Formalization of Continuous Probability Distributions

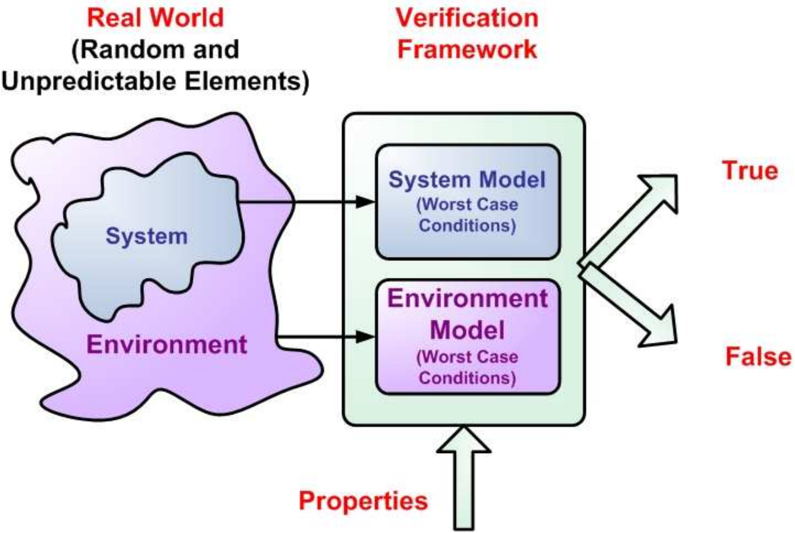
Osman Hasan   Sofiène Tahar

Hardware Verification Group
Concordia University, Montreal, Canada

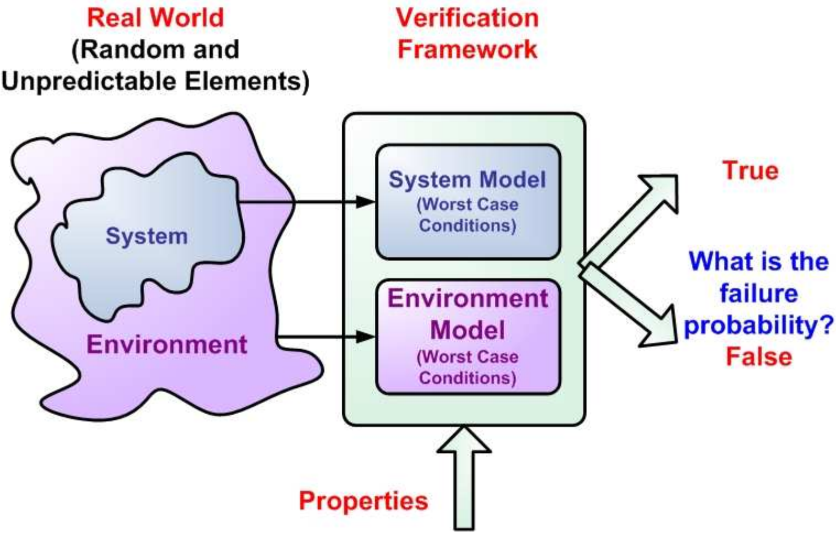
CADE, 2007

Outline

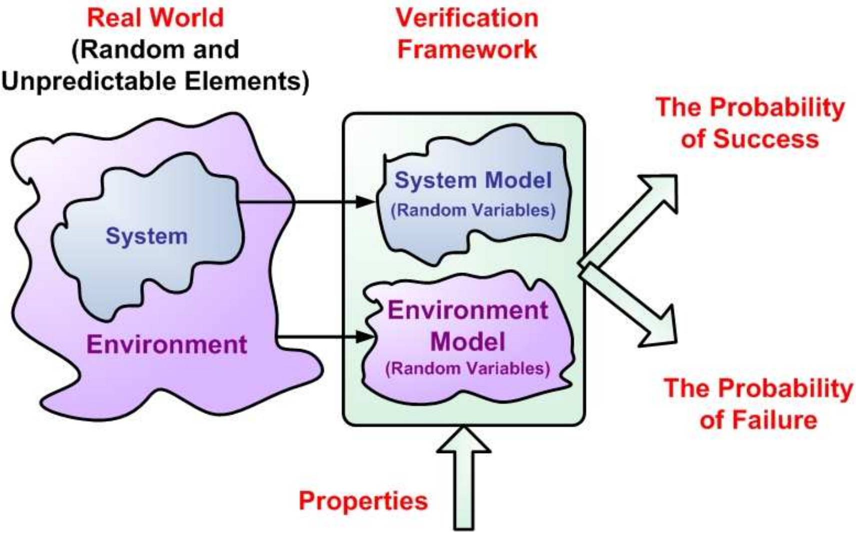
System Verification



System Verification



System Verification



Random Variables

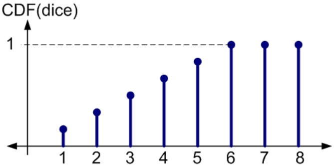
- Functions that map random events to numbers
- Discrete random variables
  - Attain a countable number of values from an interval of real numbers
  - Example: Dice
    - Interval: [1,6]
    - Possible Outcomes: {1, 2, 3, 4, 5, 6}
- Continuous random variables
  - Attain all values from an interval of real numbers
  - Example: Arithmetic Roundoff Error [-0.5, 0.5]
    - Interval [-0.5, 0.5]
    - Possible Outcomes: Infinite or Uncountable

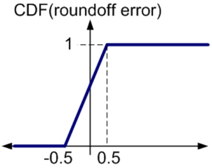
Probabilistic Properties

- Most probabilistic properties associated with a random variable can be expressed in terms of its Cumulative Distribution Function (CDF)
  - Accepts a real number x
  - Returns the probability that the random variable is less than or equal to x

$$CDF(R) = P(R \leq x)$$

- CDF can be used to characterize both Discrete and Continuous random variables





Simulation

Simulation

- Model: Using approximate random variable functions
- Verification: Analyzing a large number of samples

Simulation

- Model: Using approximate random variable functions
- Verification: Analyzing a large number of samples

Strengths

- User friendliness
- Can handle analytically complex random systems

Simulation

- Model: Using approximate random variable functions
- Verification: Analyzing a large number of samples

Strengths

