This Is the Moment for Probabilistic Loops

MARCEL MOOSBRUGGER, TU Wien, Austria
MIROSLAV STANKOVIČ, TU Wien, Austria
EZIO BARTOCCI, TU Wien, Austria
LAURA KOVÁCS, TU Wien, Austria

We present a novel static analysis technique to derive higher moments for program variables for a large class of probabilistic loops with potentially uncountable state spaces. Our approach is fully automatic, meaning it does not rely on externally provided invariants or templates. We employ algebraic techniques based on linear recurrences and introduce program transformations to simplify probabilistic programs while preserving their statistical properties. We develop power reduction techniques to further simplify the polynomial arithmetic of probabilistic programs and define the theory of moment-computable probabilistic loops for which higher moments can precisely be computed. Our work has applications towards recovering probability distributions of random variables and computing tail probabilities. The empirical evaluation of our results demonstrates the applicability of our work on many challenging examples.

CCS Concepts: • Mathematics of computing → Markov processes; • Computing methodologies → Symbolic and algebraic algorithms; • Theory of computation → Random walks and Markov chains.

Additional Key Words and Phrases: Probabilistic Programs, Higher Moments, Linear Recurrences, Distribution Recovery

1 INTRODUCTION

Probabilistic programming languages enrich classical imperative or functional languages with native primitives to draw samples from random distributions, such as Bernoulli, Uniform, and Normal distributions. The resulting probabilistic programs (PPs) [Barthe et al. 2020; Kozen 1985] embed uncertain quantities, represented by random variables, within standard program control flows. As such, PPs offer a unifying framework to naturally encode probabilistic machine learning models [Ghahramani 2015], for example Bayesian networks [Kaminski et al. 2016], into programs. Moreover, PPs enable programmers to handle uncertainty resulting from sensor measurements and environmental perturbations in cyber-physical systems [Chou et al. 2020; Selyunin et al. 2015]. Other notable examples of PPs include the implementation of cryptographic [Barthe et al. 2012a] and privacy [Barthe et al. 2012b] protocols, as well as randomized algorithms [Motwani and Raghavan 1995] such as Herman’s self-stabilization protocol [Herman 1990] for recovering from faults in a process token ring — see our example in Figure 2.

Analysis of PPs. The random nature of PPs makes their functional analysis very challenging as one needs to reason about probability distributions of random variables instead of computing with single variable values [Barthe et al. 2020]. A standard approach towards handling probability distributions associated with random variables is to estimate such distributions by sampling PPs using Monte Carlo simulation techniques [Hastings 1970]. While such approaches work well for statistical model checking [Younes and Simmons 2006], they are not suitable for the analysis of PPs with potentially infinite program loops as simulating infinite-state behavior is not always viable. Moreover, even for PPs with finitely many states, simulation-based analysis is inherently approximative.

With the aim of precisely, and not just approximately, handling random variables, probabilistic model checking [Dehnert et al. 2017; Kwiatkowska et al. 2011] became a prominent approach in
the analysis of PPs with finite state spaces. For analyzing unbounded PPs, these techniques would however require non-trivial user guidance, in terms of assertion templates and/or invariants. In this paper, we address the challenge of precisely analyzing, and even recovering, probability distributions induced by PPs with both countably and uncountably infinite state spaces. We do so by extending both expressivity and automation of the state-of-the-art in PP analysis: We (i) focus on PPs with probabilistic infinite loops (see Figure 1) and (ii) fully automate the analysis of such loops by computing exact higher-order statistical moments of program variables $x$ parameterized by a loop counter $n$.

Functional representations $f(n)$ for a program variable $x$, with $f(n)$ characterizing the $k$th moment $\mathbb{E}(x^k)$ of $x$ at iteration $n$, can be interpreted as a quantitative invariant $\mathbb{E}(x^k) - f(n) = 0$, as the equation is true for all loop iterations $n \in \mathbb{N}$. Inferring quantitative invariants is arguably not novel. On the contrary, it is one of the most challenging aspects of PP analysis, dating back to the seminal works of [Katoen et al. 2010; McIver and Morgan 2005] introducing the weakest pre-expectations calculus. Template-based approaches to discover invariants or (super-)martingales emerged [Barthe et al. 2016; Kura et al. 2019] by translating the invariant generation problem into a constraint solving one. The derived quantitative invariants are generally provided in terms of expected values [Chakarov and Sankaranarayanan 2014; Katoen et al. 2010; McIver and Morgan 2005]. Nevertheless, the expected value alone — also referred to as the first moment — provides only partial information about the underlying probability distribution. This motivates the critical importance of higher moments for PP analysis [Bartocci et al. 2020b; Kura et al. 2019; Stankovic et al. 2022; Wang et al. 2021].

Higher Moments for PP Analysis. Using concentration-of-measure inequalities [Boucheron et al. 2013], we can utilize higher moments $\mathbb{E}(X^k)$ to obtain upper and lower bounds on tail probabilities $P(X > t)$, measuring the probability that a given random variable $X$, corresponding for example to our program variables $x$ from Figure 1, surpasses some value $t$. In this paper, we also show that when a program variable $x$ admits only $k < \infty$ many values, we can fully recover its probability mass function as a closed-form expression in the loop counter $n$ using the first $k-1$ raw moments (see Section 6). Furthermore, raw moments can be used to compute central moments $\mathbb{E}((X - \mathbb{E}(X))^k)$ and thus provide insights on other important characteristics of the distribution such as the variance, skewness and kurtosis [Durrett 2019]. However, computing exact higher statistical moments for PPs is computationally expensive [Kaminski et al. 2019], a challenge which we also address in this paper, as illustrated in Figures 1–2 and described next.

Computing Higher Moments. The theory we establish in this paper describes how to compute higher moments of program variables for a large class of probabilistic loops and how to utilize these moments to gain more insights into the analyzed programs. We call this theory the theory of moment-computable probabilistic loops (Section 5). Our approach is fully automatic, meaning it does not rely on externally provided invariants or templates. Unlike constraint solving over templates [Barthe et al. 2016; Kura et al. 2019], we employ algebraic techniques based on systems of linear recurrences with constant coefficients describing so-called C-finite sequences [Kauers and Paule 2011]. Different equivalence preserving program transformations (Section 3) and power reduction of finite valued variables (Section 4) allow us to simplify PPs and represent their higher moments as linear recurrence systems in the loop counter. Figure 1 shows a PP with many unique features supported by our work towards PP analysis: it has an uncountable state-space, contains if-statements, symbolic constants, draws from continuous probability distributions with state-dependent parameters, and employs polynomial arithmetic as well as circular variable dependencies. We are not aware of other works automating the reasoning about such and similar probabilistic loops,
well as approximate methods [Moosbrugger et al. 2022]. Our experiments (Section 7) show that variable dependencies, state-dependent distribution parameters, simultaneous assignments, and variables. We strengthen the theory of [Bartocci et al. 2019] to support if-statements, circular variables are not polynomially self-dependent and all branching conditions are over finite valued variables. Our work fully automates the analysis of such and similar PP loops by computing higher moments.

This Is the Moment for Probabilistic Loops

in particular for computing precise higher moments of variables. Figure 1 lists some of the variables’ moments computed automatically by our work. Further, these moments can be used to compute tail probability bounds or central moments, such as the variance, to characterize the distribution of the program variables as the loop progresses.

Thanks to our power reduction techniques (Section 4), our approach supports arbitrary polynomial dependencies among finite valued variables. Moreover, our work can fully recover the value distributions of finite valued program variables, from finitely many higher moments, as illustrated in Section 6 for Herman’s self-stabilization algorithm from Figure 2.

Theory and Practice in Computing Higher Moments. In theory, our approach can compute any higher moment for any variable and PP of our program model, under assumptions stated in Sections 3 and 5. We also establish the necessity of these assumptions in Section 5.3. In a nutshell, the completeness theorem (Theorem 6) holds for probabilistic loops for which non-finite program variables are not polynomially self-dependent and all branching conditions are over finite valued variables. We strengthen the theory of [Bartocci et al. 2019] to support if-statements, circular variable dependencies, state-dependent distribution parameters, simultaneous assignments, and multiple assignments, and establish the necessity of our assumptions. Moreover, unlike [Wang et al. 2021], our approach does not rely on templates and provides exact closed-form representations of higher moments parameterized by the loop counter.

In practice, our approach is implemented in the Polar tool and compared against exact as well as approximate methods [Moosbrugger et al. 2022]. Our experiments (Section 7) show that Polar outperforms the state-of-the-art of moment computation for probabilistic loops in terms of supported programs and efficiency. Furthermore, Polar is able to compute exact higher moments magnitudes faster than sampling can establish reasonable confidence intervals.

Contributions. Our main contributions are listed below:

• An automated approach for computing higher moments of program variables for a large class of probabilistic loops with potentially uncountable state spaces (Sections 3-5).
• We develop power reduction techniques to reduce the degrees of finite valued program variables in polynomials (Section 4).

Fig. 1. An example of a multi-path PP loop, with Laplace and Normal distributions parametrized by program variables. Our work fully automates the analysis of such and similar PP loops by computing higher moments. Several moments for program variables in the loop counter $n$ are listed on the right. Each moment was automatically generated.
\[
x_1, x_2, x_3 = 1, 1, 1 \\
t_1, t_2, t_3 = 1, 1, 1 \\
p = 1/2; \text{ tokens } = t_1 + t_2 + t_3 \\
\textbf{while } \star : \\
x_{1o}, x_{2o}, x_{3o} = x_1, x_2, x_3 \\
\text{if } x_{1o} = x_{3o} : \text{ Bernoulli}(p) \text{ else: } x_1 = x_{3o} \text{ end} \\
\text{if } x_{2o} = x_{1o} : \text{ Bernoulli}(p) \text{ else: } x_2 = x_{1o} \text{ end} \\
\text{if } x_{3o} = x_{2o} : \text{ Bernoulli}(p) \text{ else: } x_3 = x_{2o} \text{ end} \\
\text{if } x_1 = x_3 : t_1 = 1 \text{ else: } t_1 = 0 \\
\text{if } x_2 = x_1 : t_2 = 1 \text{ else: } t_2 = 0 \\
\text{if } x_3 = x_2 : t_3 = 1 \text{ else: } t_3 = 0 \\
\text{ tokens } = t_1 + t_2 + t_3 \\
\textbf{end}
\]

\[
\mathbb{E}(\text{tokens}_n) = 1 + 2 \cdot 4^{-n} \quad | \quad \mathbb{E}(\text{tokens}^2_n) = 1 + 8 \cdot 4^{-n} \quad | \quad \mathbb{E}(\text{tokens}^3_n) = 1 + 26 \cdot 4^{-n}
\]

Fig. 2. Herman’s self stabilization algorithm with three nodes encoded as a probabilistic loop together with three moments of tokens.

- We prove completeness of our work for computing higher moments (Section 5).
- We fully recover the distributions of finite valued program variables and approximate distributions for unbounded/continuous program variables from finitely many moments (Section 6).
- We provide an implementation and empirical evaluation of our work, outperforming the state-of-the-art in PP analysis in terms of automation and expressivity (Section 7).

2 PRELIMINARIES

We use the symbol \( \mathbb{P} \) for probability measures and \( \mathbb{E} \) for the expectation operator. The support of a random variable \( X \) is denoted by \( \text{supp}(X) \).

2.1 Probability Theory

Operationally, a probabilistic program is a Markov chain with potentially uncountably many states. Let us recall some notions about Markov chains. For more details on Markov chains and probability theory in general we refer the reader to [Durrett 2019].

For a fixed set \( S \), a \( \sigma \)-algebra is a non-empty set of subsets of \( S \) closed under complementation and countable unions.

