Verification of Probabilistic Properties using the HOL Theorem Prover

Osman Hasan Sofiene Tahar

Hardware Verification Group
Concordia University
Montreal, Canada

Integrated Formal Methods, 2007

Outline

Outline

Probabilistic Analysis

- Fundamental tool for handling random or unpredictable elements
- Applications
  - Randomized Algorithms
  - Telecommunication Protocols
  - Security Protocols
  - Digital Signal Processing
  - Machine Learning
  - and many many more $\cdots$

Probabilistic Analysis

- Model random behavior with Random Variables
- Use Probabilistic Properties of random variables to evaluate performance and reliability issues

Probabilistic Analysis

- Model random behavior with Random Variables
- Use Probabilistic Properties of random variables to evaluate performance and reliability issues

- Random Variables
  - Discrete Random Variables
    - Attain a number from a countable set of numbers
    - Example: Dice [1, 2, 3, 4, 5, 6]
  - Continuous Random Variables
    - Attain a number from an uncountable set of numbers
    - Example: Precise Inter-arrival delay (in seconds) between requests in a Web Server

Simulation

Simulation

- Functions that approximately model random variables
- Verify properties by analyzing a large number of samples

Simulation

- Functions that approximately model random variables
- Verify properties by analyzing a large number of samples

Strengths

- User friendliness
- Can handle analytically complex random systems

Simulation

- Functions that approximately model random variables
- Verify properties by analyzing a large number of samples

Strengths

- User friendliness
- Can handle analytically complex random systems

Weaknesses

- Inaccurate Results
- Enormous CPU time requirements

Simulation

- Not Suitable for the performance and reliability analysis of highly sensitive and safety critical applications



Simulation

- Not Suitable for the performance and reliability analysis of highly sensitive and safety critical applications



- Solution: Formal Methods
  - Model Checking
  - Theorem Proving

Probabilistic Model Checking

Probabilistic Model Checking

- Precise mathematical model of the random system
- Exhaustive verification of formal probabilistic properties

Probabilistic Model Checking

- Precise mathematical model of the random system
- Exhaustive verification of formal probabilistic properties

Strengths

- Precise answers
- Can be automated

Probabilistic Model Checking

- Precise mathematical model of the random system
- Exhaustive verification of formal probabilistic properties

Strengths

- Precise answers
- Can be automated

Weaknesses

- State space explosion problem
  - Can be addressed through simulation-based methods
  - At the cost of accuracy

Probability and Higher-Order-Logic Theorem Proving

- Introduction to Higher Order Logic Theorem Proving

Probability and Higher-Order-Logic Theorem Proving

- Introduction to Higher Order Logic Theorem Proving
  - Higher-Order-Logic
    - System of deduction with a precise semantics
    - Can be used to represent most classical mathematical theories

Probability and Higher-Order-Logic Theorem Proving

- Introduction to Higher Order Logic Theorem Proving
  - Higher-Order-Logic
    - System of deduction with a precise semantics
    - Can be used to represent most classical mathematical theories
  - Theorem Provers
    - Computer based formal proof tools
    - Require human assistance in the case of higher-order-logic proofs

Probability and Higher-Order-Logic Theorem Proving

Probability and Higher-Order-Logic Theorem Proving

- σ-fields and Probability [Nedzusiak, 1989]
- The σ-Additive Measure Theory [Bialas, 1990]

Probability and Higher-Order-Logic Theorem Proving

- σ-fields and Probability [Nedzusiak, 1989]
- The σ-Additive Measure Theory [Bialas, 1990]
- Formal verification of Probabilistic Algorithms in HOL [Hurd, 2002]
  - Measure and Probability Theories in HOL
  - Probabilistic Algorithms
    - Deterministic functions with access to a random Boolean Sequence $\{\mathsf{T}, \mathsf{F}, \mathsf{T}, \mathsf{T}, \mathsf{F}, \mathsf{F}, \mathsf{F}, \mathsf{T}, \mathsf{F} \cdots\}$
  - Formalization of Discrete Random Variables

Probability and Higher-Order-Logic Theorem Proving

- σ-fields and Probability [Nedzusiak, 1989]
- The σ-Additive Measure Theory [Bialas, 1990]
- Formal verification of Probabilistic Algorithms in HOL [Hurd, 2002]
  - Measure and Probability Theories in HOL
  - Probabilistic Algorithms
    - Deterministic functions with access to a random Boolean Sequence $\{\mathsf{T}, \mathsf{F}, \mathsf{T}, \mathsf{T}, \mathsf{F}, \mathsf{F}, \mathsf{F}, \mathsf{T}, \mathsf{F} \cdots\}$
  - Formalization of Discrete Random Variables

Higher-Order-Logic Theorem Proving can be used to conduct Probabilistic Analysis

- Still Immature
- Significant amount of formalization required

Theorem Proving based Probabilistic Analysis

Strengths

- Precise answers
- High Expressiveness

Theorem Proving based Probabilistic Analysis

Strengths

- Precise answers
- High Expressiveness

Weaknesses

- Significant user interaction

Proposed Approach for Probabilistic Analysis

- Combined Theorem Proving and Simulation based approach
  - Complement the short comings of one another

Proposed Approach for Probabilistic Analysis

- Combined Theorem Proving and Simulation based approach
  - Complement the short comings of one another
- More efficient and precise probabilistic analysis
  - Handle the safety critical sections using Theorem Proving
    - Exact results
    - Tedious and time consuming task
  - Handle the less critical and analytically complex sections using Simulation
    - Time efficient solutions
    - Less accurate results

Outline

Extending Theorem Proving based Probabilistic Analysis Framework

- Verification of Probabilistic Properties for both Discrete and Continuous random variables
- Formalization of Continuous Random Variables

A significant step towards a successful Theorem Proving based probabilistic analysis framework

Extending Theorem Proving based Probabilistic Analysis Framework

- Verification of Probabilistic Properties for both Discrete and Continuous random variables
- Formalization of Continuous Random Variables

A significant step towards a successful Theorem Proving based probabilistic analysis framework

- The HOL Theorem Prover
  - Hurd's framework for the verification of probabilistic algorithms
  - Comprehensive library of theorems including a theory on real analysis [Harrison, 1996]

Methodology

Methodology

- Cumulative Distribution Function

  $$CDF(X) = F_X(a) = P(X \leq a), \textit{ where a is a real number}$$

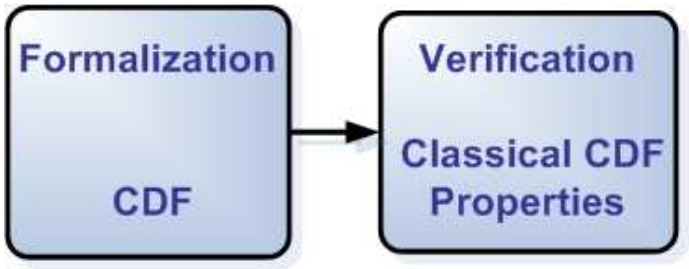
  - Completely Characterizes both Discrete and Continuous random variables

