Probabilistic programming languages (PPLs) are an expressive means for creating and reasoning about probabilistic models. Unfortunately hybrid probabilistic programs that involve both continuous and discrete structures are not well supported by today’s PPLs. In this paper we develop a new approximate inference algorithm for hybrid probabilistic programs that first discretizes the continuous distributions and then performs discrete inference on the resulting program. The key novelty is a form of discretization that we call bit blasting, which uses a binary representation of numbers such that a domain of $2^b$ discretized points can be succinctly represented as a discrete probabilistic program over $\text{poly}(b)$ Boolean random variables. Surprisingly, we prove that many common continuous distributions can be bit blasted in a manner that incurs no loss of accuracy over an explicit discretization and supports efficient probabilistic inference. We have built a probabilistic programming system for hybrid programs called HyBit, which employs bit blasting followed by discrete probabilistic inference. We empirically demonstrate the benefits of our approach over existing sampling-based and symbolic inference approaches.
Table 1. Distributions with sound and efficient bit blasting subsumed by mixed-gamma densities.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Density</th>
<th>Distribution</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>$1$</td>
<td>Gamma</td>
<td>$\frac{\beta^\alpha e^{-\beta x}}{\Gamma(\alpha)}$</td>
</tr>
<tr>
<td>Linear</td>
<td>$x$</td>
<td>Laplace</td>
<td>$\frac{1}{2b}e^{-\frac{x^2}{b^2}}$</td>
</tr>
<tr>
<td>Polynomial</td>
<td>$x^n$</td>
<td>Chi-squared</td>
<td>$\frac{1}{2\pi}\Gamma\left(\frac{1}{2}\right)x^{-0.5}e^{-\frac{x^2}{2}}$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$\lambda e^{-\lambda x}$</td>
<td>Student-T</td>
<td>$c\left(1 + \frac{x^2}{\lambda^2}\right)^{-\text{nd}}$</td>
</tr>
</tbody>
</table>

in the number of discrete variables. Other sampling-based approaches are universal and so can handle discreteness, such as importance sampling, Markov Chain Monte Carlo and Sequential Monte Carlo, etc. [Koller and Friedman 2009]. However, these algorithms are known to struggle with multimodal distributions [Yao et al. 2021], which arise through discreteness, as well as with programs that condition on low-probability events.

In this paper we develop a new inference algorithm for hybrid probabilistic programs via discretization: we convert the continuous distributions in a hybrid program to discrete distributions. This yields a fully discrete probabilistic program on which existing algorithms for discrete inference can be used. Discretization approximates a continuous distribution as a sequence of intervals, with each interval associated with the probability of the value falling in that interval. Forms of discretization have been used in prior work [Albarghouthi et al. 2017; Beutner et al. 2022; Claret et al. 2013; Huang et al. 2021] but they all scale linearly in the number of intervals. This imposes a clear tradeoff: one needs many small intervals in order to avoid losing too much precision, but the cost of inference quickly becomes prohibitive as the number of intervals grows.

We introduce a new approach to discretization that we call bit blasting, by analogy with the technique of the same name in verification [Bruttomesso and Sharygina 2009]. The key property of a bit blasted discretization is that it uses only $\text{poly}(\log n)$ Boolean random variables to represent a discretization on $n$ intervals. This is achieved by employing a binary representation of numbers and representing discretizations as discrete probabilistic programs over this binary representation. At first blush, this succinct representation would appear to lose too much accuracy to be a viable strategy, but we present both theoretical and empirical results to the contrary.

First, we prove that a large class of common continuous densities can be bit blasted soundly, that is with no loss of accuracy versus a naïve discretization. Table 1 lists example distributions that are in this class; we refer to the entire class as mixed-gamma distributions.

For instance, consider discretizing a continuous uniform distribution between 0 and 1. Naïve discretization to $2^{32}$ intervals requires enumeration of $2^{32}$ values. Instead, this distribution can be represented in binary as a tuple of 32 Bernoulli random variables of the form $\text{flip}(0.5)$, i.e., coin tosses that are equally likely to have the value 0 or 1. This observation is not new, and a similar result holds for exponential distributions as well [Marsaglia 1971]. However, the bit blasted discretizations of other mixed-gamma densities are novel. Further, unlike the case for uniforms and exponentials, these discretizations are not defined as simply a tuple of independent Bernoulli random variables but rather require full-fledged discrete probabilistic programs over such variables.

A succinct representation does not necessarily imply efficient inference, which is hard in general. As our second contribution, however, we prove that bit blasted mixed-gamma distributions are not only sound and succinct, but they also support polynomial-time inference in the number of bits of precision. Specifically, we prove that the knowledge compilation approach to discrete probabilistic
inference [Chavira and Darwiche 2008; Chavira et al. 2006; De Raedt et al. 2007; Fierens et al. 2015; Holtzen et al. 2020], which reduces inference to weighted model counting on a boolean formula, has this property for bit blasted mixed-gamma distributions. Therefore, the process of bit blasting a mixed-gamma distribution followed by discrete inference via knowledge compilation is guaranteed to take time that is polynomial in the bitwidth used for discretization.

Third, we have used the above theoretical results to design a new probabilistic programming system called HyBit that performs non-stochastic approximate inference for hybrid probabilistic programs via bit blasting. HyBit is a discrete probabilistic language that includes support for fixed-point binary numbers and associated arithmetic operations, and it is implemented as an embedded domain specific language in Julia [Bezanson et al. 2017]. The HyBit API allows users to produce bit blasted fixed-point approximations of arbitrary continuous distributions. Mixed-gamma distributions can be represented by their sound bit blasted discrete distributions. For other distributions the API enables users to employ piece-wise discrete approximations, where each piece is itself a bit blasted mixed-gamma distribution.

HyBit leverages knowledge compilation to perform exact inference on the given discrete probabilistic program. In the worst case, the inference for an arbitrary hybrid probabilistic program via bit blasting can be exponential in the number of bits, but we empirically demonstrate the benefits of our approach. We show that HyBit performs better than existing sampling-based and symbolic inference approaches on a comprehensive benchmark suite of hybrid probabilistic programs.

Overall, we present the following contributions in this paper:

1. We motivate the challenges for inference on hybrid probabilistic programs in Section 2.
2. We present a new form of discretization called bit blasting that is characterized by its succinctness in the number of discrete intervals in Section 3.
3. We present a class of continuous distributions, namely mixed-gamma distributions, for which a sound bit blasted representation exists. We formalize this construction and prove its properties in Section 3. We further prove that knowledge compilation based inference scales polynomially in the bitwidth for these distributions.
4. We describe the HyBit PPL and its new inference algorithm via bit blasting in Section 4.
5. In Section 5, we empirically compare HyBit with other PPLs on benchmarks obtained from the existing literature. We also characterize the behavior of HyBit with respect to its hyperparameters, i.e. number of bits and pieces.

HyBit is available at https://github.com/Tractables/Dice.jl/tree/hybit. The full version of this paper is available on arXiv as [Garg et al. 2023], which contains full proofs.

2 MOTIVATING EXAMPLES

This section motivates the challenge of inference for hybrid probabilistic programs using three examples. First, we present an example from computational biology with inherent logical structure. Next, we show an example from the literature with a multimodal posterior arising due to discrete control flow. Finally, we show an example of low probability observations through conjugate Gaussians. We then investigate the performance of various inference algorithms, including HyBit. The examples demonstrate the advantages of HyBit over other approximate inference algorithms. We provide a detailed comparison with several approximate and exact baselines in Section 5.

2.1 Logical Structure

We present a simplified example from computational biology which relates genetic expression with blood sugar levels. Figure 1 shows the probabilistic program, where the task is to get the updated belief of a gene’s occurrence in a patient given their blood sugar levels.
The first four lines of the probabilistic program in Figure 1 use a beta distribution as the prior probability of each of $T$ genes occurring in the general population. The syntax `flip(\theta)` denotes a Bernoulli random variable with success probability $\theta$. On line 5, the program uses the syntax `reduce(\mid, gene)` to denote the expression $\bigvee_{i=1}^T \text{gene}[i]$. In other words, the patient is considered to have diabetes if at least one of the genes is expressed. What follows is multiple readings of the patient’s blood sugar level. For each reading a random variable first defines the blood sugar depending on whether they have diabetes. Then we use the syntax `observe(y, v)` to condition on the random variable $y$ having the value $v$ — in the program this is used to condition on actual blood-sugar readings from the patient. Finally, on line 12, the program queries for the expectation of the posterior distribution of the occurrence of the first gene.