- User friendliness
- Can handle analytically complex random systems

Weaknesses

- Inaccurate results
- Enormous CPU time requirements

Probabilistic Model Checking

Probabilistic Model Checking

- Model: Probabilistic state machine
- Verification: Exhaustive

Probabilistic Model Checking

- Model: Probabilistic state machine
- Verification: Exhaustive

Strengths

- Precise answers
- Verification is automatic

Probabilistic Model Checking

- Model: Probabilistic state machine
- Verification: Exhaustive

Strengths

- Precise answers
- Verification is automatic

Weaknesses

- State space explosion problem
  - Can be addressed through simulation-based methods at the cost of accuracy
- Limited to systems that are memoryless (Markov Chains)

Probabilistic Verification and Theorem Proving

Probabilistic Verification and Theorem Proving

- Model: Using Higher-Order-Logic functions for random variables
- Verification: Theorem Proving

Probabilistic Verification and Theorem Proving

- Model: Using Higher-Order-Logic functions for random variables
- Verification: Theorem Proving

Strengths

- Precise answers
- High Expressiveness

Probabilistic Verification and Theorem Proving

- Model: Using Higher-Order-Logic functions for random variables
- Verification: Theorem Proving

Strengths

- Precise answers
- High Expressiveness

Weaknesses

- Significant user interaction
- Immature: A huge amount of formalization is required

Related Work

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

- Theorem proving with the Real Numbers [Harrison, 1996]

- Formal verification of Probabilistic Algorithms in HOL [Hurd, 2002]
  - Deterministic functions with access to a random Boolean Sequence
  - Formalization of Discrete Random Variables