**Definition 1 (Sequence Space).** Let \( (S, \mathcal{S}) \) be a measurable space, that is, \( S \) is a set with a \( \sigma \)-algebra \( \mathcal{S} \). Its sequence space is the measurable space \( (S^\omega, \mathcal{S}^\omega) \) where \( S^\omega := \{(s_1, s_2, \ldots) : s_i \in S\} \) and \( \mathcal{S}^\omega \) is the \( \sigma \)-algebra generated by the cylinder sets \( \text{Cyl}[B_1, \ldots, B_n] := \{\theta : \theta_i \in B_i, 1 \leq i \leq n\} \) for all prefixes \( B_1, \ldots, B_n \in \mathcal{S} \) and all \( n \in \mathbb{N} \).

A Markov kernel is, on a high level, a generalization of transition probabilities between states to uncountable state spaces and is required for the definition of a Markov chain.

**Definition 2 (Markov Chain).** Let \( (S, \mathcal{S}, \mathbb{P}) \) be a probability space and \( p : S \times S \to [0, 1] \) a Markov kernel. A stochastic process \( X_n \) is a Markov chain with Markov kernel \( p \) if

\[
\mathbb{P}(X_{n+1} \in B \mid X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = p(X_n, B).
\]
We briefly recall standard terminology on algebraic sequences and recurrences. For further details, with constant coefficients, we refer the reader to [Kauers and Paule 2011]. A sequence $(a_n)$ is a sequence of real numbers defined by a linear recurrence with constant coefficients, that is, $a_n = \sum_{i=1}^{r} c_i a_{n-i} + r$, for some natural numbers $r \in \mathbb{N}$ and real constants $c_i \in \mathbb{R}$. We refer to $\sum_{i=1}^{r} c_i a_{n-i} + r$ as the closed-form or the solution of the sequence $(a_n)$, or its recurrence.

Given a measurable space $(S, \mathcal{S})$, an initial distribution $\mu$, a stochastic process $X_n$ and a Markov kernel $p$, Kolmogorov’s Extension Theorem says that there is a unique measure $\mathbb{P}$ such that $X_n$ is a Markov chain in $(S^n, \mathcal{S}^n, \mathbb{P})$. For a random variable $X$, central moments $\mathbb{E}((X - \mathbb{E}(X))^k)$ can be computed from raw moments $\mathbb{E}(X^k)$ and vice versa through the transformation of center:

$$\mathbb{E}
(\mathbb{C})(X - b)^k
(\mathbb{D}) = \mathbb{E}
(\mathbb{C})(X - a + (a - b))^k
(\mathbb{D}) = \sum_{i=0}^{k} \binom{k}{i} \mathbb{E}(X - a)^i (a - b)^{k-i}. \quad (2)
$$

### 2.2 Linear Recurrences

We briefly recall standard terminology on algebraic sequences and recurrences. For further details, we refer the reader to [Kauers and Paule 2011]. A sequence $(a_n)_{n=0}^{\infty}$ is called C-finite if it obeys a linear recurrence with constant coefficients, that is, $(a_n)_{n=0}^{\infty}$ satisfies an equation of the form

$$a_{n+l} = c_l a_{n} + \cdots + c_1 a_{n-l+1} + c_0 a_{n-l},$$

for some order $l \in \mathbb{N}$, some constants $c_l \in \mathbb{R}$ and all $n \in \mathbb{N}$.

**Theorem 1** (Closed-form [Kauers and Paule 2011]). Every C-finite sequence $(a_n)_{n=0}^{\infty}$ can be written as an exponential polynomial, that is, $a_n = \sum_{i=1}^{m} d_i u_i^n$ for some natural numbers $d_i \in \mathbb{N}$ and complex numbers $u_i \in \mathbb{C}$. We refer to $\sum_{i=1}^{m} d_i u_i^n$ as the closed-form or the solution of the sequence $(a_n)$ or its recurrence.

An important fact is that closed-forms of linear recurrences with constant coefficients of any order always exist and are computable. This also holds for all variables in systems of linear recurrences with constant coefficients.

### 3 PROBABILISTIC PROGRAM MODEL

In this section we introduce our programming model (Section 3.1) and describe its semantics in terms of Markov chains (Section 3.2). Moreover, we introduce transformations (3.3) normalizing a probabilistic program to simplify its analysis.

#### 3.1 Probabilistic Program Syntax

The syntax defining our program model is given by the grammar in Figure 3. Throughout the paper, we will use the phrases (probabilistic) loops and (probabilistic) programs interchangeably for...
loops adhering to the syntax in Figure 3. In our work, we infer higher moments \( \mathbb{E}(x^n_k) \) of program variables \( x \) parameterized by the loop counter \( n \). We abstract from concrete loop guards by defining the guards of programs in our program model to be true (written as \( \star \)). Guarded loops while \( \phi \): \ldots can be modeled as an infinite loops while \( \star \): if \( \phi \): \ldots, with the limit behaviour giving the moments after termination (cf. Section 5.1).

Our program model defined in Figure 3 contains non-nested while-loops which are preceded by a loop-free initialization part. The loop-body and initialization part allow for (nested) if-statements, polynomial arithmetic, drawing from common probability distributions, and symbolic constants. Symbolic constants can be used to represent arbitrary real numbers and are also used for uninitia-\( \ldots \\)lized program variables. Categorical expressions (defined by the non-terminal \( \langle \text{categorical} \rangle \) in Figure 3) are expressions of the form \( v_l \{p_l\} \ldots v_l \{p_l\} \) such that \( \sum p_l = 1 \). Their intended meaning is that they evaluate to \( v_l \) with probability \( p_l \). The last parameter \( p_l \) can be omitted and in that case is set to \( p_l := 1 - \sum_{i=1}^{l-1} p_i \). For a program \( \mathcal{P} \) we denote with \( \text{Vars}(\mathcal{P}) \) the set of \( \mathcal{P} \)’s variables appearing in the body of a while-loop in \( \mathcal{P} \)’s loop-body. The programs of Figures 1-2 are examples of our program model defined in Figure 3. In comparison to the probabilistic Guarded Command Language (pGCL) [Barthe et al. 2020], programs of our model contain exactly one while-loop, no non-determinism \(^1\) but support continuous distributions and simultaneous assignments.

### 3.2 Program Semantics

In what follows we define the semantics of probabilistic programs in terms of Markov chains on a measurable space. We then introduce the notion of normalized probabilistic loops by means of so-called \( \mathcal{P} \)-preserving program transformations (Section 3.3).

**Definition 3 (State & Run Space).** Let \( \mathcal{P} \) be a probabilistic program with \( m \) variables. We denote by \( \text{ND}(\mathcal{P}) \) the non-probabilistic program obtained from \( \mathcal{P} \) by replacing every probabilistic choice \( C \) in \( \mathcal{P} \) by a non-deterministic choice over \( \text{supp}(C) \). Let \( \text{States}^{\mathcal{P}} \subseteq \mathbb{R}^m \) be the set of program states of \( \text{ND}(\mathcal{P}) \) reachable from any initial state. The state space of \( \mathcal{P} \) is the measurable space \( (\text{States}^{\mathcal{P}}, S_{\mathcal{P}}) \), where \( S_{\mathcal{P}} \) is the Borel \( \sigma \)-algebra on \( \mathbb{R}^m \) restricted to \( \text{States}^{\mathcal{P}} \). The run space of \( \mathcal{P} \) is the sequence space \( (\text{States}^{\mathcal{P}}, S_{\mathcal{P}})^\omega = (\text{Runs}^{\mathcal{P}}, \mathcal{R}_{\mathcal{P}}) \).

In what follows, we omit the subscript \( \mathcal{P} \) whenever the program \( \mathcal{P} \) is irrelevant or clear from the context. Executions/runs of a probabilistic program \( \mathcal{P} \) define a stochastic process, as follows.

**Definition 4 (Run Process).** Let \( \mathcal{P} \) be a probabilistic program with \( m \) variables. The run process \( \Phi_{\mathcal{P}} : \text{Runs} \rightarrow \text{States} \) is a stochastic process in the run space mapping a program run to its \( n \)th state, that means, \( \Phi_{\mathcal{P}}(\text{run}) := \text{run}_n \).

For program variable \( x \) with index \( i \geq 1 \), we denote by \( x_{\mathcal{P}} \) the projection of \( \Phi_{\mathcal{P}} \) to its \( i \)th component \( \Phi_{\mathcal{P}}(\cdot) (i) \). Given an arithmetic expression \( A \) over \( \mathcal{P} \)’s variables, we write \( A_{\mathcal{P}} \) for the stochastic process where every program variable \( x \) in \( A \) is replaced by \( x_{\mathcal{P}} \).

**Remark.** Given an initial distribution of program states \( \mu \) and a Markov kernel \( p \) defined according to the standard meaning of the program statements, by Kolmogorov’s Extension Theorem we conclude that there is a unique probability measure \( \mathbb{P}_{\mathcal{P}} \) on \( (\text{Runs}^{\mathcal{P}}, \mathcal{R}_{\mathcal{P}}) \) such that the run process is a Markov chain. \( (\text{Runs}^{\mathcal{P}}, \mathcal{R}_{\mathcal{P}}, \mathbb{P}_{\mathcal{P}}) \) is the probability space associated to program \( \mathcal{P} \). Distributions and (higher) moments of \( \mathcal{P} \)’s variables are to be understood with respect to this probability space.

For probabilistic loops according to the syntax in Figure 3, the initial distribution \( \mu \) of values of loop variables is the distribution of states after the statements \( \langle \text{statement} \rangle \) just before the while-loop.

\(^1\)Non-determinism is different from probabilistic choice. Demonic (angelic) non-determinism is concerned with the worst-case (best-case) behavior. For instance, a variable can be assigned to 0 or 1 both with probability 1/2. This is different from assigning 0 or 1 non-deterministically, where the probability is not specified.
Moreover, the loop body in Figure 3 is considered to be atomic, meaning the Markov kernel \( p \) describes the transition between full iterations in contrast to single statements.

### 3.3 \( \mathcal{P} \)-Preserving Transformations

Our probabilistic programs defined by the grammar in Figure 3 support rich arithmetic and complex probabilistic behavior/distributions. Such an expressivity of Figure 3 comes at the cost of turning the analysis of programs defined by Figure 3 cumbersome. In this section, we address this difficulty and introduce a number of program transformations that allow us to simplify our probabilistic programs to a so-called normal form while preserving the joint distribution of program variables. Normal forms allow us to extract recursive properties from the program, which we will later use to compute moments for program variables (Section 5).

**Schemas and Unification.** The program transformations we introduce in this section build on the notion of schemas and program parts. A program part is an empty word or any word resulting from any non-terminal of the grammar in Figure 3. For our purposes, a schema \( S \) is a program part with some subtrees in the program part’s syntax tree being replaced by placeholder symbols \( s_1, \ldots, s_l \). A substitution is a finite mapping \( \sigma = \{ s_1 \mapsto p_1, \ldots, s_l \mapsto p_l \} \) where \( p_1, \ldots, p_l \) are program parts. We denote by \( S[\sigma] \) the program part resulting from \( S \) by replacing every \( s_i \) by \( p_i \), assuming \( S[\sigma] \) is well-formed. For two schemas \( S_1 \) and \( S_2 \) a substitution \( u \) such that \( S_1[u] = S_2[u] \) is called a unifier (with respect to \( S_1 \) and \( S_2 \)). In this case \( S_1 \) and \( S_2 \) are called unifiable (by \( u \)).