Figure 1. Example 1: Gene Expression

```
for i in 1:T
    gene_occurrence[i] = beta(1, 1)
    gene[i] = flip(gene_occurrence[i])
end

diabetes = reduce(\mid, gene)

blood_sugar1 = if diabetes normal(80, 2)
else normal(135, 2) end

observe(blood_sugar1, 79)

blood_sugar2 = if diabetes normal(80, 2)
else normal(135, 2) end

observe(blood_sugar2, 136)

return Expectation(gene_occurrence[1])
```

Figure 2. Scaling on Logical Constraints. HyBit scales to 50 genes with the least absolute error. The graph for Psi and HyBit overlap closely.

---

In the experiment shown we use a bitwidth of 25 — each continuous distribution is discretized as a discrete program with distributions over Boolean and fixed-point numbers. Note that now `observe` on Line 8 conditions on the fact that the value of the discrete distribution for `blood_sugar1` lies in the interval corresponding to the value 79.

Our system and approach, HyBit, scales to 50 genes and has the least absolute error among approximate inference algorithms. The program, when written in HyBit, is represented by its discrete bit-level abstraction. The user, while writing the program can employ the HyBit API to replace all continuous distributions (specifically on Lines 2, 6, 9 in Figure 1) with their bit blasted discrete approximations. As a result, we now have a discrete program with distributions over Boolean and fixed-point numbers. Note that now `observe` on Line 8 conditions on the fact that the value of the discrete distribution for `blood_sugar1` lies in the interval corresponding to the value 79.

In the experiment shown we use a bitwidth of 25 — each continuous distribution is discretized as a discrete program with distributions over Boolean and fixed-point numbers. Note that now `observe` on Line 8 conditions on the fact that the value of the discrete distribution for `blood_sugar1` lies in the interval corresponding to the value 79.
program over a tuple of 25 bits interpreted as a fixed-point number. Further, we approximate these continuous distribution using 4096 pieces, each of which is a bit blasted exponential distribution. Naive discretization with 25 bits would be prohibitively slow, as it yields $2^{25}$ intervals i.e. 134M intervals. However, our bit blasted program only uses 53K coin flips (Boolean random variables) to represent them. Moreover, knowledge compilation based inference [Fierens et al. 2015; Holtzen et al. 2020] automatically identifies and exploits conditional independences in the program’s logical structure and helps to scale inference. More details of this experiment can be found in the appendix.

2.2 Handling Multimodality

This section presents an example of a multimodal distribution to highlight another challenge for inference on hybrid probabilistic programs. Multimodal distributions have multiple peaks separated by low probability regions. These distributions commonly emerge in various applications such as sensor network localization, cosmology and many more [Shaw et al. 2007; Tak et al. 2018]. We adapt an example from the existing literature [Yao et al. 2021], as shown in Figure 3.

The probabilistic program shown in Figure 3 is hard for existing probabilistic inference approaches. The datapts on Line 3 has nine entries, with two-thirds being 5 and the other one-third being $-5$. This leads to the posterior for $(\mu_1, \mu_2)$ being bimodal around $(5, -5)$ and $(-5, 5)$. As the number of data points increases with the same proportion, the posterior of $(\mu_1, \mu_2)$ converges to $(5, -5)$. However, in the presence of limited data points, the posterior for $\mu_1$ is bimodal around 5 and $-5$.

The existence of multiple modes challenges sampling-based algorithms as they tend to get stuck in one of the modes. Specifically, WebPPL using MCMC with a Metropolis Hastings kernel and both Stan and WebPPL using HMC end up arbitrarily in one of the modes and fail to explore the other mode. Figures 4a and 4b show the results obtained using WebPPL MCMC and Stan HMC respectively, where two different runs get stuck in two different modes. On the other hand, SMC and AQUA with its adaptive interval strategy only finds the more probable mode.

Figure 4. Posterior from different baselines compared with that of HyBit for Figure 3. HyBit and WebPPL SMC are able to identify both the modes in the posterior distribution. (a) Different runs of WebPPL MCMC with a Metropolis Hastings kernel converge to different modes. (b) Different runs of Stan HMC converge to different modes. (c) Different runs of WebPPL SMC are able to find both the modes. (d) AQUA with its adaptive interval strategy only finds the more probable mode.
hand, HyBit performs exact inference on its discrete abstraction and so explores the distribution globally, allowing it to identify both modes. Sequential Monte Carlo (SMC) (Figure 4c), Psi and GuBPI also explore the distribution globally and thus, are able to find both modes. Finally, to address the computational challenge of direct discretization, AQUA adapts its discretizing intervals to focus on high probability regions, and ends up only identifying the higher probability mode (Figure 4d).

2.3 Handling Low Probability Observations

This section presents conjugate Gaussians (Figure 5) with low probability observations. In Figure 5, the posterior distribution for random variable \( \mu \) is queried after conditioning on low probability data on Lines 2 and 3.

Why are low probability observations hard? Intuitively, general-purpose sampling algorithms begin sampling from the prior distribution and struggle to find samples with considerable weight. Only after a very large number of samples do these algorithms manage to sample from the true posterior. On the other hand, HyBit explores the domain of the posterior distribution globally, via exact inference on a bit blasted abstraction, and so it is insensitive to this issue.

Figure 6 plots the true prior and posterior distributions along with results from different inference algorithms. For the sampling-based algorithms MCMC with a Metropolis Hastings kernel and SMC, we obtained and plotted 1000 samples after running the corresponding WebPPL [Goodman and Stuhlmüller 2014] program. The importance sampling algorithm was not able to obtain any sample with non-negligible weight for this program. The samplers are shifting the posterior towards the true posterior distribution but require many more samples to achieve that. Even after sampling about 16M and 65K samples respectively, the expectation of the samples obtained from MCMC and SMC have absolute error of 0.549798 and 1.520776. GuBPI reports upper and lower bounds on the probability for each interval and incurs an absolute error of 2.33. On the other hand, the posterior distribution from HyBit overlaps perfectly with the ground truth. Stan HMC handles low probability observations well and obtains high accuracy. Psi also obtains the exact symbolic expression for the posterior distribution. Finally, the mean of AQUA’s reported posterior had an error of 5.66 as it fails to make any update to the prior of \( \mu \).

3 BIT BLASTING: KEY INSIGHTS

To scale inference on hybrid probabilistic programs with respect to discrete structure, we need an algorithm that treats discreteness as first class, and that discretizes away continuous structure. This section defines the semantic notion of bit blasting and sets it up as a special case of discretization with desirable properties. Then, we provide bit blasting functions for common classes of continuous
distributions. We provide discretization techniques that are sound (accurate up to $b$ bits), succinct, and amenable to efficient inference.

### 3.1 Discretization and Bit Blasting

In the standard terminology of probability theory [Rosenthal 2006], a probability space $(\Omega, \Sigma, \mu)$ consists of a sample space $\Omega$, a $\sigma$-algebra on $\Omega$ denoted $\Sigma$, and a probability measure on $\Sigma$ denoted $\mu$.

In a general sense, a discretization function takes as input such a probability space $(\Omega, \Sigma, \mu)$ and outputs a discrete probability space $(\Omega_D, \Sigma_D, \mu_D)$ where $\Omega_D$ is a countable set.

We will study a more specific notion of discretization: one that takes as input a continuous distribution over a finite interval and outputs a discrete distribution over $2^b$ points for some number of bits $b$. Formally, let $[l, u)$ be an interval with $l, u \in \mathbb{R}$. For the input probability space, we use $\mathcal{B}([l, u))$ to denote the Borel $\sigma$-algebra of subsets of interval $[l, u)$ [Rosenthal 2006]. For the output probability space, we write $\mathcal{P}(S)$ to refer to the power set (a $\sigma$-algebra) of set $S$. Moreover, we will assume the sample space to be discretized as follows.

**Definition 1** ($b$-bit interval). A $b$-bit interval $[l, u)_b$ is the set of points obtained by dividing $[l, u)$ into $2^b$ intervals: $[l, u)_b = \{r \mid r2^b \in \mathbb{Z}, r \in [l, u)\}$

We are now ready to define the notion of discretization function used in this paper.

**Definition 2** ($b$-bit discretization function). A $b$-bit discretization function takes as input a probability space $([l, u), \mathcal{B}([l, u)), \mu)$, a bit width $b \in \mathbb{Z}^+$ and outputs a discrete probability space $([l, u)_b, \mathcal{P}([l, u)_b), \mu_D)$ for some measure $\mu_D$.

**Example 1.** Let $\pi$ be a uniform distribution on the unit interval, described by the probability space $([0, 1), \mathcal{B}([0, 1)), \mu)$ such that the probability density function specified by $\mu$ is 1 on the unit interval. Consider a function $d_1$ that takes $\pi$ as input and outputs the probability space $(S, \mathcal{P}(S), \mu_D)$ where $S = \{0, 0.25, 0.5, 0.75\}$ and $\mu_D$ is a probability measure on $\mathcal{P}(S)$. Then $d_1$ is a 2-bit discretization function. As a concrete example $\mu_D$ can be defined as $\{0 \rightarrow 0.1, 0.25 \rightarrow 0.2, 0.5 \rightarrow 0.3, 0.75 \rightarrow 0.4\}$².

