



HAL
open science

Stochastic Formal Methods for Hybrid Systems

Marc Daumas, David Lester, Erik Martin-Dorel, Annick Truffert

► **To cite this version:**

Marc Daumas, David Lester, Erik Martin-Dorel, Annick Truffert. Stochastic Formal Methods for Hybrid Systems. 2009. hal-00107495v5

HAL Id: hal-00107495

<https://hal.science/hal-00107495v5>

Preprint submitted on 24 Feb 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Stochastic Formal Methods for Hybrid Systems

Marc Daumas¹, David Lester², Erik Martin-Dorel^{1,3}, and Annick Truffert³

¹ELIAUS (EA 3679 UPVD) and ³LAMPS (EA 4217 UPVD)
Perpignan, France 66860, {marc.daumas, erik.martin-dorel, truffert}@univ-perp.fr

²School of Computer Science, University of Manchester
Manchester, United Kingdom M13 9PL, david.r.lester@manchester.ac.uk

Abstract

We provide a framework to bound the probability that accumulated errors were never above a given threshold on hybrid systems. Such systems are used for example to model an aircraft or a nuclear power plant on one side and its software on the other side. This report contains simple formulas based on Lévy's and Markov's inequalities and it presents a formal theory of random variables with a special focus on producing concrete results. We selected four very common applications that fit in our framework and cover the common practices of hybrid systems that evolve for a long time. We compute the number of bits that remain continuously significant in the first two applications with a probability of failure around one against a billion, where worst case analysis considers that no significant bit remains. We are using PVS as such formal tools force explicit statement of all hypotheses and prevent incorrect uses of theorems.

1 Introduction

Formal proof assistants are used in areas where errors can cause loss of life or significant financial damage as well as in areas where common misunderstandings can falsify key assumptions. For this reason, formal proof assistants have been much used for floating point arithmetic [1, 2, 3, 4, 5, 6] and probabilistic or randomized algorithms [7, 8]. Previous references link to a few projects using proof assistants such as ACL2 [9], HOL [10], Coq [11] and PVS [12].

All the above projects that deal with floating point arithmetic aim at containing worst case behavior. Recent work has shown that worst case analysis may be meaningless for systems that evolve for a long time as encountered in the industry. A good example is a process that adds numbers in ± 2 with a measure error of $\pm 2^{-24}$. If this process adds 2^{25} items, then the accumulated error is ± 2 , and note that 10 hours of flight time at operating frequency of 1 kHz is approximately 2^{25} operations. Yet we easily agree that provided the individual errors are not correlated, the actual accumulated errors will continuously be much smaller than ± 2 .

We present in Section 2 a few examples where this work can be applied. We focus on applications for ***n* counting in billions** and a probability of failure about **one against a billion**. Should one of these constraints be removed or lessened, the problems become much simpler. The main contribution of this work is the selection of a few theorems **amenable to formal methods** in a reasonable time, their application to **software and systems reliability**, and our work with PVS. Section 3 presents the formal background on probability with Markov's and Lévy's inequality and how to use this theory to assert software and system reliability.

Doob-Kolmogorov's inequality was used in previous work [13]. It is an application of Doob's inequality that can be proved with elementary manipulations for second order moment. It is better than Lévy's inequality in the sense that it can be applied to any sum of independent and centered variables. Yet it is limited by the fact that it bounds only second order moments.

2 Applications

Lévy's inequality works with independent symmetric random variables as we safely assumed in Sections 2.1 and 2.2. Doob's inequality combined with Jensen's one will overcome this restriction in future

Listing 1: Accumulation or dot product

```

1  $a_0 = 0;$ 
2 for ( $i = 1; i \leq n; i = i + 1$ )
3    $a_i = a_{i-1} + d_i;$ 

```

formal developments for the applications presented in Section 2.3 that cannot be treated by Lévy’s inequality. Alas, we foresee that the effort to make Doob’s inequality available in any of the formal tools available today is at least a couple of years. Automatic treatment of all the following applications may use interval arithmetic that has been presented in previous publications and is now available in formal tools [4, 5, 6].

