An application to accuracy of numeric software

Authors

- Marc Daumas (Marc.Daumas@Univ-Perp.Fr)
  - Professor at the University of Perpignan
  - Perpignan, France 66860
- David Lester (David.R.Lester@Manchester.Ac.UK)
  - Invited professor at the University of Perpignan
  - School of Computer Science - University of Manchester
  - Manchester, United Kingdom M13 9PL

- Sponsors
  - CNRS PICS 2533
- Acknowledgment
  - Harold Simmons and Jean-Marc Vincent
  - Mathlogaps (European Union)

Outline

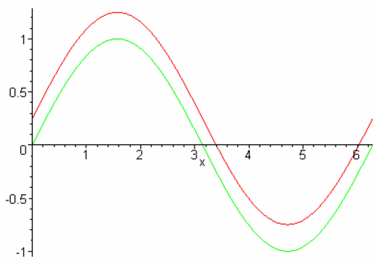
- FAA regulations for aircraft require that the probability of an error be below 10^{-9} for a 10 hour flight
  - Provides a bound on the number of numeric operations (fixed or floating point) that can safely be performed before accuracy is lost
  - Important implications for control systems with safety-critical software
  - Worst-case analysis would blindly advise the replacement of existing systems that have been successfully running for years
  - Set of formal theorems validated by the PVS proof assistant
  - Allow code analyzing tools to produce formal certificates

Systems are now running fast enough and long enough for their errors to impact on their functionality

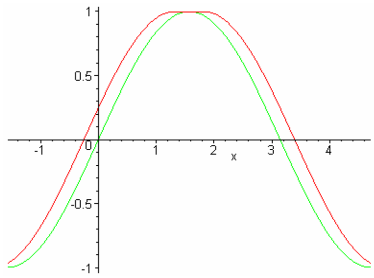
- Worst case analysis is meaningless for applications that run for a long time
- For example
  - A process adds numbers in ±1 to single precision
  - Each addition produces a round-off error of $\pm 2^{-25}$
  - This process adds 2^{25} items
  - The accumulated error is ±1
- Note that
  - 10 hours of flight time
  - At operating frequency of 1 kHz
  - Is approximately 2^{25} operations
- Provided round-off errors are not correlated, the actual accumulated error will be much smaller

Some easy ways to obtain worst case behavior

- Systematic ad-hoc errors may lead to the slow accumulation of small quantities of the same sign
  - Biased measures



  - Synchronized time shift



EVA-Flo
Perpignan
Oct 2007

Developing probabilities on floating point arithmetic

- Formal proof assistants such as ACL2, HOL, Coq and PVS are used in areas where
  - Errors can cause loss of life or significant financial damage
  - Common misunderstandings can falsify key assumptions
- Developments in probability share many features with developments in floating point arithmetic:
  - Each result usually relies on a long list of hypotheses and slight variations induce a large number of results that look almost identical
  - Most people want a trustworthy result but they are not proficient enough to either select the best scheme or detect minor faults that can quickly lead to huge problems
- Validation of a safety-critical numeric software using probability should be done using an automatic proof checker

EVA-Flo
Perpignan
Oct 2007

Related work in probability

- Asymptotic behavior
  - Continuous space Markov random walks
  - Renewal-reward processes
- We want to precisely bound the probability of remaining within bounds for a (large) given number of steps.
  - Ruin probabilities
  - Martingales (Doobs-Kolmogorov inequality)
- Created round-off and measure errors are
  - Unbiased
  - Independent random variables
  - (their expectation conditional to the previous errors is zero)

I – Stochastic model

Individual round-off errors of fixed and floating point operations

Round off errors of an accumulation loop

Individual round-off errors of fixed and floating point operations

- We use sign-magnitude or two's complement notation for the mantissa and an implicit first bit for the mantissa is in most cases

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

- One unit in the last place of v defined as above is

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

- Trailing digits of numbers randomly chosen from a logarithmic distribution are approximately uniformly distributed in $\pm ulp(v)/2$
- Sensors may be less accurate leading to a larger variance but they should not be biased
- Round-off errors created by operators are discrete and specific but expectations are 0 and we bound variances

