



Automatic Generation of Moment-Based Invariants for Prob-Solvable Loops

Ezio Bartocci¹, Laura Kovács^{1,2}(✉), and Miroslav Stankovič¹

¹ TU Wien, Vienna, Austria

laura.kovacs@tuwien.ac.at

² Chalmers, Gothenburg, Sweden

Abstract. One of the main challenges in the analysis of probabilistic programs is to compute invariant properties that summarise loop behaviours. Automation of invariant generation is still at its infancy and most of the times targets only expected values of the program variables, which is insufficient to recover the full probabilistic program behaviour. We present a method to automatically generate *moment-based invariants* of a subclass of probabilistic programs, called Prob-solvable loops, with polynomial assignments over random variables and parametrised distributions. We combine methods from symbolic summation and statistics to derive invariants as valid properties over higher-order moments, such as expected values or variances, of program variables. We successfully evaluated our work on several examples where full automation for computing higher-order moments and invariants over program variables was not yet possible.

1 Introduction

Probabilistic programs (PPs), originally employed in cryptographic/privacy protocols and randomised algorithms, are now gaining momentum due to the several emerging applications in the areas of machine learning and AI [12].

One of the main problems that arise from introducing randomness into the program is that we can no longer view variables as single values; we must think about them as distributions. Existing approaches, see e.g. [5, 19, 27] usually take into consideration only expected values, or upper and lower bounds over program variables. As argued by [29], such information is however insufficient to characterize the full value distributions of variables; (co-)variances and other higher-order moments of variables are also needed. Computing such moments is however challenging, if possible at all – see [17] for an insight on the hardness of analyzing expected values and (co-)variances of PPs. We illustrate the

This research was supported by the Austrian Science Fund (FWF) under grants S11405-N23, S11409-N23 (RiSE/SHiNE), the ERC Starting Grant 2014 SYMCAR 639270, the Wallenberg Academy Fellowship 2014 TheProSE and the Austrian FWF project W1255-N23.

© Springer Nature Switzerland AG 2019

Y.-F. Chen et al. (Eds.): ATVA 2019, LNCS 11781, pp. 255–276, 2019.

https://doi.org/10.1007/978-3-030-31784-3_15

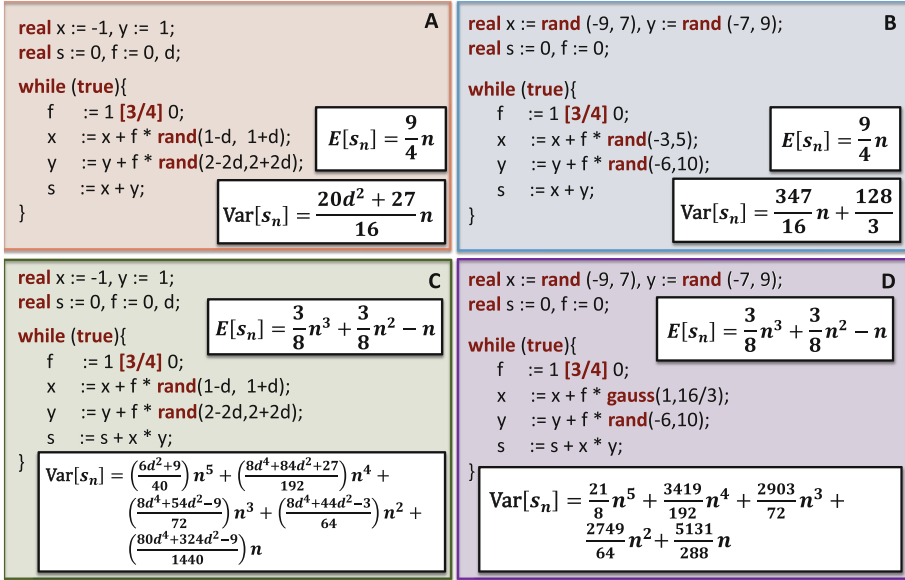


Fig. 1. Examples of four Prob-solvable loops. $f := 1[3/4]0$ is a statement that assigns to f the value 1 with probability $\frac{3}{4}$ and the value 0 with probability $1 - \frac{3}{4} = \frac{1}{4}$. The function $\text{rand}(a, b)$ samples a random number from a uniform distribution with support in the real interval $[a, b]$ and the function $\text{gauss}(\mu, \sigma^2)$ samples a random number from a normal distribution with mean μ and variance σ^2 . For each loop, we provide the moment-based invariants for the first ($E[\]$) and second moments ($\text{Var}[\]$) of s computed using our approach, where n denotes the loop counter.

importance of computing higher-order moments beyond expected values, by considering for example the PPs of Fig. 1(A) and (B): the expected value of variable s at each loop iteration is the same in both PPs, while the variance of the value distribution of s differs in general (a similar behaviour is also exploited by Fig. 1(C)–(D)). Thus, Fig. 1(A) and (B) do not have the same invariants over higher-order moments; yet, current approaches would fail identifying such differences and only compute expected values of variables.

One of the main challenges in analysing PPs and computing their higher-order moments comes with the presence of loops and the burden of computing so-called *quantitative invariants* [19]. Quantitative invariants are properties that are true before and after each loop iteration. Weakest pre-expectations [19, 27] can be used to compute quantitative invariants. This approach, supported for example in PRINSYS [13], consists in annotating a loop with a template invariant and then solve constraints over the unknown coefficients of the template. Other methods [2, 24] use martingales that are expressions over program variables whose expectations remain invariant. The aforementioned approaches are however not fully automatic since they require user guidance for providing templates and hints. In addition, they are limited to invariants over only expected

values: with the exception of [24], they do not compute higher-order moments describing the distribution generated by the PP (see Sect. 6 for more details).

In this paper we introduce a *fully automated approach* to compute invariant properties over *higher-order moments* of so-called *Prob-solvable loops*, to stand for *probabilistic P-solvable loops*. Prob-solvable loops are PPs that extend the imperative P-solvable loops described in [23] with probabilistic assignments over random variables and parametrised distributions. As such, variable updates are expressed by random polynomial, and not only affine, updates (see Sect. 3). Each program in Fig. 1 is Prob-solvable; moreover, Fig. 1(C)–(D) involve nonlinear updates over s .

Our work uses statistical properties to eliminate probabilistic choices and turn random updates into recurrence relations over higher-order moments of program variables. We show that higher-order moments of Prob-solvable loops can be described by C-finite recurrences (Theorem 1). We further solve such recurrences to derive *moment-based invariants* of Prob-solvable loops (Sect. 4). A moment-based invariant is a property that holds at arbitrary loop iterations (hence, invariants), expressing closed form solutions of higher-order moments of program variables. To the best of our knowledge, no other method is able to derive higher-order moments of PPs with infinite loops in a fully automated way – for example, the work in [11] provides the exact probabilistic inference only for PPs with bounded loops. Our work hence allows to replace, for example, the required human guidance of [13, 25] for Prob-solvable loops. We also support PPs with parametrised distributions (e.g., in Fig. 1(A)): instead of taking concrete instances of a given parametrised distribution, we automatically infer invariants of the entire class of PPs characterised by the considered parametrised distribution.

Our approach is both sound and terminating: given a Prob-solvable loops and an integer $k \geq 1$, we automatically infer the moment-based invariants over the k th moments of our input loop (see Sect. 4). Unlike the approach of [23] for deriving polynomial invariants of non-probabilistic (P-solvable) loops, our work only computes closed form expressions over higher-order moments and does not employ Gröbner basis computation to eliminate loop counters from the derived closed forms. As such, our moment-based invariants are not restrictive to polynomial properties but are linear combinations of polynomial expressions and exponential sequences over the loop counter. Moreover, Prob-solvable are more expressive than P-solvable loops as they are not restricted to deterministic updates but allow random assignments over variables.

Contributions. Our main contributions are: (1) we introduce the class of Prob-solvable loops with probabilistic assignments over random variables and distributions (Sect. 3); (2) we show that Prob-solvable loops can be modelled as C-finite recurrences over higher-order moments of variables (Theorem 1); (3) we provide a fully automated approach that derives moment-based invariants over arbitrary higher-order moments of Prob-solvable loops (Algorithm 1); (4) we implemented our work as an extension of the `Aligator` package [15] and evaluated over several challenging PPs (Sect. 5).

