WIP: Language-Agnostic Static Analysis of Probabilistic Programs

Markus Böck
TU Wien
Vienna, Austria
markus.h.boeck@tuwien.ac.at

Michael Schröder
TU Wien
Vienna, Austria
michael.schroeder@tuwien.ac.at

Jürgen Cito
TU Wien
Vienna, Austria
juergen.cito@tuwien.ac.at

***Abstract*—Probabilistic programming allows developers to focus on the modeling aspect in the Bayesian workflow by abstracting away the posterior inference machinery. In practice, however, programming errors specific to the probabilistic environment are hard to fix without deep knowledge of the underlying systems. Like in classical software engineering, static program analysis methods could be employed to catch many of these errors. In this work, we present the first framework to formulate static analyses for probabilistic programs in a language-agnostic manner: LASAPP. While prior work focused on specific languages, all analyses written with our framework can be readily applied to new languages by adding easy-to-implement API bindings. Our prototype supports five popular probabilistic programming languages out-of-the-box. We demonstrate the effectiveness and expressiveness of the LASAPP framework by presenting four sound language-agnostic probabilistic program analyses that address problems discussed in the literature.**

***Index Terms*—probabilistic programming, program analysis, language-agnostic**

I. Introduction

In Bayesian inference, we have a set of latent variables, observed data, and a statistical model which typically encodes how the observations are believed to be generated from the latents. Domain knowledge about the latents can be incorporated by specifying informative prior distributions. Posterior inference is concerned with finding or approximating the posterior probability distribution of the model—the distribution over the latent variables given the observed data. Probabilistic programming systems provide an intuitive means to specify Bayesian models as simple programs and to automatically perform posterior inference on them. These systems decouple modeling from inference by implementing general-purpose inference algorithms. While previously the analysis of complex Bayesian models was reserved for experts, probabilistic programming opens it up for users without substantial knowledge of statistics and inference techniques.

While writing probabilistic programs seems intuitive, it also leads to many counter-intuitive situations. For instance, a programmer modeling a generative process may think about it sequentially, but the execution order of sample statements during inference need not be sequential and in general depends on the applied algorithm. Also, due to the stochastic nature of probabilistic programming, bugs may not occur during every run of the program and could be hard to reproduce. Furthermore, probabilistic programming systems abstract away thousands of lines of code and one may have to trace bugs deep into the inference libraries. For these reasons, debugging probabilistic programs still requires extensive knowledge of inference algorithms and the system at hand.

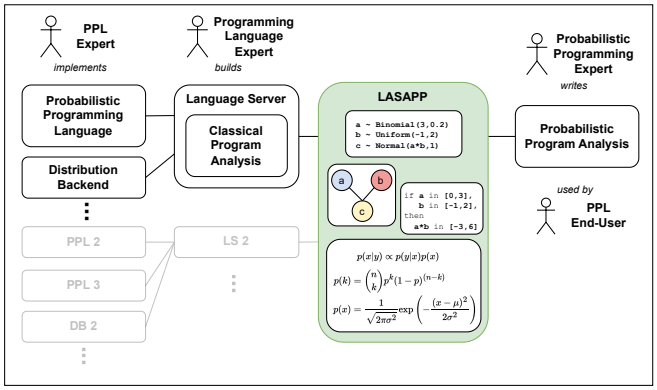


Fig. 1. Overview of the LASAPP framework: Analyses are formulated in a high-level language-agnostic API. Internally, a language server performs classic analyses like data and control flow analysis. Servers have to be built only once per host language and facilitates easy-to-implement bindings for multiple embedded probabilistic programming languages.

Program analysis can serve as a powerful technique to catch bugs statically and to improve the usability of probabilistic programming languages in general. In fact, recent research realizing this idea includes SlicStan, which relaxes the syntax of Stan [1], automatic guide generation for Pyro [2], improved debugging with probabilistic assertions [3], and slicing of probabilistic programs for improved inference [4]. However, a shared attribute of all of these methods is their limitation to a single probabilistic programming language.

In this paper, we present LASAPP (Figure 1)—the first approach to writing static analyses for probabilistic programs in a language-agnostic way. The key insight lies in the fact that while probabilistic programming languages differ in their implementation and expressiveness, the underlying meaning of these programs remains the same: a probabilistic model. By using existing program analysis techniques, like data and control flow analysis, we can put an abstraction layer on top of classical program properties and formulate analyses of probabilistic properties in the language of these abstractions at a high level. Our approach makes it easy to add support for

new languages and thus existing analyses written in LASAPP are immediately available for them. Furthermore, the clear separation of classical program analyses entails that one can write probabilistic analysis without having to be a programming languages expert. We demonstrate that our framework allows the formulation of many practical analyses that greatly improve the usability of a multitude of probabilistic programming languages at once.

The main contributions of this work are:

- The first language-agnostic static analysis framework for probabilistic programs (LASAPP).
- Four language-agnostic program analyses for improved usability of probabilistic programming systems:
  - Statistical Dependency Analysis (Section IV-B)
  - HMC Assumptions Checker (Section IV-C)
  - Parameter Constraint Verifier (Section IV-D)
  - Model-Guide Validation (Section IV-E)
- Correctness proofs of these analyses for a formal PPL,
- Prototype LASAPP bindings for the popular languages Pyro [5], PyMC [6], BeanMachine [7], Turing [8], and Gen [9], available at https://github.com/lasapp/lasapp.git,
- Evaluation of the practicability of the analyses for real-world probabilistic programs.

II. Motivation

A. Probabilistic Programming

Probabilistic programming is an intuitive means to specify Bayesian models as programs. We start by giving a general definition:

Definition 1: A *probabilistic program* is a (non-deterministic) program, for which an (unnormalized) weight can be assigned to each possible execution.

If the program returns a number, we can think of it as an (unnormalized) probability distribution over the return values. However, in probabilistic programming the local random variables of a program are also of great interest as they typically encode how observed data is believed to be generated. For each run of the program, the values of these variables are captured in an *execution trace*. Thus, probabilistic programs can also be viewed as an (unnormalized) joint probability distribution over all random variables, or as an (unnormalized) probability distribution over execution traces [10]–[14]. This distribution may also be referred to as *program density*.

In most probabilistic programming languages (PPLs), random variables of a model and their prior distributions are declared with special syntax in the form of *sample statements*. Relating these variables with program constructs implicitly defines the likelihood of the model. Furthermore, in probabilistic programming systems, there is a mechanism to constrain the values of random variables to observed data. This results in a probabilistic model *conditioned* on data. All variables that are not constrained to any observed values are called *latent variables*.

In Figure 2, you can see a classic example of a probabilistic model, namely a linear model, where the univariate data y is believed to have a noisy linear relationship with the input values x. The variables y_i are constrained to observed data and we are interested in the posterior distribution over the latents `a`, `b`, and `s2`—the distribution over the slope, intercept, and noise parameter given the data. We can inspect the posterior distribution to find the line that best fits the data and quantify the uncertainty about our conclusions.

```
def linear_model(x, y):
  a = pyro.sample("a", dist.Normal(0,10))
  b = pyro.sample("b", dist.Normal(0,10))
  s2 = pyro.sample("s2", dist.InverseGamma(1,1))
  for i in range(len(x)):
    pyro.sample(f"y_{i}", dist.Normal(a*x[i]+b,s2), obs=y[i])
```

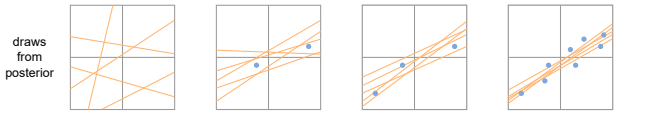


Fig. 2. Linear model implemented in Pyro on the top. Results from posterior inference with decreasing uncertainty on the bottom.