EVA-Flo
Perpignan
Oct 2007

Round off errors of an accumulation loop : Simple discrete integration

- We sum data produced by a sensor x_i with a measure error X_i
- $a_0 = 0$; for (i = 0; i < n; i = i + 1) $a_{i+1} = a_i + x_i$;
- We can safely assume that X_i are independent identical uniformly distributed random variables over $\pm ulp(x_i)/2$
- Data are fixed point
  - The sum $a_i + x_i$ does not introduce any rounding error
  - One unit in the last place does not depend on x_i
- Using the Doobs-Kolmogorov inequality where $S_i = \Sigma_{j=1}^{i} X_j$
- We have the probability that the accumulated measure error have always been constrained into user specified bounds ε for n iterations

$$P(\max_{1 \le i \le n}(|S_i|) \le \varepsilon) \ge 1 - n\, ulp^2 / (12\varepsilon^2)$$

EVA-Flo
Perpignan
Oct 2007

Round off errors of an accumulation loop : Solving initial value problem ODE (1)

- Compute an incremental slope $\Phi(t_i, h_i, x_i, f)$
  - Based on the current time, step size, value of the function and the differential equation $x'(t) = f(t, x(t))$.
  - Many methods (Euler, Runge-Kutta, implicit, adaptative…)
- for (i = 0; i < n; i = i + 1)
  - $\{x_{i+1} = x_i + h_i \cdot \Phi(t_i, h_i, x_i, f);\ \ t_{i+1} = t_i + h_i;\ h_{i+1} = h_i\ \}$
- Introduce a sequence of random variables $\{X_n\}$ that models the difference introduced by errors
- In most cases Φ introduces
  - A drift due to higher order effects
  - Correlations between the error introduced at step i+1 and errors of the previous steps
- For example, the square of a rounded value v + V where v is the stored value and V is a random variable, introduces a positive drift due to V^2 term

EVA-Flo
Perpignan
Oct 2007

Round off errors of an accumulation loop : Solving initial value problem ODE (2)

- We model the effect of errors by two terms X_i and Y_i
- We use
  - The Doobs-Kolomogorov inequality for $\{X_n\}$ that is constructed to contain only independent random variables with no drift and we only need to bound their variance
  - Worst case error analysis for $\{Y_n\}$ with interval arithmetic so that $E(X_n; X_1 \dots X_{n-1}) = 0$
- Random variables X_{i+1} and Y_{i+1} account for the errors replacing

$$x_i + X_i + Y_i + h_i \cdot \Phi\ (t_i, x_i + X_i + Y_i, h_i, f)$$

with

$$fl(x_i + h_i \cdot \Phi\ (t_i, x_i, h_i, f))$$

- fl(.) denote the evaluation of an expression on computer
- Software such as Fluctuat is already able to distinguish between first order and higher order error terms

EVA-Flo
Perpignan
Oct 2007

II – Probability distribution of being safe

Probability

A Formal Development of probability

Continuous Uniform Random Variables

Reliability of long calculations

Probability

- Two main choices in presenting an account of probability
  - One is to take an informal approach
  - The second involves taking foundational matters seriously
- I will consistently present matters informally, however the PVS system underlying these results is built on the firm foundations for probability theory (using measure theory)
- *A random variable X has distribution function F, if $P(X \leq x) = F(x)$*
- *A random variable X is continuous if its distribution function can be expressed as $F(x) = \int_{-\infty}^{x} f(x)\, dx$ for some integrable function $f: R \to [0,\infty)$ called the probability density function for the random variable X*
- *The conditional probability of "A given B" is defined as $P(A;B) = P(A \cap B)/P(B)$ whenever $P(B)>0$*

EVA-Flo
Perpignan
Oct 2007

Examples of probability

- The temperature T in an industrial process can be modeled as a continuous random variable
  - Even if an attempt is being made to hold this temperature constant, there will be minor fluctuations
- Example of conditional probability
  - Event A is "I am carrying an umbrella"
  - Event B is "it is raining"
  - P(A;B) is the probability that "I am carrying an umbrella given that it is raining"
  - Note that in general P(A;B) ≠ P(B;A)
    - Though, if you live in Perpignan or Manchester, then on most days: P(A;B) = P(B;A), though for rather different reasons
- Example of independent random variables
  - We model the outcomes of the tossing of two coins C_1 and C_2
  - We expect the result of tossing C_1 to have no effect on the result of C_2
- Consider an alternative scenario where C_1 and C_2 are dependent
  - We toss C_1 and discover that it has come up "heads"
  - We now define C_2 as "the downward facing side of the coin C_1 is tails"