Methodology

- Cumulative Distribution Function

  $$CDF(X) = F_X(a) = P(X \leq a), \textit{ where a is a real number}$$

  - Completely Characterizes both Discrete and Continuous random variables

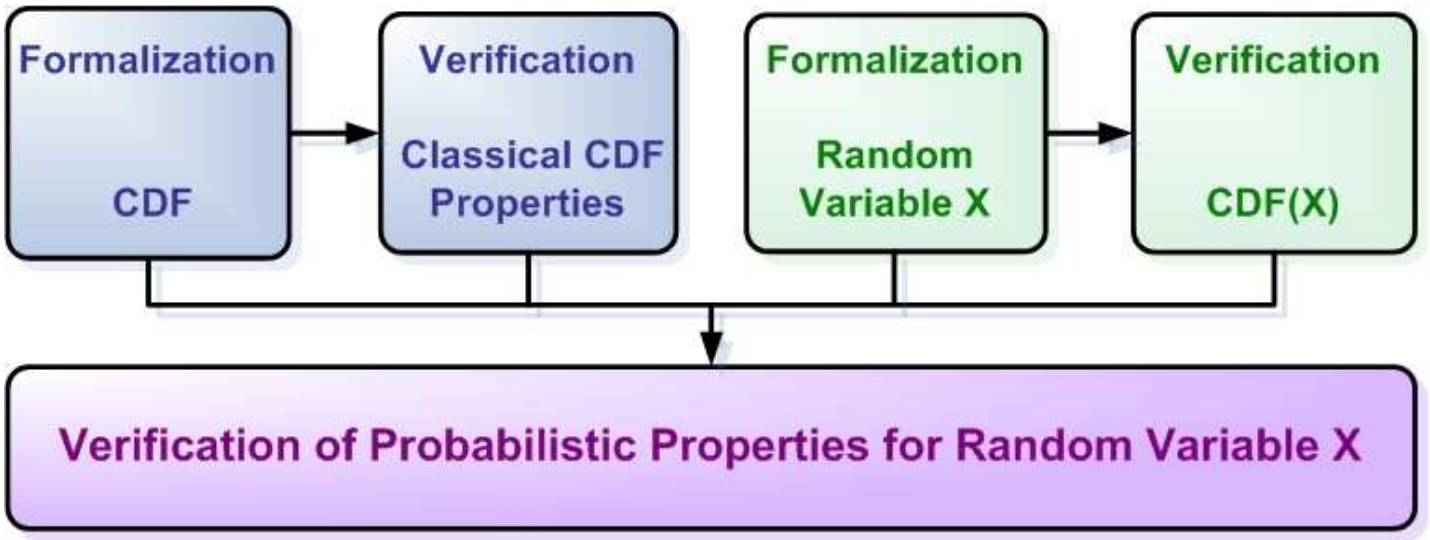


Methodology

- Cumulative Distribution Function

$$CDF(X) = F_X(a) = P(X \leq a), \textit{ where a is a real number}$$

  - Completely Characterizes both Discrete and Continuous random variables



Formalization and Verification of CDF in HOL

- Modeled as a higher-order-logic function $F_X(a)$
  - Accepts: Random Variable X, A Real Number a
  - Returns: Probability $P(X \leq a)$

Formalization and Verification of CDF in HOL

- Modeled as a higher-order-logic function $F_X(a)$
  - Accepts: Random Variable X, A Real Number a
  - Returns: Probability $P(X \leq a)$

Theorem: CDF Properties

Bounds	$\vdash \forall a, X.\ 0 \leq F_X(a) \leq 1$
Monotonic	$\vdash \forall a, b, X.\ (a < b) \Rightarrow F_X(a) \leq F_X(b)$
Positive Infinity	$\vdash \forall X.\ \lim_{n \to \infty} F_X(n) = 1$
Negative Infinity	$\vdash \forall X.\ \lim_{n \to -\infty} F_X(n) = 0$
Right Continuous	$\vdash \forall a, X.\ \lim_{n \to a^+} F_X(n) = F_X(a)$
Limit from the Left	$\vdash \forall a, X.\ \lim_{n \to a^-} F_X(n) = P(X < a)$

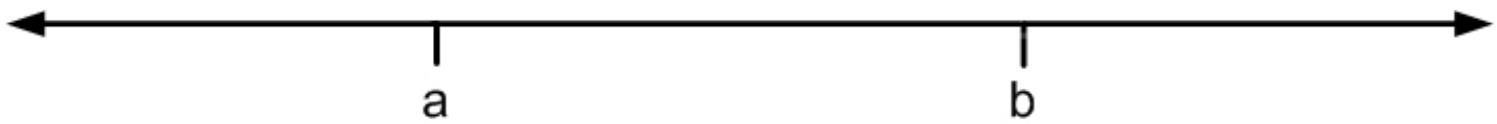
Using CDF to Verify Interval Probabilities in HOL

Probability that a Random Variable falls in any specified interval of real line

Using CDF to Verify Interval Probabilities in HOL

Probability that a Random Variable falls in any specified interval of real line

Theorem: Three Cases for Interval Probabilities

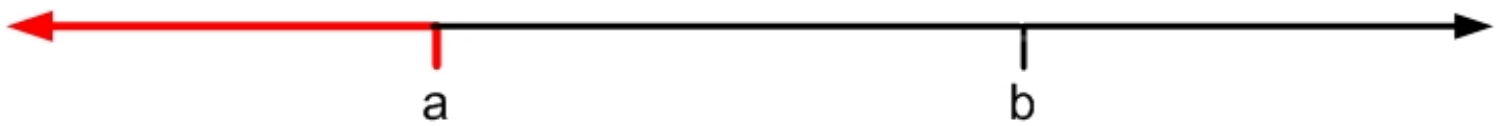


Using CDF to Verify Interval Probabilities in HOL

Probability that a Random Variable falls in any specified interval of real line

Theorem: Three Cases for Interval Probabilities

$(-\infty, a]$	$\vdash \forall a, X.\ P(X \leq a) = F_X(a)$
	$\vdash \forall a, b, X.\ P(a < X \leq b) = F_X(b) - F_X(a)$
	$\vdash \forall b, X.\ P(b < X) = 1 - F_X(b)$