**Transformations.** In what follows, we consider \( \mathcal{P} \) to be a fixed probabilistic program defined by Figure 3 and give all definitions relative to \( \mathcal{P} \). A transformation \( T \) is a mapping from program parts to program parts with respect to a schema \( \text{Old} \). \( T \) is applicable to a subprogram \( S \) of \( \mathcal{P} \) if \( S \) and \( \text{Old} \) are unifiable by the unifier \( u \). Then, the transformed subprogram is defined as \( T(S) := \text{New}[u] \) where \( \text{New} \) is a schema depending on \( \text{Old} \) and \( u \). A transformation is fully specified by defining how \( \text{New} \) results from \( \text{Old} \) and \( u \). We write \( T(\mathcal{P}, S) \) for the program resulting from \( \mathcal{P} \) by replacing the subprogram \( S \) of \( \mathcal{P} \) by \( T(S) \).

The first transformation we consider removes simultaneous assignments from \( \mathcal{P} \). For this, we store a copy of each assignment in an auxiliary variable to preserve the values used for simultaneous assignments, in case an assigned variable appears in an assignment expression. Variables are then assigned their intended value.
Definition 5 (Simultaneous Assignment Transformation). A simultaneous assignment transformation is the transformation defined by

\[ x_1, \ldots, x_l = \hat{v}_1, \ldots, \hat{v}_l \mapsto t_1 = \hat{v}_1; \ldots; t_i = \hat{v}_i; x_i = t_i; \ldots; x_l = t_l, \]

where \( t_1, \ldots, t_l \) are fresh variables.

In what follows, we assume that parameters of common distributions used in programs are constant. Nevertheless, the following transformation enables the use of some non-constant distribution parameters.

Definition 6 (Distribution Transformation). A distribution transformation is a transformation defined by either of the mappings

\[ \begin{align*}
&\hat{x} = \text{Normal}(\hat{\rho}, \hat{d}) \quad \mapsto \quad t = \text{Normal}(0, \hat{d}); \hat{x} = \hat{\rho} + t \\
&\hat{x} = \text{Uniform}(\hat{\rho}_1, \hat{\rho}_2) \quad \mapsto \quad t = \text{Uniform}(0, 1); \hat{x} = \hat{\rho}_1 + (\hat{\rho}_2 - \hat{\rho}_1)t \\
&\hat{x} = \text{Laplace}(\hat{\rho}, \hat{b}) \quad \mapsto \quad t = \text{Laplace}(0, \hat{b}); \hat{x} = \hat{\rho} + t \\
&\hat{x} = \text{Exponential}(\hat{c}, \hat{\rho}) \quad \mapsto \quad t = \text{Exponential}(\hat{c}); \hat{x} = \hat{\rho}t
\end{align*} \]

where, for every mapping, \( t \) is a fresh variable.

Example 1. Consider Figure 1. The simultaneous assignment \( l.g = \text{Laplace}(x+y, 1)\) can be transformed using the transformation rules from Definitions 5-6 as follows:

\[ \begin{align*}
t_1 &= \text{Laplace}(x+y, 1) \\
t_2 &= \text{Normal}(0, 1) \\
x &= t_1 \\
y &= t_2
\end{align*} \]

\[ \begin{align*}
t_3 &= \text{Laplace}(0, 1) \\
t_1 &= x + y + t_3 \\
x &= t_1 \\
y &= t_2
\end{align*} \]

To simplify the structure of probabilistic loops, we assume else if branches to be syntactic sugar for nested if else statements. We remove else by splitting it into if-statements (if \( C \) and if not \( C \)). Since variables in \( C \) could be changed within the first branch, we store their original values in auxiliary variables and use those for the condition \( C’ \) of the second if statement. We capture this transformation in the following definition.

Definition 7 (Else Transformation). An else transformation is the transformation

\[ \begin{align*}
\text{if } \hat{C} : & \quad \text{Branch}_1 \\
\text{else: } & \quad \text{Branch}_2 \end{align*} \]

\[ \Rightarrow \]

\[ \begin{align*}
t_1 &= x_1; \ldots; t_l = x_l \\
\text{if } \hat{C} : & \quad \text{Branch}_1 \end{align*} \]

\[ \begin{align*}
\text{if not } C' : & \quad \text{Branch}_2 \end{align*} \]

where \( x_1, \ldots, x_l \) are all variables appearing in \( \hat{C}[u] \) which are also being assigned in \( \text{Branch}_1[u] \). Every \( t_i \) is a fresh variable and \( C’ \) results from \( \hat{C}[u] \) by substituting every \( x_i \) with \( t_i \).

To further simplify the loop body into a flattened list of assignments, we equip every assignment \( a \) of form “\( x = \text{value} \)” with a condition \( C_a \) (initialized to true \( \top \)) and a default variable \( d_a \) (initialized to \( x \)), written as “\( x = \text{value} \ [C_a] d_a \)”. The semantics of the conditioned assignment is that \( x \) is assigned value if \( C_a \) holds just before the assignment and \( d_a \) otherwise. With conditioned assignments, the loop body’s structure can be flattened using the following transformation.

Definition 8 (If Transformation). An if transformation is the transformation defined by

\[ \begin{align*}
\text{if } \hat{C}_1 : & \quad t = \hat{x} \\
\hat{x} = \hat{\delta} \ [\hat{C}_2] \hat{x} , \ u \mapsto & \quad \hat{x} = \hat{\delta} \ [\hat{C}_1 \text{ and } \hat{C}_2] \hat{x} \\
\text{Rest} \end{align*} \]

\[ \begin{align*}
\text{if } C : & \quad \text{Rest} \end{align*} \]

where \( t \) is a fresh variable and \( C \) results from \( \hat{C}_1[u] \) by substituting \( \hat{x}[u] \) by \( t \). If \( \text{Rest}[u] \) is empty, the line if \( C : \text{Rest} \) end is omitted from the result. If \( \hat{x}[u] \) does not appear in \( \hat{C}_1[u] \) the line \( t = \hat{x} \) is dropped.
Example 2. The following program containing nested if-statements can be flattened as follows:

```plaintext
if x == 1:
    x = Bernoulli(1/2)
if x == 0: y = 1 end
end

if x == 1:
    x = Bernoulli(1/2)
y = 1 [x == 0] y
end
```

To bring further simplicity to our program \( P \), we ensure for each variable to be modified only once within the loop body. To remove duplicate assignments we introduce new variables \( x_1, \ldots, x_{l-1} \) to store intermediate states. Assignments to other variables, in between the updates of \( x \), will be adjusted to refer to the latest \( x_i \) instead of \( x \).

**Definition 9 (Multi-Assignment Transformation).** A multi-assignment transformation is the transformation defined by

\[
\dot{x} = \dot{v}_1 [\hat{C}_1] \dot{x} ; \text{Rest}_1;
\dot{x} = \dot{v}_2 [\hat{C}_2] \dot{x} ; \text{Rest}_2;
\ldots;
\dot{x} = \dot{v}_i [\hat{C}_1] \dot{x} ;
\]

where \( x_1, \ldots, x_{l-1} \) are fresh variables. For \( i \geq 2, v_i, C_i \) and Rest\( _i \) result from \( \dot{v}_i[u], \hat{C}_1[u] \) and Rest\( _1[u] \), respectively, by replacing \( \dot{x}[u] \) by \( x_{i-1} \).

Example 3. In the program of Figure 2, program line if \( x_1 == x_3 \): \( t_1 == 1 \) else: \( t_1 == 0 \) can be transformed using transformation rules from Definitions 7-9 as follows:

```plaintext
if x1 == x3: t1 = 1
if x1 != x3: t1 = 0
```

With program transformations defined, we can turn our attention to program properties. In particular, we show that our transformations of \( P \) do not change the joint distribution of \( P \)'s variables. Since our transformations may introduce new variables, we consider program equivalence with respect to program variables in order to ensure that the distribution of \( P \) is maintained/preserved by our transformations.

**Definition 10 (Program Equivalence).** Let \( P_1 \) and \( P_2 \) be two probabilistic programs. We define \( P_1 \) and \( P_2 \) to be equivalent with respect to a set of program variables \( X \), in symbols \( P_1 \equiv_X P_2 \), if:

1. \( X \subseteq \text{Vars}(P_1) \cap \text{Vars}(P_2) \), and
2. the joint distributions of \( X \) arising from \( P_1 \) and \( P_2 \) are equal.

To relate a program \( P \) to its transformed version \( T(P, S) \) we consider the distribution of variables of the original program \( P \). If \( P \) retains the joint distribution of its variables after applying transformation \( T \), we say that \( T \) is \( P \)-preserving.
**Definition 11 (P-Preserving Transformation).** We say that a transformation \( T \) is \( P \)-preserving if \( \mathcal{P} \equiv^\text{Vars}(\mathcal{P}) T(\mathcal{P}, S) \) for all subprograms \( S \) of \( \mathcal{P} \) which are unifiable with \( \text{Old} \).

It is not hard to argue that the transformations defined above are \( P \)-preserving, yielding the following result.

**Lemma 2.** The transformations from Definitions 5-9 are \( P \)-preserving.

By exhaustively applying the \( P \)-preserving transformations of Definitions 5-9 over \( \mathcal{P} \), we obtain a so-called normalized program \( \mathcal{P}_N \), as defined below. The normalized \( \mathcal{P}_N \) will then further be used in computing higher moments of \( \mathcal{P} \) in Sections 5, as the \( \mathcal{P}_N \) preserves the moments of \( \mathcal{P} \) (Theorem 3).

**Definition 12 (Normal Form).** A program \( \mathcal{P} \) is in normal form or a normalized program if none of the transformations from Definitions 5-9 are applicable to \( \mathcal{P} \).

**Theorem 3 (Normal Form).** For every probabilistic program \( \mathcal{P} \) there is a \( \mathcal{P}_N \) in normal form such that \( \mathcal{P} \equiv^\text{Vars}(\mathcal{P}) \mathcal{P}_N \). Moreover, \( \mathcal{P}_N \) can be effectively computed from \( \mathcal{P} \) by exhaustively applying transformations from Definitions 5-9.

**Proof.** There are two claims in the theorem, which we need to address: (i) exhaustively applying transformation rules terminates (termination), and (ii) it preserves statistical properties of the (original) program variables (correctness).

For termination, we show that programs become smaller, in some sense, after every transformation. In particular, we consider the program size to be given by a tuple \((\text{Sim}, \text{Dist}, \text{Else}, \text{If}, \text{Multi}_B, \text{Multi}_I)\), representing the number of simultaneous assignments, non-trivial distributions, else statements, assignments within if branches (weighted for nested ifs), and number of variables with multiple assignments in the loop body and initialization part, respectively. With respect to the lexicographic order, each transformation reduces the size of the program which is lower-bounded by 0.

Correctness can be shown by treating each transformation separately and showing that it does not alter the variables’ distributions after a single application (Lemma 2). This is true for all transformations from Definitions 5-9. Auxiliary variables are used to store the original value to prevent intervening variable modifications. For Multi-Assignment Transformation (Definition 9), we also revise the rest of the assignments to reflect the change of the original variable. The Distribution Transformation (Definition 6) uses statistical properties of well-known distributions.

**Properties of Normalized Programs.** Figure 4 shows a normal form for the program from Figure 1. Normalized programs have the following important properties: (1) all distribution parameters are constant; (2) the loop body is a sequence of guarded assignments; (3) every program variable is only assigned once in the loop body. Moreover for every guarded assignment \( \mathbf{v} = \text{assign}_{\mathbf{true}} \left[ \mathbf{C} \right] \text{assign}_{\mathbf{false}} \), the guard \( \mathbf{C} \) is a boolean condition and \( \text{assign}_{\mathbf{false}} \) is a single variable which is assigned to \( \mathbf{v} \) if \( \mathbf{C} \) evaluates to \( \text{false} \). If \( \mathbf{C} \) evaluates to \( \text{true} \), the variable \( \mathbf{v} \) is assigned \( \text{assign}_{\mathbf{true}} \). The expression \( \text{assign}_{\mathbf{true}} \) is either a distribution with constant parameters or a probabilistic choice of polynomials as illustrated in Figure 4.

