(X \le x) = \sum_{all\ x_i \le x} p(x_i)$$

  – Probability that the value of a RV will be less than or equal to a fixed value x

  – Properties:

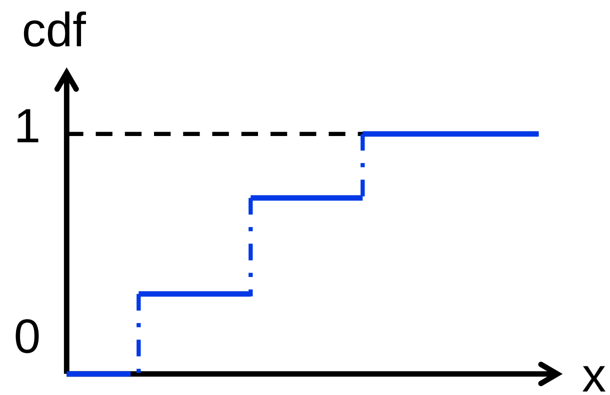
    1) $0 \le F(x) \le 1$ for all x

    2) $\lim_{x \to -\infty} F(x) = 0$

    3) $\lim_{x \to +\infty} F(x) = 1$

    4) $F(x)$ is non-decreasing in x

    5) $F(x)$ is a step function with jumps at the x_i's of height $p(x_i)$



Moments of Discrete Distributions

- Expected Value: first moment M_1; “center“ of a RV

$$M_1 = E\{X\} = \sum_{all\ i} p(x_i) x_i$$

  - Weighted average of the possible values x_i, with weights being their probability (relative likelihood) of occuring
  - The expected value (or expectation) is equivalent to the mean m
  - However, it is not the value of X that you “expect” to get - $E\{X\}$ may not even be among the possible values for X!
  - Example: data set (sample) has a “center“, i.e. an average
    - » Repeat experiment many times, observe many X_i
    - » $\bar{X}$ then converges to $E\{X\}$ as $n \to \infty$

Moments of Discrete Distributions (2)

- Variance: measure of “dispersion“ of a RV

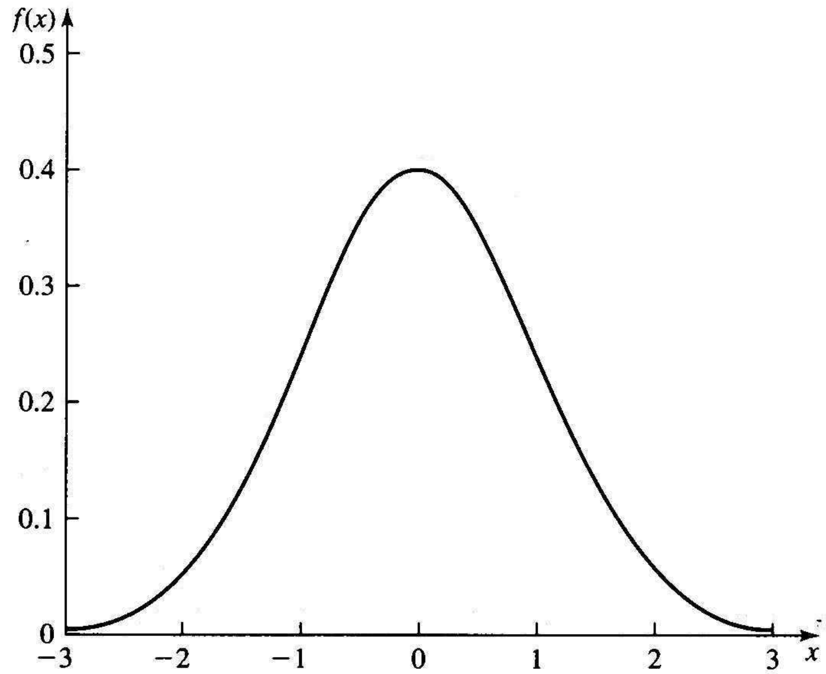
$$M_2 = V\{X\} = \sum_{all\ i} p(x_i)(x_i - E\{X\})^2$$

  - Other common notation: σ^2, ${\sigma_X}^2$, $Var(X)$
  - Weighted average of squared deviations of the possible values x_i from the mean
  - Standard deviation of X is $\sigma = \sigma_X = +\sqrt{V\{X\}}$
  - Interpretation analogous to that for $E\{X\}$

- Data set (sample) has a similar measure:
  - Sample variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
  - Sample standard deviation: $s = +\sqrt{s^2}$

Continuous Distributions

- No matter how small the range of possible values for X, the *number* of possible values for *X* is always infinite (and hence uncountable)
- $p(X = x_i)$ is always 0



- Observed X's are denser in regions where the probability density function *f(x)* is high
- Height of a density is not the probability of anything!