- Proofs of Randomized Algorithms in Coq [Audebaud *et. al*, 2006]

Related Work

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

- Theorem proving with the Real Numbers [Harrison, 1996]

- Formal verification of Probabilistic Algorithms in HOL [Hurd, 2002]
  - Deterministic functions with access to a random Boolean Sequence
  - Formalization of Discrete Random Variables

- Proofs of Randomized Algorithms in Coq [Audebaud *et. al*, 2006]

There is no machine-checked formalization of Continuous random variables

Formalization of Continuous Random Variables

- Framework for the formalization of Continuous random variables for which CDF exists in a closed mathematical form

- Minimize the formalization and verification effort
  - Reasoning based on Measure and Probability theories is not required

Formalization of Continuous Random Variables

- Framework for the formalization of Continuous random variables for which CDF exists in a closed mathematical form

- Minimize the formalization and verification effort
  - Reasoning based on Measure and Probability theories is not required

- The HOL Theorem Prover
  - Higher-Order-Logic interactive Theorem Prover
  - Hurd's framework for the verification of probabilistic algorithms
  - Comprehensive library of theorems including Harrison's theories on real analysis

Formalization of Continuous Random Variables

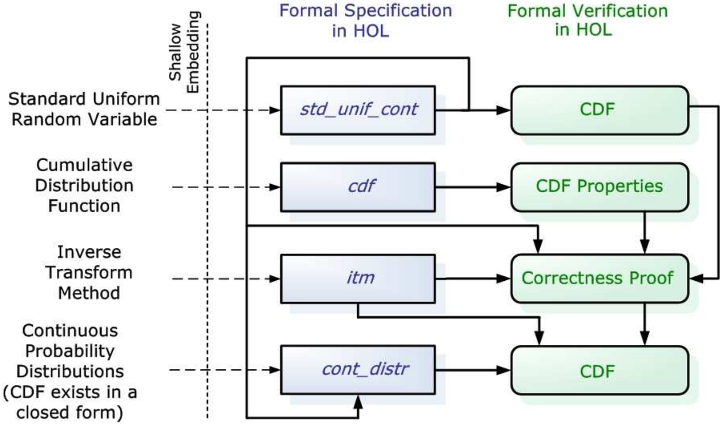
- Sampling algorithms are nonterminating
  - Tedious formalization and verification

Formalization of Continuous Random Variables

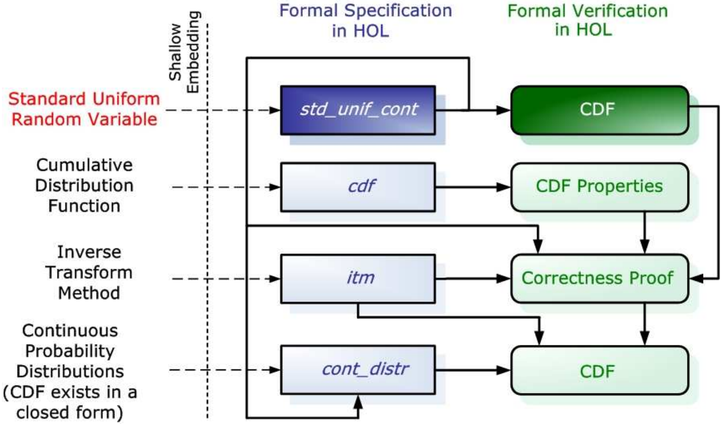
- Sampling algorithms are nonterminating
  - Tedious formalization and verification
- Inverse Transform Method
  - Extensively used method in Non-uniform random number generation

- Standard Uniform random number generator generates uniformly distributed random real numbers in the interval [0,1]

Methodology



Methodology



Formalization of Standard Uniform Random Variable

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

$$\mathsf{U} = \sum_{\mathsf{k}=1}^{\infty}\left(\frac{\mathsf{C}_\mathsf{k}}{2^\mathsf{k}}\right),\ \textit{where } C_k = 1 \textit{ if } k^{th} \textit{ coin returns a head else } 0$$