2 Preliminaries

We recall basic mathematical properties about recurrences and higher-order moments of variable values – for more details see [22, 26]. Throughout this paper, let $\mathbb{N}, \mathbb{Z}, \mathbb{R}$ denote the set of natural, integer and real numbers. We reserve capital letters to denote abstract random variables, e.g. X, Y, \dots , and use small letters to denote program variables, e.g. x, y, \dots , all possibly with indices.

2.1 C-Finite Recurrences

While sequences and recurrences are defined over arbitrary fields of characteristic zero, in our work we only focus over sequences/recurrences over \mathbb{R} .

Definition 1 (Sequence). *A univariate sequence in \mathbb{R} is a function $f : \mathbb{Z} \rightarrow \mathbb{R}$. A recurrence for a sequence $f(n)$ is*

$$f(n+r) = R(f(n), f(n+1), \dots, f(n+r-1), n), \quad \text{with } n \in \mathbb{N},$$

for some function $R : \mathbb{R}^{r+1} \rightarrow \mathbb{R}$, where $r \in \mathbb{N}$ is called the order of the recurrence.

For simplicity, we denote by $f(n)$ both the recurrence of $f(n)$ as well as the recurrence equation $f(n) = 0$. When solving the recurrence $f(n)$, one is interested in computing a *closed form* solution of $f(n)$, expressing the value of $f(n)$ as a function of n for any $n \in \mathbb{N}$. In our work we only consider the class of *linear recurrences with constant coefficients*, also called *C-finite recurrences*.

Definition 2 (C-finite recurrences). *A C-finite recurrence $f(n)$ satisfies the linear homogeneous recurrence with constant coefficients:*

$$f(n+r) = a_0 f(n) + a_1 f(n+1) + \dots + a_{r-1} f(n+r-1), \quad \text{with } r, n \in \mathbb{N}, \quad (1)$$

where r is the order of the recurrence, and $a_0, \dots, a_{r-1} \in \mathbb{R}$ are constants with $a_0 \neq 0$.

An example of a C-finite recurrence is the recurrence of Fibonacci numbers satisfying the recurrence $f(n+2) = f(n+1) + f(n)$, with initial values $f(0) = 0$ and $f(1) = 1$. Unlike arbitrary recurrences, closed forms of C-finite recurrences $f(n)$ always exist [22] and are defined as:

$$f(n) = P_1(n)\theta_1^n + \dots + P_s(n)\theta_s^n, \quad (2)$$

where $\theta_1, \dots, \theta_s \in \mathbb{R}$ are the distinct roots of the characteristic polynomial of $f(n)$ and $P_i(n)$ are polynomials in n . Closed forms of C-finite recurrences are called *C-finite expressions*. We note that, while the C-finite recurrence (1) is homogeneous, inhomogeneous C-finite recurrences can always be translated into homogeneous ones, as the inhomogeneous part of a C-finite recurrence is a C-finite expression.

In our work, we focus on the analysis of Prob-solvable loops and consider loop variables x as sequences $x(n)$, where $n \in \mathbb{N}$ denotes the loop iteration counter. Thus, $x(n)$ gives the value of the program variable x at iteration n .

2.2 Expected Values and Moments of Random Variables

Here we introduce the relevant notions from statistics that our work relies upon.

Definition 3 (Probability space). A probability space is a triple (Ω, F, P) consisting of a sample space Ω denoting the set of outcomes, where $\Omega \neq \emptyset$, a σ -algebra F with $F \subset 2^\Omega$, denoting a set of events, a probability measure $P : F \rightarrow [0, 1]$ s.t. $P(\Omega) = 1$.

We now define random variables and their higher-order moments.

Definition 4 (Random variable). A random variable $X : \Omega \rightarrow \mathbb{R}$ is a measurable function from a set Ω of possible outcomes to \mathbb{R} .

In the context of our Prob-solvable loops, for each loop variable x , we consider elements of its corresponding sequence $x(n)$ to be random variables. When working with a random variable X , one is in general interested in expected values and other moments of X .

Definition 5 (Expected value). An expected value of a random variable X defined on a probability space (Ω, F, P) is the Lebesgue integral: $E[X] = \int_{\Omega} X \cdot dP$. In the special case when Ω is discrete, that is the outcomes are X_1, \dots, X_N with corresponding probabilities p_1, \dots, p_N , we have $E[X] = \sum_{i=1}^N X_i \cdot p_i$. The expected value of X is often also referred to as the mean or μ of X .

For program variables x of Prob-solvable loops, our work computes the expected values of the corresponding sequences $x(n)$ but also higher-order and mixed moments.

Definition 6 (Higher-Order Moments). Let X be a random variable, $c \in \mathbb{R}$ and $k \in \mathbb{N}$. We write $Mom_k[c, X]$ to denote the k th moment about c of X , which is defined as:

$$Mom_k[c, X] = E[(X - c)^k] \tag{3}$$

whenever this exists.

All random distributions we consider in this paper have finite (existing) moments. In the rest of this section we also assume the moments exist and are finite. In our work we will be almost solely interested in moments about 0 (called *raw moments*) and about the mean $E[X]$ (called *central moments*). We note though that we can move to moments of X with different centers using Proposition 1.

Proposition 1 (Transformation of center). Let X be a random variable, $c, d \in \mathbb{R}$ and $k \in \mathbb{N}$. The k th moment about d of X , can be calculated from

moments about c of X by: $E[(X - d)^k] = \sum_{i=0}^k \binom{k}{i} E[(X - c)^i] (c - d)^{k-i}$.

Similarly to higher-order moments, we also consider *mixed moments*, that is $E[X \cdot Y]$, where X and Y are random variables. For arbitrary random variables X and Y , we have the following basic properties about their expected values and other moments:

- $E[c] = c$ for a constant $c \in \mathbb{R}$,
- expected value is linear, $E[X + Y] = E[X] + E[Y]$ and $E[c \cdot X] = c \cdot E[X]$,
- expected value is not multiplicative, in general $E[X \cdot Y] \neq E[X] \cdot E[Y]$
- expected value is multiplicative for independent random variables.

As a consequence of the above, expected values of monomials over arbitrary random variables, e.g. $E[X \cdot Y^2]$, cannot be in general further simplified.

The moments of a random variable X with bounded support fully characterise its value distribution. While computing all moments of X is generally very hard, knowing only a few moments of X gives useful information about its value distributions. The most common moments are variance, covariance, skewness, as defined below.

Definition 7 (Common moments). Variance measures how spread the distribution is and is defined as the second central moment: $Var[X] = Mom_2[E[X], X]$.

Covariance is a mixed moment measuring variability of two distributions and is defined as: $Cov[X, Y] = E[(X - E[X]) \cdot (Y - E[Y])]$.

Skewness measures asymmetry of the distribution and is defined as the normalised third central moment: $Skew[X] = \frac{Mom_3[E[X], X]}{(Var[X])^{3/2}}$.

Basic results about variance and covariance state: $Cov[X, X] = Var[X]$, $Var[X] = E[X^2] - (E[X])^2$, and $Cov[X, Y] = E[X \cdot Y] - E[X] \cdot E[Y]$.

Definition 8 (Moment-Generating Function (MGF)). A moment generating function of a random variable X is given by:

$$M_X(t) = E[e^{tX}], \quad \text{with } t \in \mathbb{R} \tag{4}$$

whenever this expectation exists.

Moment-generating functions, as the name suggests, can be used to compute higher-order moments of a random variable X . If we take the k th derivative of the moment-generating function of X , evaluated at 0, we get the k th moment about 0 of X , that is $Mom_k[0, X]$ ¹. For many standard distributions, including Bernoulli, uniform and normal distributions, the moment-generating function exists and gives us a way to compute the moments for random variables drawing from these distributions. Thanks to these properties, we can use common distributions in our Prob-solvable programs.

¹ due to the series expansion $e^{tX} = 1 + tE[X] + \frac{t^2 E[X^2]}{2!} + \frac{t^3 E[X^3]}{3!} + \dots$ and derivative w.r.t. t .

3 Programming Model: Prob-Solvable Programs

We now introduce our programming model of *Prob-solvable programs*, to stand for *probabilistic P-solvable programs*. P-solvable programs [23] are non-deterministic loops whose behaviour can be expressed by a system of C-finite recurrences over program variables. Prob-solvable programs build upon P-solvable programs by also allowing probabilistic assignments over random variables and distributions.

