Learning Probabilistic Sentential Decision Diagrams Under Logic Constraints by Sampling and Averaging

Renato L. Geh
renatolg@ime.usp.br

Denis D. Mauá
ddm@ime.usp.br

Department of Computer Science

Institute of Mathematics and Statistics, University of São Paulo

UAI 2021

Probabilistic Sentential Decision Diagrams

Probabilistic Sentential Decision Diagrams (PSDDs):

- Structured Decomposable probabilistic circuits
- Encode certain knowledge as logic constraints
- Encode uncertain knowledge as probabilities
- Interpretable syntax
- Many inferences are exact and tractable :

  - ☑ Evidence
  - ☑ Marginals
  - ☑ MLE Parameter Learning
  - ☑ Most Probable Explanation
  - ☑ Expectations
  - ☑ KL-divergence

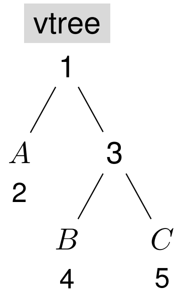
- PSDD circuit represents recursive decomposition of formula:

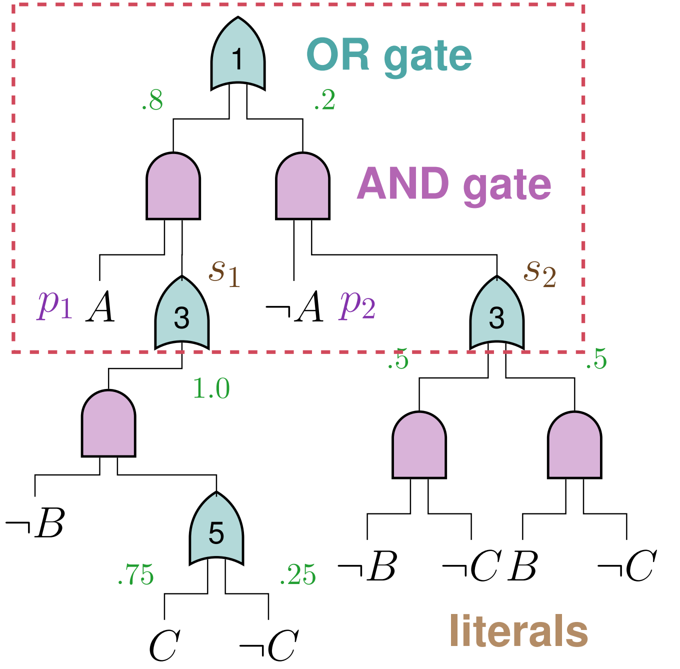
$$\bigvee_{i=1}^{k} (p_i \wedge s_i)$$, where each *prime* p_i and *sub* s_i are logical formulae

Darwiche [2011], Kisa et al. [2014]

A	B	C	Pr
0	0	0	**0.1**
0	1	0	**0.1**
1	0	0	**0.2**
1	0	1	**0.6**

s.t. $(A \to \neg B) \wedge (C \to A)$





Probabilistic Sentential Decision Diagrams

Existing PSDD learners:

LEARNPSDD (Liang et al. [2017]):

