1. Introduction

- Algorithms in AI and in the decision making literature often apply operations over random variables such as addition, multiplication, maximum, etc.

- In many cases the analysis of the complexity of these algorithms considers these operations as atomic, i.e., requiring $o(1)$ time and memory.

- When we look closer, we see that this is not always the case: The table that represents these variables (their support) may grow exponentially with convolution. This affects both memory and time.

- We, therefore, need methods for approximating random variables with large support: Given a variable X we need a way to compute a variable X' with $suppor(X') < suppor(X)$ such that if we replace X with X' in the remainder of the computation, we get a result that is close to the result that we would have gotten if we could use X.

- Another related issue is with handling continuous distributions: If we want to use them in computations, we need to convert them to a table and this requires an approximation.

- The above two points are parallel (in reverse order) to the common practice of approximating rational number with fixed point numbers (e.g., $\pi \approx 3.14$), and of truncating digits (e.g., \$$24.456 \approx$ \$$24.46$) in order to save space and computations.

- While in the context of real numbers the notion of approximation, i.e., the distance between two numbers, is clear, it is not so when dealing with random variables. There are various different distances known in the literature such as the Kolmogorov distance, the Waserstein distance, the Kanterovitch distance, etc. The choice of a specific metric to use depends on the application at hand.

- In this work we focus on the Kolmogorov distance, motivated by the problem of estimating the probability of missing a deadline in a task tree.

- More specifically, we say that X' is an ε-approximation of X if $|X - X'|_K \leq \varepsilon$ and $\forall t. F_X(t) \leq F_{X'}(t)$. The second part is because we want a safe estimation of the probability of missing a deadline, i.e., we only allow "false positive" errors where our estimation is above the real value, never below it.

- The main contribution of this work is an algorithm that, given a random variable X and an integer m, finds a random variable X' such that $\mathrm{support}(X') = m$ and there is no other variable with support of size m that approximates X better than X'.

2. Related Work

The need for approximating probability distributions by distributions that can be represented more compactly is mentioned in the literature in various contests. Typically, a continuous distribution is approximated by a discrete one that has a small support. See, for example the work of Keefer and Bodily [4, 3] on three point approximations. This, in a sense, is similar to truncating digits to obtain a fixed point representation of an irrational number. Other approximation approaches proposed in the literature include the bracket approach, discussed, e.g., in [2] and in [1], the support of the distribution is divided into several brackets (not necessary equal in probability) and the mean or the median of every bucket is chosen to be a discrete representation of that part of the target distribution in the approximation. Another approach is based on the idea that the approximation should match the moments of the original distribution. Matching the moments has been recognized to be especially important in computing value lotteries and their certain equivalents [6]. The idea is as follows: the value function frequently can be well approximated by a polynomial (with degree m) of a random variable. Thus, if that random variable is approximated by a simpler discrete variable having the same m first moments, the expected value function based on the approximation is no different from that one based on the original random variable. The key result here states that it is possible to match the first $2m-1$ moments of the target distribution by a discrete one with a support of size m, see [2] and [6]. When the original distribution is not specified completely so fewer than $2m-1$ moments of the original distribution are known, the resulting ambiguity of defining the approximation of size m was suggested to be resolved using the entropy maximization [5].

3. One sided Kolmogorov distance

Definition 1. *For a set $S \subseteq \mathbb{R}$ we say that $B \subseteq S$ is* consecutive *if any $s \in S$ that is smaller than $\max(B)$ and larger than $\min(B)$ is in B.*

Definition 2. *A partition $P = \{B_1, \ldots, B_n\}$ of a set $S \subseteq \mathbb{R}$ is called* consecutive *if all the subsets $B_1, \ldots, B_n$ are consecutive.*

Definition 3. *For a discrete real random variable X and a partition P of its support, we define a new discrete random variable X_P by:*

$$Pr(X_P = t) = \begin{cases} Pr(X \in B) & \text{if } t = min(B) \wedge B \in P, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 4. *For discrete real-valued variables X_1 and X_2, we say that X_2 is a one-sided Kolmogorov approximation of X_1 with the parameters ε and m, denoted by $X_1 \preceq_{\varepsilon,m} X_2$, if $\forall t\colon 0 \leq F_{X_2}(t) - F_{X_1}(t) \leq \varepsilon$ and $|\operatorname{support}(X_2)| \leq m$.*

Definition 5. *For a discrete real-valued random variable X and $m \in \mathbb{N}$, let $\varepsilon^* = \min\{\varepsilon\colon$ there is X' such that $X \preceq_{\varepsilon,m} X'\}$ be the best possible approximation error for X with a random variable whose support of size m.*

Theorem 1. *For any discrete real-valued random variable X and any $m \in \mathbb{N}$, there is consecutive partition P of $\operatorname{support}(X)$ such that $X \preceq_{\varepsilon^*,m} X_P$.*

Proof. Let X' be such that $X \preceq_{\varepsilon^*,m} X'$. Specifically, for all t,

$$F_X(t) \leq F_{X'}(t) \leq F_X(t) + \varepsilon^* \tag{1}$$

The proof goes in two steps: (1) we first construct a variable X'' form X' that approximate X as X' does, i.e., $X \preceq_{\varepsilon^*,m} X''$, but also has the property that its support is a subset of the support of X; (2) then, from X'', we construct another random variable, X''', that in addition to being an approximation of X with the same parameters is also equal to X_P for some consecutive partition P.

