Finite-sample Bounds for Marginal MAP

Qi Lou Rina Dechter Alexander Ihler
University of California, Irvine

SUMMARY

Our task: bounding **marginal MAP** (**MMAP**) of a discrete graphical model (exact computation intractable in general -- NP^{PP} [Park 2002]).

MMAP: $x^\star_{\mathrm{M}} = \operatorname{argmax}_{x_{\mathrm{M}} \in X_{\mathrm{M}}} \pi(x_{\mathrm{M}}), \quad \pi(x_{\mathrm{M}}) = \sum_{x_{\mathrm{S}} \in X_{\mathrm{S}}} \prod_{f_\alpha \in \mathcal{F}} f_\alpha(x_\alpha)$

where

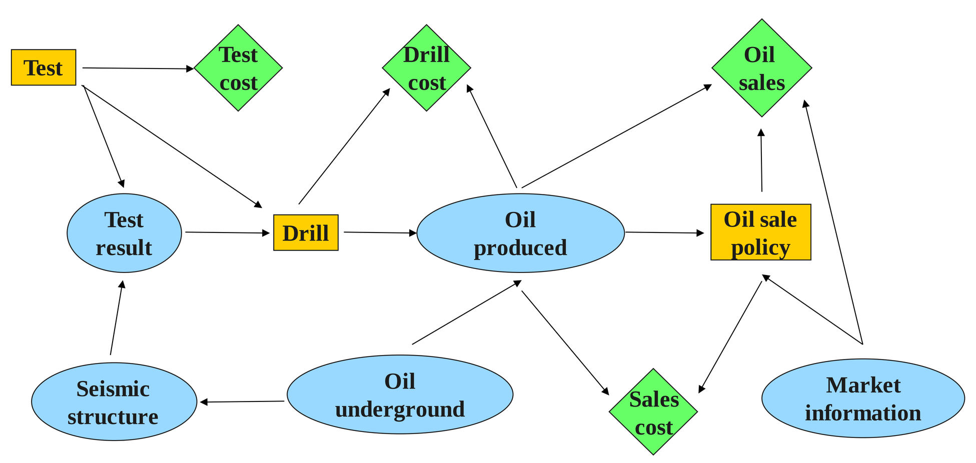
$X = \{X_1, X_2, \ldots, X_n\}$ -- discrete variables
$\mathcal{F} = \{f_1, f_2, \ldots, f_m\}$ -- non-negative functions
$X_{\mathrm{M}} \subset X$ -- maximization (MAX) variables
$X_{\mathrm{S}} = X \backslash X_{\mathrm{M}}$ -- summation (SUM) variables

It is a generalization of

MAP: $x^\star = \operatorname{argmax}_{x \in X} \prod_{f_\alpha \in \mathcal{F}} f_\alpha(x_\alpha)$

The partition function: $Z = \sum_{x \in X} \prod_{f_\alpha \in \mathcal{F}} f_\alpha(x_\alpha)$

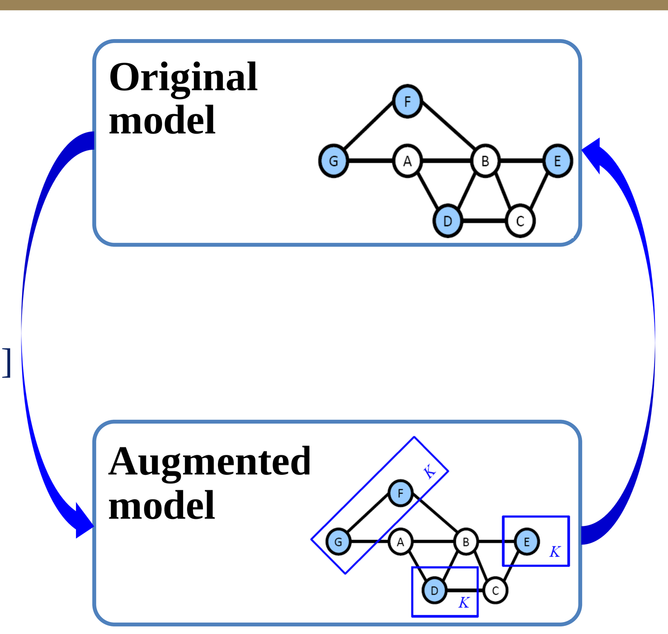
Example: influence diagrams & optimal decision-making
The “oil wildcatter” problem (e.g., [Raiffa 1968; Shachter 1986]).



