Formalization of the Heavy Hitter Problem in HOL theorem prover

Ghassen Helali, Osman Hasan, and Sofiène Tahar

Department of Electrical and Computer Engineering,
Concordia University, Montreal, Canada
{helali,o_hasan,tahar}@encs.concordia.ca

Technical Report

January, 2012

Abstract

Dynamic and real systems that exhibit probabilistic behavior represent a large class of man- made systems such as communication networks, air traffic control and other mission critical systems. Evaluation of quantitative issues like performance and dependability of these systems is of paramount importance. Probabilistic analysis is an indispensable tool for all scientists and engineers since they are often dealing with systems containing elements that exhibit random or unpredictable behavior. Traditionally, computer simulation techniques have been used to perform probabilistic analysis. However, they provide less accurate results and cannot handle large problems due to enormous computer processing time requirements. We have developed a generalized framework for the probabilistic analysis of systems using the HOL theorem prover. We present the formalization of extended reals in Higher-Order logic. This formalization is used to formalize other theories such as Measure and Probability theories. We then used our formalization of extended real numbers and the already existing formalization of Measure, Lebesgue Integration and Probability theories [14, 13] to model an algorithm for the Heavy Hitter problem. This model is then utilized to formally verify some interesting probabilistic and statistical properties associated with the heavy hitter problem in HOL.

1 Introduction

Hardware and software systems usually contain random or unpredictable components [10]. Even though the characteristics of technical systems do not drastically change over their useful life, they however do vary in reality. These realistic aspects such as variability, uncertainty, tolerance and error have to be considered in the design of reliable technical systems. Probability distributions are aimed to characterize the behavior caused by manufacturing inaccuracies, process uncertainties, environment influences, abrasion, and human factors etcetera.

The deterministic simulation cannot predict the real system behaviors, because one nominal simulation shows only one point in the design space. In addition to that, those systems are running over complex environment which themselves are also characterized by an unpredictable and random behavior due to a lot of external and internal factors such as noise, environment conditions. So for that reason, improving the quality and ensuring a higher level of reliability of such systems become expensive.

The engineering approach to analyze a system with these kinds of unavoidable elements of randomness and uncertainty is to use probabilistic analysis [10]. For that kind of systems, improving the quality of service and increasing the level of system performance can be ensured by using probabilistic analysis. In fact, the term system performance means the average time needed by a system to perform a given task, such as the average runtime of a computational algorithm or the average message delay of a telecommunication protocol.

The above averages can be computed, based on the probabilistic analysis approach, by using appropriate random variables to model inputs for the system model. Simulation is one of the most common used probabilistic analysis techniques. It allows to realize probabilistic analysis of randomized models but the time required is usually long. In fact, simulation needs a huge number of computations and repetitions to get meaningful results and the answers can not be 100% accurate. Nowadays, the results of simulation-based analysis of hardware and software systems must be precise and accurate because of the extensive usage of these systems in safety and financial critical areas, such as, medicine, transportation and stock exchange markets. This is a reason why accurate alternatives to simulation are needed for a reliable analysis of such systems.

Formal methods [8] overcome the limitations of the simulation approaches, so that, they could be allowed to conduct such precise system analysis. The principle of formal analysis is to construct a computer based mathematical model of the given system and formally verify, within a computer, that this model meets the given specifications. Two of the most commonly used formal verification methods are model checking [4] and higher-order logic theorem proving [7]. The first technique is an automated approach for systems verification that can be expressed as a finite-state machine. The second one is an interactive approach but is more flexible in terms of tackling a variety of systems.

The use of both model checking and theorem proving has showed successful results for the precise functional correctness of a lot of hardware and software systems. But on the other side, their usage for probabilistic analysis has some limitations. For the model checking, the

main limitations consist of the lack of expressibility and the inability to reason about statistical properties, and for the theorem proving, there is a lack of mathematical foundations required to conduct such proofs.

2 Related Work

A lot of work has been done in the area of probabilistic analysis using higher-order-logic theorem proving. The pioneering work in this area was done by Nedzusiak [15] and Bialas [3] proposed a formalization of measure and probability theories in Mizar theorem prover. After that, Hurd [11] implemented their work and developed a formalization of measure theory in HOL, upon which he constructed definitions of probability spaces and functions on them. Despite important contributions in the analysis of probabilistic algorithms, Hurd's work has some limitations. For example, his formalization did not include basic concepts in statistics such as the expectation of random variables which is essential in performing analysis of probabilistic systems. Besides, in Hurd's formalization, a measure space is the pair (A, μ), A is a set of subsets of X, the state space, called the set of measurable sets and μ is a measure function. In this formalization he considered the space the universal set which can not allow constructing measure spaces where the space is not the universal set.

Based on the work of Hurd [11], Richter [16] formalized the measure theory in Isabelle/HOL. Because of this, he inherited the same restriction on the measure spaces that can be constructed. Richter [16] defined the Borel sets as being generated by the intervals.
Whereas, in our proposed formalization, the Borel sigma algebra is generated by open sets and is more general as it can be applied not only to the real numbers but to any metric space, such as, complex numbers or $\mathbb{R}^n$, the n-dimensional Euclidean space. The proposed formalization also provides a unified framework to prove the measurability theorems in these spaces.

In another related work, Coble [1] generalized the measure theory formalization by Hurd [11] and then built on it to formalize the Lebesgue integration theory in HOL. He proved some properties of the Lebesgue integral but only for the class of positive simple functions. Besides, multiple theorems in Coble's work have the assumption that every set is measurable which is not correct in most cases of interest.

Hasan [10] built upon Hurd's formalizations of measure and probability theories to verify the probabilistic and statistical properties of some commonly used discrete random variables. He also formalized probabilistic properties of several continious random variables commonly used in performance analysis. The results were then utilized to formally reason about the correctness of many real-world systems that exhibit probabilistic behavior. Hasan's work inherits the above mentioned limitations of the previous works of Hurd and Coble.