Using CDF to Verify Interval Probabilities in HOL

Probability that a Random Variable falls in any specified interval of real line

Theorem: Three Cases for Interval Probabilities

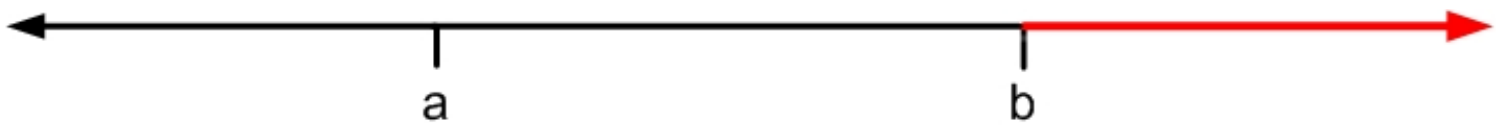
$(-\infty, a]$	$\vdash \forall a, X.\ P(X \leq a) = F_X(a)$
$(a, b]$	$\vdash \forall a, b, X.\ P(a < X \leq b) = F_X(b) - F_X(a)$
	$\vdash \forall b, X.\ P(b < X) = 1 - F_X(b)$

Using CDF to Verify Interval Probabilities in HOL

Probability that a Random Variable falls in any specified interval of real line

Theorem: Three Cases for Interval Probabilities

$(-\infty, a]$	$\vdash \forall a, X.\ P(X \leq a) = F_X(a)$
$(a, b]$	$\vdash \forall a, b, X.\ P(a < X \leq b) = F_X(b) - F_X(a)$
(b, ∞)	$\vdash \forall b, X.\ P(b < X) = 1 - F_X(b)$



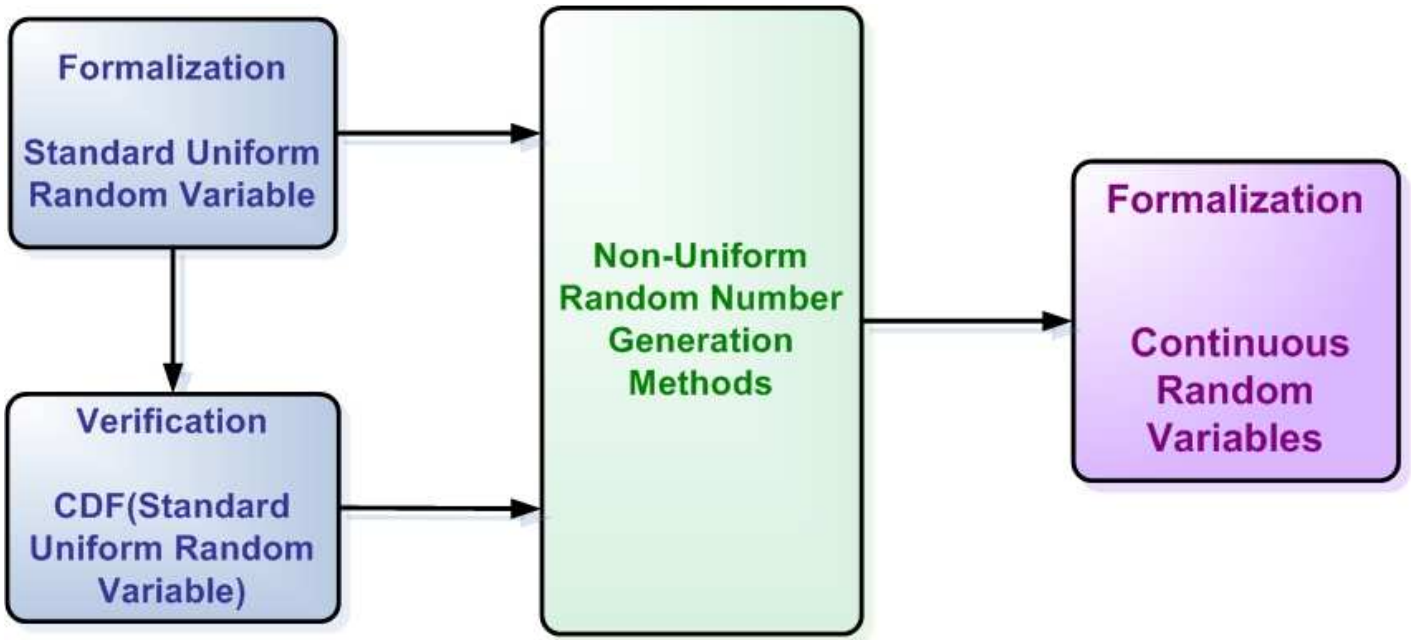
Methodology

Methodology

- Sampling algorithms are nonterminating

Methodology

- Sampling algorithms are nonterminating
- Non-Uniform Random Number Generation Techniques



Formalization of Standard Uniform Random Variable

- Continuous Uniform random variable in the interval [0,1]
- Sampling Algorithm using Infinite Coin Flips (C_k)

$$\mathsf{U} = \sum_{\mathsf{k}=1}^{\infty} (\frac{1}{2})^{\mathsf{k}} \mathsf{C_k}, \ \textit{where } C_k = \textit{if head then } 1 \textit{ else } 0$$

$$\{\mathsf{H}, \mathsf{H}, \mathsf{T}, \mathsf{T}, \mathsf{H}, \mathsf{H}, \mathsf{H}, \mathsf{T}, \mathsf{H} \cdots\} \rightarrow (0.110011101 \cdots)_2$$

Formalization of Standard Uniform Random Variable

- Continuous Uniform random variable in the interval [0,1]
- Sampling Algorithm using Infinite Coin Flips (C_k)

$$U = \sum_{k=1}^{\infty} (\frac{1}{2})^k C_k, \ \textit{where } C_k = \textit{if head then } 1 \textit{ else } 0$$

$$\{H, H, T, T, H, H, H, T, H \cdots\} \rightarrow (0.110011101 \cdots)_2$$

- Standard Uniform random variable in HOL
  - Step 1. Discrete Standard Uniform random variable

$$U_n = \sum_{k=1}^{n} (\frac{1}{2})^k B_k$$

  - Step 2. As n tends to infinity: $U = \lim_{n \rightarrow \infty} U_n$

Verification of Standard Uniform Random Variable

Theorem: CDF of Standard Uniform Random Variable

$$\vdash \forall x.\ F_U(x) = \begin{cases} 0 & \text{if } x < 0; \\ x & \text{if } 0 \leq x < 1; \\ 1 & \text{if } 1 \leq x. \end{cases}$$

Verification of Standard Uniform Random Variable

Theorem: CDF of Standard Uniform Random Variable

$$\vdash \forall x.\ F_U(x) = \begin{cases} 0 & \text{if } x < 0; \\ x & \text{if } 0 \leq x < 1; \\ 1 & \text{if } 1 \leq x. \end{cases}$$

