

# Proving uniformity and independence by self-composition and coupling

Gilles Barthe<sup>1</sup>, Thomas Espitau<sup>2</sup>, Benjamin Grégoire<sup>3</sup>  
Justin Hsu<sup>4</sup>, and Pierre-Yves Strub<sup>5</sup>

<sup>1</sup> IMDEA Software Institute    <sup>2</sup> Sorbonne Universités, UPMC Paris 6  
<sup>3</sup> Inria    <sup>4</sup> University of Pennsylvania    <sup>5</sup> École Polytechnique

## Abstract

*Proof by coupling* is a classical proof technique for establishing probabilistic properties of two probabilistic processes, like stochastic dominance and rapid mixing of Markov chains. More recently, couplings have been investigated as a useful abstraction for formal reasoning about relational properties of probabilistic programs, in particular for modeling reduction-based cryptographic proofs and for verifying differential privacy. In this paper, we demonstrate that probabilistic couplings can be used for verifying *non-relational* probabilistic properties. Specifically, we show that the program logic **pRHL**—whose proofs are formal versions of proofs by coupling—can be used for formalizing uniformity and probabilistic independence. We formally verify our main examples using the **EasyCrypt** proof assistant.

## 1 Introduction

Uniformity and probabilistic independence are two of the most useful and commonly encountered properties when analyzing randomized computations. Uniform distributions are a central building block of randomized algorithms. Arguably the simplest non-trivial distribution—the coin flip—is a uniform distribution over two values. Given access to uniform samples, there are known transformations for converting the samples to simulate more complex distributions, like Gaussian or Laplacian distributions. Conversely, turning samples from various non-uniform distributions into uniform samples is an active area of research.

Probabilistic independence is no less useful. The probability of a conjunction of independent events can be decomposed as a product of probabilities of individual events, each which can then be analyzed in isolation. Independent random variables are also needed to apply more sophisticated mathematical tools, like concentration inequalities.

Given these and other applications, it is not surprising that researchers have investigated different methods of reasoning about uniformity and independence. For instance, Pearl and Paz [18] develop an axiomatic theory based on *graphoids* for modeling conditional independence in probability theory. However, proving uniformity and independence by program verification remains a challenging task. Most verification techniques for probabilistic programs do not treat these properties as first-class assertions, and rely on reasoning principles that are cumbersome to use. Often, the only way to prove uniformity or independence is to prove exact values for the probability of specific events.

For example, consider a formal system for proving properties of the form  $\Pr_{\llbracket s \rrbracket_m} [E] = p$ , which capture the fact that the event  $E$  has probability  $p$  in the distribution obtained by executing the randomized program  $s$  on some initial memory  $m$  (many existing systems use this idea, e.g. [9, 13, 15, 17, 19, 20]). Suppose that we want to prove that a program variable  $x$  of some finite type  $A$  is uniformly distributed in the output distribution  $\llbracket s \rrbracket_m$ . The only way to show this property is to analyze the probability of each output: for every  $a \in A$ , prove that  $\Pr_{\llbracket s \rrbracket_m} [x = a] = \frac{1}{|A|}$ .

For independence, the situation is similar. Assume that we want to prove that the two program variables  $x$  and  $y$  of respective types  $A$  and  $B$  are (probabilistically) independent in the output distribution  $\llbracket s \rrbracket_m$ . This can be done by exhibiting functions  $f, g, h$  such that for every  $a \in A$  and  $b \in B$ , we have:  $\Pr_{\llbracket s \rrbracket_m} [x = a] = f(a)$ ,  $\Pr_{\llbracket s \rrbracket_m} [y = b] = g(b)$ ,  $\Pr_{\llbracket s \rrbracket_m} [x = a \wedge y = b] = h(a, b)$ . Then, independence between  $x$  and  $y$  holds by proving that  $h(a, b) = f(a) \cdot g(b)$  for every  $a \in A$  and  $b \in B$ .

While these approaches work in theory, they can be laborious in practice. It may be awkward to express the probability of  $x = a$ , and the functions  $f, g$  and  $h$  may be difficult to produce. The main contribution of this paper is an alternative method based on probabilistic couplings for proving uniformity and independence. Probabilistic couplings are a classical method for proving sophisticated probabilistic properties (e.g., stochastic dominance, rapid mixing of Markov chains, and more [16, 21, 22]). More recently, couplings have been used to reason about relational properties of probabilistic programs, notably differential privacy [5, 6]. Here we show that uniformity and independence properties can also be verified using coupling, despite being *non-relational* properties. As a consequence, our verification method inherits the many advantages of reasoning by couplings: compositional reasoning, and no need to reason directly about probabilistic events. Concretely, we show how uniformity and independence can be captured in the relational program logic pRHL [1].

In summary, our main contributions are novel methods to prove uniformity and independence properties of probabilistic programs. We prove the soundness of the methods and demonstrate their usefulness on a class of case studies.

## Detailed Contributions

**Uniformity.** Suppose we have a program  $s$  with a program variable  $x$  ranging over a finite set  $A$ , and we want to show that  $x$  is distributed uniformly over  $A$  after executing  $s$ . Rather than computing the probability of  $\Pr_{\llbracket s \rrbracket_m} [x = a]$  for each  $a \in A$ , it suffices to show that the probabilities of any two outputs are equal:

$$\forall a_1, a_2 \in A. \Pr_{\llbracket s \rrbracket_m} [x = a_1] = \Pr_{\llbracket s \rrbracket_m} [x = a_2].$$

Now, we can view uniformity as a relational property: if we consider two runs of  $s$ , then the probability of  $x$  being  $a_1$  in the first run should be equal to the probability of  $x$  being  $a_2$  in the second run. In pRHL, this property is described by the following judgment:

$$\forall a_1, a_2 \in A. \models s \sim s : \phi \implies x\langle 1 \rangle = a_1 \iff x\langle 2 \rangle = a_2$$

where the assertion  $\phi$  asserts that the initial states are equal.

**Independence.** Proving probabilistic independence is more involved. We show how to prove independence in two different ways. Assume that we want to prove that the program variables  $x$  and  $y$  of respective finite types  $A$  and  $B$  are independent. First, if the distribution of  $\langle x, y \rangle$  is

uniformly distributed over  $A \times B$ , then  $x$  and  $y$  are independent (and are themselves uniformly distributed). Indeed, assume that for all  $a \in A$  and  $b \in B$  we have  $\Pr_{\llbracket s \rrbracket_m} [x = a \wedge y = b] = \frac{1}{|A| \cdot |B|}$ . Then we have  $\Pr_{\llbracket s \rrbracket_m} [x = a] = \sum_{b \in B} \Pr_{\llbracket s \rrbracket_m} [x = a \wedge y = b] = \frac{1}{|A|}$ . A similar argument applies to the probability that  $y = b$ , from which independence follows. Thus, our first method of proving independence is by reduction to proving uniformity.

This approach is simple to use, but it only applies to proving independence of uniform random variables. A more expressive, but also slightly more complicated approach is to express probabilistic independence as a property of a modified version of the program, without any requirement on uniformity. More specifically, independence of  $x$  and  $y$  can be derived from the equality between the probabilities of  $x = a \wedge y = b$  and  $x_1 = a \wedge y_2 = b$ , where in the first case the probability is taken over the output of the original program  $s$ , and in the second case the probability is taken over the output of the program  $s_1; s_2$ , where  $s_1$  and  $s_2$  are renamings of  $s$  (we call  $s_1; s_2$  a *self-composition* of  $s$  [2, 12]). The reason is not hard to see. Since the composed programs operate on disjoint memory, the final combined output distribution models two independent runs of the original program  $s$ . So, the probability  $\Pr_{\llbracket s_1; s_2 \rrbracket_{m_1 \uplus m_2}} [x_1 = a \wedge y_2 = b]$ —where  $m_1 \uplus m_2$  is the disjoint union of two copies of  $m$ —is equal to the product of  $\Pr_{\llbracket s_1 \rrbracket_{m_1}} [x_1 = a]$  and  $\Pr_{\llbracket s_2 \rrbracket_{m_2}} [y_2 = b]$ . Since  $s_1$  and  $s_2$  are just renamed versions of the original program  $s$ , these probabilities are in turn equal to  $\Pr_{\llbracket s \rrbracket_m} [x = a]$  and  $\Pr_{\llbracket s \rrbracket_m} [y = b]$  in the original program.

Our encoding casts independence as a relational property between a program  $s$  and its self-composition  $s_1; s_2$ , a property which can be directly expressed in pRHL:

$$\forall a \in A, b \in B. \models s \sim s_1; s_2 : \phi \implies (x\langle 1 \rangle = a \wedge y\langle 1 \rangle = b) \iff (x_1\langle 2 \rangle = a \wedge y_2\langle 2 \rangle = b)$$

where the precondition  $\phi$  captures the initial conditions. We show that our approach extends to independence and conditional independence of sets of program variables.

**Outline** Section 2 and Section 3 provide the relevant mathematical background and introduce the setting of our work. Section 4, Section 5 and Section 6 respectively address the case of uniformity, independence, and conditional independence. In each case we demonstrate our method using classic examples of randomized algorithms. We conclude the paper with a discussion of alternative techniques for verifying these properties.

## 2 Mathematical Background

For the sake of simplicity, we restrict ourselves to discrete (countable) sub-distributions.