Continuous Distribution Functions

- Let X be a continuous RV

- Probability density function (PDF):

$$f(x) \geq 0 \quad \text{for all real values } x$$

  – Total area under $f(x)$ is 1:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

  – For any fixed a and b with $a \leq b$, the probability that X will fall between a and b is equal to the area under $f(x)$ between a and b:

$$p(a \leq X \leq b) = \int_a^b f(x)dx$$

Discrete Distribution Functions (2)

- Cumulative distribution function (CDF)

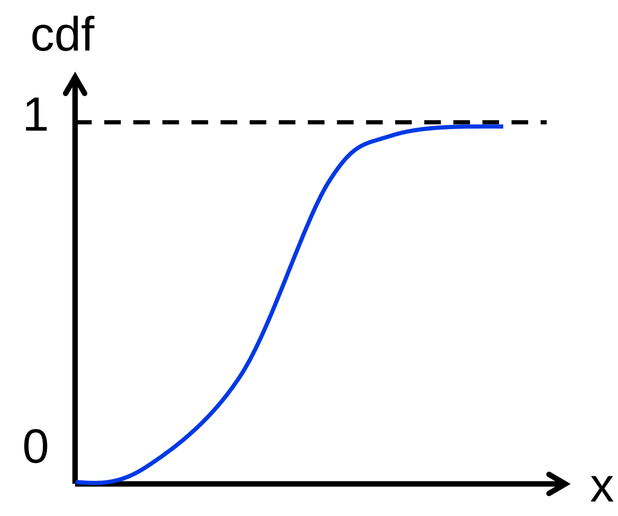
$$F(x) = p(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

  – Probability that the value of a RV will be less than or equal to a fixed value x

  – Properties:

Same as discrete!

1) $0 \leq F(x) \leq 1$ for all x
2) $\lim_{x \to -\infty} F(x) = 0$
3) $\lim_{x \to \infty} F(x) = 1$
4) $F(x)$ is non-decreasing in x
5) $F(x)$ is a continuous function with slope equal to PDF: $f(x) = F'(x)$



Moments of Continuous Distributions

- Expected Value: first moment M_1

$$M_1 = E\{X\} = \int_{-\infty}^{+\infty} x f(x) dx$$

  - The expected value (or expectation) is equivalent to the mean m
  - Roughly, a weighted "continuous" average of possible values for X
  - Same interpretation as in discrete case: average of a large number (infinite) of observations on the RV X

Moments of Continuous Distributions (2)

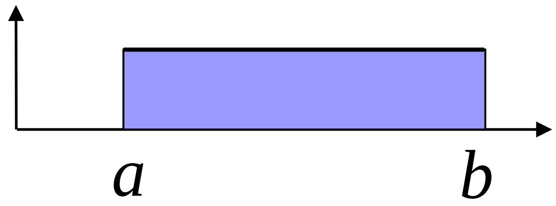
- Variance: measure of "dispersion" of a RV

$$M_2 = V\{X\} = \int_{-\infty}^{+\infty} (x - E\{X\})^2 f(x) dx$$

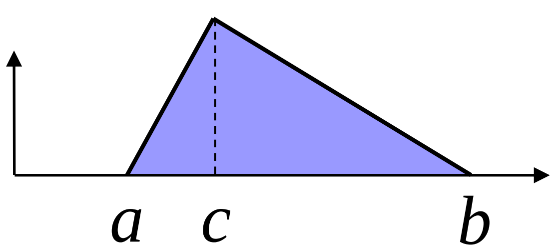
  - Other common notation: σ^2, ${\sigma_X}^2$, $Var(X)$
  - Standard deviation of X is $\sigma = \sigma_X = +\sqrt{V\{X\}}$
  - Interpretation analogous to that for $E\{X\}$

Simple Continuous Distributions

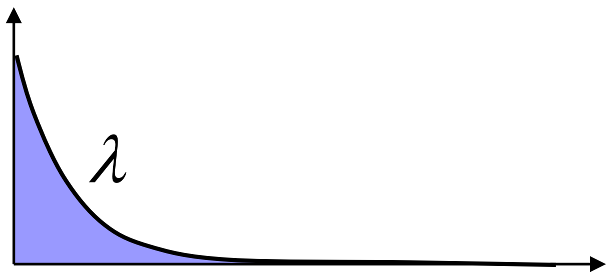
- Uniform Distribution
  - Only lower and upper bound are known



- Triangular Distribution
  - Quick first guess (lower, upper bound and mode)