Lester [12] formalized topology theory in PVS theorem prover, his work also provided the formalizations of measure and integration theories. Lester's formalization lacks the proofs of the properties of the Lebesgue integral as well as the Lebesgue convergence theorems, both of which are very important to the usability of the formalization to analyze systems properties.

T.Mhamdi [14] developed a framework for probabilistic and information theoretic analysis in the HOL theorem prover environment. This work consists on the formalization of the mathematical theories of measure, probability, Lebesgue integration, rational number, topology concepts as well as the fundamental concepts of information theory.

In that work, the Hardware Verification Group team tends to implement a robust framework based on those related work and of course overcome all the discovering shortcomings.

3 Probabilistic Analysis in HOL

This section briefly summarizes Mhamdi's formalization [13] of topology of the Measure and the probability theories and some other fundamentals probabilistic analysis fundamentals that we would be building upon to analyze the heavy hitter problem in the coming sections.

3.1 Measure Theory

A measure is a way to assign a number to a set, which can be interpreted as its size. It can be considered as a generalization of the concepts of length, area, volume, etc. Two important examples are the Lebesgue measure on an Euclidean space and the probability measure on a Borel space. A measure function is defined over a class of subsets, called the measurable sets, and assigns a non-negative real number to every measurable set. Some of the important definitions of Measure theory, formalized in [13], are given below:

- **Sigma Algebra:** *It contains the empty set $\emptyset$, is closed under countable unions and complementarity within the space $\mathcal{X}$.*
- **Measure Space:** *A triplet ($\mathcal{X}$,$\mathcal{A}$,μ) where ($\mathcal{X}$,$\mathcal{A}$) is a measurable space and μ: $\mathcal{A} \to \mathcal{R}$ is a measure.*
- **Measurable functions:** *A function f: $\mathcal{X}_1 \to \mathcal{X}_2$ is called measurable iff $f^{-1}(A) \in \mathcal{A}_1$ for all $A \in \mathcal{A}_2$.*
- **Borel Sigma Algebra:** *The* Borel *sigma algebra is the smallest sigma algebra generated by the open sets of X.*

3.2 Lebesgue Integration

Lebesgue integration [2] is a fundamental concept in many mathematical theories, such as real analysis [6], and probability [9]. The reasons for its extensive usage, compared to the commonly known *Riemann* integral, include the ability to handle a numerous classes of functions. Similar to the way in which step functions are used in the development of the *Riemann* integral, the Lebesgue integral makes use of a special class of functions called positive simple functions. It has been defined in [13] as

$$\forall x \in \mathcal{X}, g(x) = \sum_{i \in s} \alpha_i I_{\alpha_i}(x)$$

Some commonly used properties of the Lebesgue integral have also been verified in [13]

- **Lebesgue integral of positive simple functions:**

$$\int g d\mu = \Sigma \alpha_i \mu(\mathrm{a}_i)$$

- **Lebesgue integral of non-negative functions:**

$$\int_X f d\mu = sup\{\int_X g d\mu | g \leq f \ \}$$

- **Lebesgue integral of arbitrary functions:**

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu$$

$f^+(x) = max\ (f(x),\ 0)\ and\ f^-(\mathrm{x}) = \max\ (-f(x),0)$.

- **Integrable functions:** *A function f is integrable iff* $\int_X \mid f \mid d\mu < \infty$ *or equivalently iff* $\int_X f^+ < \infty$ *and* $\int_X f^- < \infty$

- **Positivity, Linearity and Monotonicity:** Let f and g two functions, we have

  1. $\forall x. 0 \leq f(x) \Rightarrow 0 \leq \int_X f d\mu$
  2. $\forall x. f(x) \leq g(x) \Rightarrow \int_X f d\mu \leq \int_X d d\mu$
  3. $\int_X c. f d\mu = c. \int_X f d\mu$
  4. $\int_X f + g d\mu = \int_X f d\mu + \int_X g d\mu$
  5. A and B disjoint sets $\Rightarrow \int_{A \bigcup B} f d\mu = \int_A f d\mu + \int_B f d\mu$

3.3 Probability Theory

Probability provides mathematical models for random phenomena and experiments. The purpose is to describe and predict relative frequencies (averages) of these experiments in terms of probabilities of events.

- **Probability Space:** *a measure space such that the measure of the state space is* 1

- **Independent events:** *Two events A and B are independent iff* $p(A \cap B) = p(A)p(B)$.

- **Random variable:** $X : \Omega \to \mathcal{R}$ *is a random variable iff X is* $(F, \mathcal{B}(\mathcal{R}))$ *measurable where F denotes the set of events.*

- **Expected value:** The properties verified for the expectation of a random variable are

  1. $E[X + Y] = E[X] + E[Y]$
  2. $E[aX] = aE[X]$
  3. $E[a] = a$
  4. $X \leq Y then E[X] \leq E[Y]$
  5. *If* X *and* Y *are independent then* $E[XY] = E[X]E[Y]$

- **Variance and Covariance:** Variance and covariance exhibit the following formally verified properties

  1. $Var(X) = E[X^2] - E[X]^2$
  2. $Cov(X, Y) = E[XY] - E[X]E[Y]$
  3. $Var(X) \geq 0$
  4. $\forall a \in R, Var(aX) = a^2 Var(X)$
  5. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

4 Extended Real Numbers

The extended real number theory allows us to describe various limiting measures in the theories of measure and integration. In this section, we give a summary of the formalization of the extended real numbers in HOL, we began by defining a new data type for extended reals, we also describe the formalization of some of the arithmetic operations as well as prove their important properties.

4.1 Type Definition

The new data type is called `extreal` which refers to the extended real. The extended real system is obtained from the real number system $\mathbb{R}$ by adding two elements $-\infty$ and $+\infty$. These new elements are not a real numbers. Thus, an extended real number can be either a standard real number, positive infinity or negative infinity.