**Definition 1.** A sub-distribution over a set  $A$  is defined by a mass function  $\mu : A \rightarrow \mathbb{R}^+$ , which gives the probability of the unitary events  $a \in A$ . This mass function must be s.t.  $\sum_{a \in A} \mu(a)$  is well-defined and its weight satisfies  $|\mu| \triangleq \sum_{a \in A} \mu(a) \leq 1$ . In particular, the support of the sub-distribution  $\text{supp}(\mu) \triangleq \{a \in A \mid \mu(a) \neq 0\}$  is discrete. When  $|\mu|$  is equal to 1, we call  $\mu$  a distribution. We let  $\mathbb{D}(A)$  denote the set of sub-distributions over  $A$ . An event over  $A$  is a predicate over  $A$ . The probability of an event  $E$  in a sub-distribution  $\mu$ , written  $\Pr_{x \sim \mu} [E]$ , is defined as  $\sum_{\{x \in A \mid E(x)\}} \mu(x)$ .

When working with sub-distributions over tuples, the probabilistic versions of the usual projections on tuples are called *marginals*. For distributions over pairs, we define the *first* and *second marginals*  $\pi_1(\mu)$  and  $\pi_2(\mu)$  of a distribution  $\mu$  over  $A \times B$  by  $\pi_1(\mu)(a) \triangleq \sum_{b \in B} \mu(a, b)$  and  $\pi_2(\mu)(b) \triangleq \sum_{a \in A} \mu(a, b)$ . We are now ready to formally define coupling.

**Definition 2.** Let  $A_1$  and  $A_2$  be two sets, and let  $\Psi \subseteq A_1 \times A_2$ . A  $\Psi$ -coupling for two sub-distributions  $\mu_1, \mu_2$  resp. over  $A_1$  and  $A_2$  is a sub-distribution  $\mu \in \mathbb{D}(A_1 \times A_2)$  such that  $\pi_1(\mu) = \mu_1$  and  $\pi_2(\mu) = \mu_2$  and  $\text{supp}(\mu) \subseteq \Psi$ . We write  $\blacktriangleleft_{\Psi} \langle \mu_1 \& \mu_2 \rangle$  to denote the existence of a  $\Psi$ -coupling.

In addition to the general definition, we shall also consider a special case of coupling: specifically, we say that  $(\mu_1, \mu_2)$  are *f-coupled* if  $f : A_1 \rightarrow A_2$  is a bijection such that  $\mu_1(x) = \mu_2(f(x))$  for every  $x \in A_1$ . In this case, we write  $f \blacktriangleleft \langle \mu_1 \& \mu_2 \rangle$ .

Previous works establish a number of basic facts about couplings, see e.g. Barthe et al. [1, 5]. In particular, one useful consequence of couplings is that they can show that one event has smaller probability than another.

**Lemma 3** (Fundamental lemma of coupling). Let  $E_1$  and  $E_2$  be predicates over  $A_1$  and  $A_2$ , and let  $\Psi \triangleq \{(x_1, x_2) \mid (x_1 \in E_1) \Rightarrow (x_2 \in E_2)\}$ . If  $\blacktriangleleft_{\Psi} \langle \mu_1 \& \mu_2 \rangle$ , then  $\Pr_{x_1 \sim \mu_1} [E_1] \leq \Pr_{x_2 \sim \mu_2} [E_2]$ .

One can immediately derive a variant of the lemma where  $\iff$  and  $=$  are used in place of  $\Rightarrow$  and  $\leq$  respectively. The following lemma provides a converse to the fundamental lemma of coupling in the special case where we are interested in proving the equality of two distributions.

**Lemma 4.** For every  $\mu_1, \mu_2 \in \mathbb{D}(A)$ , the following are equivalent:

- $\mu_1 = \mu_2$ ;
- for every  $a \in A$ ,  $\Pr_{x \sim \mu_1} [x = a] = \Pr_{x \sim \mu_2} [x = a]$ ;
- for every  $a \in A$ ,  $\blacktriangleleft_{\Psi_a} \langle \mu_1 \& \mu_2 \rangle$  where  $\Psi_a \triangleq \{(x_1, x_2) \mid x_1 = a \iff x_2 = a\}$ ;
- $\blacktriangleleft_{\Psi_A} \langle \mu_1 \& \mu_2 \rangle$  where  $\Psi_A \triangleq \{(x_1, x_2) \mid x_1 = x_2\}$ .

We note that the third item (existence of liftings for pointwise equality) is often easier to establish than the last item (existence of lifting for equality), since one can choose the coupling for each possible value of  $a$ , rather than showing a single coupling for all values of  $a$ .

### 3 Setting

We will work with a simple probabilistic imperative language. Probabilistic assignments are of the form  $x \stackrel{\$}{\leftarrow} g$ , which assigns a value sampled according to the distribution  $g$  to the program variable  $x$ . The syntax of statements is defined by the grammar:

$$s ::= \text{skip} \mid \text{abort} \mid x \leftarrow e \mid x \stackrel{\$}{\leftarrow} g \mid s; s \mid \text{if } e \text{ then } s \text{ else } s \mid \text{while } e \text{ do } s$$

where  $x, e$  and  $g$  respectively range over (typed) variables in  $\mathcal{X}$ , expressions in  $\mathcal{E}$  and distributions in  $\mathcal{D}$ . To ensure that the set of states is countable, we require that there are finitely many variables  $\mathcal{X}$ . As usual  $\mathcal{E}$  is defined inductively from  $\mathcal{X}$  and a set  $\mathcal{F}$  of simply typed function symbols. In this paper, distributions used for sampling are either uniform distributions over a finite type  $A$ , or the Bernoulli distribution with parameter  $p$ , which we denote by  $\mathbf{Bern}(p)$ . We assume that expressions and statements are typed in the usual way.

We assume we are given a set-theoretical interpretation for every type and operator of the language. We define a state as a type-preserving mapping from variables to values, and we let  $\mathbf{State}$  denote the set of states. The set of states is equipped with the usual functions for reading and writing a value; we use  $m(x)$  to denote the value of  $x$  in  $m$ , and  $m[x := v]$  to denote state update, in this case the state obtained from  $m$  by updating the value of  $x$  with  $v$ .

$$\begin{aligned}
\llbracket \mathbf{skip} \rrbracket_m &= \delta_m & \llbracket \mathbf{abort} \rrbracket_m &= \mathbf{0} \\
\llbracket x \leftarrow e \rrbracket_m &= \delta_{m[x:=\llbracket e \rrbracket_m]} & \llbracket x \leftarrow^{\#} g \rrbracket_m &= \mathbb{E}_{v \sim \llbracket g \rrbracket_m} [\delta_{m[x:=v]}] \\
\llbracket s_1; s_2 \rrbracket_m &= \mathbb{E}_{\xi \sim \llbracket s_1 \rrbracket_m} [\llbracket s_2 \rrbracket_{\xi}] \\
\llbracket \mathbf{if } e \mathbf{ then } s_1 \mathbf{ else } s_2 \rrbracket_m &= \mathbf{if } \llbracket e \rrbracket_m \mathbf{ then } \llbracket s_1 \rrbracket_m \mathbf{ else } \llbracket s_2 \rrbracket_m \\
\llbracket \mathbf{while } b \mathbf{ do } s \rrbracket_m &= \lim_{n \rightarrow \infty} \llbracket (\mathbf{if } b \mathbf{ then } s)^{[n]}; \mathbf{if } b \mathbf{ then abort} \rrbracket_m
\end{aligned}$$

where  $s^{[n]} \triangleq \overbrace{s; \dots; s}^{n \text{ times}}$ .

Figure 1: Denotational semantics of programs

One can equip  $\mathbb{D}(\mathbf{State})$  with a monadic structure, using the Dirac distributions  $\delta_x$  for the unit and *distribution expectation*  $\mathbb{E}_{x \sim \mu}[M(x)]$  for the bind, where

$$\mathbb{E}_{x \sim \mu}[M(x)] : x \mapsto \sum_a \mu(a) \cdot M(a)(x).$$

The semantics of expressions and distribution expressions is parametrized by a state  $m$ , and is defined in the usual way where we require all distribution expressions to be interpreted as proper distributions (sub-distributions with weight 1).

**Definition 5** (Semantics of statements).

- The semantics  $\llbracket s \rrbracket_m$  of a statement  $s$  w.r.t. to some initial state  $m$  is a sub-distribution over states, and is defined by the clauses of Fig. 1.
- The (lifted) semantics  $\llbracket s \rrbracket_{\mu}$  of a statement  $s$  w.r.t. to some initial sub-distribution  $\mu$  over states is a sub-distribution over states, and is defined as  $\llbracket s \rrbracket_{\mu} \triangleq \mathbb{E}_{m \sim \mu}[\llbracket s \rrbracket_m]$   $\mu \in \mathbb{D}(\mathbf{State})$ .

A basic and highly important property of probabilistic programs is termination. We say that a program  $s$  is *lossless* if for every initial memory  $m$ ,  $|\llbracket s \rrbracket_m| = 1$ . By now, there are many sophisticated techniques for proving losslessness even for languages that allow both probabilistic sampling and non-determinism (including recent advances by Chatterjee et al. [10, 11], Ferrer Fioriti and Hermanns [14]). These techniques are capable of showing losslessness for all of our examples (in some cases with a high degree of automation), so throughout the paper, we assume that all programs are lossless. This assumption is used in the rules of pRHL and the characterizations of uniformity and independence.

