Bayesian Separation Logic

A Logical Foundation and Axiomatic Semantics for Probabilistic Programming

SHING HIN HO, Imperial College London, UK
NICOLAS WU, Imperial College London, UK
AZALEA RAAD, Imperial College London, UK

Bayesian probabilistic programming languages (BPPLs) let users denote statistical models as code while the interpreter infers the posterior distribution. The semantics of BPPLs are usually mathematically complex and unable to reason about desirable properties such as expected values and independence of random variables. To reason about these properties in a non-Bayesian setting, probabilistic separation logics such as PSL and Lilac interpret separating conjunction as probabilistic independence of random variables. However, no existing separation logic can handle Bayesian updating, which is the key distinguishing feature of BPPLs.

To close this gap, we introduce Bayesian separation logic (BaSL), a probabilistic separation logic that gives semantics to BPPL. We prove an internal version of Bayes' theorem using a result in measure theory known as the Rokhlin-Simmons disintegration theorem. Consequently, BaSL can model probabilistic programming concepts such as Bayesian updating, unnormalised distribution, conditional distribution, soft constraint, conjugate prior and improper prior while maintaining modularity via the frame rule. The model of BaSL is based on a novel instantiation of Kripke resource monoid via σ-finite measure spaces over the Hilbert cube, and the semantics of Hoare triple is compatible with an existing denotational semantics of BPPL based on the category of *s*-finite kernels. Using BaSL, we then prove properties of statistical models such as the expected value of Bayesian coin flip, correlation of random variables in the collider Bayesian network, and the posterior distributions of the burglar alarm model, a parameter estimation algorithm, and the Gaussian mixture model.

1 INTRODUCTION

Statistical techniques have become increasingly prevalent with widespread applications across a multitude of domains from computer vision and data science to computational biology and social science. As a result, there is a pressing need for developing secure and explainable statistical models. A way of ensuring correctness of such systems is with *formal methods*, a collection of techniques that allows computer scientists to prove properties of algorithms/programming languages by modelling the objects of interest mathematically. In fact, researchers are increasingly applying formal methods to probabilistic/machine learning algorithms in order to provide correctness guarantees on desirable statistical properties [Cruz and Shoukry 2022; Slusarz et al. 2022]. We are interested in formal and compositional reasoning techniques for *Bayesian probabilistic programming languages* – a class of statistical programming languages that implements algorithms for *Bayesian inference*.

Bayesian inference is a statistical technique that allows users to infer unknown distributions using a statistical model and Bayes' theorem. It has a wide range of applications ranging from medicine [Muehlemann et al. 2023] to computer vision [Geman and Geman 1984]. For instance, statistical problems such as clustering and regression can be solved by reframing them as Bayesian inference problems. Computationally, *probabilistic programming* is the programming paradigm that implements Bayesian inference by using general inference algorithms such as Hamiltonian Monte Carlo and Gibbs sampling [Brooks et al. 2011], while providing a simple interface for users to specify their statistical models [van de Meent et al. 2021]. *Bayesian probabilistic programming languages* (BPPLs) such as Stan [Carpenter et al. 2017], Anglican [Tolpin et al. 2016] and Gen [Cusumano-Towner et al. 2019] are implementations of the programming paradigm and have specialised constructs for users to easily express statistical models.

Authors' addresses: Shing Hin Ho, Imperial College London, UK, shinghin.ho21@imperial.ac.uk; Nicolas Wu, Imperial College London, UK, n.wu@imperial.ac.uk; Azalea Raad, Imperial College London, UK, azalea.raad@imperial.ac.uk.

Goal. Our goal is to develop a logical foundation that allows us to reason about statistical properties of BPPL programs (e.g. independence, expected value, correlation) in a *compositional* manner. To achieve this, we use *separation logic* [Reynolds 2002], a logical system designed to allow compositional reasoning of computational resources, which in our case, are the random variables produced by the probabilistic program. In particular, we develop *Bayesian separation logic* (BaSL, pronounced 'basil'), a logical system for Bayesian reasoning based on probabilistic separation logic. From BaSL, we derive the first Hoare logic for BPPLs, which then allows us to prove the correctness/properties of statistical models such as the *Bayesian coin flip* model, the *collider network*, a *parameter estimation algorithm* and a *Gaussian-mixture-based clustering model.*

Bayesian Conditioning. Unlike *randomised* programming languages, which are languages with a sampling construct, **sample**, implemented via a pseudo-random number generator (e.g. in C and Python), a BPPL has an additional *conditioning* construct, **observe**[1], that allows users to express *conditional probability*. For example, consider an experiment where we toss a fair die X, then condition on the event $X > 4$. This can be expressed by the BPPL program in Fig. 1, which computes the distribution of X conditioning on the event $X > 4$:

```
let X = sample(FairDie) in
observe(X > 4);
return X
```

'**observe**' filters out traces that do not satisfy $X > 4$ →

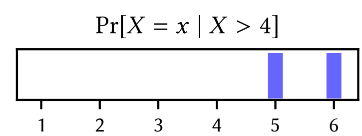


Fig. 1. Conditioning as the computational effect **observe**

Probabilistic Separation Logics. The reasoning principles for *randomised* languages have been studied in *probabilistic separation logics* [Bao et al. 2021a,b; Barthe et al. 2019; Li et al. 2023; Tassarotti and Harper 2019] through a set of axiomatic rules for deriving *Hoare triples*. A Hoare triple, $\{P\}\ M\ \{X.Q\}$, consists of four components: the program M, the precondition P, the return variable X and the postcondition Q. A Hoare triple $\{P\}\ M\ \{X.Q\}$ is *valid* when executing the program M assuming P produces a variable X and the proposition Q holds. For instance, consider the following Hoare triple (using the notation of Lilac [Li et al. 2023]):

$$\{\top\}\ \mathbf{sample}(\mathsf{FairDie})\ \{X.X \sim \mathsf{Unif}(\{1, \ldots, 6\})\}$$

The triple above describes the pre/postcondition for executing **sample**(FairDie) for rolling a fair die. Assuming a precondition with no knowledge of variables ($\top$), executing **sample**(FairDie) yields a random variable X that has a uniform distribution on $\{1, \ldots, 6\}$. Moreover, since the underlying logic of the Hoare triple is based on probabilistic *separation* logic, we can freely add independent random variables in our pre/postconditions via the *probabilistic separating conjunction* '$*$'. For example, the Hoare triple below states that given a random variable Y with distribution $\mathbb{P}$, executing **sample**(FairDie) produces an X that is *probabilistically independent* of Y:

$$\{Y \sim \mathbb{P}\}\ \mathbf{sample}(\mathsf{FairDie})\ \{X.X \sim \mathsf{Unif}(\{1, \ldots, 6\}) * Y \sim \mathbb{P}\}.$$

This is useful due to the nature of probabilistic reasoning, where a key part is to determine which random variables are independent of each other. Apart from the property of *distribution* (propositions of the form $X \sim \mathbb{P}$), probabilistic separation logic can also reason about different probabilistic properties. For example, the logical entailment

$$X \sim \mathcal{N}(0, 1) * Y \sim \mathcal{N}(0, 1) \vdash (\mathbb{E}[X] = 0 * \mathbb{E}[Y] = 0) \wedge \mathrm{Cov}[X, Y] = 0$$

[1]For readers familiar with probabilistic programming, **observe** is implemented using the soft-constraint construct **score**.

states that if X and Y are independent, normally-distributed random variables with means 0 and standard deviation 1, then they both have *expected value*[2] zero ($\mathbb{E}[X] = 0 * \mathbb{E}[Y] = 0$) and X, Y are uncorrelated (since the covariance is zero $\mathrm{Cov}[X, Y] = 0$). Having statistical propositions and combining them using the separating conjunction '$*$' allows convenient statistical reasoning in a formal setting.

Limitation of Existing Work: No Support for Bayesian Conditioning. While probabilistic separation logics are a rich field of study with plenty of variations, existing probabilistic separation logics cannot reason about the **observe** construct. For instance, following the intuition in Fig. 1, we *should* ideally have a Hoare specification as follows:

$$\{X \sim \mathsf{Unif}(\{1, ..., 6\})\}\ \mathbf{observe}(X > 4)\ \{X \sim \mathsf{Unif}(\{5, 6\})\}$$

This technique is called *Bayesian updating/Bayesian conditioning*, where we 'update' our distributions based on observations. BaSL is the *first* probabilistic separation logic that allows reasoning of Bayesian updating, and we do this by drawing an analogy between mutation of memory in standard separation logic. We achieve this by proving an *internal Bayes' theorem* (Theorem 4.11) using a result in measure theory known as the *Rokhlin-Simmons disintegration theorem.*

Limitation of Existing Work: Dependencies of Random Variables. Existing probabilistic separation logics such as Lilac and BlueBell can reason about dependent random variable via the *conditioning modality*. For instance, the proposition $x \leftarrow X \mid Y \sim \mathcal{N}(x, 1)$ means Y has distribution $\mathcal{N}(x, 1)$, conditioning on $X = x$ for some $x \in \mathbb{R}$. BaSL extends the conditioning modality to support conditional reasoning in the presence of Bayesian updating. Furthermore, the Hoare logic in BaSL supports *conditional sampling*. For example, the following Hoare triple is provable in BaSL:

$$\{X \sim \mathsf{Unif}(0, 1)\}\ \mathbf{sample}(\mathsf{Normal}(X, 1))\ \{Y.x \xleftarrow{\mathsf{Unif}(0,1)} X \mid Y \sim \mathcal{N}(x, 1)\}$$

It states that assuming $X \sim \mathsf{Unif}(0, 1)$, sampling a normal with mean X yields a random variable Y, which has a *conditional distribution* $Y \sim \mathcal{N}(x, 1)$ when $X = x$ for some $x \in [0, 1]$.

Proving Properties of the Bayesian Coin Flip Model. By supporting the above features, together with the compositionality afforded by the frame rule, BaSL can serve as a logical foundation for symbolic reasoning probabilistic programming. To demonstrate BaSL and its associated Hoare logic, we first consider a simple statistical problem known as *Bayesian coin flip.*

Problem 1. *We have a coin and we want to know if it is fair. We toss it once, it comes up heads. We toss it again, it comes up tails. Is it a fair coin?*

The problem, while simple, encapsulates the three main steps of the Bayesian method:

(1) We *assume* a *prior belief* (or *prior* in short) regarding the problem. In our case, we assume a prior regarding whether the coin is fair.
(2) We *observe* real-life data. In our case, we have two observations of coin flip.
(3) We *update* our belief based on observations. In our case, we update our belief based on *observations* using *Bayes' theorem.*

We defer the explanation of the statistical model to §2. For now, we describe the problem using statistical notation in Fig. 2, and its programmatic counterpart in Fig. 3. Fig. 2 states that X has (prior) distribution $\mathsf{Unif}(0, 1)$, $Flip_i$ has distribution $\mathsf{Bern}(x)$ when $X = x$, which is a distribution that returns 1 with x probability and 0 with $1 - x$ probability, and we have two observations. Using a BPPL, we can express the same model programmatically as shown in the BayesCoin program in Fig. 3. Programmatically, the underlying interpreter for the language is performing *rejection*

[2] The expected value $\mathbb{E}[X]$ of a random variable X is the average of the distribution of X.

$X \sim \mathsf{Unif}(0,1)$
$Flip_1 \sim \mathsf{Bern}(x)$ when $X = x$
$Flip_2 \sim \mathsf{Bern}(x)$ when $X = x$
observation: $Flip_1 = 1$
observation: $Flip_2 = 0$

Fig. 2. Bayesian solution of Problem 1

Precondition: $\top$

Postcondition: Return value X has expected value 1/2

```
let X = sample(Unif(0, 1)) in    // specify prior belief
let Flip_1 = sample(Bern(X)) in  // toss the coin
observe(Flip_1 = 1);             // assert it comes up heads
let Flip_2 = sample(Bern(X)) in  // toss again
observe(Flip_2 = 0);             // assert it comes up tails
return X                         // what is our belief now?
```

Fig. 3. BayesCoin

sampling, which, to a first approximation, executes the program many times and filters out traces that do *not* satisfy the required observations, i.e. when $Flip_1 \neq 1$ or $Flip_2 \neq 0$. Intuitively, since we observed a head and a tail, the expected value (average) of X should remain 1/2. As we show in §2, using BaSL we can encode and prove this specification easily (Fig. 4), and our proof derivation *formally* describes how the distributions of the random variables evolve.

Contributions and Roadmap. Our contributions are as follows:

- We explain how BayesCoin can be specified in BaSL (§2).
- We prove properties of five standard, but non-trivial Bayesian statistical models such as Bayesian networks (§3.3, §3.4), parameter estimation algorithm (§3.2), and the Gaussian mixture model (§3.6). The models have non-trivial features such as conditional dependencies, soft constraint, conjugate prior and improper prior (§3.5), which no existing probabilistic separation logic can verify (§3).
- For the semantics of BaSL, we describe intuitively the existing resource-theoretic model of Lilac for modelling randomness (§4.1) and motivate the need for a measure-theoretic semantics to model randomness in a Bayesian setting (§4.2).
- We develop a novel *Kripke resource monoid* for modelling randomness that supports *Bayesian updating* by using σ-finite measure spaces over the Hilbert cube (§4.3).
- We show that the Kripke resource model [Galmiche et al. 2005] of BaSL is compatible with *partially affine separation logic* [Charguéraud 2020] (§4.3).
- We generalise Lilac's *disintegration modality* so that conditional reasoning can be performed in the presence of Bayesian updating (§4.4).
- We design new logical propositions to encode the concept of *likelihood* and *normalising constants*, two key concepts in Bayesian probability (§4.4).
- We prove an internal version of Bayes' theorem and show that the concept can be encoded as logical propositions in BaSL by combining the disintegration modality and the likelihood proposition (Theorem 4.11).
- We develop the *BaSL proof system* (a set of Hoare triples) and prove that it is *sound* with respect to our Kripke resource model of BaSL (§4.4).

Finally, we discuss related and future work and conclude (§5).

2 OVERVIEW

To give an intuition of how BaSL can be used to prove properties of statistical models, we demonstrate its proof rules via BayesCoin (Problem 1). Recall that $\mathsf{Bern}(0.5)$ represents a fair coin as it returns 1 and 0 with equal probability, while $\mathsf{Bern}(0.9)$ represents a biased coin that comes up heads

1 $\{\top_1\}$
2 $\quad$ **let** X = **sample**(Unif(0, 1)) **in**
3 $\{X \sim \mathsf{Unif}(0,1)\}$
4 $\quad$ **let** $Flip_1$ = **sample**(Bern(X)) **in**
5 $\{x \xleftarrow{\mathsf{Unif}(0,1)} X \mid Flip_1 \sim \mathsf{Bern}(x)\}$
6 $\quad$ **observe**($Flip_1 = 1$);
7 $\{x \xleftarrow{\mathsf{Unif}(0,1)} X \mid Flip_1 \sim \ell_{=1} \cdot \mathsf{Bern}(x)\}$
8 $\{x \xleftarrow{\mathsf{Unif}(0,1)} X \mid \mathsf{L}(x)\}$ // likelihood of $X = x$ is x
9 $\{X \sim \mathsf{Beta}(2,1) * \mathsf{NormConst}\}$
10 $\quad\{X \sim \mathsf{Beta}(2,1)\}$
11 $\quad\quad$ **let** $Flip_2$ = **sample**(Bern(X)) **in**
12 $\quad\{x \xleftarrow{\mathsf{Beta}(2,1)} X \mid Flip_2 \sim \mathsf{Bern}(x)\}$
13 $\quad\quad$ **observe**($Flip_2 = 0$);
14 $\quad\{x \xleftarrow{\mathsf{Beta}(2,1)} X \mid Flip_2 \sim \ell_{=0} \cdot \mathsf{Bern}(x)\}$
15 $\quad\{X \sim \mathsf{Beta}(2,2) * \mathsf{NormConst}\}$
(lines 10–15: H-Frame)
16 $\{X \sim \mathsf{Beta}(2,2) * \mathsf{NormConst} * \mathsf{NormConst}\}$
17 $\{X \sim \mathsf{Beta}(2,2) * \mathsf{NormConst}\}$
18 $\quad$ **return** X
19 $\{X \sim \mathsf{Beta}(2,2) * \mathsf{NormConst}\}$
20 $\{\mathbb{E}[X] = 1/2 * \mathsf{NormConst}\}$

Fig. 4. An axiomatic description of BayesCoin

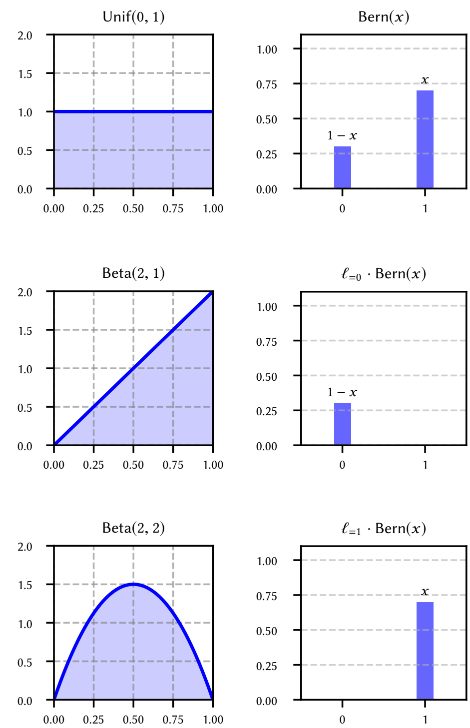


Fig. 5. Visualisation of distributions

90% of the time. Hence, we model our belief about whether our coin is fair via a random variable X that takes a value in $[0, 1]$.

Since we do not have additional information about the coin (X), we assume $X \sim \mathsf{Unif}(0,1)$ as our prior distribution and write the program described in Fig. 3. To semantically deduce that the return value X has expected value $1/2$ (see the postcondition of Fig. 3), we use BaSL to describe BayesCoin axiomatically via pre/postcondition style reasoning rules. We present the BaSL proof sketch of BayesCoin in Fig. 4. We proceed with a detailed but informal explanation of our proof.

Lines 1-3. Initially, we have no random variables, as captured by the *trivial* precondition $\top_1$, where 1 in $\top_1$ is the current *normalising constant* (we explain this on Page 7). After executing line 2 (of Fig. 4), the interpreter produces a random variable X with the desired distribution. To reason about this, we apply the H-Sample rule below at line 2, which states that sampling from a distribution $\mathbb{P}$ of type $\mathbb{R}$ returns a random variable X of type $\mathbb{R}$ distributed according to $\mathbb{P}$ (for line 2, $\mathbb{P}$ is instantiated to $\mathsf{Unif}(0,1)$):

$$\vdash \{\top_1\}\ \mathbf{sample}(\mathbb{P})\ \{X : \mathbb{R}.\ X \sim \mathbb{P}\} \quad \text{H-Sample}$$

To 'chain' the postcondition to the rest of the program, we apply the sequencing rule H-Let below, chaining the postcondition of the first program to the precondition of the second program, thus obtaining $X \sim \mathsf{Unif}(0,1)$ on line 3.

$$\frac{\vdash \{P\}\ M\ \{X : \mathcal{A}.Q\} \qquad X : \mathcal{A} \vdash \{Q\}\ N\ \{Y : \mathcal{B}.R\}}{\vdash \{P\}\ \mathbf{let}\ X = M\ \mathbf{in}\ N\ \{Y : \mathcal{B}.R\}}\ \text{H-Let}$$

Lines 4-5. At line 4 we sample a Bernoulli variable according to our sampled X. While this line certainly typechecks, it is more challenging to verify: it is ambiguous statistically, as it semantically describes a *conditional distribution*. Given $X \sim \mathsf{Unif}(0, 1)$, we write $Flip_1 | X = x \sim \mathsf{Bern}(x)$ to mean that conditioning on $X = x$ for almost all $x \in [0, 1]$, the random variable $Flip_1$ has distribution $\mathsf{Bern}(x)$. To reason about this in BaSL, we introduce our conditional sampling axiom:

$$\vdash \{X \sim \mathbb{P}\}\ \mathbf{sample}(p(X))\ \{Y : \mathbb{R}.x \overset{\mathbb{P}}{\leftarrow} X \mid Y \sim p(x)\}\ \text{H-CondSample}$$

The arrow/bar notation $x \overset{\mathbb{P}}{\leftarrow} X \mid P$ is our novel *conditioning modality*, which assumes X has distribution $\mathbb{P}$ and binds it to a deterministic name x, while the proposition P on the right is a proposition on the conditioned space after conditioning $X = x$. Specifically, $x \overset{\mathbb{P}}{\leftarrow} X \mid Y \sim p(x)$ in the postcondition is read as follows: assuming $X \sim \mathbb{P}$, if we condition on $X = x$ for some deterministic x, then the sampled Y has distribution $p(x)$. Instantiating the axiom above, line 5 says $Flip_1$ has distribution $\mathsf{Bern}(x)$ whenever $X = x$.

Lines 6-8. Line 6 asserts $Flip_1 = 1$ and rejects all program traces where $Flip_1 = 0$. Semantically, this means we apply a *likelihood function* $\ell_{=1} : \{0, 1\} \to [0, \infty)$ to $\mathsf{Bern}(x)$, where $\ell_{=1}(1) := 100\%$ and $\ell_{=1}(0) := 0\%$. Specifically, since $Flip_1 \sim \mathsf{Bern}(x)$ (as depicted on the left side of Fig. 6), applying the likelihood function $\ell_{=1}$ to $\mathsf{Bern}(x)$ *updates* the distribution to the one on the right side of Fig. 6. Notice that this operation is done in the conditioned space conditioning on $X = x$. In light of this, we introduce the H-CondObserve axiom below for *conditional observation*:

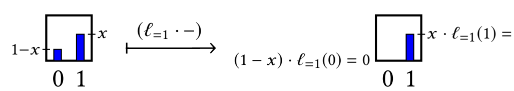


Fig. 6. Applying likelihood function

$$\vdash \{x \overset{\pi}{\leftarrow} X \mid Y \sim p(x)\}\ \mathbf{observe}(P(Y))\ \{x \overset{\pi}{\leftarrow} X \mid Y \sim \ell_P \cdot p(x)\}\ \text{H-CondObserve}$$

where P is a boolean predicate on Y. The axiom states that assuming $Y \sim p(x)$ when conditioning on $X = x$, where $X \sim \pi$ in the unconditioned space, then observing $P(Y)$ updates the distribution from $p(x)$ to $\ell_P \cdot p(x)$, where ℓ_P is the likelihood function with $\ell_P(x) := 1$ if $P(x)$ holds, and 0 otherwise. In our case, $Y = Flip_1$ and $P(Flip_1)$ is defined to be $(Flip_1 = 1)$, which yields the postcondition on line 7. For readers familiar with probabilistic programming, **observe** is implemented using the *soft constraint* construct **score**, and there is a more general rule H-CondScore as we explain in §4.

Since $\mathbf{observe}(Flip_1 = 1)$ multiplies the likelihood of $Flip_1 = 1$ by 100%, the likelihood of $Flip_1 = 1$ is $x \cdot 100\% = x$ when $X = x$. Similarly, $\mathbf{observe}(Flip_1 = 1)$ multiplies the likelihood of $Flip_1 = 0$ by 0%; i.e. $Flip_1 = 0$ has likelihood $(1 - x) \cdot 0\% = 0\%$ when $X = x$. In other words, this describes traces where $Flip_1 \sim \mathsf{Bern}(x)$ has likelihood x. Intuitively this makes sense: when $X = 0.1$, the likelihood of $Flip_1 = 1$ is 0.1, which is lower than the likelihood when $X = 0.99$. To reason about this, we introduce a *likelihood proposition* $\mathsf{L}(x)$, which denotes that the likelihood of our current state is x. Since $Flip_1 \sim \mathsf{Bern}(x)$ has likelihood x, the entailment $Flip_1 \sim \ell_{=1} \cdot \mathsf{Bern}(x) \vdash \mathsf{L}(x)$ holds, which results in the postcondition on line 8 using the standard rule of consequence H-Cons (see Fig. 9).

Line 9. In light of our observation regarding how likely $X = x$ is, we can now *update* our belief. To achieve this, we apply an internal, logical version of *Bayes' theorem*. Recall that Bayes' theorem, in the context of Bayesian statistics, states the following:

$$\text{posterior} \propto \text{prior} \cdot \text{likelihood}.$$

Or, equivalently, suppose $1/Z$ is the *normalising constant* for some $Z > 0$, then Bayes' theorem can be stated as unnormalised posterior $\cdot\, Z$ = prior $\cdot$ likelihood. Notice that line 8 has a similar structure to above – we have the prior $X \sim \mathsf{Unif}(0, 1)$ and the likelihood $\mathsf{L}(x)$. The question is:

can we find a suitable proposition that represents the (unnormalised) posterior? That is, finding a suitable proposition $?P$ such that the following entailment holds:

$$\overbrace{x \xleftarrow{\mathsf{Unif}(0,1)} X \mid}^{\text{prior}} \ \overbrace{\mathsf{L}(x)}^{\text{likelihood}} \ \vdash \ \overbrace{?P}^{\text{posterior}} \quad .$$

Before explaining what the proposition looks like, let us first intuitively visualise how our belief regarding the coin has changed.

After observing the coin landing on heads, we shift towards believing X being more likely to be larger. For instance, it is less likely for X to be close to zero, as this would make landing on heads less likely. Without doing the calculations (which can be found in a standard text on Bayesian methods, e.g. by McElreath [2020, §2.2]), the original distribution $\mathsf{Unif}(0, 1)$ (left of Fig. 7) is updated to $1/2 \cdot \mathsf{Beta}(2, 1)$ (right of Fig. 7). There is, however, a caveat: since we applied Bayes' theorem, the distribution is *unnormalised*: the area under the triangle of the graph is $1/2 \neq 1$, as denoted by $1/2$ in $1/2 \cdot \mathsf{Beta}(2, 1)$. To remedy this, we use the separating conjunction '$*$' to factor out the normalising constant. Specifically, for all $Z > 0$ (e.g. $Z = 1/2$ here), the entailment $X \sim Z \cdot \mathsf{Beta}(2, 1) \vdash X \sim \mathsf{Beta}(2, 1) * \mathsf{NormConst}$ holds. Intuitively, NormConst factors out and hides the normalising factor and asserts the existence of a non-zero normalising constant, i.e. $\mathsf{NormConst} \coloneqq \exists k : (0, \infty).\ \mathsf{L}(k)$. This answers our question: the 'posterior proposition' $?P$ is of the form:

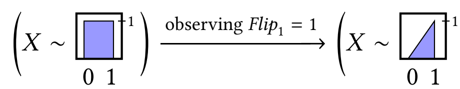


Fig. 7. Bayesian updating of $\mathsf{Unif}(0, 1)$

$$\overbrace{x \xleftarrow{\mathsf{Unif}(0,1)} X \mid}^{\text{prior}} \ \overbrace{\mathsf{L}(x)}^{\text{likelihood}} \ \vdash \ \overbrace{X \sim \mathsf{Beta}(2, 1) * \mathsf{NormConst}}^{\text{posterior}}.$$

As we will see in §4 (Theorem 4.11), the entailment is sound in BaSL via properties of *disintegration* (Lemma 4.10). Next, by applying the rule of consequence, we obtain the postcondition on line 9.

Lines 10-16. Note that lines 10-15 are similar to line 3-7: we perform another coin toss, observe it comes up tails, and return our updated belief. Intuitively, after observing the coin landing on tails, we 'shift back' our belief about X being likely to take on higher values – X is more likely to take on values in the middle (e.g. $0.4 \leq X \leq 0.6$ has higher probability than $X \geq 0.8$ or $X \leq 0.2$), i.e. we update the distribution to the one on the right of Fig. 8. This distribution is known as the $\mathsf{Beta}(2, 2)$ distribution. To prove this in BaSL, we first assume $X \sim \mathsf{Beta}(2, 1)$ as our precondition of line 10, repeat the steps above and obtain the postcondition on line 15.

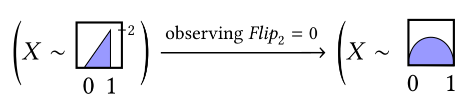


Fig. 8. Bayesian updating of $\mathsf{Beta}(2, 1)$

We now enter another key step of our proof: using the *frame rule* of separation logic:

$$\frac{\vdash \{P\}\ M\ \{X.Q\}}{\vdash \{P * F\}\ M\ \{X.Q * F\}}\ (X \notin \mathsf{fv}(F))\ \textsc{H-Frame}$$

The H-Frame rule allows us to *frame off* (factor out) propositions that are probabilistically independent of our current resources. It states that in order to prove $\{P * F\}\ M\ \{X.Q * F\}$, where F is a proposition probabilistically independent of P and Q, it suffices to prove $\{P\}\ M\ \{X.Q\}$.

As the normalising constant proposition NormConst is 'separated' from $X \sim \mathsf{Beta}(2, 1)$ on line 9, we can 'frame off' the normalising constant NormConst prior to line 10 and frame it back on after line 15 and obtain the postcondition on line 16.

Lines 17-20. At line 16 we have two normalising constants NormConst and NormConst, each created from an update of X. We now combine them: intuitively, if two independent unnormalised random variables have normalising constants Z and Z' respectively, the overall distribution has normalising constant $Z \cdot Z'$. This justifies the entailment NormConst $*$ NormConst $\vdash$ NormConst. This means we can obtain the postcondition on line 17. Returning X yields the same postcondition (see H-Ret in Fig. 9), which leads to our desired postcondition on line 19. Since from line 19 we know $X \sim \text{Beta}(2,2)$ and a Beta(2, 2)-distributed random variable has expected value (average) 1/2, the entailment $X \sim \text{Beta}(2,2) * \text{NormConst} \vdash \mathbb{E}[X] = 1/2 * \text{NormConst}$ holds, and we obtain the postcondition on line 20 using the standard rule of consequence H-Cons (see Fig. 9).

3 VERIFYING STATISTICAL MODELS WITH BASL

We demonstrate the expressivity and verification capability of BaSL by verifying five programs described below, each with distinct features. To this end, in §3.1 we first present the BaSL programming language, Bppl, a standard Bayesian probabilistic programming language, and then present the BaSL proof system as a set of Hoare triples (most of which we have described in §2).

§	Statistical model	Distinct feature(s)
§3.2	Parameter estimation algorithm	Soft constraint; conjugate priors
§3.3	The burglar alarm Bayesian network	Joint conditioning; Bayes' theorem
§3.4	The common effect Bayesian network	Conditional dependence and correlation
§3.5	The semantic Lebesgue measure	Improper prior; handling σ-finite measures
§3.6	Gaussian-mixture-based clustering	Intractable posterior; bounded loops

3.1 BaSL Programming Language and Proof System

The BaSL programming language, Bppl, is a typed, first-order language equipped with two effects: the *probabilistic sampling* effect **sample**($\mathbb{P}$), which samples from a distribution $\mathbb{P}$, and the *soft conditioning* effect **score**(ℓ), which *scales* the current distribution according to a non-negative number ℓ. In particular, the Bppl *terms and types* are defined by the following grammar, where X ranges over a countably infinite set of names, $n \in \mathbb{N}$, $r \in \mathbb{R}$, and f ranges over (measurable) functions.

$$\begin{aligned} \text{Term} \ni M ::= {} & () \mid X \mid n \mid r \mid f(M) \mid (M, M) \mid M.1 \mid M.2 \mid \mathbf{true} \mid \mathbf{false} \mid \mathbf{if}\ M\ \mathbf{then}\ M\ \mathbf{else}\ M \\ & \mid \mathbf{sample}(M) \mid \mathbf{score}(M) \mid \mathbf{return}(M) \mid \mathbf{let}\ X = M\ \mathbf{in}\ M \\ \text{Type} \ni \tau ::= {} & 1 \mid \mathbb{N} \mid \mathbb{R} \mid \mathbb{B} \mid \tau \times \tau \mid \mathbb{P}(\tau) \end{aligned}$$

Additional Encodings. We encode the following syntactic shorthands:

$$M; N := \mathbf{let}\ \_ = M\ \mathbf{in}\ N \qquad \mathbf{observe}(M) := \mathbf{score}(\mathbf{if}\ M\ \mathbf{then}\ 1\ \mathbf{else}\ 0)$$
$$\mathbf{observe}\ M\ \mathbf{from}\ \mathbb{P} := \mathbf{score}(\text{density}_{\mathbb{P}}(M))$$

The sequential composition (;) shorthand is standard; we describe the **observe** construct shortly below, and elaborate on the *soft constraint* construct '**observe from**' in §3.2.

Bppl Typing Judgements. We present the Bppl typing judgements in the technical appendix (Appendix B), where a term M is typed via a judgement $\vdash_p$. Semantically, every open term $\Delta \vdash_p M : \tau$ denotes an *s-finite kernel* $[\![M]\!] : [\![\Gamma]\!] \rightsquigarrow [\![\tau]\!]$. We refer the reader to the work of Staton [2017] for a detailed explanation of the semantics as this is not essential for understanding BaSL.

$$\frac{}{\{\top_1\}\ \mathbf{sample}(\mathbb{P})\ \{X : \mathbb{R}.X \sim \mathbb{P}\}}\ \text{H-Sample} \qquad \frac{}{\{X \sim \mathbb{P}\}\ \mathbf{sample}(p(X))\ \{Y : \mathbb{R}.x \overset{\mathbb{P}}{\leftarrow} X \mid Y \sim p(x)\}}\ \text{H-CondSample}$$

$$\frac{}{\{X \sim \pi\}\ \mathbf{score}(f(X))\ \{X \sim f \cdot \pi\}}\ \text{H-Score} \qquad \frac{}{\{x \overset{\pi}{\leftarrow} X \mid Y \sim p(x)\}\ \mathbf{score}(f(Y))\{x \overset{\pi}{\leftarrow} X \mid Y \sim f \cdot p(x)\}}\ \text{H-CondScore}$$

$$\frac{}{\{X \sim \pi\}\ \mathbf{observe}(P(X))\ \{X \sim \ell_P \cdot \pi\}}\ \text{H-Observe} \qquad \frac{}{\{x \overset{\pi}{\leftarrow} X \mid Y \sim p(x)\}\ \mathbf{observe}(P(Y))\ \{x \overset{\pi}{\leftarrow} X \mid Y \sim \ell_P \cdot p(x)\}}\ \text{H-CondObserve}$$

$$\frac{}{\{Q[\llbracket M \rrbracket/X]\}\ \mathbf{return}\ M\ \{X : \mathcal{A}.Q\}}\ \text{H-Return} \qquad \frac{\vdash \{P\}\ M\ \{X : \mathcal{A}.Q\}}{\vdash \{P * F\}\ M\ \{X : \mathcal{A}.Q * F\}}\ (X \notin \mathsf{fv}(F))\ \text{H-Frame}$$

$$\frac{\vdash \{P\}\ M\ \{X : \mathcal{A}.Q\} \qquad \vdash \forall_{\mathsf{rv}} X : \mathcal{A}.\ \{Q\}\ N\ \{Y : \mathcal{B}.R\}}{\vdash \{P\}\ \mathbf{let}\ X = M\ \mathbf{in}\ N\ \{Y : \mathcal{B}.R\}}\ \text{H-Let} \qquad \frac{P' \vdash P \qquad \vdash \{P\}\ M\ \{X : \mathcal{A}.Q\} \qquad Q \vdash Q'}{\vdash \{P'\}\ M\ \{X : \mathcal{A}.Q'\}}\ \text{H-Cons}$$

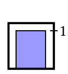
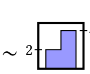
Fig. 9. BASL proof rules

BASL Assertions. The BASL *assertions* are defined by the following grammar:

$$\begin{aligned} P ::= {} & \top \mid \bot \mid P \wedge P \mid P \vee P \mid P \Rightarrow P \mid \forall x : A.P \mid \exists x : A.P \mid P * P \mid P \mathrel{-\!\!*} P \\ & \mid E \sim \pi \mid \mathbb{E}[E] = e \mid \mathsf{own}\, E \mid (x \overset{\pi}{\leftarrow} E \mid P) \mid \mathsf{L}(e) \mid \forall_{\mathsf{rv}} X : \mathcal{A}.P \mid \exists_{\mathsf{rv}} X : \mathcal{A}.P \mid \{P\}\ M\ \{X : \mathcal{A}.P\} \end{aligned}$$

where π, E, e, M range over maps defined later in Fig. 22. Intuitively, E is a random expression, e a deterministic expression, M is the syntax of BPPL, and π is a measure (distribution). The first-order and separation logic assertions (first line of the grammar) are standard. For probabilistic assertions (second line), we have described most of them intuitively in §2, but for $\forall_{\mathsf{rv}}$, $\exists_{\mathsf{rv}}$ quantifiers, which quantify over *random variables*, and the ownership assertion own, which asserts ownership of random expression E.