A Formal Development of probability

- *A <u>σ-algebra</u> over a type T, is a subset of the power-set of T, which includes the empty set {}, and is closed under the operations of complement, countable union and countable intersection*
  - For discrete random variables, T is countable and σ =P(T)
  - For continuous random variables, T is the reals and σ =B: the Borel sets
- *A <u>Measurable Space</u> (T,σ) is a set (or in PVS a type) T, and a σ-algebra over T*
- *A function $\mu:\sigma\rightarrow R_{>0}$ is a <u>Measure</u> over the σ-algebra σ , when μ({}) = 0, and for a sequence of disjoint elements $\{E_n\}$ of σ: $\mu(U_{n=0}^{\infty} E_n) = \Sigma_{n=0}^{\infty} \mu(E_n)$*
- *A <u>Measure Space</u> (T,σ,μ) is a measurable space (T,σ) equipped with a measure μ*
- *A <u>Probability Space</u> (T,σ,P) is a measure space (T,σ,P) in which the measure P is finite for any set in σ, and in which: $P(X^c) = 1\text{-}P(X)$*
- *If (T_1,σ_1,P_1) and (T_2,σ_2,P_2) are probability spaces then we can construct a <u>product probability space</u> (T_3,σ_3,P_3), where: $T_3 = T_1 \cdot T_2$, $\sigma_3 = \sigma(\sigma_1 \cdot \sigma_2)$ and $P_3'(a,b) = P_1(a)\ P_2(b)$ where P_3 is the extension of P'_3 that has the whole of σ_3 as its domain*
  - Note P_3 has the effect of declaring that the experiments carried out in probability spaces (T_1,σ_1,P_1) and (T_2,σ_2,P_2) are independent

Abbreviated probability space file in PVS

```
probability_space[T:TYPE+,        (IMPORTING finite_measure@subset_algebra_def[T]) % sample space
            S:sigma_algebra, (IMPORTING probability_measure[T,S])                  % permitted events
            P:probability_measure                                                  % probability measure
           ]: THEORY
% To discuss continuous random variables we partially instantiate this PVS file with  T = real and S =   borel_set
BEGIN
  IMPORTING finite_measure@sigma_algebra[T,S],probability_measure[T,S],continuous_functions_aux[real]
  A,B: VAR (S)
  x,y: VAR real …
  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) % Note that it DOES NOT say = 0
  random_variable?(X:[T->real]):bool = FORALL x: member({t | X(t) <= x},S)
  zero: (random_variable?) = (LAMBDA t: 0)
  random_variable: TYPE+ = (random_variable?) CONTAINING zero

  X,Y: VAR random_variable
  XS:  VAR [nat->random_variable]
  <=(X,x):(S) = {t | X(t) <= x}; …
  complement_le1: LEMMA complement(X <= x) = (x <  X) …
  +(X,Y)  :random_variable = (LAMBDA t: X(t) + Y(t)); …
  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)
  F: VAR distribution_function
  convergence_in_distribution?(XS,X):bool = …
END probability_space
```

EVA-Flo
Perpignan
Oct 2007

UPVD

Conditional probability file in PVS

```
conditional[T:TYPE+,        (IMPORTING finite_measure@subset_algebra_def[T]) % sample space
         S:sigma_algebra,    (IMPORTING probability_measure[T,S])             % permitted events
         P:probability_measure                                                % probability measure
        ]: THEORY

BEGIN
  IMPORTING probability_space[T,S,P],finite_measure@sigma_algebra[T,S]
  A,B:   VAR (S)
  n,i,j: VAR nat
  AA,BB: VAR disjoint_sequence
  P(A,B):probability = IF null?(B) THEN 0 ELSE P(intersection(A,B))/P(B) ENDIF

  conditional_complement: LEMMA P(A,B)*P(B)+P(A,complement(B))*P(complement(B)) = P(A)

  conditional_partition: LEMMA
       Union(image(BB,fullset[below[n+1]])) = fullset[T] IMPLIES
       P(A) = sigma(0,n, LAMBDA i: P(A, BB(i)) * P(BB(i)))

  bayes_theorem: THEOREM
       NOT null?(B) AND Union(image(AA,fullset[below[n+1]])) = fullset[T] IMPLIES
       P(AA(j),B) = P(B,AA(j))*P(AA(j)) / sigma(0,n, LAMBDA i: P(B, AA(i)) * P(AA(i)))
END conditional
```