Prob-Solvable Loops. Let $m \in \mathbb{N}$ and x_1, \dots, x_m denote real-valued program variables. We define Prob-solvable loops with x_1, \dots, x_m variables as programs of the form:

$$I; \text{while}(\text{true})\{U\}, \quad \text{where:} \quad (5)$$

- I is a sequence of initial assignments over x_1, \dots, x_m . That is, I is an assignments sequence $x_1 := c_1; x_2 := c_2; \dots, x_m := c_m$, with $c_i \in \mathbb{R}$ representing a number drawn from a known distribution² - in particular, c_i can be a real constant.
- U is the loop body and is a sequence of m random updates, each of the form:

$$x_i := a_i x_i + P_i(x_1, \dots, x_{i-1}) \ [p_i] \ b_i x_i + Q_i(x_1, \dots, x_{i-1}), \quad (6)$$

or, in case of a deterministic assignment,

$$x_i := a_i x_i + P_i(x_1, \dots, x_{i-1}), \quad (7)$$

where $a_i, b_i \in \mathbb{R}$ are constants and $P_i, Q_i \in \mathbb{R}[x_1, \dots, x_{i-1}]$ are polynomials over program variables x_1, \dots, x_{i-1} . Further, $p_i \in [0, 1]$ in (6) is the probability of updating x_i to $a_i x_i + P_i(x_1, \dots, x_{i-1})$, whereas the probability to update x_i to $b_i x_i + Q_i(x_1, \dots, x_{i-1})$ in (6) is $1 - p_i$.

The coefficients a_i, b_i and the coefficients of P_i and Q_i in the variable assignments (6)-(7) of Prob-solvable loops can be drawn from a random distribution as long as the moments of this distribution are known and are independent from program variables x_1, \dots, x_m . Hence, the variable updates of Prob-solvable loop can involve coefficients drawn from Bernoulli, uniform, normal, and other distributions. Moreover, Prob-solvable support parametrised distributions, for example one may have the random distribution $\text{rand}(d_1, d_2)$ with arbitrary $d_1, d_2 \in \mathbb{R}$ symbolic constants. Similarly, rather than only considering concrete numeric values of p_i , the probabilities p_i in the probabilistic updates (6) of Prob-solvable loops can also be symbolic constants. Notice that our current model assumes 'true' loop condition. We will see later that it allows us to compute invariants for any loop with set number of iterations. In the future we plan to extend the model to work with more general loop conditions.

² a known distribution is a distribution with known and computable moments.

Example 1. The programs in Fig. 1 are Prob-solvable, using uniform distributions given by `rand()`. Figure 1(D) also uses a normal distribution given by `gauss()`. Note that while the random distributions of Fig. 1(B, D) are defined in terms of concrete constants, Fig. 1(A, C) have a parametrised random distribution, defined in terms of $d \in \mathbb{R}$.

Prob-Solvable Loops and Moment-Based Recurrences. Let us now consider a Prob-solvable program with $n \in \mathbb{N}$ denoting the loop iteration counter. We show that variable updates of Prob-solvable programs yield special recurrences in n , called *moment-based recurrences*. For this, we consider program variables x_1, \dots, x_m as sequences $x_1(n), \dots, x_m(n)$ allowing us to precisely describe relations between values of x_i at different loop iterations. Using this notation, probabilistic updates (6) over x_i turn $x_i(n)$ into a random variable, yielding the relation (similarly, for deterministic updates (7)):

$$x_i(n+1) = a_i x_i(n) + P_i(x_1(n), \dots, x_{i-1}(n)) [p_i] b_i x_i(n) + Q_i(x_1(n), \dots, x_{i-1}(n)).$$

The relation above could be treated as a recurrence equation over random variables $x_i(n)$ provided the probabilistic behaviour depending on p_i is encoded (as an extension) into a recurrence equation. To analyse such probabilistic updates of Prob-solvable loops, for each random variable $x_i(n)$ we consider their expected values $E[x_i(n)]$ and create new recurrence variables from expected values of monomials over original program variables (e.g. a new variable $E[x_i \cdot x_j]$). We refer to these new recurrence variables as *E-variables*. We note that any program variable yields an E-variable, but not every E-variable corresponds to one single program variable as E-variables are expected values of monomials over program variables. We now formulate recurrence equations over E-variables rather than over program variables, yielding *moment-based recurrences*.

Definition 9 (Moment-Based Recurrences). Let $x(n)$ be a sequence of random variables. A moment-based recurrence for x is a recurrence over E-variable $E[x]$:

$$E[x(n+r)] = R(E[x(n)], E[x(n+1)], \dots, E[x(n+r-1)], n) \quad (n \in \mathbb{N}),$$

for some function $R : \mathbb{R}^{r+1} \rightarrow \mathbb{R}$, where $r \in \mathbb{N}$ is the order of the moment-based recurrence.

Similarly to [27], note that variable updates $x_i := f_1(x_i) [p_i] f_2(x_i)$ yield the relation:

$$\begin{aligned} E[x_i(n+1)] &= E[p_i \cdot f_1(x_i(n)) + (1-p_i) \cdot f_2(x_i(n))] \\ &= p_i \cdot E[f_1(x_i(n))] + (1-p_i) \cdot E[f_2(x_i(n))] \end{aligned} \tag{8}$$

Thanks to this relation, probabilistic updates (6) are rewritten into the moment-based recurrence equation:

$$\begin{aligned} E[x_i(n+1)] &= p_i \cdot E[a_i x_i(n) + P_i(x_1(n), \dots, x_{i-1}(n))] \\ &\quad + (1-p_i) \cdot E[b_i x_i(n) + Q_i(x_1(n), \dots, x_{i-1}(n))]. \end{aligned} \tag{9}$$

In particular, we have $E[x_i(n + 1)] = p_i \cdot E[a_i x_i(n) + P_i(x_1(n), \dots, x_{i-1}(n))]$ for the deterministic assignments of (7) (that is, $p_i = 1$ in (7)).

By using properties of expected values of expressions $expr_1, expr_2$ over random variables, we obtain the following simplification rules:

$$\begin{aligned}
 E[expr_1 + expr_2] &\rightarrow E[expr_1] + E[expr_2] \\
 E[expr_1 \cdot expr_2] &\rightarrow E[expr_1] \cdot E[expr_2], \text{ if } expr_1, expr_2 \text{ are independent} \\
 E[c \cdot expr_1] &\rightarrow c \cdot E[expr_1] \\
 E[c] &\rightarrow c \\
 E[\mathcal{D} \cdot expr_1] &\rightarrow E[\mathcal{D}] \cdot E[expr_1]
 \end{aligned} \tag{10}$$

where $c \in \mathbb{R}$ is a constant and \mathcal{D} is a known independent distribution.

Example 2. The moment-based recurrences of the Prob-solvable loop of Fig. 1(A) are:

$$\begin{cases} E[f(n + 1)] = \frac{3}{4}E[1] + \frac{1}{4}E[0] \\ E[x(n + 1)] = E[x(n) + f(n + 1) \cdot rand(1 - d, 1 + d)] \\ E[y(n + 1)] = E[y(n) + f(n + 1) \cdot rand(2 - 2d, 2 + 2d)] \\ E[s(n + 1)] = E[x(n + 1) + y(n + 1)] \end{cases}$$

By using the simplification rules (10) on the above recurrences, we obtain the following simplified moment-based recurrences of Fig. 1(A):

$$\begin{cases} E[f(n + 1)] = \frac{3}{4} \\ E[x(n + 1)] = E[x(n)] + E[f(n + 1)] \cdot E[rand(1 - d, 1 + d)] \\ E[y(n + 1)] = E[y(n)] + E[f(n + 1)] \cdot E[rand(2 - 2d, 2 + 2d)] \\ E[s(n + 1)] = E[x(n + 1)] + E[y(n + 1)] \end{cases} \tag{11}$$

In Sect. 4 we show that Prob-solvable loops can further be rewritten into a system of C-finite recurrences over E-variables.

Prob-Solvable Loops and Mutually Dependent Updates. Consider PP loops with mutually dependent affine updates:

$$x_i := \sum_{k=1}^m a_{i,k} x_k + c_i [p_i] \sum_{k=1}^m b_{i,k} x_k + d_i, \tag{12}$$

where $a_{i,k}, b_{i,k}, c_i, d_i \in \mathbb{R}$ are constants. While such assignments are not directly captured by updates (6) of Prob-solvable loops, this is not a restriction of our work. Variable updates given by (12) yield mutually dependent C-finite recurrences over E-variables. Using methods from [22], this coupled system of C-finite recurrences can be rewritten into an equivalent system of independent C-finite recurrences over E-variables, yielding an independent system of moment-based recurrences over which our invariant generation algorithm from Sect. 4 can be applied. Hence probabilistic loops with affine updates are special cases of Prob-solvable loops.