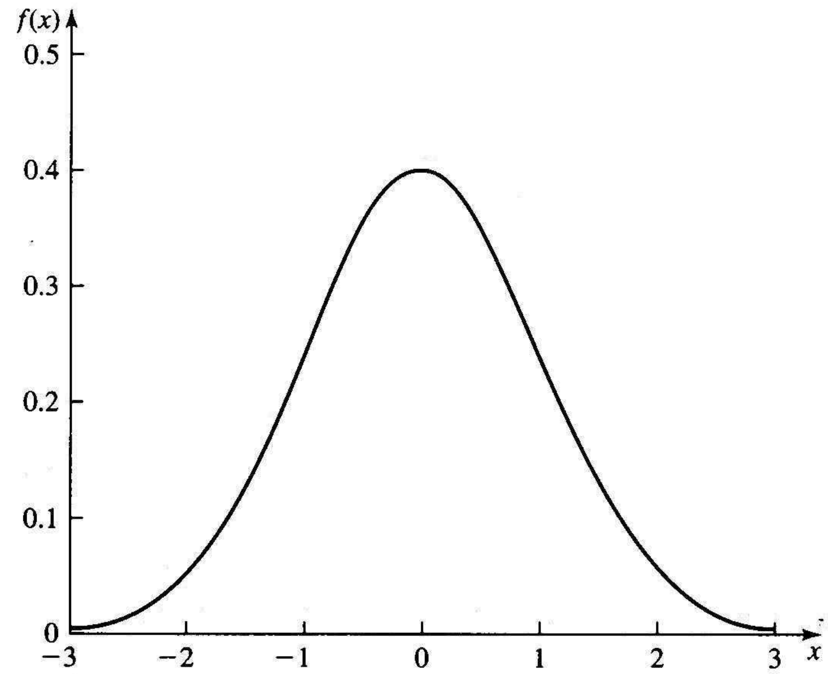
- Exponential Distribution
  - Time between random events – constant arrival rate



$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , \ x \geq 0 \\ 0 & , \ x < 0 \end{cases}$$

Normal Distribution

- Use:
  - Errors of various types
  - Sum of a large number of other quantities (central limit theorem)
  - Maximum entropy distribution when only mean and variance are known
  - NOTE: when estimating mean μ and variance σ^2, use Student distribution (t-dist)
- Parameters: m (or μ), σ



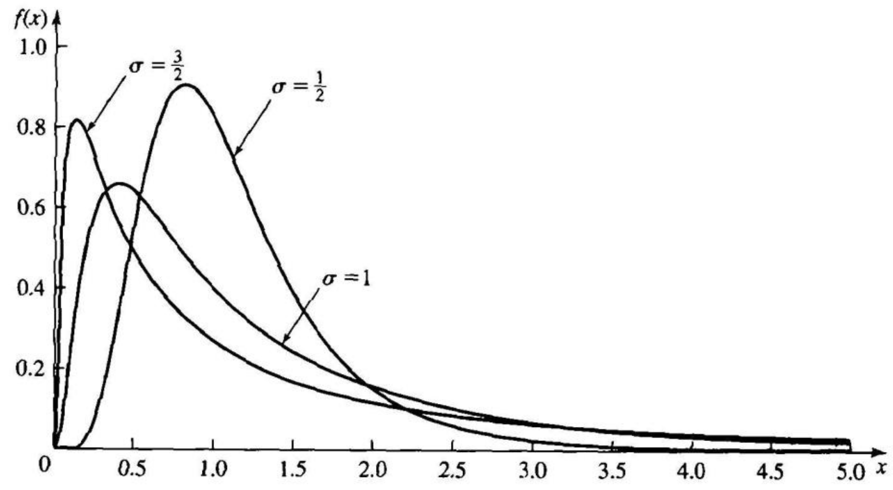
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Lognormal Distribution

- Use:
  - Time to perform a task
  - Product of a large number of other quantities
  - Quantities that are always positive and the distribution is skewed towards zero
  - Analogy to normal distribution N:

$$x \sim LN(m, \sigma^2) \Leftrightarrow \ln x \sim N(m, \sigma^2)$$

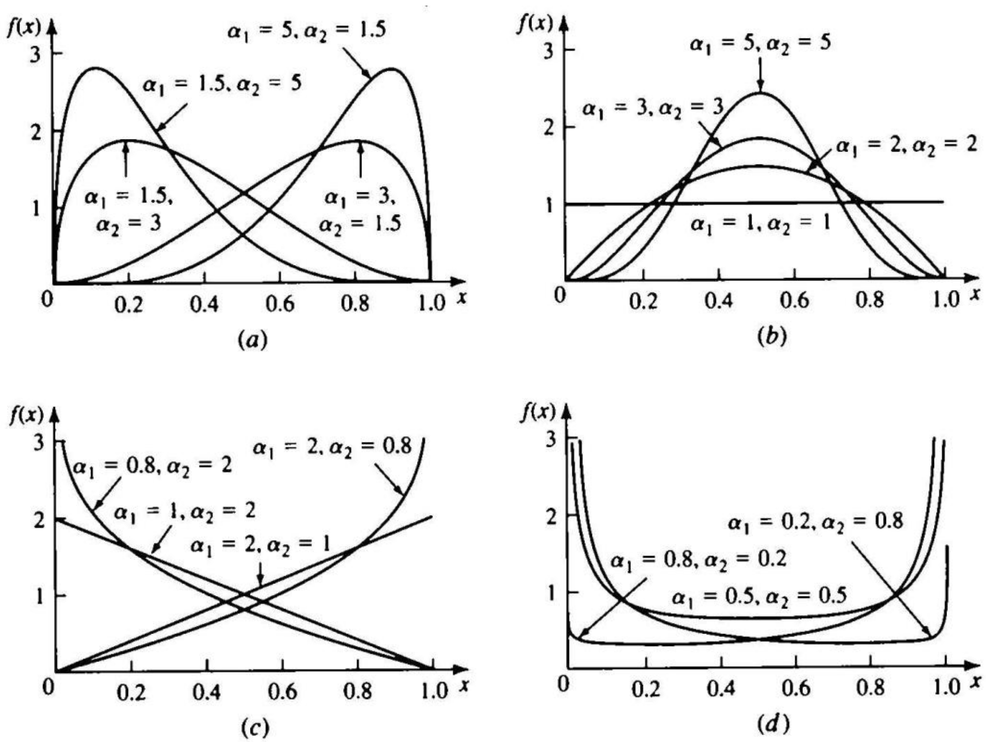
- Parameters: m (or μ), σ



$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - m)^2}{2\sigma^2}}$$

Beta Distribution

- Use:
  - When lower and upper bounds exist
  - Rough initial model – but more refined then triangular
  - Commonly used in Bayesian probability theory – easy to compute posterior distributions

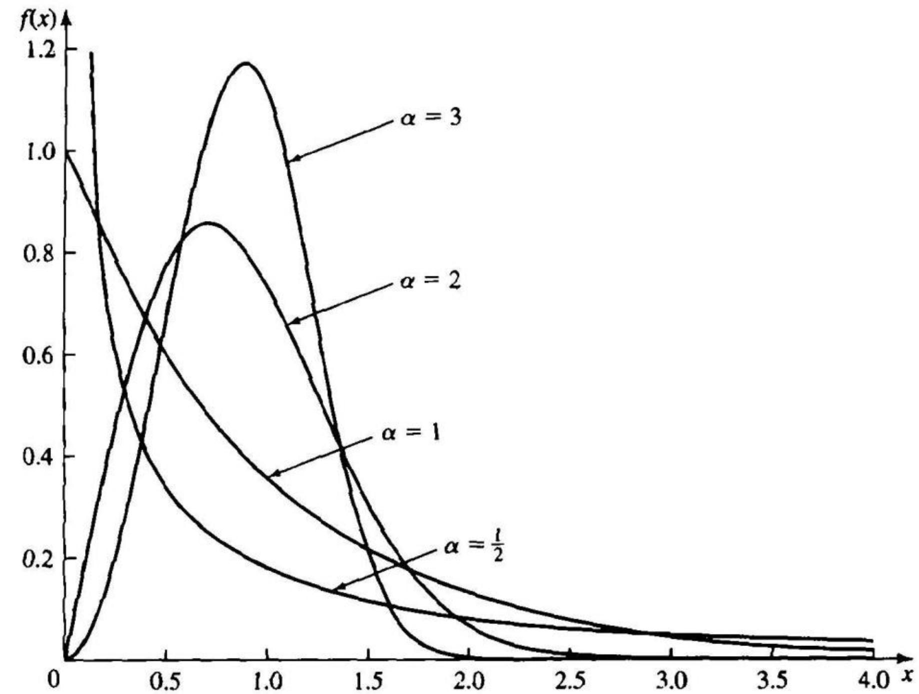


- Parameters: α_1, α_2

$$f(x) = \frac{x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1}}{B(\alpha_1, \alpha_2)} \quad (0 < x < 1)$$

Weibull Distribution

- Use:
  - Time to complete a task
  - Time to failure for a piece of equipment
  - If the failure rate decreases over time, then $\alpha < 1$
  - If the failure rate is constant over time, then $\alpha = 1$
  - If the failure rate increases over time, then $\alpha > 1$
- Parameters: α, β



$$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^{\alpha}} \quad (0 \leq x)$$

Joint Distributions: More Than 1 RV

- Jointly distributed RVs or random vectors

- Example:
  - Input: (P, W, S) = (type of part, weight, service time)
  - Output: $\{T_1, T_2, T_3, \ldots\}$ = total processing time of exiting parts

- Are the individual RVs independent of each other or related?
  - We will consider the special case of a pair of RVs (X_1, X_2) Extends naturally (but messily) to higher dimensions

Joint Distributions: More Than 1 RV

- Joint CDF of (X_1, X_2) is a function of two variables:

$$F(X_1, X_2) = p(X_1 \leq x_1 \text{ and } X_2 \leq x_2)$$
$$= p(X_1 \leq x_1, X_2 \leq x_2)$$