Definition 1: *Extended Reals data type*

```
⊢ Hol_datatype extreal = Normal of real | PosInf | NegInf
```

We will describe a number of properties and operations over the extended real type in the rest of this section.

4.2 Arithmetic Operations

The basic arithmetic operations are addition, subtraction, multiplication and division.
Addition: (+)
Addition is the most commonly used arithmetical operation. It takes two parameters of extended real numbers and returns an extended real number as the result.

Definition 2: *Addition of two extended real numbers*

```
⊢ ∀ x y.  (Normal x + Normal y = Normal (x + y)) ∧
  (Normal v0 + NegInf = NegInf) ∧
  (Normal v0 + PosInf = PosInf) ∧
  (NegInf + Normal v1 = NegInf) ∧
  (PosInf + Normal v1 = PosInf) ∧
  (NegInf + NegInf = NegInf) ∧
  (PosInf + PosInf = PosInf)
```

Using Definition 2 above we prove properties such as
$(a : extreal) + 0 = a, a + b + c = (a + b) + c = a + (b + c) and a + b = b + a$, namely *the additive identity, the associativity and the commutativity properties.*
Substraction: (-)
The substraction operation is the opposite of the addition operation. It gives the difference between two numbers. It is expressed as the addition of negative numbers (a - b = a + (-b)).

Definition 3: *Substraction of extended real numbers*

```
⊢ ∀ x y.  (Normal x - Normal y = Normal (x - y)) ∧
  (NegInf - Normal v0 = NegInf) ∧
  (PosInf - Normal v0 = PosInf) ∧
  (Normal v1 - NegInf = PosInf) ∧
  (Normal v2 - PosInf = NegInf) ∧
  (NegInf - PosInf = NegInf) ∧
  (PosInf - NegInf = PosInf)
```

A number of properties are proved using the Definition 3
$a - 0 = a, a - b = a + (-b)$.
Many other theorems and lemmas involving both addition and substraction are proved, for example $a + (b - a) = b$

Theorem 1:

```
⊢ ∀ x y.  x ≠ NegInf ∧ x ≠ PosInf ⇒
  (x + (y - x) = y)
```

Multiplication: (* or .)
Multiplication is also a basic operation commonly used in arithmetic. Multiplication also combines two numbers into a single number, *the product*. It combines two or more factors and gives an extended real as the result.

Definition 4: *Multiplication of extended real numbers*

```
⊢ ∀ x y.  (NegInf * NegInf = PosInf) ∧
  (NegInf * PosInf = NegInf) ∧
  (PosInf * NegInf = NegInf) ∧
  (PosInf * PosInf = PosInf) ∧
  (Normal x * NegInf =
    if x = 0 then Normal 0 else if 0 < x then NegInf else PosInf) ∧
  (NegInf * Normal y =
    if y = 0 then Normal 0 else if 0 < y then NegInf else PosInf) ∧
  (Normal x * PosInf =
    if x = 0 then Normal 0 else if 0 < x then PosInf else NegInf) ∧
  (PosInf * Normal y =
    if y = 0 then Normal 0 else if 0 < y then PosInf else NegInf) ∧
  (Normal x * Normal y = Normal (x * y))
```

The multiplication operation is both commutative and associative. Further, it is distributive over addition and subtraction. The multiplicative identity is 1, that is, multiplying any number by 1 yields that same number. Also, the multiplicative inverse is the reciprocal of any number (except zero that is the only number without a multiplicative inverse), that is, multiplying the reciprocal of any number by the number itself yields the multiplicative

identity.

Theorem 2: *Multiplicative identity*

```
⊢ ∀ a.  (a ≠ 0) ⇒
  (a * INV a = 1)
```

The function INV in HOL refers to the inverse of an extended real number.

Theorem 3: *Multiplicative distribution over addition*

```
⊢ ∀ x y z.  0 ≤ y ∧ 0 ≤ z ∨ y ≤ 0 ∧
  z ≤ 0 ∨ y ≠ NegInf ∧ z ≠ NegInf ∧ y ≤ 0 ∧
  z ≤ 0 ∨ y ≠ PosInf ∧ z ≠ PosInf ∧
  0 ≤ y ∧ 0 ≤ z ⇒
  (x * (y + z) = x * y + x * z)
```

Division: (/)
Division is dual operation of multiplication as substitution is to addition. Division finds the quotient of two numbers when the dividend is divided by the divisor. Any dividend divided by zero is undefined. For positive numbers, if the dividend is larger than the divisor, the quotient is greater than one, otherwise it is less than one (a similar rule applies for negative numbers). The quotient multiplied by the divisor always yields the dividend.The HOL formalization of the inversion(INV)is given by $a/b = a * (1/b)$.

Definition 5: *Division of extended real numbers*

```
⊢ ∀ x y.  x / y = x * inv y
```

The basic arithmetic operations described above can be combined with the comparison operators to formalize operators such as MIN, MAX, SUP, INF, etcetera. These operators are described in the next two subsections.

4.3 Comparison Operators

The comparison of two items is one of the most commonly used operation in real and probabilistic analysis.
For this reason many operators were formalized involving extended real numbers such as $>$ (greater than), $<$ (less than), $\geq$ (greater than or equal to), $\leq$ (less than or equal to) etcetera These operators test whether two extended real numbers involved are the same or not and the result of such a test is a boolean data-type that is either $TRUE$ or $FALSE$.

Definition 6: *Less than or equal operation*

```
⊢ ∀ x y.  (Normal x ≤ Normal y ⇔ x ≤ y) ∧   (NegInf ≤ NegInf ⇔ T) ∧
  (NegInf ≤ PosInf ⇔ T) ∧
  (NegInf ≤ Normal v5 ⇔ T) ∧
```

```
(PosInf ≤ PosInf ⇔ T) ∧
(Normal v2 ≤ PosInf ⇔ T) ∧
(PosInf ≤ NegInf ⇔ F) ∧
(Normal v3 ≤ NegInf ⇔ F) ∧
(PosInf ≤ Normal v7 ⇔ F)
```

In HOL, we formalized the less than operator based on the less or equal operator saying that,

Definition 7: *Less than operation*

```
⊢ ∀ x y.  x < y ⇔∼(y ≤ x)
```

We proved several important properties involving the comparison and the arithmetic operators, we show two such examples in the following.