2.1 Long accumulations and dot products

A floating point number represents $v = m \times 2^e$ where e is the exponent, an integer, and m is the mantissa [14]. IEEE 754 standard [15] on floating point arithmetic uses sign-magnitude notation for the mantissa and the first bit b_0 of the mantissa is implicit in most cases ($b_0 = 1$) leading to the first definition in equation (1). Some circuits such as the TMS320 [16] use two’s complement notation for m leading to the second definition in equation (1). The sign s and all the b_i are either 0 or 1 (bits).

$$v = (-1)^s \times b_0.b_1 \cdots b_{p-1} \times 2^e \quad \text{or} \quad v = (b_0.b_1 \cdots b_{p-1} - 2 \times s) \times 2^e \quad (1)$$

In fixed point notation e is a constant provided by the data type and b_0 cannot be forced to 1. We define for any representable number v , the unit in the last place function below, with the notations of equation (1).

$$\text{ulp}(v) = 2^{e-p+1}$$

The example given in Listing 1 sums n values. When the accumulation is performed with floating point arithmetic each iteration introduces a new round-off error X_i . One might assume that X_i follows a continuous or discrete uniform distribution on the range $\pm u$ with $u = \text{ulp}(a_i)/2$ as trailing digits of numbers randomly chosen from a logarithmic distribution [17, pp. 254–264] are approximately uniformly distributed [18]. A significantly different distribution may mean that the round-off error contains more than trailing digits.

Errors created by operators are discrete and they are not necessarily distributed uniformly [19]. The distribution is very specific but as soon as we verify that it is symmetric we only have to bound the moments involved in our main result as in equation (2).

$$\mathbb{E}(X_i) = 0, \quad \mathbb{E}(X_i^2) \leq \frac{u^2}{3}, \quad \mathbb{E}(X_i^4) \leq \frac{u^4}{5}, \quad \mathbb{E}(X_i^6) \leq \frac{u^6}{7}, \quad \text{and} \quad \mathbb{E}(X_i^8) \leq \frac{u^8}{9} \quad (2)$$

If a_i uses a directed rounding mode, we introduce $X'_i = X_i - \mathbb{E}(X_i)$ and we use equation (2) again. We may also assume that d_i also carries a single error $X_{2,i}$ or a linear combination of $m - 1$ round-off errors $X_{2,i}, \dots, X_{m,i}$ such that all of them satisfy equation (2) for a given u .

If d_i is a data obtained by an accurate sensor, we may assume that the difference between d_i and the actual value \bar{d}_i follows a normal distribution very close to a uniform distribution on the range $\pm u$ with some new value of u . In this cases we model the error $d_i - \bar{d}_i$ by a symmetric random variable $X_{2,i}$ and we use equation (2).

Table 1: Number of significant bits with a probability of failure $\mathbb{P}\left(\max_{1 \leq i \leq n} (|S_i|) \geq \varepsilon\right)$ bounded by P

u	n	m	P	$2k$	$\varepsilon \approx$ or $\varepsilon \sim$	$\log_2 \varepsilon \approx$ or $\log_2 \varepsilon \sim$
2^{-24}	10^9	2	10^{-9}	2	68.825	+6.10
				4	0.42832	-1.22
				6	0.085786	-3.54
				8	0.040042	-4.64
				44	0.010153	-6.62
2^{-24}	10^9	10	10^{-10}	2	486.66	+8.92
				4	1.7031	+0.768
				6	0.28156	-1.82
				8	0.11939	-3.06
				48	0.023873	-5.38
u	$u^{-3/2}$	1	$u^{3/2}$	2	$\sqrt[4]{4u^{-2}/9}$	$(-\log_2 u + 1 - \log_2 3)/2$
				4	$\sqrt[8]{4u^{-1}/9}$	$(-\log_2 u + 2 - 2\log_2 3)/8$
				6	$\sqrt[12]{100/81}$	$(\log_2 10 - 2\log_2 3)/6$
				8	$\sqrt[16]{4900u/729}$	$(\log_2 u + 2\log_2 70 - 6\log_2 3)/16$

After n iterations and assuming that all the errors introduced, $X_i, X_{2,i}, \dots, X_{m,i}$ are symmetric and independent, we want the probability that the accumulated errors have exceeded some user specified bound ε :

$$\mathbb{P}\left(\max_{1 \leq i \leq n} (|S_i|) \geq \varepsilon\right) \leq P \quad \text{with} \quad S_n = \sum_{i=1}^n \left(X_i + \begin{cases} -\mathbb{E}(X_i) & \text{if centering is needed} \\ \sum_{j=2}^m X_{j,i} & \text{if more variables are needed} \\ \dots & \dots \end{cases} \right). \quad (3)$$

Previous work used Doob-Kolmogorov's inequality. We will see in Section 3 that we can exhibit tighter bounds using Lévy's inequality followed by Markov's one. Table 1 present the number of significant bits of the results $\log_2 \varepsilon$ for some values of u, n, m , and P in equation (3). These values are obtained by using one single value of u , as large as needed. Tighter results can be obtained by using a specific value of u for each random variable and each iteration.