BASL Proof System. We present the *BASL axiomatic proof system* through a set of rules in Fig. 9. We have intuitively described most of the axioms with the coin flip example in §2 except for axioms related to the **score** construct. To understand what **score** does intuitively, suppose we have a random variable $X \sim \mathsf{Unif}(0, 1)$. Executing $\mathbf{score}(\mathbf{if}\ X < 1/2\ \mathbf{then}\ 2\ \mathbf{else}\ 4)$ then increases the likelihood of values $< 1/2$ being drawn by a factor of 2, and that of values $\geq 1/2$ by a factor of 4:

$\left\{ X \sim \right.$  $\left. \right\}$ $\mathbf{score}(\mathbf{if}\ X < 1/2\ \mathbf{then}\ 2\ \mathbf{else}\ 4)$ $\left\{ X \sim \right.$  $\left. \right\}$

Note that the distribution of X in the postcondition is *unnormalised*, i.e. the shaded region has area (normalising constant) $2 \cdot 1/2 + 4 \cdot 1/2 = 3$. Moreover, when the distribution of X is normalised, we can calculate that $\Pr[X \geq 1/2] = 2 \cdot \Pr[X < 1/2]$. Statistically speaking, it is useful to think of **score** as a way for users to specify the likelihood of an observation, allowing the interpreter to 'mutate' the current distribution. Indeed, as we described above, we encode the hard conditioning $\mathbf{observe}(M)$ construct as $\mathbf{score}(\mathbf{if}\ M\ \mathbf{then}\ 1\ \mathbf{else}\ 0)$, setting to 0 the likelihood of observations where M does not hold, rendering them impossible.

H-Norm-Conj-Norm
$\{\Theta \sim \mathcal{N}(\theta_0, \sigma_0)\}$ **observe** x **from** Normal(Θ, σ) $\{\Theta \sim \mathcal{N}(\theta^{\theta_0,\sigma_0}_{\text{new},x}, \sigma^{\theta_0,\sigma_0}_{\text{new}}) * \text{NormConst}\}$

H-Beta-Conj-Beta
$\{\Theta \sim \text{Beta}(m, n)\}$ **observe** 1 **from** Bern(Θ) $\{\Theta \sim \text{Beta}(m+1, n) * \text{NormConst}\}$

H-Gamma-Conj-Poisson
$\{\lambda \sim \Gamma(k, \theta)\}$ **observe** x **from** Poisson(λ) $\{\lambda \sim \Gamma(k + x, \theta/\theta+1) * \text{NormConst}\}$

$\{\Theta \sim \mathcal{N}(\theta_0, \sigma_0)\}$ // H-Sample
  observe x **from** Normal(Θ, σ) // desugars to **score**(normal-pdf$(x_1 \mid \Theta, \sigma)$)
$\{\Theta \sim \text{normal-pdf}(x \mid -, \sigma) \cdot \mathcal{N}(\theta_0, \sigma_0)\}$ // H-Score
$\{\Theta \sim \mathcal{N}(\theta^{\theta_0,\sigma_0}_{\text{new},x}, \sigma^{\theta_0,\sigma_0}_{\text{new}}) * \text{NormConst}\}$ // H-Cons

Fig. 11. Examples of derived conjugate priors in BaSL (above); a BaSL derivation of H-Norm-Conj-Norm (below), where the //annotation denotes the BaSL rule(s) applied to obtain the associated postcondition.

3.2 Conjugate Priors as Hoare Triples: Verifying a Parameter Estimation Algorithm

We begin our journey of verifying statistical models by considering a useful language feature known as *soft constraint* in languages such as Stan and Anglican, which allows users to specify observations even when they are drawn from *continuous* distributions. BaSL is the *first* probabilistic separation logic that can reason about soft constraints and Bayesian updating.

As described above, we encode the *soft constraint conditioning* construct as **observe** x **from** $\mathbb{P}$:= **score**(density$_{\mathbb{P}}(x)$). The **observe** x **from** $\mathbb{P}$ denotes a distribution $\mathbb{P}$ that has density with respect to either the Lebesgue measure $\lambda_{\mathbb{R}}$ or the counting measure $\#_{\mathbb{N}}$ (which includes 'common' distributions such as normal, binomial, gamma, etc.). We write density$_{\mathbb{P}}$ to denote the corresponding density function of $\mathbb{P}$ (see works of Vákár and Ong [2018, §7] and Staton [2020, §4]). Intuitively, the score is higher if the observed x is more likely to be generated from $\mathbb{P}$, and the score is 0 if the observation is not possible, e.g. **observe** 2 **from** Unif(0, 1) is tantamount to **score**(0).

```
let Θ = sample(Normal(θ0, σ0)) in
observe x1 from Normal(Θ, σ);
observe x2 from Normal(Θ, σ);
return Θ
```

Fig. 10. The GaussParam program

We next consider the GaussParam example by Lee [2012, §2] in Fig. 10, where we have a normally distributed population with a known standard deviation σ, and we would like to estimate its mean Θ. To do this, we assume a normal prior $\Theta \sim \mathcal{N}(\theta_0, \sigma_0)$, where θ_0, σ_0 are constants. Suppose we draw two samples $\{x_1, x_2\}$ from the dataset; we can then write **observe** x_i **from** Normal(Θ, σ) for $i \in \{1, 2\}$ to compute our updated belief of Θ, as shown in Fig. 10. Intuitively, the mean of Θ shifts *up* when $x_i > \Theta$ and shifts down otherwise. Using Bayes' rule, we can compute the updated mean and standard deviation as follows [Lee 2012, §2.2.1]:

$$\theta^{\theta_0,\sigma_0}_{\text{new},x} := (\sigma^{\theta_0,\sigma_0}_{\text{new}})^2 \left(\frac{\theta_0}{\sigma_0^2} + \frac{x}{\sigma^2}\right) \qquad \sigma^{\theta_0,\sigma_0}_{\text{new}} := \sqrt{\frac{1}{\sigma_0^{-2} + \sigma^2}}$$

Note that the updated distribution has a closed-form solution (which is not usually the case). Indeed, the Normal-prior/Normal-likelihood pair is an instance of a *conjugate distribution*, which is a prior/likelihood pair that leads to a closed-form posterior distribution via Bayes' rule.

In BaSL we can obtain this fact using the *derived* H-Norm-Conj-Norm rule in Fig. 11 (above), with its BaSL derivation given at the bottom of Fig. 11. In fact, the ability to derive conjugate priors

$$
\begin{array}{ll}
\{\top_1\} & \\
\quad \textbf{let } \Theta = \textbf{sample}(\mathsf{Normal}(\theta_0, \sigma_0)) \textbf{ in} & \\
\{\Theta \sim \mathcal{N}(\theta_0, \sigma_0)\} & \text{// H-SAMPLE} \\
\quad \textbf{observe } x_1 \textbf{ from } \mathsf{Normal}(\Theta, \sigma) & \text{// let } \theta_{i+1} := \theta^{\theta_i,\sigma_i}_{\text{new},x_{i+1}};\ \sigma_{i+1} := \sigma^{\theta_i,\sigma_i}_{\text{new}} \text{ for } i \geq 0 \\
\{\Theta \sim \mathcal{N}(\theta_1, \sigma_1) * \mathsf{NormConst}\} & \text{// H-NORM-CONJ-NORM} \\
\text{H-FRAME} \left| \begin{array}{l} \{\Theta \sim \mathcal{N}(\theta_1, \sigma_1)\} \\ \quad \textbf{observe } x_2 \textbf{ from } \mathsf{Normal}(\Theta, \sigma) \\ \{\Theta \sim \mathcal{N}(\theta_2, \sigma_2) * \mathsf{NormConst}\} \end{array} \right. & \text{// H-FRAME and H-NORM-CONJ-NORM} \\
\{\Theta \sim \mathcal{N}(\theta_2, \sigma_2) * \mathsf{NormConst} * \mathsf{NormConst}\} & \\
\{\Theta \sim \mathcal{N}(\theta_2, \sigma_2) * \mathsf{NormConst}\} & \text{// H-CONS} \\
\quad \textbf{return } \Theta & \\
\{\Theta \sim \mathcal{N}(\theta_2, \sigma_2) * \mathsf{NormConst}\} & \text{// H-RET}
\end{array}
$$

Fig. 12. A BaSL proof sketch of GaussParam.

in BaSL allows us to perform symbolic reasoning. To this end, in Fig. 11 we list a few common conjugate priors that can be derived in BaSL. The ability to derive conjugate priors in BaSL, along with the frame rule, gives us a foundation for *modular and symbolic reasoning* of probabilistic programs: for **observe** x_1 **from** $\mathsf{Normal}(\Theta, \sigma)$, we apply the frame rule to 'frame out' the normalising constant, then apply the conjugate distribution triple, which then allows us to symbolically derive the posterior distribution for Θ in Fig. 12.

3.3 Verifying the 'Hello World' of Probabilistic Programming: Burglar Alarm

With soft constraint conditioning defined, we can now consider a classic example in Bayesian statistics, the BurglarAlarm program at the top of Fig. 13. The problem is as follows: your home has a burglar alarm that activates when there is a burglar, but it also accidentally activates when there is an earthquake. There is a 1% probability that there is a burglar, and a 10% probability that there is an earthquake. When the alarm activates, there is a 99% probability of the phone ringing. Assuming that your phone is ringing now, what is the probability of there being an earthquake?

We use BaSL in Fig. 13 to prove that the probability of having an earthquake is roughly 84.7% by using a technique called *joint conditioning*. Note that random variable A takes a value in $\{0, 1\}$ depending on whether there is a burglar (captured by B) or an earthquake (E). By conditioning on the *joint* random variable $(B, E) = (b, e)$ for some $(b, e) \in \{0, 1\}^2$, we know A has distribution $\delta_{b \vee e}$. We then apply the rule of consequence H-Cons with our internal notion of Bayes' theorem E-Bayes (formally defined later in Theorem 4.11), we obtain a posterior distribution stating that there is an 84.7% probability of there being an earthquake, as shown below, with f defined in Fig. 13:

$$
\overbrace{(b, e) \xleftarrow{\mathsf{Bern}(0.01) \otimes \mathsf{Bern}(0.1)} (B, E)}^{\text{prior}} \mid \overbrace{\mathsf{L}(f(b, e))}^{\text{likelihood}} \vdash \overbrace{(B, E) \sim f \cdot (\mathsf{Bern}(0.01) \otimes \mathsf{Bern}(0.1))}^{\text{posterior}} \qquad \text{(Theorem 4.11)}
$$

$$
\vdash E \sim \mathsf{Bern}({}^{0.099}/{}_{0.11682}) \qquad \text{(calculation)}
$$

3.4 Reasoning about Independence and Correlations: The Collider Bayesian Network

We now consider a *Bayesian network* in which two random variables X and Y that are initially independent become negatively correlated after performing Bayesian conditioning. BaSL is the *first* logic that can reason about such conditional dependence brought about by Bayesian updating.

```
let B = sample(Bern(0.01)) in                               // probability of burglary is 1%
let E = sample(Bern(0.1)) in                                // probability of earthquake is 10%
let A = return(B ∨ E) in                                    // alarm activates in case of burglary or earthquake
observe 1 from Bern(if A then 0.99 else 0.01);              // observe the phone is ringing
return E                                                    // is there an earthquake?
```

$\{\top_1\}$
  let B = **sample**(Bern(0.01)) **in**
$\{B \sim \text{Bern}(0.01)\}$ // H-Sample
  let E = **sample**(Bern(0.1)) **in**
$\{B \sim \text{Bern}(0.01) * E \sim \text{Bern}(0.1)\}$ // H-Sample, H-Frame
$\{(B, E) \sim \text{Bern}(0.01) \otimes \text{Bern}(0.1)\}$ // H-Cons
  let A = **return**$(B \vee E)$ **in**
$\{(b, e) \xleftarrow{\text{Bern}(0.01)\otimes\text{Bern}(0.1)} (B, E) \mid A \sim \delta_{b \vee e}\}$ // H-CondSample
  observe 1 **from** Bern(**if** A **then** 0.99 **else** 0.01); // with $f(b, e) := \text{if } b \vee e \text{ then } 0.99 \text{ else } 0.1$
$\left\{(b, e) \xleftarrow{\text{Bern}(0.01)\otimes\text{Bern}(0.1)} (B, E) \;\middle|\; A \sim f \cdot \delta_{b \vee e}\right\}$ // H-CondScore
$\left\{(b, e) \xleftarrow{\text{Bern}(0.01)\otimes\text{Bern}(0.1)} (B, E) \;\middle|\; \mathsf{L}\,(f(b, e))\right\}$ // H-Cons
$\{(B, E) \sim f \cdot (\text{Bern}(0.01) \otimes \text{Bern}(0.1))\}$ // H-Cons with E-Bayes
$\{E \sim \text{Bern}(0.099/0.11682) * \text{NormConst}\}$ // H-Cons
$\{\mathbb{E}[E] = 0.099/0.11682 * \text{NormConst}\}$ // H-Frame, H-Cons
  return E
$\{\mathbb{E}[E] = 0.099/0.11682 * \text{NormConst}\}$ // H-Return
$\{\Pr[E = 1] \approx 0.847 * \text{NormConst}\}$ // H-Cons

Fig. 13. BurglarAlarm (above); BaSL proof sketch for deriving the posterior probability of earthquake (below)

```
let X = sample(Bern(1/2)) in
let Y = sample(Bern(1/2)) in
let Z = return X ∨ Y in
observe(Z = 1);
return (X, Y)
```

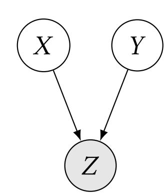


Fig. 14. Collider and its Bayesian network

A Bayesian network is a directed acyclic graph modelling relationship of random variables. A common structure in Bayesian networks is known as *colliders* (or *common effects*), where the distribution of a random variable Z is conditionally independent upon X and Y. For our example in Fig. 14, we flip two coins X and Y (taking values in $\{0, 1\}$) and let Z be the maximum of X and Y. Before observing Z, X and Y are independent. That is, we can use BaSL to prove $X \sim \text{Bern}(1/2) * Y \sim \text{Bern}(1/2)$ prior to observing Z. However, once we perform Bayesian conditioning by observing $Z = 1$, X and Y are no longer independent: knowing the value of Z gives us information about X and Y. For example, knowing $Z = 1$ and $X = 0$ gives us extra information $Y = 1$. In fact, not only X and Y are now dependent, they are also *negatively correlated*: when X has a higher value, Y is likely to be lower, and vice versa.

We use BaSL to prove this negative correlation as shown in Fig. 15. We first apply H-Sample to sample X with distribution $1/2$, then apply H-Frame and H-Sample to obtain $Y \sim 1/2$ and 'frame' Y onto X to obtain $X \sim \text{Bern}(1/2) * Y \sim \text{Bern}(1/2)$. As we show shortly in §4, the (bi)entailment $X \sim \mu * Y \sim \nu \dashv\vdash (X, Y) \sim \mu \otimes \nu$ holds and we thus use H-Cons to obtain $(X, Y) \sim \text{Bern}(1/2) \otimes \text{Bern}(1/2)$.

$\{\top_1\}$
  let X = **sample**(Exp(1)) **in**
$\{X \sim \mathsf{Exp}(1)\}$ // H-SAMPLE
  score(e^X)
$\{X \sim \exp \cdot \mathsf{Exp}(1)\}$ // H-SCORE
$\{X \sim \mathsf{Leb}_{[0,\infty)}\}$ // H-CONS

Fig. 16. LEBESGUE

$\{X \sim \mathsf{Leb}_{[0,\infty)}\}$
  observe X **from** μ
$\{X \sim \mathsf{pdf}_\mu \cdot \mathsf{Leb}_{[0,\infty)}\}$ // H-SCORE
$\{X \sim \mu\}$ // H-CONS

Fig. 17. Conditioning of improper Lebesgue prior

We next condition on (X, Y) and sample Z using the conditional sampling axiom H-CONDSAMPLE to show Z has a conditional distribution $\delta_{x \vee y}$, (jointly) conditioning on the random variable $(X, Y) = (x, y)$ for $(x, y) \in \{0, 1\}^2$. Upon subsequently observing $Z = 1$, we apply the conditional observation axiom H-CONDOBSERVE to prove the updated distribution 'ignores' the case when $X = 0$ and $Y = 0$. Let us consider the updated likelihood of all four cases of $(x, y) \in \{0, 1\}^2$: when $x = y = 0$, the likelihood is 0, otherwise it is 1. That is, the likelihood proposition L(if $x = y = 0$ then 0 else 1) holds, or equivalently, L$(1 - [x = y = 0])$ holds. Using the *internal Bayes' theorem* (formulated later in Theorem 4.11), we thus know the updated (X, Y) has distribution $(X, Y) \sim (1 - [x = y = 0]) \cdot \mathsf{Bern}(1/2)^2$. We compute the probability of (X, Y) being $(0, 1)$, $(1, 0)$ or $(1, 1)$ is $1/3$ in each case, and thus calculate the covariance $\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 1/3 - 2/3 \cdot 2/3 = -1/9$. Using H-CONS, we then formally prove $\mathrm{Cov}[X, Y] < 0$, stating that X and Y are negatively correlated.

$\{\top_1\}$
  let X = **sample**(Bern(1/2)) **in**
$\{X \sim \mathsf{Bern}(1/2)\}$
  let Y = **sample**(Bern(1/2)) **in**
$\{X \sim \mathsf{Bern}(1/2) * Y \sim \mathsf{Bern}(1/2)\}$
$\{(X, Y) \sim \mathsf{Bern}(1/2) \otimes \mathsf{Bern}(1/2)\}$
  let Z = **return** $X \vee Y$ **in**
$\{(x, y) \xleftarrow{\mathsf{Bern}(1/2) \otimes \mathsf{Bern}(1/2)} (X, Y) \mid Z \sim \delta_{x \vee y}\}$
  observe($Z = 1$);
$\{(x, y) \xleftarrow{\mathsf{Bern}(1/2) \otimes \mathsf{Bern}(1/2)} (X, Y) \mid Z \sim \ell_{=1} \cdot \delta_{x \vee y}\}$
$\{(x, y) \xleftarrow{\mathsf{Bern}(1/2) \otimes \mathsf{Bern}(1/2)} (X, Y) \mid \mathsf{L}(1 - [x = y = 0])\}$
  return (X, Y)
$\{(X, Y) \sim (1 - [x = y = 0]) \cdot \mathsf{Bern}^2(1/2)\}$
$\{(X, Y) \sim \mathsf{Unif}\{(0, 1), (1, 0), (1, 1)\} * \mathsf{NormConst}\}$
$\{\mathbb{E}[XY] = 1/3 \wedge \mathbb{E}[X] = \mathbb{E}[Y] = 2/3 * \mathsf{NormConst}\}$
$\{\mathrm{Cov}[X, Y] < 0 * \mathsf{NormConst}\}$

Fig. 15. A BASL proof sketch of COLLIDER showing that return values (X, Y) are negatively correlated

3.5 Modelling Improper Prior: Correctness of the Semantic Lebesgue Measure

We now consider a common technique in Bayesian statistical modelling known as *improper priors*. An improper prior is a distribution with an infinite normalising constant. One canonical such example [Narayanan et al. 2016; Staton 2020] is the *Lebesgue measure*, $\mathsf{Leb}_{[0,\infty)}$, which assigns length $b - a$ to any interval (a, b) with $0 \le a < b$. As shown by Staton [2020], a BPPL program can simulate the Lebesgue measure as the computation **let** X = **sample**(Exp(1)) **in score**(e^X) since the *measure* (as opposed to probability) of $X \in (a, b)$ is $\int_{(a,b)} e^x\, e^{-x} \mathrm{d}x = b - a$. In fact, X has an *improper prior* – the normalising constant is currently $\int_0^\infty e^x \cdot e^{-x}\, \mathrm{d}x = \int_0^\infty 1\, \mathrm{d}x = \infty$.

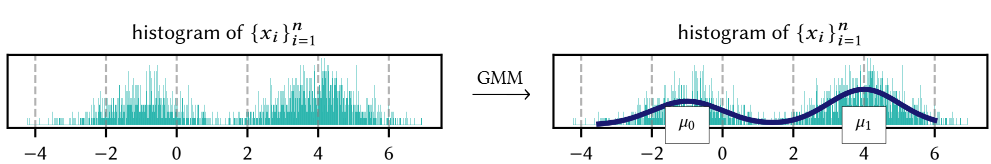
BASL is the *first* logic that can reason about improper priors because its semantic domain include *σ-finite measure spaces*. We present a BASL proof sketch of this computation in Fig. 16, proving that X is distributed according to the Lebesgue measure: $X \sim \mathsf{Leb}_{[0,\infty)}$. In fact, since $X \sim \mathsf{Leb}_{[0,\infty)}$, we know that for any measure μ with a density with respect to $\mathsf{Leb}_{[0,\infty)}$ (e.g. $\mu = \mathsf{Unif}(0, 1)$), observing X to have distribution μ results in $X \sim \mu$. We can derive this in BASL as shown in Fig. 17.

$\{\top_1\}$
  let π = **sample**(Unif(0, 1)) **in** // a prior that
$\{\pi \sim \text{Unif}(0,1)\}$ // H-Sample, H-Frame
  let μ = **sample**($\mathcal{N}(0,10) \otimes \mathcal{N}(0,10)$) **in** // sample from two normal distributions
$\{\mu \sim \mathcal{N}^2(0,10) * \pi \sim \text{Unif}(0,10)\}$ // H-Sample
  for X **in** $[x_1, \ldots, x_n]$ **do** // let $\mathbb{P}_0 \coloneqq \mathcal{N}^2(0,1) \otimes \text{Unif}(0,1)$
    $\{(\mu,\pi) \sim \mathbb{P}_i\}$ // H-Cons
      let I = **sample**(Bern(π)) **in** // let $f_i(m,p) \coloneqq (1-p)\mathcal{N}(x_i \mid m_0, 1) + p\mathcal{N}(x_i \mid m_1, 1)$
    $\{(m,p) \overset{\mathbb{P}_i}{\longleftarrow} (\mu,\pi) \mid I \sim \text{Bern}(p)\}$ // H-CondSample
      observe X **from** Normal(μ_I, 1) // let $\mathbb{P}_{i+1} \coloneqq f_i \cdot \mathbb{P}_i$
    $\{(m,p) \overset{\mathbb{P}_i}{\longleftarrow} (\mu,\pi) \mid I \sim f_i \cdot \text{Bern}(p)\}$ // H-CondScore
    $\{(\mu,\pi) \sim \mathbb{P}_i\}$ // H-Cons, E-Bayes
$\{(\mu,\pi) \sim \mathbb{P}_n\}$ // H-BoundedFor

Fig. 18. An implementation of the Gaussian mixture model and a BaSL proof sketch deriving its posterior

3.6 Representing the Posterior of a Bayesian Clustering Algorithm

For our final example, we use BaSL to calculate the posterior of a clustering algorithm that implements the Gaussian mixture model (GMM). The problem is as follows: suppose we have a dataset $\{x_i \in \mathbb{R}\}_{i=1}^n$ and we want to cluster them into two groups. The model works by assuming there are two normally-distributed latent variables $\mu_0, \mu_1 \in \mathbb{R}$ that represent the mean of the two groups and another latent variable $\pi \in [0,1]$ representing the proportion of samples belonging to the two groups. For each x_i, we sample a Bernoulli random variable N with success probability π and we assume x_i is a sample drawn from $\mathcal{N}(\mu_N, 1)$ by performing soft conditioning. The diagram below illustrates the main idea of GMM: given a dataset $\{x_i\}_{i=1}^n$, we infer the mean of clusters μ_0 and μ_1 and fit the data via the mixture of two Gaussian distributions.



Unlike the previous examples, GMM does not have a closed-form solution – we must rely on inference algorithms such as MCMC to approximate the posterior distribution of (μ,π). However, with BaSL, we can symbolically represent the formula for the posterior distribution. Similar to Lilac, we extend Bppl with a syntactic construct that encodes bounded loop over a literal list $[x_1, \ldots, x_n]$ by defining **for** X **in** $[x_1, \ldots, x_n]$ **do** $M \coloneqq M[x_1/X]; \ldots; M[x_n/X]$. By applying the sequencing rule H-Let inductively, we obtain the rule H-BoundedFor, which allows us to express GMM and represent the posterior of (μ,π) in Fig. 18. Even though the posterior distribution cannot be explicitly simplified, we can still represent it as the postcondition $(\mu,\pi) \sim \mathbb{P}_n$.

H-BoundedFor
$$\frac{\text{for } i = 1, \ldots, n,\ \{P_{i-1}\}\ M[{}^{x_i}/X]\ \{P_i\}}{\{P_0\}\ \textbf{for}\ X\ \textbf{in}\ [x_1, \ldots, x_n]\ \textbf{do}\ M\ \{P_n\}}$$

Through numerous examples showcasing the novel and hitherto-unsupported features of BaSL, we have demonstrated the expressivity of BaSL, and we believe this can serve as a logical foundation for static analysis/symbolic execution tools for probabilistic programs.

4 THE SEMANTICS OF BaSL

Now that we have demonstrated how BaSL can function as a logical framework for proving properties in statistical models, we explain its resource-theoretic semantics in §4.4. But before this, we review the Kripke resource model of the Lilac separation logic [Li et al. 2023] in §4.1, which the model of BaSL is based on, then we motivate the need for a generalised model of randomness for Bayesian updating in §4.2 and prove that it is indeed a resource model in §4.3.

4.1 Background: A Resource Monoid for Randomisation

In separation logic, *computational resources* such as heaps are modelled by *partial commutative monoids* $(\mathcal{M}, \bullet, \mathbf{1})$ [Calcagno et al. 2007]. The set $\mathcal{M}$ represents states of the resource and the partial function $(\bullet) : \mathcal{M} \times \mathcal{M} \rightharpoonup \mathcal{M}$ combines two states $m_1, m_2 \in \mathcal{M}$ if they are *compatible* (e.g. m_1 and m_2 describe different parts of the resource). For example, the heap is modelled by $\mathcal{M}_{\mathsf{heap}} := \mathsf{Loc} \rightharpoonup_{\mathsf{fin}} \mathsf{Val}$, where Loc is the set of memory addresses and Val is the set of values. For instance, $\{42 \mapsto \text{“a”}\} \in \mathcal{M}_{\mathsf{heap}}$ represents a heap where address 42 stores value “a”. Moreover, given two heaps m_1, m_2, $(\bullet)$ combines them if they contain separate addresses. For example, $m_1 := \{21 \mapsto 2025, 42 \mapsto \text{“a”}\}$ and $m_2 := \{52 \mapsto 123\}$ can be combined to $m_1 \bullet m_2 = \{21 \mapsto 2025, 42 \mapsto \text{“a”}, 52 \mapsto 123\}$ since the addresses in m_1 (21 and 42) do not overlap with 52 in m_2. There is also an identity element $\mathbf{1} \in \mathcal{M}$ that represents the empty resource. For heaps, $\mathbf{1}$ is defined to be the empty heap $\{\}$.

Applying the same intuition to probability, one reasonable model of computational resource is that of a *random number generator*. The Lilac separation logic defined a partial commutative monoid that models a random number generator (RNG) by encoding the following two properties:

(1) the *distribution* of the generated numbers, i.e. a random number generator should have information about the distribution of the numbers generated; and
(2) the *usage* of the generator, which has information regarding whether the i-th number ω_i has been generated.

To model (1) and (2), Lilac uses a *measure space*. In fact, the distribution and usage property can be modelled via the *measures* and *σ-algebras*, respectively. For readers unfamiliar with these concepts, a *measurable space* is a pair $(\Omega, \mathcal{F})$, where Ω is a set and $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is a set of subsets such that $\emptyset, \Omega \in \mathcal{F}$ and $\mathcal{F}$ is closed under complements and countable unions. We call Ω the *sample space* and $\mathcal{F}$ a *σ-algebra* of Ω. A *measure* of $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to [0, \infty]$ satisfying $\mu(\emptyset) = 0$ and $\mu(\biguplus_{i \in \mathbb{N}} U_i) = \sum_{i \in \mathbb{N}} \mu(U_i)$. The triple $(\Omega, \mathcal{F}, \mu)$ forms a *measure space*. If $\mu(\Omega) = 1$, we additionally call $(\Omega, \mathcal{F}, \mu)$ a *probability space*. For example, to model a fair die, we set $\Omega := \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} := \mathcal{P}(\Omega)$, and $\mu(U) := \sum_{i=1}^{6} \mathbf{1}_U(i) \cdot 1/6$, where $\mathbf{1}_U(i)$ is 1 when $i \in U$ and 0 when $i \notin U$. We can then compute that the probability of the die having a value greater than four is $\mu(\{5, 6\}) = 1/3$.

To understand this better, consider a simplified model of Lilac's random number generator – we define our sample space to be $\Omega := \{\mathsf{T}, \mathsf{F}\}^2$ (as opposed to the full model of Lilac, which uses $[0, 1]^{\mathbb{N}}$ as the sample space), which, intuitively, can be thought of as the computer having access to an RNG that can generate two independent booleans. The state of a random number generator can then be modelled by the following set:

$$\mathcal{M} := \{(\mathcal{F}, \mu) \mid (\Omega, \mathcal{F}, \mu) \text{ is a probability space}\}$$

To explain how $\mathcal{M}$ models random number generators, we construct the empty generator $\mathbf{1} := (\mathcal{F}_{\mathbf{1}}, \mu_{\mathbf{1}})$ – an element of $\mathcal{M}$ that represents a state where nothing has been generated. This means:

(1) We *do not* know the probability of the first boolean being F and second boolean being T. Hence, $\{(\mathsf{F}, \mathsf{T})\} \notin \mathcal{F}_{\mathbf{1}}$ and we cannot apply $\mu_{\mathbf{1}} : \mathcal{F}_{\mathbf{1}} \to [0, \infty]$ to $\{(\mathsf{F}, \mathsf{T})\}$ to get the probability. Similarly, $\{(a, b)\} \notin \mathcal{F}_{\mathbf{1}}$ for $a, b \in \{\mathsf{T}, \mathsf{F}\}$.

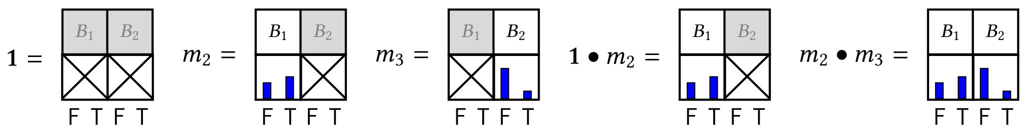


Fig. 19. Illustrations of random number generators

(2) We *do not* know the probability of the first boolean being T, which is equivalent to saying the first boolean being T and the second boolean being T *or* F. Hence, $\{(\mathsf{T},\mathsf{T}),(\mathsf{T},\mathsf{F})\} \notin \mathcal{F}_1$. Similarly, $\{(a,b),(c,d)\} \notin \mathcal{F}_1$ for $a,b,c,d \in \{\mathsf{T},\mathsf{F}\}$.
(3) Even though the first number has not been generated, we know that the first and/or second boolean will either be T or F. Hence, $\{(\mathsf{T},\mathsf{F}),(\mathsf{T},\mathsf{T}),(\mathsf{F},\mathsf{T}),(\mathsf{F},\mathsf{F})\} = \Omega \in \mathcal{F}_1$.

For $\mathcal{F}_1$ to be a σ-algebra, we need $\varnothing \in \mathcal{F}_1$; hence, $\mathcal{F}_1 = \{\varnothing, \Omega\}$. For the probability measure $\mu_1 : \mathcal{F}_1 \to [0,\infty]$, we know that the first and second boolean have a 100% probability of being T or F. Hence, $\mu_1(\{(\mathsf{T},\mathsf{F}),(\mathsf{T},\mathsf{T}),(\mathsf{F},\mathsf{T}),(\mathsf{F},\mathsf{F})\}) = 1$. See Fig. 19 for an illustration.

Next, we construct $m_2 := (\mathcal{F}_2, \mu_2) \in \mathcal{M}$ where the first boolean has a 42% probability of being T and 58% probability of being F, and only the first boolean has been generated. This means:

(1) We know the probability of the first boolean being T is 42%, which is equivalent to saying the first boolean being T and the second boolean being T or F is 42%. Hence, $\{(\mathsf{T},\mathsf{F}),(\mathsf{T},\mathsf{T})\} \in \mathcal{F}_2$ and $\mu_2(\{(\mathsf{T},\mathsf{F}),(\mathsf{T},\mathsf{T})\}) := 0.42$. Similarly, $\{(\mathsf{F},\mathsf{F}),(\mathsf{F},\mathsf{T})\} \in \mathcal{F}_2$ and $\mu_2(\{(\mathsf{F},\mathsf{F}),(\mathsf{F},\mathsf{T})\}) := 0.58$.
(2) We *do not* know the probability of the second boolean being F, which is equivalent to saying the second boolean being F and the first boolean being T or F. Hence, $\{(\mathsf{T},\mathsf{F}),(\mathsf{F},\mathsf{F})\} \notin \mathcal{F}_2$. Similarly, the singletons $\{(a,b)\}$ with $a,b \in \{\mathsf{T},\mathsf{F}\}$ and $\{(\mathsf{T},\mathsf{T}),(\mathsf{F},\mathsf{T})\}$ should not be in $\mathcal{F}_2$ since they contain information about the second boolean.

Since $\{(\mathsf{T},\mathsf{T}),(\mathsf{F},\mathsf{T})\}$ and $\{(\mathsf{F},\mathsf{F}),(\mathsf{F},\mathsf{T})\}$ are in $\mathcal{F}_2$, their union Ω is also in $\mathcal{F}_2$ and $\mu_2(\Omega) = 1$, which makes sense since the probability of either events happening is 42% + 58% = 100%.

Similarly, we can construct $m_3 = (\mathcal{F}_3, \mu_3) \in \mathcal{M}$ where only the second boolean has been generated with 30% probability of being T and 70% probability of being F. Now, since m_2 only has information about the first boolean and m_3 only has information about the second boolean, they can be combined to form $m_2 \bullet m_3 = (\mathcal{F}_{23}, \mu_{23})$ such that we can refer to e.g. the probability of the first boolean being T and the second boolean being F, i.e. $\{(\mathsf{T},\mathsf{F})\} \in \mathcal{F}_{23}$ and $\mu_{23}(\{\mathsf{T},\mathsf{F}\}) := \mu_2(\{(\mathsf{T},\mathsf{F}),(\mathsf{T},\mathsf{T})\}) \cdot \mu_3(\{(\mathsf{F},\mathsf{F}),(\mathsf{F},\mathsf{T})\}) = 42\% \cdot 70\% = 29.4\%$ (see Fig. 19). As shown by Li et al. [2023], $(\mathcal{M}, \bullet, \mathbf{1})$ forms a partial commutative monoid (PCM).

4.2 The Need for an Extended Resource Model of Randomness

We now motivate the semantics of BaSL by considering a problem we would like to solve; namely, how a random generator, such as the ones presented above, can be *updated*. In the heap model of separation logic (see first paragraph of §4.1), an update on address l with value v is modelled as an *update function* $(-)[l \mapsto v] : \mathcal{M}_{\text{heap}} \to \mathcal{M}_{\text{heap}}$ defined by $m[\ell \mapsto v] := m \cup \{\ell \mapsto v\}$. We develop an analogous operation on random number generators, for which we must model the **observe**/**score** construct. But first, let us establish the motivation: why would we want to update a distribution? And why is it such a (notoriously) hard problem? The answer is twofold: we need to handle both *unnormalised measures* and the update's effect on *dependent random variables*.

Unnormalised Measures. Consider an experiment where we flip two fair coins B_1 and B_2 (modelled as booleans). The resulting random generator can be visualised as m_{before} in Fig. 20.