As the example shows, one can come up with any arbitrary $b$-bit discretization function. To qualify them further, we need a notion of accuracy — soundness up to $b$ bits defined as follows.

**Definition 3** (soundness of $b$-bit discretization function). For any integer $b > 0$, a $b$-bit discretization function is $b$-sound for a particular input probability space $([l, u), \mathcal{B}([l, u)), \mu)$ if it outputs a discrete probability space $([l, u)_b, \mathcal{P}([l, u)_b), \mu_D)$ such that the following holds:

$$\forall x \in [l, u)_b \quad \int_x^{x + \frac{b}{2^b}} d\mu(y) = \mu_D(\{x\})$$

**Example 2.** Let $d_2$ be a $b$-bit discretization function that takes $\pi$ (as defined in Example 1) as input and outputs the probability space $(S, \mathcal{P}(S), \mu_D)$ where $S = \{0, 0.25, 0.5, 0.75\}$ and $\mu_D$ can be defined as $\{0 \rightarrow 0.25, 0.25 \rightarrow 0.25, 0.5 \rightarrow 0.25, 0.75 \rightarrow 0.25\}$. Then $d_2$ is a sound 2-bit discretization function for the probability space $\pi$.

Before we define a $b$-bit blasting function, we need to fix a generic representation of discrete probability distributions. To that purpose, we define the concept of a discrete probabilistic closure, akin to probabilistic Turing machines [Arora and Barak 2006].

Each probabilistic closure is a deterministic function from a set of biased coin flips to a discrete set. This induces a probability distribution on the output of the function through probabilities

---

²We define discrete probability measures $\mu_D$ using a mapping from single elements in the sample space $S$ to their probabilities, which can then be used to compute $\mu_D$ for any set in $\mathcal{P}(S)$. 

---

Proc. ACM Program. Lang., Vol. 8, No. PLDI, Article 182. Publication date: June 2024.
associated with coin flips. It also consists of an accepting Boolean formula that handles observations and limits the set of values that input flips can take. We provide a formal definition below:

**Definition 4 (discrete probabilistic closure).** A discrete probabilistic closure is defined as the tuple \((\varphi, \gamma, w)\) where \(w\) is a vector of Boolean random variables (or biased coin flips), \(\varphi\) is a deterministic function from \(\{T,F\}^{|w|}\) to a finitely countable set \(S\) and \(\gamma\) is a deterministic Boolean formula over variables in \(w\).

The semantics of a discrete probabilistic closure, i.e. \(\langle(\varphi, \gamma, w)\rangle\) defines a probability space \((S, \mathcal{P}(S), \mu)\) such that

\[
\forall T \subseteq S, \mu(T) = \frac{E_w((\varphi(w) \in T) \land \gamma)}{E_w(\gamma)} \quad \text{if } E_w(\gamma) \neq 0
\]

**Example 3.** From Example 2, consider \(\mu_D = \{0 \rightarrow 0.25, 0.25 \rightarrow 0.25, 0.5 \rightarrow 0.25, 0.75 \rightarrow 0.25\}\). \(\mu_D\) can be represented using a discrete probabilistic closure using the function \(\text{naïve}_\text{unif}\), the weight function \(w\) and the accepting Boolean formula \(\gamma\) as follows:

\[
\gamma = T \\
w = [f_0 \rightarrow \text{flip} \frac{1}{4}, f_1 \rightarrow \text{flip} \frac{1}{3}, f_2 \rightarrow \text{flip} \frac{1}{2}]
\]

Now, one can calculate the probability that \(\text{naïve}_\text{unif}\) returns 0.5 as follows:

\[
Pr(\text{val} = 0.5) = Pr(\neg f_0)Pr(\neg f_1)Pr(f_2) = (1 - 0.25)(1 - 0.33)0.5 = 0.25
\]

Note that as explained by Example 3, any discrete distribution \(\mu_D\) over \(2^b\) values can be represented as a discrete probabilistic closure \((\varphi, \gamma, w)\) where \(|w| = 2^b - 1\).

Dice [Holtzen et al. 2020] and Problog [Fierens et al. 2015] are examples of PPLs that directly fit into the paradigm of a discrete probabilistic closure.

We want a discrete probabilistic closure to be more succinct — of size polynomial in the number of bits of precision, \(b\). To this purpose, we define a bit blasting function.

**Definition 5 (b-bit blasting function).** A \(b\)-bit blasting function \([.]_b\) is a \(b\)-bit discretization function that outputs a discrete probabilistic closure \((\varphi, \gamma, w)\) that uses a number of Boolean random variables that is polynomial in the number of bits \(b\), that is, \(|w| \in O(\text{poly}(b))\).

It follows from Definitions 3, 4 and 5 that for an integer \(b > 0\), a \(b\)-bit blasting function is sound for a given probability space \(((\lfloor l, u\rfloor), \mathcal{B}(\lfloor l, u\rfloor), \mu)\) if

\[
\forall x \in [l, u]_b \int_x^{x + \frac{1}{2^b}} d\mu(y) = \langle[\mu]_b\rangle
\]

**Figure 7.** Commutative diagram for a \(b\)-sound bit blasting function
Example 4. Again, consider the result of the sound 2-bit discretization function from Example 2, that is \(\tilde{\mu}_D = \{0 \rightarrow 0.25, 0.25 \rightarrow 0.25, 0.5 \rightarrow 0.25, 0.75 \rightarrow 0.25\}\). Measure \(\tilde{\mu}_D\) can alternatively be represented using a discrete probabilistic closure using the function \(\text{bitblast}_\text{unif}\), the weight function \(w\) and the accepting Boolean formula \(\gamma\) as follows:

\[
\gamma = \top \quad \quad w = [f_0 \rightarrow \text{flip } \frac{1}{2}, f_1 \rightarrow \text{flip } \frac{1}{2}]
\]

Note that in \(\text{bitblast}_\text{unif}\), \(f_0\) and \(f_1\) are being used as bits in the binary representation of the variable \(\text{val}\). Now, one can calculate the probability that \(\text{bitblast}_\text{unif}\) returns 0.5 as follows:

\[
\Pr(\text{val} = 0.5) = \Pr(f_0)\Pr(\neg f_1) = (0.5)(1 - 0.5) = 0.25
\]

We would like to point out that \(\text{bitblast}_\text{unif}\) uses only 2 coin flips to represent a distribution over 4 values. It can be generalized to use \(b\) coin flips to represent a uniform distribution over \(2^b\) values. On the other hand, \(\text{naive}_\text{unif}\) uses 3 coin flips and would generalize to use \(2^b - 1\) coin flips. Hence, \(\text{bitblast}_\text{unif}\) is a viable output for a \(b\)-bit blasting function. In the next section, we provide details for the \(b\)-bit blasting functions for exponential and mixed-gamma distributions.

3.2 Concrete Bit Blasting Function: Preliminaries

Next, our goal is to provide a concrete instantiation of a sound bit blasting function for mixed gamma distributions, which have probability density functions as defined below.

**Definition 6 (generalized-gamma density).** Given parameters \(\alpha \in \mathbb{Z}^+\) and \(\beta \in \mathbb{R}\), a generalized-gamma density \(\pi_{\alpha,\beta}\) is a probability density function over the interval \([0, 1]\) of the form

\[
\pi_{\alpha,\beta}(x) = \frac{x^\alpha e^{\beta x}}{\int_0^1 y^\alpha e^{\beta y} dy}.
\]

**Definition 7 (mixed-gamma density).** Given a collection of \(N \in \mathbb{Z}^+\) generalized-gamma densities \(\pi_{\alpha_i,\beta_i}\) with their associated weights \(a_i \in [0, 1]\) such that \(\sum_{i=1}^N a_i = 1\), a mixed-gamma density \(Y\) is a probability density function over the interval \([0, 1]\) of the form

\[
Y(x) = \sum_{i=1}^N a_i \pi_{\alpha_i,\beta_i}(x).
\]

For notational convenience, we confine the continuous distributions to the unit interval to get discrete distributions over a \(b\)-bit unit interval. We generalize our approach to any finite interval for building the probabilistic programming system HyBit based on bit blasting.