Theorem 4:

```
⊢ ∀ x y z.  x ≠ NegInf ∧ x ≠ PosInf ⇒
  (y - x ≤ z ⇔ y ≤ z + x)
```

Theorem 5:

```
⊢ ∀ x y z w.  w ≤ x ∧ y ≤ z ⇒
  (w + y ≤ x + z)
```

4.4 MIN and MAX

The minimum and the maximum operators are also very widely used in real and probabilistic analysis. We formalized the minima and the maxima of two extended real numbers. For example Definition 8 gives the definition of the extended real minimum operator.

Definition 8: *Minimum operator*

```
⊢ ∀ x y.  extreal_min (x:extreal) (y:extreal) = if x ≤ y then x else y
```

The same syntax was used to define the maximum.
In this context, a variety of theorems and properties were proved. We mention for example, the minimum is less or equal than both arguments of the function min.

Theorem 6:

```
⊢ ∀ x y.  min x y ≤ x ∧ min a b ≤ y
```

In Theorem 7, we show that the maximum of a set of extended real numbers is less than or equal to an element then all the other elements of the set are also less than or equal to this element,

Theorem 7:

```
⊢ ∀ S c.  max S ≤ c ⇔ ∀ x.  x IN S ⇒ x ≤ c
```

4.5 SUP and INF

The SUPremum and the INFimum in our case are two functions defined over sets. For example given a subset S of a totally or partially ordered set T, the supremum (sup) of S, if it exists, is the least element of T which is greater than or equal to any element of S. Consequently, the supremum is also referred to as the least upper bound (lub or LUB). If the supremum exists, it is unique. If S contains a greatest element, then that element is the supremum; otherwise, the supremum does not belong to S (or it does not exist). For instance, the negative real numbers do not have a greatest element, and their supremum is 0 (which is not a negative real number).
The supremum is in a precise sense dual to the concept of an infimum. It was defined in HOL as showed below

Definition 9: *Supremum operator*

```
⊢ ∀ p.  extreal_sup p =
  if (∀a.  (∀b.  p b ⇒ b ≤ a) ⇒ (a = PosInf)) then PosInf
  else if
     (∀b.  p b ⇒ (b = NegInf)) then NegInf
  else (Normal(sup (\t.  p (Normal t))))
```

We formalized an important theorem in HOL using our formalization of extended real numbers that states that when the sup of a set of extended real numbers is less than or equal to an element then so are all the elements of this set

Theorem 8:

```
⊢ ∀ p c.  sup p ≤ c ⇔ (∀ x.  x IN p ⇒ x ≤ c)
```

where *pa* means a IN p which is a set.
We also proved an important property related to the addition of two sups of sets,

Theorem 9: *Addition of two sups of sets*

```
⊢ ∀ (f:num → extreal) (g:num → extreal).  (∀n.  (Normal 0) ≤ f n) ∧
  (∀n.  f n ≤ f (SUC n)) ∧ (∀n.  (Normal 0) ≤ g n) ∧ (∀n.  g n ≤ g (SUC
n)) ⇒
  (sup (IMAGE (\n.  f n + g n) UNIV) = sup (IMAGE f UNIV) + sup (IMAGE g
UNIV))
```

5 Topology Concepts: The Borel Theory

The Borel algebra as defined is the smallest sigma algebra generated by open sets of a space Ω. It allows us to verify a various number of properties related to the Measure Theory or the

Probability Theory, such us random variables or measurability. But in order to formalize the Borel theory we have to formalize some other concepts.

5.1 Neighborhood

In topology and related areas of mathematics, a neighbourhood (or neighborhood) is one of the basic concepts in a topological space. Intuitively speaking, a neighbourhood of a point is a set containing the point where you can move that point some amount without leaving the set. This concept is closely related to the concepts of open set and interior.
If X is a topological space and p is a point in X, a neighbourhood of p is a set V, which includes an open set U containing p, $p \in U \subseteq V$.

Definition 10: *Neighborhood*

```
⊢ ∀ A a.  extReal_NEIGHB (Normal a) A = ∃.  (Normal 0 < b) ∧
  (!y.  (Normal a) - b < y ∧ y < (Normal a) + b ==> y IN A)) ∧
  (extReal_NEIGHB posInf A = ∃b.  (∀y.  b < y ⇒ y IN A)) ∧
  (extReal_NEIGHB negInf A = ∃b.  (∀y.  y < b ⇒ y IN A)
```

5.2 Open Sets

The concept of an open set is fundamental to many areas of mathematics, a set U is open if any point x in U can be moved a small amount in any direction and still be in the set U. The notion of an open set provides a fundamental way to speak of nearness of points in a topological space, without explicitly having a concept of distance defined. Concepts that use notions of nearness, such as the continuity of functions, can be translated into the language of open sets.

Definition 11: *Open Sets*

```
⊢ ∀ A. extReal_OPEN_SET A = ∀a.  a IN A ⇒ extReal_NEIGHB a A
```

We proved some properties related to the open sets such as, the empty set is open,

Theorem 10: *Empty set is open*

```
⊢ ∀ s.  (s = {}) ⇒ extReal_OPEN_SET s
```

open intervals are open,

Theorem 11: *Open intervals are open*

```
⊢ ∀ a b:extReal.  (a≠negInf ∧ b≠posInf) ⇒
  (extReal_OPEN_SET (open_interv a b))
```

where *open_intervab* means]a,b[.