- Proof Involves
  - Mathematical theories of Boolean Algebra, Natural Numbers, Real Numbers, Sets, Pairs, Measure and Probability
  - 94 lemmas and approx. 1800 lines of HOL code

Outline

Problem Description

- A Digital processor with uniformly distributed Roundoff Error over the interval [-0.5,0.5]
  - Roundoff Error = exact value - calculated value

Problem Description

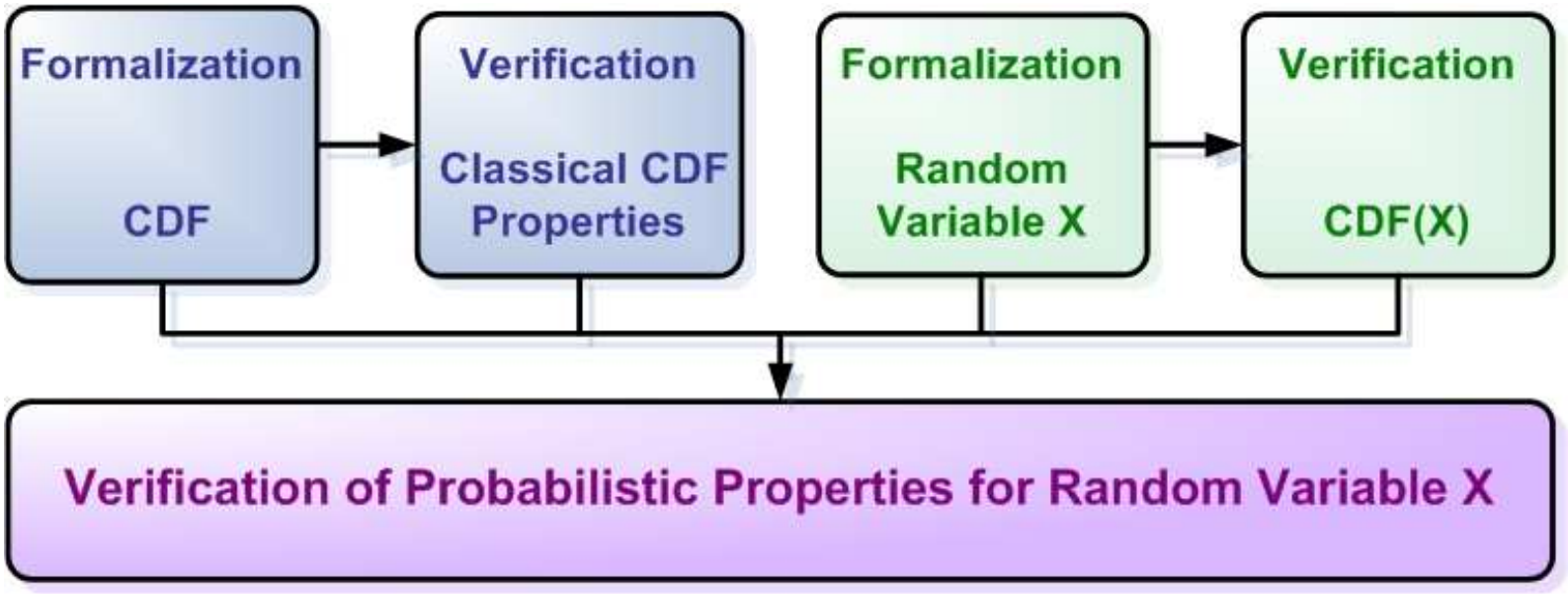
- A Digital processor with uniformly distributed Roundoff Error over the interval [-0.5,0.5]
  - Roundoff Error = exact value - calculated value
- P (Roundoff Error $>$ 0.2) $<$ $\frac{1}{3}$
- P (Result fluctuates by $\pm$0.1) = 0.2

Problem Description

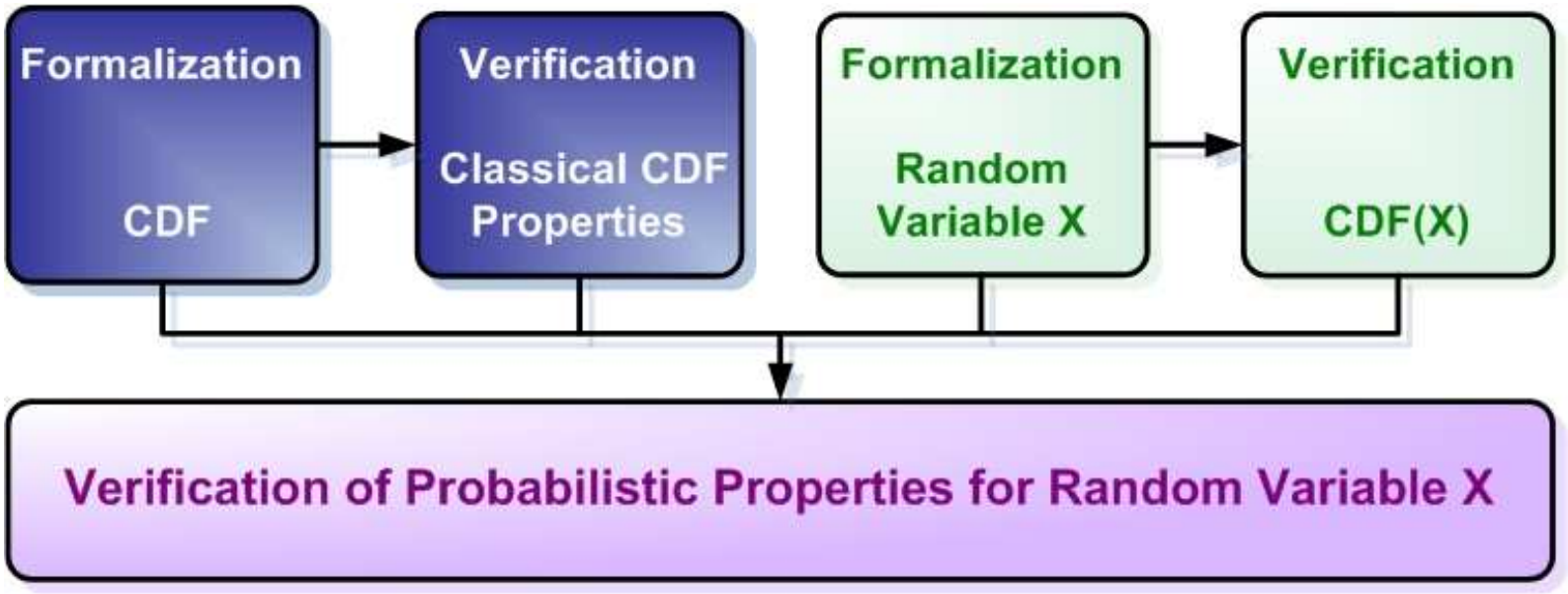
- A Digital processor with uniformly distributed Roundoff Error over the interval [-0.5,0.5]
  - Roundoff Error = exact value - calculated value
- P (Roundoff Error $>$ 0.2) $<$ $\frac{1}{3}$
- P (Result fluctuates by $\pm$0.1) = 0.2