What makes probabilistic programming special is that across different languages the semantic meaning of the programs remains the same. The languages differ in syntax and implementations, making trade-offs between the efficiency of inference algorithms and generality in terms of which models are expressible. However, the programs written in any language can all be interpreted as mathematical probabilistic models. Thus, the same probabilistic program analysis is meaningful in several languages and it is desirable to be able to apply the analysis directly without having to build up an analysis framework for each language.

B. Static Analysis of Probabilistic Programs

Consider a simple Bayesian model where a system can be in two states—5 or 6. Noisy measurements X of the state are available and with posterior inference we wish to find the most probable state of the system given the measurements and quantify the uncertainty. In Figure 3, you can see implementations of this model in Turing, BeanMachine, and Pyro.

Even though Hamiltonian Monte Carlo (HMC) [15] is a very robust and frequently recommended inference algorithm, it would be ill-advised to apply it for the described model. In fact, for all of the three implementations, it leads to cryptic runtime errors which need to be traced deep into the inference library. The trained eye will immediately spot the declarations of the discrete random variable `state` and find it incompatible with the gradient-based inference algorithm. However, we argue that this is a non-trivial bug, because the error messages provide little help, it requires knowledge of the assumptions of the inference algorithm to fix, and the same source code would be correct for different algorithms.

Having observed the differences in syntax and disregarding the implementation details of HMC, we note that the underlying root cause of the runtime error is the same for all implementations. Even more, the bug can be caught *statically*. However, implementing such a static analysis separately for

```
@model function model(X)
  state ~ Categorical([0.5, 0.5])
  if state == 1       ⇒ ERROR: InexactError:
    mu = 5.             Int64(0.13392275765318448)
  else
    mu = 6.
  end
  for i in eachindex(X)
    X[i] ~ Normal(mu, 1.)
  end
end
```

```
@bm.random_variable
def state():
  return dist.Categorical(torch.tensor([0.5,0.5]))
@bm.random_variable          ⇒ RuntimeError: only
def model(n):                  Tensors of a floating
  if state() == 1:             point and complex dtype
    mu = 5.                    can require gradients
  else:
    mu = 6.
  return dist.Normal(mu * torch.ones(n), 1.)
```

```
def model(X):
  state = pyro.sample("state", dist.Categorical([0.5,0.5]))
  if state == 1:                ⇒ RuntimeError: Boolean
    mu = 5.                       value of a Tensor with
  else:                           more than one value is
    mu = 6.                       ambiguous.
  pyro.sample("X",
    dist.Normal(mu * torch.ones(len(X)), 1.), obs=X)
```

Fig. 3. Simple model implemented in Turing, BeanMachine, and Pyro (top to bottom). Applying the HMC algorithm to these models results in different cryptic bugs. The root cause is identical for all of them: the declaration of the discrete variable `state` is not supported by HMC which requires continuous variables.

each PPL would require a separate parser for each language's unique syntax, extracting and traversing different abstract syntax trees (ASTs) to find the respective sample statements, extracting the underlying distribution and looking up its properties based on the PPL's particular semantics, and finally raising a warning if the type is discrete.

In this paper, we propose a framework where we abstract away much of the classical program analysis details like traversing the AST and parsing sample statements. As a result, the same static analysis can to be written at a higher level in just three steps:

1) Find all random variable declarations.
2) Iterate over all random variables.
3) If a variable has a discrete distribution, raise a warning.

In terms of the concrete LASAPP API, these high-level steps translate almost one-to-one to Python code:

```
import lasapp
program = lasapp.ProbabilisticProgram(path)
random_vars = program.get_random_variables()          # (1)
for rv in random_vars:                                # (2)
  props = lasapp.infer_distribution_properties(rv)
  if props.is_discrete():                             # (3)
    raise Warning("...")
```

What makes our approach powerful is that adding support for new languages is easy and all existing program analysis written in the framework are immediately available. The above analysis is written in a completely language-agnostic way and can be applied to prevent all bugs from Figure 3. In fact, it also prevents two issues reported by developers in the Turing and Pyro forum, which where the inspiration for the discussed examples [16], [17].

III. LASAPP

Figure 1 presents an overview of LASAPP, our framework for **L**anguage-**A**gnostic **S**tatic **A**nalysis for **P**robabilistic **P**rogramming. LASAPP targets PPLs that are embedded in a host language, like Pyro (which is embedded in Python), or Turing (embedded in Julia). Focusing on these types of PPLs allows us to separate probabilistic analysis from the classical program analysis machinery. LASAPP makes it possible for PPL experts to write high-level probabilistic program analyses in a language-agnostic way.

For each host language, a programming language expert implements a *language server*[1], which is responsible for parsing source code and representing it as an abstract syntax tree (AST), performing data- and control-flow analysis, abstract interpretation, or symbolic execution. Language servers need to be implemented only once per host language, and can be built using a plethora of well-established methods (Section III-B). Once a host language is supported, adding support for new PPLs embedded in that language is simple. Mostly, one has to write bindings for a PPL's special sample syntax (Section III-A1) and for its probabilistic distribution back-end (Section III-A2). For our prototype, we implemented language servers for Python and Julia, and added support for five PPLs, including two shared probability distribution back-ends (Table I).

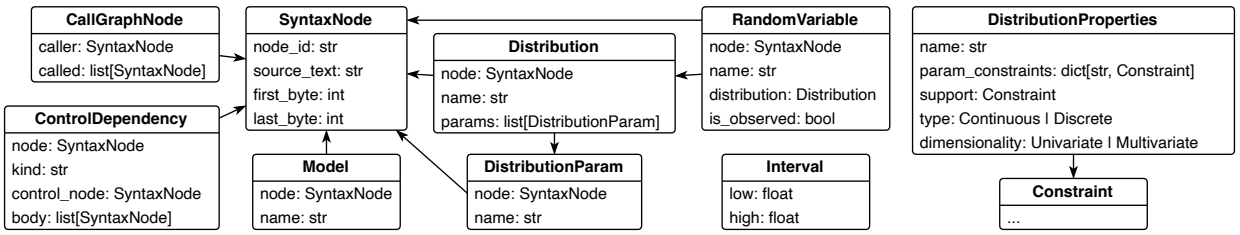
TABLE I
LINES OF CODE NEEDED TO ADD LASAPP SUPPORT FOR VARIOUS PPLS AND PROBABILITY DISTRIBUTION BACK-ENDS.