- Joint PMF for two discrete RVs:

$$p(x_1, x_2) = p(X_1 = x_1, X_2 = x_2)$$

- Joint PDF for two continuous RVs with total volume below the resulting non-negative function equal to 1:

$$p(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x_1, x_2) dx_2 \, dx_1$$

Covariance Between RVs

- Measures *linear* relation between RVs X_1 and X_2
- Covariance between X_1 and X_2 is

$$Cov(X_1, X_2) = E\{(X_1 - E\{X_1\})(X_2 - E\{X_2\})\}$$

  - If large X_1 tends to go with large X_2, then covariance > 0
  - If large X_1 tends to go with small X_2, then covariance < 0
  - If there is no tendency for X_1 and X_2 to occur jointly in agreement or disagreement over being big or small, then $Cov = 0$
- Interpreting value of covariance is difficult since it depends on units of measurement

Correlation Between RVs

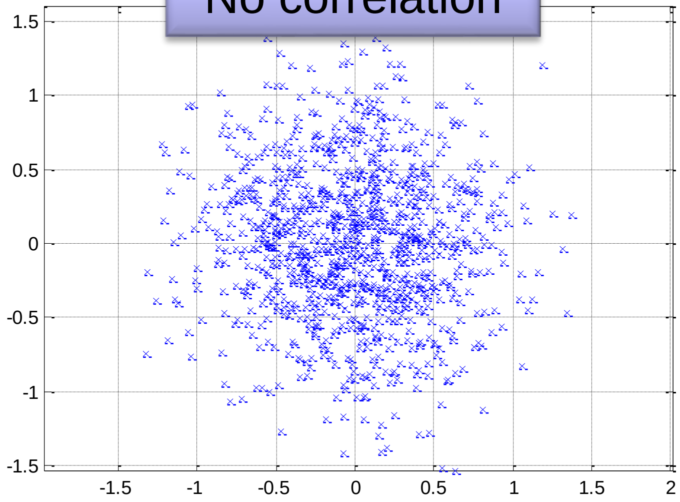
- Correlation (coefficient) between RVs X_1 and X_2 is:

$$Cor(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}}$$

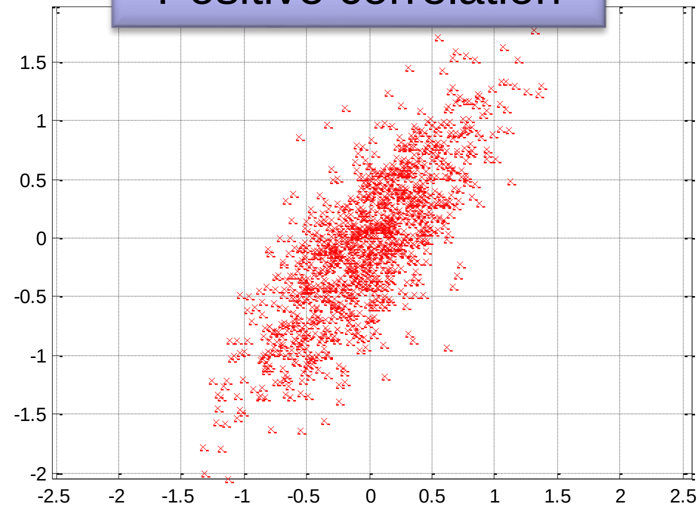
  - Has same sign as covariance
  - Always between –1 and +1
  - Numerical value does not depend on units of measurement
  - Dimensionless – universal interpretation
- Still only good for capturing *linear* relation between X_1 and X_2

Visualizing Correlations

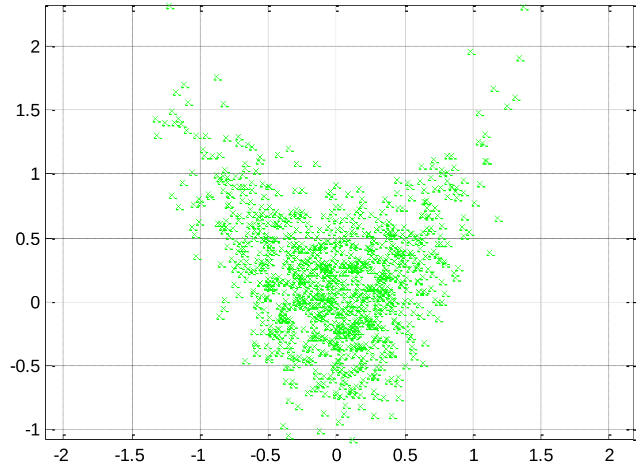
No correlation



Positive correlation



What about this one?



Statistics vs. Probability Theory

Both deal with probabilities…

…but with different interpretations!