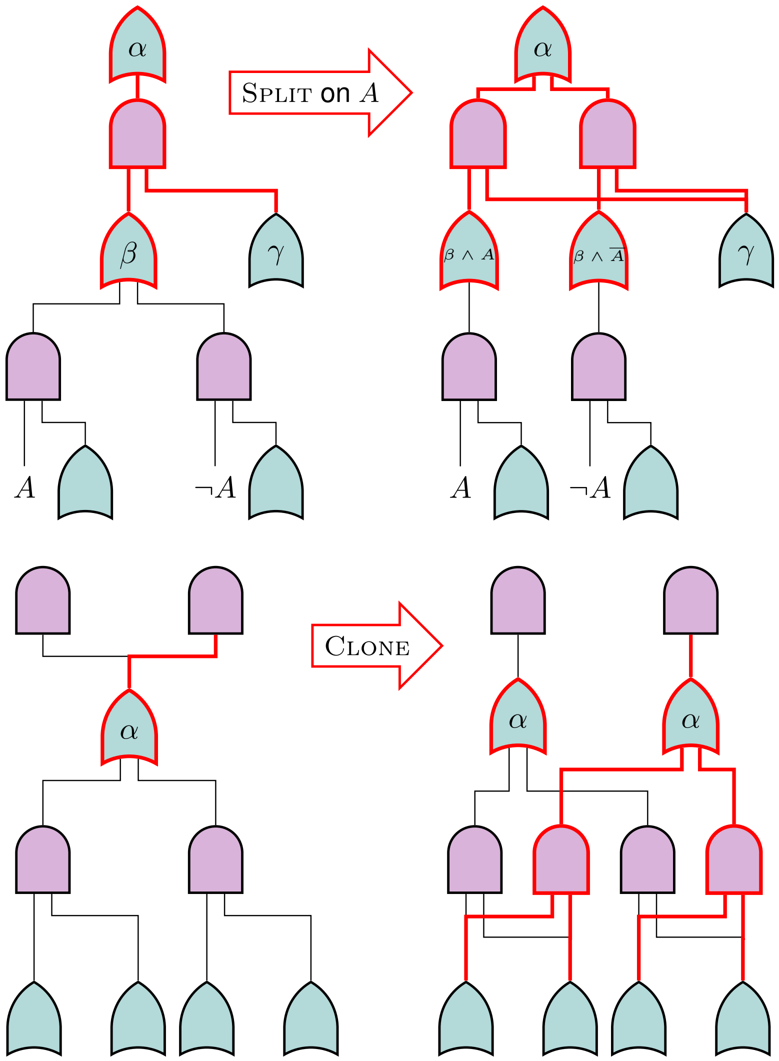
- ✗ Requires initial PSDD encoding the support...
- ✗ Scales poorly to complex formulae and/or high dimension...
- ✗ Costly whole circuit evaluation at every iteration...
- ✓ Very good performance!

STRUDEL (Dang et al. [2020]):

- ✓ Constructs an initial PSDD structure (from a CLT)!
- ✗ But does not encode constraints...
- ✓ Scales to high dimension!
- ✗ As long as the circuit doesn't get too big...

SAMPLEPSDD (this work):

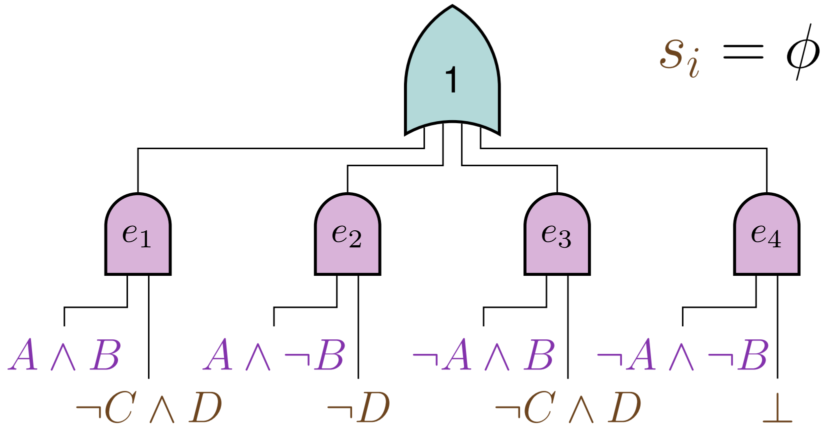
- ✓ Scales to high dimension and complex formulae!
- ✓ Constructs a structure consistent with constraints!
- ✗ But does so by relaxing the formula...
- ✗ Performance varies on set bounds and vtree structure...



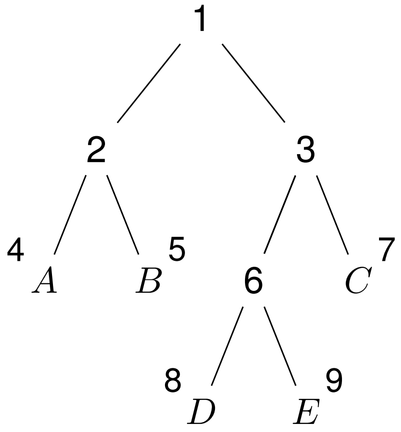
SamplePSDD

Common assumption: primes p_i are conjunctions of literals.

