Compositional Solution Space Quantification for Probabilistic Software Analysis

Mateus Borges  
Federal University of Pernambuco, Brazil

Antonio Filieri  
University of Stuttgart, Germany

Corina S. Păsăreanu  
CMU SV/NASA Ames Research Center, United States

Marcelo d’Amorim  
Federal University of Pernambuco, Brazil

Willem Visser  
University of Stellenbosch, South Africa

Abstract

Probabilistic software analysis aims at quantifying how likely a target event is to occur during program execution. Current approaches rely on symbolic execution to identify the conditions to reach the target event and try to quantify the fraction of the input domain satisfying these conditions. Precise quantification is usually limited to linear constraints, while only approximate solutions can be provided in general through statistical approaches. However, statistical approaches may fail to converge to an acceptable accuracy within a reasonable time.

We present a compositional statistical approach for the efficient quantification of solution spaces for arbitrarily complex constraints over bounded floating-point domains. The approach leverages interval constraint propagation to improve the accuracy of the estimation by focusing the sampling on the regions of the input domain containing the sought solutions. Preliminary experiments show significant improvement on previous approaches both in results accuracy and analysis time.

Categories and Subject Descriptors  D.2.4 [Software Engineering]: Software/Program Verification

Keywords Symbolic Execution, Monte Carlo Sampling, Probabilistic Analysis, Testing

1. Introduction

The pervasiveness of software, the heterogeneity of its users, and the growing complexity of interactions with third-party components are introducing a new degree of uncertainty about the execution conditions of a program, raising the quest for verification techniques able to deal with and to quantify such uncertainty, both in the problem and in the verification results. Probabilistic software analysis aims at quantifying the probability of a target event to occur, given a probabilistic characterization of the behavior of a program or of its execution environment. Examples of target events include an uncaught exception, the invocation of a certain method, or the access to confidential information.

In the past probabilistic software analysis has been mostly performed at model level [20], limiting its applicability to early software design stages or requiring explicit abstraction from the code. Some recent techniques have brought it at the code-level [11, 12, 30]. These techniques use a symbolic execution of the program to collect the symbolic constraints on the input that lead to the occurrence of the target events. The constraints are then analyzed to quantify how likely is an input, distributed according to certain usage profiles [11], to satisfy any of them.

The quantification of the solution space for a set of constraints is one of the main obstacles to the applicability of probabilistic software analysis in practice. In previous studies [11, 12], model counting techniques have been applied to count the number of points of a bounded integer domain that satisfy given linear constraints. These counts are then coupled with a probabilistic input usage profile, which establishes a probability distribution over the input domain, to assess the probability for the target event to occur in that usage profile. Sankaranarayanan et al. [30] proposed a different approach to quantify the solution space of linear constraints over bounded floating-point domains. They propose an iterative algorithm to compute tight over-approximating bounds of the actual solution space suitable for efficient volume computation. All these approaches are thus quite limited as they can only handle linear constraints. In this paper we tackle the problem for the general case of complex (non-linear, containing transcendental functions, etc.) mathematical constraints over floating-point domains with application to probabilistic software analysis.

When considering floating-point domains, the quantification of the solution space for a constraint usually involves the computation of an integral. Symbolic and numerical integration are usually impracticable in the case of complex mathematical constraints and multi-dimensional input domains due to the high computational time and memory demands of such methods [17]. Furthermore, despite the fact they could be able to provide exact results, symbolic methods cannot deal with problems whose solution cannot be expressed in analytical form, which is the case for many integration problems [15]. Statistical methods overcome the limitations in terms of feasibility and memory demand of both symbolic and numerical methods, allowing to deal with arbitrarily complex constraints. However, statistical methods are based on simulation and can only provide approximate results. While the accuracy of the results can be arbitrarily improved, the price to be paid is the in-
Probabilistic software analysis is concerned with quantifying how likely software execution is to satisfy a given property. Probabilistic software analysis is relevant in contexts where the software is designed to exhibit uncertain or randomized behavior or when the execution environment, including interactions with users, is characterized by a probabilistic profile. In these situations, it is usually more relevant to quantify the probability of satisfying (violating) a given property than to just assess the possibility of such events to occur.

3. Probabilistic Software Analysis

Probabilistic software analysis is concerned with quantifying how likely software execution is to satisfy a given property. Probabilistic software analysis is relevant in contexts where the software is designed to exhibit uncertain or randomized behavior or when the execution environment, including interactions with users, is characterized by a probabilistic profile. In these situations, it is usually more relevant to quantify the probability of satisfying (violating) a given property than to just assess the possibility of such events to occur.

Even though the low and high values of these intervals are floating-point numbers, the box denotes a 2-dimensional region of real solutions. RealPaver uses two parameters to determine how tight the boxes reported on output: the decimal bound and the time budget. The bound is the number of decimal digits that limits the size of the smallest box that RealPaver can report and the time budget limits the time for computing boxes. Note that, irrespective of these parameters, the output should always include all real solutions. The smaller the decimal bound and the higher the time budget the tighter are the boxes reported. The approach is flexible to support a wide range of mathematical functions.

Figure 1 shows the main flow of the tool that we use to support probabilistic analysis. The tool takes on input a Java program and the usage profile for the program’s input variables. The usage profile includes the domain of input variables and the probability distribution associated with each domain. The output of the tool is an estimate of the probability for satisfying (or violating) a property of interest, e.g. an assertion in the code or a certain observable event of interest. Internally, the tool uses symbolic execution to produce a set of path constraints. In particular we use Symbolic PathFinder (SPF) for the symbolic execution of Java bytecode, but other similar tools can be used. In this context, satisfaction of one individual path constraint implies the occurrence of the event of interest. The work reported in this paper focuses on the last processor in the figure pipeline, the “Probabilistic Analysis” component, which takes as input a disjunction of path conditions and, based on that, it computes the probability of the event to occur. This component is not dependent on any programming language.