	Back-end / PPL	LOC	
Python	Language Server	1131	
	PyMC [6]	153	custom back-end
	torch.distributions [18]	68	shared back-end
	Pyro [5]	63	
	BeanMachine [7]	134	
Julia	Language Server	1822	
	Distributions.jl [19]	110	shared back-end
	Gen [9]	192	
	Turing [8]	74	

A. Probabilistic Analysis API

LASAPP provides a Python API to implement static analyses of probabilistic programs. Figure 4 shows the main data types and most important API methods. We arrived at this API design by considering the common characteristics of popular PPLs to find a shared set of universal abstractions that would serve the needs of real-world probabilistic analyses (see Section IV).

[1]Not to be confused with the Language Server Protocol (LSP), which enables the use of static analysis tools independent of programming environment. One could use LSP to integrate LASAPP in an IDE. In contrast to LSP, LASAPP's API facilitates the implementation of static analyses independent of probabilistic programming language.



class ProbabilisticProgram(*filename*: str)

def get_model() → Model	# see Section III-A1
def get_random_variables() → list[RandomVariable]	# see Section III-A1, see Suppl. ??
def get_data_dependencies(*node*: SyntaxNode) → list[SyntaxNode]	# see Section III-B1
def get_control_dependencies(*node*: SyntaxNode) → list[ControlDependency]	# see Section III-B1
def estimate_value_range(*expr*: SyntaxNode, *assumptions*: dict[RandomVariable, Interval]) → Interval	# see Section III-B2
def get_path_condition(*node*: SyntaxNode, *root*: SyntaxNode, *assumptions*: dict[SyntaxNode, SExpr]) → SExpr	# see Section III-B3
def infer_properties(*rv* : RandomVariable) → DistributionProperties	# see Section III-A2

Fig. 4. Main types and methods of the LASAPP API.

1) Sample Statements: Different PPLs use different syntax to declare random variables. In these sample statements, we require a user-defined random variable name and an expression corresponding to the distribution of the variable. The variable name is also referred to as *address*—a (dynamically) unique string, with which the value of a variable is stored in the execution trace.

We give a few examples of different sample syntaxes, where we highlight the addresses, distributions, and special syntax:

- Turing has a special syntax which includes a tilde character, which is resolved with a macro before compilation. Values are assigned to program variables denoted on the left-hand side, which also serve as addresses:

```
mu ~ Normal(0., 1.)
x[i] ~ Normal(mu, 1.)
```

- In BeanMachine, a special decorator is added to function definitions to declare a random variable, which can then be referenced from anywhere with a regular function call. The above example can be rewritten in BeanMachine as:

```
@bm.random_variable
def mu() :
  return dist.Normal(0., 1.)
@bm.random_variable
def x(i) :
  return dist.Normal(mu(), 1.)
```

- In Pyro, sample statements are calls to the `pyro.sample` method, with addresses and distributions passed as arguments. In Pyro the example becomes:

```
mu = pyro.sample( "mu" , dist.Normal(0., 1.) )
x[i] = pyro.sample( f"x_{i}" , dist.Normal(mu, 1.) )
```

LASAPP abstracts all of these into the same `RandomVariable` data type, preserving the semantics of the original syntaxes.

In this work, we only consider languages where sample statements can be extracted statically from the source code. To be able to perform classical static analyses, like data flow analysis, we also require that random variables are always bound to program identifiers, e.g., variables or functions.

In addition to its distribution, it is also interesting to know whether a random variable is observed or not. Mechanism to declare an observed variable range from adding an optional "observed" parameter to the sample call, passing the observed value as a function parameter, or collecting the observations in a hash-map. For the sake of brevity we refer to the prototype implementation and PPL documentations for more details.

Lastly, probabilistic programs usually have a model entry-point or a program construct which encapsulates the sample statements. For instance, in Pyro and Turing a program function defines a model while in PyMC sample statements are listed in the body of a `with` statement.

2) Distribution Back-ends: Probabilistic programming systems either have their own implementation of common probability distributions or use a package of their host language. For instance, PyMC has its own distribution library, while Pyro and BeanMachine share the PyTorch `torch.distributions` back-end [18] and Turing and Gen share the `Distributions.jl` back-end [19]. Sharing back-ends between PPLs allows for further modularisation and reusability but introduces an additional layer of variation.

LASAPP needs to map back-end-specific distribution and parameter names to a common vocabulary. For instance, consider the `Gamma` object in the `Distributions.jl` package, where the default parameters map to a parametrization in

terms of shape and rate:

$$\texttt{Gamma}(\overset{\alpha}{1.5}, \overset{\beta}{2.0}) \quad \rightsquigarrow \quad p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\left(-\beta x\right)$$

In PyTorch, however, the `Gamma` distribution is expressed in terms of shape and scale:

$$\texttt{Gamma}(\overset{k}{1.5}, \overset{\theta}{2.0}) \quad \rightsquigarrow \quad p(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} \exp\left(-\frac{x}{\theta}\right)$$

LASAPP represents distributions with unique names for both parameters and the distribution itself.

Note that a distribution expression in a sample statement need not be a single distribution object, but could be an arbitrary expression evaluating to some distribution at runtime. LASAPP currently maps such expressions to an "Unknown" distribution type. In future work, we plan on parsing these arbitrary expression to support more distribution classes, such as mixture distributions.

B. Classical Analysis Foundation

LASAPP's probabilistic analysis API is built on a foundation of classical static analysis methods. This layered approach allows PPL experts to focus on implementing high-level language-agnostic analyses of probabilistic properties, while programming language experts can provide the necessary classical analysis foundations for each host language.

While all the classical techniques we discuss in this section are of general importance to PPL analysis, they were specifically chosen to aid in the implementation of the probabilistic analyses we present in Section IV. Other probabilistic analyses may require the support of additional classical techniques. By virtue of its modular design, LASAPP can be extended to incorporate additional static analysis methods in its language servers and expose them via the high-level API, accommodating the needs of new probabilistic analyses as they arise.

1) Data/Control Flow Analysis: Data and control flow analysis is concerned with finding the relationship between program variables and the order in which program statements are executed. We begin by informally defining data dependencies following [20].

Definition 2 (Data Dependency): A statement S_1 is *data dependent* on statement S_2, if S_2 produces a result that may be used by S_1.

This definition is related to the notion of *reaching definitions* [21]. Note that if statement S_2 is data dependent on statement S_3, then S_1 is also data dependent on S_3 by transitive closure.

Next, we give a general definition of a control-flow graph [21]. For a more formal definition, see for instance [22].

Definition 3 (Control-Flow Graph): A *control-flow graph* is a directed graph that connects basic blocks of a program (sequence of operations always executed together) and encodes all paths that might be traversed through a program during its execution.

Data dependencies and the control flow are particularly important in the analysis of probabilistic programs, because they are used to find dependencies between sample statements. In fact, knowing the dependencies of random variables allows us to represent programs as *Bayesian networks* or more generally as *probabilistic graphical models*. Graphical models are well-understood mathematical objects [23] which allows us to leverage existing methods to reason about probabilistic programs. In Section IV-B, we will show how to statically extract the model graph of a probabilistic programs following the data and control flow.