2.2 Recursive filters operating for a long time

Recursive filters are commonly used in digital signal processing and appear for example in the programs executed by Flight Control Primary Computers (FCPC) of aircraft. Finite impulse response (FIR) filters usually involve a few operations and can be treated by worst case error analysis. However infinite impulse response (IIR) filters may slowly drift.

Theory of signal processing provides that it is sufficient to study second order IIR with coefficient b_1 and b_2 such that polynomial $X^2 - b_1X - b_2$ has no zero in \mathbb{R} . Listing 2 presents the pseudo-code of one such filter. A real implementation would involve temporary registers.

When implemented with fixed or floating point operations, each iteration introduces a single error X_i or a compound one $X_{1,i} + \dots + X_{m,i}$ in y_i . As these filters are linear, we study the response to $d_0 = 1$ and

Listing 2: Infinite impulse response (IIR) filter

```

1  $y_{-1} = 0; y_0 = d_0;$ 
2 for ( $i = 1; i \leq n; i = i + 1$ )
3    $y_i = d_i - b_1 y_{i-1} - b_2 y_{i-2};$ 

```

Listing 3: Sum of squares

```

1  $a_0 = 0;$ 
2 for ( $i = 1; i \leq n; i = i + 1$ )
3    $a_i = a_{i-1} + d_i * d_i;$ 

```

$d_i = 0$ otherwise, to deduce the accumulated effect of all the errors on the output of the filter. It is defined as the sequence of real numbers such that

$$y_{-1} = 0, \quad y_0 = 1, \quad \text{and} \quad y_n = -b_1 y_{n-1} - b_2 y_{n-2} \quad \text{for all } n \geq 1.$$

This sequence can also be defined by the expression

$$y_n = b_2^{n/2} \sqrt{b_2 + 2b_1^2} \cos(\omega_0 + n\omega) \quad \text{with constants } \omega_0 \text{ and } \omega.$$

If the filter is bounded-input bounded-output (BIBO) stable, $0 < b_2 < 1$ and the accumulated effect of the round-off errors is easily bounded by $\sqrt{b_2 + 2b_1^2} / (1 - \sqrt{b_2})$. Worst case error analysis is not possible on BIBO unstable systems. Our work and the example of Table 1 can be applied to such systems.

2.3 Long sums of squares and Taylor series expansion of programs

The previous programs introduce only first order effect of the round-off errors. We present here systems that involve higher order errors such as sum of square in Listing 3 and power series of all the random variables as in equation 4.

Assuming that d_i carries an error X_i in Listing 3, its contribution to the sum of square cannot be assumed to be symmetric. Lévy's inequality cannot be applied, but Doob's inequality provides a similar result though it is out of reach with current formal tools and libraries.

The output of a system can always be seen as a function F of its input and its state (d_0, \dots, d_n) . This point of view can be extended by considering that the output of the system is also a function of the various round-off errors (X_0, \dots, X_q) introduced at run-time. Provided this function can be differentiated sufficiently, Taylor series expansion provides that

$$\begin{aligned}
F(d_0, \dots, d_n, X_0, \dots, X_q) &= F(d_0, \dots, d_n, 0, \dots, 0) \\
&+ \sum_{m=1}^r \frac{1}{m!} \left(\sum_{i=0}^q X_i \frac{\partial F}{\partial X_i} \right)^{[m]} (d_0, \dots, d_n, 0, \dots, 0) \\
&+ \left(\sum_{i=0}^q X_i \frac{\partial F}{\partial X_i} \right)^{[m+1]} (d_0, \dots, d_n, \theta_0, \dots, \theta_q),
\end{aligned} \tag{4}$$

where θ_i is between 0 and D_i , and $(\cdot)^{[m]}$ is the symbolic power defined as

$$\left(\sum_{i=0}^q X_i \frac{\partial F}{\partial X_i} \right)^{[m]} = \sum_{0 \leq i_1, \dots, i_m \leq q} X_{i_1} \cdots X_{i_m} \frac{\partial^m F}{\partial X_{i_1} \cdots \partial X_{i_m}}.$$