3.1 Computing Probabilities

Let’s classify the path constraints produced with symbolic execution in the sets $PC_T$ and $PC_F$ based on whether they lead to the occurrence of the target event (T) or not (F).

Under the assumption that symbolic execution terminates and only produces constraints for complete paths, the path constraints define disjoint input sets and they cover the whole input domain.
In detail, in order to deal with possible non-termination due to looping constructs in the code, SPF actually performs bounded symbolic execution. Hence, if an event has not occurred within the symbolic execution bound, the corresponding PC is not included in \( PC^T \). On the other hand, since hitting the execution bound is an event observable though SPF, it is possible to introduce a third set of PCs containing those where the bound has been hit and quantify the probability of such sets as well; this probability can give a measure for the confidence in the results obtained within the bound (the lower the probability the higher the confidence). This approach has been applied, for example in [11], but for the sake of space we will not consider it here.

We define the probability of satisfying a given property, as the probability of an input distributed according to the usage profile to satisfy any of the path constraints in \( PC^T \). Formally, what we aim to compute is:

\[
\int_D \mathbb{1}_{PC^T}(x) \cdot p(x) \\
\tag{1}
\]

where \( D \) is the input domain defined as the Cartesian product of the domains of the input variables, \( p(x) \) is the probability of a given \( x \) to occur for the given usage profile, and \( \mathbb{1}_{PC^T}(x) \) is the indicator function on \( PC^T \), that is a function that returns 1 when \( x \) satisfies any of the PCs in \( PC^T \), and 0 otherwise [26]. Equation (1) represents the expected probability for satisfying the target property.

### 3.2 Monte Carlo Simulation

In general, the constraints in \( PC^T \) can be non-linear or can make the integral ill-conditioned for numerical solutions [17]. To retain generality, qCORAL builds on simulation-based methods. These methods have theoretically no limitations on the complexity of the constraints they can handle. However, they can take a long time to converge, especially in the presence of large, multidimensional domains. The simplest simulation-based method applicable to the problem at hand is the Hit-Or-Miss Monte Carlo method [29]. In practice, it consists in generating a sequence of \( n \) independent inputs \( (x_0, x_1, \ldots, x_{n-1}) \), according to the usage profile, and to count the number of hits, i.e., the number of samples that satisfy the constraints under analysis. The ratio of hits (fraction of samples which are hits) provides an efficient, unbiased, and consistent estimator for the probability of satisfying the constraints [26]. In particular such estimator \( \hat{X} \) has a binomial distribution with expected values and variance given by the following equations [26]:

\[
E[\hat{X}] = \bar{x} \quad Var[\hat{X}] = \frac{\bar{x} \cdot (1 - \bar{x})}{n} \\
\tag{2}
\]

where \( \bar{x} = \frac{\sum_{i=0}^{n-1} \mathbb{1}_{PC^T}(x_i)}{n} \) is the sample mean. Note that the more samples are collected, the closer \( \hat{X} \) gets to the integral of Equation (1). In particular, this convergence can be quantified by the estimator variance \( \text{Var} \) \( \hat{X} \): the closer the variance is to 0 the more accurate is the estimation.

### 3.3 Interval Constraint Propagation and Stratified Sampling

Despite their generality, hit-or-miss Monte Carlo methods may suffer from a slow convergence rate [17], especially if the probability of the target event gets close to zero [32]. If additional information is known about the problem under analysis, it can be exploited to improve the estimation performance, i.e. to reduce the estimator variance. A well established method for variance reduction is the stratified sampling [19, 29]. This approach consists in partitioning the sample space, which in our case corresponds to the input domain \( D \), into disjoint subsets \( \{R_0, R_1, \ldots, R_m\} \), called strata. Each strata \( R_i \) can be analyzed separately via hit-or-miss Monte Carlo, obtaining the corresponding estimator \( \hat{X}_i \). Since the sampling processes within each region is independent, and assuming we take the same number of samples on each strata [29], the strata estimators can be combined to obtain an estimator \( \hat{X} \) over the original sample space, with the following expected value and variance [29]:

\[
E[\hat{X}] = \sum_i w_i \cdot E[\hat{X}_i] \\
\text{Var}[\hat{X}] = \sum_i w_i^2 \cdot \text{Var}[\hat{X}_i] \\
\tag{3}
\]

where \( w_i \) is defined as \( w_i = \text{size}(R_i)/\text{size}(D) \) and it denotes the size of region \( R_i \) relative to the size of the domain \( D \). Since the strata constitute a partition of \( D \), \( \sum_i w_i = 1 \). The expected values and variance of the estimators \( \hat{X}_i \) are obtained with Equation (2).

It has been shown that the variance of an estimator obtained through stratified sampling cannot be worse than the variance obtained running Hit-or-Miss Monte Carlo integration on the entire domain, although only a suitable choice of strata provides significant benefits [29]. Optimal stratification criteria can be defined for each problem, although this usually requires additional information about the problem, which is not always available or easy to estimate [19].