- Continuous Uniform [-0.5,0.5] Random Variable

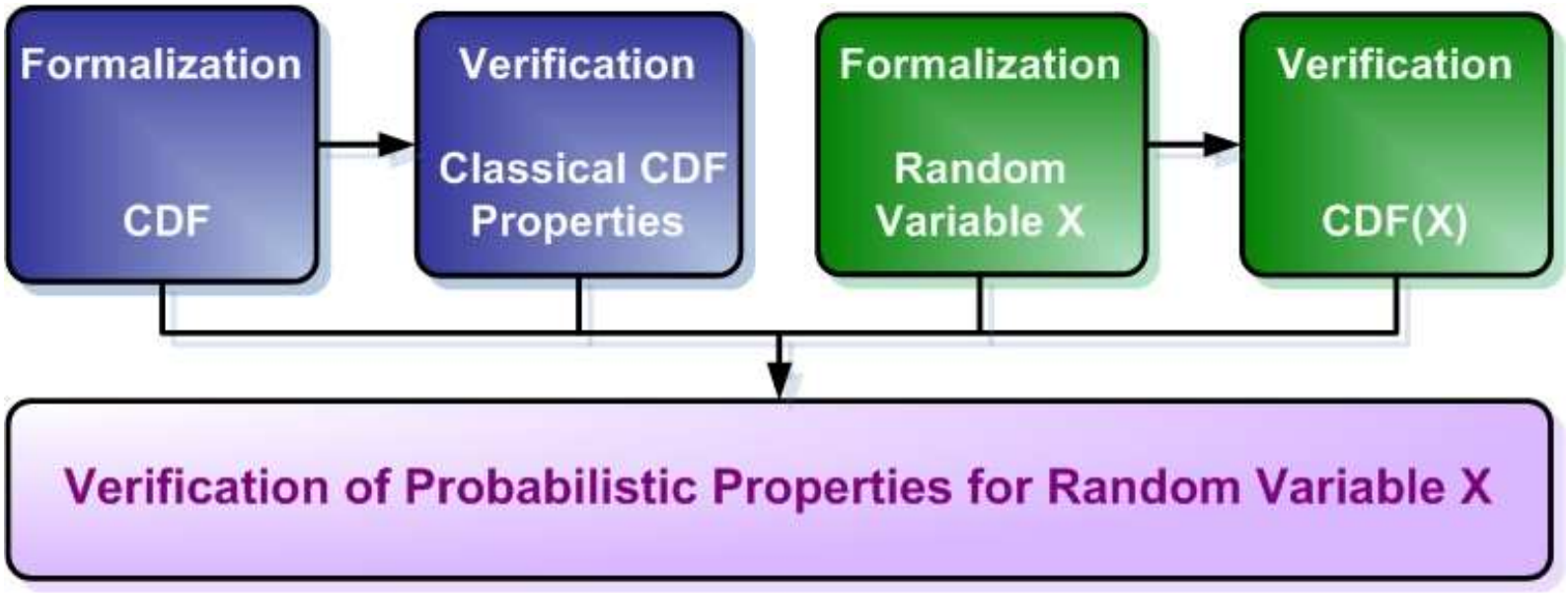
Methodology for the Verification of Probabilistic Properties