To describe our construction of the bit blasting function, we make use of Dice [Holtzen et al. 2020]. Dice already compiles its programs to weighted Boolean formulas (via the \(\rightsquigarrow\) judgement) that fit the definition of a discrete probabilistic closure.\(^3\) This allows us to only define a \(\rightsquigarrow_p\) judgment from probability density functions to Dice programs to specify a bit blasting function. Dice also defines a distributional semantics function \([\_]\) \(D : p \rightarrow V \rightarrow [0, 1]\) that takes as input a Dice program \(p\) and outputs a normalized probability distribution (described as a function from the set of values \(V\) to probabilities). We use the function \([\_]\) \(D\) to argue about soundness of our construction later. More details about syntax and semantics of Dice can be found in the appendix.

---

\(^3\)Dice compiles to weighted Boolean formulas \((\varphi, \gamma, w)\) where \(\varphi\) outputs (tuples of) Boolean values, \(\gamma\) represents observations in a Dice program and \(w\) consists of weights associated with Boolean variables (biased coin flips).
The compilation judgement for mixed-gamma densities are of the form $Y \leadsto_Y p$ where $p$ are Dice programs. We further provide the following definitions for $b$-equivalence between densities and Dice programs and $b$-succinctness of Dice programs.

**Definition 8 (binarizing function).** The binarizing function for $b$ bits, $(\cdot)_b$, takes as input a number $r \in [0, 1]_b$ and returns a $b$-bit tuple $(v_1, (v_2, \ldots, v_b))$ such that $r = \sum_{i=1}^{b} \frac{v_i}{2^i}$

**Definition 9 (b-equivalence).** A mixed-gamma density $Y$ and a Dice program $p$ are $b$-equivalent for some $b \in \mathbb{Z}^+$ if for all $r \in [0, 1]_b$

$$\int_{r^{1+b}} r^+ \ Y(y) \ dy = [p]_D ([r]_b)$$

Note that $b$-equivalence is analogous to $b$-soundness but is specialized for Dice programs.

**Definition 10.** The flip count function $flip\_count(\cdot)$ takes as input a Dice program $p$ and outputs the number of Boolean random variables (coin flips).

**Definition 11.** Compilation $\leadsto_Y$ is $b$-succinct for $Y$ if $\exists k > 0, \forall b \in \mathbb{Z}^+$ such that if $Y \leadsto_Y p$, then $flip\_count(p) \leq kb$

We define $b$-succinctness such that the Dice program $p$ employs coin flips linear in the number of bits $b$. Observe that $b$-succinctness imposes a stricter condition than that required by a $b$-bit blasting function (which requires poly($b$) coin flips). This implies that if we have a $b$-succinct judgment for a mixed-gamma density, then we can have a $b$-bit blasting function for that distribution. We describe this in more detail later.

### 3.3 Judgment $\leadsto_Y$ and bit blasting

This section describes the rules for the judgement $\leadsto_Y$. We first describe the rules for an exponential distribution and then move on to generalized gamma distributions. Finally, we describe the rule for mixed-gamma densities. For each of the rules $Y \leadsto_Y p$, we prove the $b$-equivalence of the density $Y$ and the Dice program $p$ and the $b$-succinctness of $p$. Detailed proofs can be found in the appendix. We also describe how $\leadsto_Y$ allows us to construct a bit blasting function for mixed-gamma densities.

#### 3.3.1 Exponential Distribution, $\pi_{0,b}$.

Let us first consider the uniform distribution $(\pi_{0,0})$, a special case of an exponential distribution. If we bit blast a uniform distribution using $b$ bits into $2^b$ intervals, we end up with a discrete distribution $D_b$ over $[0, 1]_b$ with $2^b$ discrete points each having probability $\frac{1}{2^b}$. A straightforward discretization strategy enumerates $2^b$ values using coin flips. But the same can be achieved using a tuple of $b$ bits, where each bit is an unbiased coin $flip(0.5)$.

The strategy to bit blast a uniform distribution using its binary representation works because of the independence between binary digits. The same strategy can be extended to general exponential distributions as well. This fact was shown in a classic paper in the statistics literature [Marsaglia 1971]. We formalize that idea using the following rule.

**Definition 12.** The function $flip\_param \colon \mathbb{R} \times \mathbb{Z}^+ \rightarrow [0, 1]$ is defined $flip\_param(\beta, b) = \frac{\beta}{1 + e^{\beta \beta}}$.
We prove the $b$-equivalence of the density $\pi_{0,\beta}$ and the Dice program in rule Expo0 and $b$-succinctness of the latter in the following lemmas. It is more straightforward to see the succinctness of the Dice program — it uses only $b$ coin flips.

**Lemma 1.** $\forall b \in \mathbb{Z}^+, \beta \in \mathbb{R}, p$, if $\pi_{0,\beta} \sim_{b} p$, then $\pi_{0,\beta}$ and $p$ are $b$-equivalent.

**Lemma 2.** $\forall \beta \in \mathbb{R}, \sim_{b}$ is $b$-succinct for $\pi_{0,\beta}$

We provide detailed proofs of the above lemmas in the appendix. For rest of the paper, we consider exponential distributions as primitives to build other distributions, since these are the only ones that enjoy the property of independent bits. But, sound bit blasting functions are still possible for other distributions, as we show next.

### 3.3.2 Gamma Distribution $\pi_{1,\beta}$

To come up with a sound bit blasting function for $\pi_{1,\beta}$, we present a key mathematical insight. Consider the program in Figure 8a. Continuous random variables $X$ and $Y$ have a uniform ($\pi_{0,0}$) and exponential ($\pi_{0,\beta}$) distribution respectively. It returns the new distribution of $X$ after conditioning on the inequality $Y < X$. It turns out that the posterior distribution is a specific gamma distribution $\pi_{1,\beta}$. We show the resulting calculation below, where $pdf$ refers to the probability density function.

$$pdf(X | Y < X) \propto \int_{y=0}^{1} pdf(Y) \cdot pdf(X) \cdot 1(Y < X) \ dy = \int_{y=0}^{X} 1 \cdot e^{\beta y} \ dy \propto xe^{\beta x} \quad (1)$$

What happens if we discretize the program in Figure 8a using $b$ bits? We get the program in Figure 8b where each continuous random variable has been replaced with its bit blasted counterpart ($X$ replace by $X_b$ and so on). We have already seen that for uniform and exponential distributions, $b$-equivalent Dice programs exist. But what about the other constructs? As Figure 8d demonstrates, observe $(Y_b < X_b)$ incurs error over its continuous counterpart observe $(X < Y)$. The good news is that we can account for the error as shown by the following equations:

$$Pr(X_b | Y < X) = Pr(X_b, Y_b < X_b | Y < X) + Pr(X_b, Y_b == X_b | Y < X)$$
$$= Pr(X_b | Y_b < X_b) \cdot Pr(Y_b < X_b | Y < X) + Pr(X_b | Y_b == X_b, Y < X) \cdot Pr(Y_b == X_b | Y < X).$$

The correction term in the above equations when computed algebraically turns out to be proportional to the exponential density ($\pi_{0,\beta}$). So now, the resulting discrete probabilistic program after correction looks as shown in Figure 8c where $\theta = Pr(Y_b == X_b | Y < X)$. 

---

Proc. ACM Program. Lang., Vol. 8, No. PLDI, Article 182. Publication date: June 2024.
Y \sim \text{uniform}(0, 1)

X \sim \text{exponential}(\beta)

\text{return}(Y < X)

\text{return } X

\text{C} \sim [\text{exponential}(\beta)]_b

\text{return}(\text{Y}_b < \text{X}_b)

\text{return } \text{X}_b

\text{Y}_b \sim [\text{uniform}(0, 1)]_b

\text{X}_b \sim [\text{exponential}(\beta)]_b

\text{C} \sim [\text{exponential}(\beta)]_b

\text{observe}(\text{Y}_b < \text{X}_b)

\text{observe}(\text{Y}_b < \text{X}_b)

\text{Z} = \text{if flip}(\theta) \text{ then } \text{C} \text{ else } \text{X}_b

\text{return } \text{Z}

(a) Continuous probabilistic program for density $xe^{bx}$

(b) bit blasted probabilistic program analogous to program 8a

(c) Sound bit blasted probabilistic program for density $xe^{bx}$

(d) Sound $b$-bit blasting of $xe^{bx}$. Prior densities for $X$ and $Y$ (shown in blue) when conditioned on $Y < X$ (shown in orange) returns the posterior for $X$ (shown in orange).

Figure 8. Key insight in bit blasting $xe^{bx}$ probability density. $[.]_b$ refers to discretization of a continuous density into $2^b$ intervals and $X_b$ refers to the discretization of $X$.