In our case, the set of constraints composing a PC are assumed to be formalized as the conjunction of mathematical inequalities, which may involve also non-linear analytical functions. By definition, the indicator function of Equation (1) evaluates to 1 iff the input satisfies the path constraint, that is, if it belongs to the locus of the solutions of the system of mathematical inequalities composing it. The exact computation of these solutions is usually infeasible, but interval constraint propagation can help to identify a set of sub-regions of the domain containing them. For this paper, we use the tool RealPaver [14], which provides interval constraint propagation for systems of non-linear inequalities over real domains. Given a set of inequalities and a bounded domain, RealPaver identifies a set of non-overlapping boxes whose union contain all the solutions of the problem. The boxes may be tight, meaning they only contain solutions, or loose, containing both solutions and other points. Since no solution exists outside those boxes, there is no need to analyze such region through hit-or-miss Monte Carlo, since we already know that the estimator would converge to 0, with variance 0, since the integral of Equation (1) evaluates to 0.

**Example.** As an example of the use of Interval Constraint Propagation (ICP) for variance reduction, let us consider the constraint \( x \leq -y \land y \leq x \), where values of variables \( x \) and \( y \) are uniformly distribution over the domain \([-1,1]\). As illustrated in Figure 2, the probability of satisfying the constraints can be easily computed as the ratio between the area of the triangle identified by the constraint and the size of the domain, and it is exactly 1/4. By using hit-or-miss Monte Carlo with 10^4 samples, we obtain mean (i.e., expected value) .2577 and variance .19131. Let’s consider that ICP produces four boxes, as shown in the figure, containing all solutions. Let’s consider we take 2,500 samples within each box, thus overall still 10^4 samples as in the case of the regular (non-stratified) hit-or-miss Monte Carlo approach. Table 1 shows the boxes coordinates for each of the boxes, size, and corresponding estimates. The part of the domain not covered by the four boxes accounts for 3/4 of the domain, and in such region we already know that both the mean and the variance of a hit-or-miss estimator would be 0, because there are no solutions in there. If we
4. Compositional Estimation for Complex Constraints

Simulation techniques are highly sensitive to the number of samples they take—see the number of samples factor “n” in the denominator of Equation 2. Efficiency is therefore essential for these techniques: improved efficiency means one can sample more within the same time budget and therefore reduce the estimate variance.

We propose a compositional approach to efficiently estimate the solution space for complex mathematical constraints, encoding the disjunction of two path constraints in this set denotes one equivalence class of the input space. We therefore decompose the estimation for each individual PC in the set.

4.1 Disjunction of Path Conditions

Consider the set $PC^T = \{PC_1^T, PC_2^T, \ldots, PC_m^T\}$ of the PCs leading to the occurrence of the target event. The elements $PC_i^T$ are pairwise disjoint by construction. Let $\hat{X}^T$ be the estimator of the probability, given an input profile, of satisfying the constraints in $PC_i^T$ (i.e. an estimator of the integral in (1)). Since there is no intersection between any two sets $PC_i^T, PC_j^T$, we can define an estimator $\hat{X}^T$ of the disjunction between $PC_i^T$ and $PC_j^T$ as follows:

\[
\hat{X}^T = \hat{X}^T + \hat{X}^T
\]

Since there is no intersection between any two sets $PC_i^T$ and $PC_j^T$, the expected value of $\hat{X}^T$ can be straightforwardly computed as [26]:

\[
E[\hat{X}^T] = E[\hat{X}^T] + E[\hat{X}^T]
\]

The following theorem gives a bound on the variance for the composed estimator $\hat{X}^T$, based on the variance computed for the individual path conditions:

**Theorem 1.** Let $PC_i^T, PC_j^T \in PC^T$, $\hat{X}^T$ and $\hat{X}^T$ estimators of the probability of satisfying $PC_i^T$ and $PC_j^T$ respectively. Let $\hat{X}^T = \hat{X}^T + \hat{X}^T$, then:

\[
\text{Var}[\hat{X}^T] \leq \text{Var}[\hat{X}^T] + \text{Var}[\hat{X}^T]
\]

**Proof.** Since $\hat{X}^T$ is defined as the sum of two random variables, the following general relation holds [26]:

\[
\text{Var}[\hat{X}^T] = \text{Var}[\hat{X}^T] + \text{Var}[\hat{X}^T] + 2 \cdot \text{Cov}[\hat{X}^T, \hat{X}^T]
\]

where the covariance $\text{Cov}[\hat{X}^T, \hat{X}^T] = E[\hat{X}^T \cdot \hat{X}^T] - E[\hat{X}^T] \cdot E[\hat{X}^T]$. We already know from Equation (1) that $E[\hat{X}^T \cdot \hat{X}^T] = \int D \cdot \mathbb{I}_{PC^T}(x) \cdot \mathbb{I}_{PC^T}(x) \cdot p(x)$ but the product of the indicator functions in the equation above is always zero because, for every input, at most one of the PCs can be satisfied. On the other hand, the probability of satisfying a PC cannot be negative, thus $\text{Cov}(\hat{X}^T, \hat{X}^T) \leq 0$.

**Theorem 1** allows us to process each PC separately (possibly in parallel) and then compose the partial results to estimate the probability of satisfying any of the PCs in $PC^T$. The price for this approach is the need to provide an overestimation of the variance of the composition, although, as will be shown in Section 6, in practical applications the benefits of the variance reduction techniques on analyzing the single disjuncts often overcome the loss due to the conservativeness we took for disjuncts composition.

4.2 Conjunction of Constraints

We now consider the problem of efficiently computing the probability of satisfying an individual path condition PC. For a real application, a path condition could be very large and may include many constraints relating many input variables. We propose to partition these constraints into smaller sets that have in common the input variables, and whose satisfiability can therefore be determined independently from one another. A path condition $PC_i$ is defined by the conjunction of several simpler constraints $c_{i1} \land \ldots \land c_{im}$. Each constraint $c_d$ can make an assertion about one or more input variables $v_j$, or functions thereof. Let indicate it by $\text{A}(c_{dk}, v_j)$.