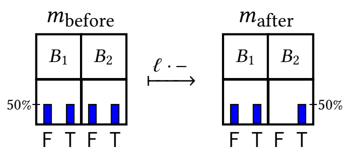


Fig. 20. Updating via likelihood function

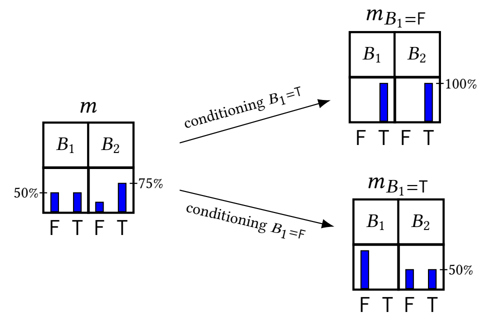


Fig. 21. A dependent random generator

Suppose that, based on experimental data, we know B_2 must be T; we can then update our belief about B_2 via a *likelihood function*, which is a function of type $\ell : \Omega \to [0, \infty)$, defined by $\ell(b_1, \mathsf{T}) := 100\%$ and $\ell(b_1, \mathsf{F}) := 0\%$. For any likelihood function ℓ and $m = (\mathcal{P}(\Omega), \mu) \in \mathcal{M}$, there is a natural update operation $(\cdot)$ such that $(\ell \cdot \mu) : \mathcal{P}(\Omega) \to [0, \infty]$ is a measure on $(\Omega, \mathcal{P}(\Omega))$ defined as follows:

$$(\ell \cdot \mu)(E) := \sum_{(b_1, b_2) \in E} \ell(b_1, b_2) \cdot \mu\{(b_1, b_2)\},$$

This update scales the measure μ according to ℓ. Let us define $(\ell \cdot m) := (\mathcal{F}, \ell \cdot \mu)$ for all $m = (\mathcal{F}, \mu) \in \mathcal{M}$. The update $m_{\text{after}} := \ell \cdot m_{\text{before}}$ then reflects our belief: the probability that $B_2 = \mathsf{F}$ is now zero since $\ell(b_1, \mathsf{F}) = 0$ and $m_{\text{after}}\{B_2 = \mathsf{F}\} = m_{\text{after}}\{(\mathsf{T}, \mathsf{F}), (\mathsf{F}, \mathsf{F})\} = 0$. However, there is a problem: m_{after} is no longer a probability measure because it does not add up to 100%:

$$\mu_{\text{after}}(\Omega) = \mu_{\text{after}}\{B_2 = \mathsf{T}\} + \mu_{\text{after}}\{B_2 = \mathsf{F}\} = 0 + 50\% = 50\%.$$

This means the current distribution is *unnormalised*, with the *normalising constant* being $1/50\% = 2$. This also means $m_{\text{after}} \notin \mathcal{M}$ as $\mathcal{M}$ only includes probability spaces. To solve this problem, we extend $\mathcal{M}$ to include non-probability measures as well.

Dependent Random Variables. A far more challenging problem is to handle dependency between random variables. Consider an experiment where we toss a fair coin B_1, and depending on the result of B_1, we obtain B_2 by sampling from two different distributions:

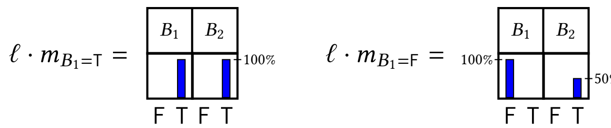
$$B_2 \sim \begin{cases} \text{toss a biased coin that is always } \mathsf{T} & \text{if } B_1 = \mathsf{T} \\ \text{toss a fair coin} & \text{if } B_1 = \mathsf{F}. \end{cases}$$

Assuming that $B_2 = \mathsf{T}$, what is the probability that $B_1 = \mathsf{T}$?

Notice that B_1 and B_2 are not probabilistically independent: knowing B_1 is T means B_2 is T 100% of the time, and knowing B_1 is F means B_2 is T 50% of the time (which means overall, B_2 is T 75% of the time). Suppose $m \in \mathcal{M}$ represents this distribution; we represent the dependency between B_1 and B_2 in m pictorially in Fig. 21.

To address the problem above, note that when we mutate the distribution of B_2 via ℓ, the distribution of B_1 changes as well: indeed, knowing $B_2 = \mathsf{T}$ makes $B_1 = \mathsf{T}$ more probable, as $B_2 = \mathsf{T}$ is more likely to be caused by flipping the biased coin that is always T, rather than the fair coin. To demonstrate how updating B_2 causes the update of B_1, we perform a 'case analysis' and apply

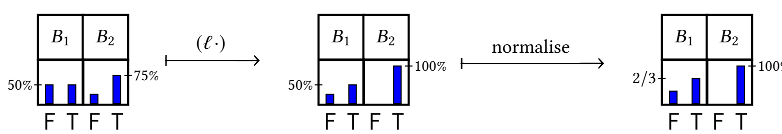
$(\ell \cdot -)$ to the conditioned space $m_{B_1=\mathsf{T}}$ and $m_{B_1=\mathsf{F}}$:



Since we now have information regarding the original distribution of B_1 (the *prior* distribution) as well as the likelihood, we can apply Bayes' theorem. Specifically, for all $b \in \{\mathsf{T}, \mathsf{F}\}$:

$$\overbrace{(\ell \cdot m)\{B_1 = b\}}^{\text{unnormalised posterior}} = \overbrace{\mu\{B_1 = b\}}^{\text{prior}} \cdot \overbrace{(\ell \cdot m_{B_1=b})\{B_2 = \mathsf{T}\}}^{\text{likelihood}}$$
$$= 1/2 \cdot (\text{if } b = \mathsf{T} \text{ then } 100\% \text{ else } 50\%).$$

This means $(\ell \cdot m)\{B_1 = \mathsf{T}\} = 1/2$ and $(\ell \cdot m)\{B_1 = \mathsf{F}\} = 1/4$ and $\ell \cdot m$ is an unnormalised measure ($1/2 + 1/4 \neq 1$). Normalising $\ell \cdot m$ yields the desired distribution – assuming $B_2 = \mathsf{T}$, B_1 has $1/3$ probability of being F, and $2/3$ probability of being T, as shown below:



To model these features in BaSL, we develop a novel resource model for randomisation as follows:

(1) To handle unnormalised measures, we extend Lilac's partial commutative monoid by allowing non-probability measure spaces and impose several *finiteness* restrictions (Definition 4.1) and show that the resulting structure remains a partial commutative monoid (Theorem 4.3), i.e. a model of separation logic. In fact (as in Lilac), it forms an even richer structure known as a *Kripke resource monoid* (Corollary 4.5).

(2) Given an unnormalised random number generator $m \in \mathcal{M}$ and a random variable X, if the underlying space of X is a *standard Borel space* (defined later in Fig. 22), then the family of generators $m_{X=x}$ conditioning on $X = x$ (formally, the *disintegration* of m over X) exists (Theorem 4.7). This allows us to perform conditional reasoning: in order to reason about the dependencies between random variables, we need to reason about conditioned spaces.

(3) Random variables arising from random generators in BaSL can be *updated* via a logical version of Bayes' theorem (Theorem 4.11). This result relies on BaSL's *partially affine* structure (Proposition 4.9) and a lemma in disintegration theory (Lemma 4.10).

With the above theorems and guarantees, BaSL admits a standard resource-theoretic semantics via a construction known as a *Kripke resource model* [Galmiche et al. 2005], which we formally develop in the rest of this section.

4.3 The Kripke Resource Model of BaSL

Recall from §2 that our sample space was set to the booleans $\{\mathsf{T}, \mathsf{F}\}$. However, in BaSL (as in Lilac) we fix the Hilbert cube $(\Omega, \Sigma_\Omega) := ([0,1]^{\mathbb{N}}, \mathcal{B}[0,1]^{\mathbb{N}})$ as our underlying sample space, which can be thought of as the random number generator having the ability to independently generate a stream of numbers between 0 and 1. We next define the Kripke resources monoid of BaSL, which intuitively comprises the unnormalised random number generators described in §4.2.

Definition 4.1. Let $\mathcal{F}$ be a σ-algebra of Ω and $\mu : \mathcal{F} \to [0, \infty]$ a measure. The pair $(\mathcal{F}, \mu)$ is a *random generator* if the following conditions hold:

(1) μ is a σ-finite measure: there exists a countable sequence $\{U_i \in \mathcal{F}\}_{i\in\mathbb{N}}$ such that $\{U_i\}_{i\in\mathbb{N}}$ covers Ω (i.e. $\bigcup_{i\in\mathbb{N}} U_i = \Omega$) and $\mu(U_i) < \infty$, for all $i \in \mathbb{N}$.
(2) μ has non-zero total measure: $\mu(\Omega) > 0$.
(3) $\mathcal{F}$ is a sub-σ-algebra of Σ_Ω: $\mathcal{F} \subseteq \Sigma_\Omega$.
(4) $\mathcal{F}$ is countably generated: there exists a countable set of subsets $\{F_i \subseteq \Omega\}_{i\in\mathbb{N}}$ such that $\mathcal{F}$ is the least σ-algebra containing $\{F_i\}_{i\in\mathbb{N}}$.
(5) $\mathcal{F}$ has a *finite footprint*, i.e. there exists a bound $n \in \mathbb{N}$ such that: $\forall F \in \mathcal{F}.\ \exists F' \subseteq [0,1]^n.\ F = F' \times \Omega$

We write $\mathcal{M}$ for the set of random generators of the Hilbert cube (Ω, Σ_Ω). Condition (1) describes the space of interest – we are not only interested in finite measures, but also measures with infinite normalising constants so that we can model *improper priors*, as demonstrated in §3.5. Condition (2) is needed so that $\mathcal{M}$ forms a PCM, and it affects the way we interpret Hoare triples (§4.3). Conditions (3) and (4) are needed so that we can consider *disintegrations* of μ with respect to its random variables (Theorem 4.7). Condition (5) is needed because we want to ensure there is enough space to generate new random variables (as in Lilac [Li et al. 2023, §2.5]).

Definition 4.2 ([Li et al. 2023, Definition 2.2]). Let $(\mathcal{F}, \mu), (\mathcal{G}, \nu) \in \mathcal{M}$. $(\mathcal{H}, \rho)$ is the *independent combination* of $(\mathcal{F}, \mu)$ and $(\mathcal{G}, \nu)$ if $\mathcal{H}$ is the smallest σ-algebra containing $\mathcal{F}$ and $\mathcal{G}$, and for all $F \in \mathcal{F}, G \in \mathcal{G}$: $\rho(F \cap G) = \mu(F) \cdot \nu(G)$.

Theorem 4.3 (PCM). *Let m be the independent combination of $m_1, m_2 \in \mathcal{M}$. Then $m \in \mathcal{M}$ and it is unique. Let $(\bullet) : \mathcal{M} \times \mathcal{M} \rightharpoonup \mathcal{M}$, mapping $m_1, m_2 \in \mathcal{M}$ to their independent combination if it exists, and $\mathbf{1}$ for the trivial probability space over Ω. Then $(\mathcal{M}, \bullet, \mathbf{1})$ is a partial commutative monoid.*

Proof notes. The proof of uniqueness, similar to Lilac, relies on the uniqueness of measures theorem, but for σ-finite measures instead. Identity and commutativity of $(\bullet)$ follow from properties of σ-algebras. For associativity, suppose $m_1, m_2, m_3 \in \mathcal{M}$ and $m_{(12)3}$ is defined; we define $m_{23} = (\mathcal{F}_{23}, \mu_{23})$ by choosing a set $V \in \mathcal{F}_1$ satisfying $0 < \mu_1(V) < \infty$ and define $\mu_{23}(U) := \frac{\mu_{(12)3}(V \cap U)}{\mu_1(V)}$. We then construct a λ-system $\Lambda \subseteq \mathcal{F}_{23}$ of $\mathcal{F}_{23}$-measurable sets such that $F_{23} \in \Lambda$ satisfies the property $\mu_{(12)3}(F_1 \cap F_{23}) = \mu_1(F_1)\mu_{23}(F_{23})$ whenever $F_1 \in \mathcal{F}_1$, $\mu_1(F_1) < \infty$ and $\mu_{23}(F_{23}) < \infty$. Finally, we apply the π-λ theorem and show that all measurable sets in $\mathcal{F}_{23}$ satisfy the property and therefore establish associativity (see Theorem C.7 in the technical appendix for the full proof).

In a logic of bunched implications (which includes separation logics), a *Kripke resource monoid* (KRM) is the basis for providing a satisfiability relation by defining a *Kripke resource model* [Galmiche et al. 2005]. We now show that our PCM can be extended to a KRM.

Definition 4.4 ([Barthe et al. 2019, Definition 1]). A *(partial) Kripke resource monoid* is a tuple $(\mathcal{M}, \bullet, \mathbf{1}, \sqsubseteq)$ such that $(\mathcal{M}, \bullet, \mathbf{1})$ is a partial commutative monoid and $(\sqsubseteq) \subseteq \mathcal{M} \times \mathcal{M}$ is a preorder such that $(\bullet)$ is bifunctorial over $(\sqsubseteq)$; i.e. for all m, n, m', n', if $m \sqsubseteq n$, $m' \sqsubseteq n'$ and $m \bullet m'$, $n \bullet n'$ are defined, then $m \bullet m' \sqsubseteq n \bullet n'$.

Corollary 4.5 (KRM). *Define a preorder $(\sqsubseteq) \subseteq \mathcal{M} \times \mathcal{M}$ by $(\mathcal{F}_1, \mu_1) \sqsubseteq (\mathcal{F}_2, \mu_2)$ if $\mathcal{F}_1 \subseteq \mathcal{F}_2$ and $\mu_2|_{\mathcal{F}_1} = \mu_1$ (Li et al. [2023, Theorem 2.4]). Then $(\mathcal{M}, \bullet, \mathbf{1}, \sqsubseteq)$ is a Kripke resource monoid.*

A desirable property of $\mathcal{M}$ is that it is closed under conditioning of random variables. That is, suppose $(\mathcal{F}, \mu) \in \mathcal{M}$, $X : \Omega \to \mathcal{A}$ is a measurable function and π is a distribution of X generated by mutating the likelihood of its current distribution; then the space $(\mathcal{F}, \mu_x^+|_{\mathcal{F}})$ conditioning on $X = x$ (almost) always exists (Theorem 4.7). To prove this, we apply the following Rokhlin-Simmons disintegration theorem, which is a variant of the disintegration theorem that holds for σ-finite measures in standard Borel spaces.

Lemma 4.6 ([Simmons 2012, Theorem 6.3]). *Let $\mu : \Sigma_\Omega \to [0,\infty]$ be a σ-finite Borel measure, $X : \Omega \to \mathcal{A}$ a $(\Sigma_\Omega, \Sigma_A)$-measurable function and $\pi : \Sigma_A \to [0,\infty]$ a σ-finite measure that dominates $X_*\mu$. Then there exists a π-almost-surely-unique (X,π)-disintegration $\{\mu_x\}_{x\in A}$.*

Theorem 4.7 (conditional). *Let $(\mathcal{F}, \mu) \in \mathcal{M}$, $X : \Omega \to \mathcal{A}$ be a $(\Sigma_\Omega, \Sigma_A)$-measurable function and $\pi : \Sigma_A \to [0,\infty]$ a σ-finite measure that dominates $X_*\mu$. Then there exists a measure $\mu^+ : \Sigma_\Omega \to [0,\infty]$ satisfying $\mu^+|_\mathcal{F} = \mu$. Further, the (X,π)-disintegration $\{\mu^+_x\}_{x\in A}$ of μ^+ exists and $(\mathcal{F}, \mu^+_x|_\mathcal{F}) \in \mathcal{M}$ for π-almost-all $x \in A$.*

Proof notes. Readers familiar with disintegration will note that this is a disintegration theorem in disguise. However, there are several non-trivialities. For existence of a Borel measure, since $\mu : \mathcal{F} \to [0,\infty]$ is assumed to be a countably-generated σ-finite measure on a sub-Borel σ-algebra $\mathcal{F}$, the extension μ^+ exists following from the result of Fremlin [2011, Proposition 433K]. We then apply the Rokhlin-Simmons disintegration theorem (Lemma 4.6) and obtain the conditional measure $\{\mu^+_x\}_{x\in A}$ and restrict them to the σ-algebra $\mathcal{F}$. See Proposition A.4 for the full proof.

4.4 Semantics of BaSL Assertions

With the Kripke resource monoid defined, we now formulate a *satisfiability relation* for BaSL assertions. In particular, we give semantics to *well-typed assertions*, where an assertion P is well-typed under context $\Gamma;\Delta$ if $\Gamma;\Delta \vdash P$, as defined in Fig. 22. For instance, the assertion $P := (X \sim \mathsf{Unif}(0,a) \Rightarrow \mathbb{E}[X] = a/2)$ is well-formed under the context $\Gamma := \Gamma', a : \mathbb{R}$ and $\Delta := \Delta', X : (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Notice that the map $(\gamma, a) \mapsto \mathsf{Unif}(0,a)$ is a function $[\![\Gamma]\!] \to \mathcal{G}(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, which means we have $\Gamma \vdash_{\mathsf{meas}} \mathsf{Unif}(0,a) : (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by T-ProbMeas. Then, notice that X is a name of Δ, which means $\Gamma;\Delta \vdash X \sim \mathsf{Unif}(0,1)$ by T-Dist. Similarly, $\Gamma;\Delta \vdash \mathbb{E}[X] = x$ by T-Expectation. Combining both assertions with T-BinModality yields the correct conclusion $\Gamma;\Delta \vdash P$.

We define the semantics of BaSL assertions through the satisfiability relation in Fig. 23. We write $(\gamma, D, m) \vDash P$ to denote that P holds in state m under the deterministic context γ and the list of random variables D. The semantics of propositional, first-order and separation logic connectives and quantifiers ($\wedge$, $\vee$, $\Rightarrow$, $*$, $-\!\!*$, $\forall$, $\exists$) is standard. In order for BaSL to be sound, the 'custom' probabilistic propositions, $\mathsf{own}\,E$, $\mathsf{L}(e)$, $\mathbb{E}[E] = e$, $x \overset{\pi}{\leftarrow} E \mid P$ and $\{P\}\,M\,\{X.Q\}$, must satisfy *Kripke monotonicity* as follows.

Proposition 4.8 (Kripke monotonicity). *Let $m \sqsubseteq m'$, $\Gamma;\Delta \vdash P$, $\gamma \in [\![\Gamma]\!]$ and $D \in \mathsf{RV}[\![\Delta]\!]$. Then $(\gamma, D, m) \vDash P$ implies $(\gamma, D, m') \vDash P$.*

We next explain the intuitive meaning of the custom probabilistic assertions in BaSL. The *ownership* $\mathsf{own}\,E$ and *distribution* $E \sim \pi$ assertions are expressed via measurability of random variables. Specifically, $(\gamma, D, m) \vDash \mathsf{own}\,E$ holds iff $E(\gamma) \circ D$ (the expression E applied to (γ, D)) is an $\mathcal{F}$-measurable function. Similarly, $(\gamma, D, m) \vDash E \sim \pi$ holds iff $E(\gamma) \circ D$ is an $\mathcal{F}$-measurable function and the pushforward of E with respect to μ is π ($\pi(\gamma) = (E(\gamma) \circ D)_*\mu$). The semantics of the expected value assertion $\mathbb{E}[E] = e$ follows its usual interpretation in statistics: a random expression has expected value e_γ if $E_{\gamma,D}$ integrates (with respect to the probability measure μ) to e_γ.

The remaining three assertions, namely the *likelihood proposition* $\mathsf{L}(e)$, the *conditioning modality* $x \overset{\pi}{\leftarrow} E \mid P$ and the *Hoare triple* $\{P\}\,M\,\{X.Q\}$ have non-trivial semantics, as we describe below.

Likelihood $\mathsf{L}(e)$. Recall from §2 that $\mathsf{L}(e)$ asserts that the current state has *likelihood* e. Indeed, a state $(\mathcal{F}, \mu) \in \mathcal{M}$ is more *likely* if the total measure $\mu(\Omega)$ is higher. Hence, $(\mathcal{F}, \mu)$ has likelihood $e(\gamma)$ when $\mu(\Omega) = e(\gamma)$. In fact, the proposition NormConst (mentioned in §2) is simply defined as having a non-zero likelihood k:

$$\mathsf{NormConst} := \exists k : (0,\infty).\mathsf{L}(k).$$

Deterministic context: $\Gamma = [x_1 : A_1, \dots, x_n : A_n]$
Probabilistic context: $\Delta = [X_1 : \mathcal{A}_1, \dots, X_n : \mathcal{A}_n]$

$$\textsc{T-RandExpr}\ \frac{E \in [\![\Gamma]\!] \to \mathbf{Meas}([\![\Delta]\!], \mathcal{A})}{\Gamma; \Delta \vdash_{\mathsf{re}} E : \mathcal{A}} \qquad \textsc{T-DetExpr}\ \frac{e \in [\![\Gamma]\!] \to A}{\Gamma \vdash_{\mathsf{de}} e : A} \qquad \textsc{T-Meas}\ \frac{\pi \in [\![\Gamma]\!] \to M_\sigma(\mathcal{A})}{\Gamma \vdash_{\mathsf{meas}} \pi : \mathcal{A}} \qquad \textsc{T-True}\ \frac{}{\Gamma; \Delta \vdash \top} \qquad \textsc{T-False}\ \frac{}{\Gamma; \Delta \vdash \bot}$$

$$\textsc{T-BinModality}\ \frac{\Gamma; \Delta \vdash P \qquad \Gamma; \Delta \vdash Q}{\Gamma; \Delta \vdash P \odot Q} \qquad \textsc{T-DetQuant}\ \frac{\Gamma, x : A; \Delta \vdash P}{\Gamma; \Delta \vdash \mathcal{Q} x : A.P} \qquad \textsc{T-RandQuant}\ \frac{\Gamma; \Delta, X : \mathcal{A} \vdash P}{\Gamma; \Delta \vdash \mathcal{Q}_{\mathsf{rv}} X : \mathcal{A}.P} \qquad \textsc{T-Expectation}\ \frac{\Gamma; \Delta \vdash_{\mathsf{re}} E : (\mathbb{R}, \mathcal{B}\mathbb{R}) \qquad \Gamma; \Delta \vdash_{\mathsf{de}} e : \mathbb{R}}{\Gamma; \Delta \vdash \mathbb{E}[E] = e}$$

$$\textsc{T-Dist}\ \frac{\Gamma; \Delta \vdash_{\mathsf{re}} E : \mathcal{A} \qquad \Gamma \vdash_{\mathsf{meas}} \pi : \mathcal{A}}{\Gamma; \Delta \vdash E \sim \pi} \qquad \textsc{T-Ownership}\ \frac{\Gamma; \Delta \vdash_{\mathsf{re}} E : \mathcal{A}}{\Gamma; \Delta \vdash \mathsf{own}\, E} \qquad \textsc{T-Conditioning}\ \frac{\Gamma; \Delta \vdash_{\mathsf{re}} E : \mathcal{A} \qquad \Gamma \vdash_{\mathsf{meas}} \pi : \mathcal{A} \qquad \Gamma, x : A; \Delta \vdash P}{\Gamma; \Delta \vdash x \overset{\pi}{\leftarrow} E \mid P}$$

$$\textsc{T-Likelihood}\ \frac{\Gamma \vdash_{\mathsf{de}} e : [0, \infty)}{\Gamma; \Delta \vdash \mathsf{L}(e)} \qquad \textsc{T-Hoare}\ \frac{\Gamma; \Delta \vdash P \qquad \Gamma; \Delta \vdash_{\mathsf{prog}} M : \mathcal{A} \qquad \Gamma; \Delta, X : \mathcal{A} \vdash Q}{\Gamma; \Delta \vdash \{P\}\, M\, \{X : \mathcal{A}.Q\}}$$

where A ranges over sets, $\mathcal{A}$ ranges over standard Borel spaces, $\odot \in \{\wedge, \vee, \Rightarrow, *, -\!\!*\}$, $\mathcal{Q} \in \{\forall, \exists\}$ and $M_\sigma(\mathcal{A})$ is the set of σ-finite measures over $\mathcal{A}$.

Fig. 22. Typing judgements for BASL assertions

Note that $\mathsf{L}(1)$ constitutes the multiplicative unit (unit of $*$) in BASL. Specifically, let $(\gamma, D, (\mathcal{F}, \mu)) \vDash X \sim \mathbb{P}$ for a probability measure $\mathbb{P}$; then $(\mathcal{F}, \mu)$ satisfies $\mathsf{L}(1)$ as the probability measure is by definition normalised. Recall that in a Kripke resource model, the multiplicative unit is a proposition I satisfying the following for any KRM $(\mathcal{M}, \bullet, \mathbf{1}, \sqsubseteq)$ [Galmiche et al. 2005, Definition 2.5]:

$$\text{for all } m \in \mathcal{M}, m \vDash I \iff e \sqsubseteq m$$

Unfolding $\sqsubseteq$ in our KRM, we know that if I is the multiplicative unit and $(\gamma, D, (\mathcal{F}, \mu)) \vDash I$, then $\mu(\Omega) = 1$. Hence, $\mathsf{L}(1)$ is the unit and we define $\top_1 := \mathsf{L}(1)$. Recall that a proposition P *entails* Q (written $P \vdash Q$) if $(\gamma, D, m) \vDash P$ implies $(\gamma, D, m) \vDash Q$, for all (γ, D, m). With $\top_1$ being the multiplicative unit, we then know the bi-entailment $P * \top_1 \dashv\vdash P$ holds.

BASL is partially affine. The fact that $\top_1$ is the multiplicative unit has significant implications on BASL's reasoning rules. In particular, this means we cannot *forget* about 'unnormalised propositions'. Recall that a separation logic is *affine* if the entailment $P * Q \vdash P \wedge Q$ holds [Galmiche et al. 2005]. For example, the IRIS separation logic is affine [Jung et al. 2018], and we can e.g. forget about pointers via the entailments $(x \hookrightarrow v * y \hookrightarrow v') \vdash (x \hookrightarrow v \wedge y \hookrightarrow v') \vdash x \hookrightarrow v$, i.e. given two addresses x and y, weakening $*$ to $\wedge$ allows us to forget information about y. On the other hand, a separation logic is *linear* (or *boolean*) when the multiplicative unit is only satisfied by the empty element $\mathbf{1} \in \mathcal{M}$, i.e. when $m \vDash I$ implies $m = \mathbf{1}$ [Galmiche et al. 2005]. Consequently, the entailment $P * Q \vdash P \wedge Q$ does not always hold, and we cannot forget information. In fact, this is the original approach to separation logic taken by O'Hearn and Pym [1999].

BASL is neither affine nor linear – the entailment $P * Q \vdash P \wedge Q$ holds in certain restricted cases. Indeed, BASL is *partially affine*, a category of logics first described by Charguéraud [2020], whereby only a class of (not necessarily all) assertions are designated as *affine*, namely those that can be

$(\gamma, D, m) \vDash \top$	*always*	
$(\gamma, D, m) \vDash \bot$	*never*	
$(\gamma, D, m) \vDash P \wedge Q$	*iff*	$(\gamma, D, m) \vDash P$ and $(\gamma, D, m) \vDash Q$
$(\gamma, D, m) \vDash P \vee Q$	*iff*	$(\gamma, D, m) \vDash P$ or $(\gamma, D, m) \vDash Q$
$(\gamma, D, m) \vDash P \Rightarrow Q$	*iff*	for all $m' \sqsupseteq m$, $(\gamma, D, m) \vDash P$ implies $(\gamma, D, m') \vDash Q$
$(\gamma, D, m) \vDash P * Q$	*iff*	there exists $m_1 \bullet m_2 \sqsubseteq m$ such that $(\gamma, D, m_1) \vDash P$, $(\gamma, D, m_2) \vDash Q$
$(\gamma, D, m) \vDash P \mathrel{-\!\!*} Q$	*iff*	if $m' \bullet m$ is defined then $(\gamma, D, m') \vDash P$ implies $(\gamma, D, m' \bullet m) \vDash Q$
$(\gamma, D, m) \vDash \forall x : A.P$	*iff*	for all $x \in A$, $((\gamma, x), D, m) \vDash P$
$(\gamma, D, m) \vDash \exists x : A.P$	*iff*	for some $x \in A$, $((\gamma, x), D, m) \vDash P$
$(\gamma, D, m) \vDash \forall_{\mathrm{rv}} X : \mathcal{A}.P$	*iff*	for all $X \in \mathsf{RV}(\mathcal{A})$, $(\gamma, (D, X), m) \vDash P$
$(\gamma, D, m) \vDash \exists_{\mathrm{rv}} X : \mathcal{A}.P$	*iff*	for some $X \in \mathsf{RV}(\mathcal{A})$, $(\gamma, (D, X), m) \vDash P$
$(\gamma, D, m) \vDash \mathsf{own}\, E$	*iff*	$E(\gamma) \circ D$ is $\mathcal{F}$-measurable
$(\gamma, D, m) \vDash E \sim \pi$	*iff*	$E(\gamma) \circ D$ is $\mathcal{F}$-measurable and $\pi(\gamma) = (E(\gamma) \circ D)_* \mu$
$(\gamma, D, m) \vDash \mathbb{E}[E] = e$	*iff*	$(\gamma, D, (\mathcal{F}, \mu)) \vDash \mathsf{L}(1)$ and $\int_\Omega E(\gamma) \circ D \,\mathrm{d}\mu = e(\gamma)$
$(\gamma, D, m) \vDash \mathsf{L}(e)$	*iff*	$\mu(\Omega) = e(\gamma)$
$(\gamma, D, (\mathcal{F}, \mu)) \vDash x \overset{\pi}{\leftarrow} E \mid P$	*iff*	$(\gamma, D, (\mathcal{F}, \mu)) \vDash \mathsf{own}\, E$ and $X_* \mu$ is absolutely continuous with respect to π and for all measures $\mu^+ : \Sigma_\Omega \to [0, \infty]$ satisfying $\mu^+\vert_{\mathcal{F}} = \mu$ and all disintegrations $\{\mu^+_x\}_{x \in \mathcal{A}}$ of μ^+ along (X, π) and for π-almost-all $x \in A$, $((\gamma, x), D, (\mathcal{F}, \mu^+_x\vert_{\mathcal{F}})) \vDash P$ where $X \coloneqq E(\gamma) \circ D$
$(\gamma, D, m) \vDash \{P\}\, M\, \{X : \mathcal{A}.Q\}$	*iff*	for all $m_{\mathrm{pre}} \in \mathcal{M}$ with $(\gamma, D, m_{\mathrm{pre}}) \vDash P$, $m_{\mathrm{fr}} \in \mathcal{M}$ with $(\mathcal{F}_0, \mu_0) \coloneqq m_{\mathrm{pre}} \bullet m_{\mathrm{fr}}$ defined, measures $\mu^+_0 : \Sigma_\Omega \to [0, \infty]$ with $\mu^+_0\vert_{\mathcal{F}_0} = \mu_0$, if $\mu^+_0(\{\omega \in \Omega \mid [\![M_\gamma]\!](D(\omega), \mathcal{A})\}) > 0$, then there exists (1) $X \in \mathsf{RV}(\mathcal{A})$, (2) $m_{\mathrm{post}} \in \mathcal{M}$ with $(\mathcal{F}_1, \mu_1) \coloneqq m_{\mathrm{post}} \bullet m_{\mathrm{fr}}$ defined, and (3) measure $\mu^+_1 : \Sigma_\Omega \to [0, \infty]$ with $\mu^+_1\vert_{\mathcal{F}_1} = \mu_1$ s.t. $(\gamma, (D, X), m_{\mathrm{post}}) \vDash Q$, and for all $f : \Omega \to \mathcal{B}$ and $U \in \Sigma_{\mathcal{B}} \otimes \Sigma_{\mathcal{A}}$, $\int_\Omega [\![M_\gamma]\!](D(\omega), \{x \in A \mid (f(\omega), x) \in U\})\, \mu^+_0(\mathrm{d}\omega) = \mu^+_1(\{\omega \in \Omega \mid (f(\omega), X(\omega)) \in U\})$

Fig. 23. Semantics of BaSL assertions

'dropped'. In BaSL, we define an assertion to be affine, written $\mathsf{affine}(P)$, as follows:

$$\mathsf{affine}(P) \overset{\text{def}}{\iff} \text{for all } \gamma, D, \mathcal{F}, \mu, \text{ if } (\gamma, D, (\mathcal{F}, \mu)) \vDash P, \text{ then } \mu(\Omega) = 1$$

Intuitively, this means only normalised assertions (e.g. $X \sim \mathbb{P}$ for a probability measure $\mathbb{P}$) are affine and assertions with unnormalised components (e.g. $\mathsf{L}(e)$ when $e \neq 1$, or $\mathsf{NormConst}$) are not affine. Affine assertions enjoy the property that they can be dropped, as stated in Proposition 4.9 below.

Proposition 4.9 (affine assertions). *The following entailments hold:*

$$\textit{E-}{*}\textit{-Weak}_1\ \frac{\mathsf{affine}(Q)}{P * Q \vdash P} \qquad \textit{E-}{*}\textit{-Weak}_2\ \frac{\mathsf{affine}(P)}{P * Q \vdash Q} \qquad \textit{E-}{*}\textit{-Weak}\ \frac{\mathsf{affine}(P) \quad \mathsf{affine}(Q)}{P * Q \vdash P \wedge Q}$$

Conditioning $x \overset{\pi}{\leftarrow} E \mid P$. As explained in §2, the conditioning assertion $x \overset{\pi}{\leftarrow} E \mid P$ lets us use the assertion P to describe the behaviour of a state μ assuming $E = x$ and $(\gamma, D, \mathcal{F}, \mu) \vDash E \sim \pi$. Its semantics is as follows. To show that P holds conditionally for almost every x, we require that $(\gamma, D, \mathcal{F}, \mu^+_x) \vDash P$ holds for almost all x, where μ^+ is a Borel measure extending μ and μ^+_x is the measure μ^+ conditioned on $E = x$. The Borel measure extension condition is a technical condition to ensure the existence of the conditioned space $\{\mu^+_x\}_{x\in A}$, which are a family of σ-finite measures that exist by Theorem 4.7. We can now see how the conditioning modality interacts with the likelihood assertion $\mathsf{L}(e)$ by revisiting BayesCoin. Suppose $X \sim \mathsf{Unif}(0,1)$ and we assert the likelihood of $X = x$ to be x as in the example of Fig. 4; then the state satisfies the proposition $x \xleftarrow{\mathsf{Unif}(0,1)} X \mid \mathsf{L}(x)$. The resulting distribution is proportional to the Beta(2, 1) distribution because Bayes' theorem states that:

$$\Pr[X \in U \mid \mathit{Flip}_1 = 1] = \int_0^1 \mathbf{1}_U(x) \cdot 2x \, \mathrm{d}x = \mathsf{Beta}(U \mid 2, 1).$$

Now, in order to derive the desired postcondition ($X \sim$ Beta(2, 1) with a normalising constant), we need to internalise a notion of Bayes' theorem within BaSL.

The Internal Bayes' Theorem. In disintegration theory, the following theorem states that the Radon-Nikodym derivative of two measures $X_*\mu \ll \pi$ is almost surely equal to the total measure of a π-disintegration (Lemma 4.10). Using this result, we can internalise Bayes' theorem as a BaSL bi-entailment (Theorem 4.11).

Lemma 4.10 ([Chang and Pollard 1997, Theorem 2]). *Let $\{\mu_x\}_{x\in A}$ be an (X,π)-disintegration of μ. Then π dominates $X_*\mu$ and for π-almost-all $x \in A$, $\frac{\mathrm{d}X_*\mu}{\mathrm{d}\pi}(x) = \mu_x(\Omega)$.*

Theorem 4.11 (internal Bayes' theorem). *For any Borel measurable function $f : \mathcal{A} \to [0,\infty)$ and random expression $\Gamma; \Delta \vdash_{\mathsf{re}} E : \mathcal{A}$, the following bi-entailment holds:*

$$\overbrace{x \xleftarrow{\pi} E}^{\text{prior}} \mid \overbrace{\mathsf{L}(f(x))}^{\text{likelihood}} \dashv\vdash \overbrace{E \sim f \cdot \pi}^{\text{posterior}}.$$

Proof. Let $m = (\mathcal{F},\mu) \in \mathcal{M}$, (γ, D, m) be a configuration and $X(\omega) := E(\gamma)(D(\omega))$. To prove the entailment from left to right, we assume $(\gamma, D, (\mathcal{F},\mu)) \vDash x \xleftarrow{\pi} E \mid \mathsf{L}(f(x))$. Then for any Borel extension μ^+ and (X,π)-disintegration $\{\mu^+_x\}_{x\in A}$ we know $(\gamma, D, (\mathcal{F}, \mu^+_x|_{\mathcal{F}})) \vDash \mathsf{L}(f(x))$ holds for π-almost-all $x \in A$. This implies $\mu^+_x(\Omega) = f(x)$ holds for π-almost-all $x \in A$, which then implies that for all $F \in \Sigma_A$:

$$X_*\mu(F) = X_*\mu^+(F) = \int_A \mu^+_x(X^{-1}(F))\,\pi(\mathrm{d}x) = \int_F \mu^+_x(\Omega)\,\pi(\mathrm{d}x) = \int_F f\,\mathrm{d}\pi = (f\cdot\pi)(F).$$

The first equality follows from our assumption that μ^+ is a Borel extension of μ. The second equality is a disintegration axiom; the third holds because $\mathbf{1}_F(x)\mu^+_x(\Omega) = X_*\mu^+_x(F)$; and the last holds because for almost all x, we have $(\gamma, D, (\mathcal{F}, \mu^+_x|_{\mathcal{F}})) \vDash \mathsf{L}(f(x))$. Hence, we know $(\gamma, D, (\mathcal{F},\mu)) \vDash E \sim f\cdot\pi$.