The rule Expo1 and Trans-expolzero (in the appendix) captures the above intuition. Here, $\text{unifObs}(y, b) = p$ is a helper judgment where $p$ constructs a uniform distribution that it conditions through observe on being less than $y$.

$$\pi_{0, \beta} \xrightarrow{b} p_1 \quad \text{unifObs}(y, b) = p_3 \quad \beta \neq 0 \quad \theta = \frac{(e^{\beta}z^{-b-1}(\beta z^{-b}-1)+1)(1-e^\theta)}{(1-e^{\beta}z^{-b})(\theta \beta^{-1}+1)}$$ (Expo1)

We prove the $b$-equivalence and $b$-succinctness. The proof for $b$-equivalence is much more involved but it is easy to see that the Dice program in the above rule uses $3b + 1$ coin flips: $b$ coin flips in each occurrence of $p_1$, $b$ coin flips in $p_1$ and 1 coin flip in the if then else guard to create a mixture.

Lemma 3. $\forall b \in \mathbb{Z}^+, \beta \neq 0 \in \mathbb{R}, p, \text{if } \pi_{1, \beta} \xrightarrow{b} p, \text{ then } \pi_{1, \beta} \text{ and } p \text{ are } b\text{-equivalent.}$

Lemma 4. $\forall \beta \in \mathbb{R}, \xrightarrow{b} \text{ is } b\text{-succinct for } \pi_{1, \beta}$
3.3.3 Generalized Gamma Distribution $\pi_{\alpha,\beta}$. The previous subsection shows how conditioning on the inequality ($Y < X$) introduces a linear factor to $\pi_{0,\beta}$ to obtain $\pi_{1,\beta}$. Note that conditioning on ($Y < X$) introduces a linear factor regardless of what initial probability density $X$ had. That is, if $X$'s initial probability density was $f(x)$, the resulting density of the following probabilistic program would be $xf(x)$. This implies that if we can bit blast $f(x)$, we can bit blast $xf(x)$. We still need to account for the error incurred by $\text{observe} (Y_b < X_b)$ i.e. $\Pr(X_b|Y_b = X_b, Y < X)$. It turns out that the correction term is a mixture of gamma distributions which can be bit blasted soundly as well. We provide the judgement rule and proof of the following lemma in the appendix.

**Lemma 5.** $\forall b, \alpha \in Z^+, \beta \in R, p$, if $\pi_{\alpha,\beta} \leadsto_b^p$ then $\pi_{\alpha,\beta}$ and $p$ are $b$-equivalent.

**Lemma 6.** $\forall \beta \in R, \alpha \in Z^+, \leadsto_b$ is $b$-succinct for $\pi_{\alpha,\beta}$

3.3.4 Mixture of Gamma Distributions $\sum_i a_i \pi_{a_i,\beta_i}$. Since generalized gamma densities $\pi_{\alpha,\beta}$ can be bit blasted, mixed gamma densities can be bit blasted as well. One bit blasts each individual generalized gamma density and creates their mixture using if then else constructs as follows.

$$\sum_{i=1}^N a_i \pi_{a_i,\beta_i} \leadsto_b$$

<table>
<thead>
<tr>
<th>$N &gt; 1$</th>
<th>$\pi_{\alpha_N,\beta_N} \leadsto_b p_1$</th>
<th>$\sum_{i=1}^{N-1} \frac{a_i}{a_N} \pi_{a_i,\beta_i} \leadsto_b p_2$</th>
<th>$\forall i, a_i \in [0, 1]$</th>
<th>$\sum_{i=1}^N a_i = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>let $y_1 = \text{flip}(a_N)$ in</td>
<td>let $y_2 = p_1$ in</td>
<td>let $y_3 = p_2$ in</td>
<td>if $y_1$ then $y_2$ else $y_3$</td>
<td></td>
</tr>
</tbody>
</table>

We prove the following theorems with details in the appendix.

**Theorem 7.** $\forall Y, b, \alpha \in Z^+, p$, if $Y \leadsto_b$ then $Y$ and $p$ are $b$-equivalent.

**Theorem 8.** $\leadsto_b$ is $b$-succinct for all mixed-gamma densities $Y$.

Now that we have specified the rules for judgment $\leadsto_b$, we specifically define a sound bit blasting function for all mixed-gamma densities $Y$, that is $\leadsto \circ \leadsto_b$.

**Theorem 9.** $\leadsto \circ \leadsto_b$ is a sound $b$-bit blasting function.

Earlier work [Holtzen et al. 2020] defines the judgment $\leadsto$ that takes as input a Dice program $p$ and outputs a weighted Boolean formula $(\varphi, \gamma, w)$ that aligns with Definition 4 of a discrete probabilistic closure. And since $\leadsto_b$ is $b$-succinct for all mixed-gamma densities by Theorem 8, $\leadsto$ always outputs a $w$ with poly($b$) coin flips. Thus, $\leadsto \circ \leadsto_b$ is a $b$-bit blasting function. Earlier work [Holtzen et al. 2020] also proves the correctness of compilation to weighted Boolean formula with respect to the semantics of the Dice program. This fact combined with Theorem 7 concludes that $\leadsto \circ \leadsto_b$ is a sound $b$-bit discretization function. Detailed proofs can be found in the appendix.

3.3.5 Example: Laplace Distribution. Previous sections described how mixed gamma densities can be bit blasted when they are confined to a unit interval. But how can distributions that are shifted or scaled to other finite intervals be bit blasted? We explain it through the example of a Laplace distribution. A Laplace distribution has two parameters: location ($\mu$) and scale ($b$) and has the probability density function as described below where $x \in R$.

$$\text{Laplace}(x|\mu, b) = \frac{1}{2b} e^{-|x-\mu|/b} = \begin{cases} \frac{1}{2b} e^{\mu} e^{-\frac{x-\mu}{b}} & x \geq \mu \\ \frac{1}{2b} e^{-\mu} e^{\frac{x-\mu}{b}} & x < \mu \end{cases}$$
Let us consider the Laplace distribution truncated at the interval \([\mu - r, \mu + r]\). We assume that \(r\) would be a suitable power of 2 allowing product with \(r\) (denoted by \(r \times p\)) to be just a decimal shift and \(\mu\) to be a \(b\)-bit representable number allowing precise shifting of \(p\). First we generate the exponentials scaled for an interval of width \(r\) instead of width 1:

\[
\pi_{0, \frac{r}{2}} \overset{\sim}{\longrightarrow}_{b} p_1 \quad \pi_{0, \frac{r}{2}} \overset{\sim}{\longrightarrow}_{b} p_2
\]

And then we create a mixture of them:

\[
\text{Laplace}(x|\mu, b) \overset{\sim}{\longrightarrow}_{b} \text{if flip 0.5 then } \mu + r \times p_1 \text{ else } \mu - r \times p_2
\]

Since \(p_1\) and \(p_2\) use \(b\) coin flips each, the program shown above uses \(2b\) coin flips and \(\overset{\sim}{\longrightarrow}_{b} \circ \overset{\sim}{\longrightarrow}_{b}\) is a sound bit blasting function for Laplace distributions as well.

### 3.4 How does bit blasting help in inference?

We have demonstrated sound bit blasting functions for mixed gamma distributions. But how does that help in inference for probabilistic programs with these distributions? To answer this question, we focus on a particular inference strategy - that is knowledge compilation. We first describe the necessary preliminaries about knowledge compilation and then argue about how the programs obtained through \(\overset{\sim}{\longrightarrow}_{b}\) are efficient for knowledge compilation.

Knowledge compilation based approaches [Fierens et al. 2015; Holtzen et al. 2020] for exact discrete probabilistic inference compile discrete probabilistic programs into weighted Boolean formulas that are represented using (reduced) ordered binary decision diagrams (OBDDs). These OBDDs are single rooted in case of a single Boolean random variable being returned and multi-rooted in case of a tuple of Boolean random variables being returned. By fixing a value for all the flips (biased coin flips with associated weights) in the program, and by traversing the OBDD following those values (solid line for true, dashed for false), we reach the terminal corresponding to the value of each bit. The operation of weighted model counting computes the probability of reaching the 1-terminal for each bit in the returned value. It is a dynamic programming algorithm that runs in time linear in the OBDD size. The size of an OBDD for a Boolean formula \(\phi\), denoted \(\text{OBDD}(\phi)\), is the number of nodes in the OBDD. Thus, if we can obtain a smaller OBDD representation for a distribution, we can efficiently compose it with other constructs in a discrete probabilistic program.

We discuss and prove formally how every program obtained through the judgement \(\overset{\sim}{\longrightarrow}_{b}\) compiles into a weighted Boolean formula that compiles to a multi-rooted OBDD that grows linearly in the number of bits as opposed to the worst case exponentially. Recall that Dice programs compile to weighted Boolean formula \((\phi, \gamma, w)\) where \(\phi\) is the (tuple of) Boolean formulae corresponding to the return value of the program, \(\gamma\) is the accepting Boolean formula to encode observations and \(w\) is the weight function with flip probabilities.