Let us introduce a dependency relation (Dep) among the variables appearing in a program:

**Definition 1.** Given the path conditions obtained by the symbolic execution of a program $\{PC\}$, such that $PC = c_{i1} \land c_{i2} \land \ldots \land c_{im}$, and let $V$ be the set of the input variables of the program. The relation $\text{Dep} \subseteq V \times V$ is recursively defined by the following statements:

\[
\forall v \in V : \text{Dep}(v, v)
\]

\[
\forall v, v_j \in V : \exists \text{c}_{ik} \ A(c_{ik}, v_j) \land A(c_{ik}, v) \implies \text{Dep}(v, v_j)
\]

Intuitively, two input variables are depending on each other if they appear together in at least one constraint in the path condition. If this is the case, to evaluate the satisfaction of such conditions the assignments to both variables has to be evaluated at the same time.

Notice that Dep is by construction an equivalence relation over $V$ and for this reason it induces a partition of $V$. Let us refer to the sets composing this partition as $\{V_0, V_1, \ldots, V_j\}$. We can now extend the definition of $A(c, v)$ by stating that $A(c_{ik}, V_j)$ holds if there exists a variable $v_j \in V_j$ such that $A(c_{ik}, v_j)$ holds.

Consider now a path condition $PC_i$. Let us define the constraints $C_{ij} = \bigwedge_{v \in A(v_j, V_j)} c_{ik}$. That is, $C_{ij}$ is the conjunction of all the constraints $c_{ik}$ occurring in $PC_i$ and containing any of the variables from $V_j \subseteq V$. The probability of satisfying $PC_i$ is the same as the probability of satisfying all the constraints $C_{ij}$ at the same time. But since two constraints $C_{ij}$ and $C_{ik}$ do not share any variable by construction, the satisfaction of one of them is independent from the satisfaction of the other. This independence is straightforwardly

\[1\] As support of this statement, notice that $Pr(C_{ij} | v_k = v_{k1}) = Pr(C_{ij} | v_k = v_{k2})$, where $v_k \notin V_j$ and $v_{k1}$ and $v_{k2}$ are any two valid assignments for $v_k$. Thus, we can conclude that $Pr(C_{ij} | v_k = v_{k1}) = Pr(C_{ij})$.\]
inherited by the hit-or-miss estimators \( \hat{X}_{ij} \) and \( \hat{X}_{ik} \) of the probability of satisfying \( C_{ij} \) and \( C_{ik} \), respectively. Indeed, the Monte Carlo procedures will generate random assignments for the sets of variables \( V_I \) and \( V_k \) independently.

Thanks to the independence of the estimators \( \hat{X}_{ij} \) and \( \hat{X}_{ik} \) of the probabilities of satisfying the constraints \( C_{ij} \) and \( C_{ik} \), the probability of \( C_{ij} \land C_{ik} \) can be estimated by \( \hat{X}_{jk} = \hat{X}_{ij} \cdot \hat{X}_{ik} \) having [26]:

\[
E[\hat{X}_{jk}] = E[\hat{X}_{ij}] \cdot E[\hat{X}_{ik}] \\
\text{Var}[\hat{X}_{jk}] = E[\hat{X}_{ij}]^2 \cdot \text{Var}[\hat{X}_{ik}] + E[\hat{X}_{ik}]^2 \cdot \text{Var}[\hat{X}_{ij}] + \text{Var}[\hat{X}_{ij}] \cdot \text{Var}[\hat{X}_{ik}]
\]

(7) (8)

4.3 Observation

By applying the composition methods introduced in this section, we may obtain the following benefits. First, we can split the analysis of a large path condition into the analysis of several simpler constraints. Second, each of the simpler constraints will possibly involve only a subset of the input variables, making the generation of the samples faster. Furthermore, reducing the input space to only the variables actually involved in the constraint, would likely result to a better coverage of the input sub-domain, for a fixed number of samples. Finally, we remark that the simpler constraints can be analyzed efficiently using the interval constraint propagation with Monte Carlo stratified sampling described in the previous section.

4.4 Example

The following code snippet mimics a safety monitor for an autonomous navigation system. If the altitude is lower than 9000 meters, the autopilot is allowed to control the position of the vehicle by manipulating the front and head flaps through floating point variables \( \text{headFlap} \) and \( \text{tailFlap} \). These variables range in the interval \([-10, 10]\).

Listing 1. Example of events and complex conditions in Java.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{if} (\text{altitude} &lt;= 9000) { ...</td>
<td></td>
</tr>
</tbody>
</table>
| \quad | \text{Math.sin(headFlap+tailFlap)>0.25) { 
| \quad | \text{callSupervisor();} 
| \quad | ... 
| \quad | } 
| \quad | } 

Finally, Equation (5) enables \( \text{qCORAL} \) to obtain the estimator \( \hat{X} \), with \( E[\hat{X}] = 0.188089 \) and \( \text{Var}[\hat{X}] = 1.64904 \cdot 10^{-6} \). In this example, the exact probability, rounded to the 6th digit is 0.737848.

5. Algorithms and Implementation

This section provides some implementation details of \text{qCORAL}. The input of \text{qCORAL} includes a set of disjoint constraints \( \text{PCs} \), representing the path conditions leading to the occurrence of the target event as obtained from the symbolic execution stage (see Section 3), and a description of the input domain, i.e., a map from floating-point input variables to their domain. The domain of each variable is a closed interval. We used SymbolicPathFinder [28] as our symbolic execution engine.