2) Abstract Interpretation: Abstract interpretation is a technique for approximating the behaviour of a program by replacing its concrete semantics with an *abstract* semantics [24]. An abstract domain is usually based on (partially) ordered sets or lattices. As an illustrative example, consider an analysis to determine whether a numerical variable is positive or negative. We can replace the concrete variables with an abstract representation of just their signs, i.e., whether the variable is positive ($+$), negative ($-$), or zero (0). To achieve soundness, the concrete semantics are over-approximated: we may need to accept that we cannot always tell the sign of a variable and have to assign it an abstract "don't know" value ($\top$).

For our purposes, we are interested in more than just the signs of program variables. As probabilistic programs are stochastic by nature and the values of random variables may be different from one execution to another, we would rather like to know the possible *range of values* of any particular variable. To achieve this, we have implemented an interval abstract domain in LASAPP to approximate value ranges with intervals. Intervals are propagated through the program by following interval arithmetic [25], [26]. In Section IV-D, we analyse the value range of distribution parameters to verify that they do not violate the imposed constraints in any possible program execution.

3) Symbolic execution: Symbolic execution is similar to abstract interpretation, in the sense that here too the concrete semantics of a program are abstracted in order to prove certain properties. Program variables or inputs are exchanged with symbols representing arbitrary values. The syntax of the programming language remains the same, but the execution now follows the symbolic semantics [27]. In contrast to abstract interpretation, symbolic execution does not necessarily explore all possible program paths, but it has the advantage that it can often avoid over-approximation.

A particular application of symbolic execution relevant for this work is concerned with so-called *path conditions*. Depending on the values of variables appearing in conditional if-statements, either the "then" or the "else" branch will be taken during execution. By representing the conditional symbolically, we can derive constraints that need to be satisfied for a certain program branch to be executed. This is especially interesting for probabilistic programs that exhibit stochastic branching. For these types of programs, random variables may appear in conditional statements where a certain fraction p of executions are expected to take the "then" branch, while the "else" branch will be taken with probability $1 - p$. Such properties are especially relevant for validating so-called guide

programs, a practical analysis we describe in Section IV-E.

IV. Sound Static Analyses in LASAPP

In this section, we present four static analyses formulated in the LASAPP framework. They are proven to be sound for small formal PPL in the supplementary material and evaluated on real-world programs in section V. For the sake of brevity, we will only give an overview description of the analyses and simplified versions of the implementations in LASAPP's API. The full implementations can be found in the replication package.

A. A Small Formal PPL: WhileProb

In this work, we not only consider the application of LASAPP to practical PPLs, but we also aim to investigate the framework from a formal stance. To this end, we introduce a small formal PPL called WhileProb, which is similar to other formal PPLs studied in prior work [1], [4]. Its complete syntax is given below.

Syntax of WhileProb:

$E ::=$	expression
$\quad c$	constant
$\quad x$	variable
$\quad g(E_1, \ldots, E_n)$	function call
$S ::=$	statement
$\quad$ skip	skip
$\quad x = E$	assignment
$\quad S_1; S_2$	sequence
$\quad$ if E then S_1 else S_2	if statement
$\quad$ while E do S	while loop
$\quad x = \text{sample}(E_0, f(E_1, \ldots, E_n))$	sample
$\quad x = \text{observe}(E_0, f(E_1, \ldots, E_n))$	observe
$P ::=$	program
$\quad$ model S	model definition
$\quad$ model S_1; guide S_2	model and guide

Due to limited space, the formal semantics of WhileProb, algorithms to implement the LASAPP API (Figure 4), and all correctness proofs will be provided in the supplementary material.

B. Statistical Dependency Analysis

Finding dependencies between sample statements is a building block of many analyses. In fact, representing a probabilistic program graphically as a Bayesian network allows us to apply algorithms designed specifically for probabilistic graphical models for further analysis [23]. The high-level overview of the algorithm to find these dependencies is as follows:

1) Initialise an empty statistical dependency graph $\mathcal{G}$.
2) Find all sample statements.
3) Filter for random variables that are reachable from the model entry point (e.g. by inspecting the call graph).
4) For each random variable, traverse the data dependency graph until we encounter another sample statement. For each data dependency d, we also need to traverse the data dependencies of every control dependency of d.
5) Add an edge between the two sample statements to $\mathcal{G}$. As the address of a random variable may be dynamic, we also need to get its dependencies.

You can see a concrete LASAPP implementation of the steps above in Listing 1 (simplified and shortened for clarity).

```
dependencies = [] # == G                                    # (1)
all_variables = program.get_random_variables()              # (2)
model = program.get_model()
random_vars = filter(all_variables, model)                  # (3)
for rv in random_vars:                                      # (4)
  queue = [rv.address_node, rv.distribution.node]
  while len(queue) > 0:
    node = queue.popleft()
    data_deps = program.get_data_dependencies(node)
    for dep in data_deps:                                   # (4)
      if dep in random_vars:
        dependencies.append((dep, rv))                      # (5)
        queue.append(dep.address_node)
      else:
        queue.append(dep)
  for dep in program.get_control_dependencies(node):# (4)
    queue.append(dep.control_node)
```

Listing 1. Model graph extraction written in LASAPP.

In Figure 5, we illustrate how the algorithm operates and why we also need to traverse the data dependencies of control nodes. To find the statistical dependencies of the random variable `x`, we start by finding the data dependencies of the distribution parameter `m`. We notice that the value of `m` depends on the conditional `b == 1` and thus, also `x` implicitly depends on the conditional. We continue to find the only data dependency of `b == 1` in form of the sample statement of `b`, stop the traversal and add the edge `b` → `x` to the statistical dependency graph. In Figure 6, you can see an example program adapted from Mak et al. [28] and its static model graph.

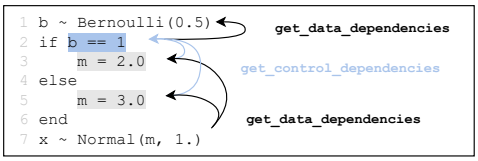


Fig. 5. Repeated application of `get_data_dependencies` and `get_control_dependencies` to find statistical dependencies in a Turing program.

The analysis described above is sound in the sense of the following proposition for programs written in WhileProb. This statement will be made more precise in the supplementary material, where we also provide the proof.

Proposition 1: Let P be a WhileProb program. Then, the model density of P factorises according to the graph extracted with the analysis outlined in Listing 1.

C. HMC Assumptions Checker

We turn back to our motivating example, see Figure 3. Recall that the runtime errors caused by applying the Hamiltonian Monte Carlo algorithm can be prevented statically

```
@model function pedestrian(end_distance)
  start ~ Uniform(0,3)
  t = 0
  position = start
  distance = 0.0
  step = Dict()
  while position > 0 && position < 10
    t = t + 1
    step[t] ~ Uniform(-1, 1)
    distance = distance + abs(step[t])
    position = position + step[t]
  end
  end_distance ~ Normal(distance, 0.1)
  return start
end
```

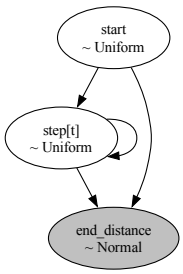


Fig. 6. Static model graph of a Turing program extracted with LASAPP.

by disallowing discrete variables in the model. We already presented the corresponding LASAPP analysis in Section II and include it again in Listing 2. However, the assumption of continuous variables is not the only one for HMC and we continue to implement more checks.