Our contributions:

- We propose a **Mixed Dynamic Importance Sampling** (**MDIS**) algorithm that provides anytime finite-sample bounds (i.e., they hold with probability $1-\delta$ for some confidence parameter δ) for MMAP.
- It provides both upper and lower bounds that are guaranteed to be tight given enough time.
- It is able to predict high-quality MAP solutions whose values converge to the optimum; the exploration-exploitation trade-off of searching MAP solutions controlled by the number of replicates of the marginalized variables.
- It runs in an anytime/anyspace manner, which gives flexible trade-offs between memory, time, and solution quality.

MAIN IDEA



Construct an augmented model [Doucet et al. 2002]

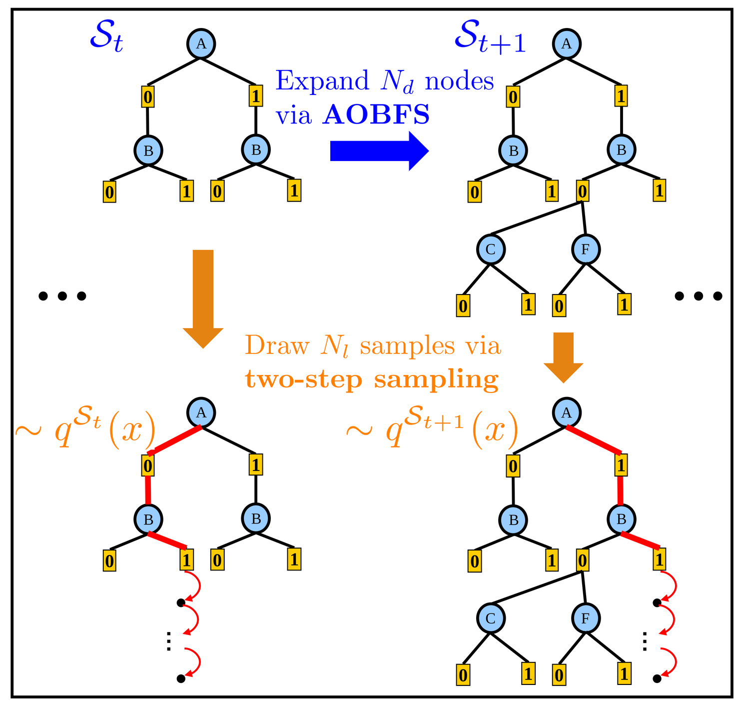
Translate the finite-sample bounds back to bound MMAP

Generalize dynamic importance sampling [Lou, Dechter, Ihler 2017] to provide finite-sample bounds for a series of summation objectives

BACKGROUND

Dynamic importance sampling (DIS):

- Provides finite-sample bounds and an unbiased estimate for the partition function.
- Interleaves search with sampling in a way that search generates a set of improving proposal distributions where samples are drawn to produce probabilistic bounds.



Sample aggregation issue for DIS:

- Samples are independent but not i.i.d.
- later samples come from improving proposals.

Weighted average of importance weights:

$$\widehat{Z} = \frac{\mathrm{HM}(\boldsymbol{U})}{N} \sum_{i=1}^{N} \frac{\widehat{Z}_i}{U_i} \qquad \mathrm{HM}(\boldsymbol{U}) = \Big[\frac{1}{N}\sum_{i=1}^{N} \frac{1}{U_i}\Big]^{-1}$$

$\widehat{Z} \le \mathrm{HM}(\boldsymbol{U})$ **(boundedness)**

$\mathbb{E}\,\widehat{Z} = Z$ **(unbiasedness)**

Finite-sample bounds of DIS:

$\Pr[Z \le \widehat{Z} + \Delta] \ge 1-\delta$ (and $\Pr[Z \ge \widehat{Z} - \Delta] \ge 1-\delta$)

$$\Delta = \mathrm{HM}(\boldsymbol{U})\Big(\sqrt{\frac{2\widehat{\mathrm{Var}}(\{\widehat{Z}_i/U_i\}_{i=1}^N)\ln(2/\delta)}{N}} + \frac{7\ln(2/\delta)}{3(N-1)}\Big)$$