To prove the entailment from right to left, we assume $(\gamma, D, (\mathcal{F},\mu)) \vDash E \sim f \cdot \pi$ holds; then $(\gamma, D, (\mathcal{F},\mu)) \vDash \mathsf{own}\ E$ by definition of $\vDash$. Next, for any Borel extension μ^+ and (X,π)-disintegration $\{\mu^+_x\}_{x\in A}$, we know $(\gamma, D, (\mathcal{F}, \mu^+_x|_{\mathcal{F}})) \vDash \mathsf{L}(f(x))$ holds for π-almost-all $x \in A$ because for any $F \in \Sigma_A$:

$$\int_F \mu^+_x(\Omega)\,\pi(\mathrm{d}x) = \int_F \frac{\mathrm{d}X_*\mu}{\mathrm{d}\pi}\,\mathrm{d}\pi = \int_F \frac{\mathrm{d}(f\cdot\pi)}{\mathrm{d}\pi}\,\mathrm{d}\pi = \int_F f\,\mathrm{d}\pi.$$

The first equality follows from Lemma 4.10; the second follows from our assumption. Note that $X_*\mu = f\cdot\pi$ is absolutely continuous with respect to π as $(f\cdot\pi)(F) = 0$ for any π-null-set F. ◻

Hoare Triple $\{P\}\ M\ \{X.Q\}$. The interpretation of a Hoare triple in BaSL is similar to that of Lilac, but with one key difference: a BaSL Hoare triple denotes *partial correctness*. A *partial correctness triple* $\{P\}\ M\ \{Q\}$ states that if a state satisfies P and executing M from that state *terminates*, then the resulting state satisfies Q. One reason of assuming termination is due to its undecidability. Similarly, determining whether a probabilistic program has a zero normalising constant is undecidable Staton [2020, §2.2.4] – this is not a relevant concern for existing probabilistic logics (including Lilac) as they do not support Bayesian updating. As such, BaSL triples denote partial correctness. Specifically, $\{P\}\ M\ \{X.Q\}$ holds iff starting from a state $m_{\text{pre}} := (\mathcal{F}_{\text{pre}}, \mu_{\text{pre}})$ satisfying the precondition P and an arbitrary frame m_{fr} (such that $m_{\text{pre}} \bullet m_{\text{fr}}$ is defined), if '$[\![M]\!]$ normalises to a non-zero constant from $m_{\text{pre}} \bullet m_{\text{fr}}$', then there must exist a random variable X and a postcondition state $m_{\text{post}} := (\mathcal{F}_{\text{post}}, \mu_{\text{post}})$ that satisfies Q and is compatible with the frame m_{fr} such that $m_{\text{post}} \bullet m_{\text{fr}}$ is also defined and 'executing M from m_{pre} results in m_{post}'.

To express that '$[\![M]\!]$ normalises to a non-zero constant from $m_{\text{pre}} \bullet m_{\text{fr}}$', we require that $\mu_0^+(\{\omega \in \Omega \mid [\![M_\gamma]\!](D(\omega), \mathcal{A})\}) > 0$, where μ_0^+ is a Borel extension of $m_{\text{pre}} \bullet m_{\text{fr}}$. This means there exists a measurable set of Ω with non-zero-μ_0^+ measure such that executing $[\![M_\gamma]\!]$ with random variables D generated from the seed ω leads to a non-zero normalising constant. Similar to Lilac, we quantify over arbitrary extensions of random variables D_{ext} to prove a substitution lemma (see Lemma D.4).

To express 'executing M from the state m_{pre} results in m_{post}', we require that the following hold for any measurable set U:

$$\int_\Omega [\![M_\gamma]\!](D(\omega), \{x \mid (D_{\text{ext}}(\omega), D(\omega), x) \in U\})\, \mu_0^+(\mathrm{d}\omega) = \mu_1^+\{\omega \mid (D_{\text{ext}}(\omega), D(\omega), X(\omega)) \in U\}.$$

where μ_0^+ is a Borel measure that extends $m_{\text{pre}} \bullet m_{\text{fr}}$, and μ_1^+ is a Borel measure that extends $\mu_{\text{post}} \bullet m_{\text{fr}}$. Comparing the BaSL semantics to the standard denotational semantics of Bppl in the category of s-finite kernels, the behaviour of the Hoare triple can be characterised by Proposition 4.12. Intuitively, suppose μ_0^+ is a state that satisfies precondition P and μ_1^+ is a state that satisfies the postcondition Q. Then the resulting measure of $[\![M_\gamma]\!]$ with random variables D along the random source μ_0^+ is equal to the 'output' random variable X with the random source μ_1^+, as stated below.

Proposition 4.12. *Let $(\gamma, D, \mu) \vDash P$, $\mathcal{M}\mathcal{A}$ be the space of measures of $\mathcal{A}$, and $[\![-]\!] : \mathbf{Syn} \to \mathbf{sfKrn}$ be the semantic functor defined in appendix B. Suppose $\Delta \vdash M : \tau$ and $\{P\}\, M\{X.Q\}$ holds with respect to μ_0^+ and μ_1^+. Then the following diagram commutes:*

$$\begin{array}{ccccc}
\mathcal{M}\Omega & \xrightarrow{D_*} & \mathcal{M}[\![\Delta]\!] & \xrightarrow{[\![M_\gamma]\!]_*} & \mathcal{M}[\![\tau]\!] \\
{\scriptstyle \mu_0^+}\uparrow & & & & \| \\
\mathbf{1} & \xrightarrow{\mu_1^+} & \mathcal{M}\Omega & \xrightarrow{X_*} & \mathcal{M}[\![\tau]\!]
\end{array}$$

We finally show that the BaSL proof system in Fig. 9 is sound, with the full proof given in appendix E. We write $\vDash \{P\}\, M\{X.Q\}$ to denote that $(\gamma, D, m) \vDash \{P\}\, M\{X.Q\}$ holds for all (γ, D, m).

Theorem 4.13 (soundness). *The BaSL proof system is* sound*: for all P, M, X, Q, if $\vdash \{P\}\, M\{X.Q\}$ is derivable using the rules in Fig. 9, then $\vDash \{P\}\, M\{X.Q\}$ holds.*

5 CONCLUSIONS, RELATED AND FUTURE WORK

We developed BaSL by extending probabilistic separation logic to verify Bayesian programs/statistical models. To this end, we devised a semantic model rich enough to encode probabilistic programming concepts such as conditional distributions, unnormalised distributions, Bayesian updating and improper priors. We then demonstrated the utility of BaSL by proving properties such as correlation, expected values and posterior distributions in various statistical models.

Related Work: Semantics of BPPLs. The semantics of randomised languages is well-established: starting from the seminal work of Kozen [1981] on linear operator semantics, there are numerous works on semantic domains for probability and non-determinism, e.g. by Jones and Plotkin [1989]. These led to the study of Bayesian inference from a programming language theory perspective [Gordon et al. 2014; van de Meent et al. 2021] with research on their operational [Borgström et al. 2016] and denotational semantics [Dahlqvist and Kozen 2019; Huot et al. 2023]. Based on the theory

	PSL	DIBI	LINA	LILAC	PSOL	BLUEBELL	BASL
Discrete distribution	✓	✓	✓	✓	✓	✓	✓
Probabilistic independence	✓	✓	✓	✓	✓	✓	✓
Continuous distribution				✓			✓
Negative dependence			✓				
Conditional independence		✓		✓	✓	✓	✓
Conditioning modality				✓	✓	✓	✓
Concurrency					✓		
Relational reasoning						✓	
Bayesian reasoning							✓

Fig. 24. Features of probabilistic separation/bunched logics

of concrete sheaves [Matache et al. 2022], Heunen et al. [2017]; Vákár et al. [2019] developed a category known as (ω-)quasi-Borel space for reasoning about higher-order, recursive Bayesian probabilistic programs and proved the existence of a strong monad of s-finite measures for a monadic semantics of higher-order BPPLs. The BASL proof system is underpinned by the category of s-finite kernels developed by Staton [2017].

Related Work: Probabilistic Separation Logics. Separation logic (SL) developed a *modular* theory for reasoning about computational resources [O'Hearn and Pym 1999] for pointer-manipulating programs. This led to the development of abstract models for SL [Calcagno et al. 2007; Galmiche et al. 2005; Jung et al. 2018]. Barthe et al. gave a probabilistic interpretation of SL in PSL (probabilistic SL) and proved the correctness of algorithms such as the one-time pad cipher.

One of the overarching themes of probabilistic separation logics is the investigation of *conditional probability*. Bao et al. [2021a] developed DIBI based on bunched implications (BI) by extending PSL with the ';' connective for describing conditional independence of random variables, while Bao et al. [2021b] developed LINA for reasoning about negative dependence of random variables. Li et al. [2023] proposed LILAC, a measure-theoretic interpretation of probabilistic SL that supports desirable features such as continuous distributions and 'mathematical' random variables. LILAC handles conditional independence by introducing the *conditioning modality*, which was later adopted by BLUEBELL [Bao et al. 2025] (a relational probabilistic SL) and PSOL [Zilberstein et al. 2024] (a concurrent probabilistic program logic). Recently, Li et al. [2024] developed a categorical model of LILAC by drawing an analogy between random sampling and fresh name generations in the theory of nominal sets. BASL generalises the model of LILAC and brings a novel perspective of conditional reasoning by supporting Bayesian updating, hence giving an axiomatic semantics of probabilistic programming via SL. We summarise the features of probabilistic separation/BI logics in Fig. 24.

Future Work. We plan to consider three avenues of future research. First, we will mechanise BASL and its soundness proof in a theorem prover such as Rocq. Second, we aim to extend BASL to add support for more sophisticated language features including (1) mutable states; (2) recursion; and (3) higher-order functions. These extensions will allow us to express programs in a class of statistical models known as *Bayesian nonparametric models* [Mak et al. 2021; Orbanz and Teh 2017], namely statistical models with an unbounded number of random variables [van de Meent et al. 2021]. Finally, we will apply BASL to develop program simplification/symbolic execution tools for probabilistic programs.

ACKNOWLEDGMENTS

We would like to thank John Li for helpful discussions on the technical aspects of the Lilac separation logic. We also thank Alberto Croquevielle for his insights on the measure-theoretic components of the proofs. We thank the members of the Veritas Lab and the functional programming group at Imperial for their feedback and comments on the earlier draft of this work.

REFERENCES

Sheldon Axler. 2019. *Measure, Integration & Real Analysis.* Springer International Publishing.

Jialu Bao, Simon Docherty, Justin Hsu, and Alexandra Silva. 2021a. A Bunched Logic for Conditional Independence. In *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS).* 1–14. doi:10.1109/LICS52264.2021.9470712

Jialu Bao, Emanuele D'Osualdo, and Azadeh Farzan. 2025. Bluebell: An Alliance of Relational Lifting and Independence for Probabilistic Reasoning. *Proc. ACM Program. Lang.* 9, POPL, Article 58 (Jan. 2025), 31 pages. doi:10.1145/3704894

Jialu Bao, Marco Gaboardi, Justin Hsu, and Joseph Tassarotti. 2021b. A Separation Logic for Negative Dependence. *CoRR* abs/2111.14917 (2021). arXiv:2111.14917 https://arxiv.org/abs/2111.14917

Gilles Barthe, Justin Hsu, and Kevin Liao. 2019. A Probabilistic Separation Logic. *CoRR* abs/1907.10708 (2019). arXiv:1907.10708 http://arxiv.org/abs/1907.10708

Patrick Billingsley. 1995. *Probability and measure* (3. ed ed.). Wiley, New York [u.a.].

Johannes Borgström, Ugo Dal Lago, Andrew D. Gordon, and Marcin Szymczak. 2016. A lambda-calculus foundation for universal probabilistic programming. *SIGPLAN Not.* 51, 9 (Sept. 2016), 33–46. doi:10.1145/3022670.2951942

Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. 2011. *Handbook of Markov Chain Monte Carlo.* Chapman and Hall/CRC. doi:10.1201/b10905

Cristiano Calcagno, Peter W. O'Hearn, and Hongseok Yang. 2007. Local Action and Abstract Separation Logic. In *22nd Annual IEEE Symposium on Logic in Computer Science (LICS 2007).* 366–378. doi:10.1109/LICS.2007.30

Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. Stan: A Probabilistic Programming Language. *Journal of Statistical Software* 76, 1 (2017), 1–32. doi:10.18637/jss.v076.i01

Joseph T. Chang and David Pollard. 1997. Conditioning as disintegration. *Statistica Neerlandica* 51, 3 (1997), 287–317. doi:10.1111/1467-9574.00056

Arthur Charguéraud. 2020. Separation logic for sequential programs (functional pearl). *Proc. ACM Program. Lang.* 4, ICFP, Article 116 (Aug. 2020), 34 pages. doi:10.1145/3408998

Ulices Santa Cruz and Yasser Shoukry. 2022. NNLander-VeriF: A Neural Network Formal Verification Framework for Vision-Based Autonomous Aircraft Landing. arXiv:2203.15841 [cs.LG] https://arxiv.org/abs/2203.15841

Marco F. Cusumano-Towner, Feras A. Saad, Alexander K. Lew, and Vikash K. Mansinghka. 2019. Gen: a general-purpose probabilistic programming system with programmable inference. In *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation* (Phoenix, AZ, USA) *(PLDI 2019).* Association for Computing Machinery, New York, NY, USA, 221–236. doi:10.1145/3314221.3314642

Fredrik Dahlqvist and Dexter Kozen. 2019. Semantics of higher-order probabilistic programs with conditioning. *Proc. ACM Program. Lang.* 4, POPL, Article 57 (Dec. 2019), 29 pages. doi:10.1145/3371125

David H. Fremlin. 2011. *Measure Theory: Topological Measure Spaces. Volume 4.* Torres Fremlin.

Didier Galmiche, Daniel Méry, and David Pym. 2005. The semantics of BI and resource tableaux. *Mathematical Structures in Computer Science* 15, 6 (2005), 1033–1088. doi:10.1017/S0960129505004858

Stuart Geman and Donald Geman. 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6, 6 (1984), 721–741. doi:10.1109/TPAMI.1984.4767596

Andrew D. Gordon, Thomas A. Henzinger, Aditya V. Nori, and Sriram K. Rajamani. 2014. Probabilistic Programming. In *Proceedings of the on Future of Software Engineering.* ACM, 167–181. doi:10.1145/2593882.2593900

Chris Heunen, Ohad Kammar, Sam Staton, and Hongseok Yang. 2017. A convenient category for higher-order probability theory. In *Proceedings of the 32nd Annual ACM/IEEE Symposium on Logic in Computer Science* (Reykjavík, Iceland) *(LICS '17).* IEEE Press, Article 77, 12 pages.

Mathieu Huot, Alexander K. Lew, Vikash K. Mansinghka, and Sam Staton. 2023. ωPAP Spaces: Reasoning Denotationally About Higher-Order, Recursive Probabilistic and Differentiable Programs. In *2023 38th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS).* 1–14. doi:10.1109/LICS56636.2023.10175739

Claire Jones and Gordon Plotkin. 1989. A probabilistic powerdomain of evaluations. In *Proceedings of the Fourth Annual Symposium on Logic in Computer Science* (Pacific Grove, California, USA). IEEE Press, 186–195.

Ralf Jung, Robbert Krebbers, Jacques-Henri Jourdan, Aleš Bizjak, Lars Birkedal, and Derek Dreyer. 2018. Iris from the ground up: A modular foundation for higher-order concurrent separation logic. *Journal of Functional Programming* 28 (2018), e20. doi:10.1017/S0956796818000151

Dexter Kozen. 1981. Semantics of probabilistic programs. *J. Comput. System Sci.* 22, 3 (1981), 328–350. doi:10.1016/0022-0000(81)90036-2

Peter M. Lee. 2012. *Bayesian Statistics: An Introduction* (4th ed.). Wiley Publishing.

John M. Li, Amal Ahmed, and Steven Holtzen. 2023. Lilac: A Modal Separation Logic for Conditional Probability. *Proc. ACM Program. Lang.* 7, PLDI, Article 112 (jun 2023), 24 pages. doi:10.1145/3591226

John M. Li, Jon Aytac, Philip Johnson-Freyd, Amal Ahmed, and Steven Holtzen. 2024. A Nominal Approach to Probabilistic Separation Logic. In *Proceedings of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science* (Tallinn, Estonia) *(LICS '24)*. Association for Computing Machinery, New York, NY, USA, Article 55, 14 pages. doi:10.1145/3661814.3662135

Carol Mak, Fabian Zaiser, and Luke Ong. 2021. Nonparametric Hamiltonian Monte Carlo. In *Proceedings of the 38th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 139)*, Marina Meila and Tong Zhang (Eds.). PMLR, 7336–7347. https://proceedings.mlr.press/v139/mak21a.html

Cristina Matache, Sean Moss, and Sam Staton. 2022. Concrete categories and higher-order recursion: With applications including probability, differentiability, and full abstraction. In *Proceedings of the 37th Annual ACM/IEEE Symposium on Logic in Computer Science* (Haifa, Israel) *(LICS '22)*. Association for Computing Machinery, New York, NY, USA, Article 57, 14 pages. doi:10.1145/3531130.3533370

Richard McElreath. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and STAN* (2nd ed.). Chapman and Hall/CRC. doi:10.1201/9780429029608

Natalia Muehlemann, Tianjian Zhou, Rajat Mukherjee, Munshi Imran Hossain, Satrajit Roychoudhury, and Estelle Russek-Cohen. 2023. A Tutorial on Modern Bayesian Methods in Clinical Trials. *Therapeutic Innovation & Regulatory Science* 57, 3 (2023), 402–416. doi:10.1007/s43441-023-00515-3

Praveen Narayanan, Jacques Carette, Wren Romano, Chung-chieh Shan, and Robert Zinkov. 2016. Probabilistic inference by program transformation in Hakaru (system description). In *International Symposium on Functional and Logic Programming - 13th International Symposium, FLOPS 2016, Kochi, Japan, March 4-6, 2016, Proceedings*. Springer, 62–79. doi:10.1007/978-3-319-29604-3_5

Peter O'Hearn and David Pym. 1999. The Logic of Bunched Implications. *Bulletin of Symbolic Logic* 5, 2 (1999), 215–244. doi:10.2307/421090

Peter Orbanz and Yee Teh. 2017. *Bayesian Nonparametric Models*. 107–116. doi:10.1007/978-1-4899-7687-1_928

John C. Reynolds. 2002. Separation Logic: A Logic for Shared Mutable Data Structures. In *Proceedings of the 17th Annual IEEE Symposium on Logic in Computer Science (LICS '02)*. IEEE Computer Society, USA, 55–74.

David Simmons. 2012. Conditional measures and conditional expectation; Rohlin's Disintegration Theorem. *Discrete and Continuous Dynamical Systems* 32 (07 2012). doi:10.3934/dcds.2012.32.2565

Natalia Slusarz, Ekaterina Komendantskaya, Matthew L. Daggitt, and Robert Stewart. 2022. Differentiable Logics for Neural Network Training and Verification. arXiv:2207.06741 [cs.AI] https://arxiv.org/abs/2207.06741

Sam Staton. 2017. Commutative Semantics for Probabilistic Programming. In *Programming Languages and Systems*, Hongseok Yang (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 855–879.

Sam Staton. 2020. Probabilistic Programs as Measures. In *Foundations of Probabilistic Programming*, Gilles Barthe, Joost-Pieter Katoen, and Alexandra Silva (Eds.). Cambridge University Press, 43–74.

Joseph Tassarotti and Robert Harper. 2019. A separation logic for concurrent randomized programs. *Proc. ACM Program. Lang.* 3, POPL, Article 64 (Jan. 2019), 30 pages. doi:10.1145/3290377

David Tolpin, Jan-Willem van de Meent, Hongseok Yang, and Frank D. Wood. 2016. Design and Implementation of Probabilistic Programming Language Anglican. *CoRR* abs/1608.05263 (2016). arXiv:1608.05263 http://arxiv.org/abs/1608.05263

Matthijs Vákár, Ohad Kammar, and Sam Staton. 2019. A domain theory for statistical probabilistic programming. *Proc. ACM Program. Lang.* 3, POPL, Article 36 (Jan. 2019), 29 pages. doi:10.1145/3290349

Jan-Willem van de Meent, Brooks Paige, Hongseok Yang, and Frank Wood. 2021. An Introduction to Probabilistic Programming. arXiv:1809.10756 [stat.ML] https://arxiv.org/abs/1809.10756

Matthijs Vákár and Luke Ong. 2018. On S-Finite Measures and Kernels. arXiv:1810.01837 [math.PR] https://arxiv.org/abs/1810.01837

Noam Zilberstein, Alexandra Silva, and Joseph Tassarotti. 2024. Probabilistic Concurrent Reasoning in Outcome Logic: Independence, Conditioning, and Invariants. arXiv:2411.11662 [cs.LO] https://arxiv.org/abs/2411.11662

A DISINTEGRATION

Let Ω be a set and Σ_Ω be a set of subsets of Ω such that Σ_Ω is closed under countable intersection and complement. Then Σ_Ω is a *σ-algebra* and the pair (Ω, Σ_Ω) is a *measurable space*. A *measure* is a function $\mu : \Sigma_\Omega \longrightarrow [0, \infty]$ satisfying $\mu(\varnothing) = 0$ and $\mu(\biguplus_{i\in\mathbb{N}} U_i) = \sum_{i\in\mathbb{N}} \mu(U_i)$. We additionally call μ a *probability measure* if $\mu(\Omega) = 1$, or *σ-finite measure* if there is a measurable cover $\{U_i \in \mathcal{F}\}_{i\in\mathbb{N}}$ such that $\mu(U_i) < \infty$ for all $i \in \mathbb{N}$. A function $f : A \longrightarrow B$ is (Σ_A, Σ_B)-measurable if for all $U \in \Sigma_B$, $f^{-1}(U) \in \Sigma_A$. We denote **Meas** the category of measurable spaces with the class of measurable spaces as objects and measurable functions as morphisms.

Definition A.1. Let (A, τ) be a topological space. The *Borel σ-algebra* of (A, τ) is the set A equipped with the smallest σ-algebra $\mathcal{B}(A, \tau)$ that contains τ.

Since σ-algebras are closed under complements, notice all closed sets are by definition part of $\mathcal{B}(A, \tau)$. By convention, we write $\mathbb{R}$, $[0, \infty)$, $[0, \infty]$ to mean the sets endowed with their usual topology. When we say a function $f : \Omega \longrightarrow \mathbb{R}$ (or with $[0, \infty)/[0, \infty]$) is *Borel measurable* when f is $(\Sigma_A, \mathcal{B}(\mathbb{R}))$-measurable.

Definition A.2. Let $\mu : \Sigma_\Omega \longrightarrow [0, \infty]$ be a measure and $f : \Omega \longrightarrow [0, \infty]$ a Borel measurable function. Then the *(Lebesgue) integral* of f with respect to μ is defined by

$$\int_\Omega f \, \mathrm{d}\mu := \sup \left\{ \sum_{i=1}^{n} \inf_{\omega\in U_i} f(\omega) \cdot \mu(U_i) \;\middle|\; \{U_i \in \mathcal{F}\}_{i=1}^{n} \text{ disjoint cover of } \Omega \right\}.$$

By convention, we write $\int_\Omega f(x)\, \mu(\mathrm{d}x)$ as a shorthand for $\int_\Omega (x \mapsto f(x))\, \mathrm{d}\mu$.

Definition A.3. Let $\mu : \Sigma_\Omega \longrightarrow [0, \infty]$ be a measure, $X : \Omega \longrightarrow A$ a $(\Sigma_\Omega, \Sigma_A)$-measurable function and $\pi : \Sigma_A \longrightarrow [0, \infty]$ a measure. An *(X, π)-disintegration of μ* is a set of measures $\{\mu_x\}_{x\in A}$ satisfying the following axioms:

- (concentration) – for π-almost-all $x \in A$, $\mu_x\{\omega \in \Omega \mid X(\omega) \neq x\} = 0$.
- (measurability) – for all $(\Sigma_\Omega, \mathcal{B}[0, \infty])$-measurable function $f : \Omega \longrightarrow [0, \infty]$, the function
$$x \longmapsto \int_\Omega f \, \mathrm{d}\mu_x$$
is $(\Sigma_A, [0, \infty])$-measurable.
- (marginalisability) – for all $(\Sigma_\Omega, \mathcal{B}[0, \infty])$-measurable function $f : \Omega \longrightarrow [0, \infty]$, the following identity holds:
$$\int_\Omega f \, \mathrm{d}\mu = \int_A \int_\Omega f(\omega)\, \mathrm{d}\mu_x(\omega)\, \mathrm{d}\pi(x).$$

When μ_x is a probability measure π-almost-surely, we call $\{\mu_x\}$ the *regular conditional distribution* of μ with respect to X and π.

A measurable space is a *standard Borel space* if it is the Borel space of a Polish space, i.e. a topological space that is separable and completely metrisable. The *Rokhlin-Simmons disintegration theorem* states that when the domain and codomain of a random variable X is standard Borel, a disintegration with respect to a σ-finite measure must exist and is almost surely unique:

Proposition A.4 ([Simmons 2012, Theorem 6.3]). *Let Ω and $\mathcal{A}$ be standard Borel spaces, $\mu : \Sigma_\Omega \longrightarrow [0, \infty]$ be a σ-finite Borel measure, $X : \Omega \longrightarrow \mathcal{A}$ a $(\Sigma_\Omega, \Sigma_A)$-measurable function and $\pi : \Sigma_A \longrightarrow [0, \infty]$ a σ-finite measure that dominates $X_*\mu$. Then there exists a π-almost-surely-unique (X, π)-disintegration $\{\mu_x\}_{x\in A}$.*

Deterministic typing judgement: $\Delta \vdash_{\mathrm{d}} M : \tau$
Probabilistic typing judgement: $\Delta \vdash_{\mathrm{p}} M : \tau$

$$\text{PT-Unit}\ \frac{}{\Delta \vdash_{\mathrm{d}} () : 1} \qquad \text{PT-Var}\ \frac{}{\Delta \vdash_{\mathrm{d}} X : \tau}\,(X : \tau \in \Delta) \qquad \text{PT-Nat}\ \frac{}{\Delta \vdash_{\mathrm{d}} n : \mathbb{N}} \qquad \text{PT-Real}\ \frac{}{\Delta \vdash_{\mathrm{d}} r : \mathbb{R}}$$

$$\text{PT-Func}\ \frac{\Delta \vdash_{\mathrm{d}} M : \sigma}{\Delta \vdash_{\mathrm{d}} f(M) : \tau}\,(f : [\![\sigma]\!] \to [\![\tau]\!]) \qquad \text{PT-Pair}\ \frac{(\Delta \vdash_{\mathrm{d}} M_i : \tau_i)_{i=1,2}}{\Delta \vdash_{\mathrm{d}} (M_1, M_2) : \tau_1 \times \tau_2} \qquad \text{PT-Fst}\ \frac{\Delta \vdash_{\mathrm{d}} M : \sigma \times \tau}{\Delta \vdash_{\mathrm{d}} M.\mathbf{1} : \sigma} \qquad \text{PT-Snd}\ \frac{\Delta \vdash_{\mathrm{d}} M : \sigma \times \tau}{\Delta \vdash_{\mathrm{d}} M.\mathbf{2} : \tau}$$

$$\text{PT-True}\ \frac{}{\Delta \vdash_{\mathrm{d}} \mathbf{true} : \mathbb{B}} \qquad \text{PT-False}\ \frac{}{\Delta \vdash_{\mathrm{d}} \mathbf{false} : \mathbb{B}} \qquad \text{PT-If}\ \frac{\Delta \vdash_{\mathrm{d}} M : \mathbb{B} \qquad \Delta \vdash_{\mathrm{z}} N_1 : \tau \qquad \Delta \vdash_{\mathrm{z}} N_2 : \tau}{\Delta \vdash_{\mathrm{z}} \mathbf{if}\ M\ \mathbf{then}\ N_1\ \mathbf{else}\ N_2 : \tau}\,(\mathrm{z} \in \{\mathrm{p}, \mathrm{d}\})$$

$$\text{PT-Sample}\ \frac{\Delta \vdash_{\mathrm{d}} M : \mathbb{P}(\tau)}{\Delta \vdash_{\mathrm{p}} \mathbf{sample}(M) : \tau} \qquad \text{PT-Score}\ \frac{\Delta \vdash_{\mathrm{d}} : M : \mathbb{R}}{\Delta \vdash_{\mathrm{p}} \mathbf{score}(M) : 1} \qquad \text{PT-Dirac}\ \frac{\Delta \vdash_{\mathrm{d}} M : \tau}{\Delta \vdash_{\mathrm{p}} \mathbf{return}(M) : \tau}$$

$$\text{PT-Let}\ \frac{\Delta \vdash_{\mathrm{p}} M : \sigma \qquad \Delta, X : \sigma \vdash_{\mathrm{p}} N : \tau}{\Delta \vdash_{\mathrm{p}} \mathbf{let}\ X = M\ \mathbf{in}\ N : \tau}$$

Fig. 25. Typing judgements for Bppl

B SYNTAX AND SEMANTICS OF Bppl

As defined in Staton [2017], Bppl is a typed first-order language equipped with two effects: $\mathbf{sample}(\mathbb{P})$, which samples from a probability distribution $\mathbb{P}$, and the soft conditioning effect $\mathbf{score}(\ell)$, which 'scales' the current distribution according to a non-negative number ℓ. Let Δ be a context $[X_1 : \mathcal{A}_1, ..., X_n : \mathcal{A}_n]$ such that X_i is a name and $\mathcal{A}_i = (A_i, \Sigma_i)$ is a measurable space.

Definition B.1. The *syntax* of Bppl terms and types are defined by the following grammar:

$$\begin{aligned} \mathit{Term} \ni M ::= {} & () \mid X \mid n \mid r \mid f(M) \mid (M, M) \mid M.\mathbf{1} \mid M.\mathbf{2} \\ \mid {} & \mathbf{true} \mid \mathbf{false} \mid \mathbf{if}\ M\ \mathbf{then}\ M\ \mathbf{else}\ M \\ \mid {} & \mathbf{sample}(M) \mid \mathbf{score}(M) \mid \mathbf{return}(M) \\ \mid {} & \mathbf{let}\ X = M\ \mathbf{in}\ M \\ \mathit{Type} \ni \tau ::= {} & 1 \mid \mathbb{N} \mid \mathbb{R} \mid \mathbb{B} \mid \tau \times \tau \mid \mathbb{P}(\tau) \end{aligned}$$

where X ranges over a countably infinite set of names, $n \in \mathbb{N}$, $r \in \mathbb{R}$ and f ranges over all measurable functions. Additionally, we define a few shorthands, all of which are standard:

$$M; N := \mathbf{let}\ \_ = M\ \mathbf{in}\ N \qquad \text{(sequencing)}$$

$$\mathbf{observe}(M) := \mathbf{score}(\mathbf{if}\ M\ \mathbf{then}\ 1\ \mathbf{else}\ 0). \qquad \text{(hard constraint)}$$

A term M is *well-typed* under context Δ if there is a type τ such that $\Delta \vdash_{\mathrm{p}} M : \tau$, where $\vdash_{\mathrm{p}}$ is defined in Fig. 25.

Fixing two measurable spaces $\mathcal{A} = (A, \Sigma_A)$ and $\mathcal{B} = (B, \Sigma_B)$. Recall a measure is *s-finite* if it is the countable sum of finite measures, i.e. $\mu : \Sigma_A \to [0, \infty]$ is *s*-finite if there exists a sequence of finite measures $\{\mu_i : \Sigma_A \longrightarrow [0, \infty]\}_{i \in \mathbb{N}}$ such that $\mu(U) = \sum_{i \in \mathbb{N}} \mu_i(U)$. A *finite kernel* from $\mathcal{A}$ to $\mathcal{B}$ is

Deterministic terms: $\Delta \vdash_{\mathrm{d}} M : \tau \implies [\![M]\!] : |[\![\Delta]\!]| \to |[\![\tau]\!]|$

$$\frac{}{[\![()]\!](\delta) := ()} \qquad \frac{}{[\![X]\!](\delta) := \delta_X} \qquad \frac{}{[\![n]\!](\delta) := n} \qquad \frac{}{[\![r]\!](\delta) := r}$$

$$\frac{[\![M]\!] : |[\![\Delta]\!]| \to |[\![\sigma]\!]| \qquad f : [\![\sigma]\!] \to [\![\tau]\!]}{[\![f(M)]\!](\delta) := f([\![M]\!](\delta))} \qquad \frac{([\![M_i]\!] : |[\![\Delta]\!]| \to |[\![\tau_i]\!]|)_{i=1,2}}{[\![(M_1, M_2)]\!](\delta) := ([\![M_1]\!](\delta), [\![M_2]\!](\delta))}$$

$$\frac{[\![M]\!] : |[\![\Delta]\!]| \to |[\![\sigma \times \tau]\!]|}{[\![M.\mathbf{1}]\!](\delta) := \mathsf{proj}_1([\![M]\!](\delta))} \qquad \frac{[\![M]\!] : |[\![\Delta]\!]| \to |[\![\sigma \times \tau]\!]|}{[\![M.\mathbf{2}]\!](\delta) := \mathsf{proj}_2([\![M]\!](\delta))} \qquad \frac{}{[\![\mathbf{true}]\!](\delta) := 1} \qquad \frac{}{[\![\mathbf{false}]\!](\delta) := 0}$$

$$\frac{[\![M]\!] : |[\![\Delta]\!]| \to |[\![\sigma]\!]| \qquad [\![N_1]\!] : |[\![\Delta]\!]| \times |[\![\sigma]\!]| \to |[\![\tau]\!]| \qquad [\![N_2]\!] : |[\![\Delta]\!]| \to |[\![\tau]\!]|}{[\![\mathbf{if}\ M\ \mathbf{then}\ N_1\ \mathbf{else}\ N_2]\!](\delta) := \mathsf{if}\ [\![M]\!] = \mathsf{inj}_2(x)\ \mathsf{then}\ [\![N_1]\!](\delta, x)\ \mathsf{else}\ [\![N_2]\!](\delta)}$$

Fig. 26. Denotation of deterministic Bppl terms

a map $\kappa : A \times \Sigma_B \to [0, \infty]$ such that $\kappa(-, U)$ is Borel measurable for all $U \in \Sigma_B$, $\kappa(x, -)$ is a finite measure for all $x \in A$, and the family of measures is uniformly finite, i.e. $\sup_{x \in A} \kappa(x, B) < \infty$. An *s-finite kernel* between $\mathcal{A}$ and $\mathcal{B}$ is a map $\kappa : A \times \Sigma_B \to [0, \infty]$ such that $\kappa = \sum_{i \in \mathbb{N}} \kappa_i$ for a sequence of finite kernels $\{\kappa_i\}_{i \in \mathbb{N}}$. The category of s-finite kernels **sfKrn** has measurable spaces as objects and s-finite kernels as morphisms. We write $\mathcal{A} \rightsquigarrow \mathcal{B}$ for the set $\mathbf{sfKrn}(\mathcal{A}, \mathcal{B})$. Given $\kappa : \mathcal{A} \rightsquigarrow \mathcal{B}$ and $\ell : \mathcal{B} \rightsquigarrow C$, the composition operator is defined by

$$(\ell \circ_{\mathbf{sfKrn}} \kappa)(x, U) := \int_B \ell(y, U)\, \kappa(x, \mathrm{d}y).$$

Definition B.2. Let **Syn** be the syntactic category of Bppl terms and types. The denotational semantics of Bppl is defined below by the semantic functor $[\![-]\!] : \mathbf{Syn} \to \mathbf{sfKrn}$, i.e. a typing judgement $\Delta \vdash_{\mathrm{p}} M : \tau$ denotes an s-finite kernel $[\![M]\!] : [\![\Delta]\!] \rightsquigarrow [\![\tau]\!]$. The denotation of types is defined inductive as follows:

$$\begin{aligned} [\![1]\!] &:= (\{*\}, \mathcal{P}(\{*\})) & [\![\mathbb{R}]\!] &:= (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ [\![\mathbb{N}]\!] &:= (\mathbb{N}, \mathcal{P}(\mathbb{N})) & [\![\mathbb{P}(\tau)]\!] &:= \mathcal{G}[\![\tau]\!] \\ [\![\mathbb{B}]\!] &:= (\mathbb{B}, \{\varnothing, \mathcal{P}(\mathbb{B})\}) & [\![\sigma \times \tau]\!] &:= [\![\sigma]\!] \times [\![\tau]\!] \end{aligned}$$

where $\mathcal{G} : \mathbf{Meas} \to \mathbf{Meas}$ is the *Giry functor* that maps $\mathcal{A}$ to a measurable space of probability measures. Note that every Bppl type denotes a *standard Borel space*, which is a measurable space generated by the open sets of a Polish topology. The denotation of context is standard: $\Delta = [X_1 : \tau_1, \ldots, X_n : \tau_n]$ is interpreted as the product measurable space $[\![\Delta]\!] := \prod_{i=1}^n [\![\tau_i]\!]$. The denotation of deterministic and probabilistic Bppl terms are listed in Fig. 26 and Fig. 27 respectively.