5.3 Rational Numbers

In mathematics, a rational number is any number that can be expressed as the quotient or fraction a/b of two integers, with the denominator b not equal to zero. Since b may be equal to 1, every integer is a rational number. The set of all rational numbers is usually denoted by a boldface Q.
The rationals are a dense subset of the real numbers: every real number has rational numbers arbitrarily close to it. A related property is that rational numbers are the only numbers with finite expansions as regular continued fractions. By virtue of their order, the rationals carry an order topology. The rational numbers, as a subspace of the real numbers, also carry a subspace topology. The rational numbers form a metric space by using the absolute difference metric $d(x, y) = |x - y|$, and this yields a third topology on Q. All three topologies coincide and turn the rationals into a topological field. The rational numbers are an important example of a space which is not locally compact. The rationals are characterized topologically as the unique countable metrizable space without isolated points.

Definition 12: *Open Sets*

```
⊢ Q_set_def = {q | ∃(m:num) n.  (q = m / n)} UNION {q | ∃(m:num) n.  (q =
-(m / n))}
```

6 Applications

In this chapter we are going to illustrate the previous theoretical work with an application, The Heavy Hitter problem. Before dealing with that application, we should first formalize the Chebychev's inequality which in turn needs the formalization of the Markov's inequality.

6.1 Markov's inequality

In probability theory, Markov's inequality gives an upper bound for the probability that a non-negative function of a random variable is greater than or equal to some positive constant. It is named after the Russian mathematician Andrey Markov, although it appeared earlier in the work of Pafnuty Chebyshev (Markov's teacher), and many sources, especially in analysis, refer to it as Chebyshev's inequality or Bienaym's inequality.

Markov's inequality (and other similar inequalities) relate probabilities to expectations, and provide (frequently) loose but still useful bounds for the cumulative distribution function of a random variable.
An example of an application of Markov's inequality is the fact that (assuming incomes are non-negative) no more than $\frac{1}{5}$ of the population can have more than 5 times the average income.

If X is any random variable and $a > 0$, then

$$Pr(|X| \geq a) \leq \frac{E(|X|)}{a} \tag{1}$$

In HOL it is formalized as followed:

Theorem 12: *Markov inequality (probability statement)*

```
⊢ ∀p.  prob p {x | x IN p_space p ∧ a ≤ abs (X x)} ≤
  (1/a)*expectation (λx.  abs (X x))
```

Where X is a random variable, p a probability space and *prob* is a probability measure.

In the language of measure theory, MARKOV's inequality states that if (X, Ω, μ) is a measure space, f is a measurable extended real-valued function, and $t > 0$, then

$$\mu(\{x \in X : f(x) \geq t\}) \leq \frac{1}{t} \int_X |f| d\mu \tag{2}$$

- **Proofs:**

We separate the case in which the measure space is a probability space from the more general case because the probability case is more accessible for the general reader.

- **In the language of probability theory:**

For any event E, let I_E be the indicator random variable of E, that is, $I_E = 1$ if E occurs and $I_E = 0$ otherwise. Thus $I_{(|X| \geq a)} = 1$ if the event $|X| \geq a$ occurs, and $I_{(|X| \geq a)} = 0$ if $|X| < a$. Then, given $a > 0$,

$$aI_{(|X| \geq a)} \leq |X| \tag{3}$$

which is clear if we consider the two possible values of $I_{(|X| \geq a)}$. Either $|X| < a$ and thus $I_{(|X| \geq a)} = 0$, or $I_{(|X| \geq a)} = 1$ and by the condition of $I_{(|X| \geq a)}$, the inequality must be true. Therefore,

$$E(aI_{(|X| \geq a)}) \leq E(|X|) \tag{4}$$

Now, using linearity of expectations, the left side of this inequality is the same as

$$aE(I_{(|X| \geq a)}) = aPr(|X| \geq a) \tag{5}$$

Thus we have

$$aPr(|X| \geq a) \leq E(|X|) \tag{6}$$

and since $a > 0$, we can divide both sides by a.

- **In the language of measure theory:**

For any measurable set A, let 1_A be its indicator function, that is, $1_{A(x)} = 1$ if $x \in A$, and 0 otherwise. If A_t is defined as $A_t = x \in X || f(x)| \geq t$, then

$$0 \leq t1_{A_t} \leq |f|1_{A_t} \leq |f| \tag{7}$$

Therefore, by monotonicity of the Lebesgue integral

$$\int_X t1_{A_t} d\mu \leq \int_{A_t} |f| 1_{A_t} d\mu \leq \int_X |f| d\mu \tag{8}$$

Now, note that the left side of this inequality is the same as

$$t \int_X 1_{A_t} d\mu = t\mu(A_t) \tag{9}$$

Thus we have

$$t\mu(\{x \in X || f(x)| \geq t\}) \leq \int_X |f| d\mu \tag{10}$$

and since $t > 0$, both sides can be divided by t, obtaining

$$\mu(\{x \in X || f(x)| \geq t\}) \leq \frac{1}{t} \int_X |f| d\mu \tag{11}$$

- **In HOL:**

In HOL, we first prove that inequality in the measure statement using a number of proved theorems in the measure theory. The proof steps are exactly the same as those shown previously. Some used theorems such as `GSPEC_AND, GSPECIFICATION, indicator_fn_def, GSPEC_ID,`
`IN_MEASURABLE_BOREL_ABS`,... to reduce the main goal into other specific subgoals that in turn could prove the initial goal. All those theorems were used within some tactics such as `RW_TAC, FULL_SIMP_TAC, MP_TAC, METIS_TAC`,...

Concerning the probability statement, the initial goal was a specific case of the measure statement, so that it was proved using `MP_TAC markov_inequality_mea-`

`sure`and the specification using the `Q.GSPECL`

6.2 Chebyshev's inequality

In probability theory, Chebyshev's inequality guarantees that in any data sample or probability distribution,nearly all values are close to the mean - the precise statement being that no more than $1/k^2$ of the distribution's values can be more than k standard deviations away from the mean. The inequality has great utility because it can be applied to completely arbitrary distributions (unknown except for mean and variance), for example it can be used to prove the weak law of large numbers.