- Statistics:

  Analysis of frequencies of ***past*** events

- Probability Theory:

  Predicting the likelihood of ***future*** events

- For a good overview of the meaning of probability see: "Philosophical Foundations of Probability Theory" by Roy Weatherford

Fundamental Interpretations of Probabilities

- Many interpretations argued throughout history
- You probably learned about the frequentist interpretation
  - Probability of an event is the limit value of long run frequency of outcomes
  - E.g., coin toss: p(heads) ≈ # heads / # tosses
- Frequentist interpretation breaks down when accepting that every event is unique — no repetition ever occurs
  - Probability of rain tomorrow
  - Probability of GT winning against Virginia on Saturday
- Is probability meaningful beyond relative frequencies?

Subjective Probabilities

- Probability expresses your willingness to bet or act
- Probability of an event = relative amount you are willing to pay to engage in a bet that...
  - – Pays $1 if the event occurs
  - – Pays $0 otherwise
  - ⇒ Probability = $bet / $1
- You should determine the amount for which you are willing to both buy and sell the bet – the fair price
- Subjective, but:
  - – Unambiguous, since the meaning is well defined and consistent across different events
  - – Operational definition, which is especially important in support of decision making

Criteria for Acceptable Probability Values

- Beliefs must be internally consistent / coherent
- Example: GT plays against Virginia
  - I believe GT has a 50% chance of winning
  - I believe Virginia has a 40% chance of winning
- Are these acceptable probability values?
  - No! Must satisfy no-sure-loss criterion — see "Dutch book" argument in textbook Chapter 4.1
- Beliefs must adhere to Kolmogorov's axioms:
  - For any event E: $0 \leq p(E) \leq 1$
  - For the space S of all possible events: $p(S) = 1$
  - For disjoint events: $p(E_1 \cup E_2 \cup E_3 \ldots \cup E_n) = \sum_{i=1}^{n} p(E_i)$

Criteria for Acceptable Probability Values

- Rational beliefs must also be externally consistent
- What is the wrong with the following belief: ”I am willing to pay $0.6 for a bet that pays $1 if a fair coin flip results in heads, $0 otherwise”
- What is the relationship between a frequentist interpretation of an inherently random event and a subjective probability of that event?
  - Beliefs should be consistent with scientific, factual information, i.e., observations of nature
- How much would you be willing to pay for a coin flip with a bent coin?

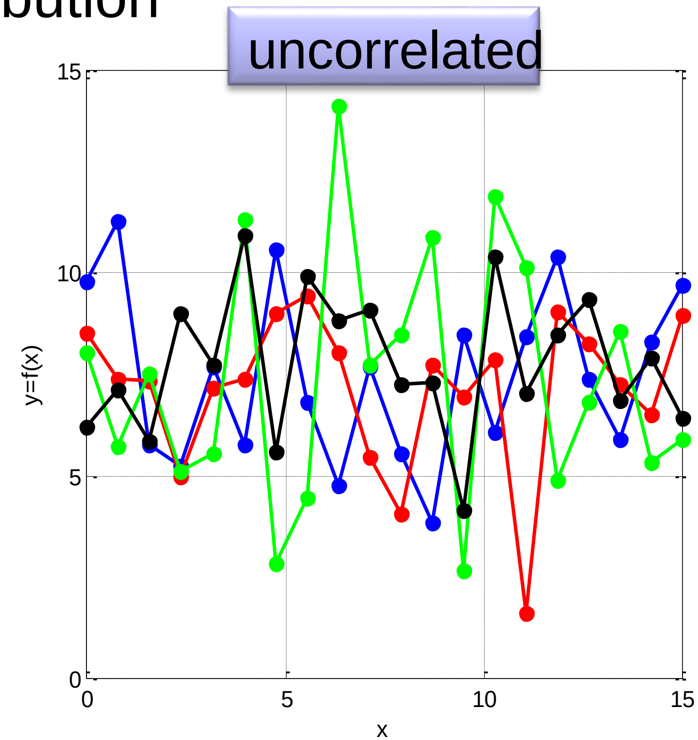
What is a Stochastic/Random Process?

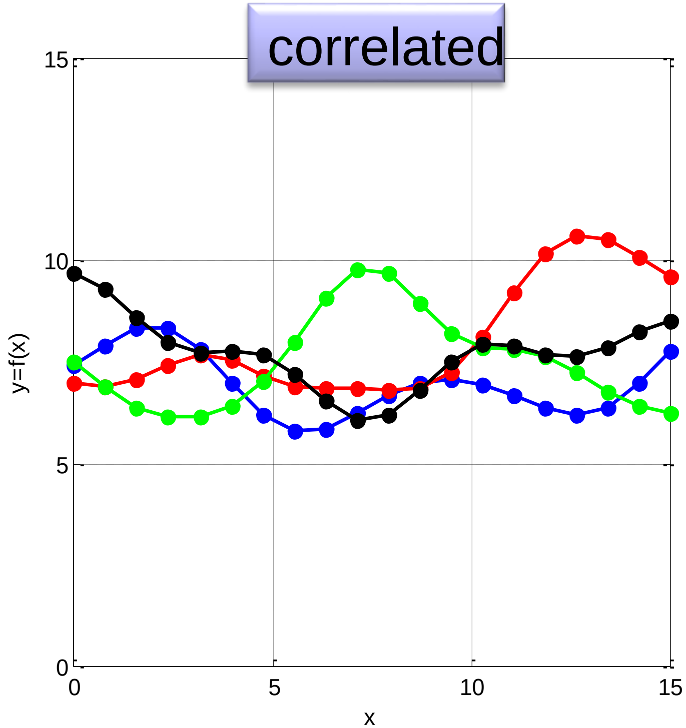
- A stochastic process or random process is a collection of random variables — in our case, one random variable, y, for each value of x (for time series, use t instead of x)
- x can be continuous or discrete
- y can be a multidimensional vector
- A specific sample of a random process is called a *realization*