EVA-Flo
Perpignan
Oct 2007

UPVD

Continuous Uniform Random Variables

- If X is a continuous random variable distributed uniformly over the interval [a,b], then informally it takes any value within the interval [a,b] with equal probability
- *The characteristic function of a set S is the function χ_S, which takes the values 1 when it is applied to a member of S and 0 otherwise*
- The probability density function f is $1/(b-a)\ \chi_{(a,b]}$
- The distribution function is $F(x) = \int_{-\infty}^{x} f(x)\,dx$
- The probability $P(x<X<=y) = F(y)-F(x)$
- If X is distributed $U_{[a,b]}$
  - $E(X)=(a+b)/2$ and $V(X)=(a-b)^2/12$
  - With a=0, b=1 we get $E(X)=1/2$ and $V(X)=1/12$

Sums of Continuous Random Variables

- *We have a sequence of continuous random variables $\{X_n\}$*
- *We define their partial sums as a sequence of continuous random variables $\{S_n\}$ with the property $S_n = \Sigma_{i=1}^{n}$ X_i.*
- *If continuous random variables X and Y have joint probability density functions f, then Z=X+Y has probability density function*

$$f_Z(z) = \int_{-\infty}^{\infty} f(x,z-x)\,dx$$

- *Continuous Convolution Theorem: If continuous random variables X and Y are independent and have probability density functions f_X and f_Y respectively, then Z=X+Y has probability density function*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx = \int_{-\infty}^{\infty} f_X(z-x) f_Y(x)\,dx$$

Reliability of long calculations $P(\max_{1 \le i \le n}(|S_i|) \le \varepsilon)$

- *A sequence $\{S_n\}$ is a <u>martingale</u> with respect to the sequence $\{X_n\}$, if for all n: $E(|S_n|)<\infty$ and $E(S_{n+1};X_1,X_2,\ldots,X_n) = S_n$*
- The sequence $\{S_n\}$, where $S_n= \Sigma_{i=1}^{n} X_i$ is martingale with respect to the sequence $\{X_n\}$ if X_n are independent random variables with $E(X_n)=0$ or for all i, $E(X_i) = 0$ and $E(X_i;X_1 \ldots X_{i-1}) = 0$
- *<u>Doobs-Kolmogorov Inequality</u>: If $\{S_n\}$ is a martingale with respect to $\{X_n\}$ then, provided that $\varepsilon>0$: $P(\max_{1 \le i \le n}(|S_i|) \ge \varepsilon) \le E(S_n^2) / \varepsilon^2$*
- When each X_i is an independent random variable with $E(X_i)=0$, we observe that $P(\max_{1 \le i \le n}(|S_i|) \le \varepsilon) \ge 1 - 1/\varepsilon^2\, \Sigma_{i=1}^{n} V(X_i)^2$
- Eventually errors will accumulate and overwhelm the accuracy of any numerical software
  - If ε is large enough and each of the $V(X_i)^2$ are small enough
  - The number of iterations required for this to occur will be high enough to be of no practical significance
- Crucially, the results hinge critically on the errors $\{X_n\}$ being independent