Multi-path Prob-Solvable Loops. While (5) defines Prob-solvable programs as single-path loops, the following class of multi-path loops can naturally be modeled by Prob-solvable programs:

$$I; \text{while}(\text{true})\{\text{if } t \text{ then } U_1 \text{ else } U_2\}, \quad \text{where:} \quad (13)$$

I is as in (5), t is a boolean-valued random variable, and U_1 and U_2 are respectively sequences of deterministic updates $x_i := a_i x_i + P_i(x_1, \dots, x_{i-1})$ and $x_i := b_i x_i + Q_i(x_1, \dots, x_{i-1})$ as in (7). PPs (13) can be rewritten to equivalent Prob-solvable loops, as follows. A pair of updates $x := u_1[p]v_1$ from U_1 and $x := u_2[p]v_2$ from U_2 is rewritten by the following sequence of updates:

$$\begin{aligned} f &:= 1[p]0; \\ g &:= 1[p]0; \\ x &:= t(u_1 f + v_1(1 - f)) + (1 - t)(u_2 g + v_2(1 - g)) \end{aligned} \quad (14)$$

with f, g fresh program variables. The resulting program is Prob-solvable and we can thus compute moment-based invariants of multi-path loops as in (13). The programs COUPON, RANDOM_WALK_2D of Table 1 are Prob-solvable loops corresponding to such multi-path loops. Notice that the last term in (14) grows exponentially in terms of branch degree. This is similar to the *weakest pre-expectation* reasoning approach [19, 27].

4 Moment-Based Invariants of Prob-solvable Loops

Thanks to probabilistic updates, the values of program variables of Prob-solvable loops after a specific number of loop iterations are not a priori determined. The value distributions $x_i(n)$ of program variables x_i are therefore random variables. When analysing Prob-solvable loops, and in general probabilistic programs, one is therefore required to capture relevant properties over expected values and higher moments of the variables in order to precisely summarise the value distribution of program variables.

Moment-Based Invariants. We are interested in automatically generating so-called *moment-based invariants* of Prob-solvable loops. Moment-based invariants are properties over expected values and higher moments of program variables such that these properties hold at arbitrary loop iterations (and hence are invariants).

Automated Generation of Moment-Based Invariants of Prob-Solvable Loops. Our method for generating moment-based invariants of Prob-solvable loops is summarized in Algorithm 1. Algorithm 1 takes as input a Prob-solvable loop \mathcal{P} and a natural number $k \geq 1$ and returns *moment-based invariants over the k th moments* of the program variables $\{x_1, \dots, x_m\}$. We denote by n the loop counter of \mathcal{P} .

Algorithm 1. Moment-Based Invariants of Prob-solvable Loops

Input: Prob-solvable loop \mathcal{P} as defined in (5), with variables $\{x_1, \dots, x_m\}$, and $k \geq 1$
Output: Set MI of Moment-based invariants of \mathcal{P} over the k th moments of $\{x_1, \dots, x_m\}$
Assumptions: $n \in \mathbb{N}$ is the loop counter of \mathcal{P}

1: Extract the moment-based recurrence relations of \mathcal{P} , for $i = 1, \dots, m$:

$$E[x_i(n+1)] = p_i \cdot E[a_i x_i(n) + P_i(x_1(n), \dots, x_{i-1}(n))] \\ + (1 - p_i) \cdot E[b_i x_i(n) + Q_i(x_1(n), \dots, x_{i-1}(n))].$$

2: $MBRecs = \{E[x_i(n+1)] \mid i = 1, \dots, m\}$ \triangleright initial set of moment-based recurrences

3: $S := \{x_1^k, \dots, x_m^k\}$ \triangleright initial set of monomials of E-variables
as $Mom_k[0, x_i(n)] = E[x_i(n)^k]$

4: **while** $S \neq \emptyset$ **do**

5: $M := \prod_{i=1}^m x_i^{\alpha_i}$, where and $M \in S$ $\alpha_i \in \mathbb{N}$

6: $S := S \setminus \{M\}$

7: $M' = M[x_i^{\alpha_i} \leftarrow upd_i]$, for each $i = m, \dots, 1$ \triangleright replace each $x_i^{\alpha_i}$ in M with upd_i

where upd_i denotes:

$$p_i \cdot (a_i x_i + P_i(x_1, \dots, x_{i-1}))^{\alpha_i} + (1 - p_i) \cdot (b_i x_i + Q_i(x_1, \dots, x_{i-1}))^{\alpha_i}$$

8: Rewrite M' as $M' = \sum N_j$ for monomials N_j over x_1, \dots, x_m

9: Simplify the moment-based recurrence $E[M(n+1)] = E[\sum N_j]$ using the rules (10)
 $\triangleright M(n+1)$ denotes $\prod_{i=1}^m x_i(n+1)^{\alpha_i}$

10: $MBRecs = MBRecs \cup \{E[M(n+1)]\}$
 \triangleright add $E[M(n+1)]$ to the set of moment-based recurrences

11: **for** each monomial N_j in M **do**

12: **if** $E[N_j] \notin MBRecs$ **then** \triangleright there is no moment-based recurrence for N_j

13: $S = S \cup \{N_j\}$ \triangleright add N_j to S

14: **end while**

15: Solve the system of moment-based recurrences $MBRecs$

16: $MI = \{E[x_i(n)^k] - CF_i(k, n) = 0 \mid i = 1, \dots, m\}$
 $\triangleright CF_i(k, n)$ is the closed form solution of $E[x_i^k]$

17: **return** the set MI of moment based invariants of \mathcal{P} for the k th moments of x_1, \dots, x_m

Theorem 1. Higher-order moments of variables in Prob-solvable loops can be modeled by C-finite recurrences over E-variables.

Proof. We want to show that $E[x_i^{\alpha_i}]$ can be expressed using recurrence equations. The idea is to express $x_i^{\alpha_i}(n+1)$ in terms of the value of x_i at the n -th iteration. Value of $x_i(n+1)$ is $a_i x_i(n) + P_i(x_1(n+1), \dots, x_{i-1}(n+1))$ with probability p_i and $b_i x_i(n) + Q_i(x_1(n+1), \dots, x_{i-1}(n+1))$ with probability $(1 - p_i)$. From here we can derive that $E[x_i^{\alpha_i}(n+1)] = E[p_i \cdot (a_i x_i + P_i(x_1, \dots, x_{i-1}))^{\alpha_i} + (1 - p_i) \cdot (b_i x_i + Q_i(x_1, \dots, x_{i-1}))^{\alpha_i}]$. For arbitrary monomial $M = \prod x_i^{\alpha_i}(n+1)$ we can express $E[M]$ by substituting each $x_i^{\alpha_i}(n+1)$ as above. This process is captured by line 7 of Algorithm 1. The new equations can be further simplified using properties of expected values and the simplification rules (10) to give recurrence equations over E-variables.

We now describe Algorithm 1. Our algorithm first rewrites \mathcal{P} into a set $MBRecs$ of moment-based recurrences, as described in Sect. 3. That is, program variables x_i are turned into random variables $x_i(n)$ and variable updates over x_i become moment-based recurrences over E-variables by using the relation of (8) (lines 1–2) of Algorithm 1).