In the formulation of program logics, it is often the case that a program contains some metavariables. We distinguish these names by using lowercase letters, as opposed to uppercase letters for random names. For example, **let** X = **sample**($\mathsf{Unif}(0, a)$) **in** $X + Y$ contains three kinds of names: the bound random name $X : \mathbb{R}$, the free random name $Y : \mathbb{R}$, and a metavariable $a : \mathbb{R}$. Technically speaking, the above (pre-)term takes an ambient metavariable context Γ as argument and outputs another term open under a context $\Delta := [Y : \mathbb{R}]$. To formally model programs that may contain metavariables in the separation logic, we consider a new judgement $\vdash_{\mathrm{prog}}$, introduced by

Probabilistic terms: $\Delta \vdash_p M : \tau \implies [\![M]\!] \in [\![\Delta]\!] \rightsquigarrow [\![\tau]\!] \subseteq \{\mu : [\![\Delta]\!] \times \Sigma_{[\![\tau]\!]} \to [0, \infty]\}$

$$\frac{[\![M]\!] : |[\![\Delta]\!]| \to |\mathcal{G}([\![\tau]\!])|}{[\![\mathbf{sample}(M)]\!](\delta, U) := [\![M]\!](\delta)(U)} \qquad \frac{[\![M]\!] : |[\![\Delta]\!]| \to \mathbb{R}}{[\![\mathbf{score}(M)]\!](\delta, U) := \max(0, [\![M]\!](\delta)) \cdot \mathbf{1}_U(*)}$$

$$\frac{[\![M]\!] : |[\![\Delta]\!]| \to |[\![\tau]\!]|}{[\![\mathbf{return}(M)]\!](\delta, U) := \mathbf{1}_U([\![M]\!](\delta))} \qquad \frac{[\![M]\!] : [\![\Delta]\!] \rightsquigarrow [\![\sigma]\!] \qquad [\![N]\!] : [\![\Delta]\!] \times [\![\sigma]\!] \rightsquigarrow [\![\tau]\!]}{[\![\mathbf{let}\ X = M\ \mathbf{in}\ N]\!](\delta, U) := \int_{[\![\sigma]\!]} [\![N]\!]((\delta, x), U)[\![M]\!](\delta, dx)}$$

$$\frac{[\![M]\!] : |[\![\Delta]\!]| \to |[\![\sigma]\!]| \qquad [\![N_1]\!] : [\![\Delta]\!] \times [\![\sigma]\!] \rightsquigarrow [\![\tau]\!] \qquad [\![N_2]\!] : [\![\Delta]\!] \rightsquigarrow [\![\tau]\!]}{[\![\mathbf{if}\ M\ \mathbf{then}\ N_1\ \mathbf{else}\ N_2]\!](\delta, U) := \text{if } [\![M]\!](\delta) \text{ then } [\![N_1]\!](\delta, U) \text{ else } [\![N_2]\!](\delta, U)}$$

Fig. 27. Denotation of probabilistic Bppl terms

the following rule:

T-Prog
$$\frac{\forall \gamma_i \in \mathit{Term}.\ \Delta \vdash_{z_i} \gamma_i : \tau_i \implies \Delta \vdash_p M[\gamma_1/x_1, ..., \gamma_n/x_n] : \tau}{\Gamma, x_1 : \tau_1, ..., x_n : \tau_n; \Delta \vdash_{\mathrm{prog}} M[x_1, ..., x_n] : \tau}\ (z_i \in \{\mathrm{d}, \mathrm{p}\})$$

where $M[x_1, ..., x_n] \in \mathit{Term}[x_1, ..., x_n]$. For example, here is a program with a metavariable $x : \mathbb{R}$:

$$\frac{\forall \gamma.\ \Delta \vdash_d \gamma : \mathbb{R} \implies \Delta \vdash_p \overbrace{(\mathbf{let}\ X = \mathbf{sample}(\mathsf{Unif}(0, x))\ \mathbf{in}\ X + Y)}^{M[x]}[\gamma/x] : \mathbb{R}}{[x : \mathbb{R}]; \Delta \vdash_{\mathrm{prog}} \mathbf{let}\ X = \mathbf{sample}(\mathsf{Unif}(0, x))\ \mathbf{in}\ X + Y : \mathbb{R}}$$

Intuitively, the above expression is a term of $\mathbb{R}$ with a free variable $Y : \mathbb{R}$ and a metavariable $x \in \mathbb{R}$. If $\Gamma; \Delta \vdash_{\mathrm{prog}} M : A$, we say that M is a program of type A under the program variable context Δ and contains metavariables specified in Γ. In cases where it is clear, we simply refer to M as a program. Notice that we write $\mathbb{R}$ instead of the measurable space $(\mathbb{R}, \mathcal{B}\mathbb{R})$ as a shorthand. We denote $\mathrm{fv}(M)$ the set of free random variables for a well-formed program (that may include metavariables) $\Gamma; \Delta \vdash_{\mathrm{prog}} M : \mathcal{A}$. For instance, $\mathrm{fv}(\mathbf{let}\ X = \mathbf{sample}(\mathsf{Bern}(a))\ \mathbf{in}\ X + Y) = \{Y\}$, X is not included because it is bound to **let**, a is not included because it is a metavariable.

C RESOURCE MONOID FOR RANDOMNESS

To develop the Kripke resource monoid for BaSL, we fix the Hilbert cube as our underlying measurable space, i.e. we define $(\Omega, \Sigma_\Omega) \coloneqq ([0,1]^{\mathbb{N}}, \mathcal{B}[0,1]^{\mathbb{N}})$, where $\mathcal{B}[0,1]^{\mathbb{N}}$ is the Borel σ-algebra generated by iterating the product measure on Borel sets of $[0,1]$, i.e. $\mathcal{B}[0,1]^{\mathbb{N}} \coloneqq \bigotimes_{i\in\mathbb{N}} \mathcal{B}[0,1]$. We also adopt standard conventions in measure theory – multiplication $(\cdot) : [0,\infty] \times [0,\infty] \to [0,\infty]$ is defined to satisfy $0 \cdot \infty = \infty \cdot 0 = 0$. Recall a measure $\mu : \mathcal{F} \to [0,\infty]$ is *σ-finite* if there exists a sequence of disjoint measurable sets $\{E_i \in \mathcal{F}\}_{i\in\mathbb{N}}$ such that $\biguplus_{i\in\mathbb{N}} E_i = \Omega$ and $\mu(E_i) < \infty$ for every $i \in \mathbb{N}$. Every finite measure is trivially σ-finite. The Lebesgue measure on $(\mathbb{R}, \mathcal{B}\mathbb{R})$ that returns the length of a subset (e.g. $\lambda_{\mathbb{R}}([a,b]) = b - a$) and the counting measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ that returns the cardinality of the subset (e.g. $\#_{\mathbb{N}}(\{2025, 2026, 2027\}) = 3$) are examples of σ-finite measures with total measure infinity.

Definition C.1. A *random generator* is a pair $(\mathcal{F}, \mu)$ such that $(\Omega, \mathcal{F}, \mu)$ is a σ-finite measure space, $\mathcal{F} \subseteq \Sigma_\Omega$ and $\mu(\Omega) > 0$. We write $\mathcal{M}_1$ for the set of such pairs. That is, we define

$$\mathcal{M}_1 \coloneqq \{(\mathcal{F}, \mu) \mid (\Omega, \mathcal{F}, \mu)\ \sigma\text{-finite measure space}, \mathcal{F} \subseteq \Sigma_\Omega, \mu(\Omega) > 0\}.$$

Notice we used the notation $\mathcal{M}_1$ instead of $\mathcal{M}$ because $\mathcal{M}_1$ is not the final definition – it needs to be further restricted in order to have desirable properties, namely the existence of disintegration and Borel measure extensions.

Definition C.2 ([Li et al. 2023, Definition 2.2]). Let $(\Omega, \mathcal{H}, \rho)$ be a σ-finite measure space. We say that $(\mathcal{H}, \rho)$ is the *independent combination* of $(\mathcal{F}, \mu)$ and $(\mathcal{G}, \nu)$ if (1) $\mathcal{H}$ is the smallest σ-algebra containing $\mathcal{F}$ and $\mathcal{G}$, and (2) for all $F \in \mathcal{F}$ and $G \in \mathcal{G}$, $\rho(F \cap G) = \mu(F)\nu(G)$.

Recall that a *π-system* over a set A is a collection of subsets $\Pi \subseteq \mathcal{P}(A)$ such that Π is closed under intersection. Similarly, a *λ-system* over A is a collection of subsets $\Lambda \subseteq \mathcal{P}(A)$ such that $\emptyset \in \Lambda$, Λ is closed under complements and countable disjoint union. Let $C_i \subseteq \mathcal{P}(\Omega)$ be a set of subsets of Ω for $i \in \{1, \ldots, n\}$. We write $\sigma\{C_1, \ldots, C_n\}$ for the smallest σ-algebra generated by $C_1 \cup \ldots \cup C_n$.

Proposition C.3 ([Li et al. 2023, Lemma B.2]). *Let $\mathcal{F}, \mathcal{G}$ be σ-algebras of Ω. Then $\Pi \subseteq \mathcal{P}(A)$ defined by $\Pi \coloneqq \{F \cap G \mid F \in \mathcal{F}, G \in \mathcal{G}\}$ is a π-system that generates the smallest σ-algebra that contains $\mathcal{F}$ and $\mathcal{G}$.*

Proposition C.4 ([Billingsley 1995, Theorem 10.3]). *Let $\Pi \subseteq \mathcal{P}(\Omega)$ be a π-system, $\sigma\{\Pi\}$ be the smallest σ-algebra containing Π, and $\mu, \nu : \sigma\{\Pi\} \to [0,\infty]$ be σ-finite measures. If $\mu|_\Pi = \nu|_\Pi$ and there exists a sequence $\{U_i \in \Pi\}_{i\in\mathbb{N}}$ such that $\mu(U_i) < \infty$ for all $i \in \mathbb{N}$ and $\bigcup_{i\in\mathbb{N}} U_i = \Omega$, then $\mu = \nu$.*

Lemma C.5 (uniqueness). *Let $(\mathcal{H}, \rho)$ and $(\mathcal{H}', \rho')$ be independent combinations of $(\mathcal{F}, \mu)$ and $(\mathcal{G}, \nu)$, then $\mathcal{H} = \mathcal{H}'$ and $\rho = \rho'$.*

Proof. The σ-algebras $\mathcal{H}$ and $\mathcal{H}'$ are equal as both are defined to be the smallest σ-algebra containing $\mathcal{F}$ and $\mathcal{G}$. By Proposition C.4, ρ and ρ' are equal when (1) there is a π-system Π such that Π generates $\mathcal{H}$ and $\rho|_\Pi = \rho'|_\Pi$, and (2) there is a sequence $\{W_i \in \Pi\}_{i\in\mathbb{N}}$ such that $\bigcup_{i\in\mathbb{N}} W_i = \Omega$ and $\rho(W_i) < \infty$ for all $i \in \mathbb{N}$. By Proposition C.3, the π-system $\Pi \coloneqq \{F \cap G \mid F \in \mathcal{F}, G \in \mathcal{G}\}$ generates $\sigma\{\mathcal{F}, \mathcal{G}\}$. To show (1), notice for any $F \in \mathcal{F}$ and $G \in \mathcal{G}$, $\rho(F \cap G) = \mu(F)\nu(G) = \rho'(F \cap G)$. For (2), notice there are two sequences of disjoint covers $\{U_i \in \mathcal{F}\}_{i\in\mathbb{N}}$ and $\{V_i \in \mathcal{G}\}_{i\in\mathbb{N}}$ such that $\mu(U_i), \nu(V_i) < \infty$. Consider the countable set $\{U_i \cap V_j \in C\}_{i,j\in\mathbb{N}}$, clearly $\bigcup_{i,j} U_i \cap V_j = \Omega$ and for all $i, j \in \mathbb{N}$, $\rho(U_i \cap V_i) = \mu(U_i)\nu(V_i) < \infty$ by assumption. ◻

Lemma C.6 (closure). *If $(\mathcal{H}, \rho)$ is the independent combination of $(\mathcal{F}, \mu) \in \mathcal{M}_1$ and $(\mathcal{G}, \nu) \in \mathcal{M}_1$, then $(\mathcal{H}, \rho) \in \mathcal{M}_1$.*

Proof. We need to show that (1) $\mathcal{H} \subseteq \Sigma_\Omega$, (2) $\rho(\Omega) > 0$, and (3) $\rho : \mathcal{H} \to [0, \infty]$ is σ-finite. For (1), $\mathcal{F} \subseteq \Sigma_\Omega$ and $\mathcal{G} \subseteq \Sigma_\Omega$ implies $\mathcal{F} \cup \mathcal{G} \subseteq \Sigma_\Omega$, hence the generated σ-algebra $\mathcal{H}$ is a subset of Σ_Ω. For (2), $\rho(\Omega) > 0$ because $\rho(\Omega \cap \Omega) = \mu(\Omega) \cdot \nu(\Omega) > 0$ by definition of $\mathcal{M}_1$; For (3), the sequences of disjoint covers $\{U_i \in \mathcal{F}\}_{i\in\mathbb{N}}$ and $\{V_i \in \mathcal{G}\}_{i\in\mathbb{N}}$ allow us to define the countable set $\{W_{i,j} \in \mathcal{H}\}_{i,j\in\mathbb{N}}$ with $W_{i,j} := U_i \cap V_j$. Notice $\bigcup_{i,j\in\mathbb{N}} W_{i,j} = \Omega$ and $\rho(W_{i,j}) = \mu(U_i)\nu(V_j) < \infty$, which means ρ is σ-finite. □

Due to the uniqueness and closure lemmas (Lemmas C.5 and C.6), there exists a unique partial function $(\oplus) : \mathcal{M}_1 \times \mathcal{M}_1 \rightharpoonup \mathcal{M}_1$ that maps two elements to their independent combination when it exists. Moreover, let $\mu_1 : \{\emptyset, \Omega\} \to [0, \infty]$ be the trivial probability measure with $\mu_1(\Omega) := 1$ and $\mathbf{1} := (\{\emptyset, \Omega\}, \mu_1)$. Then $(\mathcal{M}_1, \oplus, \mathbf{1})$ is a partial commutative monoid:

Theorem C.7 (PCM). *$(\mathcal{M}_1, \oplus, \mathbf{1})$ is a partial commutative monoid.*

Proof. To prove identity, we need to show that when $m := (\mathcal{F}, \mu)$, then $\mathbf{1} \oplus m$ is defined and $\mathbf{1} \oplus m = m$. For the σ-algebra, compute that $\sigma\{\{\emptyset, \Omega\}, \mathcal{F}\} = \mathcal{F}$. For the measure, notice for all $E \in \{\emptyset, \Omega\}$ and $F \in \mathcal{F}$, $\mu(E \cap F) = \mu_1(E)\mu(F)$ – when $E = \emptyset$, we have $\mu(E \cap F) = 0 = \mu_1(E)\mu(F)$ (recall $0 \cdot \infty$ is defined to be 0 in measure theory). When $E = \Omega$, we have $\mu(E \cap F) = \mu(F) = \mu_1(E)\mu(F)$. Hence, the combination $\mathbf{1} \oplus m$ is defined and is equal to m.

To prove commutativity, let $m_1 := (\mathcal{F}_1, \mu_1)$ and $m_2 := (\mathcal{F}_2, \mu_2)$, we need to show that if $m_1 \oplus m_2$ is defined then $m_2 \oplus m_1$ is defined and $m_1 \oplus m_2 = m_2 \oplus m_1$. For the σ-algebra, compute that $\sigma\{\mathcal{F}_1, \mathcal{F}_2\} = \sigma\{\mathcal{F}_2, \mathcal{F}_1\}$. For the measures, let $m_{12} := (\mathcal{F}_{12}, \mu_{12}) := m_1 \oplus m_2$, notice for all $F_1 \in \mathcal{F}_1$ and $F_2 \in \mathcal{F}_2$, we have $\mu_{12}(F_2 \cap F_1) = \mu_{12}(F_1 \cap F_2) = \mu_1(F_1)\mu_2(F_2) = \mu_2(F_2)\mu_1(F_1)$. Hence, m_{12} witnesses the independent combination of m_2 and m_1.

To prove associativity, let $m_1 := (\mathcal{F}_1, \mu_1)$, $m_2 := (\mathcal{F}_2, \mu_2)$, $m_3 := (\mathcal{F}_3, \mu_3)$, $m_{12} := (\mathcal{F}_{12}, \mu_{12}) := m_1 \oplus m_2$, $m_{(12)3} := (\mathcal{F}_{(12)3}, \mu_{(12)3}) := m_{12} \oplus m_3$. We need to show that if $m_1 \oplus m_2$ and $(m_1 \oplus m_2) \oplus m_3$ are defined, then $m_2 \oplus m_3$ and $m_1 \oplus (m_2 \oplus m_3)$ are defined, and $(m_1 \oplus m_2) \oplus m_3 = m_1 \oplus (m_2 \oplus m_3)$. We first show $m_2 \oplus m_3$ is defined: any element $\mathcal{M}_1$ is σ-finite and has non-zero total measure by definition. Hence, there exists a set $V \in \mathcal{F}_1$ such that $0 < \mu_1(V) < \infty$. Consider a function $\rho : \sigma\{\mathcal{F}_2, \mathcal{F}_3\} \to [0, \infty]$ defined by $\rho(F_{23}) := \frac{\mu_{(12)3}(V \cap F_{23})}{\mu_1(V)}$. Notice that ρ is a measure: when $F_{23} = \emptyset$, we have $\rho(F_{23}) = 0$, when $F_{23} = \biguplus_{i\in\mathbb{N}} F_i$, we have $\rho(\biguplus_{i\in\mathbb{N}} F_i) = \sum_{i\in\mathbb{N}} \mu_{(12)3}(V \cap F_i)/\mu_1(V) = \sum_{i\in\mathbb{N}} \rho(F_i)$. Also, ρ is the independent combination measure of $F_2 \in \mathcal{F}_2$ and $F_3 \in \mathcal{F}_3$ because $\rho(F_2 \cap F_3) = \mu_{(12)3}(V \cap F_2 \cap F_3)/\mu_1(V) = \mu_1(V)\mu_2(F_2)\mu_3(F_3)/\mu_1(V) = \mu_2(F_2)\mu_3(F_3)$. Hence, $m_2 \oplus m_3$ is defined and is equal to $(\sigma\{\mathcal{F}_2, \mathcal{F}_3\}, \rho)$. Note that the choice of V does not matter, they all yield the same measure by the uniqueness lemma (Lemma C.5). We write $m_{23} := (\mathcal{F}_{23}, \mu_{23}) := m_2 \oplus m_3$.

We now show that $m_1 \oplus (m_2 \oplus m_3)$ is defined by proving $\mathcal{F}_{1(23)} = \sigma\{\mathcal{F}_{12}, \mathcal{F}_3\}$ and $\mu_{1(23)} = \mu_{(12)3}$. This means we need to show that $\mu_{(12)3}$ is the independent combination of μ_1 and μ_{23}, i.e. proving for all $F_1 \in \mathcal{F}_1$ and $F_{23} \in \mathcal{F}_{23}$, $\mu_{(12)3}(F_1 \cap F_{23}) = \mu_1(F_1)\mu_{23}(F_{23})$. We first define $\Lambda, \Lambda' \subseteq \mathcal{P}(\Omega)$ by

$$\Lambda := \left\{F_{23} \in \mathcal{F}_{23} \,\middle|\, \text{for all } F_1 \in \mathcal{F}_1, \mu_{(12)3}(F_1 \cap F_{23}) = \mu_1(F_1)\mu_{23}(F_{23})\right\},$$

$$\Lambda' := \left\{F_{23} \in \mathcal{F}_{23} \,\middle|\, \begin{array}{l}\text{for all } F_1 \in \mathcal{F}_1, (\mu_1(F_1) < \infty \text{ and } \mu_{23}(F_{23}) < \infty) \\ \text{implies } \mu_{(12)3}(F_1 \cap F_{23}) = \mu_1(F_1)\mu_{23}(F_{23})\end{array}\right\}.$$

We now show that $\Lambda = \Lambda'$ and Λ forms a λ-system – $\Lambda \subseteq \Lambda'$ is trivially true. To show that $\Lambda' \subseteq \Lambda$, let $F_{23} \in \Lambda'$, we consider two cases: (1) when $\mu_{23}(F_{23}) < \infty$, and (2) when $\mu_{23}(F_{23}) = \infty$. For (1), let $F_1 \in \mathcal{F}_1$ and $\{U_i \in \mathcal{F}_1\}_{i\in\mathbb{N}}$ be a disjoint cover of Ω with $\mu_1(F_1) < \infty$ (which exists because μ_1 is σ-finite). Then

$$\mu_{(12)3}(F_1 \cap F_{23}) = \sum_{i\in\mathbb{N}} \mu_{(12)3}(F_1 \cap U_i \cap F_{23}) \qquad \text{(additivity)}$$

$$= \sum_{i\in\mathbb{N}} \mu_1(F_1 \cap U_i)\mu_{23}(F_{23}) \qquad (\mu_1(F_1 \cap U_i) < \infty)$$
$$= \mu_1(F_1)\mu_{23}(F_{23}). \qquad \text{(additivity)}$$

For (2), consider a disjoint cover $\{V_i \in \mathcal{F}_{23}\}_{i\in\mathbb{N}}$ with $\mu_{23}(V_i) < \infty$. Then

$$\begin{aligned}
\mu_{(12)3}(F_1 \cap F_{23}) &= \sum_{i\in\mathbb{N}}\sum_{j\in\mathbb{N}} \mu_{(12)3}(F_1 \cap U_i \cap F_{23} \cap V_j) && (\sigma\text{-finite measure})\\
&= \sum_{i\in\mathbb{N}}\sum_{j\in\mathbb{N}} \mu_1(F_1 \cap U_i)\mu_{23}(F_{23} \cap V_j) && (\mu_{23}(F_{23} \cap V_j) < \infty)\\
&= \sum_{i\in\mathbb{N}} \mu_1(F_1 \cap U_i) \sum_{j\in\mathbb{N}} \mu_{23}(F_{23} \cap V_j) && (\text{distributivity in } [0,\infty])\\
&= \mu_1(F_1)\mu_{23}(F_{23}). && (\text{additivity})
\end{aligned}$$

Hence, $\Lambda = \Lambda'$. To show that Λ forms a λ-system, we need to prove that (i) $\emptyset \in \Lambda'$, (ii) Λ' is closed under complements, and (iii) Λ' is closed under countable disjoint unions. For (i), it is true because $\mu_{(12)3}(\emptyset \cap F_{23}) = 0 = \mu_1(\emptyset)\mu_{23}(F_{23})$; For (ii), suppose $F_{23} \in \Lambda'$, then for all $F_1 \in \mathcal{F}_1$, if $0 < \mu_1(F_1) < \infty$, then

$$\mu_{(12)3}(F_1 \cap (\Omega \setminus F_{23})) = \mu_1(F_1)\frac{\mu_{(12)3}(F_1 \cap (\Omega \setminus F_{23}))}{\mu_1(F_1)} = \mu_1(F_1)\mu_{23}(\Omega \setminus F_{23}).$$

The second equality holds because we assumed $\mu_1(F_1) > 0$, which means the measure $F_{23} \mapsto {\mu_{(12)3}(F_1 \cap F_{23})}/{\mu_1(F_1)}$ witnesses the independent combination of $\mathcal{F}_2$ and $\mathcal{F}_3$. By Lemma C.5, we have $\mu_{23} = {\mu_{(12)3}(F_1 \cap -)}/{\mu_1(F_1)}$. Next, assuming $\mu_1(F_1) = 0$, then

$$\mu_{(12)3}(F_1 \cap (\Omega \setminus F_{23})) \le \mu_{(12)3}(F_1) = \mu_1(F_1)\mu_{23}(\Omega) = 0.$$

Hence, $\mu_{(12)3}(F_1 \cap (\Omega \setminus F_{23})) = 0 = \mu_1(F_1)\mu_{23}(\Omega \setminus F_{23})$ and (ii) holds. For (iii), assuming $\{F_i \in \Lambda'\}_{i\in\mathbb{N}}$ is a sequence of disjoint sets, then for any $G \in \mathcal{F}_1$ with $\mu_1(G) < \infty$ and $\mu_{23}(\biguplus_{i\in\mathbb{N}} F_i) < \infty$, we have

$$\mu_{(12)3}\left(G \cap \biguplus_{i\in\mathbb{N}} F_i\right) = \sum_{i\in\mathbb{N}} \mu_{(12)3}(G \cap F_i) = \sum_{i\in\mathbb{N}} \mu_1(G)\mu_{23}(F_i) = \mu_1(G)\mu_{23}\left(\biguplus_{i\in\mathbb{N}} F_i\right).$$

Hence, $\Lambda = \Lambda'$ forms a λ-system. By the π-λ theorem, to show that $\mathcal{F}_{23} \subseteq \Lambda'$, it suffices to show that $\Pi \subseteq \Lambda'$ for some π-system Π that generates $\mathcal{F}_{23}$. By Proposition C.3, we know that $\Pi := \{F_2 \cap F_3 \mid F_2 \in \mathcal{F}_2, F_3 \in \mathcal{F}_3\}$ generates $\mathcal{F}_{23}$. As a consequence, if $\Pi \subseteq \Lambda'$, then $\mathcal{F}_{23} \subseteq \Lambda$. Consider any element $F_2 \cap F_3 \in \Pi$, we need to show that for all $F_1 \in \mathcal{F}_1$, if $\mu_1(F_1) < \infty$ and $\mu_{23}(F_2 \cap F_3) < \infty$, then $\mu_{(12)3}(F_1 \cap (F_2 \cap F_3)) = \mu_1(F_1)\mu_{23}(F_2 \cap F_3)$, which is trivially true by definition of independent combination. This means $\mathcal{F}_{23} = \Lambda' = \Lambda$, which implies associativity. Hence, $(\mathcal{M}_1, \oplus, \mathbf{1})$ is a partial commutative monoid. □

Definition C.8 ([Li et al. 2023, Definition 2.6]). A σ-algebra $\mathcal{F}$ of Ω is said to have *finite footprint* if every element is of the form $F' \times [0,1]^{\mathbb{N}}$ for some $F' \subseteq [0,1]^n$ and $n \in \mathbb{N}$. A random variable $X : (\Omega, \Sigma_\Omega) \to (A, \Sigma_A)$ is said to have finite footprint if its pullback σ-algebra $\{X^{-1}(F) \mid F \in \Sigma_A\}$ has finite footprint. We denote $\mathrm{RV}(A, \Sigma_A)$ the set of random variables from (Ω, Σ_Ω) with finite footprint.

Definition C.9. Recall that a σ-algebra $\mathcal{F}$ is countably generated if there exists a set of subsets $C \subseteq \mathcal{P}(\Omega)$ such $\sigma\{C\} = \mathcal{F}$. We say that *random generator* $\mathcal{M}$ is the restriction of $\mathcal{M}_1$ where the σ-algebras are countably generated and have finite footprint, i.e. the set is defined by

$$\mathcal{M} := \left\{ (\mathcal{F}, \mu) \;\middle|\; \begin{array}{l} (\Omega, \mathcal{F}, \mu)\ \sigma\text{-finite measure space} \\ \mathcal{F} \subseteq \Sigma_\Omega, \mu(\Omega) > 0 \\ \mathcal{F} \text{ finite footprint and countably generated} \end{array} \right\}.$$

Proposition C.10 (Borel measure extension). *Let $(\mathcal{F}, \mu) \in \mathcal{M}$. Then there exists a σ-finite Borel measure $\mu^+ : \Sigma_\Omega \to [0, \infty]$ satisfying $\mu^+|_{\mathcal{F}} = \mu$.*

Proof. Since (Ω, Σ_Ω) is a standard Borel space and $\mathcal{F}$ is a countably generated sub-σ-algebra of Σ_Ω. By Fremlin [2011, Proposition 433K], the extension $\mu^+ : \Sigma_\Omega \to [0, \infty]$ exists. By σ-finiteness of μ, there is a sequence of sets $\{F_i \in \mathcal{F} \subseteq \Sigma_\Omega\}_{i\in\mathbb{N}}$ that satisfies $\bigcup_{i\in\mathbb{N}} F_i = \Omega$ and $\mu^+(F_i) = \mu^+|_{\mathcal{F}}(F_i) = \mu(F_i) < \infty$. □

Lemma C.11 (BaSL PCM). *Let $(\bullet) : \mathcal{M} \times \mathcal{M} \rightharpoonup \mathcal{M}$ be a partial function defined by*

$$m_1 \bullet m_2 := \begin{cases} m_1 \oplus m_2 & \text{if } m_1 \oplus m_2 \in \mathcal{M} \\ \text{undefined} & \text{otherwise.} \end{cases}$$

Then $(\mathcal{M}, \bullet, \mathbf{1})$ is a partial commutative monoid.

Proof. Notice $\mathbf{1} \in \mathcal{M}$ because it has finite footprint 0 and the σ-algebra is finite, which means countably generated. Next, we show that $(\oplus)$ is closed under $\mathcal{M}$: Let $m_1 := (\mathcal{F}_1, \mu_1)$ and $m_2 := (\mathcal{F}_1, \mu_2) \in \mathcal{M}$, assuming $(\mathcal{F}_{12}, \mu_{12}) := m_1 \bullet m_2$ exists and $\mathcal{F}, \mathcal{G}$ are generated by $\{F_i\}_{i\in\mathbb{N}}$ and $\{G_i\}_{i\in\mathbb{N}}$ respectively, then $\mathcal{F}_{12}$ is generated by $\{F_i\}_{i\in\mathbb{N}} \cup \{G_i\}_{i\in\mathbb{N}}$, which is a countable set. We also need to check that $\mathcal{F}_{12}$ has finite footprint when $\mathcal{F}_1$ and $\mathcal{F}_2$ have finite footprints, this property is analogous to Li et al. [2023, Lemma B.7], where $\max(m, n)$ witnesses the finite footprint of $\mathcal{F}_{12}$ when $\mathcal{F}_1$ and $\mathcal{F}_2$ have finite footprint m and n respectively. □

Definition C.12 (KRM). A *(partial commutative) Kripke resource monoid* is a quadruple $(A, \cdot, 1, \sqsubseteq)$ such that $(A, \cdot, 1)$ is a partial commutative monoid, $(\sqsubseteq) \subseteq A \times A$ is a partial order such that $(\cdot)$ is bifunctorial with respect to $(\sqsubseteq)$, i.e. if $m_1 \sqsubseteq n_1$, $m_2 \sqsubseteq n_2$, and both $m_1 \cdot m_2$ and $n_1 \cdot n_2$ are defined, then $m_1 \cdot m_2 \sqsubseteq n_1 \cdot n_2$.

Theorem C.13 (BaSL KRM). *Let $(\sqsubseteq) \subseteq \mathcal{M} \times \mathcal{M}$ be a partial order defined by*

$$(\mathcal{F}_1, \mu_1) \sqsubseteq (\mathcal{F}_2, \mu_2) \text{ if and only if } \mathcal{F}_1 \subseteq \mathcal{F}_2 \text{ and } \mu_2|_{\mathcal{F}_1} = \mu_1.$$

Then $(\mathcal{M}, \bullet, \mathbf{1}, \sqsubseteq)$ is a Kripke resource monoid.