  - $\{\mathsf{H}, \mathsf{H}, \mathsf{T}, \mathsf{H}, \cdots\} \rightarrow (\frac{1}{2^1} + \frac{1}{2^2} + \frac{0}{2^3} + \frac{1}{2^4} + \cdots) = (0.1101\cdots)_2$

Formalization of Standard Uniform Random Variable

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

$$\mathrm{U} = \sum_{\mathrm{k}=1}^{\infty}\left(\frac{\mathrm{C_k}}{2^{\mathrm{k}}}\right), \textit{ where } C_k = 1 \textit{ if } k^{th} \textit{ coin returns a head else } 0$$

  - $\{\mathrm{H}, \mathrm{H}, \mathrm{T}, \mathrm{H}, \cdots\} \rightarrow \left(\frac{1}{2^1} + \frac{1}{2^2} + \frac{0}{2^3} + \frac{1}{2^4} + \cdots\right) = (0.1101\cdots)_2$

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

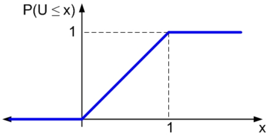
$$\mathrm{U_n} = \sum_{\mathrm{k}=1}^{\mathrm{n}}\left(\frac{\mathrm{B_k}}{2^{\mathrm{k}}}\right), \textit{ where } B_k = 1 \textit{ if } k^{th} \textit{ random bit is a True else } 0$$

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

Verification of Standard Uniform Random Variable

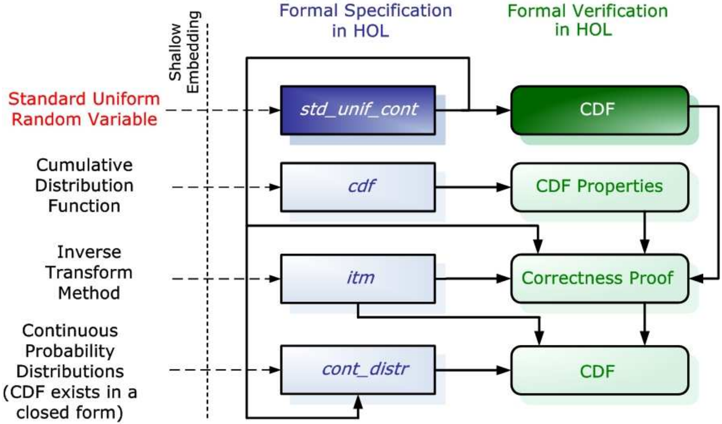
Theorem: CDF of Standard Uniform Random Variable

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

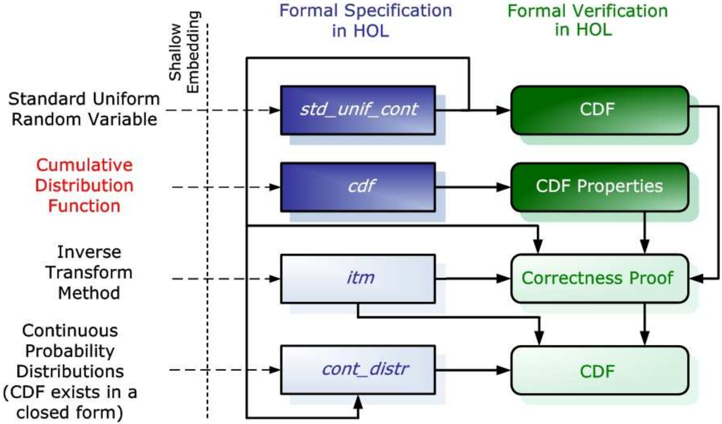


- Proof Sketch
  - Verify CDF for the discrete Standard Uniform random variable
  - Take the limit as n approaches infinity