Assume that $t_0, t_1, \ldots, t_n$ are all the elements in the support of X in ascending order. Define the random variable X'' by

$$f_{X''}(t) = \begin{cases} Pr(X' \leq t_0) & \text{if } t = t_0 \\ Pr(t_{i-1} < X' \leq t_i) & \text{if } t = t_i \text{ for some } i \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

We will show now that: (1) $\operatorname{support}(X'') \subseteq \operatorname{support}(X)$; (2) $X \preceq_{\varepsilon^*,m} X''$. Since we only assign a non-zero probability to $f_{X''}(t)$ if $t = t_0$ or if $t = t_i$ for some i, i.e., only if t is in the support of X, we have that $\operatorname{support}(X'') \subseteq \operatorname{support}(X)$. Furthermore, if $t_i \in \operatorname{support}(X'')$ then $Pr(t_{i-1} < X' \leq t_i) \neq 0$ which means that there is some $t_{i-1} < t' \leq t_i$ such that $t' \in \operatorname{support}(X')$. To also handle the case where $i = 0$, we denote $t_{-1} = -\infty$. This (unique) mapping gives us that $|\operatorname{support}(X'')| \leq$

$|\operatorname{support}(X')| \leq m$. To complete the proof of the properties of X'', we will show now that $F_X(t) \leq F_{X''}(t) \leq F_{X'}(t)$ for all t by examining the different ts as follows:

Case $t < t_0$: $F_{X''}(t) = F_X(t) = 0$. Since $F_{X'}(t) \geq 0$ for all t, we get that $F_X(t) \leq F_{X''}(t) \leq F_{X'}(t)$.

Case $t = t_i$: $F_{X'}(t) = F_{X''}(t)$ and $F_X(t) \leq F_{X'}(t)$ by Eq. (1).

Case $t_{i-1} < t < t_i$: $F_{X''}(t) = F_{X''}(t_{i-1})$ and $F_X(t) = F_X(t_{i-1})$. Since we already have that $F_X(t_{i-1}) \leq F_{X''}(t_{i-1}) \leq F_{X'}(t_{i-1})$, we get that $F_X(t) \leq F_{X''}(t) \leq F_{X'}(t_{i-1})$. By monotonicity of CDF, $F_{X'}(t_{i-1}) \leq F_{X'}(t)$ therefore $F_X(t) \leq F_{X''}(t) \leq F_{X'}(t)$.

Case $t > t_n$: $F_X(t) = F_{X''}(t) = 1$ and, by Eq. (1), since CDFs are always bounded by one, also $F_{X'}(t) = 1$.

From the four different cases of t, as we already established that $|\operatorname{support}(X'')| \leq m$, we get that $X \preceq_{\varepsilon^*,m} X''$.

Let $s_0, s_1, \ldots, s_k$ be the elements in the support of X'' in ascending order $k \leq m$. Define the random variable X'''

$$f_{X'''}(t) = \begin{cases} Pr(s_i \leq X < s_{i+1}) & \text{if } t = s_i \text{ for some } i < k \\ Pr(X \geq s_k) & \text{if } t = s_k \\ 0 & \text{otherwise} \end{cases}$$

We will show that: (1)$X \preceq_{\varepsilon^*,m} X'''$; (2) There is a partition P such that $X''' = X_P$. Again, we will show that $F_X(t) \leq F_{X'''}(t) \leq F_{X''}(t)$ for all t by examining the different values of t as follows:

Case $t < s_0$: $F_{X'''}(t) = F_{X''}(t) = F_X(t) = 0$.

Case $t = s_i$: First, $F_X(t) \leq F_{X'''}(t)$ since $F_{X'''}(t) = F_X(t) + Pr(s_i < X < s_{i+1})$. Second we show that $F_{X'''}(s_i) \leq F_{X''}(s_i)$. Since $X \preceq_{\varepsilon^*,m} X''$, $F_X(s_i) + Pr(s_i < X < s_{i+1}) \leq Pr(X'' < s_{i+1})$. As $s_1, \ldots, s_m$ is the support of X'', $Pr(X'' < s_{i+1}) = F_{X''}(s_i)$. By definition $F_{X'''}(s_i) = F_X(s_i) + Pr(s_i < X < s_{i+1})$. Together we get that $F_{X'''}(s_i) \leq F_{X''}(s_i)$. For the case $t = s_m$, the argument holds with the notation $m + 1 = \infty$.

Case $s_{i-1} < t < s_i$: $F_{X''}(t) = F_{X''}(s_{i-1})$ and $F_{X'''}(t) = F_{X'''}(s_{i-1})$ therefore $F_{X'''}(t) \leq F_{X''}(t)$. Also, $F_X(t) \leq Pr(X < s_i) = F_{X'''}(t)$.