**Remark.** Based on the order in which transformations are applied to a program \( \mathcal{P} \) and the names used for auxiliary variables, several different normalized programs can be achieved for \( \mathcal{P} \). In this work, only the existence of a normal form is relevant. Moreover, from the definitions of the transformation, it is apparent that exhaustively applying them leads to a normal form whose size is linear in the size of the original program.
toggle = 0; sum = 0
x = 1; y = 1; z = 1

while ★:
    toggle = 1 - toggle
    x = 1 + x \{1/2\} 2 + x \{\text{toggle} == 0\} x
    y = y + z + x \{1/3\} - x + y + z \{\text{toggle} == 0\} y
    z = z + y \{1/4\} z - y \{\text{toggle} == 0\} z
    t1 = \text{Laplace}(0, 1)
    l = t1 + x + y
    g = \text{Normal}(0, 1)
    sum = sum + x \{g < 1/2\} sum

end

Fig. 4. A normal form for the program in Figure 1.

4 FINITE TYPES IN PROBABILISTIC PROGRAMS

Given a probabilistic program \( P \) in our programming model, the transformations of Section 3.3 simplify \( P \) by computing its normalized form while maintaining the distribution (and hence also moments) of \( P \). Nevertheless, the normalized form of \( P \) contains computationally expensive polynomial arithmetic, potentially hindering the automated analysis of \( P \) in Section 5 due to a computational blowup. Therefore, we introduce further simplifications for \( P \) by means of power reduction techniques.

Example 4. Consider Figure 1 and assume we are interested in the \( k \)th power of variable toggle and deriving the raw moment \( E(toggle_n^k) \). Our analysis relies on replacing variables with their assignments (see Section 5), leading to the expression \((1 - toggle_n) \cdot 1^{k+1} \) monomials:

\[
(1 - toggle_n)^k = \sum_{i=0}^{k} \binom{k}{i} (-1)^i toggle_{n-1}^i
\]

Higher moments, together with the aforementioned replacements, may lead to blowups of the number of monomials to consider. However, observing that the variable toggle is binary, we have \( toggle_n^k = toggle_{n-1} \) for any \( k \geq 0 \). Arbitrary powers of the finite variable toggle with 2 possible values can be written in terms of powers smaller than 2. In the rest of this section, we show that this phenomenon generalizes from binary variables to arbitrary finite valued variables, thus simplifying the analysis of higher moments of finite valued program variables.

As defined in Definition 4, for an arithmetic expression \( X \) over program variables, \( X_n \) denotes the stochastic process mapping a program run to the value of \( X \) after iteration \( n \).

Definition 13 (Finite Expression). Let \( P \) be a probabilistic program and \( X \) an arithmetic expression over the variables of \( P \). We say that \( X \) is finite if there exist \( a_1, \ldots, a_m \in \mathbb{R} \) such that for all \( n \in \mathbb{N} : X_n \in \{a_1, \ldots, a_m\} \).

4.1 Power Reduction for Finite Types

As established in [Bartocci et al. 2020a, Lemma 1], high powers \( k \) of a random variable \( X \) over a finite set can be reduced. We adapt their result to our setting as follows.
Theorem 4 (Finite Power Reduction). Let \( m, k \in \mathbb{Z} \) and \( X \) be a discrete random variable over \( A = \{a_1, \ldots, a_m\} \). Then we can rewrite \( X^k \) as a linear combination of \( 1, X, X^2, \ldots, X^{m-1} \). Furthermore,
\[
X^k = a^k M^{-1} \overline{X},
\]
where \( a^k = (a_1^k, \ldots, a_m^k) \), \( M \) is an \( m \times m \) matrix with \( M_{ij} = a_j^{-1} \) (with \( 0^0 := 1 \)), and \( \overline{X} = (X^0, \ldots, X^{m-1})^T \).

In other words, Theorem 4 implies that any higher moment of \( X \), can be computed from just its first \( m-1 \) moments. Furthermore, we build on Theorem 4 and establish the inverse of matrix \( M \) explicitly \((M \text{ is explicit in Theorem 4 but its inverse is implicit})\).

Theorem 5 (Reduction Formula). Recall that the \( k \)th elementary symmetric polynomial with respect to a set \( V = \{v_1, \ldots, v_n\} \) is \( e_k(V) = \sum_{1 \leq j_1 < \cdots < j_k \leq n} v_{j_1} \cdots v_{j_k} \) and let \( A_- = A \setminus \{a_j\} \). Then the inverse of \( M \) in (3) is given by
\[
M^{-1}_{ij} = \frac{(-1)^i e_{m-i}(A_-)}{\prod_{a \in A_+} (a - a_i)}.
\]

Proof. Let \( MN = B \) for \( M \) as of (3) and \( N \) as of (4). We show that \( B = I \) by showing that \( B_{ij} = 1 \) if \( i = j \) and \( B_{ij} = 0 \) otherwise. We have
\[
B_{ij} = \sum_{1 \leq k \leq m} N_{ik} M_{kj} = \sum_{1 \leq k \leq m} \frac{(-1)^k e_{m-k}(A_-)}{\prod_{a \in A_+} (a - a_i)} a_j^{k-1}
\]

\[
eq \frac{1}{\prod_{a \in A_+} (a - a_i)} \sum_{1 \leq k \leq m} (-1)^k a_j^{k-1} e_{m-k}(A_-)
\]

\[
eq \frac{1}{\prod_{a \in A_+} (a - a_i)} \sum_{0 \leq k \leq m-1} (-a_j)^k e_{m-k-1}(A_-)
\]

\[
eq \frac{1}{\prod_{a \in A_+} (a - a_i)} \prod_{a \in A_+} (a - a_j),
\]

where the last equation comes from the expansion of product \( \prod_{a \in A_+} (a - a_j) \) and grouping by the exponent of \( a_j \). We can clearly see that the last expression is 1 if \( i = j \) and 0 otherwise. \( \Box \)

Example 5. Let \( X \) be a random variable over \( A := \{-2, 0, 1, 3\} \). Using Theorem 4-5, we obtain the 10th power of \( X \) as:

\[
X^{10} = \begin{pmatrix}
0 & -1/10 & 2/5 & -1/30 \\
1 & -5/6 & -1/3 & 1/6 \\
0 & 1 & 1/6 & -1/6 \\
0 & -1/15 & 1/30 & 1/30
\end{pmatrix}
\begin{pmatrix}
X^0 \\
X^1 \\
X^2 \\
X^3
\end{pmatrix}
\]

\[
= 1934X^3 + 2105X^2 - 4038X.
\]

5 Computing Higher Moments of Probabilistic Programs

We now bring together the results from Sections 3-4 to develop the theory of moment-computability for probabilistic loops. We establish the technical details leading to sufficient conditions that ensure moment-computability, culminating in the proof of Theorem 6. The main ideas of our method are illustrated on the probabilistic loop from Figure 1 in Example 6 at the end of this section.

Definition 14 (Moment-Computability). A probabilistic loop \( \mathcal{P} \) is moment-computable if a closed-form (according to Theorem 1) of \( \mathbb{E}(x^k_n) \) exists and is computable for all \( x \in \text{Vars}(\mathcal{P}) \) and \( k \in \mathbb{N} \).
We will describe the class of moment-computable probabilistic loops through the properties of the dependencies between program variables.

**Definition 15 (Variable Dependency).** Let \( \mathcal{P} \) be a probabilistic loop and \( x, y \in \text{Vars}(\mathcal{P}) \). We define:

- \( y \) depends conditionally on \( x \), if there is an assignment of \( y \) within an if-else-statement and \( x \) appears in the if-condition.
- \( y \) depends finitely on \( x \), if \( x \) is finite and appears in an assignment of \( y \).
- \( y \) depends linearly on \( x \), if \( x \) appears only linearly in every assignment of \( y \).
- \( y \) depends polynomially on \( x \), if there is an assignment of \( y \) in which \( x \) appears non-linearly and \( x \) is not finite (motivated by Section 4.1).
- \( y \) depends on \( x \) if it depends on \( x \) conditionally, finitely, linearly, or polynomially.

Furthermore, we consider the transitive closure for variable dependency as follows: If \( z \) depends on \( y \) and \( y \) depends on \( x \), then \( z \) depends on \( x \). If one of the two dependencies is polynomial, then \( z \) depends polynomially on \( x \).

A crucial point to highlight in Definition 15 is that due to transitivity, variables can depend on themselves. For instance, if variable \( x \) depends on \( y \) and \( y \) on \( x \), then \( x \) is self-dependent. Moreover, if either of the dependencies between \( x \) and \( y \) is non-linear, \( x \) depends, by Definition 15, polynomially on itself. The absence of such polynomial self-dependencies is a central condition for our notion of moment-computable loops.

**Theorem 6 (Moment-Computability).** A probabilistic loop \( \mathcal{P} \) is moment-computable if (1) none of its non-finite variables depends on itself polynomially, and (2) if the variables in all if-conditions are finite.

Note that none of the program transformations from Section 3.3 can introduce a polynomial (self-)dependence. We capture this in the following lemma:

**Lemma 7 (Non-Dependency Preservation).** If a variable \( x \in \text{Vars}(\mathcal{P}) \) does not depend on itself polynomially, neither does \( x \in \text{Vars}(\mathcal{P}_N) \).

Before we prove Theorem 6, let us first show its validity for programs in normal form (as defined in Section 3.3). Recall that a normalized program’s loop body is a flat list of (guarded) assignments, one for every (possibly auxiliary) program variable.

**Lemma 8 (Normal Moment-Computability).** The Moment-Computability Theorem (Theorem 6) holds for loops in normal form.

**Proof.** We have to show that for an arbitrary normalized program \( \mathcal{P} \) satisfying the conditions of Theorem 6, all \( x \in \text{Vars}(\mathcal{P}) \) and all \( k, n \in \mathbb{N} \), the \( k \)th moment of \( x \) (that is \( \mathbb{E}(x^k) \)) admits a closed-form as an exponential polynomial. \( \mathbb{E}(x^k) \) admits a closed-form as an exponential polynomial in \( n \) if it satisfies a linear recurrence. We show a slightly more general statement. That is, we show that for any monomial of program variables \( M \) (and hence also for \( x^k \) \( \mathbb{E}(M) \)) satisfies a linear recurrence. The main idea of the proof is to show that \( \mathbb{E}(M) \) only depends on a finite set of monomials, each (in some sense) not larger than \( M \) itself. Intuitively, the finite set of monomials on which \( \mathbb{E}(M) \) depends on are all monomials of program variables such that their expected values determine \( \mathbb{E}(M) \). We will show that this set of monomials exists, is finite, and leads to a system of linear recurrences containing \( \mathbb{E}(M) \), implying a computable exponential polynomial closed-form for \( \mathbb{E}(M) \) by Theorem 1.