As a gradient-based method, HMC requires the program density to be differentiable with respect to all random variables. Stochastic control flow often leads to discontinuous densities and can be statically avoided, if we check whether a random variable is control dependent on another variable. This can be done with an algorithm that is a slight modification of the dependency analysis of Section IV-B, where only those dependencies are recorded that affect the control flow. We abbreviate this subroutine with `has_random_control_deps`. The full implementation can be found in the supplied artifact.

```
random_variables = program.get_random_variables()
for rv in random_variables:
  props = lasapp.infer_distribution_properties(rv)
  if props.is_discrete():
    raise Warning("Discrete Variables ...")
  if has_random_control_deps(program, rv):
    raise Warning("Stochastic Control Flow ...")
```

Listing 2. HMC assumption checker written in LASAPP.

Again, we can make a formal statement about the soundness of the HMC assumption checker for WHILEPROB programs. See the supplementary material for the proof.

Proposition 2: Let P be a WHILEPROB program that only uses differentiable primitive functions. If the analysis outlined in this section does not produce warnings, then the model density defined by P is differentiable.

D. Parameter Constraint Verifier

One of the bigger challenges when starting with probabilistic programming is that probabilistic programs are typically not executed sequentially and values of variables are random. This may lead to runtime errors that occur with low probability which are hard to reproduce and debug. Therefore, it is often productive to think about the entire *range of values* a random variable can take (its support) to verify that a program runs without errors for every possible execution trace.

This line of thought motivated the use of abstract interpretation in form of interval analysis as described in Section III-B2 to verify that the parameters passed to distributions satisfy constraints. For example, the standard deviation parameter of a Normal distribution has to be strictly positive. This check, among others, is also included in Stan's pedantic mode implemented directly in the compiler [29]. In contrast, the analysis presented in the following is written in the LASAPP framework and can be applied to multiple PPLs at once.

We give a brief overview of the probabilistic analysis to find constraint violations in distribution arguments:

1) Initialise an empty "assumptions map" $\mathcal{A}$.
2) Find all sample statements.
3) For each random variable R, retrieve its value range V and add $R \mapsto V$ to $\mathcal{A}$.
4) For each random variable, retrieve its parameter constraints.
5) Check for violations by comparing against the estimated range for the parameter node in $\mathcal{A}$.

The LASAPP implementation can be seen in Listing 3.

```
assumptions = {}                                              # (1)
random_variables = program.get_random_variables()             # (2)
for rv in random_variables:
  props = lasapp.infer_distribution_properties(rv)
  assumptions[rv] = Interval(props.support)                   # (3)

for rv in random_variables:
  props = lasapp.infer_distribution_properties(rv)
  for param in rv.distribution.params:
    constraint = props.param_constraints[param]               # (4)
    estimated_range = program.estimate_value_range(           # (5)
      expr=param.node,
      assumptions=assumptions
    )
   if not estimated_range.is_subset_of(constraint):           # (5)
     raise Warning("Constraint violation ...")
```

Listing 3. Parameter constraint verifier written in LASAPP.

The analysis is best explained by considering the example program in Figure 7. We estimate the value range of the variable `prob`. The variable `z` is masked with $[-\infty, \infty]$ and `u` is masked with $[0, 1]$. Then, these intervals are transformed according to the mathematical expressions by applying interval arithmetic rules. Lastly, since either of the two branches could be taken, we compute the union of the estimated ranges for `prob` to be $[0, 1.5]$. This violates the constraint for the "success probability" of the Geometric distribution, which has to be less than 1. Note that the program would result in a runtime error with a probability of less than 0.1%.

In the following proposition, we state the soundness of the constraint verifier for WHILEPROB programs, which is proven in the supplementary material.

Proposition 3: Let P be a WHILEPROB program that only uses univariate random variables. If the analysis outlined in this section does not produce warnings, then P is free of distribution parameter constraint violations.

E. Model-Guide Validator

Many posterior inference algorithms targeting the density p make use of a reference distribution q. For instance, MCMC

```
@gen function model()
  b ~ bernoulli(0.999)                          # [0,1]
  if b
    z ~ normal(0.,1.)                           # Real
    prob = 1/(1+exp(z))                         # [0,1]
  else
    u ~ beta(1,1)                               # [0,1]
    prob = 1.5 * u                              # [0,1.5]
  end
  x ~ geometric(prob)                           # [0,1.5]
end
WARNING: Parameter p of Geometric
distribution has constraint [0, 1],
but values are estimated to be in [0.0, 1.5].
```

Fig. 7. Gen program with parameter constraint violations.

algorithms benefit from making random value draws not from the distribution defined in the model, but from a proposal distributions q. For instance, in the linear model in Figure 2, we define broad uninformative priors for the slope and intercept parameters as to not bias the model towards specific values for these parameters. However, in inference, it makes sense to sample values, for instance, in the neighbourhood of the least squares solution to the line fitting problem, because we know that the posterior distribution will have low probability outside of this neighbourhood. Furthermore, variational inference methods fit a so-called variational distribution q to be close to p. In these algorithms, the density ratios $p(x)/q(x)$ and $q(x)/p(x)$ occur, which are only well-defined (in a probabilistic sense, $0/0 = 0$) if p and q satisfy the absolute continuity property. We say p is *absolutely continuous* with respect to q and write $p \ll q$ if

$$p(x) > 0 \implies q(x) > 0 \text{ for all } x.$$

This condition makes $p(x)/q(x)$ well-defined, whereas $q \ll p$ makes $q(x)/p(x)$ well-defined.

In probabilistic programming, these reference distributions come in form of so-called *guide programs* [2], [30], which are typically defined by the user. To make posterior inference sound the guide must match the model in terms of absolute continuity. This motivated the formulation of a static analysis in the LASAPP framework, which automatically verifies that a model and guide program satisfy the absolute continuity property. This analysis is described briefly in five steps, where we assume that all random variables are univariate and that all addresses are static.

1) Find all sample statements for the model and guide program and group them by address (name). To check $\texttt{model} \ll \texttt{guide}$ take $p = \texttt{model}$ and $q = \texttt{guide}$. To check $\texttt{guide} \ll \texttt{model}$ take $q = \texttt{model}$ and $p = \texttt{guide}$.
2) Get the path condition $\texttt{pc}(\texttt{stmt})$ for each sample statement in terms random variables.
3) Express the support interval $[a, b]$ of sample statements with address X symbolically in form of distribution constraints $\text{dc}(\text{stmt}) = a \leq X \wedge X \leq b$.
4) For each address X of program P, combine the constraints of all of its sample statements:

$$\text{constraints}(p, X) = \bigvee_{\substack{\text{stmt} \in \text{sample-} \\ \text{stmts of } X \text{ in } p}} \text{pc}(\text{stmt}) \wedge \text{dc}(\text{stmt})$$