Brownian motion: a realization of a 2-dimensional random process

What is a Gaussian Process?

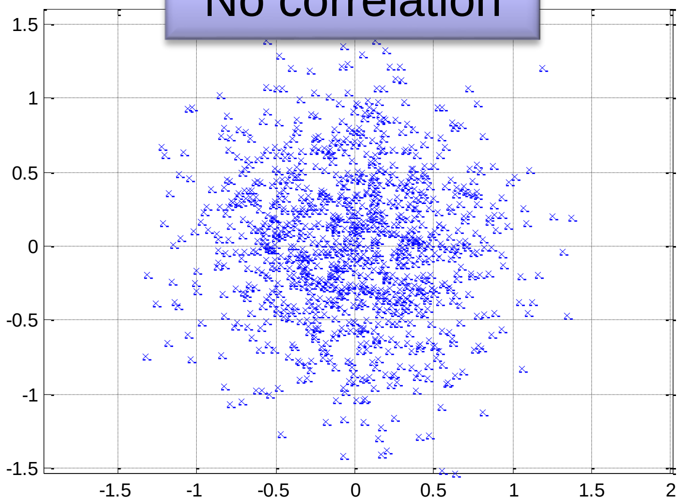
- A stochastic process or random process is a collection of random variables — in our case, one random variable, y, for each value of x
- For a Zero-Mean Gaussian Process, $y \sim N(0, \sigma)$ for each x
- In addition, the Gaussian process is characterized by its covariance function
- A vector, $y(x_i)$, is characterized by a multivariate Gaussian distribution