**Theorem 10.** \(\forall \gamma, p, \phi, \gamma, w, \exists k, \forall b, \text{if } \gamma \overset{\sim}{\longrightarrow}_{b} p \text{ and } p \overset{\sim}{\longrightarrow}_{b} (\phi, \gamma, w)\), then there exists a variable order \(\Pi\) of Boolean random variables in \(w\) such that \(\text{OBDD}(\phi) + \text{OBDD}(\gamma) \leq kb\)

We now provide intuition for the proof of the above theorem. Note that in the programs obtained through the judgement \(\overset{\sim}{\longrightarrow}_{b}\), there are only two constructs that depend on the number of bits: (1) construction of exponential distribution \(\pi_{0, \beta}\), and (2) conditioning on inequality between an exponential and a uniform distribution through \(\text{unifObs}(\gamma, b)\). We provide intuition how the OBDD size for these constructs increase linearly in the number of bits, \(b\).

#### 3.4.1 Exponential Distribution

In Figure 9, we 3-bit blast an exponential distribution, i.e. we have a discrete exponential distribution over 8 values \((0, 0.125, 0.25, \ldots, 0.875)\). The figure shows a 3-rooted OBDD where each root labeled as \(b_1, b_2\) and \(b_3\) represents a bit in the returned value of 3 bits. Consider that we fix the value of the flip corresponding to the node \(f_1\) to be true, then for \(b_1\), we would
reach the terminal 1 and its value would be assigned 1. The operation of weighted model counting (WMC) would calculate the probability of $b_1$ to be 1 as $6.14e-6$ as node $f_1$ has probability $6.14e-6$ to be true. Similarly WMC can be used for other roots of this OBDD. Since each bit needs one OBDD node, the overall OBDD size grows linearly with the number of bits. Since WMC runs in time linear in the OBDD size, probabilistic inference for an exponential distribution would run linear in the number of bits $O(b)$. Another example of OBDD for a uniform distribution can be found in the appendix.

For all programs $p$ obtained through the judgement $\leadsto_b$, if $p \leadsto_b (\varphi, y, w)$, then $\varphi$ is the tuple of Boolean formulae representing the return value of the program. We argue that the return value of $p$ is always a mixture of exponential distributions making its OBDD size linear in the number of bits.

3.4.2 Conditioning on inequality between an exponential and a uniform distribution. The rules for judgment $\leadsto_b$ use the helper judgment $\text{unifObs}(y, b) = p$ to condition on an inequality between binary representations of a uniform distribution and an exponential distribution.

**Definition 13 (Inequality function).** A $b$-bit inequality function, $\text{LT}_b : \{0, 1\}^b \times \{0, 1\}^b \rightarrow \{0, 1\}$ takes as input two $b$-bit numbers $x, y$ and outputs 1 if $x < y$ and 0 otherwise. Thus,

\[
\text{LT}_b((x_1, \ldots, x_b), (y_1, \ldots, y_b)) = \begin{cases} 
\neg x_1 y_1 & b = 1 \\
\neg x_1 y_1 + (\neg x_1 \neg y_1 + x_1 y_1) \text{LT}_{b-1}(((x_2, \ldots, x_b), (y_2, \ldots, y_b)) & b > 1 
\end{cases}
\]

We prove the following lemma which states that the OBDD size for the inequality function grows linearly with the number of bits.

**Lemma 11.** $\exists k, \forall b$, for the variable order $x_1, y_1, x_2, y_2, \ldots, x_b, y_b$, the size of the OBDD, that is $\text{OBDD}(\text{LT}((x_1, x_2, \ldots, x_b), (y_1, y_2, \ldots, y_b))) \leq kb$.

Since the only constructs that depend on the number of bits (exponential distributions and inequalities) grow linearly with the number of bits, Theorem 10 holds intuitively. We provide a formal proof in the appendix.

4 HYBIT: A PROBABILISTIC PROGRAMMING SYSTEM

The previous section described how to bit blast mixed-gamma distributions. We further use it to build a probabilistic program system HyBit for hybrid probabilistic programs. This section describes its syntax and implementation and elaborates on two important aspects: piece-wise approximations of continuous distributions and advantages of a binary representation.

---

1. $r ::= \text{Bool} | \text{DistFix}(n, n)$
2. $v ::= T | F | \text{DistFix}(n, n)(r)$
3. $e ::= x | v | \text{flip} \theta | \text{generalGamma}(n, n, n, r, r, r) | \text{bitblast}(n, \text{pdf}, n, b, r, r)$
   $\quad | \text{if e then e else e} | \text{observe e} | \text{op}^n(e_1, \ldots, e_n)$

Figure 10. Syntax for the core HyBit expressions. \text{DistFix}(n, n) refers to the type of fixed-point numbers. The metavariable $r$ ranges over real numbers, $n$ over integers, $b$ over Booleans, $x$ over variable names, and $\theta$ over real numbers in the range $[0, 1]$. The metavariable \text{pdf} ranges over continuous density functions \text{pdf} : R \rightarrow R. \text{op}^n ranges over arithmetical operations with $n$-arity. We explain the API \text{DistFix}, \text{generalGamma} and \text{bitblast} and their arguments in Figure 11.

Proc. ACM Program. Lang., Vol. 8, No. PLDI, Article 182. Publication date: June 2024.
4.1 HyBit — Syntax and Implementation Details

We build a probabilistic programming system HyBit around sound bit blasting of mixed-gamma densities and approximate bit blasting of other continuous distributions. HyBit has been implemented as a shallow embedded domain specific language in Julia [Bezanson et al. 2017].

The core syntax of HyBit expressions is given in Figure 10. It provides support for distributions over Booleans (flip θ) and fixed-point numbers (general_gamma and bitblast). It supports Boolean operations (¬, ∧, ∨) and arithmetic operations (+, -, *, /, %, <, ==) over these distributions as well as hard observations for probabilistic conditioning (observe). For all the constructs in Figure 10, HyBit performs a non-standard execution and compiles them to OBDDs to perform probabilistic inference. Since HyBit has been implemented as a library in Julia, HyBit programmers can also make use of Julia constructs such as (bounded) loops, tuples and functions. As an example, a for loop from Julia can be used with HyBit constructs in the loop body to build a probabilistic model. HyBit is available as an open source repository with a comprehensive set of examples. 4

Figure 11 contains more details on the API of HyBit. DistFix{W, F} is the type of fixed point numbers of bitwidth W with F bits after the binary point. The function general_gamma performs sound bit blasting of a specified generalized gamma density to the given fixed-point W and F. Sound bit blasting of mixed-gamma densities is achieved by using the if-then-else construct over generalized gamma densities. The function bitblast is used for bit blasting any arbitrary continuous distribution using piece-wise approximation, employing the bits and pieces specified using the parameters W, F, and pieces. The API also allows the user to choose the type of discrete distribution — linear or exponential — for the piece-wise approximation. The parameters of linear

---

4https://github.com/Tractables/Dice.jl/tree/hybit
(slope) and exponential ($\beta$) pieces are automatically chosen such that the ratio of probabilities of the first and last interval is the same as that for the naïve discretization. Finally, the API also provides functions for querying the probability distribution, expectation and the variance of a random variable. The next subsections describe piece-wise approximations and the computation of expectation and variance in more detail.

4.2 Piece-wise Approximations

Even though mixed-gamma distributions capture many natural distributions, there are other commonly occurring ones, such as the Gaussian. It is an open problem as to whether Gaussians admit a sound bit blasting function, let alone one that compiles to a compact OBDD. For such distributions, one can instead use a piece-wise approximation.

Let $C$ be an arbitrary continuous probability distribution over the interval $[l, u)$. To bit blast $C$ using a piece-wise distribution with $t$ pieces, $C$ is approximated using a mixture of $t$ discrete probability distributions over disjoint intervals. For every piece, one creates a shifted and scaled instance of a bit blasted mixed-gamma density and then creates a mixture of them. Note that since each piece uses $O(b)$ coin flips, a piece-wise distribution with $t$ pieces uses $O(\frac{tb}{b})$ coin flips. Section 5 shows empirical advantages of this approach. This piece-wise approximation using linear or exponential pieces can be easily built using the bitblast API available in HyBit (Figure 11). Figure 12 shows bit blasting of a Gaussian distribution using 2, 4, 8 and 16 pieces, where each piece is a bit blasted exponential. This provides the user with the conventional trade off between accuracy and performance. We elaborate more on this in Section 5.2.