Methodology



Methodology



Formalization and Verification of CDF

- 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

- 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	$\vdash \forall a, b, X.\ (a < b) \Rightarrow F_X(a) \leq F_X(b)$
Interval Probability	$\vdash \forall a, b, X.\ (a < b) \Rightarrow P(a < X \leq b) = F_X(b) - F_X(a)$
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	$\vdash \forall a, X.\ \lim_{n \to a^-} F_X(n) = P(X < a)$

Formalization and Verification of CDF

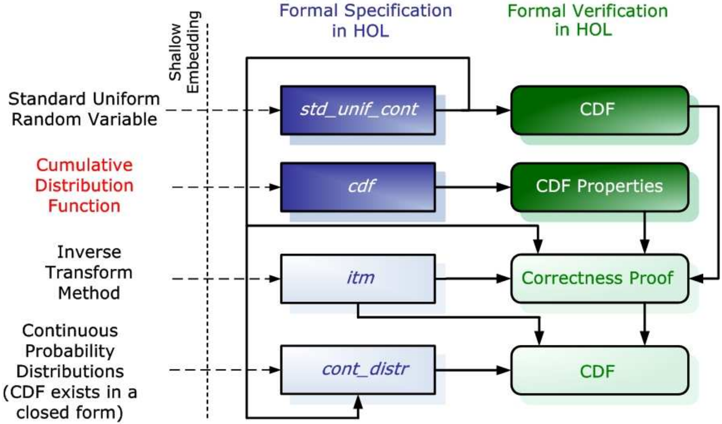
- Modeled as a higher-order-logic function $F_X(a)$
  - Accepts: Random Variable X, A Real Number a
  - Returns: Probability $P(X \leq a)$

Theorem: CDF Properties

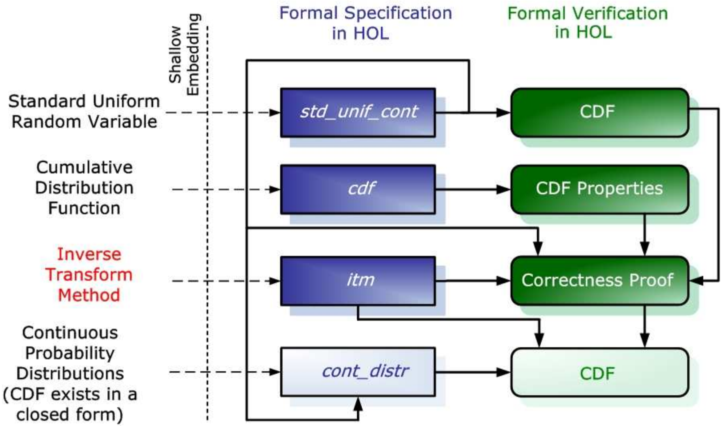
Bounds	$\vdash \forall a, X.\ 0 \leq F_X(a) \leq 1$
Monotonic	$\vdash \forall a, b, X.\ (a < b) \Rightarrow F_X(a) \leq F_X(b)$
Interval Probability	$\vdash \forall a, b, X.\ (a < b) \Rightarrow P(a < X \leq b) = F_X(b) - F_X(a)$
Positive Infinity	$\vdash \forall X.\ \lim_{n \to \infty} F_X(n) = 1$
Negative 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	$\vdash \forall a, X.\ \lim_{n \to a^-} F_X(n) = P(X < a)$

- Verification of Probabilistic Properties in HOL, IFM 07

Methodology



Methodology



Inverse Transform Method

- A random variable, X, with well-defined CDF F

$$X = F^{-1}(U)$$

  - U = Standard Uniform random variable
  - F^{-1} = Inverse function of F

Inverse Transform Method

- A random variable, X, with well-defined CDF F

$$X = F^{-1}(U)$$

  - U = Standard Uniform random variable
  - F^{-1} = Inverse function of F

Predicate	Input	Data type	True
is_cdf	g	$(real \rightarrow real)$	If g is a valid CDF
inv_fn	f, g	$(real \rightarrow real)$	If f is the inverse function of g