The algorithm next proceeds with computing the moment-based recurrences of the k th moments of x_1, \dots, x_m . Recall that the k th moment of x_i is given by:

$$Mom_k[0, x_i(n)] = E[x_i(n)^k].$$

Hence, the set S of monomials yielding E-variables for which moment-based recurrences need to be solved is initialized to $\{x_1^k, \dots, x_m^k\}$ (line 3 of Algorithm 1). Note that by considering the resulting E-variables $E[x_i^k]$ and solving the moment-based recurrences of $E[x_i^k]$, we derive closed forms of the k th moments of $\{x_1, \dots, x_m\}$ (line 16 of Algorithm 1). To this end, Algorithm 1 recursively computes the moment-based recurrences of every E-variable arising from the moment-based recurrences of $E[x_i^k]$ (lines 4–14 of Algorithm 1), thus ultimately computing closed forms for $E[x_i^k]$. One can then use transformations described in Proposition 1 to compute closed forms for other moments, such as variance and covariance. In more detail,

- for each monomial $M = \prod x_j^{\alpha_j}$ from S , we substitute $x_i^{\alpha_i}$ in M by its probabilistic behaviour. That is, the update of x_i in the Prob-solvable loop \mathcal{P} is rewritten, according to (8), into the sum of its two probabilistic updates, weighted by their respective probabilities (lines 5–7 of Algorithm 1). Rewriting in line 7 of Algorithm 1 represents the most non-trivial step in our algorithm, combining non-deterministic nature of our program with polynomial properties. The resulting polynomial M' from M is then reordered to be expressed as a sum of new monomials N_j (line 8 of Algorithm 1); such a sum always exists as M' involves only addition and multiplication over x_1, \dots, x_m (recall that P_i and Q_i are polynomials over x_1, \dots, x_m).
- By applying the simplification rules(10) of E-variables over the moment-based recurrence of $E[\sum N_j]$, the recurrence of $E[M(n+1)]$ is obtained and added to the set $MBRecs$. Here, $M(n+1)$ denotes $\prod_{i=1}^m x_i(n+1)^{\alpha_i}$. As the recurrence of $E[M(n+1)]$ depends on $E[N_j]$, moment-based recurrences of $E[N_j]$ need also be computed and hence S is enlarged by N_j (lines 9–13 of Algorithm 1).

As a result, the set $MBRecs$ of moment-based recurrences of E-variables corresponding to S are obtained. These recurrences are C-finite expressions over E-variables (see correctness argument of Theorem 3) and hence their closed form solutions exist. In particular, the closed forms $CF_i(k, n)$ of $E[x_i(n)^k]$ is derived, turning $E[x_i(n)^k] - CF_i(k, n) = 0$ into an inductive property that holds at arbitrary loop iterations and is hence a moment-based invariant of \mathcal{P} over the k th moment of x_i (line 16 of Algorithm 1).

Theorem 2 (Soundness). *Consider a Prob-solvable loop \mathcal{P} with program variables x_1, \dots, x_m and let k be a non-negative integer with $k \geq 1$. Algorithm 1 generates moment-based invariants of \mathcal{P} over the k th moments of x_1, \dots, x_m .*

Note when $k = 1$, Algorithm 1 computes the moment-based invariants as invariant relations over the closed form solutions of expected values of x_1, \dots, x_m . In this case, our moment-based invariants are quantitative invariants as in [19].

Example 3. We illustrate Algorithm 1 for computing the second moments (i.e. $k = 2$) of the Prob-solvable loop of Fig. 1(A).

Our algorithm initializes $MBRecs = \{E[f(n+1)], E[x(n+1)], E[y(n+1)], E[s(n+1)]\}$ and $S = \{f^2, x^2, y^2, s^2\}$.

We next (arbitrarily) choose M to be the monomial f^2 from S . Thus, $S = \{x^2, y^2, s^2\}$. Using the probabilistic update of f , we replace f^2 by $\frac{3}{4} \cdot 1^2 + (1 - \frac{3}{4}) \cdot 0^2$, that is by $\frac{3}{4}$. As a result, $MBRecs = MBRecs \cup \{E[f(n+1)^2] = \frac{3}{4}\}$ and S remains unchanged.

We next choose M to be x^2 and set $S = \{y^2, s^2\}$. We replace x^2 by its randomised behaviour, yielding $E[M(n+1)] = E[x(n+1)^2] = E[(x(n) + f(n+1) \cdot \mathbf{rand}(1 - \mathbf{d}, 1 + \mathbf{d}))^2]$. By the simplification rules (10) over E-variables, we obtain:

$$E[x(n+1)^2] = E[x(n)^2] + 2 \cdot E[x(n)] \cdot E[f(n+1)] + E[f(n+1)^2] \cdot \frac{1}{3}(d^2 + 3), \quad (15)$$

as $f(n+1)$ is independent from $x(n)$ and $E[\mathbf{rand}(1 - \mathbf{d}, 1 + \mathbf{d})^2] = \frac{1}{3}(d^2 + 3)$. We add the recurrence (15) to $MBRecs$ and keep S unchanged as the E-variables $E[x(n)], E[f(n+1)], E[f(n+1)^2]$ have their recurrences already in $MBRecs$.

We next set M to y^2 and change $S = \{s^2\}$. Similarly to $E[x(n+1)^2]$, we get:

$$E[y(n+1)^2] = E[y(n)^2] + 4 \cdot E[y(n)] \cdot E[f(n+1)] + E[f(n+1)^2] \cdot \frac{4}{3}(d^2 + 3), \quad (16)$$

by using that $f(n+1)$ is independent from $y(n)$ and $E[\mathbf{rand}(2 - 2\mathbf{d}, 2 + 2\mathbf{d})^2] = \frac{4}{3}(d^2 + 3)$. We add the recurrence (16) to $MBRecs$ and keep S unchanged. We set M to s^2 , yielding $S = \emptyset$. We extend $MBRecs$ with the recurrence:

$$E[s(n+1)^2] = E[(x(n+1) + y(n+1))^2] = E[x(n+1)^2] + 2E[(xy)(n+1)] + E[y(n+1)^2]$$

and add xy to S . We therefore consider M to be xy and set $S = \emptyset$. We obtain:

$$E[(xy)(n+1)] = E[(xy)(n)] + 2 \cdot E[x(n)] \cdot E[f(n+1)] + E[y(n)] \cdot E[f(n+1)] + 2 \cdot E[f(n+1)^2],$$

by using that $E[\mathbf{rand}(1 - \mathbf{d}, 1 + \mathbf{d})] = 1$ and $E[\mathbf{rand}(2 - 2\mathbf{d}, 2 + 2\mathbf{d})] = 2$. We add the recurrence of $E[(xy)(n+1)]$ to $MBRecs$ and keep $S = \emptyset$.

As a result, we proceed to solve the moment-based recurrences of $MBRecs$. We focus first on the recurrences over expected values:

$$\begin{aligned} E[f(n+1)] &= \frac{3}{4} \\ E[x(n+1)] &= E[x(n)] + E[f(n+1) \cdot \mathbf{rand}(1 - \mathbf{d}, 1 + \mathbf{d})] = E[x(n)] + \frac{3}{4} \\ E[y(n+1)] &= E[y(n)] + E[f(n+1) \cdot \mathbf{rand}(2 - 2\mathbf{d}, 2 + 2\mathbf{d})] = E[x(n)] + 2 \cdot \frac{3}{4} \\ E[s(n+1)] &= E[x(n+1)] + E[y(n+1)] \end{aligned}$$

Note that the above recurrences are C-finite recurrences over E-variables. For computing closed forms, we respectively substitute $E[f(n + 1)]$ by its closed form in $E[y(n + 1)]$ and $E[x(n + 1)]$, yielding closed forms for $E[y(n + 1)]$ and $E[x(n + 1)]$, and hence for $E[s(n + 1)]$. By also using the initial values of Fig. 1, we derive the closed forms:

$$\begin{aligned} E[f(n)] &= \frac{3}{4} & E[s(n)] &= \frac{9}{4}n \\ E[x(n)] &= \frac{3}{4}n - 1 & E[y(n)] &= \frac{3}{2}n + 1 \end{aligned}$$

We next similarly derive the closed forms for higher-order and mixed moments:

$$\begin{aligned} E[f(n)^2] &= \frac{3}{4} & E[s(n)^2] &= \frac{81}{16}n^2 + \frac{20d^2+27}{16}n \\ E[x(n)^2] &= \frac{9}{16}n^2 + \frac{4d^2-21}{16}n + 1 & E[y(n)^2] &= \frac{9}{4}n^2 + \frac{4d^2+15}{4}n + 1 \\ E[(xy)(n)] &= \frac{9}{8}n^2 - \frac{3}{8}n - 1 \end{aligned}$$

yielding hence the moment-based invariants over the second moments of variables of Fig. 1. Using Proposition 1 and Definition 7, we derive the variance of $s(n)$ as $Var(s(n)) = \frac{20d^2+27}{16}n$. □

Let us finally note that the termination of Algorithm 1 depends on whether for every monomial M (from the set S , line 4 of Algorithm 1) the moment-based recurrence equation over the corresponding E-variable $E[M(n + 1)]$ can be computed as a C-finite recurrence over E-variables. We prove this using transfinite induction over monomials and properties of inhomogeneous C-finite recurrences.