$$\phi(A, B, C, D) = (A \wedge \neg B \wedge \neg D) \vee (B \wedge \neg C \wedge D)$$



$$s_i = \phi|_{p_i}$$

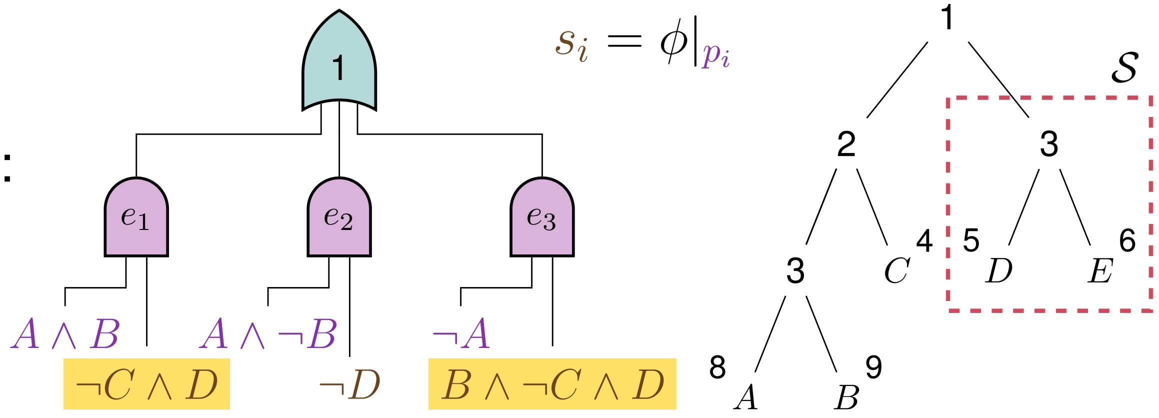


Problem: size of circuit is exponential in the size of p_i

SAMPLEPSDD

Solution: randomly sample a bounded number (k) of p_i

Example, $k = 3$:



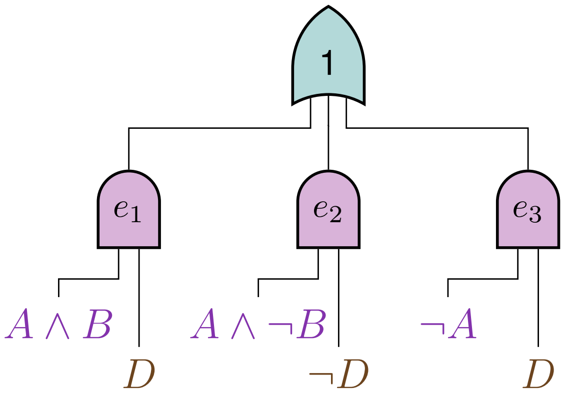
But: this violates structure decomposability

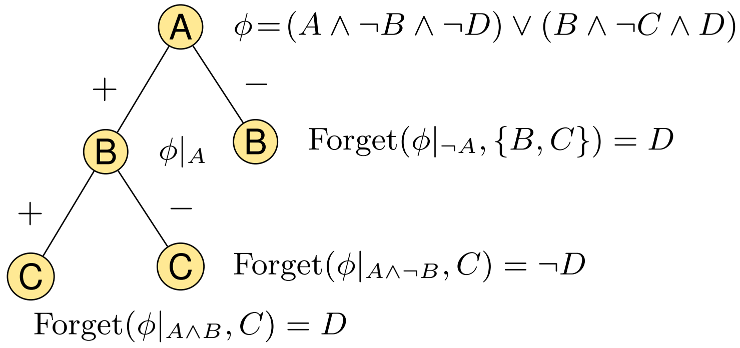
$\neg C \wedge D$ contains C, and $C \notin \mathcal{S}$

$\neg B \wedge \neg C \wedge D$ contains B and C, and $B, C \notin \mathcal{S}$

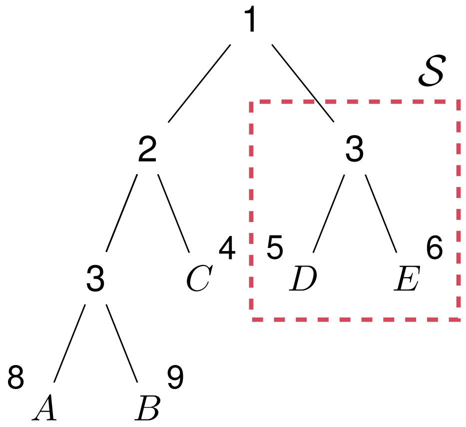
SamplePSDD

New solution: relax logical constraints ϕ





Now all s_i respect $\mathcal{S}$



Experiments

Evaluation: we sample 30 PSDDs and use 5 ensemble strategies:

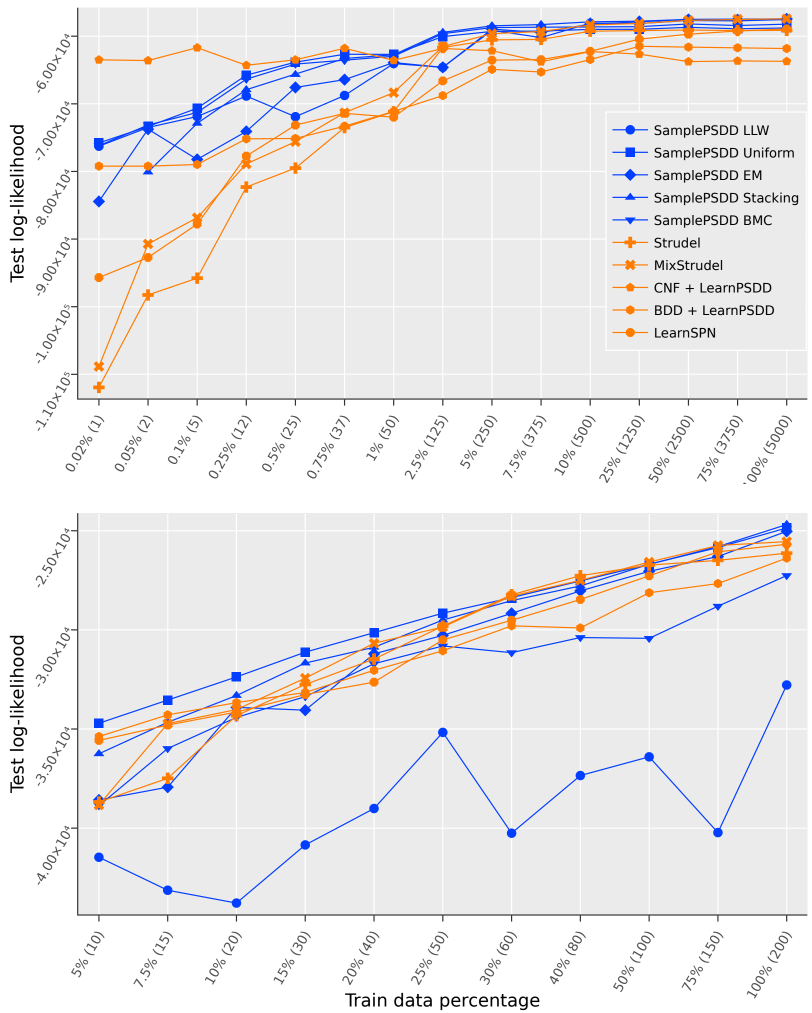
- ● **Likelihood weighting (LLW)**,
- ■ **Uniform weights**,
- ◆ **Expectation Maximization (EM)**,
- ▲ **Stacking**,
- ▼ **Bayesian Model Combination (BMC)**;

comparing against **STRUDEL**, **LEARNPSDD** and **LEARNSPN**.

Datasets: we evaluate with 5 data + knowledge as logic constraints:

	Dataset	#vars	#train	ϕ's size
$\Rightarrow$	LED	14	5000	23
$\Rightarrow$	LED + IMAGES	157	700	39899
	SUSHI RANKING	100	3500	17413
	SUSHI TOP 5	10	3500	37
	DOTA 2 GAMES	227	92650	1308

Our approach **fares better with fewer data**, yet **remains competitive under lots of data**.



Mattei et al. [2020], Kamishima [2003], Shen et al. [2017], Choi et al. [2015], Gens and Domingos [2013], Dang et al. [2020]

Experiments

Evaluation: we sample 30 PSDDs and use 5 ensemble strategies:

- ● **Likelihood weighting (LLW)**,
- ■ **Uniform weights**,
- ◆ **Expectation Maximization (EM)**,
- ▲ **Stacking**,
- ▼ **Bayesian Model Combination (BMC)**;