Algorithm 1 describes the main loop of \text{qCORAL}. It iterates over the input path constraints, processes each one, and combines the partial results as described in Section 4.1. Note that \text{qCORAL} returns an upper bound for the variance of a disjunction of path constraints, as defined in Equation (6). The procedure \text{computeDependencyRelation} returns the partition of the input variables according to Definition 1. To efficiently implement this procedure we rely on the Jung graph library\(^2\). We create an undirected graph, where a node corresponds to an input variable and an edge corresponds to a dependence between two variables. This procedure visits every path condition and when two variables appear in the same constraint it adds an edge between their corresponding nodes. The procedure then computes the weakly connected components of the graph (with complexity at most quadratic in the number of input variables).

Algorithm 2 computes results for one conjunctive clause. The procedure \text{extractRelatedConstraints} goes through all the constraints.

\[^2\text{http://jung.sourceforge.net}\]
Algorithm 2: analyzeConjunction

Input: pc, D, dep  
Output: mean, var

mean ← 1; var ← Nil
for varSet ∈ dep do
    part ← extractRelatedConstraints(pc, varSet)
    if !cache.contains(part) then
        (partMean, partVar) ← stratSampling(part, varSet, D)
        cache.put(part, (partMean, partVar))
    (partMean, partVar) ← cache.get(part)
    mean ← mean + partMean
    if var = Nil then
        var ← partVar
    else
        mean ← mean + partVar + partMean + var + var * partVar
return (mean, var);

Algorithm 3: stratSampling

Input: part, varSet, D  
Output: mean, var

mean ← 0; var ← 0
boxes ← icp(part, varSet, D)
for box ∈ boxes do
    ⟨boxMean, boxVar⟩ ← hitOrMiss(part, varSet, D)
    boxWeight ← size(box)/size(D)
    mean ← mean + boxWeight * boxMean
    var ← var + boxWeight^2 * boxVar
return (mean, var);

6. Evaluation

Our evaluation addresses the following research questions:

• RQ1. What is the accuracy of qCORAL estimates?
• RQ2. How qCORAL compares with the built-in numerical integration routines of Mathematica and with VolComp [30] when handling linear constraints?
• RQ3. How the different features of qCORAL affect accuracy and time when handling complex constraints?

All experiments have been run on an Intel Core i7 920 64-bit machine, with 2.67Ghz and 8GB, running Ubuntu 12.04.

6.1 RQ1: What is the accuracy of qCORAL estimates?

This experiment evaluates how accurate qCORAL is for computing the volume of several geometric figures for which analytical solutions are widely known. We used the symbolic integration routines of Mathematica to obtain the exact volumes of these solids, while for qCORAL we computed them as the fraction of a domain of known size.

Table 2 summarizes the results. We evaluated the accuracy of our approach for 10^3, 10^4, 10^5, and 10^6 as maximum number of samples for the Monte Carlo procedure. We run 30 times each configuration and reported the average value and standard deviation over the population of estimated volumes.

We grouped our subjects in three groups: convex polyhedra, solids of revolution, and intersection of solids. Except for the convex polyhedra cases, all other subjects contain non-linear constraints and mathematical functions, namely exponentiation and square root. For these subjects the compositional approach does not provide benefits, since the three variables characterizing each solid are tightly dependent one another. We will come back to compositionality later in this section.

All the experiments completed within 2 seconds, with the exception of Icosahedron and Rhombicuboctahedron for 10^6 samples (4 and 7 seconds, respectively). Notably, thanks to Interval Constraint Propagation (ICP) and stratified sampling, even with a relatively small number of samples, qCORAL provided a reasonably accurate result for most of the subjects.

Finally, consider the case of Cube: the standard deviation is 0 because, regardless the number of samples, qCORAL was always able to find the exact solution thanks to ICP. Indeed, the real subject a 3D box and RealPaver can exactly identify it, driving the estimation error to zero.

1http://www.wolfram.com/mathematica
Table 2. Microbenchmarks.