Theorem 3 (Termination). *For any non-negative integer k with $k \geq 1$ and any Prob-solvable loop \mathcal{P} with program variables x_1, \dots, x_m , Algorithm 1 terminates. Moreover, Algorithm 1 terminates in at most $\mathcal{O}(k^m \cdot d_m^{m-1} \cdot d_{m-1}^{m-2} \cdots d_2^1)$ steps, where $d_i = \max\{\deg(P_i), \deg(Q_i), 1\}$ with $\deg(P_i), \deg(Q_i)$ denoting the degree of polynomials P_i and Q_i of the variable updates (6).*

Proof. We associate every monomial with an ordinal number as follows:

$$x_k^{\alpha_k} \cdot x_{k-1}^{\alpha_{k-1}} \cdots x_1^{\alpha_1} \xrightarrow{\sigma} \omega^k \cdot \alpha_k + \omega^{k-1} \cdot \alpha_{k-1} \cdots + \alpha_1,$$

and order monomials M, N such that $M > N$ if $\sigma(M) > \sigma(N)$. Algorithm 1 terminates if for every monomial M (from the set S , line 4 of Algorithm 1) the moment-based recurrence equation over the corresponding E-variable $E[M(n+1)]$ can be computed as a C-finite recurrence over E-variables. We will show that this is indeed the case by transfinite induction over monomials.

Let $M = \prod_{k=1}^K x_k^{\alpha_k}$ be a monomial and assume that every smaller monomial has a closed form solution in form of a C-finite expression. Let

$$x_i^{\alpha_i} := (c_i x_i + P_i(x_1, \dots, x_{i-1}))^{\alpha_i} \tag{17}$$

be the updates of our variables after removing the probabilistic choice, as in line 5 of Algorithm 1. Then the recurrence for M is:

$$\begin{aligned}
 E[M(n+1)] &= E\left[\prod_{i=1}^K \left(p_i \cdot (a_i x_i + P_i(x_1, \dots, x_{i-1}))^{\alpha_i} \right. \right. \\
 &\quad \left. \left. + (1 - p_i) \cdot (b_i x_i + Q_i(x_1, \dots, x_{i-1}))^{\alpha_i}\right)(n)\right] \\
 &= E[M(n)] + \sum_{j=1}^J b_j \cdot E[N_j(n)] \tag{18}
 \end{aligned}$$

for some J , constants b_i and monomials N_1, \dots, N_J all different than M . By Lemma 1, we have an inhomogeneous C-finite recurrence relation $E[M(n+1)] = E[M(n)] + \gamma$, for some C-finite expression γ . Hence, the closed form of $E[M(n+1)]$ exists and is a C-finite expression. \square

We finally prove our auxiliary lemma, used in the above proof of Theorem 3.

Lemma 1. *In the recurrence (18) over E -variables, we have $M > N_j$ for all $j \leq J$.*

Proof. Let $M = \prod_{k=1}^K x_k^{\alpha_k}$ and have $N_j = \prod_{k=1}^K x_k^{\beta_k}$ coming from

$$\prod_{i=1}^K (c_i x_i + P_i(x_1, \dots, x_{i-1}))^{\alpha_i}. \tag{19}$$

Assume $M \leq N_j$, i.e. $\omega^K \cdot \alpha_K + \dots + \alpha_1 \leq \omega^K \cdot \beta_K + \dots + \beta_1$, so we have $\alpha_K \leq \beta_K$. Note that in (19) x_K only appears in factor $c_K x_K + P_K(x_1, \dots, x_{K-1})$. Considering the multiplicity, we get at most α_K th power of x_K , hence $\alpha_K \geq \beta_K$. Thus $\alpha_K = \beta_K$. So for $M \leq N_j$ we need N_j from $(c_K x_K)^{\alpha_K} \cdot \prod_{i=1}^{K-1} (c_i x_i + P_i(x_1, \dots, x_{i-1}))^{\alpha_i}$.

Proceeding similarly for x_{K-1}, x_{K-2}, \dots , we get that for each $k \leq K$ we have $\alpha_k = \beta_k$, which contradicts the assumption, thus $M > N_j$ as needed.

Regarding the termination time of Algorithm 1, let us look at what monomials can possibly be added to S . Let $M = \prod x_i^{\alpha_i} \in S$. Based on the above reasoning, it is clear that in case $i = m$ we have $\alpha_m \leq k$. For any $i < m$ the maximum value of α_i is $\alpha_{i+1} \cdot d_{i+1}$. Therefore, we have $\alpha_i \leq k \cdot \prod_{j=i+1}^m d_j$. Thus, we can count all possible monomials and hence derived the upper bound on the time complexity of Algorithm 1 as the product of these upper bounds. That is, the upper bound on the time complexity of Algorithm 1 is given by $k^m \cdot d_m^{m-1} \cdot d_{m-1}^{m-2} \dots \cdot d_2^1$. \square

5 Implementation and Experiments

We implemented our work in the Julia language, using `Aligator`[15] for handling and solving recurrences. We evaluated our work on several challenging probabilistic programs with parametrised distributions, symbolic probabilities and/or

both discrete and continuous random variables. All our experiments were run on MacBook Pro 2017 with 2.3 GHz Intel Core i5 and 8GB RAM. Our implementation and benchmarks are available at: github.com/miroslav21/aligator.

Benchmarks. We evaluated our work on 13 probabilistic programs, as follows. We used 7 programs from works [5, 7, 9, 19, 24] on invariant generation. These examples are given in lines 1–7 of Table 1; we note though that BINOMIAL (“ p ”) represents our generalisation of a binomial distribution example taken from [7, 9, 19] to a probabilistic program with parametrised probability p . We further crafted 6 examples of our own, illustrating the distinctive features of our work. These examples are listed in lines 8–13 of Table 1: lines 8–11 correspond to the examples of Fig. 1; line 12 of Table 1 shows a variation of Fig. 1, with a parametrized distribution p ; line 13 corresponds to a non-linear Prob-solvable loop computing squares. All our benchmarks are available at the aforementioned url.

Experimental Results with Moment-Based Invariants. Results of our evaluation are presented in Table 1. While Algorithm 1 can compute invariants over arbitrary k th higher-order moments, due to lack of space and readability, Table 1 lists only our moment-based invariants up to the third moment (i.e. $k \leq 3$), that is for expected values, second- and third-order moments. The first column of Table 1 lists the benchmark name, whereas the second column gives the degree of the moments (i.e. $k = 1, 2, 3$) for which we compute invariants. The third column reports the timings (in seconds) our implementation needed to derive invariants. The last column shows our moment-based invariants; for readability, we decided to omit intermediary invariants (up to 30 for some programs) and only show the most relevant invariants.

We could not perform a fair practical comparison with other existing methods: to the best of our knowledge, existing works, such as [2, 13, 19, 24], require user guidance/templates/hints. Further, most of the existing techniques do not support symbolic probabilities and/or parametrised distributions - which are, for example, required in the analysis of programs STUTTERINGA, STUTTERINGC, STUTTERINGP of Table 1. We also note that examples COUPON, STUTTERINGC, STUTTERINGP involve non-linear probabilistic updates hindering automation in existing methods, while such updates can naturally be encoded as moment-based recurrences in our framework. We finally note that while second-order moments are computed only by [24], but with the help of user-provided templates, no existing approaches compute moments for $k \geq 3$. Our experiments show that inferring third-order moments are in general not expensive. Yet, for examples STUTTERINGA, STUTTERINGC, STUTTERINGP with parametrized distributions/probabilities more computation time is needed. This increase in time comes from handling more complex symbolic expressions due to the non-linear updates and parametrized distributions, and from non-optimal recurrence solving.

Table 1. Moment-based invariants of Prob-solvable loops, where n is the loop counter.