Let \( \mathcal{P} \) be a normalized program satisfying the conditions of Theorem 6, \( x \in \text{Vars}(\mathcal{P}) \), \( k, n \in \mathbb{N} \) arbitrary, and \( M \) the set of all monomials over \( \text{Vars}(\mathcal{P}) \) with the powers of every finite variable \( d \) bounded by the number of possible values of \( d \) (higher powers can be reduced as of Theorem 4).
Recurrences over Moments. Given the syntax of probabilistic programs and properties of expectation, for any monomial $M \in \mathcal{M}$ there is a natural way to express the expected value of $M$ in iteration $n+1$, that is $\mathbb{E}(M_{n+1})$, as a linear combination of expectations of monomials in iteration $n$:

$$\mathbb{E}(M_{n+1}) = \sum_{N \in M^*} c_N \mathbb{E}(N_n),$$

(5)

for some finite set $M^* \subseteq \mathcal{M}$, and non-zero constants $c_N$. Equation (5) is called the recurrence of $\mathbb{E}(M)$. The set $M^*$ is the set of monomials that appear in the recurrence of $\mathbb{E}(M)$. We define the operator to give such a set for any monomial and extend the definition to sets by

$$S^* = \bigcup_{M \in S} M^*.$$

The exact recurrence can be computed from $\mathbb{E}(M_{n+1})$ by replacing variables appearing in $M$ by their assignments and using the linearity of $\mathbb{E}$ to convert expected values of polynomials to linear combinations of expected monomials. Recall that for a program in normal form, every program variable is only assigned once, all distribution parameters are constant, and the loop body is a flat list of guarded assignments (Section 3.3). The guarded assignments are of the form

$$x = a_0 \{p_0\} \ldots \{p_{i-1}\} a_i [C_x] d_x,$$

for polynomials of program variables $a_0, \ldots, a_i$ and constant probabilities $p_0, \ldots, p_{i-1}$, or

$$x = \text{Dist}[C_x] d_x$$

for some admissible distribution $\text{Dist}$. The guard $[C_x]$ is a boolean condition and for normalized programs, $d_x$ is always a program variable. Assume, that the variable $x$ appears in the monomial $M$. Hence, $M = M' \cdot x_{n+1}^k$ for some monomial $M'$ not containing $x$. If the single assignment of $x$ is a (guarded) probabilistic choice of polynomials we rewrite $\mathbb{E}(M_{n+1})$ to

$$\mathbb{E}(M_{n+1}) = \mathbb{E}(M' \cdot x_{n+1}^k) = \mathbb{E} \left( M' \left( d_x [-C_x] + \sum p_i a_i^k [C_x] \right) \right).$$

(6)

If the single assignment of $x$ is a (guarded) draw from a distribution we rewrite $\mathbb{E}(M_{n+1})$ to

$$\mathbb{E}(M_{n+1}) = \mathbb{E}(M' \cdot x_{n+1}^k) = \mathbb{E} \left( M' d_x [-C_x] \right) + \mathbb{E} \left( M' [C_x] \right) \mathbb{E} \left( \text{Dist}^k \right).$$

(7)

Variables in conditions $[C_x]$ are all finite since they come from branch conditions. We further simplify the expressions of Equations (6)-(7) by replacing the logical conditions $[C_x]$ by polynomials that evaluate to 1 whenever variables satisfy the condition $[C_x]$ and to 0 otherwise. It is possible to write any logical condition over finitely valued variables as such a polynomial ([Stankovic et al. 2022]), with $[x = c] := \prod_{d \in \omega(x) \setminus \{c\}} \frac{x - d}{c - d}, [-C] = 1 - [C]$, and $[C_1 \land C_2] = [C_1] \cdot [C_2]$, where $\omega(x)$ is the set of possible values of $x$.

Converting conditions to polynomials in the equation above leads to polynomials over moments of program variables. We can compute the recurrence of $\mathbb{E}(M)$ in Equation (5) by replacing all variables in $M$ of iteration $n+1$ from last to first (by their appearance in $P$’s loop body). Throughout, the linearity of $\mathbb{E}$ is used to convert expected values of polynomials to linear combinations of expected monomials. In Example 6, we illustrate this computation on the program from Figure 1 with its normal form from Figure 4.

Because negation and conjunction are functionally complete for propositional logic, we can also handle disjunctions using De Morgan’s laws: $[P \lor Q] = [\neg (\neg P \land \neg Q)] = 1 - (1 - [P]) \cdot (1 - [Q])$. Inequalities can then be transformed into a disjunction of equalities since we assume that all variables appearing in if-conditions are finitely valued.
Ordering. Now that we can compute the recurrences, we need to introduce the order for monomials, such that the monomials appearing in the recurrence for $M$ are not larger than $M$. We will need this to show that computing the recurrences as described above is, indeed, a finite process. Intuitively, a variable $y$ is larger than (or equal to) $x$ if it depends on $x$. Mutually dependent variables will form an equivalence class. We then extend the order to monomials based on their degrees with respect to the variables’ equivalence classes.

More formally, let $\leq$ be a smallest total preorder on variables such that $x \leq y$ whenever $y$ depends on $x$. We write $x < y$ iff $x \leq y$ and $y \not< x$, and $x \sim y$ iff $x \leq y$ and $y \leq x$. Let $\bar{x}$ be the equivalence class of $x$ induced by $\sim$. Note, that all variables in an equivalence class are mutually dependent. However, because of our assumptions on the program $P$, the mutual dependencies among non-finite variables are all linear (as there are no polynomial self-dependencies).

We extend $\leq$ to a preorder on the set of monomials $M$. For every monomial $M$ and non-finite variable $x$, we consider the degree $\text{deg}(\bar{x}, M)$ of $M$ in the equivalence class of $\bar{x}$. We associate $M$ with the sequence of $\text{deg}(\bar{x}, M)$ for equivalence classes of all non-finite variables, ordered reverse-lexicographically with respect to $\leq$. Then the relations $<, \sim$, and equivalence classes ($\sim$) follow naturally from $\leq$.

By Theorem 4 and the definition of $\leq$, the equivalence class $\bar{M}$ is finite for each $M \in M$. Let $M_\sim$ be the set of equivalence classes of $\sim$. The preorder $\leq$ induces a partial order $\leq_\sim$ on $M_\sim$. Note that monomials only contain non-negative powers and a finite number of variables. These facts together with $\leq$ being total imply that $\leq_\sim$ is a well-order. We will write $\leq$ instead of $\leq_\sim$ when the meaning is clear from the context.

With these orders defined, we have $N \leq M$ for any $N \in M^*$ and $M \in M$. Intuitively, this means that for every monomial $M$, the monomials occurring in the recurrence of $M$ are not larger than $M$ itself. This is true because the order on variables is defined according to variable dependencies, the order on monomials is a reverse-lexicographic extension of the order on variables, and the fact that polynomial self-dependencies are not allowed by assumption.

Finite $S^{(Q)}$. We show, that for any monomial $Q$, there is a finite set of monomials $S^{(Q)} \subset M$ containing $Q$ such that $M^* \subset S^{(Q)}$ for any $M \in S^{(Q)}$. Because $M^*$ is the set of all monomials in the recurrence of $M$, this means that $S^{(Q)}$ contains all monomials necessary to construct a system of linear recurrences containing $E(Q)$ (if it exists and is finite).

Let

$$S^{(Q)} := Q \cup \bigcup_{A \in Q^*, A < Q} S^{(A)}.$$  

Clearly $Q \in S^{(Q)}$ and $M^* \subset S^{(Q)}$ for any $M \in S^{(Q)}$ by construction. We are left to show that $S^{(Q)}$ is finite for all $Q \in Q \in M_\sim$, which we can do by transfinite induction over $M_\sim$.

For the base case, we have to show that $S^{(Q)}$ is finite for all $Q$ in $\bar{1}$ (the trivial monomial). Since $\bar{1} = \{ \prod_{d \in F} d^{\lambda_d} \mid \lambda_d \leq m_d \}$, where $F$ is the set of finite program variables and $m_d$ are upper bounds on their powers as of Theorem 4, $S^{(Q)} = \bar{1}$ is finite for all $Q \in \bar{1}$. Suppose $S^{(A)}$ is finite for all $A < Q$. Since $Q^*$ is finite, so is the union in (8) and, as a result, $S^{(Q)}$.

Moments. A system of linear recurrences with constant coefficients can be constructed for monomials in $S^{(Q)}$. Therefore, the closed-form of any $E(M) \in M$ exists and is computable. \hfill \[ \square \]

We now turn back to Theorem 6 and establish its validity. The crux of our proof below relies on the fact that our transformations computing normal forms of $P$ (see Section 3.3) are $P$-preserving.

\footnote{$\text{deg}(\bar{x}, M) := \sum_{x \in \bar{x}} \text{deg}(x, M)$, where $\text{deg}(x, M)$ is the degree of $x$ in $M$.}
Proof (of Theorem 6). By Theorem 3 (Normal Form Termination), \( \mathcal{P} \) can be transformed to a normalized loop \( \mathcal{P}_N \). By Lemma 7 (Non-Dependency Preservation), \( \mathcal{P}_N \) satisfies all conditions from Lemma 8 (Normal Moment-Computability). Thus, \( \mathcal{P}_N \) is moment-computable and the moments are equivalent to those of \( \mathcal{P} \) for \( x \in \text{Vars}(\mathcal{P}) \) by Theorem 3 (Normal Form Correctness).

The proofs of Theorem 6 and Lemma 8 are constructive and describe a procedure to compute (higher) moments of program variables by (1) transforming a probabilistic loop into a normal form according to Theorem 3, (2) constructing a system of linear recurrences as in the proof of Lemma 8 and (3) solving the system of linear recurrences with constant coefficients. In the following example, we illustrate the whole procedure on the running example from Figure 1.

Example 6. We return to the probabilistic loop \( \mathcal{P} \) from Figure 1. A normal form \( \mathcal{P}_N \) for \( \mathcal{P} \) was given in Figure 4. To compute a closed-form of the expected value of the program variable \( z \), we will model \( \mathbb{E}(z_n) \) as a system of linear recurrences according to the proof of Lemma 8. For a cleaner presentation we will refer to the variable toggle by \( t \). Note that \( t \) is binary. To construct the recurrence for \( \mathbb{E}(z_n) \), we start with \( \mathbb{E}(z_{n+1}) \) at the assignment of \( z \) in \( \mathcal{P}_N \) and repeatedly replace variables by the right-hand side of their assignment starting from \( z \)'s assignment and stopping at the top of the loop body. Throughout the process we use the linearity of expectation to convert expected values of polynomials to linear combinations of expected monomials (as required by Equation 5):

\[
\begin{align*}
\mathbb{E}(z_{n+1}) \\
\downarrow \text{assignment of } z \\
\mathbb{E}((1-t_{n+1})(z_n+y_{n+1}) + (1-t_{n+1})(z_n-y_{n+1})) + t_{n+1}z_n \\
\downarrow \text{replace (in)equalities by polynomials} \\
\mathbb{E}((1-t_{n+1})(z_n+y_{n+1}) + (1-t_{n+1})(z_n-y_{n+1})) + t_{n+1}z_n \\
\downarrow \mathbb{E} \text{ on categorical} \\
\frac{1}{4} \mathbb{E}((1-t_{n+1})(z_n+y_{n+1})) + \frac{3}{4} \mathbb{E}((1-t_{n+1})(z_n-y_{n+1})) + \mathbb{E}(t_{n+1}z_n) \\
\downarrow \text{simplify; } \mathbb{E} \text{ linearity} \\
\mathbb{E}(z_n) + \frac{1}{2} \mathbb{E}(t_{n+1}y_{n+1}) - \frac{1}{2} \mathbb{E}(y_{n+1}) \\
\downarrow, \text{ similarly, replace } y_{n+1}, x_{n+1}, t_{n+1} \\
\mathbb{E}(z_n) - \frac{1}{6} \mathbb{E}(t_nx_n) - \frac{1}{2} \mathbb{E}(t_ny_n) - \frac{1}{6} \mathbb{E}(t_nx_n^2) + \frac{1}{12} \mathbb{E}(t_n) + \frac{1}{6} \mathbb{E}(t_nz_n).
\end{align*}
\]

The last line of the calculation represents the recurrence equation of the expected value of \( z \). For every monomial in the recurrence equation of \( \mathbb{E}(z_n) \) (that is: \( z^* = \{tx, ty, tx^2, t, tz\} \)), we compute its respective recurrence equation and recursively repeat this procedure which eventually terminates according to Lemma 8.