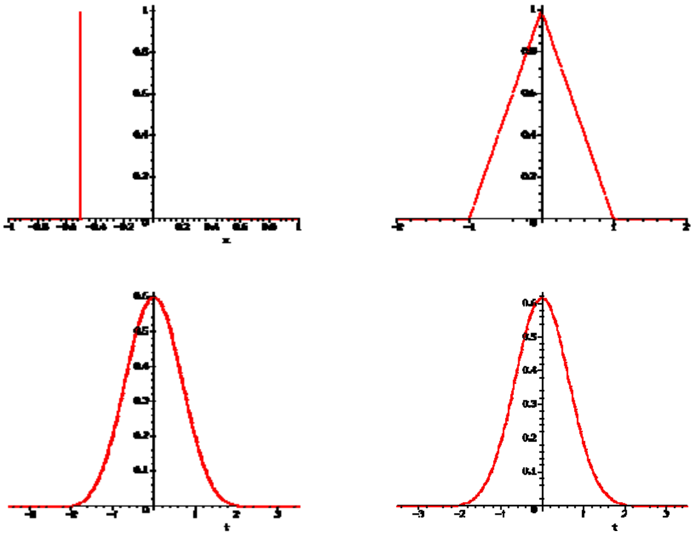
III – Concluding remarks

The Central Limit Theorem in action

Future work

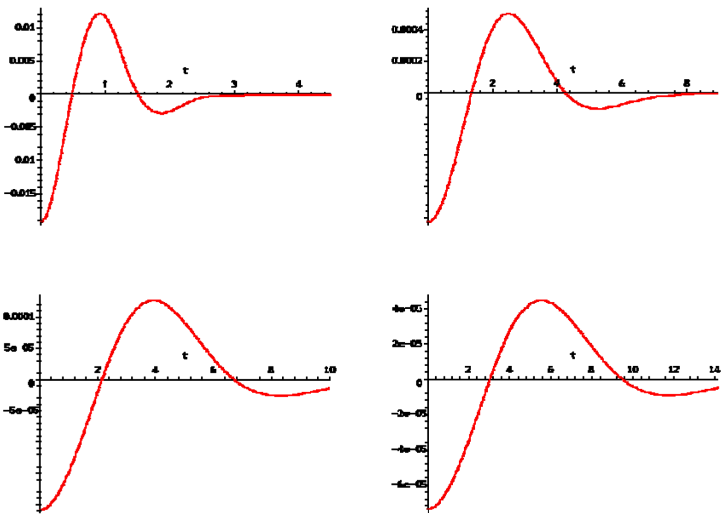
Conclusions

The Central Limit Theorem in action (n = 1, 2 or 5)



EVA-Flo
Perpignan
Oct 2007

Limitations of the Central Limit Theorem to target probability 10^{-9} (n = 5, 40, 100 or 200)



EVA-Flo
Perpignan
Oct 2007

Future work
1 - Invisible formal methods (Shankar & Rushby™)

- Modify Fluctuat to generate theorems that can be checked automatically by PVS using ProofLite
  - Collaboration with the developers of
    - Fluctuat (CEA)
    - ProofLite (NASA & NIA)
- Conservatively estimate the final effect of the error introduced by each individual floating point operations
- Compute upper bounds of their variances
- Obtain tighter results with tools that are able to infer and solve inductions on variances of random variables

Future work
2 – Contribute theories, theorems and facts

- Develop and validate in PVS accurate proofs about the round-off errors of operations
- Handle random variables with a drift through Wald Identity
  - Two's complement operation of TMS320 may truncate results
  - Address higher order error terms
- Library and future work will be included into NASA Langley PVS library

EVA-Flo

Perpignan

Oct 2007

Conclusions
1 - First generic formal development in PVS

- Able to handle random variables
  - Continuous
  - Discrete
  - Non-continuous non-discrete
- Previous developments in higher order logic were
  - Targeting other applications
  - Using other proof assistants (Coq, HOL or Mizar)
- See
  - Hurd's PhD and references herein (Cambridge→Oxford)
  - ALEA library by Audebaud and Paulin (ENS Lyon & Orsay)

EVA-Flo

Perpignan

Oct 2007

Conclusions
2 – One last warning

- First application of the Doobs-Kolmogorov Inequality to software reliability
- The limit on the reliability of a piece of numeric software can be expressed succinctly
- Even with a high tolerance of error, and with independent errors, we will still eventually fail
- Our results permit the development of safe upper limits on the number of operations that a piece of numeric software should be permitted to undertake similar to what was done in Gappa
- Violating our assumptions (independence of errors, and zero drift) would lead to worse results, so one should treat the limits we have deduced with caution, should these assumptions not be met