### 3.1 Self-Composition of Programs

For every program  $s$  and  $n \in \mathbb{N}$ , we let  $s^{(n)}$  denote the  $n$ -fold self-composition of  $s$ , i.e.  $s^{(n)} \triangleq s_1; \dots; s_n$ , where each  $s_i$  is a copy of  $s$  where all variables are tagged with a superscript  $i$ . In order to state the main property of self-composition, we define the self-composition of a state; given a state  $m$ , we define its  $n$ -fold self-composition  $m^{(n)}$  as the state from  $\mathcal{X}^{(n)}$  to values, where  $\mathcal{X}^{(n)} \triangleq \{x^i \mid x \in \mathcal{X}, 1 \leq i \leq n\}$  such that for every  $x$  and  $i$ ,  $m^{(n)}(x^i) \triangleq m(x)$ . Given a state  $m$  from  $\mathcal{X}^{(n)}$ , we denote by  $m_i$  the  $i$ -th projection of  $m$ .

**Proposition 6.** *For every program  $s$  and state  $m$ , we have*

$$\Pr_{\llbracket s^{(n)} \rrbracket_{m^{(n)}}} [\wedge_{1 \leq i \leq n} E_i] = \prod_{1 \leq i \leq n} \Pr_{\llbracket s \rrbracket_m} [E_i]$$

where the event  $E^i$  is defined by  $E^i(m^{(n)}) \triangleq E(\pi_i(m'))$  for every  $i$  and  $\pi_i$  is the projection from a self-composed state to its  $i$ -th component.

### 3.2 Probabilistic Relational Hoare Logic

Probabilistic Relational Hoare Logic (pRHL) is a program logic for reasoning about relational properties of probabilistic programs. Its judgments are of the form  $\models s_1 \sim s_2 : \phi \Longrightarrow \psi$ , where  $s_1$  and  $s_2$  are commands and the pre-condition  $\phi$  and the post-condition  $\psi$  are relational assertions, i.e. first-order formulae built over generalized expressions. The latter are similar to expressions, except that each variable is tagged with  $\langle 1 \rangle$  or  $\langle 2 \rangle$  to indicate the execution that it belongs to; we call the two executions *left* and *right*. Generalized expressions are interpreted w.r.t. a pair  $(m_1, m_2)$  of states, where the interpretation of the tagged variables  $x\langle 1 \rangle$  and  $x\langle 2 \rangle$  are  $m_1(x)$  and  $m_2(x)$  respectively. We write  $(m_1, m_2) \models \phi$  to denote that the interpretation of the assertion  $\phi$  w.r.t.  $(m_1, m_2)$  is valid.

**Definition 7.** A judgment  $\models s_1 \sim s_2 : \phi \Longrightarrow \psi$  is valid iff for every states  $m_1$  and  $m_2$ ,  $(m_1, m_2) \models \phi$  implies  $\blacktriangleleft_{\{(m'_1, m'_2) \mid (m'_1, m'_2) \models \psi\}} \llbracket s_1 \rrbracket_{m_1} \& \llbracket s_2 \rrbracket_{m_2}$ .

Fig. 2 presents the main rules of the logic; see Barthe et al. [1, 7] for the full system. The logic includes *two-sided* rules, which operate on both programs, and *one-sided* rules, which operate on a single program (left or right).

The [CONSEQ] rule is the rule of consequence, and reflects that validity is preserved by weakening the post-condition and strengthening the pre-condition. The [CASE] rule allows proving a judgment by case analysis; specifically, the validity of a judgment with pre-condition  $\Phi$  can be established from the validity of two judgments, one where the pre-condition is strengthened with  $\Xi$  and the other where the pre-condition is strengthened with  $\neg \Xi$ .

The [STRUCT] rule allows replacing programs by provably equivalent programs. The rules for proving program equivalence are given in Fig. 3, and manipulate judgments of the form  $\Phi \vdash c \equiv c'$ , where  $\Phi$  is a relational assertion. The first rule ([WHILE-SPLIT]) splits a single loop into two loops (the first running while  $e'$  is true, and the second running for the remaining iterations); this transformation is useful for selecting different couplings in different program iterations. The second rule ([SWAP]) reorders two instructions, as long as they modify disjoint variables. This allows us to couple sampling instructions that may come from two different parts of the two programs.

Moving on to the two-sided rules, the [SEQ] rule for sequential composition simply reflects the compositional property of couplings. The [ASSG] rule is standard. The [RAND] rule informally takes a coupling between the two distributions used for sampling in the left and right programs, and requires that every element in the support of the coupling validates the post-condition. The rule is parametrized by a bijective function  $f$  from the domain of the first distribution to the domain of the second distribution. This bijection gives us the freedom to specify the relation between the two samples when we couple them. The [COND] rule states that two *synchronized if* statements can be related if their respective branches are also related. The [WHILE] rule is the standard while rule adapted to pRHL. Note that we require the guard of the two commands to be equal—so in particular the two loops must make the same number of iterations—and  $\Phi$  plays the role of the while loop invariant as usual.

The one-sided rules ASSG-L, RAND-L, COND-L and WHILE-L are similar to their two-sided variant, but only operate on the left program. The full system includes mirrored versions of each one-sided rule, for reasoning about the right program.

$$\begin{array}{c}
\text{CONSEQ} \frac{\begin{array}{c} \vdash s_1 \sim s_2 : \Phi \Longrightarrow \Psi \\ \Phi' \Longrightarrow \Phi \quad \Psi \Longrightarrow \Psi' \end{array}}{\vdash s_1 \sim s_2 : \Phi' \Longrightarrow \Psi'} \\
\text{CASE} \frac{\begin{array}{c} \vdash s_1 \sim s_2 : \Phi \wedge \Xi \Longrightarrow \Psi \\ \vdash s_1 \sim s_2 : \Phi \wedge \neg \Xi \Longrightarrow \Psi \end{array}}{\vdash s_1 \sim s_2 : \Phi \Longrightarrow \Psi} \\
\text{ASSG} \frac{\Phi \triangleq \Psi[e_1\langle 1 \rangle/x_1\langle 1 \rangle, e_2\langle 2 \rangle/x_2\langle 2 \rangle]}{\vdash x_1 \leftarrow e_1 \sim x_2 \leftarrow e_2 : \Phi \Longrightarrow \Psi} \\
\text{COND} \frac{\begin{array}{c} \Phi \Longrightarrow e_1 = e_2 \\ \vdash s_1 \sim s_2 : \Phi \wedge e_1 \Longrightarrow \Psi s \quad \vdash s'_1 \sim s'_2 : \Phi \wedge \neg e_1 \Longrightarrow \Psi s' \end{array}}{\vdash \mathbf{if } e_1 \mathbf{ then } s_1 \mathbf{ else } s'_1 \sim \mathbf{if } e_2 \mathbf{ then } s_2 \mathbf{ else } s'_2 : \Phi \Longrightarrow \Psi} \\
\text{WHILE} \frac{\vdash s_1 \sim s_2 : \Psi \wedge e_1\langle 1 \rangle \wedge e_2\langle 2 \rangle \Longrightarrow \Psi \wedge e_1\langle 1 \rangle = e_2\langle 2 \rangle}{\vdash \mathbf{while } e_1 \mathbf{ do } s_1 \sim \mathbf{while } e_2 \mathbf{ do } s_2 : \Psi \wedge e_1\langle 1 \rangle = e_2\langle 2 \rangle \Longrightarrow \Psi \wedge \neg e_1\langle 1 \rangle \wedge \neg e_2\langle 2 \rangle} \\
\text{ASSG-L} \frac{\Phi \triangleq \Psi[e_1\langle 1 \rangle/x_1\langle 1 \rangle]}{\vdash x_1 \leftarrow e_1 \sim \mathbf{skip} : \Phi \Longrightarrow \Psi} \\
\text{RAND-L} \frac{\Phi \triangleq \forall v_1 \in \text{supp}(g_1), \Psi[v_1/x_1\langle 1 \rangle]}{\vdash x_1 \stackrel{\$}{\leftarrow} g_1 \sim \mathbf{skip} : \Phi \Longrightarrow \Psi} \\
\text{COND-L} \frac{\begin{array}{c} \vdash s_1 \sim s_2 : \Phi \wedge e_1\langle 1 \rangle \Longrightarrow \Psi \\ \vdash s'_1 \sim s_2 : \Phi \wedge \neg e_1\langle 1 \rangle \Longrightarrow \Psi \end{array}}{\vdash \mathbf{if } e_1 \mathbf{ then } s_1 \mathbf{ else } s'_1 \sim s_2 : \Phi \Longrightarrow \Psi} \\
\text{WHILE-L} \frac{\vdash s_1 \sim \mathbf{skip} : \Psi \wedge e_1\langle 1 \rangle \Longrightarrow \Psi}{\vdash \mathbf{while } e_1 \mathbf{ do } s_1 \sim \mathbf{skip} : \Psi \Longrightarrow \Psi \wedge \neg e_1\langle 1 \rangle} \\
\text{STRUCT} \frac{\begin{array}{c} \vdash s_1 \sim s_2 : \Phi \Longrightarrow \Psi \\ \Phi \vdash s_1 \equiv s'_1 \quad \Phi \vdash s_2 \equiv s'_2 \end{array}}{\vdash s'_1 \sim s'_2 : \Phi \Longrightarrow \Psi} \\
\text{SEQ} \frac{\begin{array}{c} \vdash s_1 \sim s_2 : \Phi \Longrightarrow \Xi \\ \vdash s'_1 \sim s'_2 : \Xi \Longrightarrow \Psi \end{array}}{\vdash s_1; s'_1 \sim s_2; s'_2 : \Phi \Longrightarrow \Psi} \\
\text{RAND} \frac{f \blacktriangleleft \langle g_1 \& g_2 \rangle}{\Phi \triangleq \forall v. \Psi[v/x_1\langle 1 \rangle, f(v)/x_2\langle 2 \rangle]} \\
\text{RAND} \frac{\Phi \triangleq \forall v. \Psi[v/x_1\langle 1 \rangle, f(v)/x_2\langle 2 \rangle]}{\vdash x_1 \stackrel{\$}{\leftarrow} g_1 \sim x_2 \stackrel{\$}{\leftarrow} g_2 : \Phi \Longrightarrow \Psi}
\end{array}$$