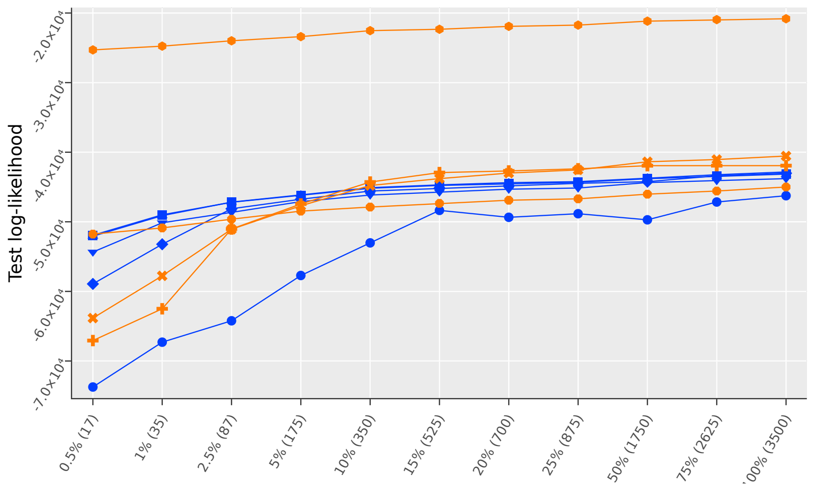
comparing against **Strudel**, **LearnPSDD** and **LearnSPN**.

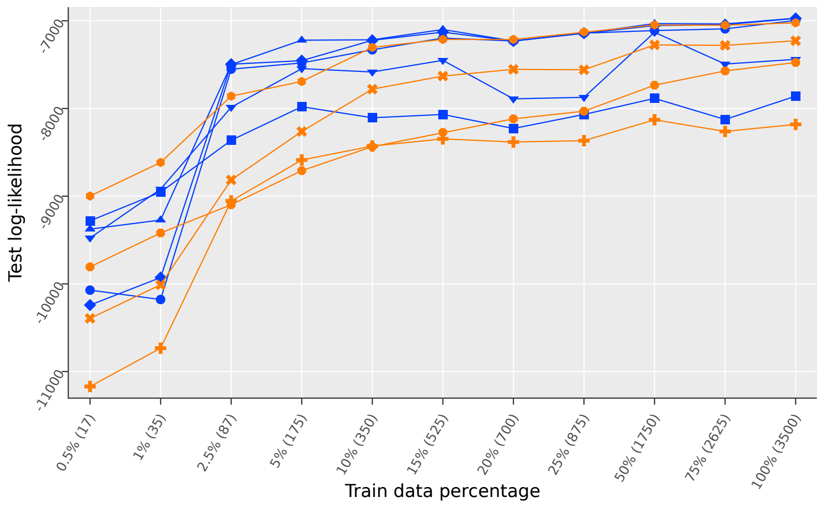
Datasets: we evaluate with 5 data + knowledge as logic constraints:

	Dataset	#vars	#train	ϕ's size
	LED	14	5000	23
	LED + Images	157	700	39899
$\Rightarrow$	Sushi Ranking	100	3500	17413
$\Rightarrow$	Sushi Top 5	10	3500	37
	Dota 2 Games	227	92650	1308

Our approach fares **better** with **fewer** data, yet remains **competitive** under **lots of data**.

Mattei et al. [2020], Kamishima [2003], Shen et al. [2017], Choi et al. [2015], Gens and Domingos [2013], Dang et al. [2020]





Experiments

Evaluation: we sample 30 PSDDs and use 5 ensemble strategies:

- ● **Likelihood weighting (LLW)**,
- ■ **Uniform weights**,
- ◆ **Expectation Maximization (EM)**,
- ▲ **Stacking**,
- ▼ **Bayesian Model Combination (BMC)**;

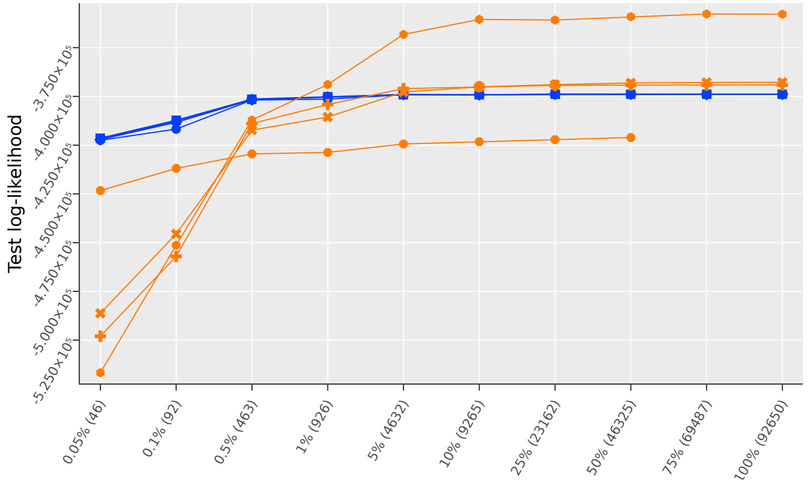
comparing against **Strudel**, **LearnPSDD** and **LearnSPN**.