When the Taylor series is stopped after $m = 2$, we can use Doob's inequality to provide results similar to the ones presented in Table 1 provided X_i are symmetric and independent. Higher order Taylor series don't necessarily create sub-martingales but weaker results can be obtained by combining inequalities on sub-martingales.

3 Formal background on probability

3.1 A generic and formal theory of probability

We rebuilt the previously published theory of probability spaces [13] as a theory of Lebesgue's integration recently became fully available. The new PVS development in Figure 1, still takes three parameters: T , the sample space, \mathcal{S} , a σ -algebra of permitted events, and \mathbb{P} , a probability measure, which assigns to each permitted event in \mathcal{S} , a probability between 0 and 1. Properties of probability that are independent of the particular details of T , \mathcal{S} , and \mathbb{P} are then provided in this file.

A random variable X is a measurable application from (T, \mathcal{S}) to any other measurable space (T', \mathcal{S}') . In most theoretical developments of probability T , \mathcal{S} , and \mathbb{P} remain generic as computations are carried on T' . Results on real random variables use $T' = \mathbb{R}$ whereas results on random vectors use $T' = \mathbb{R}^n$. Yet both theories refer explicitly to the Borel sets of T' .

As the Borel sets of \mathbb{R}^n are difficult to grasp, most authors consider finite T and $\mathcal{S} = \mathcal{P}(T)$ for discrete random variables in introductory classes. This simpler analysis is meant only for educational purposes and most results of probability considered for formal methods can be implemented with generic T , \mathcal{S} , and \mathbb{P} parameters.

Handling discrete and continuous random variables through different T and \mathcal{S} parameters is not necessary and it is contrary to most uses of probability spaces in mathematics. Such variables can be described on the same generic T , \mathcal{S} , and \mathbb{P} parameters in spite of their differences. In practice, we use $T' = \mathbb{R}$ or $T' = \mathbb{R}^n$, and for discrete variables we can choose countable codomains.

Similarly, many authors work on sections $\{X \leq x\}$ rather than using *the inverse images of Borel sets* of T' because the latter are difficult to visualize. Such a simplification is valid thanks to Dynkin's systems. But using abstract Borel sets rather than sections in formal methods often leads to easier proofs.

3.2 A concrete theory of expectation

The previous theory of random variables [13] made it possible to define them and to use and derive their properties. Very few results were enabling users to actually compute concrete results on random variables. Most of such results lie on a solid theory of the expected value. As most theorems in the later theory are corollaries of a good theory of Lebesgue's integration, we have developed a formal measure theory based on Lebesgue's integration and we develop formal theorems on expected values as needed in our applications.

The expected value is the (unique) linear and monotonous operator \mathbb{E} on the set of \mathbb{P} -integrable random variables that satisfies Beppo-Lévy's property and such that $\mathbb{E}(\chi_A) = \mathbb{P}(A)$ for all $A \in \mathcal{S}$. We can also use the following definition when Lebesgue's integral exists:

$$\mathbb{E}(X) = \int_T X \, d\mathbb{P}.$$

```

probability_space[T:TYPE+,
                (IMPORTING topology@subset_algebra_def[T])
                S:sigma_algebra, (IMPORTING probability_measure[T,S])
                P:probability_measure]: THEORY
BEGIN
  IMPORTING topology@sigma_algebra[T,S],
            probability_measure[T,S],
            continuous_functions_aux[real],
            measure_theory@measure_space[T,S],
            measure_theory@measure_props[T,S,to_measure(P)]

  limit: MACRO [(convergence_sequences.convergent)->real]
          = convergence_sequences.limit

  h : VAR borel_function
  A,B: VAR (S)
  x,y: VAR real
  n0z: VAR nzreal
  t: VAR T
  n: VAR nat
  X,Y: VAR random_variable
  XS: VAR [nat->random_variable]

  null?(A) :bool = P(A) = 0
  non_null?(A) :bool = NOT null?(A)
  independent?(A,B):bool = P(intersection(A,B)) = P(A) * P(B)

  zero: random_variable = (LAMBDA t: 0)
  one: random_variable = (LAMBDA t: 1)

  <=(X,x):(S) = {t | X(t) <= x}; % Needed for syntax purposes! < > = /= >= omitted

  complement_le1: LEMMA complement(X <= x) = (x < X) % More omitted

  +(X,x) :random_variable = (LAMBDA t: X(t) + x); % Needed for syntax purposes! More omitted

  borel_comp_rv_is_rv: JUDGEMENT o(h,X) HAS_TYPE random_variable

  partial_sum_is_random_variable:
    LEMMA random_variable?(LAMBDA t: sigma(0,n,LAMBDA n: XS(n)(t)))

  distribution_function?(F:[real->probability]):bool
    = EXISTS X: FORALL x: F(x) = P(X <= x)
  distribution_function: TYPE+ = (distribution_function?) CONTAINING
    (LAMBDA x: IF x < 0 THEN 0 ELSE 1 ENDIF)
  distribution_function(X)(x):probability = P(X <= x)

  convergence_in_distribution?(XS,X):bool
    = FORALL x: continuous(distribution_function(X),x) IMPLIES
      convergence((LAMBDA n: distribution_function(XS(n))(x)),
                  distribution_function(X)(x))

  invert_distribution: LEMMA LET F = distribution_function(X) IN
                      P(x < X) = 1 - F(x)
  interval_distribution: LEMMA LET F = distribution_function(X) IN
                          x <= y IMPLIES
                          P(intersection(x < X, X <= y)) = F(y) - F(x)
  limit_distribution: LEMMA LET F = distribution_function(X) IN
                      P(X = x) = F(x) - limit(LAMBDA n: F(x-1/(n+1)))

  F: VAR distribution_function

  distribution_0: LEMMA convergence(F o (lambda (n:nat): -n),0)
  distribution_1: LEMMA convergence(F,1)
  distribution_increasing: LEMMA increasing?[real](F)
  distribution_right_continuous: LEMMA right_continuous(F)
END probability_space

```

Figure 1: Abbreviated probability space file in PVS

Markov's inequality below is heavily used to obtain concrete properties on random variables.

Theorem 1 (Markov's inequality). *For any random variable X and any constant ε ,*

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}(|X|)}{\varepsilon}.$$

Many theorems relate to independent random variables and their proof are much easier once independence is well defined. The family (X_1, \dots, X_n) is independent if and only if, for any family of Borel sets (B_1, \dots, B_n) ,

$$\mathbb{P}\left(\bigcap_{i=1}^n (X_i \in B_i)\right) = \prod_{i=1}^n \mathbb{P}(X_i \in B_i).$$

The following characteristic property is used a lot on families of independent variables: For any family of Borelean functions (h_1, \dots, h_n) such that the $h_i(X_i)$ are \mathbb{P} -integrable,

$$\mathbb{E}\left(\prod_{i=1}^n h_i(X_i)\right) = \prod_{i=1}^n \mathbb{E}(h_i(X_i)).$$

It is worth noting that the fact that n random variables are independent is not equivalent to the fact that any pair of variables is independent and cannot be built recursively from $n - 1$ independent random variables.

Future work may lead us to implement a theory of the law \mathbb{P}_X associated to each random vector $X : \mathbb{T} \rightarrow \mathbb{R}^n$, with a "transfer" theorem for any Borelean function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ below and most properties of Lebesgue's integral including Fubini's theorem.

$$\mathbb{E}(h(X)) = \int_{\mathbb{T}} h(X) \, d\mathbb{P} = \int_{\mathbb{R}^n} h \, d\mathbb{P}_X$$

3.3 Almost certain a priori error bound

What we are actually interested in is whether a series of calculations might accumulate a sufficiently large error to become meaningless. In the language we have developed, we are computing the probability that a sequence of n calculations has failed because it has exceeded the ε error-bound somewhere.