The ordering of variables and monomials, which is essential for proving termination in Lemma 8 comes from an ordering of variables \( t, x, y, z, l, g, \text{sum} \). Based on the program assignments, we need \( t \leq x < y \leq l; y \leq z; t \leq z \leq \text{sum} < y \leq \text{sum} < g \leq \text{sum} \). Variables \( y \) and \( z \) form an equivalence class, since \( y \leq z \) and \( z \leq y \). Notice that some variables are not ordered, for example \( t \) and \( g \). This gives some freedom in choosing the total preorder, any will work. Let us have \( t < g < x < \text{sum} < y \). This gives an order on the equivalence classes \( \{t\} < \{g\} < \{x\} < \{\text{sum}\} < \{y, z\} < \{l\} \). The ordering on monomials then considers the classes of non-finite program variables, i.e. all equivalence classes except \( \{t\} \). A monomial is then assigned a sequence of degrees with respect to each equivalence class,
e.g. \( z \rightarrow (0, 0, 0, 1, 0), tz \rightarrow (0, 0, 0, 1, 0), \) and \( x^2 \rightarrow (0, 0, 2, 0, 0), \) with monomials ordered reverse-lexicographically with respect to their sequence. For the monomials in this example: \( 1 \sim t < x \sim tx < x^2 \sim tx^2 < y \sim z \sim ty \sim tz. \)

After computing all necessary recurrence equations, we are faced with a system of linear recurrences of the expected values of the monomials \( z, tx, ty, x, y, tx^2, t, tz, x^2 \) and \( 1 \) with recurrence matrix

\[
\begin{bmatrix}
1 & -\frac{1}{6} & -\frac{1}{2} & 0 & 0 & -\frac{1}{6} & \frac{1}{12} & \frac{1}{6} & 0 & 0 \\
0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & \frac{3}{2} & 0 & 0 & 0 \\
0 & \frac{1}{2} & 0 & 0 & 1 & \frac{1}{2} & -\frac{1}{6} & -\frac{1}{2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
0 & 3 & 0 & 0 & 0 & 0 & \frac{5}{2} & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

Using any computer algebra system, such as sympy\(^4\), we arrive at the closed-form solution for the expected value of \( z, \mathbb{E}(z_n) \), parameterized by the loop counter \( n \):

\[
\mathbb{E}(z_n) = \frac{883}{32} + \frac{29n}{16} - \frac{201}{20} 2^{-n} \frac{6}{2} - \frac{67}{10} 2^{-n} \frac{14n}{2} - \frac{37}{20} 3^{\frac{1}{3}} 3^{-n} \frac{6}{2} \\
- \frac{37}{20} 3^{-n} \frac{14n}{2} - \frac{201}{20} 2^{-n} \frac{6}{2} \left( \frac{-1}{2} \right)^n - \frac{37}{10} 2^{\frac{2}{3}} \left( \frac{-1}{3} \right)^n + \frac{67}{20} \left( \frac{-1}{2} \right)^n \\
+ \frac{37}{20} 3^{-n} \frac{14n}{2} \left( \frac{-1}{3} \right)^n + \frac{9}{16} n (-1)^n + \frac{29}{32} (-1)^n + \frac{9}{16} n^2.
\]

This example highlights the strength of algebraic techniques for probabilistic program analysis in comparison to constraint-based methods employing templates. We are not aware of any template-based method able to handle functions of the complexity of \( \mathbb{E}(z_n) \). Our tool Polar is able to find the closed-form of \( \mathbb{E}(z_n) \) in under one second (see Section 7).

### 5.1 Guarded Loops

When we model a guarded probabilistic loop \textbf{while} \( \phi: \ldots \) as an infinite loop \textbf{while} \( \ast: \text{if } \phi: \ldots \), we impose the same restrictions on \( \phi \) as on if-conditions (that means \( \phi \) only contains finite variables) to guarantee computability and correctness. The \( k \)th moment of \( x \) after termination is then given by

\[
\lim_{n \to \infty} \mathbb{E}(x_n^k \mid -\phi_n) = \lim_{n \to \infty} \frac{\mathbb{E}(x_n^k \cdot [-\phi_n])}{\mathbb{E}([-\phi_n])}.
\]

If the limit exists, we can use standard methods from computer algebra to compute it, as the (higher) moments our approach computes are given as exponential polynomials [Gruntz 1996].

\(^4\)https://www.sympy.org
Example 7. Consider the following loop, in which \( x \) after termination is geometrically distributed with parameter \( 1/2 \):

\[
x, \text{ stop} = 0, 0 \\
\text{while} \; \text{stop} == 0: \\
\quad \text{stop} = \text{Bernoulli}(1/2) \\
\quad x = x + 1 \\
\end{while}
\]

With Equation 9 and the techniques from this section we get:

\[
\mathbb{E}(x) = \lim_{n \to \infty} \mathbb{E}(x_n \mid \text{stop}_n = 1) = \lim_{n \to \infty} \frac{\mathbb{E}(x_n \cdot \text{stop}_n)}{\mathbb{E}(\text{stop}_n)} = \lim_{n \to \infty} \frac{-n2^{-n} - 2^{1-n} + 2}{1 - 2^{-n}} = 2.
\]

Moreover, whenever the variables in the loop guard are not probabilistic, traditional techniques can be applied to determine the number of loop iterations \( n \) which can then be plugged into the (higher) moments computed by our work. Apart from the guarded loops, many systems show the type of infinite behavior naturally modeled with infinite loops, such as probabilistic protocols or dynamical systems.

5.2 Infinite If-Conditions

Theorem 6 on moment-computability requires the variables in all if-conditions to be finite. Nevertheless, in some cases, if-conditions containing infinite variables can be handled by our approach. Let \( P \) be a probabilistic loop containing an if-statement with condition \( F \) and \( I \) where \( F \) contains only finite variables and \( I \) contains infinite variables. Without loss of generality, no variable in \( I \) is assigned in or after the if-statement. Let the transformation removing \( I \) be defined by

\[
\text{if } F \text{ and } I : \text{Branch} \\ t = \text{Bernoulli}(p) \\ \text{if } F \text{ and } t == 1: \\ \text{Branch} \end{if}
\]

where \( t \) is a fresh variable and \( p := \mathbb{P}(I) \) (potentially symbolic). Then, the transformation preserves the distributions of all \( x \in \text{Vars}(P) \) under the following assumptions:

1. \( I \) is iteration independent, meaning for every variable \( x \) in \( I \) neither \( x \) nor any variable \( x \) depends on (as of Definition 15) has a self-dependency.
2. \( I \) is statistically independent from \( F \) and all conditions in \( \text{Branch} \).
3. For every assignment \( A \) in \( \text{Branch} \) and every variable \( x \) in \( A \) which has been assigned before \( A \), it holds that \( I \) and \( x \) are statistically independent.

Assumption 1 ensures that \( \mathbb{P}(I) = \mathbb{E}([I]) \) is constant. Assumption 2 and 3 further ensure that \( \mathbb{E}([I]) \) can always be "pulled out" (that means \( \mathbb{E}([I]x) = \mathbb{E}([I])\mathbb{E}(x) \)) in the construction of the recurrences. Assumptions 1-3 can often be checked automatically.

Example 8. Consider the statement \( \text{if } g < 1/2: \; \text{sum} = \text{sum} + x \) of the program from Figure 1, where the value of \( g \) is drawn from a standard normal distribution. In this case, the transformation’s parameter \( p \) represents \( \mathbb{P}((\text{Normal}(0, 1) < 1/2) \) but is left symbolic for the moment computation. The integral \( \mathbb{P}((\text{Normal}(0, 1) < 1/2) \) can be solved separately and the result be substituted for \( p \).

5.3 On the Necessity of the Conditions Ensuring Moment-Computability

Theorem 6 states two conditions that are sufficient to ensure that the closed-forms of the program variables’ higher moments always exist and are computable. Condition 1 enforces that there is no variable with potentially infinite values with a polynomial self-dependency. Condition 2 demands that all variables appearing in if-conditions are finite. Our approach for computing the moments of
variables of probabilistic loops can handle precisely the programs that satisfy these two conditions. We argue that both conditions are necessary, in the sense that if either of the conditions does not hold, the existence or computability of the variable moments’ closed-forms as exponential polynomials cannot be guaranteed for all programs from our program model when one or both conditions are removed from Theorem 6.

**Condition 1.** Relaxing condition 1 of Theorem 6 means that we allow for polynomial self-dependencies of non-finite variables. The logistic map [May 1976] is a quadratic first-order recurrence defined by $x_{n+1} = r \cdot x_n (1 - x_n)$ and well-known for its chaotic behavior. A famous fact about the logistic map is that it does not have an analytical solution for most values of $r$ [Maritz 2020]. By neglecting condition 1, we can easily devise a loop modeling the logistic map:

```plaintext
while ★:
  x = r \cdot x (1 - x)
end
```

The value of the program variable $x$ after iteration $n$ is equal to the $n$th term of the logistic map. This means, for most values of $r$ and initial values of $x$, there does not exist an analytical closed-form solution for the program variable $x$. Moreover, our counter-example illustrates that condition 1 is necessary already for programs with a single variable and without stochasticity and if-statements.

**Condition 2.** Loosening condition 2 of Theorem 6 and allowing for non-finite variables in if-conditions renders our programming model Turing-complete. Intuitively, one can model a Turing machine’s tape with two variables $l$ and $r$ such that the binary representation of $l$ represents the tape’s content left of the read-write-head. The binary representation of $r$ represents the tape’s content at the position of the read-write-head and towards the right. The least significant bit of $r$ is the current symbol the Turing machine is reading. We can extract the least significant bit of $r$ in our programming model (and neglecting condition 2) by introducing a variable $\text{lsb}$ and using a single if-statement involving non-finite variables: whenever the loop changes the value of $r$, we set $\text{lsb} := r$. The while-loop’s body is of the form “if $\text{lsb} > 1$: $\text{lsb}=\text{lsb} - 2$ else transitions end”. The Turing machine’s transition table can be encoded using if-statements. Writing and shifting can be accommodated for by multiplying by 2 or $1/2$ and using addition and subtraction. The Turing machine’s state can be modelled by a single finite variable. Therefore, by dropping condition 2, being able to model the program variables by linear recurrences would give rise to a decision procedure for the Halting problem: assume we introduce a variable $\text{terminated}$ which is initialized to 0 before the loop and set to 1 whenever the Turing-machine terminates. If $\text{terminated}$ can be modelled by a linear recurrence of order $k$, it suffices to check the first $k$ values of the recurrence to determine whether or not $\text{terminated}$ is always 0 [Kauers and Paule 2011] and the Turing-machine does not terminate. As the Halting problem is well-known to be undecidable, condition 2 is necessary to guarantee that the program variables can be modelled by linear recurrences, even without stochasticity and polynomial arithmetic.

**Remark (Sequential & Nested Loops).** Our program model consists of single non-nested loops. Sequential loops can be analyzed one by one with the same techniques as presented in this section. For nested loops, one could design a program transformation transforming a nested loop into a non-nested loop and then apply the techniques presented in this section. Alternatively, we conjecture that the approach presented for guarded loops (Section 5.1) could be used to first compute the moments of the most inner loops and then use the obtained information to compute the moments of the outer loops. The main challenge lies in ensuring the moment-computability conditions for the outer loops (Theorem 6) once the inner loops have been analyzed.
6 USE-CASES OF HIGHER MOMENTS

For probabilistic loops, computing closed-forms of the variables’ (higher) moments poses a technique for synthesizing quantitative invariants: Given a program variable $x$ and a closed-form $f(n)$ of its $k$th moment, the equation $\mathbb{E}(x^k) = f(n) = 0$ is an invariant. Moreover, closed-forms of raw moments can be converted into closed-forms of central moments, such as variance, skewness or kurtosis (cf. Section 2.1). In addition, this section provides hints on two further use-cases of higher moments of probabilistic loops: (i) deriving tail probabilities (Section 6.1) and (ii) inferring distributions of random variables from their moments (Section 6.2).