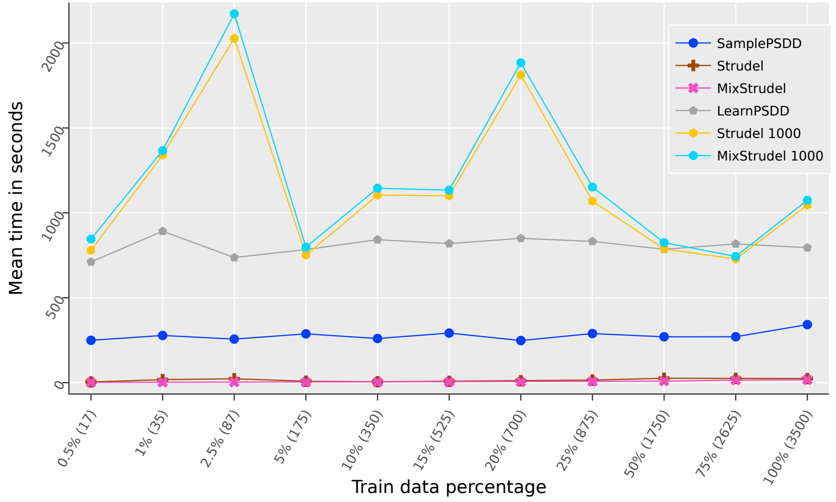
Datasets: we evaluate with 5 data + knowledge as logic constraints:

Dataset	#vars	#train	ϕ's size
LED	14	5000	23
LED + Images	157	700	39899
Sushi Ranking	100	3500	17413
Sushi Top 5	10	3500	37
$\Rightarrow$ Dota 2 Games	227	92650	1308

Our approach fares **better** with **fewer** data, yet
remains **competitive** under **lots of data**.

Mattei et al. [2020], Kamishima [2003], Shen et al. [2017],
Choi et al. [2015], Gens and Domingos [2013], Dang et al. [2020]

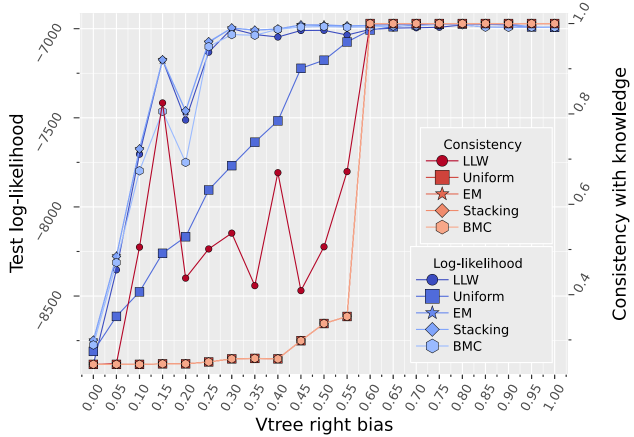


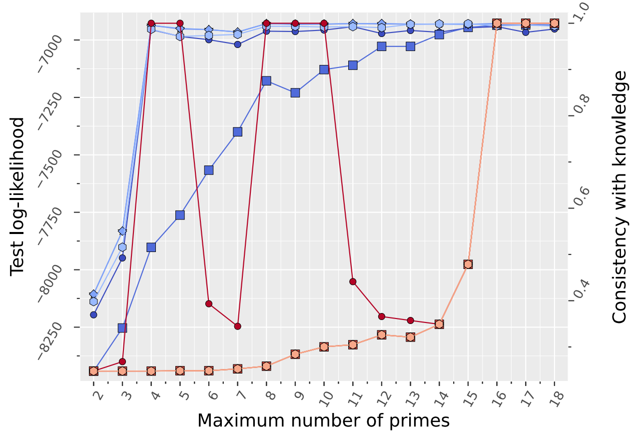


Experiments

What is the impact of higher k's and right-leaning vtrees

in log-likelihood and consistency?





Samples perform **better** with higher k's and right-leaning vtrees...

...but at a **cost** to complexity.