Figure 2: Proof rules (selection)

$$\text{WHILE-SPLIT} \frac{}{\Phi \vdash \mathbf{while } e \mathbf{ do } s \equiv \mathbf{while } e \wedge e' \mathbf{ do } s; \mathbf{while } e \mathbf{ do } s} \quad \text{SWAP} \frac{\text{var}(s_1) \cap \text{var}(s_2) = \emptyset}{\Phi \vdash s_1; s_2 \equiv s_2; s_1}$$

Figure 3: Equivalence rules (selection)

Throughout the paper, we often assert that the left and the right copies of a state are equal. This is captured by the relational assertion  $\text{EqMem} \triangleq \bigwedge_{x \in \mathcal{X}} x\langle 1 \rangle = x\langle 2 \rangle$ . We also often assert cross-equality on  $n$ -fold composition of states  $\text{EqMem}^{(p), (q)} \triangleq \bigwedge_{x \in \mathcal{X}, 1 \leq i \leq p, 1 \leq j \leq q} x^i\langle 1 \rangle = x^j\langle 2 \rangle$ .

## 4 Uniformity

Reasoning about probabilistic programs often requires establishing that a set of program variables (each ranging over a finite type) is uniformly distributed:

**Definition 8.** A set  $X = \{x_1, \dots, x_n\}$  of program variables of finite types  $A_1, \dots, A_n$  is uniformly distributed in a distribution  $\mu \in \mathbb{D}(\mathbf{State})$  iff for every  $(a_1, \dots, a_n) \in A_1 \times \dots \times A_n$ :

$$\Pr_\mu \left[ \bigwedge_{1 \leq i \leq n} x_i = a_i \right] = \prod_{1 \leq i \leq n} \frac{1}{|A_i|}$$

Note that the definition of uniformity (and as we will see in later sections, the definition of independence) naturally extends to sets of expressions, and so do our characterizations.

### 4.1 Characterization

The following proposition characterizes uniformity in terms of couplings.

**Proposition 9** (Uniformity by coupling). *Let  $X = \{x_1, \dots, x_n\}$  be a set of variables of respective finite types  $A_1, \dots, A_n$ . For every program  $s$ , the following are equivalent:*

1. for every state  $m$ ,  $X$  is uniformly distributed in  $\llbracket s \rrbracket_m$ ;
2. for every two tuples  $(a_1, \dots, a_n), (a'_1, \dots, a'_n) \in A_1 \times \dots \times A_n$ , we have

$$\models s \sim s : \text{EqMem} \implies \left( \bigwedge_{1 \leq i \leq n} x_i\langle 1 \rangle = a_i \right) \iff \left( \bigwedge_{1 \leq i \leq n} x_i\langle 2 \rangle = a'_i \right).$$

*Proof.* [1.  $\Rightarrow$  2.] Let  $m$  be a memory and assume that  $X$  is uniformly distributed in  $\llbracket s \rrbracket_m$ . Let  $(a_1, \dots, a_n), (a'_1, \dots, a'_n) \in A_1 \times \dots \times A_n$ . We denote by  $f : \mathbf{State} \rightarrow \mathbf{State}$  the bijection defined by

$$\begin{cases} f(m) = m[x_i \leftarrow a'_i]_{1 \leq i \leq n} & \text{if } \forall i. m[x_i] = a_i \\ f(m) = m[x_i \leftarrow a_i]_{1 \leq i \leq n} & \text{if } \forall i. m[x_i] = a'_i \\ f(m) = m & \text{otherwise.} \end{cases}$$

Let  $\eta \in \mathbb{D}(\mathbf{State} \times \mathbf{State})$  be the distribution defined by  $\eta(m_1, m_2) = \llbracket s \rrbracket_m(m_1)$  if  $m_2 = f(m_1)$ , and  $\eta(m_1, m_2) = 0$  otherwise. We prove that  $\eta$  is a  $\Psi$ -coupling for  $\llbracket s \rrbracket_m$ , where

$$\psi \triangleq \left( \bigwedge_{1 \leq i \leq n} x_i\langle 1 \rangle = a_i \right) \iff \left( \bigwedge_{1 \leq i \leq n} x_i\langle 2 \rangle = a'_i \right).$$

Regarding the marginals, we have:

$$\begin{aligned} \pi_1(\eta)(m_1) &= \sum_{m_2} \eta(m_1, m_2) = \eta(m_1, f(m_1)) = \llbracket s \rrbracket_m(m_1) \\ \pi_2(\eta)(m_2) &= \sum_{m_1} \eta(m_1, m_2) = \eta(f^{-1}(m_2), m_2) = \llbracket s \rrbracket_m(f^{-1}(m_2)) \\ &= \llbracket s \rrbracket_m(f(m_2)) = \llbracket s \rrbracket_m(m_2), \end{aligned}$$



the last equality being a consequence of  $X$  being uniformly distributed in  $\llbracket s \rrbracket_m$ . Moreover, for  $(m_1, m_2) \in \text{supp}(\eta)$ , we have  $m_2 = f(m_1)$ . Thus,  $m_1[x_i] = a_i$  iff  $m_2[x_i] = a'_i$ , and  $m_1, m_2 \models \Psi$ .

[2.  $\Rightarrow$  1.] Let  $(a_1, \dots, a_n), (a'_1, \dots, a'_n) \in A_1 \times \dots \times A_n$  and assume that  $\models s \sim s : \text{EqMem} \implies \Psi$ , where  $\Psi$  is defined as in the previous case. Since  $m, m \models \text{EqMem}$ , by Lemma 3 we have:

$$\Pr_{\llbracket s \rrbracket_m} \left[ \bigwedge_{1 \leq i \leq n} x_i = a_i \right] = \Pr_{\llbracket s \rrbracket_m} \left[ \bigwedge_{1 \leq i \leq n} x_i = a'_i \right],$$

showing that  $X$  is uniform in  $\llbracket s \rrbracket_m$ . □

By expressing uniformity as a coupling property, we can use pRHL to prove uniformity. To demonstrate the technique, we consider classical examples from the theory of randomized algorithms.

## 4.2 Simulating a Fair Coin

```

x ← 0;
y ← 0;
while x = y do
  x  $\stackrel{\$}{\leftarrow}$  Bern(p);
  y  $\stackrel{\$}{\leftarrow}$  Bern(p);

```

Figure 4: Bernoulli uniformizer

This example considers a process for simulating a fair coin using a biased coin. The idea is simple: 1) toss the coin twice; 2) if the two outcomes differ, return the value of the first coin; 3) if the two outcomes match, repeat from step 1. The algorithm does not require the bias of the coin to be known, as long as it is some constant bias and there is positive probability of returning 0 and 1. This process can be modelled by the program  $s$  from Fig. 4, where  $0 < p < 1$  is a real parameter modeling the probability of the biased coin to return 0 (tail). Our goal is to establish the trivial judgment  $\{\top\} s \{\top\}$  and the following pRHL judgment:

$$\models s \sim s : \top \implies x\langle 1 \rangle \iff \neg x\langle 2 \rangle$$

By the fundamental lemma of coupling, this implies that  $\Pr_{\llbracket s \rrbracket_m} [x = 1] = \Pr_{\llbracket s \rrbracket_m} [x = 0]$ , and hence that  $x$  is uniformly distributed upon termination. The proof proceeds by establishing the following invariant:

$$x\langle 2 \rangle = \text{if } x\langle 1 \rangle = y\langle 1 \rangle \text{ then } y\langle 2 \rangle \text{ else } \neg x\langle 1 \rangle$$

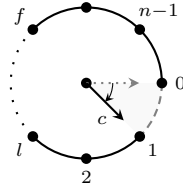
Validity of the invariant entails that the desired postcondition holds when the program exits, as the invariant and the negation of the loop guard both hold. The invariant holds when entering the loop, so we only need to prove that it is preserved by the loop body. The proof proceeds as follows: first, we swap the two random assignments on the right, leading to the judgment:

$$\models (x \stackrel{\$}{\leftarrow} \mathbf{Bern}(p); \quad y \stackrel{\$}{\leftarrow} \mathbf{Bern}(p)) \sim (y \stackrel{\$}{\leftarrow} \mathbf{Bern}(p); \quad x \stackrel{\$}{\leftarrow} \mathbf{Bern}(p)) : \phi' \implies \phi$$

where  $\phi$  denotes the loop invariant and  $\phi'$  denotes its strengthening by the loop guard—we do not need the precondition, since the values are freshly sampled in the body. Next, we apply the [RAND] rule twice, with the identity bijection. The required pre-condition

$$\forall v_1, v_2, v_2 = (\text{if } v_1 = v_2 \text{ then } v_1 \text{ else } \neg v_1)$$

is clearly true.



```

d ← 0; c ← 0; f ← 0; l ← 0;
while l + 1 ≤ f do
  d  $\stackrel{\$}{\leftarrow}$   $\mathcal{U}_{\{-1,1\}}$ ;
  if c = l ∧ d = 1 then l ← l + 1;
  if c = f ∧ d = -1 then f ← f - 1;
  c ← c + d;
ret ← (l, l + 1)
    
```