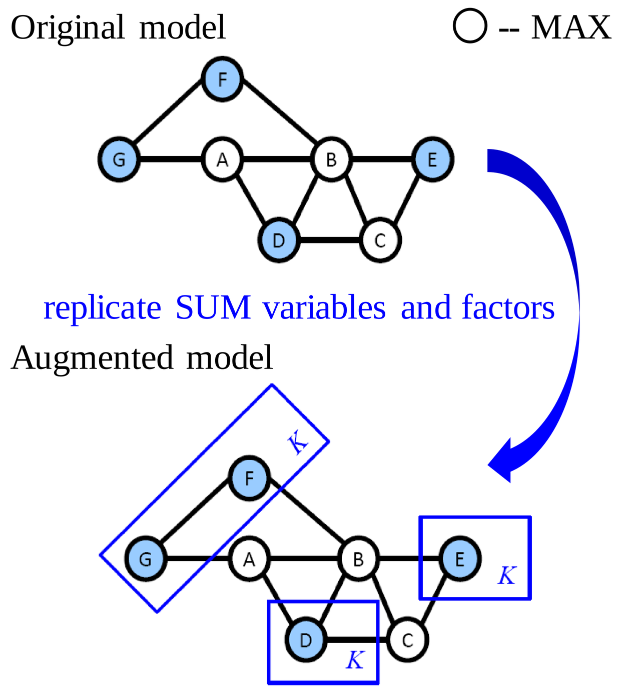
$\widehat{\mathrm{Var}}(\{\widehat{Z}_i/U_i\}_{i=1}^N)$: empirical variance of $\{\widehat{Z}_i/U_i\}_{i=1}^N$.

OUR ALGORITHM

Mixed dynamic importance sampling (MDIS):

- Connect MMAP to a pure summation task of an augmented model by replicating the summation variables and their associated factors.

Original model ○ -- MAX variable ● -- SUM variable



Let $X_{\mathrm{aug}} = (X_{\mathrm{M}}, \overbrace{X^1_{\mathrm{S}}, \ldots, X^K_{\mathrm{S}}}^{\text{copies of } X_{\mathrm{S}}})$
and $f_{\mathrm{aug}}(x_{\mathrm{aug}}) = \prod_{k=1}^K f(x_{\mathrm{M}}, x^k_{\mathrm{S}})$

The partition function of the augmented model:

$$Z_{\mathrm{aug}} = \sum_{x_{\mathrm{aug}}} \prod_{k=1}^{K} f_{\mathrm{aug}}(x_{\mathrm{aug}}) = \sum_{x_{\mathrm{M}}} \pi^K(x_{\mathrm{M}})$$

Bound the MMAP optimum using the partition function of the augmented model:

$$(Z_{\mathrm{aug}}/\underbrace{|X_{\mathrm{M}}|}_{\text{size of the MAP space}})^{1/K} \le \pi(x^\star_{\mathrm{M}}) \le Z^{1/K}_{\mathrm{aug}}$$

- DIS applicable to the augmented model to bound its partition function.
  - Complexity independent of the replicates K.
  - NOT compatible to pruning of the MAP space during search.

Key observation:

$x^\star_{\mathrm{M}} \in \mathcal{A} \implies Z^{\mathcal{A}}_{\mathrm{aug}}/|\mathcal{A}| \le \pi^K(x^\star_{\mathrm{M}}) \le Z^{\mathcal{A}}_{\mathrm{aug}}$, where $Z^{\mathcal{A}}_{\mathrm{aug}} = \sum_{x_{\mathrm{M}} \in \mathcal{A}} \pi^K(x_{\mathrm{M}})$.

Connect the MMAP optimum to a series of summation objectives:

$$\frac{\mathrm{HM}(\boldsymbol{U}/|\boldsymbol{\mathcal{A}}|)}{\mathrm{HM}(\boldsymbol{U})}\,\mathbb{E}\big[\widehat{Z}_{\mathrm{aug}}\big] \le \pi^K(x^\star_{\mathrm{M}}) \le \mathbb{E}\big[\widehat{Z}_{\mathrm{aug}}\big] \qquad \widehat{Z}_{\mathrm{aug}} = \frac{\mathrm{HM}(\boldsymbol{U})}{N}\sum_{i=1}^{N}\frac{\widehat{Z}^i_{\mathrm{aug}}}{U_i}$$

$$\mathrm{HM}(\boldsymbol{U}) = \Big[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{U_i}\Big]^{-1}, \quad \mathrm{HM}(\boldsymbol{U}/|\boldsymbol{\mathcal{A}}|) = \Big[\frac{1}{N}\sum_{i=1}^{N}\frac{|\mathcal{A}_i|}{U_i}\Big]^{-1} \qquad \mathbb{E}\big[\widehat{Z}^i_{\mathrm{aug}}\big] = Z^{\mathcal{A}_i}_{\mathrm{aug}} \le U_i$$