4.3 Advantages of the Binary Representation

The binary representation has important advantages for probabilistic reasoning beyond the succinctness that bit blasting provides. First, many hybrid probabilistic models involve arithmetic operation on continuous random variables. Since we use a binary representation of fixed point numbers, arithmetic operations such as $+$, $\times$, $\div$, $<$ are compiled as Boolean formulas over binary numbers (similar to ALU circuits in architecture). This representation allows probabilistic inference (specifically the knowledge compilation approach that we employ) to identify and exploit the structure that exists in arithmetic, such as conditional independences among the resulting bits in a computation. Recent work \cite{Caoetal2023} described this compilation and showed its advantage empirically for integers; HyBit leverages these advantages for computations over fixed-point numbers.

The binary representation also enables an optimized computation of expectation and variance. Naïve computation of expectation and variance for a distribution over $2^b$ values requires one to compute probability of $2^b$ values. Bitwise representation allows one to achieve this computation by only
computing probability of $b$ bits which gives an exponential improvement. Note that in the worst case for an arbitrary hybrid probabilistic program, getting the corresponding OBDD for the binary representation can itself be exponential in the number of bits. But for the class of mixed-gamma distributions, this conversion is linear in the number of bits (Theorem 10). We formalize the computation of expectation and variance in the following two theorems and provide proofs in the appendix.

**Theorem 12.** Let $D$ be a discrete probability distribution over the interval $[0, 2^n]$ represented as a distribution over $n$ bits as $(b_n, b_{n-1}, \ldots, b_1)$, then the expectation of $D$ can be computed using linearity of expectation as follows:

$$E[D] = \sum_{i=0}^{2^n-1} i \cdot pr(i) = E(\sum_{j=1}^{n} 2^{j-1}b_j) = \sum_{j=1}^{n} 2^{j-1} \cdot pr(b_j)$$

**Theorem 13.** Let $D$ be a discrete probability distribution over the interval $[0, 2^n]$ represented as a distribution over $n$ bits as $(b_n, b_{n-1}, \ldots, b_1)$, then the variance of $D$ can be computed as follows:

$$Var[D] = \sum_{i=0}^{2^n-1} i^2 pr(i) - (E[D])^2 = Var(\sum_{j=1}^{n} 2^{j-1}b_j) = \sum_{k=1}^{n} \sum_{l=1}^{n} 2^{l+k-2}[pr(b_l \land b_k) - pr(b_l)pr(b_k)]$$

**Example.** Consider a discrete uniform distribution $U_4$ over the integers $\{0, 1, 2, 3\}$ represented using two bits, $(X_2, X_1)$. The direct way of calculating the expectation and variance of this uniform distribution requires inferring the probability of all the integers in the domain. But Theorems 12 and 13 allow us to compute these quantities by using only the probabilities of the individual bits.

$$E[U_4] = \sum_{j=1}^{2} 2^{j-1} \cdot pr(b_j) = 1.5 \quad Var[U_4] = \sum_{k=1}^{2} \sum_{l=1}^{2} 2^{l+k-2}[pr(b_l \land b_k) - pr(b_l)pr(b_k)] = 1.25$$

Figure 13 empirically shows the performance benefits in computing expectation and variance of a distribution as we increase the number of bits in bit blasting a standard normal distribution.

## 5 EMPIRICAL EVALUATION

We evaluate the practicality of bit blasting on real-life probabilistic programs. We have carried out relevant experiments to investigate the following questions:

- **Q1:** How does HyBit perform in comparison to existing inference algorithms? Section 5.1
- **Q2:** How effective is the piece-wise approximation? Section 5.2

### 5.1 Comparison with existing inference algorithms

#### 5.1.1 Approximate Inference Algorithms

We evaluate HyBit against two classes of approximate inference algorithms.

**Sampling Methods** We compare against WebPPL rejection sampling, MCMC sampling (with a Metropolis Hastings kernel), SMC sampling and Stan HMC as representatives of this class.

**Discretization Methods** We compare against AQUA and GuBPI in this class of inference algorithms [Beutner et al. 2022; Huang et al. 2021].

Comparing performance of different probabilistic programming systems is a challenging task since performance is directly affected by the structure of the program. We write equivalent programs for these benchmarks in each system and put in our best effort to optimize them. The tables in this section and subsequent sections report the mean value of absolute error over 10 runs for stochastic algorithms. For other inference algorithms, we report output of a single run. All experiments were single-threaded and were carried out on a server with 2.4 GHz CPU and 512 GB RAM.
Table 2 reports the results of performance evaluation of HyBi t against other approximate inference algorithms. We take all the hybrid and continuous benchmarks on which Psi [Gehr et al. 2016] was evaluated and a few more relevant benchmarks from existing work [Huang et al. 2021]. We put in our best effort to compute ground truth for these benchmarks either analytically or using computer algebra systems. We include only those benchmarks in our evaluation for which we were able to compute the ground truth reliably. We report the absolute error with respect to the ground truth for all the benchmarks. For benchmarks that returned a non-boolean value, we compute the absolute error of expectation for each of the approaches. We report the minimum error achieved by an inference algorithm within a timeout of 20 minutes.

For all the benchmarks, HyBi t replaced mixed-gamma distributions with their sound bit blasted distribution and all other distributions with their linear piece-wise approximation \( \pi_{1,0} \). The employed bits and pieces for each benchmark are reported in Table 2. To run Stan on these benchmarks, we make use of SlicStan [Gorinova et al. 2020] to get the Stan program with marginalized discrete random variables. For all WebPPL baselines, default settings were used for all the sampling algorithms with maximum number of samples within 20 minutes.

Table 3. Comparison of HyBi t with Psi, an exact inference PPL. ‘\( \phi \)’ denotes a timeout, ‘\( \times \)’ denotes result unsolved by Mathematica.

<table>
<thead>
<tr>
<th>Benchmarks</th>
<th>HyBi t Bit</th>
<th>AQUA Pieces</th>
<th>rejection</th>
<th>WebPPL MCMC</th>
<th>GuBPI</th>
<th>Stan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pi [@10kdiver [n. d.]]</td>
<td>1.05E-04</td>
<td>14</td>
<td>–</td>
<td>( \times )</td>
<td>8.30E-05</td>
<td>9.66E-05</td>
</tr>
<tr>
<td>weekend [Gehr et al. 2016]</td>
<td>2.08E-08</td>
<td>24</td>
<td>4096</td>
<td>( \times )</td>
<td>1.57E-02</td>
<td>1.57E-02</td>
</tr>
<tr>
<td>spacesc [canyon289 2022]</td>
<td>6.94E-04</td>
<td>19</td>
<td>52</td>
<td>( \times )</td>
<td>9.06E-04</td>
<td>3.24E-03</td>
</tr>
<tr>
<td>GPA [Wu et al. 2018]</td>
<td>2.22E-16</td>
<td>25</td>
<td>4096</td>
<td>3.62E-01</td>
<td>1.70E-02</td>
<td>9.39E-03</td>
</tr>
<tr>
<td>Tug of war [Huang et al. 2021]</td>
<td>4.50E-07</td>
<td>22</td>
<td>16</td>
<td>( \times )</td>
<td>6.93E-04</td>
<td>6.94E-04</td>
</tr>
<tr>
<td>alternmu2 [Nishihara et al. 2013]</td>
<td>3.48E-06</td>
<td>17</td>
<td>256</td>
<td>3.41E-07</td>
<td>( \phi )</td>
<td>4.61E-01</td>
</tr>
<tr>
<td>conjugate gaussians [Jordan 2010]</td>
<td>4.92E-06</td>
<td>23</td>
<td>16</td>
<td>0.99</td>
<td>2.19E-04</td>
<td>3.53E-04</td>
</tr>
<tr>
<td>normal_mixture (θ) [Huang et al. 2021]</td>
<td>5.49E-09</td>
<td>5</td>
<td>64</td>
<td>4.13E-07</td>
<td>( \phi )</td>
<td>3.90E-04</td>
</tr>
<tr>
<td>normal_mixture (μ) [Huang et al. 2021]</td>
<td>5.20E-03</td>
<td>9</td>
<td>16</td>
<td>7.55E-06</td>
<td>( \phi )</td>
<td>1.36E-03</td>
</tr>
<tr>
<td>normal_mixture (μ) [Huang et al. 2021]</td>
<td>3.92E-03</td>
<td>9</td>
<td>32</td>
<td>8.65E-06</td>
<td>( \phi )</td>
<td>7.11E-04</td>
</tr>
<tr>
<td>zeroone (w1) [Bissiri et al. 2016]</td>
<td>9.40E-05</td>
<td>16</td>
<td>–</td>
<td>5.66E-02</td>
<td>( \phi )</td>
<td>( \phi )</td>
</tr>
<tr>
<td>zeroone (w2) [Bissiri et al. 2016]</td>
<td>4.51E-04</td>
<td>19</td>
<td>–</td>
<td>3.69E+00</td>
<td>1.64E+00</td>
<td>1.64E+00</td>
</tr>
<tr>
<td>coinBias [Gehr et al. 2016]</td>
<td>2.02E-07</td>
<td>22</td>
<td>4096</td>
<td>6.25E-02</td>
<td>1.69E-05</td>
<td>7.73E-05</td>
</tr>
<tr>
<td>Addfun/sum [Gehr et al. 2016]</td>
<td>3.81E-06</td>
<td>23</td>
<td>16</td>
<td>( \times )</td>
<td>4.41E-04</td>
<td>1.69E-03</td>
</tr>
<tr>
<td>ClickGraph [Gehr et al. 2016]</td>
<td>1.75E-03</td>
<td>10</td>
<td>–</td>
<td>( \times )</td>
<td>7.29E-04</td>
<td>1.22E-03</td>
</tr>
<tr>
<td>trueSkill [Gehr et al. 2016]</td>
<td>3.05E-03</td>
<td>10</td>
<td>16</td>
<td>( \times )</td>
<td>1.81E-04</td>
<td>4.22E-04</td>
</tr>
<tr>
<td>clinicaltrial1 [Gehr et al. 2016]</td>
<td>5.27E-16</td>
<td>8</td>
<td>–</td>
<td>( \times )</td>
<td>1.51E-01</td>
<td>1.53E-01</td>
</tr>
<tr>
<td>clinicaltrial2 [Gehr et al. 2016]</td>
<td>6.81E-07</td>
<td>12</td>
<td>–</td>
<td>( \times )</td>
<td>1.42E-01</td>
<td>1.43E-01</td>
</tr>
</tbody>
</table>
As Table 2 shows, HyBit with bit blasting is comparable with the existing approaches on all the benchmarks, even better on 11/19 of them. For the other 8 benchmarks, HyBit achieves a very close accuracy. AQUA performs better on only four of the benchmarks and GuBPI fails to obtain good accuracy. This is primarily because their enumerative discretization does not scale well for higher precision. WebPPL and Stan (equipped with automated marginalization through SlicStan) support most of the benchmarks but do not achieve good accuracy within the threshold time. This is because sampling based algorithms are stochastic and cannot obtain sufficiently many samples from the true posterior in limited time.