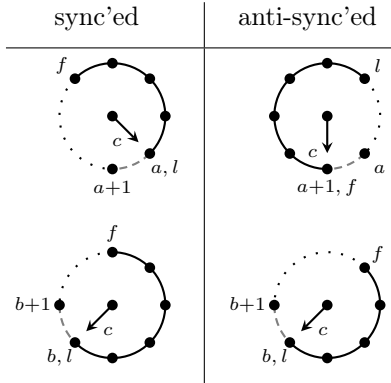
Figure 5: Cyclic random walk

### 4.3 Cyclic Random Walk

Consider a random walk over a cyclic path composed of  $n$  nodes labeled  $0, 1, \dots, n - 1$ : starting from position 0, at each step, we flip a fair coin over  $\{-1, 1\}$  and update the position accordingly to the result of the coin flip. To take into account that we are on a cyclic structure, all arithmetical operations are in the cyclic ring  $\mathbb{Z}/n\mathbb{Z}$ —i.e. are performed modulo  $n$ . At each iteration, when moving between two contiguous positions over the circle, we consider that the random walk visited the arc between the two nodes. We want to show that the last visited arc is uniformly distributed. Fig. 5 (left) gives a graphical representation of the random walk, where  $c$  is the random walk position and the dashed arc is the last visited arc when  $c$  moved from 0 to 1.

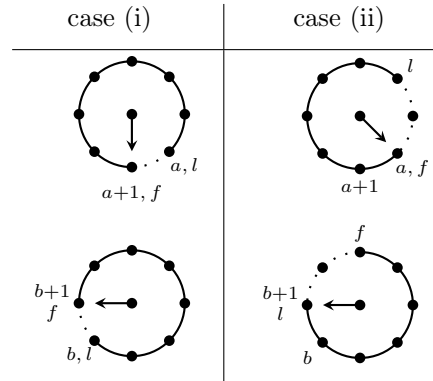
This process can be seen as a simple version of an algorithm that samples a uniformly random spanning tree on a graph—when the graph is a cycle, a spanning tree visits all but one of the edges. While Broder [8] analyzes the general problem, we can verify uniformity for the cyclic random walk with couplings.

The proof proceeds as follows. We imagine executing two random walks, from the same initial position. The goal is to couple the walks so that  $(a, a + 1)$  is the last arc in the first walk if and only if  $(b, b + 1)$  is the last arc in the second walk. If we can show this property for all  $a, b$ , then this coupling argument shows that any two arcs have the same probability of being the last arc, hence the last arc must be uniformly distributed.



To describe the coupling informally, we first execute asynchronously the two random walks until they eventually synchronize respectively on the arcs  $(a, a + 1)$  and  $(b, b + 1)$ . At that point, we are in one of the following cases: either the random walks synchronize on the same side of the arcs  $(a, a + 1)$  and  $(b, b + 1)$ , or they synchronize on opposite sides. (These cases are depicted on the left diagrams above, where the arc we want to synchronize on is dashed.) From that point, we execute the two processes resp. in lock-step (if they synchronized on the same side) or anti-lock-step (if they did not).

At some point, both processes will visit the other side of the arcs  $(a, a + 1)$  and  $(b, b + 1)$ , and since they execute in (anti)-lock-step, these events will occur synchronously.



At that point, either the processes finished their walk and they resp. return the arcs  $(a, a + 1)$  and  $(b, b + 1)$  as their result (case (i) of the right diagram above), or they have other nodes to visit and so they *will not* resp. return the arcs  $(a, a + 1)$  and  $(b, b + 1)$  (case (ii) of the same diagram).

We now detail the formal proof. Consider the program of Fig. 5, where all arithmetical operations are done modulo  $n$ . This algorithm instruments the random walk with two points  $f$  and  $l$  representing the range  $[f, l]$  (using clockwise ordering) of all the points that have been visited by the walk. When all nodes of the cycle have been visited (i.e. when  $l + 1 = f$ ), the arc between  $l$  and  $l + 1$  is the only arc that has not been visited by the walk. Let  $s$  be the program of Fig. 5 and  $s'$  the loop body of the single loop of  $s$ . We want to show that the final arc  $ret$ —the only arc that has not been marked—is uniformly distributed among all arcs. This follows by the judgment:

$$\forall a, b \in \mathbb{Z}/n\mathbb{Z}, \models s \sim s : \top \Longrightarrow ret\langle 1 \rangle = (a, a + 1) \iff ret\langle 2 \rangle = (b, b + 1).$$

First, we make use of the loop splitting equivalence rule ([WHILE-SPLIT]) to transform the main loop into three pieces. In the left program:

```

while ( $|v| < n \wedge a, a + 1 \notin [f, l]$ ) do  $s'$ ;
while ( $|v| < n \wedge \neg(a, a + 1 \in [f, l])$ ) do  $s'$ ;
while ( $|v| < n$ ) do  $s'$ 

```

where  $[f, l]$  represent the range  $f, f + 1, \dots, l$ . We use a similar transformation on the right program, with  $b$  in place of  $a$ . To carry out the proof, we first use the one-sided loop rules ([WHILE-L] and the corresponding version [WHILE-R]) on the first loops of the left and right programs. This part of the proof correspond to the walks synchronization as described above. By a straightforward loop invariant, we can show that

$$\Phi \wedge P(a)\langle 1 \rangle \wedge P(b)\langle 2 \rangle$$

holds after the first loops, where  $P(x) \triangleq (x \in [f, l] \oplus (x + 1) \in [f, l])$  and  $\Phi \triangleq \forall i \in \{1, 2\}. (c \in [f, l])\langle i \rangle$  indicates that the current positions  $(c\langle 1 \rangle, c\langle 2 \rangle)$  are contained in the range of visited arcs. Next, we show that after the two second loops the following relational invariant is satisfied:

$$(a, a + 1 \in [f, l])\langle 1 \rangle \wedge (b, b + 1 \in [f, l])\langle 2 \rangle \wedge (l\langle 1 \rangle = a \iff l\langle 2 \rangle = b)$$

After the second loop, there are two cases for the third loop. If  $l\langle 1 \rangle = a$ , we have  $l\langle 2 \rangle = b$ ,  $f\langle 1 \rangle = a + 1$  (since  $a + 1$  is visited in  $\langle 1 \rangle$ ) and  $f\langle 2 \rangle = b + 1$ . In this case, which corresponds to the case (i) of the last diagram, the third loops both exit immediately and the random walks resp. return the arcs  $(a, a + 1)$  and  $(b, b + 1)$ . Otherwise, we have  $l\langle 1 \rangle \neq a$  and  $l\langle 2 \rangle \neq b$ , and we can show, using the rules [WHILE-L] and [WHILE-R], that  $l\langle 1 \rangle$  (resp.  $l\langle 2 \rangle$ ) will never be set to  $a$  (resp.  $b$ ). In this case, which corresponds to the case (ii) of the last diagram, we can show that the walks resp. return arcs distinct from  $(a, a + 1)$  and  $(b, b + 1)$ .

We now focus on the second loops, relating them with the two-sided variant of the [WHILE] rule. The particular coupling we choose will depend on the current positions in the two sides at the start of the second loops. If  $a \in [f, l]\langle 1 \rangle$  and  $b \in [f, l]\langle 2 \rangle$  then we have  $l\langle 1 \rangle = a$  and  $l\langle 2 \rangle = b$  (since  $a + 1 \notin [f, l]\langle 1 \rangle$ ) and we couple the walks to make identical moves. In that case, the key part of the loop invariant is:

$$\bigwedge \left\{ \begin{array}{l} \Phi \wedge (a \in [f, l])\langle 1 \rangle \wedge (b \in [f, l])\langle 2 \rangle \\ c\langle 1 \rangle - a = c\langle 2 \rangle - b \\ l\langle 1 \rangle = a \iff l\langle 2 \rangle = b \end{array} \right\}$$

The first line enforces some structural invariant and the second line enforces that both walks make identical moves relative to  $a$  and  $b$ . The main difficulty is to show that both loops are synchronized. Note that there are two reasons the loop may exit. If  $l$  has been incremented, then the increment will be done on both side. Otherwise, if  $f$  has been decremented to  $a + 1$ , then we have  $c(1) = f(1) = a + 2$ , so  $c(1) - a = c(2) - b = 2$  and  $c(2) = b + 2$  and the right loop will also decrement  $f$  to  $b + 1$ . The case  $a + 1 \in [f, l](1)$  and  $b + 1 \in [f, l](2)$  is very similar, by reversing the roles of  $f$  and  $l$ . The remaining two cases,  $a + 1 \in [f, l](1)$  and  $b \in [f, l](2)$  or  $a \in [f, l](1)$  and  $b + 1 \in [f, l](2)$  is similar except that we force the walks to be execute in anti-lock-step. Using the rule [CASE], we put together these four cases and we conclude by application of the rule for sequence.