The theorem is named after Russian mathematician Pafnuty Chebyshev, although it was first formulated by his friend and colleague Irne-Jules Bienaym [**?**]. It can be stated quite generally using measure theory, the statement in the language of probability theory then follows as a particular case, for a space of measure 1. The term Chebyshev's inequality may also refer to the Markov's inequality, especially in the context of analysis.

Let (X, Ω, μ) be a measure space, and let f be an extended real-valued measurable function defined on X. Then for any real number $t > 0$, In the language of measure theory we have,

$$\mu(\{x \in X : |f(x)| \geq t\}) \leq \frac{1}{t^2} \int_X |f|^2 d\mu \tag{12}$$

More generally, if g is an extended real-valued measurable function, nonnegative and nondecreasing on the range of f, then

$$\mu(\{x \in X : |f(x)| \geq t\}) \leq \frac{1}{g(t)} \int_X f \circ g d\mu \tag{13}$$

With $g(t) = t^2$ if $t \geq 0$ and 0 otherwise and taking $|f|$ instead of f.

In HOL, the measure statement of the CHEBYSHEV's inequality is formalized this way,

Theorem 13: *Chebychev inequality (measure statement)*

```
⊢ ∀m.  measure m ({x|x IN m_space m ∧ t ≤ abs (fx)}) ≤
  (1/(t pow 2))*pos_fn_integral m (λ x.  (abs (f x)) pow 2)
```

In the probability statement, let X be a random variable with expected value μ and finite variance σ^2. Then for any real number $k > 0$,

$$Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{14}$$

Only the case $k \geq 1$ provides useful information (when $k < 1$ the right-hand side is greater than one, so the inequality becomes vacuous, as the probability of any event cannot be greater than one).

The HOL version of that,

Theorem 14: *Chebychev inequality (probability statement)*

```
⊢ ∀p.  prob p ({x|x IN p_space p ∧ a≤ abs(X x - expectation p X)}) ≤
  (1/(a pow 2))*expectation p (λx.  (abs (X x - expectation p X)) pow 2)
```

- **Proofs:**

For that proof we will be using the MARKOV's inequality previously proved. The same thing will be for the HOL proof.

- **In the language of measure theory:**

Let A_t be defined as $A_t = \{x \in X | f(x) \geq t\}$, and let 1_{A_t} be the indicator function of the set A_t. Then, it is easy to check that

$$0 \leq g(t)1_{A_t} \leq g \circ f 1_{A_t} \leq g \circ f \tag{15}$$

And therefore,

$$g(t)\mu(A_t) = \int_X g(t)1_{A_t} d\mu \leq \int_{A_t} g \circ f d\mu \leq \int_X g \circ f d\mu \tag{16}$$

The desired inequality follows from dividing the above inequality by $g(t)$.

- **In the language of probability theory:**

MARKOV's inequality states that for any real-valued random variable Y and any positive number a, we have $Pr(|Y| > a) \leq E(|Y|)/a$. One way to prove CHEBYSHEV's inequality is to apply MARKOV's inequality to the random variable $Y = (X - \mu)^2$ with $a = (\sigma k)^2$. It can also be proved directly. For any event A, let I_A be the indicator random variable of A, i.e. I_A equals 1 if A occurs and 0 otherwise. Then

$$\begin{aligned} \Pr(|X - \mu| \geq k\sigma) &= E(I_{|X-\mu| \geq k\sigma}) = E(I_{[(X-\mu)/(k\sigma)]^2 \geq 1}) \\ &\leq E((\frac{X-\mu}{k\sigma})^2) = \frac{1}{k^2}\frac{E((X-\mu)^2)}{\sigma^2} = \frac{1}{k^2} \end{aligned} \tag{17}$$

- **In HOL:**(*see more in annexe*)

To prove the Chebyshev's inequality in HOL we need the Markov's inequality already proved. So, we use for that the `MP_TAC` tactic which refers to the modus ponens rule. That way, by a simple specification using the `Q.SPQCL` we extract our new variables from those in the Markov's inequality.

Once done with the proof in the measure environment, we use that theorem to prove the probability statement inequality and also using the `MP_TAC` and the `Q.SPECL`.
The link between the probability and measure statement is ensured by the fact that the probability theory is a special variant of the measure theory.

6.3 The Heavy Hitter Problem

In this section, we will use the formalized work presented in the previously in order to conduct a probabilistic analysis of the Heavy Hitter problem.

6.3.1 Problem Description

An element of the universe U is called α-heavy hitter in an insertion-only data stream of length n, if it appears at least α n times in the stream. In the heavy hitter problem we are interested to report the set of λ-heavy hitters in the stream, that means if an item occurs more than λ n times, the algorithm returns it and if it occurs less than (α - ϵ)n times, it doesn't return it. One motivavtion to study the heavy hitter problem is the search for hot items in streams, for example, heavily traded stocks in streams of financial transactions. Another example is the problem of detecting spreading viruses in network traffic.

Problem: (Relaxed Heavy Hitter Problem) As mentioned in [5], *The ϵ-relaxed α-heavy hitter problem is to find a set $H \subseteq U$ such that*

1. *U contains every α-heavy hitter;*

2. *U contains no item that appears less than $(\alpha - \epsilon)n$*

We will now consider the following simple approach to the heavy hitter problem. We maintain a random sample of *cs* elements from the stream chosen independently and uniformly at random with repetition. Then
ReportHeavyHitter(α, ϵ)
s: Set of *cs* random elements chosen independently and uniformly at random with repetition
1. Report all elements that occur more than $(\lambda - \epsilon/2)$cs times in S.