Proof. $(\mathcal{M}, \bullet, \mathbf{1})$ is a PCM by Lemma C.11. For $(\sqsubseteq)$, it suffices to check that $(\bullet)$ is bifunctorial with respect to $(\sqsubseteq)$. Let $m_i := (\mathcal{F}_i, \mu_i)$ and $n_i = (\mathcal{G}_i, \nu_i)$ for $i = 1, 2$. Suppose $m_1 \sqsubseteq n_1$, $m_2 \sqsubseteq n_2$ and $(\mathcal{G}_{12}, \mu_{12}) := n_1 \bullet n_2$ is defined, then by definition of independent combination, we know for all $G_1 \in \mathcal{G}_1$ and $G_2 \in \mathcal{G}_2$, $\mu_{12}(G_1 \cap G_2) := \mu_1(G_1) \cdot \mu_2(G_2)$. Since $\mathcal{F}_1 \subseteq \mathcal{G}_1$ and $\mathcal{F}_2 \subseteq \mathcal{G}_2$ by assumption, we know $\mu_{12}|_{\sigma\{\mathcal{F}_1,\mathcal{F}_2\}}(F_1 \cap F_2) = \mu_1(F_1) \cdot \mu_2(F_2)$ for all $F_1 \in \mathcal{F}_1$ and $F_2 \in \mathcal{F}_2$. This means $\mu_{12}|_{\sigma\{\mathcal{F}_1,\mathcal{F}_2\}}$ witnesses the independent combination of m_1 and m_2, which implies $m_1 \bullet m_2 = (\sigma\{\mathcal{F}_1, \mathcal{F}_2\}, \mu_{12}|_{\sigma\{\mathcal{F}_1,\mathcal{F}_2\}}) \sqsubseteq n_1 \bullet n_2$. □

Lemma C.14. *Let $m_i := (\mathcal{F}_i, \mu_i) \in \mathcal{M}$ for $i = 1, 2, 3$. Suppose $m_1 \bullet m_2$ is defined and $m_3 \sqsupseteq m_1 \bullet m_2$. Let $f : \Omega \to [0, \infty]$ be an $(\mathcal{F}_1, \mathcal{B}[0, \infty])$-measurable function and $F_2 \in \mathcal{F}_2$. Then*

$$\int_{F_2} f \, \mathrm{d}\mu_3 = \int_\Omega f \, \mathrm{d}\mu_1 \cdot \mu_2(F_2).$$

Proof. Suppose $f = \mathbf{1}_{E_1}$ for some $E_1 \in \mathcal{F}_1$, then $\int_{F_2} \mathbf{1}_{E_1} \, \mathrm{d}\mu_3 = \mu_3(F_1 \cap F_2) = \mu_1(F_1)\mu_2(F_2) = \int_\Omega \mathbf{1}_{E_1} \, \mathrm{d}\mu_1 \cdot \mu_2(F_2)$. Suppose f is a $[0, \infty)$-valued simple function $\sum_{i=1}^n c_i \mathbf{1}_{E_i}$, then

$$\int_{F_2} \sum_{i=1}^n c_i \mathbf{1}_{E_i} \, \mathrm{d}\mu_3 = \sum_{i=1}^n c_i \int_{F_2} \mathbf{1}_{E_i} \, \mathrm{d}\mu_3 = \sum_{i=1}^n c_i \mu_1(E_i)\mu_2(F_2) = \int_\Omega \sum_{i=1}^n c_i \mathbf{1}_{E_i} \, \mathrm{d}\mu_1 \cdot \mu_2(E_2).$$

By the simple function approximation theorem for $[0,\infty]$, every measurable function $f : \Omega \to [0,\infty]$ has a sequence of $[0,\infty)$-simple measurable functions $\{f_i : \Omega \to [0,\infty)\}_{i\in\mathbb{N}}$ that converges pointwise to f, i.e. $f = \lim_{n\to\infty} f_n$. By the monotone convergence theorem, we have

$$\begin{aligned}\int_{F_2} \lim_{n\to\infty} f_n \,\mathrm{d}\mu_3 &= \lim_{n\to\infty} \int_{F_2} f_n \,\mathrm{d}\mu_3 \\ &= \lim_{n\to\infty} \int_{\Omega} f_n \,\mathrm{d}\mu_1 \cdot \mu_2(F_2) \\ &= \int_{\Omega} \lim_{n\to\infty} f_n \,\mathrm{d}\mu_1 \cdot \mu_2(F_2) \\ &= \int_{\Omega} f \,\mathrm{d}\mu_1 \cdot \mu_2(F_2).\end{aligned}$$

Hence, the property holds for all measurable functions. □

Lemma C.15. *Let $m_i := (\mathcal{F}_i, \mu_i) \in \mathcal{M}$ for $i = 1, 2$. Suppose $m_1 \bullet m_2$ is defined and $m_3 \sqsupseteq m_1 \bullet m_2$. Let $f_i : \Omega \to [0,\infty]$ be an $(\mathcal{F}_i, \mathcal{B}[0,\infty])$-measurable function. Then*

$$\int_{\Omega} f_1 f_2 \,\mathrm{d}\mu_3 = \int_{\Omega} f_1 \,\mathrm{d}\mu_1 \cdot \int_{\Omega} f_2 \,\mathrm{d}\mu_2.$$

Proof. Suppose $f_2 = \mathbf{1}_F$ for some $E \in \mathcal{F}_2$, then we apply Lemma C.14 and yield $\int_\Omega f_1 \mathbf{1}_E \,\mathrm{d}\mu_3 = \int_\Omega f_1 \,\mathrm{d}\mu_1 \cdot \int_\Omega \mathbf{1}_E \,\mathrm{d}\mu_2$. The next steps are similar to the proof of Lemma C.14, suppose $f_2 = \sum_{i=1}^n c_i \mathbf{1}_{E_i}$, then linearity of integral in $[0,\infty]$ ensures the property holds. For the limiting case $f_2 = \lim_{n\to\infty} f_n'$, we apply the monotone convergence theorem. □

Lemma C.16. *Let $m_i := (\mathcal{F}_i, \mu_i) \in \mathcal{M}$ for $i = 1, \ldots, n$. Suppose $m_1 \bullet \ldots \bullet m_n$ is defined and $m \sqsupseteq m_1 \bullet \ldots \bullet m_n$. Let $f_i : \Omega \to [0,\infty]$ be an $(\mathcal{F}_i, \mathcal{B}[0,\infty])$-measurable function. Then*

$$\int_{\Omega} \prod_{i=1}^n f_i \,\mathrm{d}\mu = \prod_{i=1}^n \int_{\Omega} f_i \,\mathrm{d}\mu_i.$$

Proof. By induction on $n \in \mathbb{N}_1$. The base case is equal by definition. For the inductive case, suppose $m_1 \bullet \ldots \bullet m_n$ is defined, then $m_1 \bullet \ldots \bullet m_{n-1}$ is defined with the σ-algebra $\mathcal{F}_{1..n-1} := \sigma\{\mathcal{F}_1, \ldots, \mathcal{F}_{n-1}\}$. It follows from Lemma C.15 that

$$\int_{\Omega} f_n \prod_{i=1}^{n-1} f_i \,\mathrm{d}\mu = \int_{\Omega} f_n \,\mathrm{d}\mu_n \cdot \int_{\Omega} \prod_{i=1}^{n-1} f_i \,\mathrm{d}\mu = \prod_{i=1}^{n} \int_{\Omega} f_i \,\mathrm{d}\mu_i.$$

□

Proposition C.17 (mutual independence). *Let $m_i := (\mathcal{F}_i, \mu_i) \in \mathcal{M}$ and $X_i : \Omega \to A_i$ for $i = 1, \ldots, n$ and suppose $(\mathcal{F}, \mu) := m_1 \bullet \ldots \bullet m_n$ is defined. If for $i = 1, \ldots, n$, X_i is $(\mathcal{F}_i, \Sigma_{A_i})$-measurable and $E_i \in \Sigma_{A_i}$, then*

$$\mu\{X_1 \in E_1, \ldots, X_n \in E_n\} = \prod_{i=1}^n \mu_i\{X_i \in E_i\}.$$

Proof. By direct computation:

$$\begin{aligned}\mu\{X_1 \in E_1, \ldots, X_n \in E_n\} &= \mu\{\omega \in \Omega \mid \forall i \in \{1, \ldots, n\}, X_i(\omega) \in E_i\} \\ &= \int_{\Omega} \mathbf{1}_{\bigcap_{i=1}^n X_i^{-1}(E_i)} \,\mathrm{d}\mu\end{aligned}$$

$$= \int_\Omega \prod_{i=1}^{n} \mathbf{1}_{X_i^{-1}(E_i)} \, \mathrm{d}\mu$$

$$= \prod_{i=1}^{n} \int_\Omega \mathbf{1}_{X_i^{-1}(E_i)} \, \mathrm{d}\mu_i \qquad \text{(Lemma C.16)}$$

$$= \prod_{i=1}^{n} \mu_i\{X_i \in E_i\}. \qquad \square$$

Lemma C.18. *Let $\varphi : \Omega \to [0, \infty)$ be an $(\mathcal{F}, \mathcal{B}[0, \infty))$-measurable function and $\mu : \mathcal{F} \to [0, \infty]$ a σ-finite measure. Then the induced measure $(\varphi \cdot \mu) : \mathcal{F} \to [0, \infty]$ defined by*

$$(\varphi \cdot \mu)(U) := \int_U \varphi \, \mathrm{d}\mu$$

is σ-finite. Moreover, if $(\mathcal{F}, \mu) \in \mathcal{M}$ and $(\varphi \cdot \mu)(\Omega) > 0$, then $(\mathcal{F}, \varphi \cdot \mu) \in \mathcal{M}$.

Proof. Since μ is σ-finite, there is a cover $\{U_i \in \mathcal{F}\}_{i \in \mathbb{N}}$ with $\mu(U_i) < \infty$. For each $i \in \mathbb{N}$, we define the sequence $\{U_{i,j} \subseteq \Omega\}_{j \in \mathbb{N}}$ with $U_{i,j} := \{\omega \in U_i \mid \varphi(\omega) \leq j\}$. Notice (1) $\bigcup_{j \in \mathbb{N}} U_{i,j} = U_i$, and (2) $U_{i,j} \in \mathcal{F}$: (1) is true as φ has finite range, meaning that for any $\omega \in \Omega$, there exists an $N \in \mathbb{N}$ such that $\varphi(\omega) < N$. Hence, $\bigcup_{j \in \mathbb{N}} U_{i,j}$ 'covers' all of U_i. For (2), notice that $U_{i,j} = U_i \cap \varphi^{-1}([0, j])$. By measurability of φ, we know $\varphi^{-1}([0, j]) \in \mathcal{F}$, which implies $U_i \cap \varphi^{-1}([0, j]) = U_i \in \mathcal{F}$. Next, notice the restriction $\varphi|_{U_{i,j}} : U_{i,j} \to [0, \infty)$ is bounded above by j. Since $\mu(U_i) < \infty$, $U_{i,j} \in \mathcal{F}$ and $U_{i,j} \subseteq U_i$, we know $\mu(U_{i,j}) \leq \mu(U_i) < \infty$. Then

$$(\varphi \cdot \mu)(U_{i,j}) = \int_{U_{i,j}} \varphi \, \mathrm{d}\mu \leq \mu(U_{i,j}) \cdot \sup \varphi|_{U_{i,j}} \leq \mu(U_i) \cdot j < \infty.$$

This implies $\{U_{i,j} \in \mathcal{F}\}_{i,j \in \mathbb{N}}$ is a countable cover of $\mathcal{F}$-measurable sets with each $(\varphi \cdot \mu)(U_{i,j}) < \infty$, which means $\varphi \cdot \mu$ is σ-finite. Assuming $(\mathcal{F}, \mu) \in \mathcal{M}$ and $(\varphi \cdot \mu)(\Omega) > 0$, since we have proven that $\varphi \cdot \mu$ is σ-finite and $\mathcal{F}$ stays the same, we know $(\mathcal{F}, \varphi \cdot \mu) \in \mathcal{M}$. $\square$

D SYNTAX AND SEMANTICS OF BASL

In this section, we range Γ and Δ over two kinds of context: the *deterministic context* $\Gamma = [x_1 : A_1, \ldots, x_n : A_n]$, which is a list of names x_i and their underlying *sets* A_i, and the *probabilistic context* $\Delta = [X_1 : \mathcal{A}_1, \ldots, X_n : \mathcal{A}_n]$, which is a list of names X_i and their underlying *measurable spaces* $\mathcal{A}_i = (A_i, \Sigma_{A_i})$. The contexts admit a standard interpretation (in the sense of denotational semantics): $[\![\Gamma]\!]$ is interpreted as the (**Set**-)cartesian product by $A_1 \times \ldots \times A_n$, while $[\![\Delta]\!]$ is interpreted as the product measurable space $\mathcal{A}_1 \times \ldots \times \mathcal{A}_n$.

Definition D.1. The *syntax* of a BaSL proposition is defined by the following grammar:

$$\begin{aligned} P ::= {} & \top \mid \bot \mid P \wedge P \mid P \vee P \mid P \Rightarrow P \mid P * P \mid P \mathrel{-\!\!*} P \\ & \mid \forall x : A.P \mid \exists x : A.P \mid \forall_{\mathsf{rv}} X : \mathcal{A}.P \mid \exists_{\mathsf{rv}} X : \mathcal{A}.P \\ & \mid E \sim \pi \mid \mathbb{E}[E] = e \mid \mathsf{own}\, E \mid (x \overset{\pi}{\leftarrow} E \mid P) \mid \mathsf{L}(e) \\ & \mid \{P\}\, M\, \{X : \mathcal{A}.P\} \end{aligned}$$

where π, E, e, M range over maps defined in Fig. 22, along with the definition of a well-typed proposition, i.e. a formula P is *well-typed* under $\Gamma; \Delta$ if $\Gamma; \Delta \vdash P$.

$$\text{T-RandExpr}\ \frac{E \in [\![\Gamma]\!] \to \mathbf{Meas}([\![\Delta]\!], \mathcal{A})}{\Gamma; \Delta \vdash_{\mathsf{re}} E : \mathcal{A}} \qquad \text{T-DetExpr}\ \frac{e \in [\![\Gamma]\!] \to A}{\Gamma \vdash_{\mathsf{de}} e : A} \qquad \text{T-Meas}\ \frac{\pi \in [\![\Gamma]\!] \to M_\sigma(\mathcal{A})}{\Gamma \vdash_{\mathsf{meas}} \pi : \mathcal{A}} \qquad \text{T-True}\ \frac{}{\Gamma; \Delta \vdash \top}$$

$$\text{T-False}\ \frac{}{\Gamma; \Delta \vdash \bot} \qquad \text{T-BinModality}\ \frac{\Gamma; \Delta \vdash P \qquad \Gamma; \Delta \vdash Q}{\Gamma; \Delta \vdash P \odot Q} \qquad \text{T-DetQuant}\ \frac{\Gamma, x : A; \Delta \vdash P}{\Gamma; \Delta \vdash \mathcal{Q} x : A.P} \qquad \text{T-RandQuant}\ \frac{\Gamma; \Delta, X : \mathcal{A} \vdash P}{\Gamma; \Delta \vdash \mathcal{Q}_{\mathsf{rv}} X : \mathcal{A}.P}$$

$$\text{T-Expectation}\ \frac{\Gamma; \Delta \vdash_{\mathsf{re}} E : (\mathbb{R}, \mathcal{B}\mathbb{R}) \qquad \Gamma; \Delta \vdash_{\mathsf{de}} e : \mathbb{R}}{\Gamma; \Delta \vdash \mathbb{E}[E] = e} \qquad \text{T-Dist}\ \frac{\Gamma; \Delta \vdash_{\mathsf{re}} E : \mathcal{A} \qquad \Gamma \vdash_{\mathsf{meas}} \pi : \mathcal{A}}{\Gamma; \Delta \vdash E \sim \pi}$$

$$\text{T-Ownership}\ \frac{\Gamma; \Delta \vdash_{\mathsf{re}} E : \mathcal{A}}{\Gamma; \Delta \vdash \mathsf{own}\, E} \qquad \text{T-Conditioning}\ \frac{\Gamma; \Delta \vdash_{\mathsf{re}} E : \mathcal{A} \qquad \Gamma \vdash_{\mathsf{meas}} \pi : \mathcal{A} \qquad \Gamma, x : A; \Delta \vdash P}{\Gamma; \Delta \vdash x \overset{\pi}{\leftarrow} E \mid P} \qquad \text{T-Likelihood}\ \frac{\Gamma \vdash_{\mathsf{de}} e : [0, \infty)}{\Gamma; \Delta \vdash \mathsf{L}(e)}$$

$$\text{T-Hoare}\ \frac{\Gamma; \Delta \vdash P \qquad \Gamma; \Delta \vdash_{\mathsf{prog}} M : \mathcal{A} \qquad \Gamma; \Delta, X : \mathcal{A} \vdash Q}{\Gamma; \Delta \vdash \{P\}\, M\, \{X : \mathcal{A}.Q\}}$$

where $\odot \in \{\wedge, \vee, \Rightarrow, *, \mathrel{-\!\!*}\}$, $\mathcal{Q} \in \{\forall, \exists\}$ and $M_\sigma(\mathcal{A})$ is the set of σ-finite measures over $\mathcal{A}$.

Fig. 28. Typing judgements for BaSL propositions

Lemma D.2 (Kripke monotonicity). *Let $\gamma \in [\![\Gamma]\!]$ and $D \in [\![\Delta]\!]$. If $(\gamma, D, m) \vDash P$ and $m \sqsubseteq m'$, then $(\gamma, D, m') \vDash P$.*

Proof. By structural induction on P. The cases $\top$, $\bot$, $\wedge$, $\Rightarrow$, $\vee$, $*$, $-\!\!*$, $\forall$, $\exists$, $\forall_{\mathsf{rv}}$, $\exists_{\mathsf{rv}}$ are straightforward. We focus on the probabilistic propositions. Let $(\mathcal{F}, \mu) := m$ and $(\mathcal{F}', \mu') := m'$. When $P = E \sim \mathbb{P}$, $P = \mathsf{rv}\ E$, or $P = \mathbb{E}[E] = e$, notice that having a finer σ-algebra $\mathcal{F}'$ preserves measurability and the integral, and consequently the distribution as well. When $P = \mathsf{L}(e)$, $\mu'(\Omega) = \mu(\Omega)$ by definition of $\sqsubseteq$. When $P = \{P'\}M\{X.Q\}$, as the semantics does not depend on m, it holds for m' as well. When $P = x \xleftarrow{\pi} E \,|\, P'$, any Borel extension μ'^{+} of μ' is a Borel extension of μ, and consequently $X_{*}\mu'^{+}$ is absolutely continuous with respect to π. ◻

Definition D.3 (§B.4.1, [Li et al. 2023]). A *substitution* from (Γ, Δ) to (Γ', Δ') is a pair (s, S) where $s : [\![\Gamma']\!] \to [\![\Gamma]\!]$ and $S \in \mathbf{Meas}([\![\Delta']\!], [\![\Delta]\!])$. Given a well-formed BaSL formula $\Gamma; \Delta \vdash P$, we define the *syntactic substitution* $\Gamma'; \Delta' \vdash P[s, S]$ inductively as follows:

S-Form

$$\frac{s \in [\![\Gamma']\!] \to [\![\Gamma]\!] \qquad S \in \mathbf{Meas}([\![\Delta']\!], [\![\Delta]\!])}{\Gamma'; \Delta' \vdash (s, S) : \Gamma; \Delta} \qquad ([\![\Gamma']\!], [\![\Delta']\!]) \xrightarrow{(s,S)} ([\![\Gamma]\!], [\![\Delta]\!])$$

S-RandExpr

$$\frac{\Gamma; \Delta \vdash_{\mathsf{re}} E : \mathcal{A} \qquad \Gamma'; \Delta' \vdash (s, S) : \Gamma; \Delta}{\Gamma'; \Delta' \vdash_{\mathsf{re}} E[s, S] : \mathcal{A}} \qquad E[s, S](\gamma) := \quad [\![\Delta']\!] \xrightarrow{S} [\![\Delta]\!] \xrightarrow{E(s(\gamma))} \mathcal{A}$$

S-DetExpr

$$\frac{\Gamma \vdash_{\mathsf{de}} e : A \qquad s \in [\![\Gamma']\!] \to [\![\Gamma]\!]}{\Gamma' \vdash_{\mathsf{de}} e[s] : A} \qquad e[s] := \quad [\![\Gamma']\!] \xrightarrow{s} [\![\Gamma]\!] \xrightarrow{e} A$$

S-Meas

$$\frac{\Gamma \vdash_{\mathsf{meas}} \pi : \mathcal{A} \qquad s \in [\![\Gamma']\!] \to [\![\Gamma]\!]}{\Gamma' \vdash_{\mathsf{meas}} \pi[s] : \mathcal{A}} \qquad \pi[s] := \quad [\![\Gamma']\!] \xrightarrow{s} [\![\Gamma]\!] \xrightarrow{\pi} M_{\sigma}(\mathcal{A})$$

S-Prog

$$\frac{\Gamma; \Delta \vdash_{\mathsf{prog}} M : \mathcal{A} \qquad \Gamma'; \Delta' \vdash (s, S) : \Gamma; \Delta}{\Gamma'; \Delta' \vdash_{\mathsf{prog}} M[s, S] : \mathcal{A}} \quad M[s, S](\gamma) := \quad ([\![\Gamma']\!], [\![\Delta']\!]) \xrightarrow{1 \times S} ([\![\Gamma']\!], [\![\Delta]\!]) \xrightarrow{M(s(\gamma))} ([\![\Gamma]\!], [\![\Delta]\!])$$

S-Prop

$$\frac{\Gamma; \Delta \vdash P \qquad \Gamma'; \Delta' \vdash (s, S) : \Gamma; \Delta}{\Gamma; \Delta \vdash P[s, S]} \qquad P[s, S] \text{ is defined inductively below}$$

$$\begin{aligned}
\top[s, S] &:= \top \\
\bot[s, S] &:= \bot \\
(E \sim \pi)[s, S] &:= E[s, S] \sim \pi[s] \\
(\mathsf{own}\, E)[s, S] &:= \mathsf{own}\, E[s, S] \\
(\mathsf{L}(e))[s, S] &:= \mathsf{L}(e[s])
\end{aligned}
\qquad
\begin{aligned}
(P \mathcal{B} Q)[s, S] &:= P[s, S] \mathcal{B} Q[s, S] \\
(\mathcal{Q}x : A.P)[s, S] &:= \mathcal{Q}x : A.P[s \times 1_A, S] \\
(\mathcal{Q}_{\mathsf{rv}}X : \mathcal{A}.P)[s, S] &:= \mathcal{Q}_{\mathsf{rv}}X : A.P[s, S \times 1_{\mathcal{A}}] \\
(\mathbb{E}[E] = e)[s, S] &:= \mathbb{E}[E[s, S]] = e[s]
\end{aligned}$$

$$(\{P\}\, M\, \{X : \mathcal{A}.Q\})[s, S] := \{P[s, S]\}\, M[s, S]\, \{X : \mathcal{A}.Q[s, S \times 1_{\mathcal{A}}]\}$$

$$(x \xleftarrow{\pi} E : \mathcal{A} \mid P)[s, S] := x \xleftarrow{\pi[s]} E[s, S] \mid P[s \times 1_A, S]$$

where $\mathcal{B} \in \{\wedge, \vee, \to, *, -\!\!*\}$ and $\mathcal{Q} \in \{\forall, \exists\}$.

Lemma D.4 (substitution). *Let (γ, D, m) be a configuration and P a well-formed proposition. Then*

$$(\gamma, D, m) \vDash P[s, S] \iff (s(\gamma), S \circ D, m) \vDash P.$$

Proof. Most cases follow from structural induction, we focus on the case for Hoare triple $\{P\}\, M\, \{X : \mathcal{A}.Q\}$. The proof for $\{P\}\, M\, \{X : \mathcal{A}.Q\}$ is similar to the proof in Lilac (see Li et al.

[2023, Lemma B.10]), except for the 'normalising' assumption, which is equivalent as both sides yield the following condition:

$$\int_{\Omega} [\![M_{s(\gamma)}]\!](S(D(\omega)), A)\, \mu_0^+(\mathrm{d}\omega) > 0.$$

The rest of the proof follows from instantiating the extension D'_{ext} to (D_{ext}, D). □

Proposition D.5 (entailments). *Define semantic entailment $P \vdash Q$ to mean $(\gamma, D, m) \vDash P$ implies $(\gamma, D, m) \vDash Q$ for all (γ, D, m). The following entailments are sound:*

E-∗-Ident $P * \top_1 \vdash P$

E-∗-Comm $P * Q \dashv\vdash Q * P$

E-∗-Assoc $(P * Q) * R \dashv\vdash P * (Q * R)$

E-True $\dfrac{\mathsf{affine}(P)}{P \vdash \top_1}$

E-False $\mathsf{L}(0) \vdash \bot$

E-∗-Weak₁ $\dfrac{\mathsf{affine}(Q)}{P * Q \vdash P}$

E-∗-Weak₂ $\dfrac{\mathsf{affine}(P)}{P * Q \vdash Q}$

E-∗-Weak $\dfrac{\mathsf{affine}(P) \qquad \mathsf{affine}(Q)}{P * Q \vdash P \wedge Q}$

E-∗-Cong $\dfrac{P \vdash Q}{P * R \vdash Q * R}$

E-∗-AdjointL $\dfrac{P * Q \vdash R}{P \vdash Q \twoheadrightarrow R}$

E-∗-AdjointR $\dfrac{P \vdash Q \twoheadrightarrow R}{P * Q \vdash R}$

E-∧-Elim₁ $P \wedge Q \vdash P$

E-∧-Elim₂ $P \wedge Q \vdash Q$

E-∨-Intro₁ $P \vdash P \vee Q$

E-∨-Intro₂ $Q \vdash P \vee Q$

E-NormConst $\mathsf{NormConst} * \mathsf{NormConst} \vdash \mathsf{NormConst}$

Proof. We focus on rules that involve affine propositions. The rule E-True follows from definition of an affine proposition – P being affine and $(\gamma, D, (\mathcal{F}, \mu)) \vDash P$ imply $\mu(\Omega) = 1$, which satisfies of $\top_1 = \mathsf{L}(1)$. For E-∗-Weak₁, we assume $(\gamma, D, m) \vDash P * Q$ and $m_1 \bullet m_2 \sqsubseteq m$ such that $(\gamma, D, m_1) \vDash P$ and $(\gamma, D, m_2) \vDash Q$. Notice m_2 being a probability space implies $m_1 \sqsubseteq m_2$ – by definition of independent combination, we have $\mu(F_1) = \mu(F_1 \cap \Omega) = \mu_1(F_1)\mu_2(\Omega) = \mu_1(F_1)$ for any $F_1 \in \mathcal{F}_1$. Then, notice Kripke monotonicity holds (Lemma D.2), which means $(\gamma, D, m_1) \vDash P$ implies $(\gamma, D, m) \vDash P$. The proof for E-∗-Weak₂ is similar. For E-∗-Weak, it follows directly from E-∗-Weak₁ and E-∗-Weak₂.

For E-NormConst, suppose $(\gamma, D, (\mathcal{F}, \mu)) \vDash \exists k : (0, \infty).\mathsf{L}(k) * \exists k' : (0, \infty).\mathsf{L}(k')$, then there exists $m_1 = (\mathcal{F}_1, \mu_1), m_2 = (\mathcal{F}_2, \mu_2) \in \mathcal{M}$ such that $m_1 \bullet m_2 \sqsubseteq m$ and $(\gamma, D, (\mathcal{F}, m_1)) \vDash \mathsf{L}(k)$ for some $k \in (0, \infty)$ and $(\gamma, D, (\mathcal{F}, m_2)) \vDash \mathsf{L}(k')$ for some $k' \in (0, \infty)$. This implies $\mu_1(\Omega) = k$ and $\mu_2(\Omega) = k'$. By definition of independent combination, we know $\mu(\Omega) = \mu(\Omega \cap \Omega) = \mu_1(\Omega)\mu_2(\Omega) = k \cdot k'$, which means $(\gamma, D, (\mathcal{F}, \mu)) \vDash \mathsf{L}(k \cdot k')$, implying NormConst. □

Lemma D.6 ([Chang and Pollard 1997, Theorem 2]). *Let $\{\mu_x\}_{x \in A}$ be an (X, π)-disintegration of $\mu : \Sigma_\Omega \longrightarrow [0, \infty]$. Then $X_*\mu$ is absolutely continuous with respect to π and for π-almost-all $x \in A$,*

$$\frac{\mathrm{d}X_*\mu(x)}{\mathrm{d}\pi} = \mu_x(\Omega).$$

Also, if $\pi = X_*\mu$, then μ_x is a probability measure for π-almost-all $x \in A$.

Theorem D.7 (Bayes' theorem). *The following entailment is sound:*

E-Bayes $E \sim f \cdot \pi \dashv\vdash x \overset{\pi}{\leftarrow} E \mid \mathsf{L}(f(x))$

Proof. Let (γ, D, m) be a configuration, $m = (\mathcal{F}, \mu)$ and write $X(\omega) := E(\gamma)(D(\omega))$. We prove the entailment in two directions. From left to right, we assume $(\gamma, D, (\mathcal{F}, \mu)) \vDash E \sim f \cdot \pi$, then $(\gamma, D, (\mathcal{F}, \mu)) \vDash \mathsf{own}\ E$ by definition of $\vDash$. Next, for any Borel extension μ^+ and any (X, π)-disintegration $\{\mu_x^+\}_{x \in A}$, we know $(\gamma, D, (\mathcal{F}, \mu_x^+|_{\mathcal{F}})) \vDash \mathsf{L}(f(x))$ holds for π-almost-all $x \in A$ because for any $F \in \Sigma_A$,

$$\begin{aligned}
\int_F \mu_x^+(\Omega)\, \pi(\mathrm{d}x) &= \int_F \frac{\mathrm{d}X_*\mu(x)}{\mathrm{d}\pi}\, \pi(\mathrm{d}x) && \text{(Lemma D.6)} \\
&= \int_F \frac{\mathrm{d}(f \cdot \pi)(x)}{\mathrm{d}\pi}\, \pi(\mathrm{d}x) && ((\gamma, D, m) \vDash E \sim f \cdot \pi) \\
&= \int_F f(x)\, \pi(\mathrm{d}x). && (\tfrac{\mathrm{d}(f \cdot \pi)}{\mathrm{d}\pi} = f \text{ almost surely})
\end{aligned}$$

Also, $X_*\mu = f \cdot \pi$ is absolutely continuous with respect to π. Let F be a π-null-set, then $(f \cdot \pi)(F) = 0$. From right to left, we assume $(\gamma, D, (\mathcal{F}, \mu)) \vDash x \overset{\pi}{\leftarrow} E \mid \mathsf{L}(f(x))$, then for any Borel extension μ^+ and (X, π)-disintegration $\{\mu_x^+\}_{x \in A}$, we know the following statement holds for π-almost-all $x \in A$:

$$(\gamma, D, (\mathcal{F}, \mu_x^+|_{\mathcal{F}})) \vDash \mathsf{L}(f(x)).$$

This implies $\mu_x^+(\Omega) = f(x)$ holds for π-almost-all $x \in A$, which then implies for all $F \in \Sigma_A$,

$$\begin{aligned}
X_*\mu(F) &= X_*\mu^+(F) && \text{(Borel extension)} \\
&= \int_\Omega \mathbf{1}_{X^{-1}(F)}\, \mathrm{d}\mu^+ && \text{(by definition)} \\
&= \int_A \mu_x(X^{-1}(F))\, \pi(\mathrm{d}x) && ((X, \pi)\text{-disintegration}) \\
&= \int_F \mu_x(\Omega)\, \pi(\mathrm{d}x) && (\mathbf{1}_F(x)\mu_x(\Omega) = \mu_x(X^{-1}(E))) \\
&= \int_F f\, \mathrm{d}\pi && ((\gamma, D, (\mathcal{F}, \mu_x^+|_{\mathcal{F}})) \vDash \mathsf{L}(f(x)) \text{ a.e.}) \\
&= (f \cdot \pi)(F), && \text{(by definition)}
\end{aligned}$$

which yields $(\gamma, D, (\mathcal{F}, \mu)) \vDash E \sim f \cdot \pi$. □

E PROOF RULES

THEOREM E.1 (SOUNDNESS). *The following Hoare triples are sound, i.e. for all proof rules defined below,* $\vdash \{P\}\ M\ \{X : \mathcal{A}.Q\}$ *implies* $\vDash \{P\}\ M\ \{X : \mathcal{A}.Q\}$.

H-Sample
$$\{\top_1\}\ \mathbf{sample}(\mathbb{P})\ \{X : \mathbb{R}.X \sim \mathbb{P}\}$$

H-CondSample
$$\{X \sim \mathbb{P}\}\ \mathbf{sample}(p(X))\ \{Y : \mathbb{R}.x \overset{\mathbb{P}}{\leftarrow} X \mid Y \sim p(x)\}$$

H-Score
$$\{X \sim \pi\}\ \mathbf{score}(f(X))\ \{X \sim f \cdot \pi\}$$

H-CondScore
$$\{x \overset{\pi}{\leftarrow} X \mid Y \sim p(x)\}\ \mathbf{score}(f(Y))\{x \overset{\pi}{\leftarrow} X \mid Y \sim f \cdot p(x)\}$$

H-Observe
$$\{X \sim \pi\}\ \mathbf{observe}(P(X))\ \{X \sim \ell_P \cdot \pi\}$$

H-CondObserve
$$\{x \overset{\pi}{\leftarrow} X \mid Y \sim p(x)\}\ \mathbf{observe}(P(Y))\ \{x \overset{\pi}{\leftarrow} X \mid Y \sim \ell_P \cdot p(x)\}$$

H-Return
$$\{Q[\llbracket M \rrbracket/X]\}\ \mathbf{return}\ M\ \{X : \mathcal{A}.Q\}$$

H-Frame
$$\frac{\vdash \{P\}\ M\ \{X : \mathcal{A}.Q\}}{\vdash \{P * F\}\ M\ \{X : \mathcal{A}.Q * F\}}\ (X \notin \mathsf{fv}(F))$$

H-Let
$$\frac{\vdash \{P\}\ M\ \{X : \mathcal{A}.Q\} \qquad \vdash \forall_{\mathrm{rv}} X : \mathcal{A}.\ \{Q\}\ N\ \{Y : \mathcal{B}.R\}}{\vdash \{P\}\ \mathbf{let}\ X = M\ \mathbf{in}\ N\ \{Y : \mathcal{B}.R\}}$$

H-Cons
$$\frac{P' \vdash P \qquad \vdash \{P\}\ M\ \{X : \mathcal{A}.Q\} \qquad Q \vdash Q'}{\vdash \{P'\}\ M\ \{X : \mathcal{A}.Q'\}}$$

.

LEMMA E.2. *Let* $(\mathcal{F}, \mu) \in \mathcal{M}$ *such that* $m \in \mathbb{N}$ *witnesses its finite footprint. For any* $n \geq m$, *define* $\mathcal{F}|_n := \pi_{1..n}[\mathcal{F}] = \{\{\omega_{1..n} \mid \omega \in F\} \mid F \in \mathcal{F}\}$ *and* $\mu|_n : \mathcal{F}|_n \to [0, \infty]$ *by* $\mu|_n(F) := \mu(F \times \Omega)$. *Then:*

1. $\mathcal{F}|_n$ *is a sub-*σ*-algebra of* $\mathcal{B}[0,1]^n$,
2. *the map* $\pi_{1..n} : \Omega \to [0,1]^n$ *is* $(\mathcal{F}, \mathcal{F}|_n)$*-measurable,*
3. $\mu|_n = \pi_{1..n*}\mu$, *and*
4. *for any standard Borel* (A, Σ_A) *and* $(\mathcal{F}, \Sigma_A)$*-measurable* $f : \Omega \to A$, *the map* $f|_n : [0,1]^n \to A$ *defined by* $f|_n(\omega_{1..n}) := f(\omega_{1..n}, x)$ *for some* $x \in \Omega$ *is* $(\mathcal{F}|_n, \Sigma_A)$*-measurable. Also, for all* $\omega_{1..n} \in [0,1]^n$, $x, y \in \Omega$, $f(\omega_{1..n}, x) = f(\omega_{1..n}, y)$.

PROOF. For (1), we must first show that $\mathcal{F}|_n$ is a σ-algebra. Since $\emptyset = \emptyset \times \Omega \in \mathcal{F}$, $\emptyset \in \mathcal{F}|_n$. Next, for every $F' \in \mathcal{F}|_n$, we know $F' \times \Omega \in \mathcal{F}$, which implies $\Omega \setminus (F' \times \Omega) = ([0,1]^n \setminus F') \times \Omega$ and $[0,1]^n \setminus F' \in \mathcal{F}|_n$. For $\{F'_i \in \mathcal{F}|_n\}_{i \in \mathbb{N}}$, notice $\biguplus_{i \in \mathbb{N}}(F'_i \times \Omega) = (\biguplus_{i \in \mathbb{N}} F'_i) \times \Omega$, which implies $\biguplus_{i \in \mathbb{N}} F'_i \in \mathcal{F}|_n$.

For (2), suppose $F' \in \mathcal{F}|_n$, then $\pi^{-1}_{1..n}(F') = \{\omega \in \Omega \mid \omega_{1..n} \in F'\} = \{\omega \in \Omega \mid \omega \in F' \times \Omega\} \in \mathcal{F}$.

For (3), since $\pi_{1..n} : \Omega \to [0,1]^n$ is $(\mathcal{F}, \mathcal{F}|_n)$-measurable, the pushforward measure $\pi_{1..n*}\mu : \mathcal{F}|_n \to [0, \infty]$ is well-defined. Moreover, $\mu|_n(F') = \mu(F' \times \Omega) = \mu(\pi^{-1}_{1..n}(F')) = \pi_{1..n*}\mu(F')$.