#### 4.4 Ballot Theorem

So far, we have shown how couplings can be used to prove that a set of program variables is uniformly distributed. Couplings can also be used for showing that two events have the same probability, such as in the following example.

**Example 10** (Ballot Theorem). *Assume that voters must choose between two candidates  $A$  and  $B$ . The outcome of the vote is  $n_A$  votes for  $A$  and  $n_B$  votes for  $B$ , with  $n_A > n_B$ . Assuming that the order in which the votes are cast is uniformly random, the probability that  $A$  is always strictly ahead in partial counts is  $(n_A - n_B) / (n_A + n_B)$ .*

The process can be formalized by the program from Fig. 6. Here we use the list  $l$  to store intermediate results. Using  $l_i$  to denote the  $i$ -th element of the list  $l$ , the Ballot Theorem is captured by the statement:

$$\forall n_A, n_B. n_A > n_B \implies \Pr_{\llbracket s \rrbracket_m} \left[ \bigwedge_{1 \leq i \leq n} l_i \neq 0 \mid x_A = n_A \wedge x_B = n_B \right] = \frac{n_A - n_B}{n_A + n_B}.$$

```

r ← 0; x_A ← 0; x_B ← 0; l ← ε;
while |l| ≤ n do
  r ← {A, B};
  if r = A then ;
    x_A ← x_A + 1;
  else
    x_B ← x_B + 1;
  l ← l :: (x_A - x_B)

```

There exist many proofs of the Ballot Theorem; we formalize a proof that is sometimes called Andre’s reflection principle. The crux of the method is a coupling proof of the following fact: “bad” sequences starting with a vote to the loser are equi-probable with “bad” sequences starting with a vote to the winner, where a sequence of votes is “bad” if there is a tie at some point in the partial counts. Let  $\phi \triangleq (\bigvee_{1 \leq i \leq n} l_i = 0)$  and  $\psi \triangleq x_A = n_A \wedge x_B = n_B$ . The above facts are captured by the pRHL judgment (universally quantified over  $n_A$  and  $n_B$  such that  $n_A > n_B$ ):  $\models s \sim s : \top \implies \xi$  where

$$\xi \triangleq (l_1 \cdot l_n > 0 \wedge \phi \wedge \psi)(1) \iff (l_1 \cdot l_n < 0 \wedge \phi \wedge \psi)(2).$$

It follows from the properties of coupling that for every  $n_A$  and  $n_B$  such that  $n_A > n_B$ ,  $\Pr_{\llbracket s \rrbracket_m} [l_1 \cdot l_n > 0 \wedge \phi \wedge l_n = k] = \Pr_{\llbracket s \rrbracket_m} [l_1 \cdot l_n < 0 \wedge \phi \wedge l_n = k]$ . In terms of conditional probabilities, we have  $\Pr_{\llbracket s \rrbracket_m} [l_1 \cdot l_n > 0 \wedge \phi \mid \psi] = \Pr_{\llbracket s \rrbracket_m} [l_1 \cdot l_n < 0 \wedge \phi \mid \psi]$ . Now observe that any sequence that starts with a vote to  $B$  (i.e. the loser) is necessarily bad. Therefore,  $\Pr_{\llbracket s \rrbracket_m} [l_1 \cdot l_n < 0 \wedge \phi \mid \psi] = \Pr_{\llbracket s \rrbracket_m} [l_1 \cdot l_n < 0 \mid \psi]$ . By the above and elementary properties

of conditional independence:

$$\begin{aligned} \Pr_{\llbracket s \rrbracket_m} [\phi \mid \psi] &= \Pr_{\llbracket s \rrbracket_m} [l_1 \cdot l_n > 0 \wedge \phi \mid \psi] + \Pr_{\llbracket s \rrbracket_m} [l_1 \cdot l_n < 0 \wedge \phi \mid \psi] \\ &= 2 \cdot \Pr_{\llbracket s \rrbracket_m} [l_1 \cdot l_n < 0 \mid \psi]. \end{aligned}$$

Note that the probability in the right-hand side of the last equation represents the probability that the first vote goes to the loser, conditional on  $\psi$ . This turns out to be exactly  $\frac{n_B}{n_A+n_B}$ , so we conclude that  $\Pr_{\llbracket s \rrbracket_m} [\phi \mid \psi] = 2 \cdot \frac{n_B}{n_A+n_B}$  or equivalently  $\Pr_{\llbracket s \rrbracket_m} [-\phi \mid \psi] = \frac{n_A-n_B}{n_A+n_B}$  as desired.

We now turn to the proof of the **pRHL** judgments. By symmetry it suffices to consider the first judgment. Using the rule of consequence and the elimination rule for universal quantification, it suffices to prove for every  $i$ :

$$\models s \sim s : \top \implies l_1\langle 1 \rangle \cdot l_n\langle 1 \rangle > 0 \wedge l_i\langle 1 \rangle = 0 \wedge \psi\langle 1 \rangle \implies l_1\langle 2 \rangle \cdot l_n\langle 2 \rangle < 0 \wedge l_i\langle 2 \rangle = 0 \wedge \psi\langle 2 \rangle$$

We couple the samplings of  $x$  using the negation function until  $|l| = i$ , and then with the identity bijection. This establishes the following loop invariant, from which we can conclude:

$$(\forall j \leq i. l_j\langle 1 \rangle = -l_j\langle 2 \rangle) \wedge l_i\langle 1 \rangle = l_i\langle 2 \rangle = 0 \wedge (\forall j > i. l_j\langle 1 \rangle = l_j\langle 2 \rangle).$$

## 5 Independence

We now turn to characterizing probabilistic independence using couplings. We focus on probabilistic independence of program variables, a common task when reasoning about randomized computations. In our setting, the textbook definition of probabilistic independence can be cast as follows:

**Definition 11.** *A set  $X = \{x_1, \dots, x_n\}$  of program variables of types  $A_1, \dots, A_n$  is probabilistically independent in a distribution  $\mu \in \mathbb{D}(\mathbf{State})$  iff for every  $(a_1, \dots, a_n) \in A_1 \times \dots \times A_n$ :*

$$\Pr_\mu \left[ \bigwedge_{1 \leq i \leq n} x_i = a_i \right] = \prod_{1 \leq i \leq n} \Pr_\mu [x_i = a_i].$$

### 5.1 Characterization

Our first characterization of independence is based on the observation that uniformity entails independence.

**Fact 12** (Independence from uniformity). *From every state  $m$ , if  $X$  is uniformly distributed in  $\llbracket s \rrbracket_m$  then  $X$  is independent in  $\llbracket s \rrbracket_m$ .*

This observation enables proving independence by coupling, in the special case where variables are uniform and independent. For the general case, we will use an alternative characterization based on self-composition.

**Proposition 13** (Independence by coupling). *The following are equivalent:*

1. *for every state  $m$ ,  $X$  is independent in  $\llbracket s \rrbracket_m$ ;*
2. *the following judgment, between a single copy of the program on the one hand and a  $n$ -fold copy on the other hand, is derivable for every  $(a_1, \dots, a_n) \in A_1 \times \dots \times A_n$ :*

$$\models s \sim s^{(n)} : \text{EqMem}^{(1), (n)} \implies \bigwedge_{1 \leq i \leq n} x_i\langle 1 \rangle = a_i \iff \bigwedge_{1 \leq i \leq n} x_i^i\langle 2 \rangle = a_i.$$

*Proof.* The validity of the universally quantified pRHL judgment is equivalent to the following statement: for every  $a_1, \dots, a_n$  and  $n$ -fold copy  $m^{(n)}$  of some initial state  $m$ ,

$$\Pr_{\llbracket s \rrbracket_m} \left[ \bigwedge_{1 \leq i \leq n} x_i = a_i \right] = \Pr_{\llbracket s^{(n)} \rrbracket_{m^{(n)}}} \left[ \bigwedge_{1 \leq i \leq n} x_i^i = a_i \right] = \prod_{1 \leq i \leq n} \Pr_{\llbracket s \rrbracket_m} [x_i = a_i].$$

The last equality comes from the property of  $n$ -fold self-composition (Proposition 6).  $\square$

## 5.2 Pairwise Independence of Bits

Our first example is a well-known algorithm for generating  $2^n$  pairwise independent bits. The algorithm first samples  $n$  independent bits  $b_1 \dots b_n$ , and then defines for every subset  $X \subseteq \{1, \dots, n\}$  the bit  $z_X = \bigoplus_{i \in X} b_i$ .

```

for  $i = 1$  to  $n$  do
   $b_i \leftarrow \{0, 1\}$ ;
for  $j = 0$  to  $2^n - 1$  do
   $z_j \leftarrow \bigoplus_{k \in \text{bits}(j)} b_k$ ;

```

Figure 7: Pairwise independence

By our characterization based on self-composition, pairwise independence of  $z_j$  and  $z_{j'}$  for every  $j \neq j'$  is equivalent to the (universally quantified) pRHL judgment

$$\models s \sim s_1; s_2 : \top \Longrightarrow z_j \langle 1 \rangle = a \wedge z_{j'} \langle 1 \rangle = a' \iff z_j^1 \langle 2 \rangle = a \wedge z_{j'}^2 \langle 2 \rangle = a'.$$