Theorem 2 (Corollary of Lévy's inequality). *Provided the (X_n) are independent and symmetric the following property holds for any constant ε .*

$$\mathbb{P}\left(\max_{1 \leq i \leq n} |S_i| \geq \varepsilon\right) \leq 2\mathbb{P}(|S_n| \geq \varepsilon)$$

Proof. We use a proof path similar to the one published in [20]. We define $S_n^{(j)}$ below with Dirichlet's operator δ_P that is equal to 1 if the predicate holds and 0 otherwise. As the X_n are symmetric, the random variables S_n and $S_n^{(j)}$ share the same probability density function.

$$S_n^{(j)} = \sum_{i=1}^n (-1)^{\delta_{i>j}} X_i$$

We now define $N = \inf\{k \text{ such that } |S_k| \geq \varepsilon\}$ with the addition that $\inf \emptyset = +\infty$ and similarly $N^{(j)} = \inf\{k \text{ such that } |S_k^{(j)}| \geq \varepsilon\}$. Events $\max_{1 \leq i \leq n} (|S_i|) \geq \varepsilon$ and $N \leq n$ are identical. Furthermore

$$\mathbb{P}(|S_n| \geq \varepsilon) = \sum_{j=1}^n \mathbb{P}(|S_n| \geq \varepsilon \cap N = j) = \sum_{j=1}^n \mathbb{P}(|S_n^{(j)}| \geq \varepsilon \cap N = j).$$

As soon as $j \leq n$, $2S_j = S_n + S_n^{(j)}$ and $2|S_j| = |S_n| + |S_n^{(j)}|$. Therefore, the event $\{|S_j| \geq \varepsilon\}$ is included in $\{|S_n| \geq \varepsilon\} \cup \{|S_n^{(j)}| \geq \varepsilon\}$ and

$$\mathbb{P}(N \leq n) = \sum_{j=1}^n \mathbb{P}(|S_j| \geq \varepsilon \cap N = j) \leq \sum_{j=1}^n \mathbb{P}(|S_n| \geq \varepsilon \cap N = j) + \sum_{j=1}^n \mathbb{P}(|S_n^{(j)}| \geq \varepsilon \cap N = j).$$

This ends the proof of Lévy's inequality. \square

Should we need to provide some formula beyond the hypotheses of Lévy's inequality, we may have to prove Doob's original inequality for martingales and sub-martingales [21] in PVS. It follows a proof path very different from Doob-Kolmogorov's inequality but it is not limited to second order moment and it can be applied to any sub-martingale S_i^{2k} with $k \geq 1$ to lead to

$$\mathbb{P}\left(\max_{1 \leq i \leq n} (|S_i|) \geq \varepsilon\right) \leq \frac{\mathbb{E}(S_n^{2k})}{\varepsilon^{2k}}.$$

Shall we need to create a sub-martingale different from S_i^{2k} , we may have to prove Jensen's conditional inequality that let us introduce $h(|S_i|)$ where $h: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is convex. The bound becomes $\mathbb{E}(h(|S_n|)) / h(\varepsilon)$.

We use Markov's inequality applied to S_n^k in order to obtain the results of Table 1:

$$\mathbb{P}(|S_n| \geq \varepsilon) = \mathbb{P}(|S_n^k| \geq \varepsilon^k) \leq \mathbb{E}(|S_n^k|) / \varepsilon^k.$$

Formulas

$$\mathbb{E}(S_n^2) = u^2 \left(\frac{1}{3}n\right)$$

$$\mathbb{E}(S_n^4) = u^4 \left(\frac{1}{5}n + \frac{1}{3}n(n-1)\right)$$

$$\mathbb{E}(S_n^6) = u^6 \left(\frac{1}{7}n + n(n-1) + \frac{5}{9}n(n-1)(n-2)\right)$$

$$\mathbb{E}(S_n^8) = u^8 \left(\frac{1}{9}n + \frac{41}{15}n(n-1) + \frac{14}{3}n(n-1)(n-2) + \frac{35}{27}n(n-1)(n-2)(n-3)\right)$$

are based on the binomial formula for independent symmetric random variables

$$\mathbb{E}(S_n^{2k}) = \sum_{k_1+k_2+\dots+k_n=k} (2k)! \frac{\mathbb{E}(X_1^{2k_1})}{(2k_1)!} \frac{\mathbb{E}(X_2^{2k_2})}{(2k_2)!} \dots \frac{\mathbb{E}(X_n^{2k_n})}{(2k_n)!}.$$