Algorithm 1: *The λ-Heavy Hitter Problem*

Input: the frequency λ, a data stream DS and a Universe U with length n
Output: the list L of elements from U occurring at least $\lambda * n$ times in DS

$L \leftarrow [\,]$
for $i = 1 \rightarrow n$ **do**
  if $freq(DS[i]) \geq \lambda$ **then**
    $L \leftarrow L$ ***INSERT*** $DS[i]$
  end if
end for

6.3.2 Formal Specification in HOL

The Heavy Hitter Problem can be formalized in HOL by modeling the sample set of elements and the data stream as lists. Then we model a function, `freq_def`, that returns the frequency of an element in a list which is defined below,

Definition 13: *Frequency of an element in a set*

```
⊢ ∀ e L. freq e L =
  Normal ((LENGTH (FILTER (λr.  r = e) L))) / Normal (LENGTH L)
```

where the HOL function `LENGTH` returns the length of a list, and FILTER returns from a list another filtered based on such function.
The above function will be needed later, to report the list of the λ-heavy hitter elements. In fact, we model another function, `HeavyHitter_lst_def`, which takes as parameters two lists and a real value, thus

Definition 14: *Heavy Hitter list*

```
⊢ ∀ L M λ.  HeavyHitter_lst L M α =
  FILTER(λr.  α ≤ (freq (EL r L)M))L
```

After modeling the problem we will now tackle the probabilistic analysis part.

6.4 Analysis Using Chebyshev's Inequality

We apply a probabilistic analysis to the Heavy Hitter Algorithm. Let us fix a stream $\sigma = (\sigma_1, ..., \sigma_n)$ and let us consider an arbitrary $x \in U$ and assume that x occurs $\lambda^* n$ times in σ for some $\lambda* \in [0, 1]$. We will first determine the probability that x is reported by the algorithm. For that reason let X_i denote the indicator random variable for the event that the i-th element of the sample is taken uniformly at random from the stream.
First of all, we model our random variable X which is a Bernoulli random variable, i.e, X could be 1 if the considered element occurs into the data stream and 0 otherwise.
For that reason we will instantiate a new random variable from the major definition in the probability theory. This new random variable needs a new probability space that have the $\{0, 1\}$ as a space and power set, POW $\{0, 1\}$, as the events space and the probability measure will be a new function that returns pr if the set in parameter implements the fact that $f(x)$, which refers to our random variable in this case, is equal to 1, and returns $1 - pr$ otherwise. In HOL, the probability measure is defined as below

Definition 15: *Heavy Hitter probability measure*

```
⊢ ∀ i X. (HH_rv (X i) pr) ⇒
  (prob (HH_prob_space pr (X i)) {x | X i x = 1} = pr)
```

where the new probability space $HH\_prob\_spaceprf$ is the following

Definition 16: *Heavy Hitter probability space*

```
⊢ HH_prob_space pr g = ({0;1},POW {0;1},mu g pr)
```

thus the new random variable will be modeled as

Definition 17: *Heavy Hitter random variable*

```
⊢ HH_rv X pb = random_variable X (HH_prob_space pb X) Borel
```

Since the element of the sample is taken uniformly at random, $Pr[X_i = 1] = \frac{\alpha^* n}{n} = \alpha^*$, which is represented in the HOL specification by pr. This theorem can be easily proved in HOL using only the definition of the new random variable.

Theorem 15:

```
⊢ ∀i X. (HH_rv (X i) pr) ⇒
  (prob (HH_prob_space pr (X i)) {x | X i x = 1} = pr)
```

From the previous equality, with $\alpha^* = pr$ it immediately follows that

$$E[X_i] = 0.Pr[X_i = 0] + 1.Pr[X_i = 1] = \alpha^* \tag{18}$$

The linearity of the expectation implies that $E[\sum_{i=1}^{cs} X_i] = \lambda^* cs$

Theorem 16: *Expectation of the Heavy Hitter random variable*

```
⊢ ∀s.  (FINITE s) ∧ (pr ≠ PosInf) ∧ (∀i.  (∀x.  0 ≤ X i x)) ∧
  (HH_rv (X i') pr) ∧ (∀i.  i IN s ⇒
```

```
    (X i IN measurable (m_space (HH_prob_space pr (X i')),
      measurable_sets (HH_prob_space pr (X i')))) Borel)) ⇒
  expectation (HH_prob_space pr (X i')) (αx.  SIGMA (αi.  X i x) s) =
    pr * (CARD s)
```

Now, let us consider the case that x is a heavy hitter, i.e. $\alpha^* \geq \alpha$. Choosing the cardinal of s $cs = \frac{4}{\delta.\epsilon^2}$. We would like to prove, using all the previously proved theorems as well as the Chebyshev's inequality and some real analysis, the property: the probability of reporting x is at least $(1 - \delta)$.

$$\begin{aligned}
\mathbf{Pr}[\textstyle\sum_i X_i > (\lambda - \epsilon/2)] &\geq \mathbf{Pr}[\textstyle\sum_i X_i > \mathbf{E}[\sum_i X_i] - \epsilon.cs/2] \\
&= \mathbf{Pr}[\mathbf{E}[\textstyle\sum_i X_i] - \sum_i X_i < \epsilon.cs/2] \\
&= 1 - \mathbf{Pr}[\mathbf{E}[\textstyle\sum_i X_i] - \sum_i X_i \geq \epsilon.cs/2] \\
&\geq 1 - \mathbf{Pr}[|\textstyle\sum_i X_i - \mathbf{E}[\sum_i X_i]| \geq \epsilon.cs/2] \\
&= 1 \text{ - } \delta
\end{aligned}$$

In HOL, when we were dealing with that property, we met a number of theorems that had to be proved before, as for example the equality of the 2 sets $\{$`x|`$(pr - (e/2)) * cs$ `<` $\sum_i X_i$ $\}$ and $\{$`x|E[`$\sum_i X_i$`] - `$(e/2) * cs)$ `<` $\sum_i X_i$ $\}$.