Since  $j \neq j'$ , the two sets  $\text{bits}(j)$  and  $\text{bits}(j')$  must differ in at least one element. Let  $k_0$  be the smallest element in which they differ. Without loss of generality, we can assume that  $k_0 \notin \text{bits}(j)$  and  $k_0 \in \text{bits}(j')$ . The crux of the proof is to establish the following judgment:

$$\models s_l \sim s_r : \top \Longrightarrow z \langle 1 \rangle = a \wedge z' \langle 1 \rangle = a' \iff z \langle 2 \rangle = a \wedge z'' \langle 2 \rangle = a'$$

where  $z = \bigoplus_{k \in \text{bits}(j)} b_k$ ,  $z' = \bigoplus_{k \in \text{bits}(j')} b_k$  and  $z'' = \bigoplus_{k \in \text{bits}(j')} b'_k$  and

$$s_l \triangleq \text{for } i \in [1 \dots n] \setminus k_0 \text{ do } b_i \leftarrow \{0, 1\}; b_{k_0} \leftarrow \{0, 1\}$$

$$s_r \triangleq \text{for } i \in [1 \dots n] \setminus k_0 \text{ do } (b_i \leftarrow \{0, 1\}; b'_i \leftarrow \{0, 1\}); b_{k_0} \leftarrow \{0, 1\}; b'_{k_0} \leftarrow \{0, 1\}.$$

This is proved by coupling the variables of the two programs in an appropriate way. We couple the random samplings as follows:

- for every  $k \neq k_0$ , we couple  $b_k \langle 1 \rangle$  and  $b_k \langle 2 \rangle$  using the identity sampling;
- we use the RND-R rule for  $b'_k \langle 2 \rangle$  for every  $k \neq k_0$ ;
- we couple  $b_{k_0} \langle 1 \rangle$  and  $b'_{k_0} \langle 2 \rangle$  with the bijection which ensures

$$b_{k_0} \langle 1 \rangle \oplus \left( \bigoplus_{k \in \text{bits}(j') \setminus \{k_0\}} b_k \langle 1 \rangle \right) = b'_{k_0} \langle 2 \rangle \oplus \left( \bigoplus_{k \in \text{bits}(j') \setminus \{k_0\}} b'_k \langle 2 \rangle \right).$$

Putting everything together, the final proof obligation follows from the algebraic properties of  $\oplus$ .

### 5.3 $k$ -wise Independence

The previous example can be generalized to achieve  $k$ -wise independence for general  $k$ . Suppose we wish to generate  $n$  random variables that are  $k$ -wise independent. We will work in  $\mathbb{Z}/p\mathbb{Z}$ , the field of integers modulo a prime  $p$ , such that  $k \leq p$ . Let  $a_0, \dots, a_{k-1}$  be drawn uniformly at random from  $\mathbb{Z}/p\mathbb{Z}$  and define the family of random variables for every  $m \in \{1, \dots, n\}$ :

$$x_m = \sum_{j=0}^{k-1} a_j \cdot m^j,$$

where we take  $0^0 = 1$  by convention. The corresponding code is given in Fig. 8. Then, we can show that any collection of  $k$  distinct variables  $\{x_i\}_i$  is independent.

```

for  $i = 1$  to  $n$  do
   $a_i \leftarrow \mathbb{Z}/p\mathbb{Z}$ ;
for  $m = 0$  to  $n - 1$  do
   $x_m \leftarrow 0$ ;
  for  $j = 0$  to  $k - 1$  do
     $x_m \leftarrow a_j \cdot m^j$ ;
```

Figure 8:  $k$ -wise independence

For simplicity, we will show that the first  $k$  elements  $x_0, \dots, x_{k-1}$  are uniform, and hence independent. Let  $v_0, \dots, v_{k-1} \in \mathbb{Z}/n\mathbb{Z}$  be arbitrary elements of the field. Then the probability that  $(x_0, \dots, x_{k-1}) = (v_0, \dots, v_{k-1})$  is equal to  $p^{-k}$ . Indeed, the equations

$$\begin{cases} v_0 = a_0 \\ \vdots \\ v_{k-1} = \sum_{j=0}^{k-1} a_j \cdot (k-1)^j \end{cases}$$

define a system of linear equations with variables  $a_0, \dots, a_{k-1}$ . By basic linear algebra the system of equations has a unique solution for the variables  $a_0, \dots, a_{k-1}$ ,<sup>1</sup> which we denote  $(v_0^*, \dots, v_{k-1}^*)$ . Now consider the pRHL judgment that establishes uniformity:

$$\models s \sim s : \top \Longrightarrow x_0 \langle 1 \rangle = v_0 \wedge \dots \wedge x_{k-1} \langle 1 \rangle = v_{k-1} \iff x_0 \langle 2 \rangle = w_0 \wedge \dots \wedge x_{k-1} \langle 2 \rangle = w_{k-1}$$

By applying (relational) weakest precondition on the deterministic fragments of the program, the judgment is reduced to

$$\models s \sim s : \top \Longrightarrow a_0 \langle 1 \rangle = v_0^* \wedge \dots \wedge a_{k-1} \langle 1 \rangle = v_{k-1}^* \iff a_0 \langle 2 \rangle = w_0^* \wedge \dots \wedge a_{k-1} \langle 2 \rangle = w_{k-1}^*$$

We then repeatedly apply the rule for random sampling, with the permutation on  $\mathbb{Z}/p\mathbb{Z}$  that exchanges  $(v_i^*, w_i^*)$ .

## 6 Conditional Independence

Finally, we consider how to show conditional independence. Recall that the conditional probability  $\Pr_{x \sim \mu} [A \mid B]$  is defined when  $\Pr_{x \sim \mu} [B] \neq 0$  and satisfies  $\Pr_{x \sim \mu} [A \mid B] \triangleq \frac{\Pr_{x \sim \mu} [A \wedge B]}{\Pr_{x \sim \mu} [B]}$ .

<sup>1</sup>Let the Vandermonde matrix  $V(1, \dots, k-1) = (i^j)_{i,j}$  be

$$V(1, \dots, k-1) \cdot (a_0, \dots, a_{k-1})^T = (v_0, \dots, v_{k-1})^T.$$

The system of equations has a unique solution if and only if the matrix  $V(1, \dots, k-1)$  is invertible in the space of matrices over  $\mathbb{Z}/n\mathbb{Z}$ , which happens if and only if its determinant is non-zero mod  $n$ . Expanding,

$$\det(V(1, \dots, k-1)) = \prod_{i \neq j} (i-j) = \prod_{i=2}^{k-1} i!$$

Note that  $p$  does not divide the determinant by Gauss' lemma, since  $n$  can't divide any of the terms  $i!$  for any  $i < n$ . Therefore, the system of equations has a unique solution.

**Definition 14.** Let  $X = \{x_1, \dots, x_n\}$  be a set of program variables of types  $A_1, \dots, A_n$  and let  $E$  be an event. We say that  $X$  is independent conditioned on  $E$  in a distribution  $\mu \in \mathbb{D}(\mathbf{State})$  iff for every  $(a_1, \dots, a_n) \in A_1 \times \dots \times A_n$ :

$$\Pr_\mu \left[ \bigwedge_{1 \leq i \leq n} x_i = a_i \mid E \right] = \prod_{1 \leq i \leq n} \Pr_\mu [x_i = a_i \mid E].$$

(For this definition to make sense, we are implicitly assuming that  $\Pr_\mu [E] \neq 0$ .)

The following lemma unfolds the definition of conditional independence and is useful for the characterization of the next section.

**Lemma 15.** A set of variables  $X$  is independent conditioned on an event  $E$  in  $\mu$  iff for every  $a_1 \in A_1, \dots, a_n \in A_n$ :

$$\Pr_\mu \left[ \bigwedge_{1 \leq i \leq n} x_i = a_i \wedge E \right] \cdot (\Pr_\mu [E])^{n-1} = \prod_{1 \leq i \leq n} \Pr_\mu [x_i = a_i \wedge E].$$

## 6.1 Characterization

The characterization of independence based on self-composition can be extended as follows.

**Proposition 16** (Conditional independence by coupling). *The following are equivalent:*

1. for every state  $m$ ,  $X$  is independent conditioned on  $E$  in  $\llbracket s \rrbracket_m$ ;
2.  $\models s^{(n)} \sim s^{(n)} : \text{EqMem}^{(n),(n)} \implies (\phi_1 \wedge \mathbb{E}\langle 1 \rangle) \iff (\phi_2 \wedge \mathbb{E}\langle 2 \rangle)$  where  $\mathbb{E} \triangleq \bigwedge_{1 \leq i \leq n} E^i$ ,  
 $\phi_1 \triangleq \bigwedge_{1 \leq i \leq n} (x_i^1 \langle 1 \rangle = a_i)$  and  $\phi_2 \triangleq \bigwedge_{1 \leq i \leq n} (x_i^i \langle 2 \rangle = a_i)$ .

*Proof.* The proof is similar to the case of independence (Proposition 13). □

## 6.2 Example: Conditional Independence

$$\begin{array}{l} x \stackrel{s}{\leftarrow} \mu; \\ y \stackrel{s}{\leftarrow} \mu'; \\ z \stackrel{s}{\leftarrow} \mu''; \\ w \leftarrow f(x, y); \\ w' \leftarrow g(y, z); \end{array}$$

We consider a simple example often used to illustrate Bayesian networks models. Let  $x, y, z, w$  and  $w'$  be random variables, where  $x, y$  and  $z$  are sampled from distributions  $\mu, \mu'$  and  $\mu''$  respectively, and  $w$  and  $w'$  are defined by their respective assignments. Both  $w$  and  $w'$  depend on  $y$ , along with independent sources of randomness, respectively  $x$  and  $z$ . While  $w$  and  $w'$  are not independent—they share dependence on  $y$ —if we *condition* on a particular value of  $y$ , then  $w$  and  $w'$  are independent.