Program	Moment	Runtime (s)	Computed Moment-Based Invariants
COUPON [24]	1	0.37	$E[c(n)] = (2^n - 1)/(2^n)$
	2	0.40	$E[c^2(n)] = (2^n - 1)/(2^n)$
	3	0.34	$E[c^3(n)] = (2^n - 1)/(2^n)$
COUPON4 [24]	1	0.90	$E[c(n)] = (4^n - 3^3)/(4^n)$
	2	1.1	$E[c^2(n)] = (4^n - 3^3)/(4^n)$
	3	1.3	$E[c^3(n)] = (4^n - 3^3)/(4^n)$
RANDOM_WALK_1D_CTS [24]	1	0.12	$E[x(n)] = n/5$
	2	0.45	$E[x^2(n)] = n^2/25 + 22n/75$
	3	1.00	$E[x^3(n)] = n^3/125 + n^2/22/125 - n/21/250$
SUM_RND_SERIES [7]	1	0.31	$E[x(n)] = n^2/4 + n/4$
	2	2.89	$E[x^2(n)] = n^4/16 + 5n^3/24 + 3n^2/16 + n/24$
	3	17.7	$E[x^3(n)] = n^6/64 + 7n^5/64 + 13n^4/64 + 9n^3/64 + n^2/32$
PRODUCT_DEP_VAR [7]	1	0.65	$E[p(n)] = n^2/4 - n/4$
	2	6.27	$E[p^2(n)] = n^4/16 - n^3/8 + 3n^2/16 - n/8$
	3	37.5	$E[p^3(n)] = n^6/64 - 3n^5/64 + 9n^4/64 - 21n^3/64 + 15n^2/32 - n/4$
RANDOM_WALK_2D [5, 24]	1	0.07	$E[x(n)] = 0$
	2	0.26	$E[x^2(n)] = n/2$
	3	0.49	$E[x^3(n)] = 0$
BINOMIAL("p") [7, 9, 19]	1	0.17	$E[x(n)] = np$
	2	0.47	$E[x^2(n)] = n^2p^2 + np(1 - p)$
	3	1.6	$E[x^3(n)] = n^3p^3 - 3n^2p^3 + 3n^2p^2 + 2np^3 - 3np^2 + np$
STUTTERINGA – FIG. 1(A)	1	0.44	$E[s(n)] = 9n/4$
	2	2.2	$E[s^2(n)] = 81n^2/16 + (20d^2 + 27)/16n$
	3	8.48	$E[s^3(n)] = 81d^2n^2/16 + 63d^2n/16 + 729n^3/64 + 9n^2(4d^2 - 9)/32 + 9n^2(4d^2 + 9)/16 + 567n^2/64 + 3n(-6d^2 - 21)/8 + 3n(6d^2 - 12)/16 + 243n/32$
STUTTERINGB – FIG. 1(B)	1	0.49	$E[s(n)] = 9n/4$
	2	2.03	$E[s^2(n)] = 81n^2/16 + 347/16n + 128/3$
	3	7.43	$E[s^3(n)] = 729n^3/64 + 9369n^2/64 + 1359n/32 =$
STUTTERINGC – FIG. 1(C)	1	1.8	$E[s(n)] = 3n^3/8 + 3n^2/8 - n$
	2	72.5	$E[s^2(n)] = 9n^6/64 + 3n^5(8d^2 + 27)/160 + n^4(8d^4 + 84d^2 - 90)/192 + n^3(32d^4 + 216d^2 - 252)/288 + n^2(8d^4 + 44d^2 + 61)/64 + n(80d^4 + 324d^2 - 9)/1440$
	3	2144	$E[s^3(n)] = 27n^9/512 + 27n^8(16d^2 + 39)/2560 + 3n^7(824d^4 + 6444d^2 + 1242)/17920 + n^6(1900d^4 + 3996d^2 - 4365)/2560 + n^5(2004d^4 + 1704d^2 - 54)/2560 + n^4(-1900d^4 - 7056d^2 + 13446)/7680 + n^3(-6948d^4 - 12708d^2 - 6969)/7680 + n^2(-1900d^4 - 3114d^2 - 315)/3840 + n(-108d^4 - 603d^2 + 288)/6720$

(continued)

Table 1. (*continued*)

Program	Moment	Runtime (s)	Computed Moment-Based Invariants
STUTTERINGD – FIG. 1(D)	1	1.92	$E[s(n)] = 3n^3/8 + 3n^2/8 - n$
	2	46.3	$E[s^2(n)] = 9n^6/64 + 93n^5/32 + 1651n^4/96 + 2849n^3/72 + 2813n^2/64 + 5131n/288$
	3	2076	$E[s^3(n)] = 27n^9/512 + 1593n^8/512 + 94587n^7/1792 + 545971n^6/2560 + 270117n^5/1280 - 58585n^4/768 - 132599n^3/512 - 536539n^2/3840 - 771n/140$
STUTTERINGP	1	0.28	$E[s(n)] = 3np$
	2	1.68	$E[s^2(n)] = 11n^2p^2 + 3np(-2p + 1) + np(-p - 1) + 4np(-p + 2) - 1$
	3	6.05	$E[s^3(n)] = 27n_1^3p^3 - 3n_1^2p^3 + 3n_1^2p^2(-6p + 3) + 12n_1^2p^2(-3p + 3) + 12n_1^2p^2(-2p + 3) + 3n_1p(4p^2 - 3p + 3) + 3n_1p(8p^2 - 12p + 9) + n_1p(p^2 - 3p(-p - 1) - 3p + 2)/2 + 2n_1p(2p^2 - 6p(-p + 2) - 6p + 13) + 6$
SQUARE	1	0.38	$E[y(n)] = n^2 + n$
	2	2.46	$E[y^2(n)] = n^4 + 6 * n^3 + 3 * n^2 - 2 * n$
	3	8.70	$E[y^3(n)] = n^6 + 15 * n^5 + 45 * n^4 - 15 * n^3 - 30 * n^2 + 16 * n$

6 Related Work

Despite the impressive advancements [14, 16, 21], probabilistic model checking [1] tools [8, 20, 25] are in general not able to handle programs with unbounded and real variables. Model checking algorithms suffer from the state explosion problem and their performance in terms of time and memory consumption degrades as the number of reachable states to be considered increases. Furthermore, probabilistic model checking tools have no support for invariant generation. Our approach, based on symbolic summation over probabilistic expressions, can instead analyse probabilistic programs with a potentially infinite number of reachable states.

In [27], one of the first deductive frameworks to reason about probabilistic programs was proposed by annotating probabilistic programs with real-valued expressions over the expected values of program variables. Of particular interest are the annotations as *quantitative invariants*, summarising loop behaviors. The setting of [27] considers probabilistic programs where the stochastic inputs are restricted to discrete distributions with finite support and can deal also with demonic non-deterministic choice. Although our approach does not yet support demonic non-determinism, we are not restricted to discrete input distributions as long as we know their moments (e.g., the Gaussian distribution is characterised only by two moments: the mean and the variance). Moreover, our work is not restricted to quantitative invariants as invariants over expected values of program variables. Rather, we generate moment-based invariants that precisely capture invariant properties of higher-order and mixed moments of program variables.

Katoen et al. provided in [19] the first semi-automatic and complete method synthesising the linear quantitative invariants defined in [27]. The work of [19], implemented in PRINSYS [13], consists in annotating a loop with a linear template invariants and uses a constraint solver to find the parameters for which the template yields an invariant. The works [7,9] synthesize non-linear quantitative invariants.

In [3] the authors consider PPs with loops where the assignments in each loop iteration is statistically independent and identically distributed. This restriction is powerful enough to encode Bayesian networks in PPs and to obtain automatically a closed-form expression over the expected values of the program variables. Although the expression in the loop-guard can be more complex than in our setting, our approach can handle also assignments that depend on previous iterations.

The work in [11] proposes the PSI tool, a symbolic analysis system for exact inference in probabilistic programs with both continuous and discrete random variables. PSI can compute succinct symbolic representations of the joint posterior distribution represented by a given PP. However, the tool supports the analysis only of PP with specified number of loop iterations, while our approach can handle arbitrary number of loop iterations and also infinite loops.

Another related line of research is given in [2], where martingales are used to compute invariants of probabilistic programs. The martingales generated by [2] however heavily depend on the user-provided hints and hence less generic hints yield less expressive/precise invariants. Moreover, of [2] mainly focuses on invariants over expected values and it remains unclear which extensions of martingales need to be considered to compute higher-order moments. The work of [24] addresses such generalizations of martingales for computing higher-order moments of program variables, with the overall goal of approximating runtimes of randomized programs. The approach in [24] is however again restricted to user-provided templates. Unlike the works of [2, 7, 9, 13, 19, 24], our work does not rely on a priori given templates/hints, but computes the most precise invariant expression over higher-order or mixed moments of program variables. To do so, we use symbolic summation to compute closed forms of higher-order moments. In addition, Prob-solvable loops support parametrized distributions and symbolic probabilities, which is not the case of [2, 24].