Case $t > s_m$**:** $F_{X''}(t) = F_{X'''}(t) = 1$. Since CDFs are always smaller or equal to one, also $F_X(t) \leq 1$.

From the four different cases of t and that $\mathrm{support}(X'') = \mathrm{support}(X''')$ we established that $X \preceq_{\varepsilon^*,m} X'''$. The next step is to prove that $X''' = X_P$, by presenting a partition P. As shown before, $\mathrm{support}(X) = \{t_0, t_1, \ldots, t_n\}$, $\mathrm{support}(X''') = \{s_0, s_1, \ldots, s_k\}$, so $\mathrm{support}(X''') \subseteq \mathrm{support}(X)$. In addition, $\forall 0 \leq i \leq m$, $Pr(X''' = s_i) = Pr(s_i \leq X \leq s_{i+1})$ therefore $P = \{s_0, s_1, \ldots, s_k\}$. By definition 3, $X''' = X_P$, moreover, by definition 2 P is a consecutive partition. □

Algorithm 1: $OptTrim(X, m)$

1 $S = \mathrm{support}(X) \cup \{\infty\}$
2 $G = (V, E) = (S, \{(i,j) \in S^2 \colon j > i\})$
3 **foreach** $e = (i,j) \in E$ **do**
4 $\quad w(e) = Pr(i < X < j)$
5 /* The following can be obtained, e.g., using the Bellman-Ford algorithm */
6 $l = \mathrm{argmin}_{l \in paths(G), |l|=m} \max\{w(e) \colon e \in l\}$
7 **foreach** $e = (i,j) \in l$ **do**
8 $\quad f_{X'}(i) = Pr(i \leq X < j)$
9 **return** X'

Theorem 2. $X \preceq_{\varepsilon^*,m} OptTrim(X, m)$.

Proof. As proved in Theorem 1 there is a consecutive partition P for which $X \preceq_{\varepsilon^*,m} X_P$. For every consecutive partition P there is a path l, $l \in paths(G)$, $|l| = m$, such that the X' generated in lines 7-8 in the algorithm satisfies $X' = X_P$ and $X \preceq_{\varepsilon,m} X'$ where $\varepsilon = \max\{w(e) \colon e \in l\}$. By using for instance the Bellman-Ford algorithm as in line 6, allow us to get the path l^* containing the minimal edge among all maximal edges of all the other paths in G. The consecutive partition P associated with this "lightest" path l^*, resulted with X_P eventually $X_P = OptTrim(X, m)$ and $X \preceq_{\varepsilon^*,m} X_P$. □

The following example shows that even if $\mathrm{support}(X'') \subseteq \mathrm{support}(X)$ that is not enough to establish that $X'' = X_P$. For example, given the random variables X and X''. X'' is an optimal approximation of X such that $X \preceq_{\varepsilon^*,m} X''$ but $X'' \neq X_P$.

Example 1.

$$f_X(t) = \begin{cases} 1/3 & \textit{if } t = 1 \\ 1/3 & \textit{if } t = 2 \\ 1/6 & \textit{if } t = 3 \\ 1/6 & \textit{if } t = 4 \\ 0 & \textit{otherwise} \end{cases}$$

$$f_{X''}(t) = \begin{cases} 2/3 & \textit{if } t = 1 \\ 1/3 & \textit{if } t = 2 \\ 0 & \textit{otherwise} \end{cases}$$

Lemma 1. $\varepsilon^* \leq \frac{1}{m}$

Proof. Assume that $\varepsilon = 1/m$, then from [??] $X' = Trim(X, 1/m)$, Lemma 1 and Lemma 2 in that paper establish that $X \preceq_{1/m,m} X'$. Since ε^* is the minimal distance between X and X' then $\varepsilon^* \leq 1/m$ □

Lemma 2. *If* $X' = OptTrim(X, 1/\varepsilon)$ *then* $X \preceq_\varepsilon X'$

Proof. $X \preceq_{\varepsilon^*,1/\varepsilon} X' \Rightarrow X \preceq_{\varepsilon^*} X' \Rightarrow X \preceq_\varepsilon X'$. □

Not proved yet issue: 1) Why is the first value gives us the minimal partition?

[1] RK Hammond and JE Bickel. Reexamining discrete approximations to continuous distributions. *Decis. Anal.*, 2013.

[2] AC Miller III and TR Rice. Discrete approximations of probability distributions. *Manage. Sci.*, 1983.

[3] DL Keefer. Certainty equivalents for three-point discrete-distribution approximations. *Manage. Sci.*, 1994.

[4] DL Keefer and SE Bodily. Three-point approximations for continuous random variables. *Manage. Sci.*, 1983.

[5] E Rosenblueth and HP Hong. Maximum entropy and discretization of probability distributions. *Probabilistic Eng. Mech.*, 1987.

[6] JE Smith. Moment methods for decision analysis. *Manage. Sci.*, 1993.