Inverse Transform Method

- A random variable, X, with well-defined CDF F

$$X = F^{-1}(U)$$

  - U = Standard Uniform random variable
  - F^{-1} = Inverse function of F

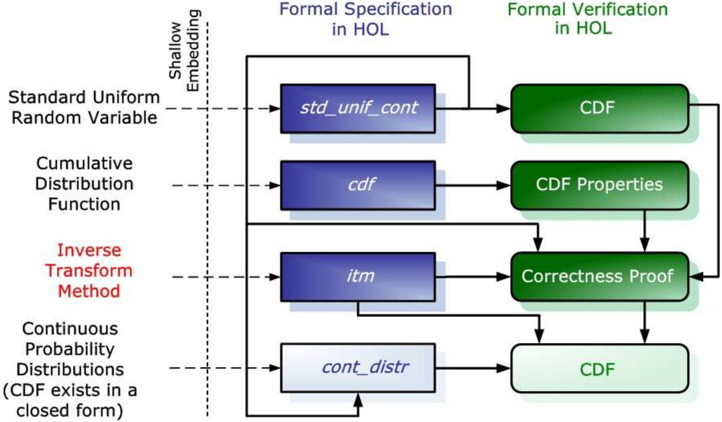
Predicate	Input	Data type	True
is_cdf	g	$(real \rightarrow real)$	If g is a valid CDF
inv_fn	f, g	$(real \rightarrow real)$	If f is the inverse function of g

Theorem: Inverse Transform Method

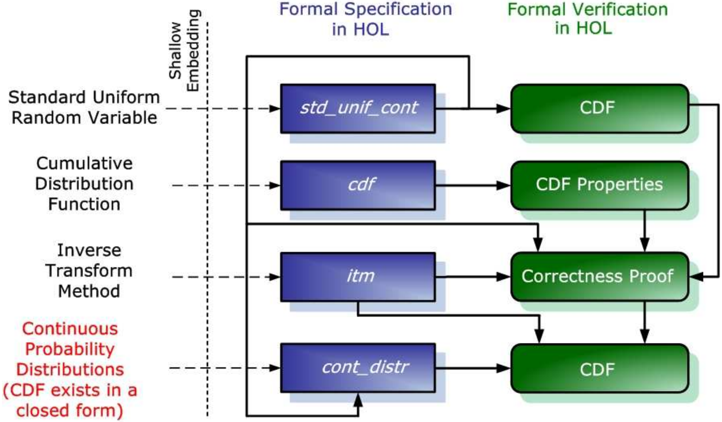
$$\vdash \forall \mathtt{f}, \mathtt{g}, \mathtt{x}.\ (\mathtt{is\_cdf}\ \mathtt{g}) \wedge (\mathtt{inv\_fn}\ \mathtt{f}\ \mathtt{g}) \Rightarrow (\mathtt{F}_{\mathtt{f(U)}}(\mathtt{x}) = \mathtt{g}(\mathtt{x}))$$