There are two orthogonal problems related to quantitative invariants generation: program termination [10, 28] and worst-case execution [4, 6, 18]. The first is to assess whether a probabilistic program terminates with probability 1 or if the expected time of termination is bounded. In principle, one can use our approach to solve this class of problems for Prob-solvable loops, but this is not the focus of this paper. The second class of problems is related to finding bounds over the expected values. In [4] the authors consider bounds also over higher-order moments for a specific class of probabilistic programs with probabilistic affine assignments. This approach can handle also nonlinear terms using interval arithmetic and fresh variables, at the price to produce very conservative bounds. On the contrary our approach supports natively probabilistic polyno-

mial assignments (in the form of Prob-solvable loops) and provides a precise symbolic expression over higher-order moments.

7 Conclusion

We introduced a novel approach for automatically generating *moment-based invariants* of a subclass of probabilistic programs (PPs), called Prob-solvable loops, with polynomial assignments over random variables and parametrised distributions. We combine methods from symbolic summation and statistics to derive invariants over higher-order moments, such as expected values or variances, of program variables. To the best of our knowledge, our approach is the first fully automated method computing higher-order moments of PPs with infinite loops and polynomial assignments over random variables and parametrised distributions. Extending our approach to a richer class of PPs, in particular by supporting nested loops and demonic non-determinism, is an interesting line for future work.

Acknowledgements. We would like to thank Joost-Pieter Katoen for his constructive feedback on a preliminary version of the manuscript.

References

1. Baier, C., Katoen, J.P.: Principles of Model Checking. The MIT Press, Cambridge (2008)
2. Barthe, G., Espitau, T., Ferrer Fioriti, L.M., Hsu, J.: Synthesizing probabilistic invariants via Doob’s decomposition. In: Chaudhuri, S., Farzan, A. (eds.) CAV 2016. LNCS, vol. 9779, pp. 43–61. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41528-4_3
3. Batz, K., Kaminski, B.L., Katoen, J.-P., Matheja, C.: How long, O Bayesian network, will I sample thee? In: Ahmed, A. (ed.) ESOP 2018. LNCS, vol. 10801, pp. 186–213. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-89884-1_7
4. Bouissou, O., Goubault, E., Putot, S., Chakarov, A., Sankaranarayanan, S.: Uncertainty propagation using probabilistic affine forms and concentration of measure inequalities. In: Chechik, M., Raskin, J.-F. (eds.) TACAS 2016. LNCS, vol. 9636, pp. 225–243. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49674-9_13
5. Chakarov, A., Sankaranarayanan, S.: Expectation invariants for probabilistic program loops as fixed points. In: Müller-Olm, M., Seidl, H. (eds.) SAS 2014. LNCS, vol. 8723, pp. 85–100. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10936-7_6
6. Chatterjee, K., Fu, H., Goharshady, A.K., Goharshady, E.K.: Polynomial invariant generation for non-deterministic recursive programs. In: PLDI (2019, to appear)
7. Chen, Y.-F., Hong, C.-D., Wang, B.-Y., Zhang, L.: Counterexample-guided polynomial loop invariant generation by lagrange interpolation. In: Kroening, D., Păsăreanu, C.S. (eds.) CAV 2015. LNCS, vol. 9206, pp. 658–674. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21690-4_44

8. Dehnert, C., Junges, S., Katoen, J.-P., Volk, M.: A storm is coming: a modern probabilistic model checker. In: Majumdar, R., Kunčák, V. (eds.) CAV 2017. LNCS, vol. 10427, pp. 592–600. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63390-9_31
9. Feng, Y., Zhang, L., Jansen, D.N., Zhan, N., Xia, B.: Finding polynomial loop invariants for probabilistic programs. In: D’Souza, D., Narayan Kumar, K. (eds.) ATVA 2017. LNCS, vol. 10482, pp. 400–416. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68167-2_26
10. Fu, H., Chatterjee, K.: Termination of nondeterministic probabilistic programs. In: Enea, C., Piskac, R. (eds.) VMCAI 2019. LNCS, vol. 11388, pp. 468–490. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11245-5_22
11. Gehr, T., Misailovic, S., Vechev, M.: PSI: exact symbolic inference for probabilistic programs. In: Chaudhuri, S., Farzan, A. (eds.) CAV 2016. LNCS, vol. 9779, pp. 62–83. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41528-4_4
12. Ghahramani, Z.: Probabilistic machine learning and artificial intelligence. *Nature* **521**(7553), 452–459 (2015)
13. Gretz, F., Katoen, J.-P., McIver, A.: Prinsys—on a quest for probabilistic loop invariants. In: Joshi, K., Siegle, M., Stoelinga, M., D’Argenio, P.R. (eds.) QEST 2013. LNCS, vol. 8054, pp. 193–208. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40196-1_17
14. Hermanns, H., Wachter, B., Zhang, L.: Probabilistic CEGAR. In: Gupta, A., Malik, S. (eds.) CAV 2008. LNCS, vol. 5123, pp. 162–175. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-70545-1_16
15. Humenberger, A., Jaroschek, M., Kovács, L.: Aligator.jl – a Julia package for loop invariant generation. In: Rabe, F., Farmer, W.M., Passmore, G.O., Youssef, A. (eds.) CISM 2018. LNCS (LNAI), vol. 11006, pp. 111–117. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-96812-4_10
16. Jansen, N., Dehnert, C., Kaminski, B.L., Katoen, J.-P., Westhofen, L.: Bounded model checking for probabilistic programs. In: Artho, C., Legay, A., Peled, D. (eds.) ATVA 2016. LNCS, vol. 9938, pp. 68–85. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46520-3_5
17. Kaminski, B.L., Katoen, J., Matheja, C.: On the hardness of analyzing probabilistic programs. *Acta Inf.* **56**(3), 255–285 (2019)
18. Karp, R.M.: Probabilistic recurrence relations. *J. ACM* **41**(6), 1136–1150 (1994)
19. Katoen, J.-P., McIver, A.K., Meinicke, L.A., Morgan, C.C.: Linear-invariant generation for probabilistic programs: automated support for proof-based methods. In: Cousot, R., Martel, M. (eds.) SAS 2010. LNCS, vol. 6337, pp. 390–406. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15769-1_24
20. Katoen, J.P., Zapreev, I.S., Hahn, E.M., Hermanns, H., Jansen, D.N.: The ins and outs of the probabilistic model checker MRMC. *Perform. Eval.* **68**(2), 90–104 (2011)
21. Kattenbelt, M., Kwiatkowska, M.Z., Norman, G., Parker, D.: A game-based abstraction-refinement framework for Markov decision processes. *Formal Methods Syst. Des.* **36**(3), 246–280 (2010)
22. Kauers, M., Paule, P.: *The Concrete Tetrahedron - Symbolic Sums, Recurrence Equations, Generating Functions, Asymptotic Estimates*. Texts & Monographs in Symbolic Computation. Springer, Heidelberg (2011)
23. Kovács, L.: Reasoning algebraically about P-solvable loops. In: Ramakrishnan, C.R., Rehof, J. (eds.) TACAS 2008. LNCS, vol. 4963, pp. 249–264. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-78800-3_18

24. Kura, S., Urabe, N., Hasuo, I.: Tail probabilities for randomized program runtimes via martingales for higher moments. In: Vojnar, T., Zhang, L. (eds.) TACAS 2019. LNCS, vol. 11428, pp. 135–153. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-17465-1_8
25. Kwiatkowska, M., Norman, G., Parker, D.: PRISM 4.0: verification of probabilistic real-time systems. In: Gopalakrishnan, G., Qadeer, S. (eds.) CAV 2011. LNCS, vol. 6806, pp. 585–591. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22110-1_47
26. Lin, G.L.: Characterizations of Distributions via Moments. Indian Statistical Institute (1992)
27. McIver, A., Morgan, C.: Abstraction, Refinement and Proof for Probabilistic Systems. Monographs in Computer Science. Springer, Heidelberg (2005)
28. McIver, A., Morgan, C., Kaminski, B.L., Katoen, J.P.: A new proof rule for almost-sure termination. PACMPL **2**(POPL), 33:1–33:28 (2018)
29. Novi Inverardi, P.L., Tagliani, A.: Discrete distributions from moment generating function. Appl. Math. Comput. **182**(1), 200–209 (2006)