<table>
<thead>
<tr>
<th>Assertion</th>
<th>Num. Paths</th>
<th>Num. Ar. Ops.</th>
<th>NIntegrate</th>
<th>VolComp</th>
<th>qCORAL[STRAT,PARTCACHE]-Mk</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARTRIAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>points ≥ 10</td>
<td>442</td>
<td>1,484 (3)</td>
<td>0.1243</td>
<td>3.03</td>
<td>[0.1234, 0.1254]</td>
</tr>
<tr>
<td>points ≥ 10</td>
<td>5</td>
<td>1740</td>
<td>89.26</td>
<td>3.66</td>
<td>[0.0000, 0.0005]</td>
</tr>
<tr>
<td>points ≥ 5</td>
<td>2,260</td>
<td>68,630 (19,125)</td>
<td>0.9350</td>
<td>4.17936</td>
<td>[0.9340, 0.9364]</td>
</tr>
<tr>
<td>CART</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>count ≥ 3</td>
<td>44</td>
<td>1,209</td>
<td>638 (3)</td>
<td>0.9746</td>
<td>7.26</td>
</tr>
<tr>
<td>count ≥ 5</td>
<td>47</td>
<td>1,296</td>
<td>681 (3)</td>
<td>0.9828</td>
<td>7.66</td>
</tr>
<tr>
<td>COROTARY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>temp ≥ 5</td>
<td>120</td>
<td>195</td>
<td>62 (3)</td>
<td>0.0006</td>
<td>3.44</td>
</tr>
<tr>
<td>temp ≥ 5</td>
<td>274</td>
<td>31</td>
<td>8 (3)</td>
<td>0.0001</td>
<td>0.86</td>
</tr>
<tr>
<td>EGFR EPI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f1 - f ≥ 0.1</td>
<td>45</td>
<td>547</td>
<td>31 (3)</td>
<td>0.1264</td>
<td>1.98</td>
</tr>
<tr>
<td>f1 - f ≥ 0.1</td>
<td>44</td>
<td>422</td>
<td>33 (2)</td>
<td>0.0986</td>
<td>1.69</td>
</tr>
<tr>
<td>EGFR EPI (SIMPLE)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f1 ≥ 4.6 &amp; f ≤ 4.6</td>
<td>13</td>
<td>163</td>
<td>12 (3)</td>
<td>0.5382</td>
<td>0.83</td>
</tr>
<tr>
<td>f1 ≥ 4.6 &amp; f ≤ 4.6</td>
<td>14</td>
<td>101</td>
<td>9 (3)</td>
<td>0.3012</td>
<td>0.65</td>
</tr>
<tr>
<td>INVPEND</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pAvg &lt;= 1</td>
<td>1</td>
<td>54</td>
<td>229 (3)</td>
<td>0.0509</td>
<td>1.15</td>
</tr>
<tr>
<td>PACK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>count ≥ 5</td>
<td>1,103</td>
<td>16,414</td>
<td>0 (0)</td>
<td>0.9546</td>
<td>57.44</td>
</tr>
<tr>
<td>count ≥ 7</td>
<td>906</td>
<td>12,080</td>
<td>0 (0)</td>
<td>0.3898</td>
<td>41.76</td>
</tr>
<tr>
<td>count ≥ 7</td>
<td>924</td>
<td>0</td>
<td>0 (0)</td>
<td>0.1428</td>
<td>45.48</td>
</tr>
<tr>
<td>count ≥ 10</td>
<td>821</td>
<td>840</td>
<td>0 (0)</td>
<td>4.41</td>
<td>[0.0002, 0.0002]</td>
</tr>
<tr>
<td>totalWeight ≥ 6</td>
<td>954</td>
<td>14,850</td>
<td>6,948 (2)</td>
<td>0.2462</td>
<td>5066.20</td>
</tr>
<tr>
<td>totalWeight ≥ 4</td>
<td>1,030</td>
<td>16,186</td>
<td>7,578 (2)</td>
<td>0.6771</td>
<td>70.16</td>
</tr>
<tr>
<td>totalWeight ≥ 4</td>
<td>1,132</td>
<td>17,972</td>
<td>8,420 (2)</td>
<td>0.9592</td>
<td>54.48</td>
</tr>
<tr>
<td>VOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>count ≥ 20</td>
<td>24</td>
<td>13,824</td>
<td>882,508 (3)</td>
<td>1.0005</td>
<td>1,245.30</td>
</tr>
</tbody>
</table>

Table 3. Comparison of NIntegrate (default numerical integration method from Mathematica [1]), VolComp, and qCORAL with features STRAT and PARTCACHE enabled. Note that the comparison is restricted to linear constraints and the comparison metrics are not the same.

6.2 RQ2: How qCORAL compares with Mathematica and VolComp?

This experiment evaluates how qCORAL (with both stratified sampling and the compositional analysis) compares with existing techniques for quantifying solution spaces of linear constraints. As baseline for comparison, we use the built-in procedure for numerical integration of the commercial tool Mathematica, NIntegrate. This procedure is available off-the-shelf and we run it with its default settings. We also include VolComp, a recently developed tool producing as output a tight closed interval over the real numbers containing the requested solution.

The built-in numerical integration routine of Mathematica performs a recursive procedure called Global Adaptive Integration [21]. On each recursive call, it analyzes the behavior of the integrand function within the integration domain and automatically selects an integration rule to apply. Mathematica supports a broad variety of integration rules, e.g. Trapezoidal, Simpson, or Newton [1]. After each iteration, the procedure analyzes the obtained results and if they do not meet the accuracy goals, the integration domain is bisected and the integration routine is invoked on each of the two parts. Partial results are then combined, similarly to the case of stratified sampling. The procedure terminates when the default accuracy requirements are met or when the recursion depth limit has been reached. Since NIntegrate guarantees the accuracy of the result (or gives a notification when its requirements are not met), it provides a reference to evaluate how tight are the intervals produced by VolComp and how precise are the estimates produced by qCORAL.

The techniques we evaluate in this experiment report solutions in different formats. Mathematica reports a single point solution,
which is exact up to an informed bound. VolComp reports an interval solution that bounds the exact point solution, and qCORAL reports an estimate on the exact solution and a statistical variance for that estimate. In place of the variance, for qCORAL we report the standard deviation (square root of the variance), which is in the same unit scale of the estimate. Furthermore, since qCORAL implements a randomized algorithm, the estimate and the standard deviation we report are averaged over 30 executions.

To compare the tools, we used the VolComp benchmark, which is publicly available [2]. The subjects we selected from the benchmark are: the Framingham atrial fibrillation risk calculator (ATRIAL), a steering controller to deal with wind disturbances (CART), the Framingham coronary risk calculator (CORONARY), an eGFR estimator of chronic kidney’s disease (EGFR EPI and EGFR EPI SIMPLE), an inverted pendulum (INVPEND), a model of a robot that packs objects of different weights in a carton with the goal to maximize total weight within a certain limit (PACK), and a controller for filling up a tank with fluid (VOL).

We re-ran the benchmarks from [2] using their scripts. The PCs were obtained by translating PCs produced with their frontend. All experiments ran until completion. Table 3 shows the comparison between qCORAL and VolComp [2]. Notice that the two techniques provide different types of results. VolComp uses iterative bounding to return an interval containing the solution. qCORAL returns a statistical estimator characterized by its expected value and variance: the expected value is the most likely value as it comes out from qCORAL analysis; the variance provides a measure of the uncertainty of the estimate; such uncertainty could be used to quantify the probability the real value belongs to an interval, for example by using Chebyshev’s inequality [26].