Methodology for the Verification of Probabilistic Properties

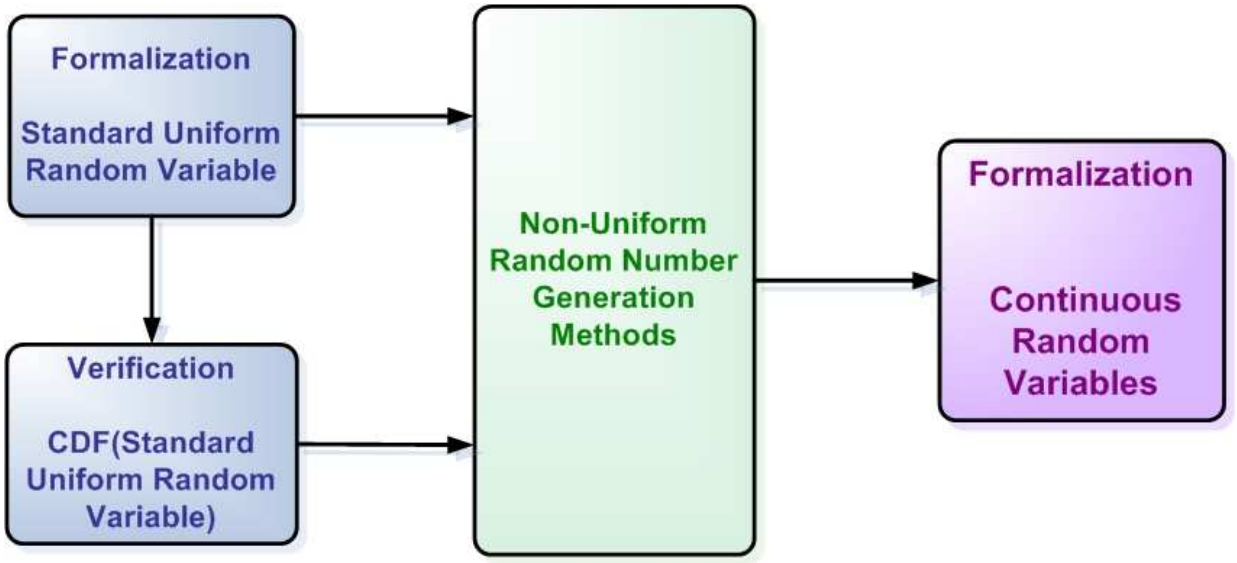


Methodology for the Verification of Probabilistic Properties

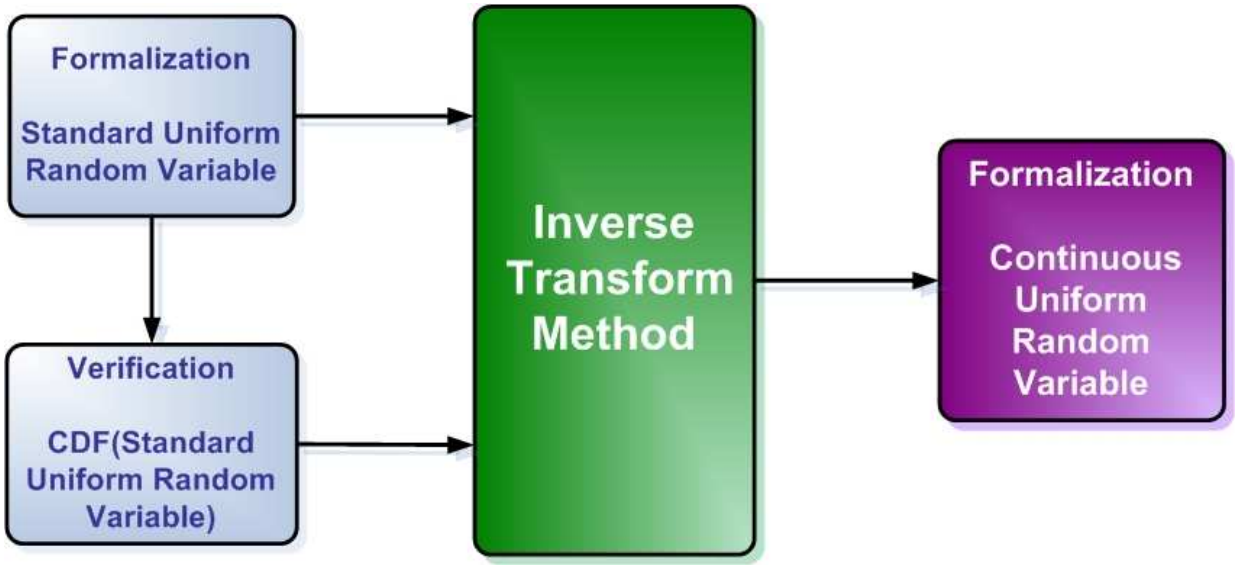


Continuous Uniform [a,b] Random Variable in HOL

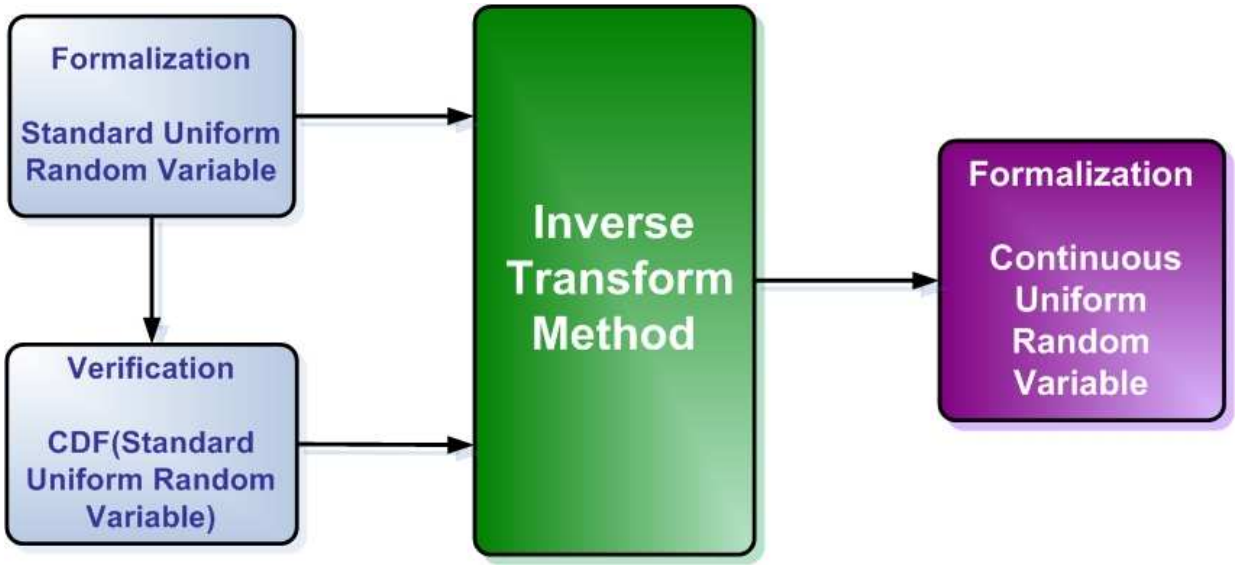
Continuous Uniform [a,b] Random Variable in HOL