For (4), notice that the domain of f is measurably isomorphic to the product with $(\Omega, \mathcal{F}) \cong ([0,1]^n, \mathcal{F}|_n) \times (\Omega, \{\emptyset, \Omega\})$. Recall that the slice of a product measurable function is measurable, which means $\omega_{1..n} \mapsto f(\omega_{1..n}, x)$ is $(\mathcal{F}|_n, \Sigma_A)$-measurable for any $x \in \Omega$. Next, assume by contradiction that there is some $y \in A$ and $\gamma \in [0,1]^n$ such that $x \neq y$ and $f(\gamma, x) \neq f(\gamma, y)$. Observe that $f(\gamma, -) : \Omega \to A$ is $(\{\emptyset, \Omega\}, \Sigma_A)$-measurable by property of product σ-algebra. Since (A, Σ_A) is standard Borel, it contains all singleton sets and we have $E := f^{-1}(\gamma, \{f(\gamma, x)\}) \in \{\emptyset, \Omega\}$. Notice $x \in E$, which means $E = \Omega$ by definition of $(\{\emptyset, \Omega\}, \Sigma_A)$-measurable functions. This implies $y \in E$ and $f(\gamma, x) = f(\gamma, y)$, which is a contradiction. □

LEMMA E.3. *The following Hoare triple is sound:*

H-Sample
$$\vdash \{\top_1\}\ \mathbf{sample}(\mathbb{P})\ \{X : \mathbb{R}.X \sim \mathbb{P}\}$$

Proof. Let $m_{\text{pre}} := (\mathcal{F}_{\text{pre}}, \mu_{\text{pre}})$, $(\gamma, D, m_{\text{pre}})$ be a configuration, $(\gamma, D, m_{\text{pre}}) \vDash \top_1$ and $m_{\text{fr}} \in \mathcal{M}$ such that $(\mathcal{F}_0, \mu_0) := m_{\text{pre}} \bullet m_{\text{fr}}$ is defined. Consider a Borel measure μ_0^+ that extends $m_{\text{pre}} \bullet m_{\text{fr}}$ (i.e. $\mu_0^+|_{\mathcal{F}_0} = \mu_0$), a probabilistic context Δ_{ext} and its random variables $D_{\text{ext}} \in \mathsf{RV}[\![\Delta_{\text{ext}}]\!]$. Notice the condition regarding the program integrating to a positive number is always true because $\mu_0(\Omega)$ is non-zero by definition of $\mathcal{M}$:

$$\int_\Omega [\![\textbf{sample}\ \mathbb{P}]\!](D(\omega), A)\, \mu_0^+(\mathrm{d}\omega) = \int_\Omega 1\, \mu_0^+(\mathrm{d}\omega) = \mu_0^+(\mathrm{d}\omega) > 0.$$

Denote n the maximum dimension of the Hilbert cube used in m_{pre}, m_{fr}, D and D_{ext}, i.e. n is the maximum of the witness of their finite footprint property. We define the σ-algebra $\mathcal{F}_{\text{post}}$ and its measure μ_{post} by

$$\mathcal{F}_{\text{post}} := \mathcal{F}_{\text{pre}}|_n \otimes \mathcal{B}[0,1] \otimes \{\varnothing, \Omega\}$$
$$\mu_{\text{post}} := \mu_{\text{pre}}|_n \otimes \lambda_{[0,1]} \otimes p$$

where $p : \{\varnothing, \Omega\} \to [0,\infty]$ is the trivial probability measure with $p(\Omega) = 1$. We first check that the above definition is well-formed: if $\mathcal{F}$ is a σ-algebra with finite footprint m, then for any $n \geq m$, its restriction to n dimensions

$$\mathcal{F}|_n := \pi_{1..n}(\mathcal{F}) := \{\{\pi_{1..n}(\omega) \mid \omega \in F\} \mid F \in \mathcal{F}\}$$

is a σ-algebra as every dimension greater n can be written as $[0,1]^{\mathbb{N}}$. Similarly, the restriction $\mu|_n : \mathcal{F}|_n \to [0,\infty]$ is defined by

$$\mu|_n(F) := \mu(F \times [0,1]^{\mathbb{N}}).$$

By the randomisation lemma (Lemma E.5), there exists a $(\mathcal{B}[0,1], \mathcal{B}\mathbb{R})$-measurable function $f : [0,1] \to \mathbb{R}$ such that $f_*\lambda_{[0,1]} = \mathbb{P}(\gamma)$. Define $X \in \mathsf{RV}(\mathcal{A})$ by $X(\omega) := f(\omega_{n+1}) = (f \circ \pi_{n+1})(\omega)$, which is the composition of two measurable functions, and hence, is measurable. Notice X has finite footprint $n+1$ as the pullback σ-algebra is of the form

$$\begin{aligned} X^{-1}(\mathcal{B}\mathbb{R}) &= \{\pi_{n+1}^{-1}(f^{-1}(U)) \mid U \in \mathcal{B}\mathbb{R}\} \\ &= \{\{(\omega_1, \omega_2, \ldots) \in \Omega \mid \omega_{n+1} \in f^{-1}(U)\} \mid U \in \mathcal{B}\mathbb{R}\} \\ &= \{[0,1]^n \times f^{-1}(U) \times \Omega \mid U \in \mathcal{B}\mathbb{R}\}. \end{aligned}$$

Also, define $\mathcal{F}_1 := \langle \mathcal{F}_{\text{post}}, \mathcal{F}_{\text{fr}} \rangle$ and $\mu_1 : \mathcal{F}_1 \to [0,\infty]$ by $\mu_1 := \mu_0|_n \otimes \lambda_{[0,1]} \otimes p$. Notice $(\mathcal{F}_1, \mu_1)$ is well-defined: since n witnesses the finite footprint of $\mathcal{F}_{\text{fr}}$, $\mu_0|_n$ is a well-defined measure on $\mathcal{F}_0|_n$, which is a σ-algebra of $[0,1]^n$. Additionally, $(\mathcal{F}_1, \mu_1) := m_{\text{post}} \bullet m_{\text{fr}}$ because for all $F_{\text{post}} \in \mathcal{F}_{\text{post}}$ and $F_{\text{fr}} \in \mathcal{F}_{\text{fr}}$, we have $F_{\text{post}} = F_{\text{post}}|_{n+1} \times [0,1]^{\mathbb{N}}$, $F_{\text{fr}} = F_{\text{fr}}|_n \times [0,1]^{\mathbb{N}}$ and

$$\begin{aligned} \mu_1(F_{\text{post}} \cap F_{\text{fr}}) &= (\mu_0|_n \otimes \lambda_{[0,1]} \otimes p)(F_{\text{post}} \cap F_{\text{fr}}) \\ &= (\mu_{\text{pre}}|_n \otimes \lambda_{[0,1]} \otimes p)(F_{\text{post}}) \cdot \mu_{\text{fr}}(F_{\text{fr}}|_{n+1}) \\ &= \mu_{\text{post}}(F_{\text{post}}) \cdot \mu_{\text{fr}}(F_{\text{fr}}). \end{aligned}$$

Also, $(\gamma, (D, X), m_{\text{post}}) \vDash X \sim \mathbb{P}$: notice that $X = f \circ \pi_{n+1}$ is $(\mathcal{F}_{\text{post}}, \Sigma_A)$-measurable because π_{n+1} is $(\mathcal{F}_{\text{post}}, \mathcal{B}[0,1])$ by construction of the product measure and f is $(\mathcal{B}[0,1], \Sigma_A)$-measurable by assumption, and for every $E \in \Sigma_A$, we have

$$\begin{aligned} X_*\mu_{\text{post}}(E) &= \mu_{\text{post}}(\{\omega \in \Omega \mid \omega_{n+1} \in f^{-1}(E)\}) \\ &= \mu_{\text{post}}([0,1]^n \times f^{-1}(E) \times [0,1]^{\mathbb{N}}) \\ &= \mu_{\text{pre}}([0,1]^{\mathbb{N}}) \cdot f_*\lambda_{[0,1]}(E) \\ &= \mathbb{P}(\gamma)(E). && (\mu_{\text{pre}}([0,1]^{\mathbb{N}}) = 1 \text{ by assumption}) \end{aligned}$$

Finally, consider a Borel measure $\mu_1^+ : \Sigma_\Omega \to [0,\infty]$ that satisfies $\mu_1^+|_{\mathcal{F}_1} = \mu_1$ (which is guaranteed to exist by Proposition C.10), then the following identity holds for all $U \in \Sigma_{[\![\Delta_{\text{ext}}]\!]} \otimes \Sigma_{[\![\Delta]\!]} \otimes \Sigma_A$:

$$\begin{aligned}
&\int_\Omega \int_\mathbb{R} \mathbf{1}_U(D_{\text{ext}}(\omega), D(\omega), x)\, [\![\mathbf{sample}(\mathbb{P}_\gamma)]\!](D(\omega), \mathrm{d}x)\, \mu_0^+(\mathrm{d}\omega) \\
&= \int_\Omega \int_\mathbb{R} \mathbf{1}_U(D_{\text{ext}}(\omega), D(\omega), x)\, \mathbb{P}_\gamma(\mathrm{d}x)\, \mu_0^+(\mathrm{d}\omega) && \text{(semantics of } \mathbf{sample}\text{)} \\
&= \int_\Omega \int_\mathbb{R} \mathbf{1}_U(D_{\text{ext}}|_n(\pi_{1..n}(\omega)), D|_n(\pi_{1..n}(\omega)), x)\, \mathbb{P}_\gamma(\mathrm{d}x)\, \mu_0^+(\mathrm{d}\omega) && \text{(Lemma E.2)} \\
&= \int_{[0,1]^n} \int_{[0,1]} \mathbf{1}_U(D_{\text{ext}}|_n, D|_n, f(x))\, \lambda_{[0,1]}(\mathrm{d}x)\, \mathrm{d}\mu_0|_n && \text{(Lemma E.2)} \\
&= \int_{[0,1]^{n+1}} \mathbf{1}_U(D_{\text{ext}}|_n, D|_n, f \circ \pi_{n+1}) \circ \pi_{1..n+1}\, \mathrm{d}(\mu_0|_n \otimes \lambda_{[0,1]}) && \text{(Fubini)} \\
&= \int_\Omega \mathbf{1}_U(D_{\text{ext}}, D, X)\, \mathrm{d}\mu_1^+, && \text{(by definition of } \mu_1\text{)}
\end{aligned}$$

which yields the postcondition $X \sim \mathbb{P}$. □

Lemma E.4. *The following Hoare triple is sound:*

$$\frac{\textit{H-Score}}{\vdash \{X \sim \pi\}\ \mathbf{score}(f(X))\ \{X \sim f \cdot \pi\}}$$

Proof. Let $(\gamma, (D, X), \_)$ be a configuration, $m_{\text{pre}} \in \mathcal{M}$ such that $(\gamma, (D, X), m_{\text{pre}}) \vDash X \sim \pi$, and all m_{fr} with $(\mathcal{F}_0, \mu_0) := m_{\text{pre}} \bullet m_{\text{fr}}$ defined, and all measures μ_0^+ satisfying $\mu_0^+|_{\mathcal{F}_0} = \mu_0$, and all probabilistic contexts Δ_{ext} and all $D_{\text{ext}} \in \mathrm{RV}[\![\Delta_{\text{ext}}]\!]$. We assume

$$\int_\Omega [\![f(X)]\!](D(\omega), 1)\, \mu_0^+(\mathrm{d}\omega) = \int_\Omega f(X(\omega))\, \mu_0^+(\mathrm{d}\omega) > 0,$$

Define $Y \in \mathrm{RV}(\mathbf{1})$ by $Y(\omega) := *$, $\mathcal{F}_{\text{post}} = \mathcal{F}_{\text{pre}}$ and $\mu_{\text{post}} : \mathcal{F}_{\text{post}} \to [0,\infty]$ by

$$\mu_{\text{post}}(F) := \int_F f(X(\omega))\, \mu_{\text{pre}}(\mathrm{d}\omega).$$

Notice m_{post} is well-defined: $f \circ X : \Omega \to [0,\infty]$ is $(\mathcal{F}_{\text{pre}}, \mathcal{B}[0,\infty])$-measurable since X is assumed to be $(\mathcal{F}_{\text{pre}}, \Sigma_A)$-measurable via $(\gamma, D, m_{\text{pre}}) \vDash X \sim \pi$ and f is a $(\Sigma_A, \mathcal{B}[0,\infty])$-measurable function. Moreover, there is a measure $\mu_1 : \langle \mathcal{F}_{\text{post}}, \mathcal{F}_{\text{fr}} \rangle \to [0,\infty]$ defined by

$$\mu_1(F) := \int_F f(X(\omega))\, \mu_0(\mathrm{d}\omega)$$

such that μ_1 witnesses the independent combination of m_{post} and m_{fr}: for all $F_1 \in \mathcal{F}_{\text{post}} = \mathcal{F}_{\text{pre}}$ and $F_2 \in \mathcal{F}_{\text{fr}}$, we have

$$\begin{aligned}
\mu_1(F_1 \cap F_2) &= \int_{F_1 \cap F_2} f \circ X\, \mathrm{d}\mu_0 && \text{(definition of } \mu_1\text{)} \\
&= \int_\Omega \mathbf{1}_{F_1} \cdot (f \circ X) \cdot \mathbf{1}_{F_2}\, \mathrm{d}\mu_0 && (\mathbf{1}_{F_1 \cap F_2} = \mathbf{1}_{F_1} \cdot \mathbf{1}_{F_2}) \\
&= \int_\Omega \mathbf{1}_{F_1} \cdot (f \circ X)\, \mathrm{d}\mu_{\text{pre}} \cdot \int_\Omega \mathbf{1}_{F_2}\, \mathrm{d}\mu_{\text{fr}} && \text{(Lemma C.15)} \\
&= \mu_{\text{post}}(F_1) \cdot \mu_{\text{fr}}(F_2). && \text{(definition of } \mu_{\text{post}}\text{)}
\end{aligned}$$

We now show that $(\gamma, (D, X, Y), m_{\text{post}}) \vDash x \overset{\pi}{\leftarrow} X \mid (\mathsf{L}(f(x)))$: write $\pi_\gamma := \pi(\gamma)$, π_γ is a σ-finite measure by assumption, and $X_* \mu_{\text{post}}$ is absolutely continuous with respect to π_γ because for every $E \in \Sigma_A$ such that $\pi_\gamma(E) = 0$, we know

$$
\begin{aligned}
X_* \mu_{\text{post}}(E) &= \mu_{\text{post}}(X^{-1}(E)) \\
&= \int_{X^{-1}(E)} f \circ X \, \mathrm{d}\mu_{\text{pre}} && \text{(definition of } m_{\text{post}}) \\
&= \int_E f \, \mathrm{d}X_* \mu_{\text{pre}} && \text{(change of variable)} \\
&= \int_E f \, \mathrm{d}\pi_\gamma && ((\gamma, D, m_{\text{pre}}) \vDash X \sim \pi) \\
&= 0. && (\textstyle\int_E \mathrm{d}\pi_\gamma = 0)
\end{aligned}
$$

Consider an arbitrary Borel measure extension $\mu^+_{\text{post}} : \mathcal{B}[0, \infty] \to [0, \infty]$ and a disintegration $\{\mu^+_{\text{post},x}\}_{x \in A}$ of μ^+_{post} along X and π_γ. Notice, via the above derivation, that $X_* \mu_{\text{post}}$ has a density f with respect to π_γ. By Lemma D.6, we know for π-almost-all $x \in A$,

$$
\mu^+_{\text{post},x}|_{\mathcal{F}_{\text{post}}}(\Omega) = \mu^+_{\text{post},x}(\Omega) = \frac{\mathrm{d}X_* \mu^+_{\text{post}}}{\mathrm{d}\pi_\gamma}(x) = \frac{\mathrm{d}X_* \mu_{\text{post}}}{\mathrm{d}\pi_\gamma}(x) = f(x).
$$

Hence, $((\gamma, x), (D, Y), m_{\text{post},x}) \vDash \mathsf{L}(f(x))$ and $(\gamma, (D, Y), m_{\text{post}}) \vDash x \overset{\pi}{\leftarrow} X \mid \mathsf{L}(f(x))$. Finally, consider a Borel measure $\mu^+_1 : \mathcal{B}[0, \infty] \to [0, \infty]$ defined by

$$
\mu^+_1(U) := \int_U f \circ X \, \mathrm{d}\mu^+_0.
$$

Notice that $\mu^+_1|_{\mathcal{F}_1}(F) = \int_F f \circ X \, \mathrm{d}\mu^+_0 = \int_F f \circ X \, \mathrm{d}\mu_0 = \mu_1(F)$ and $f \circ X$ is the density of μ^+_1 with respect to μ^+_0. Hence, the following identity holds:

$$
\begin{aligned}
&\int_\Omega \int_{\mathcal{A}} \mathbf{1}_U(D_{\text{ext}}(\omega), D(\omega), y) \, [\![\mathbf{score}(f(X))]\!](D(\omega), \mathrm{d}y) \, \mu^+_0(\mathrm{d}\omega) \\
&= \int_\Omega \mathbf{1}_U(D_{\text{ext}}(\omega), D(\omega), *) f(X(\omega)) \, \mu^+_0(\mathrm{d}\omega) && \text{(semantics of } \mathbf{score}) \\
&= \int_\Omega \mathbf{1}_U(D_{\text{ext}}(\omega), D(\omega), *) \frac{\mathrm{d}\mu^+_1(\omega)}{\mathrm{d}\mu^+_0} \, \mu^+_0(\mathrm{d}\omega) && \text{(density)} \\
&= \int_\Omega \mathbf{1}_U(D_{\text{ext}}(\omega), D(\omega), *) \, \mu^+_1(\mathrm{d}\omega) \\
&= \int_\Omega \mathbf{1}_U(D_{\text{ext}}(\omega), D(\omega), Y(\omega)) \, \mu^+_1(\mathrm{d}\omega).
\end{aligned}
$$

Notice own X also holds under $(\gamma, (D, X, Y), m_{\text{post}})$. By Theorem D.7, we know $x \overset{\pi}{\leftarrow} X \mid \mathsf{L}(f(x)) \vdash X \sim f \cdot \pi$. Now, we apply the consequence rule H-Cons (Lemma E.18), which yields the desired triple. □

Lemma E.5 (randomisation; [Heunen et al. 2017, Proposition 24]). *Let* (A, Σ_A) *be a measurable space and* $\kappa : A \times \mathcal{B}\mathbb{R} \to [0, 1]$ *a probability kernel. Then there is a measurable function* $f : A \times \mathbb{R} \to \mathbb{R}$ *such that*

$$
\kappa(x, U) = \lambda_{[0,1]}\{r \in \mathbb{R} \mid f(x, r) \in U\}.
$$

Lemma E.6. *Let (A, Σ_A) and (B, Σ_B) be standard Borel spaces, $(\mathcal{F}, \mu) \in \mathcal{M}$ such that μ is a probability measure, $X : \Omega \to A$ an $(\mathcal{F}, \Sigma_A)$-measurable function and $Y : \Omega \to B$ an $(\mathcal{F}, \Sigma_B)$-measurable function. For any Borel extension $\mu^+ : \Sigma_\Omega \to [0, 1]$ and its $(X, X_*\mu^+)$-disintegration $\{\mu^+_x\}_{x\in A}$, X and Y are μ^+_x-independent for $X_*\mu^+$-almost-all $x \in A$. Moreover, if X and Y are μ-independent, then $Y_*\mu^+_x = Y_*\mu$ holds for $X_*\mu^+$-almost-all $x \in A$.*

Proof. Define $\pi := X_*\mu$. We need to show that for π-almost-all $x \in A$, $U \in \Sigma_A$ and $V \in \Sigma_B$, the following identity holds:

$$\mu^+_x\{X \in U, Y \in V\} = \mu^+_x\{X \in U\} \cdot \mu^+_x\{Y \in V\}.$$

Suppose $x \notin U$, then both L.H.S. and R.H.S. are zero because $\{X \in U\} = \{\omega \mid X(\omega) \in U\} \subseteq \{\omega \mid X(\omega) \neq x\} = \{X \neq x\}$, which means $\mu^+_x\{X \in U\} = 0$ by the concentration axiom. Suppose $x \in U$, then $\{X \in U\} = \{X = x\} \uplus \{X \in U \setminus \{x\}\}$. By additivity, we know

$$\mu^+_x\{X \in U\} = \mu^+_x\{X = x\} + \mu^+_x\{X \in U \setminus \{x\}\} = 1 + 0 = 1.$$

Next, compute that

$$\begin{aligned}
\mu^+_x\{X \in U, Y \in V\} &= \mu^+_x(X^{-1}(U) \cap Y^{-1}(V)) && \text{(by definition)}\\
&= \mu^+_x((X^{-1}(\{x\}) \uplus X^{-1}(U \setminus \{x\})) \cap Y^{-1}(V))\\
&= \mu^+_x((X^{-1}(\{x\}) \cap Y^{-1}(V)) \uplus (X^{-1}(U \setminus \{x\}) \cap Y^{-1}(V)))\\
&= \mu^+_x\{X = x, Y \in V\} + \mu^+_x\{X \in U \setminus \{x\}, Y \in V\} && \text{(additivity)}\\
&= \mu^+_x\{X = x, Y \in V\} && (\mu^+_x\{X \neq x\} = 0)\\
&= \mu^+_x\{X = x, Y \in V\} + \mu^+_x\{X \neq x, Y \in V\} && (\mu^+_x\{X \neq x\} = 0)\\
&= \mu^+_x\{Y \in V\} && \text{(additivity)}\\
&= \mu^+_x\{X \in U\} \cdot \mu^+_x\{Y \in V\}. && (\mu^+_x\{X \in U\} = 1)
\end{aligned}$$

Next, assuming X and Y are μ-independent, by the above derivation they must be μ^+_x-independent as well. Suppose $U \in \Sigma_A$, then $\mu^+_x\{X \in U\} = \mathbf{1}_U(x)$ and $\mu^+_-\{X \in U\} = \mathbf{1}_U$. This implies

$$\begin{aligned}
\int_U Y_*\mu^+_x(V)\, \pi(\mathrm{d}x) &= \int_A \mathbf{1}_U(x) \cdot \mu^+_x\{Y \in V\}\, \pi(\mathrm{d}x) && \text{(by definition)}\\
&= \int_A \mu^+_x\{X \in U\} \cdot \mu^+_x\{Y \in V\}\, \pi(\mathrm{d}x) && (\mu^+_x\{X \in U\} = 1)\\
&= \int_A \mu^+_x\{X \in U, Y \in V\}\, \pi(\mathrm{d}x) && \text{(see above)}\\
&= \mu^+\{X \in U, Y \in V\} && \text{(disintegration)}\\
&= \mu\{X \in U, Y \in V\} && (\mu^+ \text{ extends } \mu)\\
&= \mu\{X \in U\} \cdot \mu\{Y \in V\} && (\mu\text{-independence})\\
&= \int_U Y_*\mu(V)\, \pi(\mathrm{d}x). && (\pi = X_*\mu)
\end{aligned}$$

Hence, $Y_*\mu^+_x$ and $Y_*\mu$ are π-almost-surely equal. □

Lemma E.7. *Let $(A, \Sigma_A), (B, \Sigma_B)$ be measurable spaces, $(\mathcal{F}, \mu) \in \mathcal{M}$, $X : \Omega \to A$ an $(\mathcal{F}, \Sigma_A)$-measurable function and $Y : \Omega \to B$ an $(\mathcal{F}, \Sigma_B)$-measurable function. Then for any Borel extension*

μ^+ and $(X, X_*\mu^+)$-disintegration $\{\mu^+_x\}_{x\in A}$, the following identity holds for all $(\Sigma_A \otimes \Sigma_B, \mathcal{B}[0,\infty])$-measurable $f : A \times B \to [0,\infty]$ and almost all $x \in A$:

$$\int_\Omega f(X(\omega), Y(\omega))\, \mu^+_x(\mathrm{d}\omega) = \int_B f(x,y)\, Y_*\mu^+_x(\mathrm{d}y).$$

Proof. For $X_*\mu$-almost-all $x \in A$, the following identity holds:

$$\begin{aligned}
\int_\Omega f(X(\omega), Y(\omega))\, \mu^+_x(\mathrm{d}\omega) &= \int_{A\times B} f\, \mathrm{d}(X,Y)_*\mu^+_x && \text{(change of variable)}\\
&= \int_{A\times B} f\, \mathrm{d}(X_*\mu^+_x \otimes Y_*\mu^+_x) && \text{(Lemma E.6)}\\
&= \int_{A\times B} f\, \mathrm{d}(\delta_x \otimes Y_*\mu^+_x) && \text{(disintegration)}\\
&= \int_{A\times B} f\, \mathrm{d}(\delta_x \otimes p) && \text{(Lemma E.6)}\\
&= \int_B f(x,y)\, p(\mathrm{d}y). && \left(\textstyle\int_A f(-,y)\, \mathrm{d}\delta_x = f(x,y)\right)
\end{aligned}$$

This completes the proof. □

Lemma E.8. *Let (A, Σ_A) be a standard Borel space, $p : (A, \Sigma_A) \to \mathcal{G}(\mathbb{R})$ be a probability kernel. Then the following triple is sound:*

H-CondSample

$$\{X \sim \pi\}\ \mathbf{sample}(p(X))\ \{Y : \mathbb{R}.\ x \overset{\pi}{\leftarrow} X \mid Y \sim p(x)\}$$

Proof. Let $m_{\mathrm{pre}} := (\mathcal{F}_{\mathrm{pre}}, \mu_{\mathrm{pre}}) \in \mathcal{M}$ such that $(\gamma, D, m_{\mathrm{pre}}) \vDash X \sim \pi$, $m_{\mathrm{fr}} \in \mathcal{M}$ such that $(\mathcal{F}_0, \mu_0) := m_{\mathrm{pre}} \bullet m_{\mathrm{fr}}$ is defined, $\mu^+_0 : \Sigma_\Omega \to [0,\infty]$ a measure satisfying $\mu^+_0|_{\mathcal{F}_0} = \mu_0$, and all probabilistic contexts Δ_{ext} and all $D_{\mathrm{ext}} \in \mathrm{RV}[\![\Delta_{\mathrm{ext}}]\!]$. Notice the assumption that the program integrates to non-zero is always true:

$$\int_\Omega [\![\mathbf{sample}(p(X))]\!]((D,X)(\omega), \mathbb{R})\, \mu^+_0(\mathrm{d}\omega) = \int_\Omega 1\, \mu^+_0(\mathrm{d}\omega) = \mu^+_0(\Omega)$$

Denote n the maximum dimension of $m_{\mathrm{pre}} \bullet m_{\mathrm{fr}}$, D and D_{ext}'s finite footprint. Define a measurable function $Y : \Omega \to \mathbb{R}$ by

$$Y(\omega) := f(X(\omega), \pi_{n+1}(\omega))$$

where $f : A \times [0,1] \to \mathbb{R}$ is the $(\Sigma_A \otimes \mathcal{B}[0,1], \mathcal{B}\mathbb{R})$-measurable function obtained via Lemma E.5 that generates the distribution via a number uniformly distributed in $[0,1]$. Next, we define $\mathcal{F}_{\mathrm{post}} := \mathcal{F}_{\mathrm{pre}}|_n \otimes \mathcal{B}[0,1] \otimes \{\emptyset, \Omega\}$ and $m_{\mathrm{post}} := (\mathcal{F}_{\mathrm{post}}, \mu_{\mathrm{post}})$. Notice

$$\begin{aligned}
\mathcal{F}_{\mathrm{post}} &= \langle \mathcal{F}_{\mathrm{pre}}|_n \times \{\emptyset, \Omega\}, \{\emptyset, [0,1]^n\} \times \mathcal{B}[0,1] \times \{\emptyset, \Omega\}\rangle\\
&= \sigma\{F' \times U \times \Omega \mid F' \in \mathcal{F}_{\mathrm{pre}}|_n, U \in \mathcal{B}[0,1]\}.
\end{aligned}$$

Define μ_{post} to be the product measure $\mu_{\mathrm{post}} := \mu_{\mathrm{pre}}|_n \otimes \lambda_{[0,1]} \otimes p$ where p is the trivial probability measure on $\{\emptyset, \Omega\}$. By definition of a product measure, we know when $F \in \mathcal{F}_{\mathrm{pre}}$ and $U \in \mathcal{B}[0,1]$, $\mu_{\mathrm{post}}(F|_n \times U \times \Omega) = \mu_{\mathrm{pre}}(F) \cdot \lambda_{[0,1]}(U)$.

We now show that $(\gamma, (D,X,Y), m_{\mathrm{post}}) \vDash x \overset{\pi}{\leftarrow} X \mid Y \sim p(x)$. Let $\mu^+_{\mathrm{post}} : \Sigma_\Omega \to [0,\infty]$ be a Borel extension of μ_{post} and consider any (X,π)-disintegration $\{\mu^+_{\mathrm{post},x}\}_{x\in A}$, we need to show that the following holds for π-almost-all $x \in A$:

$$((\gamma, x), (D,X,Y), (\mathcal{F}_{\mathrm{post}}, \mu^+_{\mathrm{post},x}|_{\mathcal{F}_{\mathrm{post}}})) \vDash Y \sim p(x).$$

Also, $X_*\mu^+_{\text{post}} \ll \pi$ since X is $(\mathcal{F}_{\text{pre}}, \Sigma_A)$-measurable and $(\gamma, (D, X, Y), m_{\text{pre}}) \vDash X \sim \pi$, which implies for any π-null-set $E \in \Sigma_A$, we have $X_*\mu^+_{\text{post}}(E) = X_*\mu_{\text{pre}}(E) = \pi(E) = 0$. Next, notice that $X : \Omega \to A$ and $\pi_{n+1} : \Omega \to [0, 1]$ are μ_{post}-independent: for any $E \in \Sigma_A$ and $F \in \mathcal{B}[0, 1]$, we know

$$
\begin{aligned}
\mu_{\text{post}}\{X \in E, \pi_{n+1} \in F\} &= \mu_{\text{post}}(X^{-1}(E) \cap \pi^{-1}_{n+1}(F)) \\
&= \mu_{\text{post}}((E'|_n \times \Omega) \cap ([0, 1]^n \times F' \times \Omega)) && \text{(measurability)} \\
&= \mu_{\text{post}}(E'|_n \times F' \times \Omega) \\
&= \mu_{\text{pre}}(E') \cdot \lambda_{[0,1]}(F') && \text{(see above)} \\
&= \mu_{\text{post}}(E') \cdot \mu_{\text{post}}([0, 1]^n \times F' \times \Omega) && (\mu_{\text{post}}|_{\mathcal{F}_{\text{pre}}} = \mu_{\text{pre}}) \\
&= \mu_{\text{post}}\{X \in E\} \cdot \mu_{\text{post}}\{\pi_{n+1} \in F\}.
\end{aligned}
$$

Hence, $p(x) = Y_*\mu^+_{\text{post},x}|_{\mathcal{F}_{\text{post}}}$ for almost all $x \in A$: let $E \in \mathcal{B}\mathbb{R}$ and $F \in \Sigma_A$, we can then calculate

$$
\begin{aligned}
\int_F Y_*\mu^+_{\text{post},x}(E)\, \pi(\mathrm{d}x) &= \int_F \int_\Omega \mathbf{1}_E(f(X(\omega), \omega_{n+1}))\, \mu^+_{\text{post},x}(\mathrm{d}\omega)\, \pi(\mathrm{d}x) \\
&= \int_F \int_\Omega \mathbf{1}_E(f(x, r))\, \pi_{n+1*}\mu^+_{\text{post},x}(\mathrm{d}r)\, \pi(\mathrm{d}x) && \text{(Lemma E.7)} \\
&= \int_F \int_\Omega \mathbf{1}_E(f(x, r))\, \lambda_{[0,1]}(\mathrm{d}r)\, \pi(\mathrm{d}x) && \text{(Lemma E.7)} \\
&= \int_F \lambda_{[0,1]}\{r \in [0, 1] \mid f(x, r) \in E\}\, \pi(\mathrm{d}x) \\
&= \int_F p(x, E)\, \pi(\mathrm{d}x). && \text{(Lemma E.5)}
\end{aligned}
$$

Hence, $(\gamma, x), (D, X, Y), (\mathcal{F}_{\text{post}}, \mu^+_{\text{post},x}|_{\mathcal{F}_{\text{post}}}) \vDash Y \sim p(x)$ holds for π-almost-all $x \in A$. Notice μ_1 is always defined since n witnesses the finite footprint of the frame $(\mathcal{F}_0, \mu_0)$. Choose an extension $\mu^+_1 : \Sigma_\Omega \to [0, \infty]$ that satisfies $\mu^+_1|_{\mathcal{B}[0,1]^{n+1} \otimes \{\varnothing, \Omega\}} = \mu^+_0|_n \otimes \lambda_{[0,1]} \otimes p$. Then the following identity holds:

$$
\begin{aligned}
&\int_\Omega \int_\mathbb{R} \mathbf{1}_U(D_{\text{ext}}(\omega), D(\omega), y) [\![p(X)]\!]((D, X)(\omega), \mathrm{d}y)\, \mu^+_0(\mathrm{d}\omega) \\
&= \int_\Omega \int_\mathbb{R} \mathbf{1}_U(D_{\text{ext}}(\omega), D(\omega), y)\, p(X(\omega), \mathrm{d}y)\, \mu^+_0(\mathrm{d}\omega) \\
&= \int_\Omega p(X(\omega), E_\omega)\, \mu^+_0(\mathrm{d}\omega) \\
&= \int_\Omega \lambda_{[0,1]}\{r \in \mathbb{R} \mid f(X(\omega), r) \in E_\omega\}\, \mu^+_0(\mathrm{d}\omega) && \text{(Lemma E.5)} \\
&= \int_{[0,1]^n} \lambda_{[0,1]}\{r \in \mathbb{R} \mid f(X|_n(\omega_{1..n}), r) \in E_{\omega_{1..n}}\}\, \mu^+_0|_n(\mathrm{d}\omega_{1..n}) && \text{(Lemma E.2)} \\
&= \int_{[0,1]^{n+1}} \mathbf{1}_{E_{\omega_{1..n}}}(f(X(\omega_{1..n}), r))\, (\mu^+_0|_n \otimes \lambda_{[0,1]})(\mathrm{d}\omega_{1..n}, \mathrm{d}r) && \text{(Fubini)} \\
&= \int_\Omega \mathbf{1}_{E_\omega}(Y(\omega))\, \mu^+_1(\mathrm{d}\omega) \\
&= \int_\Omega \mathbf{1}_U(D_{\text{ext}}(\omega), (D, X)(\omega), Y(\omega))\, \mu^+_1(\mathrm{d}\omega),
\end{aligned}
$$

where $E_\omega \coloneqq \{y \in \mathbb{R} \mid (D_{\text{ext}}(\omega), D(\omega), y) \in U\}$. □

Lemma E.9. *Let $(\mathcal{F}, \mu) \in \mathcal{M}$, $f : \Omega \to [0, \infty)$ an $(\mathcal{F}, \mathcal{B}[0, \infty))$-measurable function and $(f \cdot \mu)^+ : \Sigma_\Omega \to [0, \infty]$ a Borel measure that extends a σ-finite measure $f \cdot \mu : \mathcal{F} \to [0, \infty]$. Then there is a Borel measure ν such that ν extends μ and*

$$(f \cdot \mu)^+ = f \cdot \nu.$$

Further, let $X : \Omega \to A$ be a $(\Sigma_\Omega, \Sigma_A)$-measurable function and $\{\nu_x\}_{x \in A}$ an (X, π)-disintegration of ν. Then $\{f \cdot \nu_x\}_{x \in A}$ is an (X, π)-disintegration of $(f \cdot \mu)^+$.