The first 4 columns of Table 3 state the targeted assertion, how many paths reach the assertion true, how many conjuncts (and) are there on these paths and lastly how many arithmetic operations (with the number of unique operations). For each tool we then show the estimated probability of satisfying the assertions specified in the first column. We make the following observations from these results:

- **ATRIAL.** For one of the three assertions Mathematica takes very long to produce results and VolComp takes more than 10m to finish. qCORAL produces an accurate result very quickly.
- **CART.** This subject produces a highly skewed polynomial [30], which is known to be a case where branch-and-bound techniques do not perform well. Indeed, results of VolComp are not good for this case. qCORAL also suffers as it relies on RealPaver which also uses branch-and-bound techniques. For this reason, we observe a high standard deviation in our results relative to other subjects. Note that despite this limitation qCORAL can still report an estimate close to the exact solution. Mathematica can easily handle the subject in this case.
- **CORONARY and EGFR.** These are the best cases for qCORAL. As compared to other tools, qCORAL reports results very quickly and precisely.
- **PACK.** qCORAL is slower for the last two cases of PACK. The reason for this high cost is the high number of paths and the high interdependence among the variables on each path constraint, which reduces the impact of our divide and conquer strategy for conjuncts. We observed that for one case in this subject Mathematica actually misses the interval; this is highlighted in grey color. This happened because the default settings of \(^\text{NIntegrate}\) does not allow to collect enough points for the numerical integration to converge. This situation is reported by the tool. Note that on some cases rounding the result to four decimal digits made the results of VolComp be different than those of Mathematica and qCORAL by the last digit. These have to be considered just rounding error.

- **VOL** is a case that stresses qCORAL. Mathematica finishes in more than 20m and VolComp returns the full range 0.0-1.0, i.e. it did not perform any pruning. qCORAL reports the expected result of \(\sim 1.0\). Note that the estimate that qCORAL reports is actually slightly greater than 1, for this case. This is due to the propagation of errors in the estimation which magnifies when the exact probability is close to the corner case 1. Mathematica also reports a result > 1 due to the finite accuracy of numerical integration, partially due to the use of off-the-shelf settings for such complex subject.

In summary, we observed that qCORAL reports estimates very close to the exact point solutions that Mathematica’s numerical integration method reports (apart from the case of PACK where Mathematica reported no convergence and a wrong). However, numerical integration is potentially expensive when the number of variables grows [27]. This observation is confirmed for the cases of ATRIAL and PACK, where Mathematica takes respectively more than 1h to complete. In most of the experiments, especially those on complex subjects, qCORAL resulted faster than both off-the-shelf use of Mathematica’s numerical integration function and VolComp. For specific problems advanced settings of Mathematica may improve its performance, though a deeper understanding of the mathematical nature of the problem might be required, which might be not straightforwardly derived from the code under analysis. We also notice that the results of qCORAL have been consistent with those of VolComp: the qCORAL estimates fall within the corresponding VolComp intervals almost always, up to the accuracy (last decimal digit might be dangling because of rounding); the only exception is \(f_1, f_2 \geq 0.1\), where the variance is anyway large enough to account for the deviation (3.29 \(10^{-4}\)). Concluding, we observe that qCORAL has provided a reasonable balance between time efficiency and precision when handling linear constraints on this benchmark suite.

### 6.3 RQ3: How the different features of qCORAL affect accuracy and time when handling complex constraints?

This experiment evaluates how different configurations of qCORAL compare with respect to precision and time. We considered two implementations of Monte Carlo Hit-or-Miss: one from Mathematica (baseline) and one from qCORAL: qCORAL{}{}. The empty braces indicate that no feature from qCORAL has been enabled in this configuration. The configuration qCORAL[\text{STRAT}] incorporates stratified sampling in the analysis of individual path conditions. The configuration qCORAL[\text{STRAT}, PARTCACHE] cumulatively incorporates partitioning and caching. We considered the following subjects from the aerospace domain in this experiment:

- **Apollo.** The Apollo Lunar Autopilot is a Simulink model that was automatically translated to Java using the Vanderbilt toolset [23]. The model is available from MathWorks \(^6\). It contains both Simulink blocks and Stateflow diagrams and makes use of complex Math functions (e.g. Math.sqrt). The code contains 2.6KLOC, deployed in a single package with 54 classes. We analyzed 5,779 path constraints for this subject.
- **TSAFE.** The Tactical Separation Assisted Flight Environment (TSAFE) is designed to prevent near misses and collisions of

---

\(^4\)http://www.framinghamheartstudy.org

\(^3\)http://nephron.com/epi_equation

\(^6\)http://www.mathworks.com/products/simulink/demos.html
aircraft that are predicted to happen in the near future (from 30s to 3m). The Conflict Probe module of TSAFE tests for conflicts between a pair of aircraft within a safe time horizon. The two aircraft may either be flying level or engaged in turns of constant radius. The following math functions appear in the path constraints of this subject: cos, pow, sin, sqrt, and tan. The Turn Logic module of TSAFE computes the change in heading required once an impending loss of separation between two aircraft is detected. It assumes a constant turning radius for the aircraft making the maneuver. Path constraints for this subject contain the atan2 function. Each of these modules is about 50 LOC. We analyzed respectively 23 path constraints on Conflict and 225 path constraints on Turn Logic.