5) For each address X, use an SMT-solver to prove that the following predicate is unsatisfiable:

$$\neg[\text{constraints}(p, X) \implies \text{constraints}(q, X)]$$

Again, we show a simplified version of the LASAPP implementation in Listing 4.

```
def get_dist_constraint(rv):
  props = lasapp.infer_distribution_properties(rv)
  a, b = Interval(props.support)
  X = SExpr(rv)
  return And(a < X, X < b)

rvs = program.get_random_variables()                    # (1)
P = program.get_model(); Q = program.get_guide();
# Q = program.get_model(); P = program.get_guide();
P_rvs = ...;              Q_rvs = ...;
P_rvs_by_name = ...;      Q_rvs_by_name = ...;

A = {rv.node: SExpr(rv) for rv in rvs} # symbolic masks
pc = {                                                  # (2)
**{rv: program.get_path_condition(rv.node, P.node, A)
                                        for rv in P_rvs},
**{rv: program.get_path_condition(rv.node, Q.node, A)
                                        for rv in Q_rvs}
}

dc = {rv: get_dist_constraint(rv) for rv in rvs}        # (3)

for name, P_stmts in P_rvs_by_name.items():
  Q_stmts = Q_rvs_by_name[name]
  impl = Implies(                                       # (4)
    Or([And(pc[stmt],dc[stmt]) for stmt in P_stmts]),
    Or([And(pc[stmt],dc[stmt]) for stmt in Q_stmts])
  )
  if check(Not(impl)) == satisfiable:                   # (5)
    raise Warning("Absolute continuity violation ...")
```

Listing 4. Model-guide validation written in LASAPP.

Note that $\bigwedge_{X \in \text{used-addresses}} \text{constraints}(p, X)$ precisely captures the condition $p(x) > 0$ symbolically. This makes the analysis sound, which is proven for WHILEPROB programs in the supplementary material and is summarised in following proposition. We assume that the programs do not contain while loops to make exact path conditions tangible and the proof more straightforward.

Proposition 4: Let P be a WHILEPROB program that only uses univariate random variables with static addresses. Further, assume that P does not contain while loops. Let p be the model density and q the guide density defined by P. If the analysis outlined in this section does not produce warnings, then $p \ll q$.

V. EVALUATION

While we have established correctness results of the presented static analyses for the small formal PPL WHILEPROB, it is natural to ask whether these analyses have also a practical benefit for the more complex real-world PPLs. To answer this

question, we test the prototype implementation of the analyses on a plethora of real-world probabilistic programs.

To evaluate the statistical dependency analysis (section IV-B) and constraint verifier (section IV-D), we gathered a total of 117 Turing and 97 PyMC programs from open-source repositories comprised mostly of implementations of models from textbooks on applied Bayesian inference [31]–[34].

By manual inspection, we verified that the statistical dependency analysis correctly extracted the model graph of 111 Turing models. It failed on 6 programs, where programming constructs (generators and lambda functions) were used, which are not supported by our prototype yet. For the PyMC programs, we compared the static model graph against the dynamically computed graph by PyMC. There were only errors for 2 models, due to arbitrary distribution expressions used in sample statements. While the error rate is already low, it could be further lowered by improving our prototype implementation as the errors are not inherit limitations of the LASAPP framework itself. We also note that with this analysis we even found a bug in the implementation of one textbook model. In this program a random variable node was disconnected from all other variables in the graph, because a different variable was mistakenly used in its place in the computation of an intermediate value.

With the constraint verification analysis, we were able to assert that 80/117 Turing programs and 65/97 PyMC are free of parameter constraint violations. For the rest of programs, this assertion could not be established, because either 1) data read from an external source was required to satisfy constraints or 2) multivariate distributions were present where the constraints cannot be verified with interval arithmetic. The second reason could be mitigated by extending LASAPPs classical analysis foundation with a new abstract interpretation domain suited for the constraints of multivariate distributions. This exemplifies the benefits of the modular design of LASAPP.

To evaluate the HMC assumption checker (section IV-C), we cannot use any arbitrary probabilistic program, since not every program is meant to work with HMC. Instead, we collected 8 probabilistic programs with discontinuous densities from literature. For 6 of them the discontinuity arises from random control flow, for 1 of them from the use of discrete variables, and for 1 of them from the use of the discontinuous `floor` rounding function. 7 of these programs served as motivating examples for more sophisticated inference methods that overcome the shortcomings of HMC [28], [35]–[37] and 1 program motivated the static analysis for random control flow in the Stan compiler [29]. We implemented these models in Gen and found that our static analysis gives the correct warning that HMC is not applicable for 7/8 programs. It fails to give the warning for the program which uses the `floor` function as our analysis does not verify the continuity properties of primitive functions used in the program.

Lastly, we evaluate the model-guide validator (section IV-E), on Pyro programs, as Pyro heavily relies on model-guide pairs to perform stochastic variational inference (SVI). In fact, Lee et al [38] developed a static analysis with the same goal and provide a benchmark set of 8 Pyro programs, which we use for this evaluation. Even though their method is formally sound, more sophisticated, and specifically tailored to Pyro, our more light-weight language agnostic analysis achieves the same result on this benchmark set. Namely, it asserts the validity of 6 model-guide pairs and raises the correct warnings for the `br` and `lda` programs[2].

We end this section by noting that running all analyses mentioned in this section takes a combined time of less than 15 seconds on a modern laptop.

VI. Related Work

A. Language-Agnostic Program Analysis

Research on static program analysis is plentiful [39], [40]. Therefore, in this section, we focus on program analysis methods that aim to support language-agnostic analyses to some extent.

In 2015, Google presented the Tricorder framework [41]. Similar to the LASAPP language servers, the Tricorder ecosystem consists of analyzer worker services written in different languages that all implement a common language-agnostic API. Analyzers may be written in any language and may analyze any language. The analysis driver calls out to the language- or compiler-specific workers to run the program analyses.

The Cobra tool by Hoffmann [42] achieves language-agnosticism by interpreting programs simply as a list of tokens. These tokens are annotated with categories (identifier, keyword, type, etc.) and with additional context information. A special query language was developed to formulate program analyses on these lists of tokens.

The domain-specific language and infrastructure Boa [43], [44] was developed for easier and reproducible analyses of software repositories. Boa implements a representation of source code as abstract syntax trees with generic nodes like declaration, type, method, etc. It provides language features inspired by visitor patterns to allow users to query and analyse the ASTs of large-scale projects.

The methods listed above focus on program analyses for general program properties that are applicable to many languages and are not powerful enough to implement the analyses presented in this work. In contrast, our framework is designed specifically for the analysis of probabilistic programs written in different host-languages.

Tangentially related, there are efforts to create language-agnostic representations of source code for machine learning models. For instance, Zügner et al. [45] presented a multilingual code summarization model.

Lastly, we like to mention the compilation framework LLVM [46]. LLVM is designed around a language-independent intermediate representation to facilitate program analyses. However, this representation is too low-level for the analyses considered in this work.

[2]In the `lda` program, the model-guide mismatch is intentional, because variational expectation-maximisation was employed instead of SVI.

B. Static Analysis For Probabilistic Programs

The field of static analysis for probabilistic programming is young and largely under-explored. In this section, we will highlight prior work related to the probabilistic program analyses presented in Section IV and refer to Bernstein's survey [47] for more methods and details.

In 2014, [4] presented a way to reduce probabilistic programs to only those statements that are relevant to estimate the distribution of the return expression. This program slicing approach is a semantics-preserving transformation which can speed up posterior inference.