Proof. We first prove the formula below by induction on n for any exponent m .

$$\mathbb{E}(S_n^m) = \sum_{m_1+m_2+\dots+m_n=m} m! \frac{\mathbb{E}(X_1^{m_1})}{m_1!} \frac{\mathbb{E}(X_2^{m_2})}{m_2!} \dots \frac{\mathbb{E}(X_n^{m_n})}{m_n!}$$

It holds for $n = 1$. We now write the following identity based on the facts that X_n are independent and symmetric. $\mathbb{E}(S_{n+1}^m) = \mathbb{E}((S_n + X_{n+1})^m)$ is also equal to

$$\mathbb{E}\left(\sum_{m_{n+1}=0}^m \frac{m!}{(m-m_{n+1})!m_{n+1}!} X_{n+1}^{m_{n+1}} S_n^{m-m_{n+1}}\right) = \sum_{m_{n+1}=0}^m \frac{m!}{(m-m_{n+1})!m_{n+1}!} \mathbb{E}(X_{n+1}^{m_{n+1}}) \mathbb{E}(S_n^{m-m_{n+1}})$$

We expand the terms of the sum for $m_{n+1} = 0, \dots, p$ and $\sum_{i=1}^n m_i = m - m_{n+1}$

$$\frac{m!}{(m - m_{n+1})! m_{n+1}!} \frac{(m - m_{n+1})!}{\prod_{i=1}^n m_i!} \prod_{i=1}^{n+1} \mathbb{E}(X_i^{m_i})$$

We end the proof for the even values of m after noticing that $\mathbb{E}(X_i^{2k+1}) = 0$ for any i and any k since X_n are symmetric. \square

4 Perspectives and concluding remarks

To the best of our knowledge this paper presents the first application of Lévy's inequality to software and system reliability of very long processes with an extremely low rate of failure. Our results allow any one to develop safe upper limits on the number of operations that a piece of numeric software should be permitted to undertake. In addition, we are finishing certification of our results with PVS. The major restriction lies in the fact that the slow process of proof checking has forced us to insist that individual errors are symmetric.

At the time we are submitting this work, the bottleneck is the full certification of more results using PVS proof assistant. Yet this step is compulsory to provide full certification to future industrial uses. We anticipate no problem as these results are gathered in textbooks in computer science and mathematics. This library and future work will be included into NASA Langley PVS library¹ as soon as it becomes stable.

The main contribution of this work is that we selected theorems that produce significant results for extremely low probabilities of failure of systems that run for a long time and that are amenable to formal methods. During our work, we discarded many mathematical methods that would need too many operations or that would be too technical to be implemented with existing formal tools.

Notice that this work can be applied to any sequence of independent and symmetric random variables that satisfy equation (2). It is worth pointing out one more time that violating our assumption (independence of errors) would lead to worse results, so one should treat the limit we have deduced with caution, should this assumption not be met.

Acknowledgment

This work has been partially funded by CNRS PICS 2533 and by the EVA-Flo project of the ANR. It was initiated while one of the authors was an invited professor at the University of Perpignan Via Domitia.

References

- [1] D. M. Russinoff, "A mechanically checked proof of IEEE compliance of the floating point multiplication, division and square root algorithms of the AMD-K7 processor," *LMS Journal of Computation and Mathematics*, vol. 1, pp. 148–200, 1998. [Online]. Available: <http://www.onr.com/user/russ/david/k7-div-sqrt.ps>
- [2] J. Harrison, "Formal verification of floating point trigonometric functions," in *Proceedings of the Third International Conference on Formal Methods in Computer-Aided Design*, W. A. Hunt and S. D. Johnson, Eds., Austin, Texas, 2000, pp. 217–233. [Online]. Available: <http://www.springerlink.com/link.asp?id=wxvaqu9wjrgc8199>

¹<http://shemesh.larc.nasa.gov/fm/ftp/larc/PVS-library/pvslib.html>.