- Finite-sample bounds for MMAP:

$$\Pr\big[\pi(x^\star_{\mathrm{M}}) \le (\widehat{Z}_{\mathrm{aug}}+\Delta)^{\frac{1}{K}}\big] \ge 1-\delta, \quad \Pr\big[\pi(x^\star_{\mathrm{M}}) \ge \big(\frac{(\widehat{Z}_{\mathrm{aug}}-\Delta)\,\mathrm{HM}(\boldsymbol{U}/|\boldsymbol{\mathcal{A}}|)}{\mathrm{HM}(\boldsymbol{U})}\big)^{\frac{1}{K}}\big] \ge 1-\delta.$$

$$\Delta = \mathrm{HM}(\boldsymbol{U})\Big(\sqrt{\frac{2\widehat{\mathrm{Var}}\ln(2/\delta)}{N}} + \frac{7\ln(2/\delta)}{3(N-1)}\Big)$$

EMPIRICAL EVALUATION

Experimental settings:

- **Baselines:** two state-of-the-art search algorithms: **UBFS** [Lou et al. 2018], a unified best-first search algorithm that emphasizes rapidly tightening the upper bound. **AAOBF** [Marinescu et al. 2017], a best-first/depth-first hybrid search algorithm that balances upper bound quality with generating and evaluating potential solutions.
- **Benchmarks:** four benchmarks; three out of the four formed by instances selected from recent UAI competitions, where 10% variables randomly set to MAX variables. The fourth benchmark formed by instances from probabilistic conformant planning with a finite-time horizon [Lee et al. 2016a]. Statistics on the right.

	grid	promedas	protein	planning
# instances	50	50	44	15
avg. # variables	1248.20	982.10	109.55	1122.33
avg. % of MAX vars	10%	10%	10%	12%
avg. # of factors	1248.20	994.76	394.64	1127.67
avg. max domain size	2.00	2.00	81.00	3.00
avg. max scope	3.00	3.00	2.00	5.00
avg. induced width	124.82	108.14	15.84	165.00
avg. pseudo tree depth	228.92	158.78	33.52	799.33
avg. ind. width of sum	43.44	40.32	10.20	49.67

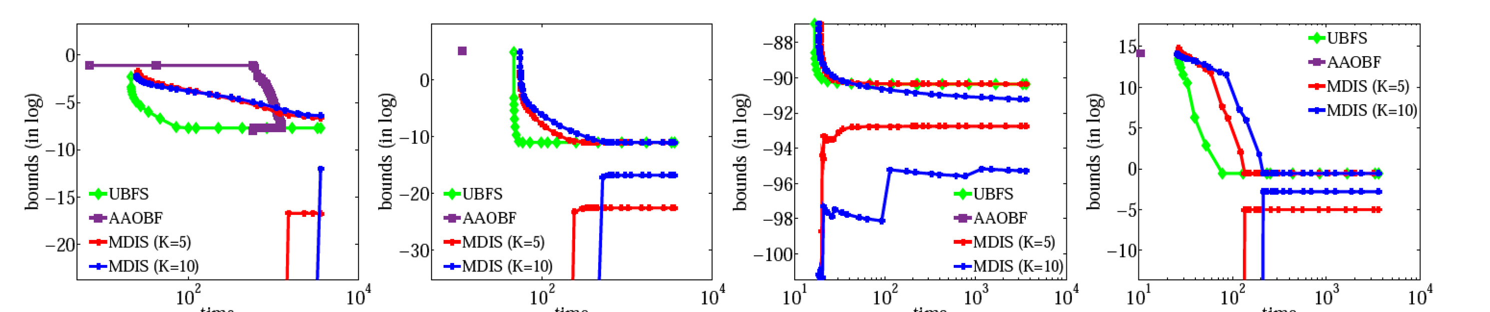
- **Other settings:** δ=0.025; memory: 4GB; runtime: 1hr; implementation: all in C/C++ by the original authors.