In 2014, [48] introduced probabilistic assertion statements and an evaluation approach to either statically or dynamically verify them. In 2018, [49] presented a method to automatically analyze the source code of density functions for conjugacy structure and leveraged the structure for improved inference.

Stan has received the most attention with regards to static analysis. In 2018, SlicStan [1], a compositional version of the probabilistic programming language, was published. By relaxing the strict block syntax of Stan, it makes the code more flexible, reusable, and beginner friendly. Also, a semantics-preserving translation to Stan was developed. In the following years, Stan was also extended with a pedantic compilation mode [29]. Similar to the analyses in Section IV-C and IV-D, warnings will be raised if there are any distribution usage issues or stochastic control flow. A static transformation method relying on the extraction of the factor graph was developed to make draws from the prior and posterior predictive distribution of Stan programs [50]. Lastly, a compilation sheme was designed to translate a Stan program to a generative PPL [51].

Most recently [52] developed a type system to enforce absolute continuity of Pyro guide programs and [2] presented an automatic approach to generate sound guide programs for a given Pyro program. Lastly, [53] make use of a semantics-preserving compilation to a first-order functional language to design an inference algorithm for their synchronous PPL called ProbZelus.

While the analyses of Section IV have some overlap with the methods above, the key difference is that through the LASAPP framework our analyses are not restricted to one particular probabilistic programming language, but are written in a language-agnostic way and can be applied to any PPL which implements the API bindings.

VII. Limitations

One potential limiting factor of our approach is that the framework and its abstractions may be only suitable for the selected set of probabilistic programming languages and the implemented program analyses. However, we argue that LASAPP generalises to new PPLs and analyses. The fact that the framework supports many different kinds of sample statements (function calls, function definitions, special syntax) shows its flexibility and adaptability to new languages. Also, the classic static analyses that are the backbone of the API are versatile. In particular, data and control flow was important for almost every program analysis presented in Section IV. Furthermore, the modularity of the LASAPP framework allows for the extension with additional classic static analysis methods if needed. Lastly, while the analyses are proven to be sound for the considered formal PPL and are evaluated to be practical for real-world programs, slight differences in semantics of real-world PPLs may lead to incorrect results.

VIII. Conclusion

In this work, we presented the first language-agnostic static analysis framework for probabilistic programs LASAPP. We have demonstrate the effectiveness and expressiveness of the LASAPP framework by presenting four program analyses that address problems discussed in literature and applied them to five state-of-the-art probabilistic program languages.

To conclude, we describe the key characteristics of our language-agnostic static analysis framework for probabilistic programming by reflecting on the insights gained from designing LASAPP. First, probabilistic programming languages are special in the sense that their programs can be universally interpreted as probabilistic models. They only differ in the underlying posterior inference machinery and expressiveness. This justifies the goal of a language-agnostic framework as the same analysis for some probabilistic program properties is meaningful for many PPLs.

It is clear that probabilistic program analyses rely on classical static analysis methods. By designing LASAPP in a modular fashion, we made an effort to clearly separate the two types of analysis. This allows programming languages experts to build the language servers and implement the needed classical static analysis techniques like data and control flow analysis. The probabilistic programming systems experts can then provide the bindings for their PPL, where we have identified sample statements as the core language construct. The complexity of these implementations is then abstracted away by LASAPP's API. We argue that the selection of LASAPP's API methods forms a set of essential abstractions. Finding dependencies of random variables via data and control flow is important for almost all analyses. Further, interval arithmetic and symbolic execution lends itself to handle the stochastic nature of probabilistic programs.

We have demonstrated the versatility of these abstraction by implementing four analyses tackling common problems in probabilistic programming including checking the assumptions of the HMC inference algorithm and verifying the compatibility of guide programs. We have proven the analyses to be sound for the formal PPL WhileProb and evaluated them on over 200 real-world probabilistic programs implemented in four different languages demonstrating their practical benefits.

References

[1] M. I. Gorinova, A. D. Gordon, and C. Sutton, "Slicstan: Improving probabilistic programming using information flow analysis," in *Workshop on Probabilistic Programming Languages, Semantics, and Systems (PPS). https://pps2018. soic. indiana. edu/files/2017/12/SlicStanPPS. pdf*, 2018.

[2] J. Li, L. Ven, P. Shi, and Y. Zhang, "Type-preserving, dependence-aware guide generation for sound, effective amortized probabilistic inference," *Proceedings of the ACM on Programming Languages*, vol. 7, no. POPL, pp. 1454–1482, 2023.

[3] C. Nandi, D. Grossman, A. Sampson, T. Mytkowicz, and K. S. McKinley, "Debugging probabilistic programs," in *Proceedings of the 1st ACM SIGPLAN International Workshop on Machine Learning and Programming Languages*, pp. 18–26, 2017.

[4] C.-K. Hur, A. V. Nori, S. K. Rajamani, and S. Samuel, "Slicing probabilistic programs," *ACM SIGPLAN Notices*, vol. 49, no. 6, pp. 133–144, 2014.

[5] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. Szerlip, P. Horsfall, and N. D. Goodman, "Pyro: Deep universal probabilistic programming," *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 973–978, 2019.

[6] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in python using pymc3," *PeerJ Computer Science*, vol. 2, p. e55, 2016.

[7] N. Tehrani, N. S. Arora, Y. L. Li, K. D. Shah, D. Noursi, M. Tingley, N. Torabi, E. Lippert, E. Meijer, *et al.*, "Bean machine: A declarative probabilistic programming language for efficient programmable inference," in *International Conference on Probabilistic Graphical Models*, pp. 485–496, PMLR, 2020.

[8] H. Ge, K. Xu, and Z. Ghahramani, "Turing: a language for flexible probabilistic inference," in *International conference on artificial intelligence and statistics*, pp. 1682–1690, PMLR, 2018.

[9] M. F. Cusumano-Towner, F. A. Saad, A. K. Lew, and V. K. Mansinghka, "Gen: a general-purpose probabilistic programming system with programmable inference," in *Proceedings of the 40th acm sigplan conference on programming language design and implementation*, pp. 221–236, 2019.

[10] A. D. Gordon, T. A. Henzinger, A. V. Nori, and S. K. Rajamani, "Probabilistic programming," in *Future of Software Engineering Proceedings*, pp. 167–181, 2014.

[11] G. Barthe, J.-P. Katoen, and A. Silva, *Foundations of Probabilistic Programming*. Cambridge University Press, 2020.

[12] S. Staton, H. Yang, F. Wood, C. Heunen, and O. Kammar, "Semantics for probabilistic programming: higher-order functions, continuous distributions, and soft constraints," in *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*, pp. 525–534, 2016.

[13] M. Vákár, O. Kammar, and S. Staton, "A domain theory for statistical probabilistic programming," *Proceedings of the ACM on Programming Languages*, vol. 3, no. POPL, pp. 1–29, 2019.

[14] M. Cusumano-Towner, A. K. Lew, and V. K. Mansinghka, "Automating involutive mcmc using probabilistic and differentiable programming," *arXiv preprint arXiv:2007.09871*, 2020.