6.1 From Moments to Tail Probabilities

Tail probabilities measure the probability that a random variable surpasses some value. The mathematical literature contains several inequalities providing upper- or lower bounds on tail probabilities of random variables from their moments (Section 6.2).

**Theorem 9 (Markov’s Inequality).** Let $X$ be a non-negative random variable, and $t \geq 0$, then

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X^k)}{t^k}.$$ 

**Theorem 10 (Paley-Zygmund Inequality).** Let $X$ be a random variable with $X \geq t$ almost-surely. Then

$$\mathbb{P}(X > t) \geq \frac{\mathbb{E}(X^2) - \mathbb{E}(X)^2}{t^2}.$$ 

**Example 9.** For Herman’s Self-Stabilization program from Figure 2 almost-surely tokens $\in \{0, 1, 2, 3\}$. With the techniques from previous sections, we can compute the first two moments $\mathbb{E}(\text{tokens}_n) = 1 + 2 \cdot 4^{-n}$ and $\mathbb{E}(\text{tokens}_n^2) = 1 + 8 \cdot 4^{-n}$. Markov’s inequality (Theorem 9) gives us the upper bound $\mathbb{P}(\text{tokens}_n \geq 2) \leq \frac{1}{2} + 4^{-n}$ using the first moment and $\mathbb{P}(\text{tokens}_n \geq 2) \leq \frac{1}{4} + 2 \cdot 4^{-n}$ utilizing the second moment.

For the Paley-Zygmund inequality (Theorem 10) both the first and the second moment are required, yielding the lower bound $\mathbb{P}(\text{tokens}_n \geq 2) = \mathbb{P}(\text{tokens}_n > 1) \geq 4^{-n}$. The theorem’s precondition that almost-surely tokens $\geq 1$ might not be apparent at first sight. We take this for granted for now and will clarify this fact in Example 10.

Markov’s inequality and the Paley-Zygmund inequality are just two examples showing that our technique for moment computation can be leveraged for further program analysis using known results from probability theory. Our approach computes the exact moments of variables in probabilistic loops instead of just approximations or bounds on moments. This enables our technique to be readily combined with results from probability theory that require exact moments.

6.2 From Moments to Distributions

For finite random variables, their full distribution can be recovered from finitely many moments. More precisely, given a random variable $X$ with $m$ possible values, the distribution of $X$ can be recovered from its first $m-1$ moments, as the following theorem states:

**Theorem 11.** Let $X$ be a random variable over $\{a_1, \ldots, a_m\}$ and $p_i := P(X = a_i)$. The values $p_i$ are the solutions of the system of linear equations given by $\sum_{i=1}^{m} p_i a_i^j = \mathbb{E}(X^j)$ for $0 \leq j < m$.

**Proof.** There are $m$ unknowns $p_i$ for $1 \leq p_i \leq m$. Note that all $a_i$ are constant and that the first $m-1$ raw moments of $X$ are fixed. Using the definition of raw moments, we get $m$ linear equations.
Fig. 5. The empirical density of program variable $x$ for the benchmark *Bimodal* (cf. Table 1) and loop iterations 10 (left) and 100 (right) obtained by $10^5$ samples, together with two approximations using the Gram-Charlier A Series with 6 (red dashed lines) and 12 (blue solid lines) moments.

$$\sum_{i=1}^{m} p_i a^j_i = \mathbb{E}(X^j) \text{ for } 0 \leq j < m.$$ The solutions of the system of $m$ linear equations are the values $p_i$ that determine the distribution of $X$. \hfill \Box

**Example 10.** Consider Herman’s Self-Stabilization program from Figure 2. In Example 9 we obtained upper and lower bounds for tail probabilities of the tokens variable using the first one or two moments. With the first three moments we can fully recover the distribution of the tokens variable. We have that tokens $\in \{0, 1, 2, 3\}$. Let $p_i := \mathbb{P}(\text{tokens}_n = i)$ for $0 \leq i \leq 3$. By Theorem 11, we get the following system:

\begin{align*}
p_0 + p_1 + p_2 + p_3 &= \mathbb{E}(\text{tokens}_n^0) = 1, \\
p_1 + 2p_2 + 3p_3 &= \mathbb{E}(\text{tokens}_n) = 1 + 2 \cdot 4^{-n}, \\
p_1 + 4p_2 + 9p_3 &= \mathbb{E}(\text{tokens}_n^2) = 1 + 8 \cdot 4^{-n}, \\
p_1 + 8p_2 + 27p_3 &= \mathbb{E}(\text{tokens}_n^3) = 1 + 26 \cdot 4^{-n}.
\end{align*}

The solution can be obtained using standard techniques and tools, yielding $p_0 = 0; p_1 = 1 - 4^{-n}; p_2 = 0; p_3 = 4^{-n}$.

Note that probabilities are given as functions of the loop iteration $n$. Moreover, the solution shows that almost-surely tokens $\geq 1$, which we assumed to be true in Example 9.

The distributions of program variables with potentially infinitely many values, including continuous variables, cannot be, in general, fully reconstructed from finitely many moments. However, expansions such as the *Gram–Charlier A Series* [Kolassa 2006] can be used to approximate a probability density function using finitely many moments. Figure 5 illustrates how exact moments computed by our approach can be used to approximate unknown probability density functions of program variables. While Figure 5 shows approximations for specific loop iterations, we emphasize that the symbolic nature of our approach allows for approximating the densities of program variables for all – potentially infinitely many – loop iterations simultaneously. Therefore, our technique is constant in the number of loop iterations, whereas sampling has linear complexity. We compute the approximations from Figure 5 for all infinitely many loop iterations in $\sim 22$ seconds, with the experimental setup from Section 7. In comparison, sampling the loop $10^5$ times takes $\sim 3.6$ minutes for loop iteration 10 and $\sim 33$ minutes for loop iteration 100.
7 IMPLEMENTATION AND EVALUATION

Implementation. The program transformations (Section 3) and (higher) moment computation (Section 5) are implemented in the new tool $\texttt{Polar}$ [Moosbrugger et al. 2022]. The experiments can be reproduced using the corresponding artifact. For automatically inferring finiteness of program variables, we use a standard approach based on abstract interpretation. $\texttt{Polar}$ is implemented in python3, consisting of $\sim 3300$ LoC, and uses the packages sympy and symengine for symbolic manipulation of mathematical expressions. Together with all our benchmarks, $\texttt{Polar}$ is publicly available at https://github.com/probing-lab/polar.

Experimental Setting and Evaluation. The evaluation of our work is split into three parts. First, we evaluate $\texttt{Polar}$ on the ability of computing higher moments for 15 probabilistic programs exhibiting different characteristics (Section 7.1). Second, we compare $\texttt{Polar}$ to the exact tool $\texttt{Mora}$ [Bartocci et al. 2020b] which computes so-called moment-based invariants for a subset of our programming model (Section 7.2). Third, we compare our tool to approximate methods estimating program variable moments by confidence intervals through sampling (Section 7.3). All experiments have been run on a machine with a 2.6 GHz Intel i7 (Gen 10) processor and 32 GB of RAM. Runtime measurements are averaged over 10 executions.

Table 1. Evaluation of $\texttt{Polar}$ on 15 benchmarks. All times are in seconds. #V = number of variables in benchmark; C = benchmark contains circular dependencies; If = benchmark contains if-statements; S = benchmark contains symbolic constants; INF = benchmark’s states space is infinite; CONT = benchmark’s state space is continuous; Moment = Moment to compute; RT = Total runtime.

| Benchmark                          | #V | C/If/S/INF/CONT | Moment   | RT  |
|-----------------------------------|----|----------------|----------|-----|
| Running-Example (Fig. 1)          | 7  | ✓/✓/✓/✓/✓      | $E(z)$   | 0.67|
| Herman-3                          | 10 | ✓/✓/✓/✓/✓      | $E(\text{tokens})$ | 0.58|
| Las-Vegas-Search                  | 3  | X/✓/✓/✓/✓      | $E(\text{found}_3)$ | 0.36|
| Pi-Approximation                  | 4  | X/✓/✓/✓/✓      | $E(\text{count}_3)$ | 0.47|
| 50-Coin-Flips                     | 101| X/✓/✓/✓/✓      | $E(\text{total})$ | 0.91|
| Gambler-Ruin-Momentum             | 4  | ✓/✓/✓/✓/✓      | $E(x^3)$ | 2.89|
| Hawk-Dove-Symbolic                | 5  | X/✓/✓/✓/✓      | $E(p1bal^4)$ | 2.00|
| Variable-Swap                     | 4  | ✓/✓/✓/✓/✓      | $E(x^{30})$ | 2.42|
| Retransmission-Protocol           | 4  | X/✓/✓/✓/✓      | $E(\text{fail}_3)$ | 1.62|
| Randomized-Response               | 7  | X/✓/✓/✓/✓      | $E(p1^7)$ | 0.59|
| Duelling-Cowboys                  | 4  | ✓/✓/✓/✓/✓      | $E(\text{ahit})$ | 1.14|
| Martingale-Bet                    | 4  | X/✓/✓/✓/✓      | $E(\text{capital}_3)$ | 8.44|
| Bimodal                           | 5  | X/✓/✓/✓/✓      | $E(x^{10})$ | 4.50|
| DBN-Umbrella                      | 2  | X/✓/✓/✓/✓      | $E(\text{umbrella}_5)$ | 0.77|
| DBN-Component-Health              | 3  | X/✓/✓/✓/✓      | $E(\text{obs}_5)$ | 0.26|

5https://doi.org/10.5281/zenodo.7055030
6https://www.sympy.org
7https://github.com/symengine
7.1 Experimental Results with Higher Moments

Table 1 shows the evaluation of Pola on the program from Figure 1 and 14 benchmarks which are either from the literature on probabilistic programming [Kwiatkowska et al. 2012] (Herman-3), [Mclver and Morgan 2005] (Duelling-Cowboys), [Gretz et al. 2013] (Martingale-Bet), [Chakarov and Sankaranarayanan 2014] (Hawk-Dove-Symbolic, Variable-Swap), [Barthe et al. 2016] (Gambler-Ruin-Momentum), [Batz et al. 2021] (Retransmission-Protocol), differential privacy schemes [Warner 1965] (Randomized-Response), Dynamic Bayesian Networks (DBN-Umbrella, DBN-Component-Health) or well-known stochastic processes (Las-Vegas-Search, Pi-Approximation, Bimodal). The benchmarks Retransmission-Protocol and Hawk-Dove-Symbolic were further generalized from their original definition by replacing concrete numbers with symbolic constants. This makes these benchmarks only harder as solutions to the generalized versions are solutions for the concretizations. Table 1 illustrates that Pola can compute higher moments for various probabilistic programs exhibiting different features, like circular variable dependencies, if-statements, and symbolic constants with finite, infinite, continuous, and discrete state spaces. Moreover, the table shows that the number of program variables is not the primary factor for the complexity of computing moments. For instance, the benchmarks 50-Coin-Flips and Duelling-Cowboys have 101 and 4 program variables respectively. Nevertheless, the runtimes for computing first moments for the two benchmarks only differ by 0.23 s. The complexity of computing moments lies in the complexity of the resulting systems of recurrences which depend on the concrete features present in the benchmarks like specific variable dependencies, symbolic constants, or degrees of polynomials.