Figure 9: Conditional indep. The code of the corresponding program  $s$  is given in Fig. 9.

We want to show that  $w$  and  $w'$  are independent conditioned on  $y = c$  for every  $c$ . Using our characterization based on self-composition, it amounts to proving the following (universally quantified) pRHL judgment:

$$\models s^{(2)} \sim s^{(2)} : \text{EqMem}^{(2),(2)} \implies \phi\langle 1 \rangle \iff \psi\langle 2 \rangle$$

$$\text{where } \begin{cases} \phi \triangleq w^1 = a \wedge w'^1 = b \wedge y^1 = c \wedge y^2 = c \\ \psi \triangleq w^1 = a \wedge w'^2 = b \wedge y^1 = c \wedge y^2 = c. \end{cases}$$



The proof proceeds by moving the samplings of  $z^1$  and  $z^2$  in both programs to the front of the program, and then swapping samplings in the left program (we can use the rule [SWAP] to reorder the instructions, as the sampling instructions for  $z^1$  and  $z^2$  operate on different variables). Then, we couple  $z^1\langle 1 \rangle$  to be equal to  $z^2\langle 2 \rangle$ , and  $z^2\langle 1 \rangle$  to be equal to  $z^1\langle 2 \rangle$ . We apply the identity coupling to all other random samplings.

## 7 Formalization

EasyCrypt [3, 4] is an interactive proof assistant that supports reasoning about (relational) properties of probabilistic programs, using the pRHL logic. We have applied EasyCrypt to the main examples of this paper. For uniformity, it suffices to establish the required pRHL judgment; in contrast, independence via self-composition requires to build the self-composed program, which we have done manually. The main challenges for the verification are:

1. Restructuring the code of the program to make the rules of the logic applicable; this is done by applying the equivalent of the [STRUCT] rules.
2. Discovering and establishing the correct proof invariants. The current version of EasyCrypt requires that invariants are produced by the users.
3. Building an appropriate coupling, primarily through applying the rule for random samplings with carefully chosen bijections.

Our examples are formalized in about 1,000 lines of proof script in EasyCrypt.<sup>2</sup> The most complex example, and the one where the three challenges are most pronounced, is the random walk over a cycle. This example is formalized in about 500 lines of EasyCrypt code, out of which the statement, including the definition of the program, takes about 50 lines. The remaining 90% of the formalization covers the notions used in the proof and the proof itself.

## 8 Conclusion

We have proposed a new method based on probabilistic couplings for formally verifying uniformity and independence properties of probabilistic programs. Our method complements the existing range of techniques for probabilistic reasoning, and has many potential applications in program verification, security, and privacy.

## 9 Acknowledgments

We thank the anonymous reviewers for their detailed comments. This work was partially supported by NSF grants TC-1065060 and TWC-1513694, by the European Union’s H2020 Programme under grant agreement number ICT-644209, and a grant from the Simons Foundation (#360368 to Justin Hsu).

## References

- [1] G. Barthe, B. Grégoire, and S. Zanella-Béguelin. Formal certification of code-based cryptographic proofs. In *ACM SIGPLAN–SIGACT Symposium on Principles of Pro-*

---

<sup>2</sup>Proofs are available at the following link: <https://gitlab.com/easycrypt/indep>

- programming Languages (POPL)*, Savannah, Georgia, pages 90–101, New York, 2009. URL <http://certicrypt.gforge.inria.fr/2013.Journal.pdf>.
- [2] G. Barthe, P. R. D’Argenio, and T. Rezk. Secure information flow by self-composition. *Mathematical Structures in Computer Science*, 21(06):1207–1252, 2011.
- [3] G. Barthe, B. Grégoire, S. Héraud, and S. Zanella-Béguelin. Computer-aided security proofs for the working cryptographer. In *IACR International Cryptology Conference (CRYPTO)*, Santa Barbara, California, volume 6841 of *Lecture Notes in Computer Science*, pages 71–90. Springer-Verlag, 2011. URL [http://dx.doi.org/10.1007/978-3-642-22792-9\\_5](http://dx.doi.org/10.1007/978-3-642-22792-9_5).
- [4] G. Barthe, F. Dupressoir, B. Grégoire, C. Kunz, B. Schmidt, and P.-Y. Strub. EasyCrypt: A tutorial. In A. Aldini, J. Lopez, and F. Martinelli, editors, *Foundations of Security Analysis and Design VII*, volume 8604 of *Lecture Notes in Computer Science*, pages 146–166. Springer-Verlag, 2014. ISBN 978-3-319-10081-4. doi: 10.1007/978-3-319-10082-1\_6. URL [http://dx.doi.org/10.1007/978-3-319-10082-1\\_6](http://dx.doi.org/10.1007/978-3-319-10082-1_6).
- [5] G. Barthe, T. Espitau, B. Grégoire, J. Hsu, L. Stefanescu, and P.-Y. Strub. Relational reasoning via probabilistic coupling. In *International Conference on Logic for Programming, Artificial Intelligence and Reasoning (LPAR)*, Suva, Fiji, volume 9450 of *Lecture Notes in Computer Science*, pages 387–401. Springer-Verlag, 2015. URL <http://arxiv.org/abs/1509.03476>.
- [6] G. Barthe, M. Gaboardi, B. Grégoire, J. Hsu, and P. Strub. Proving differential privacy via probabilistic couplings. In *IEEE Symposium on Logic in Computer Science (LICS)*, New York, New York, pages 749–758, 2016. URL <http://arxiv.org/abs/1601.05047>.
- [7] G. Barthe, B. Grégoire, J. Hsu, and P.-Y. Strub. Coupling proofs are probabilistic product programs. In *ACM SIGPLAN–SIGACT Symposium on Principles of Programming Languages (POPL)*, Paris, France, 2017. URL <http://arxiv.org/abs/1607.03455>.
- [8] A. Z. Broder. Generating random spanning trees. In *ACM SIGACT Symposium on Theory of Computing (STOC)*, Seattle, Washington, pages 442–447, 1989. doi: 10.1109/SFCS.1989.63516. URL <http://dx.doi.org/10.1109/SFCS.1989.63516>.
- [9] R. Chadha, L. Cruz-Filipe, P. Mateus, and A. Sernadas. Reasoning about probabilistic sequential programs. *Theoretical Computer Science*, 379(1–2):142–165, 2007.
- [10] K. Chatterjee, H. Fu, P. Novotný, and R. Hasheminezhad. Algorithmic analysis of qualitative and quantitative termination problems for affine probabilistic programs. In *ACM SIGPLAN–SIGACT Symposium on Principles of Programming Languages (POPL)*, Saint Petersburg, Florida, 2016. URL <http://arxiv.org/abs/1510.08517>.
- [11] K. Chatterjee, P. Novotný, and D. Zikelic. Stochastic invariants for probabilistic termination. In *ACM SIGPLAN–SIGACT Symposium on Principles of Programming Languages (POPL)*, Paris, France, 2017. URL <http://arxiv.org/abs/1611.01063>.
- [12] Á. Darvas, R. Hähnle, and D. Sands. A theorem proving approach to analysis of secure information flow. In *Security in Pervasive Computing (SPC)*, Boppard, Germany, volume 3450 of *Lecture Notes in Computer Science*, pages 193–209. Springer-Verlag, 2005.
- [13] J. den Hartog. *Probabilistic extensions of semantical models*. PhD thesis, Vrije Universiteit Amsterdam, 2002.

- [14] L. M. Ferrer Fioriti and H. Hermanns. Probabilistic termination: Soundness, completeness, and compositionality. In *ACM SIGPLAN–SIGACT Symposium on Principles of Programming Languages (POPL), Mumbai, India*, pages 489–501, 2015.
- [15] D. Kozen. A probabilistic PDL. *J. Comput. Syst. Sci.*, 30(2):162–178, 1985.
- [16] T. Lindvall. *Lectures on the coupling method*. Courier Corporation, 2002.
- [17] C. Morgan, A. McIver, and K. Seidel. Probabilistic predicate transformers. *ACM Transactions on Programming Languages and Systems*, 18(3):325–353, 1996.
- [18] J. Pearl and A. Paz. Graphoids: Graph-based logic for reasoning about relevance relations. Technical report, University of California, Los Angeles, 1985. URL [http://ftp.cs.ucla.edu/pub/stat\\_ser/r53-L.pdf](http://ftp.cs.ucla.edu/pub/stat_ser/r53-L.pdf).
- [19] L. H. Ramshaw. *Formalizing the Analysis of Algorithms*. PhD thesis, Computer Science, 1979.
- [20] R. Rand and S. Zdancewic. VPHL: A verified partial-correctness logic for probabilistic programs. In *Conference on the Mathematical Foundations of Programming Semantics (MFPS), Nijmegen, The Netherlands*, 2015.
- [21] H. Thorisson. *Coupling, Stationarity, and Regeneration*. Springer-Verlag, 2000.
- [22] C. Villani. *Optimal transport: old and new*. Springer-Verlag, 2008.