- Proof utilizes CDF of the Standard Uniform random variable and CDF properties

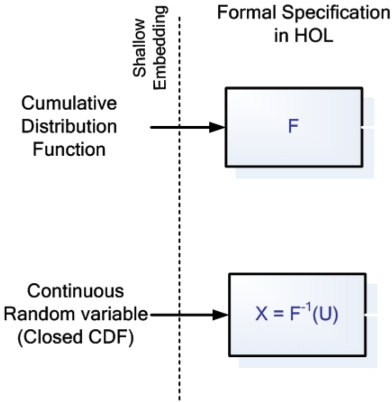
Methodology



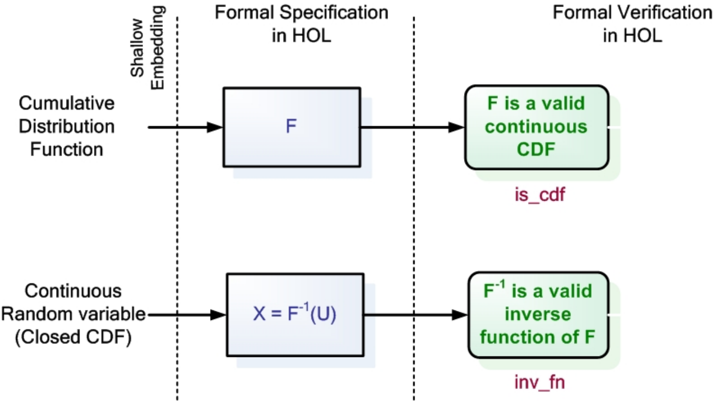
Methodology



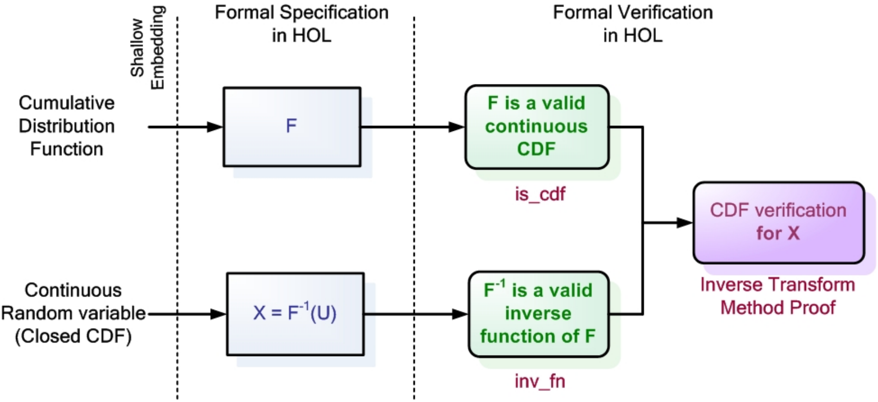
Continuous Random Variables



Continuous Random Variables



Continuous Random Variables



Continuous Random Variables

Theorem: Continuous Random Variables

Distribution	**CDF**	**Random Variable**
Exponential(l)	$0 \quad x \leq 0$ $1 - e^{-lx} \quad 0 < x$	$(\lambda x. -\frac{1}{l}\ln(1-x))U$
Uniform(a, b)	$0 \quad x \leq a$ $\frac{x-a}{b-a} \quad a < x \leq b$ $1 \quad b < x$	$(\lambda x.(b-a)x + a)U$
Rayleigh(l)	$0 \quad x \leq 0$ $1 - e^{\frac{-x^2}{2l^2}} \quad 0 < x$	$(\lambda x.l\sqrt{-2\ln(1-x)})U$
Triangular(a)	$0 \quad x \leq 0$ $\frac{2}{a}(x - \frac{x^2}{2a}) \quad x < a$ $1 \quad a \leq x$	$(\lambda x.a(1 - \sqrt{1-x}))U$

Applications: Continuous Random Variables

Applications: Continuous Random Variables

- Sources of Error in Computer Arithmetic
  - Uniform random variable
- Inter-Arrival and Service times in Telecommunication Networks
  - Exponential random variable
- Noise signal in Telecommunication Receivers
  - Rayleigh random variable
- Randomized Algorithms
- Security Protocols
- Machine Learning
- and many many more $\cdots$

Conclusions

Conclusions

- Summary
  - Formalization framework for Continuous random variables in HOL
  - Simple to use approach
  - Precise Probabilistic Analysis

Conclusions

- Summary
  - Formalization framework for Continuous random variables in HOL
  - Simple to use approach
  - Precise Probabilistic Analysis

- Future Work
  - Verification of Statistical properties (Mean, Variance)
  - Multiple random variables
  - Case studies: Suggestions are welcome

Thank You

More details and HOL sources

HVG Concordia: http://hvg.ece.concordia.ca

Contact: o_hasan@ece.concordia.ca