Continuous Uniform [a,b] Random Variable in HOL

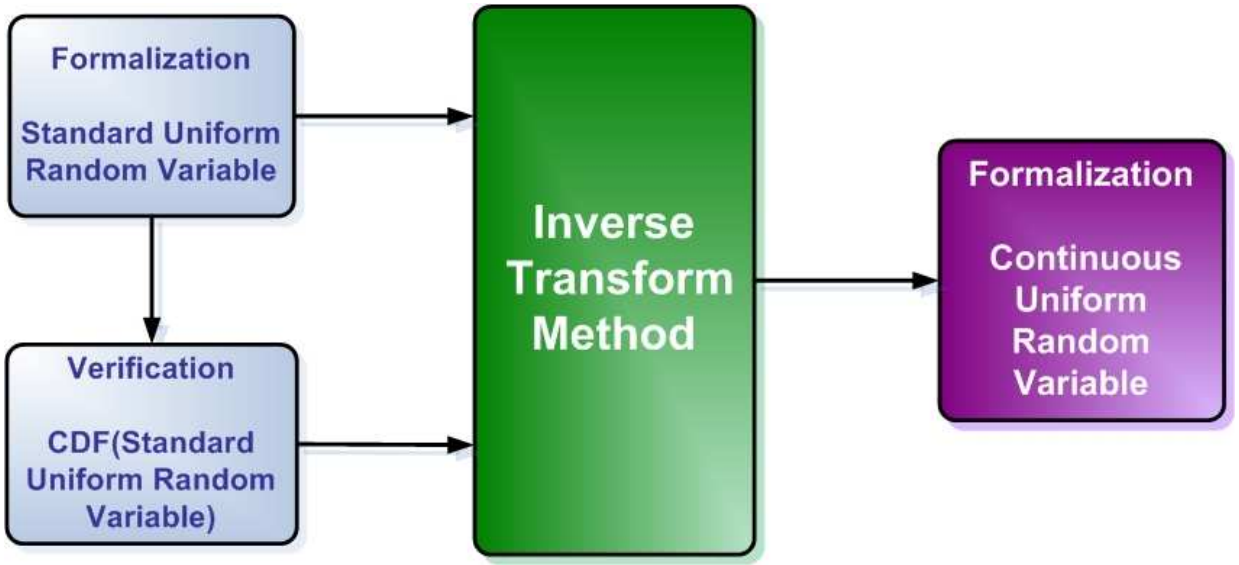


Continuous Uniform [a,b] Random Variable in HOL



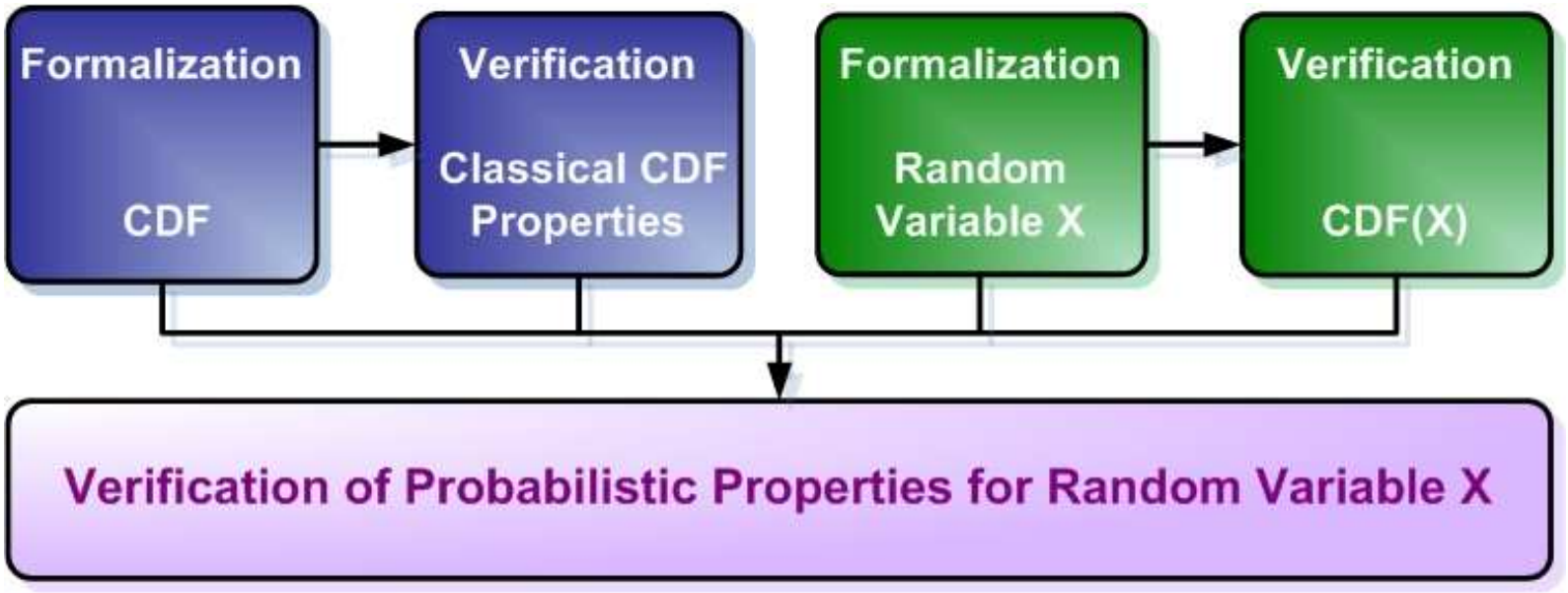
- $U_{[a,b]} = (b - a)U + a$

Continuous Uniform [a,b] Random Variable in HOL

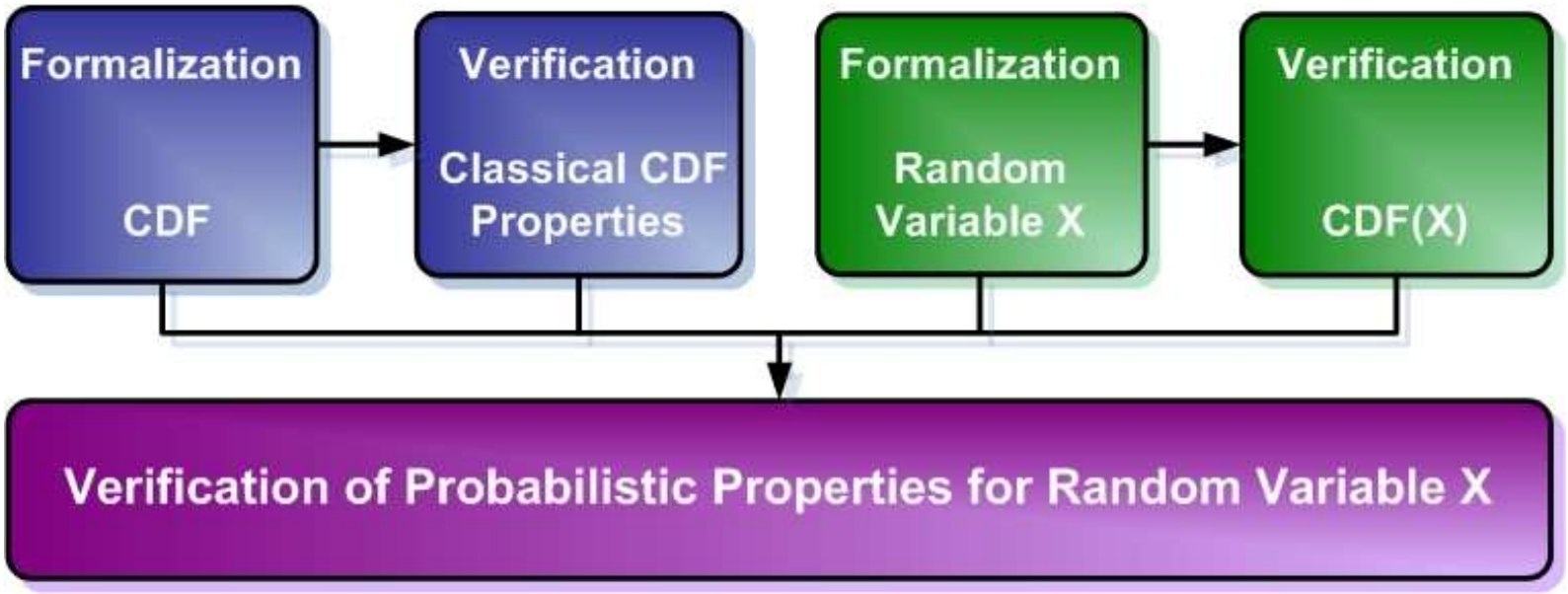


- $U_{[a,b]} = (b - a)U + a$
- CDF verification

Methodology for the Verification of Probabilistic Properties



Methodology for the Verification of Probabilistic Properties



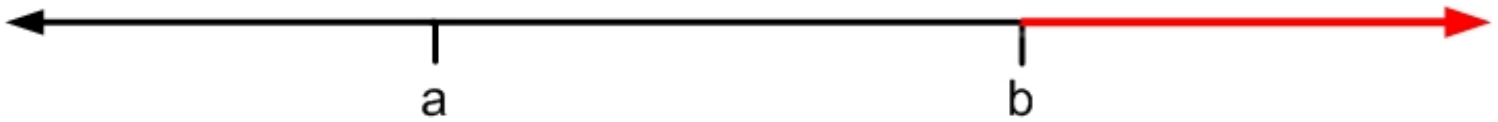
Verification of Probabilistic Properties