These case studies contain complex constraints. They have 2.6KLOC for Apollo and 50 each for the two TSAFE components analyzed. Since no properties were defined, instead of fabricating a property, we generated all the PCs (using Symbolic PathFinder with search bound equal to 50) and selected a percentage of PCs to quantify the path probabilities. We generated 5779 PCs for Apollo and 225 for TSAFE; the latter is smaller but involves complex mathematical functions and has high dependence among variables. We arbitrarily picked the first 70% of the PCs in a bounded depth-first order. We picked 70% of the paths so to avoid obtaining a probability close to 0 or 1, to not bias the evaluation for the Monte Carlo estimation towards its worst cases (some extreme cases have already been evaluated in Table 3). This selection mimics a property that is satisfied on some of the paths and not on the rest.

Table 4 shows experimental results. Note the reported estimations are not exactly 70% because different PCs have different solution spaces. Different columns show different configurations of qCORAL, each line denotes one different subject, and each group of lines denotes a different maximum number of samples allowed for simulation in qCORAL. Results show that the addition of feature STRAT results in a significant reduction in the variance of the estimate at the expense of an overhead in time. This additional cost is justified by the multiple invocations of the ICP solver; one call per disjunctive clause of the input formula. Adding the feature PARTCACHE, i.e., performing compositional analysis, may further improve precision (see Section 4), however, considering the most time-consuming cases, it always reduces time. Increasing the maximum sample size from 1K to 100K results in higher time savings as the cost for analyzing each subproblem increases. While the relationship between execution time and maximum number of samples is approximatively linear for the basic configuration, the impact of stratified sampling and compositionality on execution time depends on the specific subject.

7. Related Work

There are many related works to ours, including probabilistic abstract interpretation [22], probabilistic model checking [16] and volume computations [9]. A comprehensive list of related work that is relevant here too can be found in [30]. Here we focus on the works we consider the most closely related.

We calculate the probability of a path, or more generally a set of paths, being executed. Typically these paths lead to an event of interest, such as an assertion violation for example. Techniques for doing such probabilistic analyses differ in the type of input distributions they consider, the language features supported, and the approach used to calculate the number of solutions.

One approach in this area was that of Geldenhuys et al. [12] that considered uniform distributions for the inputs, linear integer arithmetic constraints, and used LatTE Machiato [9] to count solutions of path conditions produced during symbolic execution. Their goal was to calculate the probability of covering branches and assertion violations in the code. One of the main technical differences between this work and others based on symbolic execution (including the present paper) is that probabilities are calculated at each branching point rather than after symbolic execution is finished. Here we support complex constraints, including non-linear constraints, and we use a statistical approach to count solutions. Our proposed approach is compositional and hence we believe it is applicable in the incremental setting from [12].

Sankaranarayanan et al. [30] and Filieri et al. [11] recently proposed similar techniques to compute probabilities of violating state assertions. Both techniques remove the restriction of uniform distributions, although in the latter case it is by discretizing the domain into small uniform regions. As with [12] both approaches only consider linear constraints. Their techniques also build on symbolic execution to compute a relevant set of symbolic paths of a program that leads execution to some assertion of interest. Both techniques estimate probability of exercising those paths and violating the corresponding assertion from such a set. Sankaranarayanan et al. developed a customized algorithm for under and over-approximations of probabilities. They use Linear Programming (LP) solvers to compute over-approximations and heuristics-based “ray-shooting” algorithms to compute under-approximations, which is applicable for convex polyhedra. Filieri et al. used the LatTE Machiato [9] tool to compute probabilities. Our technique complements the works of Sankaranarayanan et al. and Filieri et al. by supporting a wider range of constraints. Experimental results show that although there is a potential loss in precision by using our analysis, when that loss occurs it is small.

Bouissou et al. [5] handles non-linear constraints with a combination of abstraction based on affine and p-box arithmetic. The approach relies on the use of noise variables to represent the un-
certainty of non-linear computations. However they cannot handle conditional branches as we can here. More recently Adje et al. [3] extended this work to allow conditional statements as well. The main difference with our approach lies in that they use an abstraction based approach whereas we use a statistical approach. We can thus handle a wider set of non-linear constraints such as complex mathematical functions (sine, cosine, etc.). In future work we would like to do an empirical comparison between these two approaches on the examples that both can handle.

Pavese et al. [25] propose the use Monte Carlo simulation and model inference [10] to optimize probabilistic model checkers, such as PRISM [16]. Conceptually, the idea is to obtain a symbolic computational model from a set of traces obtained with sampling and feed that model to the model checker. Experimental results show that their technique can obtain tighter lower bounds for the mean time to first failure (MTTF) of network protocol models with rare failures. Our work is orthogonal to theirs. Our technique applies to imperative programs, such as those for controlling vehicles while their technique applies to probabilistic programs. Our technique expects as input a summary of the behavior characterized with a set of complex constraints.

8. Conclusion
We have developed a compositional statistical approach for the quantification of solution spaces for arbitrary complex mathematical constraints with application to probabilistic software analysis. The approach is enhanced by an acceleration procedure to improve the accuracy of the simulation based analysis with a limited number of additional experiments. Our initial experimental results are promising. In principle, our statistical approach can be applied to both integer and floating-point domains, but in practice it performs poorly for integer constraints if compared to counting-based techniques (particularly problematic are the equality constraints). Our future work will therefore focus on handling mixed integer and floating-point constraints, possibly through a combination with exact volume computations or guidance heuristics. We also plan to use the compositional approach in an incremental symbolic execution setting and to combine the technique with a statistical (as opposed to systematic) exploration of program paths as guided by the estimated conditional probabilities for the conditions in the code.

Acknowledgments
This work is partly supported by NSF Awards CCF-1329278 and CCF-1319858 (Pääräneu). Mateus Borges is supported by FACEPE graduate fellowship number IBPG-0668-1/03/12.

References