Table: Number of instances that an algorithm achieves the best *lower/upper* bounds at each timestamp (1 min, 10 min, and 1 hour) for each benchmark. Entries for UBFS are blank because UBFS does not provide lower bounds.

	grid	promedas	protein	planning
# instances	50	50	44	15
Timestamp: 1min/10min/1hr				
MDIS (K=5)	**47/44/45**	**32/34/31**	**31/27/28**	**14/13/13**
MDIS (K=10)	3/2/1	4/5/6	11/13/14	1/2/2
UBFS	-/-/-	-/-/-	-/-/-	-/-/-
AAOBF	0/4/4	16/21/24	2/4/4	0/0/0

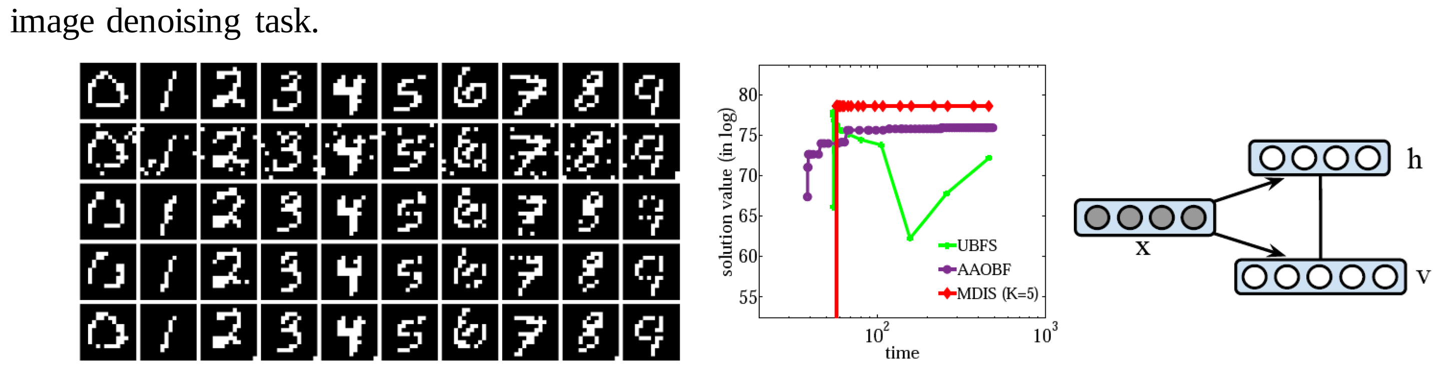
	grid	promedas	protein	planning
# instances	50	50	44	15
Timestamp: 1min/10min/1hr				
MDIS (K=5)	0/0/0	9/12/13	5/9/15	1/1/1
MDIS (K=10)	0/0/0	10/13/14	9/10/13	1/2/3
UBFS	**50/50/50**	**50/50/50**	**36/32/26**	**14/14/13**
AAOBF	0/0/1	2/4/6	2/2/2	1/1/1

Figure 1: Anytime bounds for MMAP on instances from four benchmarks. The max domain sizes of those instances from (a)-(d) are 2, 2, 81, 3 respectively, and the induced widths of the internal summation problems are 25, 28, 8, 24, respectively.



(a) grid/90-25-5 (b) promedas/or_chain_48.fg (c) protein/1skz (d) planning/bw2_2_4_4

Figure 2: (a) Image denoising results for one instance per digit. The first row is for the ground truth images. The second row is for the noisy inputs created from the ground truth by randomly flipping 5% pixels. Below the first two rows are denoised images from UBFS, AAOBF, MDIS (K=5) respectively. (b) An example on MAP solution quality comparison. (c) Illustration of the CRBM model used for the image denoising task.



(a) (b) (c) CRBM

RELATED WORK

- Deterministic approaches:
  - Exact solvers based on depth-first branch and bound, e.g., [Park & Darwiche 2003; Yuan & Hansen 2009].
  - Search equipped with variational heuristics, e.g., [Marinescu et al. 2014; Lee et al. 2016b; Marinescu et al. 2017; Lou et al. 2018].
  - Variational methods, e.g., [Liu & Ihler 2013; Ping et al. 2015].
  - Factor set elimination based [Mauá & de Campos 2012].
- Monte Carlo approaches:
  - Random hashing based, e.g., [Xue et al. 2016].
  - Markov chain Monte Carlo based, e.g., [Yuan et al. 2004; Doucet et al. 2002].