Proof. First, we define a Borel measure $\pi : \Sigma_\Omega \to [0, \infty]$ by

$$\pi(U) := \int_U f' \, \mathrm{d}(f \cdot \mu)^+ \qquad f'(\omega) := \mathbf{1}_{(0,\infty)}(f(\omega)) \cdot \frac{1}{f(\omega)}.$$

By Lemma C.18, the measure π is σ-finite. Notice for any $F \in \mathcal{F}$, π satisfies

$$\begin{aligned}
\pi(F) &= \int_F f' \, \mathrm{d}(f \cdot \mu)^+ \\
&= \int_F f' \, \mathrm{d}(f \cdot \mu) \\
&= \int_F f' \cdot f \, \mathrm{d}\mu \\
&= \int_F \mathbf{1}_{f^{-1}(0,\infty)} \, \mathrm{d}\mu \\
&= \mu(F \cap f^{-1}(0, \infty)).
\end{aligned}$$

By Proposition C.10, μ has a Borel measure extension $\mu^+ : \Sigma_\Omega \to [0, \infty]$. We can therefore define a σ-finite Borel measure $\pi' : \Sigma_\Omega \to [0, \infty]$ by $\pi'(U) := \mu^+(U \cap f^{-1}\{0\})$. Now, the sum of two σ-finite measures remains a σ-finite measure and we define $\nu : \Sigma_\Omega \to [0, \infty]$ by $\nu := \pi + \pi'$. Notice for any $F \in \mathcal{F}$, we have

$$\begin{aligned}
\nu(F) &= \pi(F) + \pi'(F) \\
&= \mu(F \cap f^{-1}(0, \infty)) + \mu(F \cap f^{-1}\{0\}) \\
&= \pi(F).
\end{aligned}$$

Hence, ν is a Borel extension of μ. Next, notice for any $U \in \Sigma_\Omega$, we have

$$\begin{aligned}
(f \cdot \nu)(U) &= \int_U f \, \mathrm{d}(\pi + \pi') \\
&= \int_U f \, \mathrm{d}\pi + \int_U f \, \mathrm{d}\pi' \\
&= \int_U f \, \mathrm{d}\pi + \int_U f \, \mu^+(\mathrm{d}\omega \cap f^{-1}\{0\}) \\
&= \int_U f \, \mathrm{d}\pi \\
&= (f \cdot \pi)(U).
\end{aligned}$$

Also, since $f^{-1}\{0\}$ is a $(f \cdot \mu)^+$-null-set, the following identity holds:

$$(f \cdot \mu)^+(U) = (f \cdot \mu)^+(U \cap f^{-1}(0, \infty)) + (f \cdot \mu)^+(U \cap f^{-1}\{0\}) = (f \cdot \mu)^+(U \cap f^{-1}(0, \infty)).$$

Hence, to show $(f \cdot \mu)^+ = f \cdot \nu$, it suffices to prove that $(f \cdot \mu)^+(U \cap f^{-1}(0,\infty)) = (f \cdot \pi)(U)$. Notice

$$\begin{aligned}(f \cdot \mu)^+(U \cap f^{-1}(0,\infty)) &= \int_{U \cap f^{-1}(0,\infty)} \mathrm{d}(f \cdot \mu)^+ \\ &= \int_U f \cdot f' \,\mathrm{d}(f \cdot \mu)^+ \\ &= \int_U f \,\mathrm{d}(f' \cdot (f \cdot \mu)^+) \\ &= (f \cdot \pi)(U).\end{aligned}$$

Next, we show that any (X, π)-disintegration of ν induces an (X, π)-disintegration of $(f \cdot \mu)^+$. Let $\{\nu_x : \Sigma_\Omega \longrightarrow [0,\infty]\}_{x \in A}$ be an (X, π)-disintegration of ν. Then $\{f \cdot \nu_x\}_{x \in A}$ is an (X, π)-disintegration of $f \cdot \nu$. Notice that the concentration property of disintegration holds:

$$(f \cdot \nu_x)\{X \neq x\} = \int_{\{X \neq x\}} f \,\mathrm{d}\nu_x \leq \sup f \cdot \nu_x\{X \neq x\} = 0.$$

Also, for every $x \in A$ and $(\Sigma_\Omega, \mathcal{B}[0,\infty])$-measurable $f : \Omega \longrightarrow [0,\infty]$, the map $x \longmapsto \int_\Omega f' \,\mathrm{d}(f \cdot \nu_x)$ is measurable – since $x \longmapsto \int_\Omega f'' \,\mathrm{d}\nu_x$ is measurable for any $(\Sigma_\Omega, \mathcal{B}[0,\infty])$-measurable $f'' : \Omega \longrightarrow [0,\infty]$, instantiating $f'' = f' \cdot f$ makes f' measurable. Finally, notice that the 'nesting' property of disintegration holds:

$$\int_\Omega f' \,\mathrm{d}(f \cdot \nu) = \int_\Omega f' \cdot f \,\mathrm{d}\nu = \int_A \int_\Omega f' \cdot f \,\mathrm{d}\nu_x \,\mu(\mathrm{d}x) = \int_A \int_\Omega f' \,\mathrm{d}(f \cdot \nu_x) \,\mu(\mathrm{d}x)$$

for all $(\Sigma_\Omega, \mathcal{B}[0,\infty])$-measurable $f' : \Omega \longrightarrow [0,\infty]$, which concludes the proof. □

Lemma E.10. *Let (A, Σ_A) be a standard Borel space, $p : (A, \Sigma_A) \longrightarrow \mathcal{G}(\mathbb{R})$ be a probability kernel. Then the following triple is sound:*

H-CondScore

$$\{x \overset{\pi}{\leftarrow} X \mid Y \sim p(x)\}\ \mathbf{score}(f(Y))\ \{x \overset{\pi}{\leftarrow} X \mid Y \sim f \cdot p(x)\}$$

Proof. Let $m_{\mathrm{pre}} := (\mathcal{F}_{\mathrm{pre}}, \mu_{\mathrm{pre}}) \in \mathcal{M}$ such that $(\gamma, D, m_{\mathrm{pre}}) \vDash x \overset{\pi}{\leftarrow} X \mid Y \sim p(x)$, $m_{\mathrm{fr}} \in \mathcal{M}$ such that $(\mathcal{F}_0, \mu_0) := m_{\mathrm{pre}} \bullet m_{\mathrm{fr}}$ is defined, $\mu_0^+ : \Sigma_\Omega \longrightarrow [0,\infty]$ a measure satisfying $\mu_0^+|_{\mathcal{F}_0} = \mu_0$, and all probabilistic contexts Δ_{ext} and all $D_{\mathrm{ext}} \in \mathrm{RV}[\![\Delta_{\mathrm{ext}}]\!]$. We assume that the following is true:

$$\int_\Omega [\![\mathbf{score}(f(Y))]\!](D, X, Y, \mathbf{1}) \,\mathrm{d}\mu_0^+ = \int_\Omega f(Y) \,\mathrm{d}\mu_0^+ > 0.$$

Define $Z : \Omega \longrightarrow \mathbf{1}$ by $Z(\omega) := *$, $\mathcal{F}_{\mathrm{post}} := \mathcal{F}_{\mathrm{pre}}$ and $\mu_{\mathrm{post}} : \mathcal{F}_{\mathrm{post}} \longrightarrow [0,\infty]$ by $\mu_{\mathrm{post}} := (f \circ Y) \cdot \mu_{\mathrm{pre}}$. We now show that $(\gamma, (D, X, Y, Z), m_{\mathrm{post}}) \vDash x \overset{\pi}{\leftarrow} X \mid Y \sim (f \cdot p)(x)$. Let $\mu_{\mathrm{post}}^+ : \Sigma_\Omega \longrightarrow [0,\infty]$ be a Borel extension of μ_{post} and consider any (X, π)-disintegration $\{\mu_{\mathrm{post},x}^+\}_{x \in A}$, we need to show that the following holds for π-almost-all $x \in A$:

$$((\gamma, x), (D, X, Y, Z), (\mathcal{F}_{\mathrm{pre}}, \mu_{\mathrm{post},x}^+|_{\mathcal{F}_{\mathrm{pre}}})) \vDash Y \sim f \cdot p(x).$$

The function Y is $(\mathcal{F}_{\mathrm{pre}}, \mathcal{B}\mathbb{R})$-measurable by assumption. Also, for any $U \in \mathcal{B}\mathbb{R}$, the following identity holds:

$$\begin{aligned}Y_* \mu_{\mathrm{post},x}^+|_{\mathcal{F}_{\mathrm{pre}}}(U) &= \mu_{\mathrm{post},x}^+(Y^{-1}(U)) \\ &= ((f \circ Y) \cdot \mu_{\mathrm{pre},x})^+(Y^{-1}(U)) && \text{(by definition)} \\ &= \int_{Y^{-1}(U)} f(Y(\omega)) \,\mu_{\mathrm{pre},x}^+(\mathrm{d}\omega) && \text{(see below)}\end{aligned}$$

$$= \int_U f \, \mathrm{d}Y_* \mu^+_{\mathrm{pre},x} \qquad \text{(change of variable)}$$

$$= \int_U f(y)\, p(x, \mathrm{d}y) \qquad \text{(Lemma E.7)}$$

$$= (f \cdot p(x))(U). \qquad \text{(by definition)}$$

where $\mu^+_{\mathrm{pre},x}$ is the disintegration of a Borel measure μ^+_{pre} satisfying $((f \circ Y) \cdot \mu_{\mathrm{pre}})^+ = (f \circ Y) \cdot \mu^+_{\mathrm{pre}}$ (which exists by Lemma E.9). Notice $(\mathcal{F}_1, \mu_1) := m_{\mathrm{post}} \bullet m_{\mathrm{fr}}$ is well-defined: $\mathcal{F}_1 = \mathcal{F}_0$ because $\langle \mathcal{F}_{\mathrm{post}}, \mathcal{F}_{\mathrm{fr}} \rangle = \langle \mathcal{F}_{\mathrm{pre}}, \mathcal{F}_{\mathrm{fr}} \rangle = \mathcal{F}_0$, and $\mu_1 = (f \circ Y) \cdot \mu_0$. For any $F \in \mathcal{F}_{\mathrm{post}}$ and $G \in \mathcal{F}_{\mathrm{fr}}$, we have

$$\mu_1(F \cap G) = \int_{F \cap G} f \circ Y \, \mathrm{d}\mu_0$$

$$= \int_F f \circ Y \mathrm{d}\mu_{\mathrm{pre}} \cdot \mu_{\mathrm{fr}}(G) \qquad \text{(Lemma C.15)}$$

$$= \mu_{\mathrm{post}}(F) \cdot \mu_{\mathrm{fr}}(G).$$

Define $\mu^+_1 := (f \circ Y) \cdot \mu_0$. Then for any $U \in \Sigma_{[\![\Delta_{\mathrm{ext}}]\!]} \otimes \Sigma_{[\![\Delta, X, Y]\!]}$, we have

$$\int_\Omega \mathbf{1}_U(D_{\mathrm{ext}}(\omega), (D, X, Y)(\omega), *)\, f(Y(\omega))\, \mu^+_0(\mathrm{d}\omega) = \int_\Omega \mathbf{1}_U(D_{\mathrm{ext}}(\omega), D(\omega), *)\, \mu^+_1(\mathrm{d}\omega),$$

which yields the postcondition $x \overset{\pi}{\leftarrow} X \mid Y \sim f \cdot p(x)$. □

Lemma E.11. *The following Hoare triple is sound:*

H-Ret
$\{P[[\![M]\!]/X]\}$ **return** M $\{X.P\}$

Proof. Let $m_{\mathrm{pre}} := (\mathcal{F}_{\mathrm{pre}}, \mu_{\mathrm{pre}}) \in \mathcal{M}$, $(\gamma, D, m_{\mathrm{pre}}) \vDash P[[\![M]\!]/X]$, $m_{\mathrm{fr}} := (\mathcal{F}_{\mathrm{fr}}, \mu_{\mathrm{fr}})$ such that $m_0 := m_{\mathrm{pre}} \bullet m_{\mathrm{fr}}$ is defined, and all m_{fr} with $(\mathcal{F}_0, \mu_0) := m_{\mathrm{pre}} \bullet m_{\mathrm{fr}}$ defined, and all measures μ^+_0 satisfying $\mu^+_0|_{\mathcal{F}_0} = \mu_0$, and all probabilistic contexts Δ_{ext} and all $D_{\mathrm{ext}} \in \mathrm{RV}[\![\Delta_{\mathrm{ext}}]\!]$. By the substitution lemma (Lemma D.4), we know

$$(\gamma, D, m_{\mathrm{pre}}) \vDash P[[\![M]\!]/X] \iff (\gamma, (D, [\![M]\!](\gamma)), m_{\mathrm{pre}}) \vDash P.$$

Define $X : \Omega \to A$ by $X(\omega) := [\![M(\gamma)]\!](D(\omega))$, $m_{\mathrm{post}} := m_{\mathrm{pre}}$, and $\mu^+_1 := \mu^+_0$. Then $(\gamma, (D, X), m_{\mathrm{pre}}) \vDash P$ and for all $U \in [\![\Delta_{\mathrm{ext}}]\!] \otimes [\![\Delta]\!] \otimes [\![A]\!]$, the following identity holds:

$$\int_\Omega \int_A \mathbf{1}_U(D_{\mathrm{ext}}(\omega), D(\omega), x)\, [\![\mathbf{return}\ M(\gamma)]\!](D(\omega), \mathrm{d}x)\, \mu^+_0(\mathrm{d}\omega)$$

$$= \int_\Omega \mathbf{1}_U(D_{\mathrm{ext}}(\omega), D(\omega), [\![M(\gamma)]\!](D(\omega)))\, \mu^+_0(\mathrm{d}\omega) \qquad \text{(semantics of } \mathbf{return}\text{)}$$

$$= \int_\Omega \mathbf{1}_U(D_{\mathrm{ext}}(\omega), D(\omega), X(\omega))\, \mu^+_1(\mathrm{d}\omega). \qquad \text{(by definition)}$$

which yields the postcondition P. □

Lemma E.12. *Let $\mu : \Sigma_A \to [0, \infty]$ be a measure and $f : A \to [0, \infty]$ be a $(\Sigma_A, \mathcal{B}[0, \infty])$-measurable function. Then $\int_A f \, \mathrm{d}\mu > 0$ if and only if there exists $U \in \Sigma_A$ such that $\mu(U) > 0$ and $f(x) > 0$ for all $x \in U$.*

Proof. Since the codomain of f is $[0, \infty]$, $f(x) = 0$ holds for almost all $x \in A$ iff $\int_A f \, \mathrm{d}\mu = 0$. This means $\mu\{x \in A \mid f(x) \neq 0\} = 0$ iff $\int_A f \, \mathrm{d}\mu = 0$. Negating L.H.S. yields $\mu\{x \in A \mid f(x) \neq 0\} \neq 0$, which is equivalent to $\mu\{x \in A \mid f(x) > 0\} > 0$. By contrapositivity, we know $\int_A f \, \mathrm{d}\mu > 0$ if and only if $\mu\{x \in A \mid f(x) > 0\} > 0$. This proves the direction from left to right by defining

$U := \{x \in A \mid f(x) > 0\}$. From right to left, assuming such a set U exists, then $0 < \mu(U) = \mu\{x \in U \mid f(x) > 0\} \leq \mu\{x \in A \mid f(x) > 0\}$. By the above derivation, we have $\int_A f \, d\mu > 0$. □

Corollary E.13. *Let $\mu : \Sigma_\Omega \to [0, \infty]$ be a measure, $\kappa : \Omega \times \Sigma_A \to [0, \infty]$ a measure kernel, and $\varphi : \Omega \times A \to [0, \infty]$ a measurable function. Then*

$$\int_\Omega \int_A \varphi(\omega, x) \, \kappa(\omega, dx) \, \mu(d\omega) > 0 \implies \int_\Omega \kappa(\omega, A) \, \mu(d\omega) > 0.$$

Proof. By Lemma E.12, we know there exists $B \in \Sigma_\Omega$ such that $\mu(B) > 0$ and $\int_A \varphi(\omega, x) \, \kappa(\omega, dx) > 0$ for all $\omega \in B$. Also, we know there exists, for every $\omega \in B$, a set $U_\omega \in \Sigma_A$ such that $\kappa(\omega, U_\omega) > 0$ and $\varphi(\omega, x) > 0$ for all $x \in U_\omega$. To show that $\int_\Omega \kappa(\omega, A) \, \mu(d\omega) > 0$, notice that $\kappa(\omega, A) \geq \kappa(\omega, U_\omega) > 0$ for all $\omega \in B$. By Lemma E.12, the desired inequality holds. □

Lemma E.14. *Let $\mu, \nu : \Sigma_\Omega \to [0, \infty]$ be measures and $\kappa : \Omega \times \Sigma_A \to [0, \infty]$ a measure kernel. If for every $U \in \Sigma_\Omega \otimes \Sigma_A$ and $(\Sigma_\Omega, \Sigma_A)$-measurable $X : \Omega \to A$,*

$$\int_\Omega \int_A \mathbf{1}_U(\omega, x) \, \kappa(\omega, dx) \, \mu(d\omega) = \int_\Omega \mathbf{1}_U(\omega, X(\omega)) \, \nu(d\omega),$$

then for any $(\Sigma_\Omega \otimes \Sigma_A, \mathcal{B}[0, \infty])$-measurable $\varphi : \Omega \times A \to [0, \infty]$, the following identity holds:

$$\int_\Omega \int_A \varphi(\omega, x) \, \kappa(\omega, dx) \, \mu(d\omega) = \int_\Omega \varphi(\omega, X(\omega)) \, \nu(d\omega).$$

Proof. For every Borel set $V \in \mathcal{B}[0, \infty]$, the following identity holds by assumption:

$$\begin{aligned}
&\int_\Omega \int_A \mathbf{1}_V(\varphi(\omega, x)) \, \kappa(\omega, dx) \, \mu(d\omega) \\
&\quad = \int_\Omega \int_A \mathbf{1}_{\varphi^{-1}(V)}(\omega, x) \, \kappa(\omega, dx) \, \mu(d\omega) && \text{(inverse image)} \\
&\quad = \int_\Omega \mathbf{1}_{\varphi^{-1}(V)}(\omega, X(\omega)) \, \nu(d\omega) && \text{(assumption)} \\
&\quad = \int_\Omega \mathbf{1}_V(\varphi(\omega, X(\omega))) \, \nu(d\omega). && \text{(inverse image)}
\end{aligned}$$

Assume that $\varphi = \mathbf{1}_W$ for some measurable $W \in \Sigma_\Omega \otimes \Sigma_A$, then by instantiating $V := \{1\}$, we know $\mathbf{1}_V \circ \mathbf{1}_W = \mathbf{1}_W$, which implies

$$\int_\Omega \int_A \mathbf{1}_W(\omega, x) \, \kappa(\omega, dx) \, \mu(d\omega) = \int_\Omega \mathbf{1}_W(\omega, X(\omega)) \, \nu(d\omega).$$

Hence, all characteristic functions satisfy the property. Next, assume that $\varphi = \sum_{i=1}^n c_i \mathbf{1}_{W_i}$ with $c_i \in [0, \infty)$ and $\{W_i \in \Sigma_\Omega \otimes \Sigma_A\}_{i=1}^n$ is a disjoint measurable partition. We can assume that $c_i \neq c_j$ when $i \neq j$ (if $i \neq j$ and $c_i = c_j$, φ can alternatively be represented by $\sum_{i=1}^{n-1} c_i \mathbf{1}_{W'_i}$ for some partition $\{W'_k\}_{k=1}^{n-1}$ with $W'_i := W_i \cup W_j$ and 'removing' W_j). Our assumption is now

$$\int_\Omega \int_A \mathbf{1}_V\left(\sum_{i=1}^n c_i \mathbf{1}_{W_i}(\omega, x)\right) \kappa(\omega, dx) \, \mu(d\omega) = \int_\Omega \mathbf{1}_V\left(\sum_{i=1}^n c_i \mathbf{1}_{W_i}(\omega, X(\omega))\right) \nu(d\omega).$$

for all $V \in \mathcal{B}[0, \infty]$. For each $i = 1, \ldots, n$, instantiating $V := \{c_i\}$ implies $\mathbf{1}_V(\sum_{i=1}^n c_i \mathbf{1}_{W_i}(\omega, x)) = \mathbf{1}_{W_i}(\omega, x)$ for all $\omega \in \Omega$ and $x \in A$, which then implies

$$\int_\Omega \int_A \mathbf{1}_{W_i}(\omega, x) \, \kappa(\omega, dx) = \int_\Omega \mathbf{1}_{W_i}(\omega, X(\omega)) \, \nu(d\omega).$$

Since the integrands are both non-negative, multiplying both sides by c_i yields

$$\int_\Omega \int_A c_i \mathbf{1}_{W_i}(\omega, x)\, \kappa(\omega, \mathrm{d}x)\, \mu(\mathrm{d}\omega) = \int_\Omega c_i \mathbf{1}_{W_i}(\omega, X(\omega))\, \nu(\mathrm{d}\omega).$$

An integral is always defined, and hence, linear, when the codomain is $[0, \infty]$ [Axler 2019, Theorem 3.16], which implies

$$\begin{aligned}
&\int_\Omega \int_A \sum_{i=1}^n c_i \mathbf{1}_{W_i}(\omega, x)\, \kappa(\omega, \mathrm{d}x)\, \mu(\mathrm{d}\omega) \\
&= \sum_{i=1}^n \int_\Omega \int_A c_i \mathbf{1}_{W_i}(\omega, x)\, \kappa(\omega, \mathrm{d}x)\, \mu(\mathrm{d}\omega) \\
&= \sum_{i=1}^n \int_\Omega c_i \mathbf{1}_{W_i}(\omega, X(\omega))\, \nu(\mathrm{d}\omega) && \text{(see above)} \\
&= \int_\Omega \sum_{i=1}^n c_i \mathbf{1}_{W_i}(\omega, X(\omega))\, \nu(\mathrm{d}\omega).
\end{aligned}$$

Hence, every non-negative measurable simple function $\varphi : \Omega \to [0, \infty)$ satisfies the desired property. Next, we assume $\varphi : \Omega \to [0, \infty]$ is an arbitrary $(\Sigma_\Omega \otimes \Sigma_A, \mathcal{B}[0, \infty])$-measurable function. By the simple function approximation theorem for $[0, \infty]$ [Axler 2019, Theorem 2.89], there exists a sequence of measurable simple functions $\{\varphi_i : \Omega \to [0, \infty)\}_{i \in \mathbb{N}}$ such that $\varphi_i \leq \varphi_{i+1}$ for all $i \in \mathbb{N}$ and φ_i converges pointwise to φ, i.e. $\lim_{i \to \infty} \varphi_i(\omega, x) = \varphi(\omega, x)$ for all $\omega \in \Omega$ and $x \in A$. Then

$$\begin{aligned}
&\int_\Omega \int_A \lim_{i \to \infty} \varphi_i(\omega, x)\, \kappa(\omega, \mathrm{d}x)\, \mu(\mathrm{d}\omega) \\
&= \lim_{i \to \infty} \int_\Omega \int_A \varphi_i(\omega, x)\, \kappa(\omega, \mathrm{d}x)\, \mu(\mathrm{d}\omega) && \text{(monotone convergence)} \\
&= \lim_{i \to \infty} \int_\Omega \varphi_i(\omega, X(\omega))\, \nu(\mathrm{d}\omega) && \text{(see above)} \\
&= \int_\Omega \lim_{i \to \infty} \varphi_i(\omega, X(\omega))\, \nu(\mathrm{d}\omega). && \text{(monotone convergence)}
\end{aligned}$$

Hence, every $(\Sigma_\Omega \otimes \Sigma_A, \mathcal{B}[0, \infty])$-measurable function satisfies the property. □

Lemma E.15. *Let $\mu : \Sigma_A \to [0, \infty]$ be a measure, $\kappa : A \times \Sigma_B \to [0, \infty]$ a measure kernel. Consider a measure $\nu : \Sigma_B \to [0, \infty]$ defined by $\nu(U) := \int_A \kappa(x, U)\, \mu(\mathrm{d}x)$, then for any $(\Sigma_B, \mathcal{B}[0, \infty])$-measurable $\varphi : B \to [0, \infty]$,*

$$\int_B \varphi\, \mathrm{d}\nu = \int_A \int_B \varphi(y)\, \kappa(x, \mathrm{d}y)\, \mu(\mathrm{d}x).$$

Proof. Suppose $\varphi : \Omega \to [0, \infty)$ is a simple function $\varphi = \sum_{i=1}^n c_i \mathbf{1}_{U_i}$, notice that

$$\int_B \varphi\, \mathrm{d}\nu = \sum_{i=1}^n c_i \nu(U_i) = \sum_{i=1}^n c_i \int_A \int_{U_i} \kappa(x, \mathrm{d}y)\, \mu(\mathrm{d}x) = \int_A \int_B \varphi(y)\, \kappa(x, \mathrm{d}y)\, \mu(\mathrm{d}x),$$

where additivity and multiplicity follow from the same theorems as the ones in Lemma E.14. Now, consider a $(\Sigma_B, \mathcal{B}[0, \infty])$-measurable function $\varphi : B \to [0, \infty]$ approximated by the sequence of measurable simple functions $\{\varphi_i : B \to [0, \infty)\}_{i \in \mathbb{N}}$ via the simple function approximation theorem for $[0, \infty]$ [Axler 2019, Theorem 2.89]. By the monotone convergence theorem, we know

$$\int_B \varphi\, \mathrm{d}\nu = \int_B \lim_{i \to \infty} \varphi_i\, \mathrm{d}\nu = \lim_{i \to \infty} \int_B \varphi_i\, \mathrm{d}\nu = \lim_{i \to \infty} \int_A \int_B \varphi_i(y)\, \kappa(x, \mathrm{d}y)\, \mu(\mathrm{d}x)$$

$$= \int_A \int_B \varphi(y)\, \kappa(x, \mathrm{d}y)\, \mu(\mathrm{d}x).$$

Hence, the desired property holds for all $(\Sigma_B, \mathcal{B}[0,\infty])$-measurable functions. □

Lemma E.16. *The following Hoare triple is sound:*

$$\frac{\vdash \{P\}\, M\, \{X : \mathcal{A}.Q\} \qquad \vdash \forall_{\mathrm{rv}} X : \mathcal{A}.\, \{Q\}\, N\, \{Y : \mathcal{B}.R\}}{\vdash \{P\}\, \mathbf{let}\ X = M\ \mathbf{in}\ N\, \{Y : \mathcal{B}.R\}}\ \text{H-Let}$$

Proof. Let (γ, D, m) be a configuration such that $(\gamma, D, m) \vDash \{P\}\, M\, \{X : \mathcal{A}.Q\}$ and $(\gamma, D, m) \vDash \forall_{\mathrm{rv}} X.\, \{Q\}\, N\, \{Y : \mathcal{B}.R\}$. For all $m_{\mathrm{pre}} \in \mathcal{M}$ with $(\gamma, D, m_{\mathrm{pre}}) \vDash P$, $m_{\mathrm{fr}} \in \mathcal{M}$ with $(\mathcal{F}_0, \mu_0) := m_{\mathrm{pre}} \bullet m_{\mathrm{fr}}$ defined, Borel measures $\mu_0^+ : \Sigma_\Omega \to [0,\infty]$ satisfying $\mu_0^+|_{\mathcal{F}_0} = \mu_0$, and all probabilistic contexts Δ_{ext} and all $D_{\mathrm{ext}} \in \mathrm{RV}[\![\Delta_{\mathrm{ext}}]\!]$. Assuming $\int_\Omega [\![\mathbf{let}\ X = M(\gamma)\ \mathbf{in}\ N(\gamma)]\!](D(\omega), \mathcal{B})\, \mu_0^+(\mathrm{d}\omega) > 0$, then by definition of the semantics for **let**,

$$\int_\Omega [\![\mathbf{let}\ X = M(\gamma)\ \mathbf{in}\ N(\gamma)]\!](D(\omega), \mathcal{B})\, \mu_0^+(\mathrm{d}\omega) = \int_\Omega \int_A \varphi(\omega, x)\, [\![M(\gamma)]\!](D(\omega), \mathrm{d}x)\, \mu_0^+(\mathrm{d}\omega) > 0,$$

where $\varphi(\omega, x) := [\![N(\gamma)]\!]((D(\omega), x), \mathcal{B})$. By Corollary E.13, we know $\int_\Omega [\![M(\gamma)]\!](D(\omega), \mathcal{A})\, \mu_0^+(\mathrm{d}\omega) > 0$. Hence, there exists (1) $X \in \mathrm{RV}(\mathcal{A})$, (2) $m_{\mathrm{post}} \in \mathcal{M}$ such that $(\mathcal{F}_1, \mu_1) := m_{\mathrm{post}}, m_{\mathrm{fr}}$ is defined, and (3) Borel measure $\mu_1^+ : \Sigma_\Omega \to [0,\infty]$ that extends μ_1.

Moving to the premises of $\forall_{\mathrm{rv}} X : \mathcal{A}.\, \{Q\}\, M\, \{Y : \mathcal{B}.R\}$, we instantiate $X \in \mathrm{RV}(\mathcal{A})$ to the random variable obtained above, $m'_{\mathrm{pre}} := m_{\mathrm{post}}$, $m'_{\mathrm{fr}} := m_{\mathrm{fr}}$, $\mu_0'^+ := \mu_1^+$, $\Delta'_{\mathrm{ext}} := \Delta_{\mathrm{ext}}$ and $D'_{\mathrm{ext}} := D_{\mathrm{ext}}$. Notice that $\int_\Omega [\![N(\gamma)]\!]((D, X)(\omega), \mathcal{B})\, \mu_1^+(\mathrm{d}\omega) = \int_\Omega \varphi(\omega, X(\omega))\, \mu_1^+(\mathrm{d}\omega) > 0$: by instantiating $\Delta_{\mathrm{ext}} := \Omega$, we know, from the conclusion of $\{P\}\, M\, \{X.Q\}$, that the following identity holds for every $U \in \Sigma_\Omega \otimes \Sigma_A$:

$$\int_\Omega \int_A \mathbf{1}_U(\omega, x)\, [\![M(\gamma)]\!](D(\omega), \mathrm{d}x)\, \mu_0^+(\mathrm{d}\omega) = \int_\Omega \mathbf{1}_U(\omega, X(\omega))\, \mu_1^+(\mathrm{d}\omega).$$

Notice that $(\omega, V) \mapsto [\![M(\gamma)]\!](D(\omega), V)$ is a kernel from Ω to $\mathcal{A}$. By Lemma E.14, we know

$$\int_\Omega \varphi(\omega, X(\omega))\, \mu_1^+(\mathrm{d}\omega) = \int_\Omega \int_A \varphi(\omega, x)\, [\![M(\gamma)]\!](D(\omega), \mathrm{d}x)\, \mu_0^+(\mathrm{d}\omega) > 0.$$

Hence, there exists (1) $Y \in \mathrm{RV}(\mathcal{B})$, (2) $m'_{\mathrm{post}} \in \mathcal{M}$ such that $(\mathcal{F}'_1, \mu'_1) := m'_{\mathrm{post}} \bullet m_{\mathrm{fr}}$ is defined, and (3) a Borel measure $\mu_1'^+$ such that $\mu_1'^+|_{\mathcal{F}'_1} = \mu'_1$ s.t. $(\gamma, (D, Y), m'_{\mathrm{post}}) \vDash R$, and for all $U \in \Sigma_{[\![\Delta_{\mathrm{ext}}]\!]} \otimes \Sigma_{[\![\Delta]\!]} \otimes \Sigma_A \otimes \Sigma_B$,

$$\int_\Omega \int_{\mathcal{B}} \mathbf{1}_U(D_{\mathrm{ext}}, D, X, y)\, [\![N(\gamma)]\!](D, X, \mathrm{d}y)\, \mathrm{d}\mu_1^+ = \int_\Omega \mathbf{1}_U(D_{\mathrm{ext}}, D, X, Y)\, \mathrm{d}\mu_1'^+.$$

With the same $Y, m'_{\mathrm{post}}, \mu_1'^+$, notice that

$$\begin{aligned}
&\int_\Omega \int_B \mathbf{1}_U(D_{\mathrm{ext}}, D, y)\, [\![\mathbf{let}\ X = M(\gamma)\ \mathbf{in}\ N(\gamma)]\!](D, \mathrm{d}y)\, \mathrm{d}\mu_0^+ \\
&= \int_\Omega \int_B \mathbf{1}_U(D_{\mathrm{ext}}, D, y)\, \nu(D, \mathrm{d}y)\, \mathrm{d}\mu_0^+ && \text{(semantics of } \mathbf{let}) \\
&= \int_\Omega \int_A \int_B \mathbf{1}_{A\times U}(x, D_{\mathrm{ext}}, y)\, [\![N(\gamma)]\!](D, x, \mathrm{d}y)\, [\![M(\gamma)]\!](D, \mathrm{d}x)\, \mathrm{d}\mu_0^+ && \text{(Lemma E.15)} \\
&= \int_\Omega \int_B \mathbf{1}_{A\times U}(X, D_{\mathrm{ext}}, y)\, [\![N(\gamma)]\!](D, X, \mathrm{d}y)\, \mathrm{d}\mu_1^+ && \text{(Lemma E.14)} \\
&= \int_\Omega \mathbf{1}_{A\times U}(X, D_{\mathrm{ext}}, D, Y)\, \mathrm{d}\mu_1'^+ && (\{Q\}\, N\, \{Y.R\} \text{ holds})
\end{aligned}$$

$$= \int_{\Omega} \mathbf{1}_U(D_{\text{ext}}(\omega), D(\omega), Y(\omega))\, \mu_1'^{+}(\mathrm{d}\omega).$$

where $\nu : [\![\Delta]\!] \times \Sigma_B \to [0, \infty]$ is an s-finite kernel defined by

$$\nu(\delta, E) := \int_A [\![N(\gamma)]\!](\delta, x, E)\, [\![M(\gamma)]\!](\delta, \mathrm{d}x). \qquad \square$$

Lemma E.17. *The following Hoare triple is sound:*

$$\text{H-Frame}\quad \frac{\vdash \{P\}\, M\, \{X : \mathcal{A}.Q\}}{\vdash \{P * F\}\, M\, \{X : \mathcal{A}.Q * F\}}\ (X \notin \mathsf{fv}(F))$$

Proof. Let $(\gamma, D, m) \vDash \{P\}\, M\, \{X.Q\}$ and F be a proposition that does not contain X as a free random variable. We need to show that $(\gamma, D, m) \vDash \{P * F\}\, M\, \{X.Q * F\}$. For every m_{pre} such that $(\gamma, D, m_{\text{pre}}) \vDash P * F$, we know, by definition of the satisfaction relation, there exists $m_{\text{pre1}}, m_{\text{pre2}}$ such that $(\gamma, D, m_{\text{pre1}}) \vDash P$, $(\gamma, D, m_{\text{pre2}}) \vDash F$, and $m_{\text{pre1}} \bullet m_{\text{pre2}} \sqsubseteq m_{\text{pre}}$. Now, suppose $m_{\text{fr}} \in \mathcal{M}$ with $(\mathcal{F}_0, \mu_0) := m_{\text{pre}} \bullet m_{\text{fr}} = (m_{\text{pre1}} \bullet m_{\text{pre2}}) \bullet m_{\text{fr}} = m_{\text{pre1}} \bullet (m_{\text{pre2}} \bullet m_{\text{fr}})$ defined. We instantiate the semantics of $(\gamma, D, m) \vDash \{P\}\, M\, \{X.Q\}$ with $m'_{\text{pre}} := m_{\text{pre1}}$, $m'_{\text{fr}} := m_{\text{pre2}} \bullet m_{\text{fr}}$. Then, we know that for every Borel measures $\mu_0^+ : \Sigma_\Omega \to [0, \infty]$ satisfying $\mu_0^+|_{\mathcal{F}_0}$ and $D_{\text{ext}} \in \mathsf{RV}[\![\Delta_{\text{ext}}]\!]$, there exists $X \in \mathsf{RV}(A)$, $m_{\text{post}} \in \mathcal{M}$ with $(\mathcal{F}_1, \mu_1) := m_{\text{post}} \bullet m'_{\text{fr}} = m_{\text{post}} \bullet (m_{\text{pre2}} \bullet m_{\text{fr}})$ defined, and Borel measure $\mu_1^+ : \Sigma_\Omega \to [0, \infty]$ with $\mu_1^+|_{\mathcal{F}_1} = \mu_1$ such that $(\gamma, (D, X), m_{\text{post}}) \vDash Q$. Since $m_{\text{post}} \bullet (m_{\text{pre2}} \bullet m_{\text{fr}}) = (m_{\text{post}} \bullet m_{\text{pre2}}) \bullet m_{\text{fr}}$ is defined, $m_{\text{post}} \bullet m_{\text{pre2}}$ is also defined. Also, since X is not a free variable of F, $(\gamma, D, m_{\text{pre2}}) \vDash F$ implies $(\gamma, (D, X), m_{\text{pre2}}) \vDash F$. By definition of the satisfaction relation, $(\gamma, (D, X), m_{\text{post}}) \vDash Q$ and $(\gamma, (D, X), m_{\text{pre2}}) \vDash F$ implies $(\gamma, D, m_{\text{post}} \bullet m_{\text{fr}}) \vDash Q * F$. $\square$

Lemma E.18. *The following Hoare triple is sound:*

$$\text{H-Cons}\quad \frac{P' \vdash P \qquad \vdash \{P\}\, M\, \{X : \mathcal{A}.Q\} \qquad Q \vdash Q'}{\vdash \{P'\}\, M\, \{X : \mathcal{A}.Q'\}}$$

Proof. Assuming $P' \vdash P$, $(\gamma, D, m) \vDash \{P\}\, M\, \{X.Q\}$ and $Q \vdash Q'$. For every m_{pre} such that $(\gamma, D, m_{\text{pre}}) \vDash P'$, we know $(\gamma, D, m_{\text{pre}}) \vDash P$ by definition of $P' \vdash P$. This means for all m_{fr} with $(\mathcal{F}_0, \mu_0) := m_{\text{pre}} \bullet m_{\text{fr}}$ defined, Borel measure $\mu_0^+ : \Sigma_\Omega \to [0, \infty]$ satisfying $\mu_0^+|_{\mathcal{F}_0} = \mu_0$ and random variables $D_{\text{ext}} \in \mathsf{RV}[\![\Delta_{\text{ext}}]\!]$. Assuming $\int_\Omega [\![M(\gamma)]\!](D(\omega), A)\, \mu_0^+(\mathrm{d}\omega) > 0$, then there exists (1) $X \in \mathsf{RV}(A)$, (2) $m_{\text{post}} \in \mathcal{M}$ with $(\mathcal{F}_1, \mu_1) := m_{\text{post}} \bullet m_{\text{fr}}$ defined, and measure $\mu_1^+ : \Sigma_\Omega \to [0, \infty]$ with $\mu_1^+|_{\mathcal{F}_1} = \mu_1$ such that $(\gamma, (D, X), m_{\text{post}}) \vDash Q$. By definition of $Q \vdash Q'$, $(\gamma, (D, X), m_{\text{post}}) \vDash Q'$. $\square$