5.1.2 Exact Inference Algorithms. Table 3 compares HyBit against a probabilistic programming system that performs exact inference using algebraic methods i.e. Psi [Gehr et al. 2016]. We put in our best effort to translate the benchmarks for optimal performance in Psi. It would often output a symbolic expression which we would feed to Mathematica for further simplification. Computing and simplifying these algebraic expressions is not a trivial task and hence, Psi timed out on 6 of these benchmarks and Mathematica failed to simplify 4 of these benchmarks. HyBit works for all 19 benchmarks as it reduces the computation to discrete inference on Boolean random variables and approximates the inference query.

5.2 How effective is piece-wise approximation?

We analyze the tradeoff between performance and accuracy when using different numbers of pieces to approximate the continuous distribution. Figure 14 demonstrates the trends of runtime and accuracy with the increase in pieces for different bitwidths for four benchmarks. As the number of linear pieces increases, runtime tends to first decrease and then increase. As the number of pieces increases, the accuracy tends to improve as shown by the lower four plots. This is because as we increase the number of pieces, continuous distributions are replaced with more accurate bit blasted distributions. That accuracy improvement comes at the cost of increased runtime after a certain sweet spot. The appendix provides additional experiments that also justify the usage of piece-wise approximations over an approach based on the central limit theorem.

6 RELATED WORK

Probabilistic programming has been an an active area of research both from the perspective of semantics and inference [Dahlqvist et al. 2023; Milch et al. 2005]. This section positions HyBit with respect related work. At a high level, the key distinction in HyBit is the development of bit blasting for succinct discretization of hybrid probabilistic programs.
Discretization approaches. Prior approaches that discretize continuous or hybrid probabilistic programs estimate the posterior by exhaustively enumerating all of the discretized values [Huang et al. 2021], which does not scale to provide sufficient accuracy in many cases. One prior discretization technique also employs a bit representation [Claret et al. 2013]. However, their approach is not a form of bit blasting, since it is not succinct and still in general produces a representation that is proportional to the number of discretized points. Finally, a recent approach uses discretization to produce upper and lower bounds on the posterior of a probabilistic program [Beutner et al. 2022].

Inference algorithms for hybrid probabilistic programs. Other research specifically targets hybrid probabilistic programs. Leios [Laurel and Misailovic 2020] continualizes the hybrid probabilistic program in order to harness the power of existing continuous inference algorithms. HyBi t, on the other hand discretizes the hybrid programs which helps in scaling inference for hybrid programs specifically with respect to the discrete structure. SPPL supports hybrid programs by translating them to specific representations for inference [Saad et al. 2021]. However, these representations constrain the hybrid programs that can be supported. For instance, SPPL does not support arithmetic on continuous random variables while HyBi t can. Finally, probabilistic logic programming languages have been extended to support hybrid models using interval traces [Gutmann et al. 2011].

Algebraic approaches. Some PPL inference algorithms produce closed form algebraic expressions to encode probability distributions and then use symbolic techniques to perform exact inference [Gehr et al. 2016; Hur et al. 2014; Narayanan et al. 2016]. However, these systems are necessarily limited in their expressivity and the programs that they can handle, as shown in Table 3.

Path based inference algorithms. A common class of inference algorithms for PPLs are operational: they record traces on the program by using concrete values of the random variables. This includes sampling algorithms and variational approximations [Bingham et al. 2019; Carpenter et al. 2017; Chaganty et al. 2013; Dillon et al. 2017; Goodman et al. 2008; Hur et al. 2015; Kucukelbir et al. 2015; Mansinghka et al. 2013, 2018; Minka et al. 2014; Pfeffer 2007; Saad and Mansinghka 2016; Tristan et al. 2014; van de Meent et al. 2015; Wingate and Weber 2013; Wood et al. 2014]. Sampling algorithms like rejection sampling and MCMC methods are universal but have known limitations such as difficulty in handling multi-modality and low-probability evidence, as described in Section 2. More sophisticated techniques like Hamiltonian Monte Carlo and variational approximation address these limitations but impose constraints of continuity and almost-everywhere differentiability, so they must resort to marginalizing out all discrete structure.

Use of a binary representation. Bit blasting has been a widespread technique in software verification, used in constraint solvers to reason about arithmetic using a bit representation [Bruttomesso and Sharygina 2009]. Recent work in scaling inference for probabilistic programs over integers also employs a binary representation for numbers [Cao et al. 2023], in order to exploit conditional independences in that representation. The bit blasting in HyBi t is inspired by these techniques but has a different purpose and hence a very different technical approach: to develop succinct, and in many cases provably sound, approximations of continuous probability distributions.

7 CONCLUSION AND FUTURE WORK
In this work, we motivated the need for new inference methods for hybrid probabilistic programs. We described bit blasting, whereby hybrid probabilistic programs are succinctly discretized and then analyzed using algorithms for discrete inference. We characterized a class of continuous densities — mixed-gamma densities — for which bit blasting is not only succinct but also sound relative to an explicit discretization approach as well as provably efficient to analyze. We then presented a new PPL HyBi t that employs a novel inference algorithm for hybrid programs based on bit blasting. We demonstrated the performance benefits of HyBi t over existing approximate inference algorithms.
In future work, we hope to expand the class of distributions that can be bit blasted soundly. We plan to investigate how HyBi t can be extended to support hierarchical Bayesian models. We plan to enhance its usability by not requiring user to specify the hyperparameters for every probabilistic program. We are also interested to explore the integration of HyBi t with other inference approaches, to leverage their relative strengths for support of a wider range of hybrid probabilistic programs.

ACKNOWLEDGMENTS

The authors would like to thank previous and current members of Star AI lab for helpful discussions and emotional support. This work was funded in part by the DARPA PTG Program under award HR00112220005, the DARPA ANSR program under award FA8750-23-2-0004, NSF grants #IIS-1943641, #IIS-1956441, #CCF-1837129, and a gift from RelationalAI. GVdB discloses a financial interest in RelationalAI.

ARTIFACT

HyBi t probabilistic programming system is available as an open source repository on GitHub at https://github.com/Tractables/Dice.jl/tree/hybit with thorough documentation for reusability. It is also available as an archived version on Zenodo [Garg et al. 2024] with comprehensive instructions and scripts for reproducibility of experiments reported in the paper.

REFERENCES

@10kdiver. [n. d.]. @10kdiver. https://twitter.com/10kdiver/status/1503307755976765443?s=20&t=Ux1C55thW5qnCPVOHosguxQ. [Online; accessed 14-March-2022].