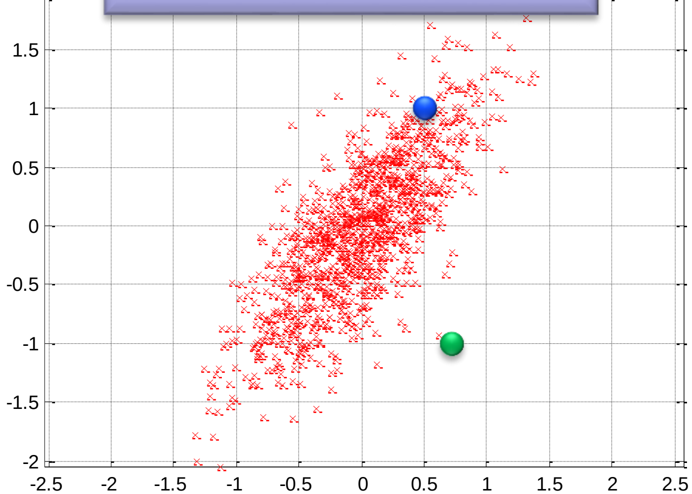


Meaning of Correlation

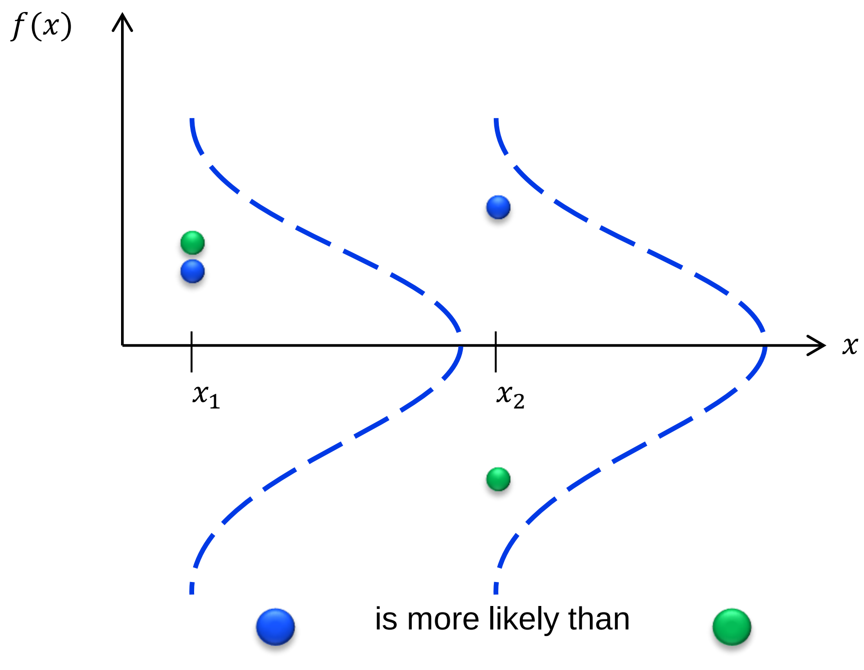
No correlation



Positive correlation

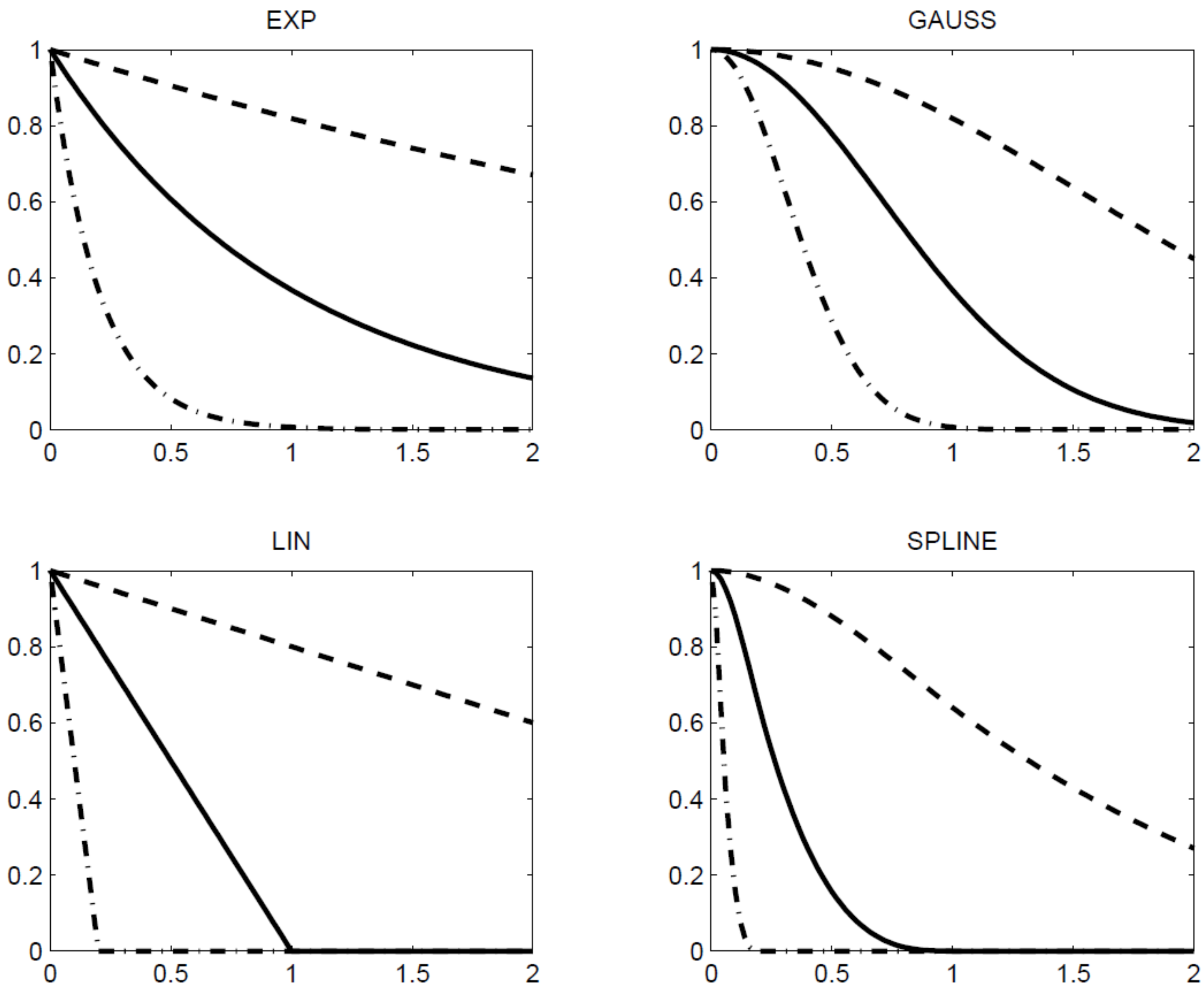


Positive correlation: If $f(x_1)$ is above the mean then $f(x_2)$ is more likely to be above the mean also



Typically: the closer x_2 is to x_1, the stronger the correlation between $f(x_2)$ and $f(x_1)$

Typical Correlation Kernels for Kriging



Influence of Gaussian Process Parameters