[15] M. D. Hoffman, A. Gelman, *et al.*, "The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo.," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1593–1623, 2014.

[16] EvoArt, "Hmc issue in the julia turing forum." https://discourse.julialang.org/t/turing-inexacterror-for-discreteuniform-distribution-with-nuts-sampler/52820. Accessed: 2023-07-24.

[17] jianlin, "Hmc issue in the pyro forum." https://forum.pyro.ai/t/mcmc-discrete-rv-parallelization-is-there-anyway-to-stop-pyro-automatically-vectorizing-tensors/4160. Accessed: 2023-07-24.

[18] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," 2019.

[19] M. Besançon, T. Papamarkou, D. Anthoff, A. Arslan, S. Byrne, D. Lin, and J. Pearson, "Distributions.jl: Definition and modeling of probability distributions in the juliastats ecosystem," *Journal of Statistical Software*, vol. 98, no. 16, pp. 1–30, 2021.

[20] J. L. Hennessy and D. A. Patterson, *Computer architecture: a quantitative approach*. Elsevier, 2011.

[21] K. D. Cooper and L. Torczon, *Engineering a compiler*. Elsevier, 2011.

[22] R. Cytron, J. Ferrante, B. K. Rosen, M. N. Wegman, and F. K. Zadeck, "Efficiently computing static single assignment form and the control dependence graph," *ACM Transactions on Programming Languages and Systems (TOPLAS)*, vol. 13, no. 4, pp. 451–490, 1991.

[23] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[24] P. Cousot and R. Cousot, "Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints," in *Proceedings of the 4th ACM SIGACT-SIGPLAN symposium on Principles of programming languages*, pp. 238–252, 1977.

[25] R. E. Moore, *Interval analysis*, vol. 4. Prentice-Hall Englewood Cliffs, 1966.

[26] G. Alefeld and G. Mayer, "Interval analysis: theory and applications," *Journal of computational and applied mathematics*, vol. 121, no. 1-2, pp. 421–464, 2000.

[27] J. C. King, "Symbolic execution and program testing," *Communications of the ACM*, vol. 19, no. 7, pp. 385–394, 1976.

[28] C. Mak, F. Zaiser, and L. Ong, "Nonparametric hamiltonian monte carlo," in *International Conference on Machine Learning*, pp. 7336–7347, PMLR, 2021.

[29] R. Bernstein, *Abstractions for Probabilistic Programming to Support Model Development*. PhD thesis, Columbia University, 2023.

[30] M. F. Cusumano-Towner and V. K. Mansinghka, "Using probabilistic programs as proposals," *arXiv preprint arXiv:1801.03612*, 2018.

[31] R. McElreath, *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC, 2018.

[32] M. D. Lee and E.-J. Wagenmakers, *Bayesian cognitive modeling: A practical course*. Cambridge university press, 2014.

[33] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

[34] A. A. Johnson, M. Q. Ott, and M. Dogucu, *Bayes rules!: An introduction to applied Bayesian modeling*. Chapman and Hall/CRC, 2022.

[35] H. Mohasel Afshar and J. Domke, "Reflection, refraction, and hamiltonian monte carlo," *Advances in neural information processing systems*, vol. 28, 2015.

[36] A. Nishimura, D. B. Dunson, and J. Lu, "Discontinuous hamiltonian monte carlo for discrete parameters and discontinuous likelihoods," *Biometrika*, vol. 107, no. 2, pp. 365–380, 2020.

[37] Y. Zhou, H. Yang, Y. W. Teh, and T. Rainforth, "Divide, conquer, and combine: a new inference strategy for probabilistic programs with stochastic support," in *International Conference on Machine Learning*, pp. 11534–11545, PMLR, 2020.

[38] W. Lee, H. Yu, X. Rival, and H. Yang, "Towards verified stochastic variational inference for probabilistic programs," *Proceedings of the ACM on Programming Languages*, vol. 4, no. POPL, pp. 1–33, 2019.

[39] V. D'silva, D. Kroening, and G. Weissenbacher, "A survey of automated techniques for formal software verification," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 7, pp. 1165–1178, 2008.

[40] A. Gosain and G. Sharma, "Static analysis: A survey of techniques and tools," in *Intelligent Computing and Applications: Proceedings of the International Conference on ICA, 22-24 December 2014*, pp. 581–591, Springer, 2015.

[41] C. Sadowski, J. Van Gogh, C. Jaspan, E. Soderberg, and C. Winter, "Tricorder: Building a program analysis ecosystem," in *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, vol. 1, pp. 598–608, IEEE, 2015.

[42] G. J. Holzmann, "Cobra: a light-weight tool for static and dynamic program analysis," *Innovations in Systems and Software Engineering*, vol. 13, no. 1, pp. 35–49, 2017.

[43] R. Dyer, H. A. Nguyen, H. Rajan, and T. N. Nguyen, "Boa: Ultra-large-scale software repository and source-code mining," *ACM Transactions on Software Engineering and Methodology (TOSEM)*, vol. 25, no. 1, pp. 1–34, 2015.

[44] R. Dyer, H. A. Nguyen, H. Rajan, and T. N. Nguyen, "Boa: A language and infrastructure for analyzing ultra-large-scale software repositories," in *2013 35th International Conference on Software Engineering (ICSE)*, pp. 422–431, IEEE, 2013.

[45] D. Zügner, T. Kirschstein, M. Catasta, J. Leskovec, and S. Günnemann, "Language-agnostic representation learning of source code from structure and context," *arXiv preprint arXiv:2103.11318*, 2021.

[46] C. Lattner and V. Adve, "Llvm: A compilation framework for lifelong program analysis & transformation," in *International symposium on code generation and optimization, 2004. CGO 2004.*, pp. 75–86, IEEE, 2004.

[47] R. Bernstein, "Static analysis for probabilistic programs," *arXiv preprint arXiv:1909.05076*, 2019.

[48] A. Sampson, P. Panchekha, T. Mytkowicz, K. S. McKinley, D. Grossman, and L. Ceze, "Expressing and verifying probabilistic assertions," in *Proceedings of the 35th ACM SIGPLAN Conference on Programming Language Design and Implementation*, pp. 112–122, 2014.

[49] M. D. Hoffman, M. J. Johnson, and D. Tran, "Autoconj: recognizing and exploiting conjugacy without a domain-specific language," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[50] R. Bernstein, M. Vákár, and J. Wing, "Transforming probabilistic programs for model checking," in *Proceedings of the 2020 ACM-IMS on Foundations of Data Science Conference*, pp. 149–159, 2020.

[51] G. Baudart, J. Burroni, M. Hirzel, L. Mandel, and A. Shinnar, "Compiling stan to generative probabilistic languages and extension to deep probabilistic programming," in *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*, pp. 497–510, 2021.

[52] D. Wang, J. Hoffmann, and T. Reps, "Sound probabilistic inference via guide types," in *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*, pp. 788–803, 2021.

[53] G. Baudart, L. Mandel, E. Atkinson, B. Sherman, M. Pouzet, and M. Carbin, "Reactive probabilistic programming," in *Proceedings of the 41st ACM SIGPLAN Conference on Programming Language Design and Implementation*, pp. 898–912, 2020.