- [3] S. Boldo and M. Daumas, “Representable correcting terms for possibly underflowing floating point operations,” in *Proceedings of the 16th Symposium on Computer Arithmetic*, J.-C. Bajard and M. Schulte, Eds., Santiago de Compostela, Spain, 2003, pp. 79–86. [Online]. Available: <http://perso.ens-lyon.fr/marc.daumas/SoftArith/BolDau03.pdf>
- [4] M. Daumas and G. Melquiond, “Generating formally certified bounds on values and round-off errors,” in *Real Numbers and Computers*, Dagstuhl, Germany, 2004, pp. 55–70. [Online]. Available: <http://hal.inria.fr/inria-00070739>
- [5] —, “Certification of bounds on expressions involving rounded operators,” *ACM Transactions on Mathematical Software*, 2009, to appear. [Online]. Available: <http://hal.archives-ouvertes.fr/hal-00127769>
- [6] M. Daumas, D. Lester, and C. Muñoz, “Verified real number calculations: A library for interval arithmetic,” *IEEE Transactions on Computers*, 2008, to appear.
- [7] J. Hurd, “Formal verification of probabilistic algorithms,” Ph.D. dissertation, University of Cambridge, 2002. [Online]. Available: <http://www.cl.cam.ac.uk/~jeh1004/research/papers/thesis.pdf>
- [8] P. Audebaud and C. Paulin-Mohring, “Proofs of randomized algorithms in Coq,” in *Proceedings of the 8th International Conference on Mathematics of Program Construction*, T. Uustalu, Ed., Kuressaare, Estonia, 2006, pp. 49–68. [Online]. Available: http://dx.doi.org/10.1007/11783596_6
- [9] M. Kaufmann, P. Manolios, and J. S. Moore, *Computer-Aided Reasoning: An Approach*. Kluwer Academic Publishers, 2000.
- [10] M. J. C. Gordon and T. F. Melham, Eds., *Introduction to HOL: A theorem proving environment for higher order logic*. Cambridge University Press, 1993.
- [11] G. Huet, G. Kahn, and C. Paulin-Mohring, *The Coq proof assistant: a tutorial: version 8.0*, 2004. [Online]. Available: <ftp://ftp.inria.fr/INRIA/coq/current/doc/Tutorial.pdf.gz>
- [12] S. Owre, J. M. Rushby, and N. Shankar, “PVS: a prototype verification system,” in *11th International Conference on Automated Deduction*, D. Kapur, Ed. Saratoga, New-York: Springer-Verlag, 1992, pp. 748–752. [Online]. Available: <http://pvs.csl.sri.com/papers/cade92-pvs/cade92-pvs.ps>
- [13] M. Daumas and D. Lester, “Stochastic formal methods: an application to accuracy of numeric software,” in *Proceedings of the 40th IEEE Annual Hawaii International Conference on System Sciences*, Waikoloa, Hawaii, 2007, p. 7 p. [Online]. Available: <http://hal.ccsd.cnrs.fr/ccsd-00081413>
- [14] D. Goldberg, “What every computer scientist should know about floating point arithmetic,” *ACM Computing Surveys*, vol. 23, no. 1, pp. 5–47, 1991. [Online]. Available: <http://doi.acm.org/10.1145/103162.103163>
- [15] D. Stevenson *et al.*, “An American national standard: IEEE standard for binary floating point arithmetic,” *ACM SIGPLAN Notices*, vol. 22, no. 2, pp. 9–25, 1987.
- [16] *TMS320C3x — User’s guide*, Texas Instruments, 1997. [Online]. Available: <http://www-s.ti.com/sc/psheets/spru031e/spru031e.pdf>
- [17] D. E. Knuth, *The Art of Computer Programming: Seminumerical Algorithms*. Addison-Wesley, 1997, third edition.
- [18] A. Feldstein and R. Goodman, “Convergence estimates for the distribution of trailing digits,” *Journal of the ACM*, vol. 23, no. 2, pp. 287–297, 1976. [Online]. Available: <http://doi.acm.org/10.1145/321941.321948>
- [19] J. Bustoz, A. Feldstein, R. Goodman, and S. Linnainmaa, “Improved trailing digits estimates applied to optimal computer arithmetic,” *Journal of the ACM*, vol. 26, no. 4, pp. 716 – 730, 1979. [Online]. Available: <http://doi.acm.org/10.1145/322154.322162>
- [20] J. Bertoin, “Probabilités,” 2001, cours de licence de mathématiques appliquées. [Online]. Available: <http://www.proba.jussieu.fr/cours/bertoin.pdf>
- [21] J. Neveu, Ed., *Martingales à temps discret*. Masson, 1972.