7.2 Experimental Comparison to Exact Methods

To the best of our knowledge, Mora is the only other tool capable of computing higher moments for variables of probabilistic loops without templates – as described in [Bartocci et al. 2019]. Mora operates on so-called Prob-solvable loops which form a strict subset of our program model (Section 3). Prob-solvable loops do not admit circular variable dependencies, if-statements, or state-dependent distribution parameters. We compare Pola against Mora on the Mora benchmarks taken from [Chakarov and Sankaranarayanan 2014; Chen et al. 2015; Katoen et al. 2010; Kura et al. 2019]. Details can be found in Table 2. The experiments illustrate that Pola can handle all programs and moments that Mora can, however, cannot compute any moment for any program in Table 1. On simple benchmarks, Pola is slightly slower than Mora due to the constant overhead of the program transformations and type inference to identify finite valued variables. On complex benchmarks Pola provides a significant speedup compared to Mora. For instance, for the STUTTERING_C benchmark Pola computes the moment $E(s^3)$ in about 8 seconds, whereas Mora needs over one minute.

7.3 Experimental Comparison with Sampling

For a probabilistic loop with program variable $x$ the moment $E(x_k^n)$ can be approximated for fixed $k$ and $n$ by sampling $x_k^n$ and calculating the sample average or confidence intervals. Table 3 compares Pola to computing confidence intervals by sampling for $k = 1$ and $n = 10$. The table shows that our tool is able to compute precise moments in a fraction of the time needed to sample programs to achieve satisfactory confidence intervals. An advantage of sampling is that it is applicable for any probabilistic loop. However, by its nature, sampling fails to give any formal guarantees or hard bounds on the approximated moments. This is critical if the loop body contains branches that are executed with low probability. If applicable, Pola can provide exact moments for symbolic $n$ (and involving other symbolic constants) faster than sampling can establish acceptable approximations. Moreover, even if the sampling of the loops is sped up by using a more efficient implementation,
Table 2. Comparison of Polar to Mora. The runtimes are in seconds per tool, benchmark, and moment. For Polar the comparison contains in brackets the seconds spent on parsing, normalizing, and type inference.

| Benchmark          | Mora  | Polar |
|--------------------|-------|-------|
| COUPON             |       |       |
| $E(c)$             | 0.25  | 0.29  |
| $E(c^2)$           | 0.27  | 0.29  |
| $E(c^3)$           | 0.29  | 0.29  |
| COUPON4            |       |       |
| $E(c)$             | 0.71  | 0.36  |
| $E(c^2)$           | 0.87  | 0.36  |
| $E(c^3)$           | 1.22  | 0.36  |
| RANDOM_WALK_1D     |       |       |
| $E(x)$             | 0.07  | 0.12  |
| $E(x^2)$           | 0.11  | 0.23  |
| $E(x^3)$           | 0.10  | 0.24  |
| SUM_RND_SERIES     |       |       |
| $E(x)$             | 0.27  | 0.27  |
| $E(x^2)$           | 0.97  | 0.43  |
| $E(x^3)$           | 2.48  | 0.79  |
| PRODUCT_DEP_VAR    |       |       |
| $E(p)$             | 0.37  | 0.28  |
| $E(p^2)$           | 1.41  | 0.46  |
| $E(p^3)$           | 4.03  | 0.97  |
| RANDOM_WALK_2D     |       |       |
| $E(x)$             | 0.10  | 0.12  |
| $E(x^2)$           | 0.21  | 0.24  |
| $E(x^3)$           | 0.17  | 0.24  |
| BINOMIAL($p$)      |       |       |
| $E(x)$             | 0.12  | 0.25  |
| $E(x^2)$           | 0.32  | 0.29  |
| $E(x^3)$           | 0.79  | 0.44  |

| Benchmark          | Mora  | Polar |
|--------------------|-------|-------|
| STUTTERING_A       |       |       |
| $E(s)$             | 0.29  | 0.26  |
| $E(s^2)$           | 1.13  | 0.43  |
| $E(s^3)$           | 3.32  | 0.97  |
| STUTTERING_B       |       |       |
| $E(s)$             | 0.26  | 0.27  |
| $E(s^2)$           | 0.94  | 0.39  |
| $E(s^3)$           | 2.26  | 0.79  |
| STUTTERING_C       |       |       |
| $E(s)$             | 0.76  | 0.40  |
| $E(s^2)$           | 12.43 | 1.94  |
| $E(s^3)$           | 74.83 | 8.19  |
| STUTTERING_D       |       |       |
| $E(s)$             | 0.76  | 0.43  |
| $E(s^2)$           | 8.19  | 1.34  |
| $E(s^3)$           | 25.67 | 4.33  |
| STUTTERING_P       |       |       |
| $E(s)$             | 0.26  | 0.28  |
| $E(s^2)$           | 1.17  | 0.49  |
| $E(s^3)$           | 3.53  | 1.17  |
| SQUARE             |       |       |
| $E(y)$             | 0.30  | 0.29  |
| $E(y^2)$           | 0.88  | 0.45  |
| $E(y^3)$           | 1.98  | 0.65  |

Table 3. Comparison of Polar to approximation through sampling. Polar = the tools runtime to compute the precise moment; CI N = an approximated 0.95-CI-interval from N samples; $T_{100,000}$ = the runtime for CI 100.000. The symbolic constant $p$ in Retransmission-Protocol is set to 0.9 and for Hawk-Dove-Symbolic we set $\nu = 4$ and $c = 8$.

| Benchmark            | Moment | CI 100   | CI 1.000 | CI 100.000 | $T_{100,000}$ | Polar |
|----------------------|--------|----------|----------|------------|---------------|-------|
| Running-Example (Fig. 1) | $E(z_{10})$ | (-55.6, 6.55) | (-58.5, -39.3) | (-45.6, -43.7) | 545.6s | 0.67s |
| Retransmission-Protocol | $E(fail_{10})$ | (0.09, 0.25) | (0.11, 0.16) | (0.109, 0.114) | 146.2s | 0.30s |
| Variable-Swap        | $E(y_{10})$ | (4.47, 5.75) | (5.06, 5.45) | (5.50, 5.54) | 245.8s | 0.13s |
| Hawk-Dove-Symbolic   | $E(p_{bal_{10}})$ | (8.28, 12.3) | (9.07, 10.5) | (9.86, 10.0) | 347.4s | 0.27s |
Polar enjoys complexity theoretical advantages. The complexity of sampling is linear in both the number of samples and the number of loop iterations. In contrast, Polar does not need to take multiple samples for higher precision as it symbolically computes the exact moments. Additionally, our method is constant in the number of loop iterations. With Polar, computing the moment for a specific loop iteration, say $10^5$, just amounts to evaluate the closed-form at $10^5$.

Figure 6 illustrates the importance of higher moments for probabilistic loops. The first moment provides a center of mass but contains no information on how the mass is distributed around this center. For this purpose higher moments are essential.

### 7.4 Evaluation Summary

Our experimental evaluation demonstrates that: (1) Polar can compute higher moments for a rich class of probabilistic loops with various characteristics, (2) Polar outperforms the state-of-the-art of moment computation for probabilistic loops in terms of supported programs and efficiency, and (3) Polar computes exact moments magnitudes faster than sampling can establish reasonable approximations.

### 8 RELATED WORK

Using recurrence equations to extract closed-forms for variables and quantitative invariants of loops is a well-studied technique for non-probabilistic programs [Breck et al. 2020; de Oliveira et al. 2016; Farzan and Kincaid 2015; Humenberger et al. 2017, 2018; Kincaid et al. 2019, 2018; Kovács 2008; Rodríguez-carbonell and Kapur 2004]. Because a classical program is a special case of a probabilistic program, our technique presented in Section 5 is a generalization of the closed-form computation for classical programs to probabilistic programs. Moreover, the generalization to probabilistic programs is not trivial. One reason for this is that for classical programs the closed form for $x^p$ is just the closed-form for $x$ to the power $p$. However, this fails for moments, as in general $E(x^p)$ is not equal to $E(x)^p$.

A common approach to quantitatively and exactly analyze probabilistic programs is to employ probabilistic model checking techniques [Baier and Katoen 2008; Dehnert et al. 2017; Holtzen et al. 2021; Katoen et al. 2011; Kwiatkowska et al. 2011].

Exact inference for computing precise posterior distributions for probabilistic programs has been studied in [Claret et al. 2013; Gehr et al. 2016; Holtzen et al. 2020; Narayanan et al. 2016; Saad et al.
An interesting direction for future research is using our techniques to assist probabilistic inference in the presence of loops.

A different approach to characterize the distributions of program variables are statistical methods such as Monte Carlo and hypothesis testing [Younes and Simmons 2006]. Simulations are however performed on a chosen finite number of program steps and do not provide guarantees over a potentially infinite execution, such as unbounded loops, limiting thus their use (if at all) for invariant generation.

In [McIver and Morgan 2005], a deductive approach, the weakest pre-expectation calculus, for reasoning about PPs with discrete program variables is introduced. Based on the weakest pre-expectation calculus, [Katoen et al. 2010] presents the first template-based approach for generating linear quantitative invariants for PPs. Other works [Chen et al. 2015; Feng et al. 2017] also address the synthesis of non-linear invariants or employ martingale expressions [Barthe et al. 2016]. All of these works target a slightly different problem and, unlike our approach, rely on templates. The first data-driven technique for invariant generation for PPs is presented in [Bao et al. 2022].

Another line of related work comes with computing bounds over expected values [Bouissou et al. 2016; Chatterjee et al. 2020; Karp 1994] and higher moments [Kura et al. 2019; Wang et al. 2021]. The approach in [Bouissou et al. 2016] can provide bounds for higher moments and can handle non-linear terms at the price of producing more conservative bounds. In contrast, our approach natively supports probabilistic polynomial assignments and provides a precise symbolic expression for higher moments.

The technique presented in [Bartocci et al. 2019] automates the generation of so-called moment-based invariants for a subclass of PPs with polynomial probabilistic updates and sets the basis for fully automatic exact higher moment computation. Relative to our approach, [Bartocci et al. 2019] supports neither if-statements (thus also no guarded loops), state-dependent distribution parameters, nor circular variable dependencies. Our work establishes stronger theoretical foundations.

9 CONCLUSION

We describe a fully automated approach for inferring exact higher moments for program variables of a large class of probabilistic loops with complex control flow, polynomial assignments, symbolic constants, circular dependencies among variables, and potentially uncountable state spaces. Our work uses program transformations to normalize and simplify probabilistic programs while preserving the joint distribution of program variables. We propose a power reduction technique for finite program variables to ease the complex polynomial arithmetic of probabilistic programs. We prove soundness and completeness of our approach, by establishing the theory of moment-computable probabilistic loops. We demonstrate use cases of exact higher moments in the context of computing tail probabilities and recovering distributions from moments. Our experimental evaluation illustrates the applicability of our work, solving several examples whose automation so far was not yet supported by the state-of-the-art in probabilistic program analysis.

DATA AVAILABILITY STATEMENT

The tool Po1ar together with all benchmarks and scripts necessary to reproduce the results reported in this paper are available through an openly accessible artifact [Moosbrugger et al. 2022].

ACKNOWLEDGMENTS

This research was supported by the WWTF ICT19-018 grant ProbInG, the ERC Consolidator Grant ARTIST 101002685, the Austrian FWF project W1255-N23, and the SecInt Doctoral College funded by TU Wien. We thank the anonymous reviewers for their outstanding and detailed feedback.
Di Wang, Jan Hoffmann, and Thomas Reps. 2021. Central Moment Analysis for Cost Accumulators in Probabilistic Programs. In Proc. of PLDI. https://doi.org/10.1145/3453483.3454062

Stanley L. Warner. 1965. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. J. Am. Stat. Assoc. (1965). https://doi.org/10.1080/01621459.1965.10480775

Håkan L. S. Younes and Reid G. Simmons. 2006. Statistical probabilistic model checking with a focus on time-bounded properties. Inf. Comput. (2006). https://doi.org/10.1016/j.ic.2006.05.002