Verification of Probabilistic Properties

- P (Roundoff error > 0.2) $< \frac{1}{3}$

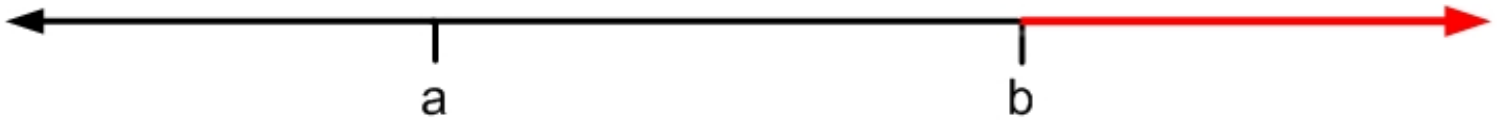
Verification of Probabilistic Properties

- $P(\text{Roundoff error} > 0.2) < \frac{1}{3}$



Verification of Probabilistic Properties

- P (Roundoff error $> 0.2) < \frac{1}{3}$

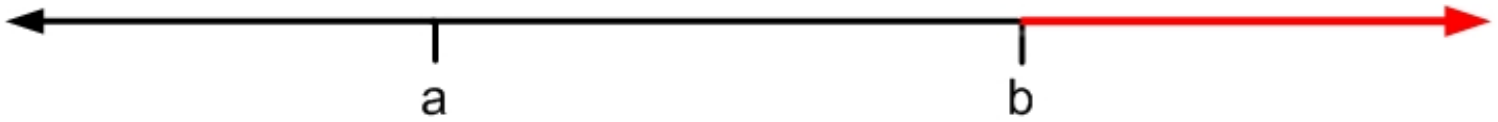


Theorem: Property 1

$P(U_{[-0.5,0.5]} > 0.2) < \frac{1}{3}$

Verification of Probabilistic Properties

- P (Roundoff error > 0.2) $< \frac{1}{3}$



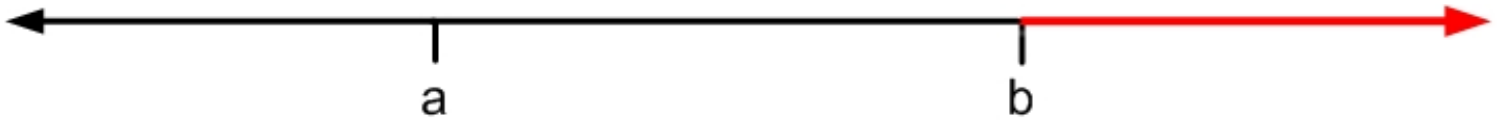
Theorem: Property 1

$P(U_{[-0.5,0.5]} > 0.2) < \frac{1}{3}$

- P (Result fluctuates by ± 0.1) = 0.2

Verification of Probabilistic Properties

- P (Roundoff error > 0.2) $< \frac{1}{3}$



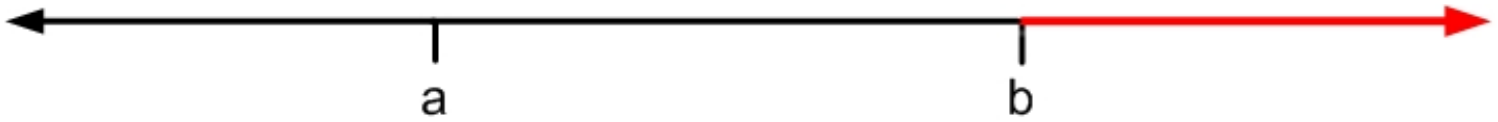
Theorem: Property 1

$P(U_{[-0.5,0.5]} > 0.2) < \frac{1}{3}$

- P (Result fluctuates by ± 0.1) = 0.2

Verification of Probabilistic Properties

- P (Roundoff error $> 0.2) < \frac{1}{3}$



Theorem: Property 1

$$P(U_{[-0.5,0.5]} > 0.2) < \frac{1}{3}$$

- P (Result fluctuates by ± 0.1) = 0.2

Theorem: Property 2

$$P(-0.1 < U_{[-0.5,0.5]} \leq 0.1) = 0.2$$

Outline

Conclusions

Conclusions

- Verification of Probabilistic Properties in HOL
- Formalization of Continuous Random Variables in HOL

Conclusions

- Verification of Probabilistic Properties in HOL
- Formalization of Continuous Random Variables in HOL
- Tedious and time consuming task
- Precise Probabilistic Analysis
- Theorems can be reused to formalize more complex mathematical concepts

Conclusions

- Verification of Probabilistic Properties in HOL
- Formalization of Continuous Random Variables in HOL
- Tedious and time consuming task
- Precise Probabilistic Analysis
- Theorems can be reused to formalize more complex mathematical concepts

- Good indicator of the State-of-the-art in Higher-Order-Logic Theorem Proving

Ongoing/Future Work

Ongoing/Future Work

- Formalization of Non-Uniform Random Generation Methods
  - Inverse Transform Method
  - Acceptance/Rejection
  - Box-Muller

Ongoing/Future Work

- Formalization of Non-Uniform Random Generation Methods
  - Inverse Transform Method
  - Acceptance/Rejection
  - Box-Muller
- Library of Continuous random variables
  - Uniform, Exponential, Rayleigh and Triangular
  - Normal

Ongoing/Future Work

- Formalization of Non-Uniform Random Generation Methods
  - Inverse Transform Method
  - Acceptance/Rejection
  - Box-Muller
- Library of Continuous random variables
  - Uniform, Exponential, Rayleigh and Triangular
  - Normal
- Verification of Moments and Bounds
  - Mean
  - Variance

Ongoing/Future Work

- Formalization of Non-Uniform Random Generation Methods
  - Inverse Transform Method
  - Acceptance/Rejection
  - Box-Muller
- Library of Continuous random variables
  - Uniform, Exponential, Rayleigh and Triangular
  - Normal
- Verification of Moments and Bounds
  - Mean
  - Variance
- Case Studies
  - Coupon Collector's problem
  - Suggestions are welcome!

More details and HOL sources

- HVG Concordia: http://hvg.ece.concordia.ca
- Contact: o_hasan@ece.concodia.ca

Thank You!