Theorem 17:

```
⊢ (∀ e s pr.  (0 < e ∧ e ≠ PosInf) ∧ (0 < pr ∧ pr < 1) ∧ FINITE s ∧
  (CARD s = cs) ∧ (cs≠PosInf ∧ cs≠NegInf) ∧ (∀i x.  0 ≤ X i x) ∧
  HH_rv (X i') pr ∧ (∀i.  i IN s ⇒ X i IN measurable (m_space (HH_prob_space
pr (X i')),
    measurable_sets (HH_prob_space pr (X i'))) Borel)) ⇒
  ({x|x IN p_space (HH_prob_space pr (X i')) ∧ (pr - (e/2))*cs <
    (∑ (λi.  X i x) s)} =
  {x|x IN p_space (HH_prob_space pr (X i')) ∧
    (expectation (HH_prob_space pr (X i')) (λx.  ∑ (λi.  X i x) s) -
    (e/2)*cs)<(∑ (λi.  X i x) s)})
```

The proof for this theorem includes the definition of the `HH_expectation` already proved, `GSPECIFICATION` and `EXTENSION` which transform the set on its properties.
In addition, other theorems was proved in the middle such as

$$Pr[|X| \geq a] \geq Pr[X \geq a] \tag{19}$$

In HOL, the property above, was transformed into a measure of set relatively to a probability space, thus its equivalent in HOL is

Theorem 18:

```
⊢ ∀ e s pr.  (HH_rv (X i') pr) ∧ (0 < e ∧ e≠PosInf) ∧
```

```
(cs≠PosInf ∧ cs≠NegInf)) ⇒
(prob (HH_prob_space pr (X i'))
  {x|x IN p_space (HH_prob_space pr (X i')) ∧ (e/2)*cs ≤
    (expectation (HH_prob_space pr (X i')) (λx.  ∑ (λi.  X i x) s) -
    ∑ (λi.  X i x) s)} ≤
( prob (HH_prob_space pr (X i'))
  {x|x IN p_space (HH_prob_space pr (X i')) ∧ (e/2)*cs ≤
    abs((expectation (HH_prob_space pr (X i')) (λx.  ∑ (λi.  X i x) s) -
    (∑ (λi.  X i x) s)))})
```

The proof of the last theorem was dealt using the property of *if P is a SUBSET of S then prob p S ≥ prob p P*.
We applied then the CHEBYSHEV's inequality to what we got, and then with some real analysis we get our main expected result.

7 Conclusion

In this report, we presented an overview of the formalization of the extended real theory, which was need in the formalization of the probability theory, in the HOL theorem prover. To formalize this theory, we defined a new data-type, and proved numerous properties over it, such as arithmetic operations, comparison operations, and a few other fundamental topology concepts.

And to show the usefulness of the formalization in improving the existing formalization of the probability theory in higher-order logic, we formalized and proved the Markov and Chebychev inequalities. We then used these properties in The Heavy Hitter Problem and verified the performance results in the cove of HOL theorem prover. The HOL signature of the formalized properties and proofs are available on **??**.

Our formalization contains around 500 lines of code for the application part and around 2850 for the formalization of the extended real theory and its related properties. This work required deep mathematical knowledge of probability theory, real analysis, set properties and topology concepts. The formalization in HOL was also challenging.

As a future work we are planning to tackle the information theory section by doing an Information Theoretical Analysis of Data Compression using Theorem Proving. Also, we are planning to apply our developed method on a multimedia case study.

References

[1] Coble A. R. *Anonymity, Information, and Machine-Assisted Proof.* PhD thesis, 2010.

[2] K. Berberian. *Fundamentals of Real Analysis.* 1988.

[3] Jzef Biaas. The σ-additive measure theory. *Journal of Formalized Mathematics*, 2, 1990.

[4] Edmund M. Clarke, Orna Grumberg, and David E. Long. Verification tools for finite-state concurrent systems. In *A Decade of Concurrency, Reflections and Perspectives, REX School/Symposium*, pages 124–175. Springer-Verlag, 1994.

[5] C.Sohler. The heavy hitter problem. In *Streaming Algorithms*, pages 25–28.

[6] R.R. Goldberg. *Methods of Real Analysis*. Oxford & IBH Pub., 1970.

[7] M.J.C. Gordon. *Mechanizing programming logics in higher order logic*. University of Cambridge, Computer Laboratory, 1988.

[8] Aarti Gupta. Formal hardware verification methods: A survey. *Formal Methods in System Design*, 1(2/3):151–238, 1992.

[9] P.R. Halmos. *Naive Set Theory*. Van Nostrand, 1960.

[10] Osman Hasan. *Formal probabilistic analysis using theorem proving*. PhD thesis, Montreal, P.Q., Canada, Canada, 2008.

[11] Joe Hurd. A formal approach to probabilistic termination. pages 230–245.

[12] David R Lester. Topology in pvs: continuous mathematics with applications. In *Proceedings of the second workshop on Automated formal methods*, AFM '07, pages 11–20. ACM, 2007.

[13] T. Mhamdi, O. Hasan, and S. Tahar. Formalization of entropy measures in hol. In *Interactive Theorem Proving*, volume 6898 of *LNCS*, pages 233–248, 2011.

[14] Tarek Mhamdi, Osman Hasan, and Sofiène Tahar. On the formalization of the lebesgue integration theory in hol. In *ITP*, pages 387–402, 2010.

[15] Andrzej Nędzusiak. σ-Fields and Probability. *Journal of Formalized Mathematics*, 1989.

[16] S. Richter. Formalizing integration theory with an application to probabilistic algorithms. In *Theorem Proving in Higher Order Logics, 17th International Conference, TPHOLs 2004, Park City, Utah, USA, September 14-17, 2004, Proceedings*, volume 3223 of *Lecture Notes in Computer Science*, pages 271–286. Springer, 2004.