Taming Discrete Integration via the Boon of Dimensionality

Jeffrey M. Dudek

Rice University

Dror Fried

The Open University of Israel

Kuldeep S. Meel

National University of Singapore







What is the probability that an input sampled **uniformly from all inputs** is adversarial for a given neural network?

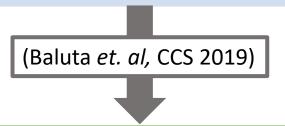
What is the probability that an input sampled **uniformly from all inputs** is adversarial for a given neural network?

(Baluta *et. al,* CCS 2019)

Unweighted Projected Model Counting

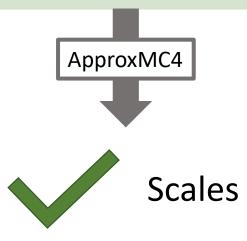
How many discrete solutions does a set of equations have?

What is the probability that an input sampled **uniformly from all inputs** is adversarial for a given neural network?

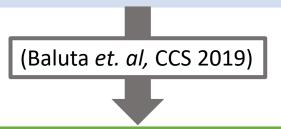


Unweighted Projected Model Counting

How many discrete solutions does a set of equations have?

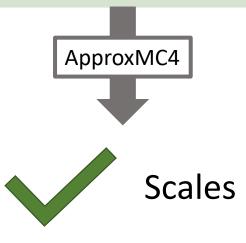


What is the probability that an input sampled **uniformly from all inputs** is adversarial for a given neural network?



Unweighted Projected Model Counting

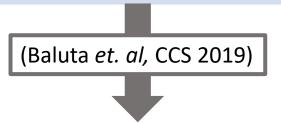
How many discrete solutions does a set of equations have?



Neural Network (Log-Linear) Robustness

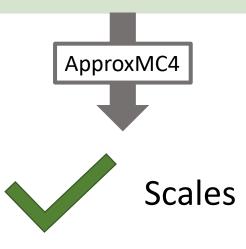
What is the probability that an input sampled **from a log-linear distribution** is adversarial for a given neural network?

What is the probability that an input sampled **uniformly from all inputs** is adversarial for a given neural network?



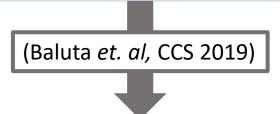
Unweighted Projected Model Counting

How many discrete solutions does a set of equations have?



Neural Network (Log-Linear) Robustness

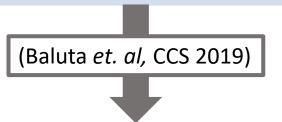
What is the probability that an input sampled **from a log-linear distribution** is adversarial for a given neural network?



Discrete Integration

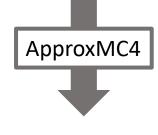
How many **weighted** discrete solutions does a set of equations have?

What is the probability that an input sampled **uniformly from all inputs** is adversarial for a given neural network?



Unweighted Projected Model Counting

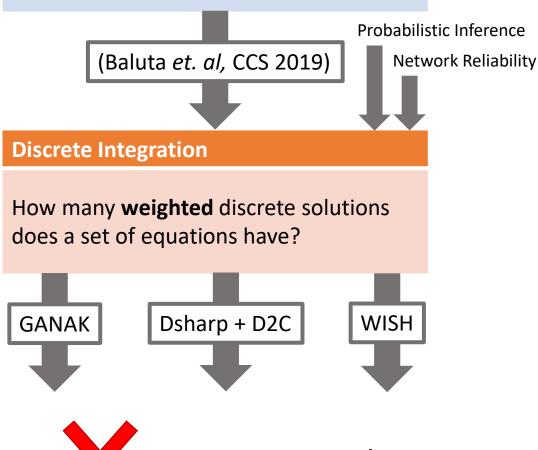
How many discrete solutions does a set of equations have?

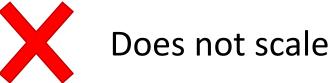




Neural Network (Log-Linear) Robustness

What is the probability that an input sampled **from a log-linear distribution** is adversarial for a given neural network?





What is the probability that an input sampled **uniformly from all inputs** is adversarial for a given neural network?

(Baluta *et. al,* CCS 2019)

Unweighted Projected Model Counting

How many discrete solutions does a set of equations have?

ApproxMC4



Neural Network (Log-Linear) Robustness

What is the probability that an input sampled **from a log-linear distribution** is adversarial for a given neural network?

(Baluta et. al, CCS 2019)

Probabilistic Inference
Network Reliability

Discrete Integration

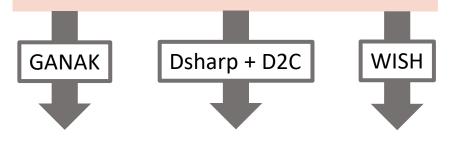
Our contribution

Add new variables

simulate weights

to **exactly**

How many **weighted** discrete solutions does a set of equations have?





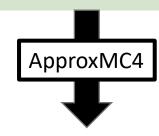
Does not scale

What is the probability that an input sampled **uniformly from all inputs** is adversarial for a given neural network?

(Baluta et. al, CCS 2019)

Unweighted Projected Model Counting

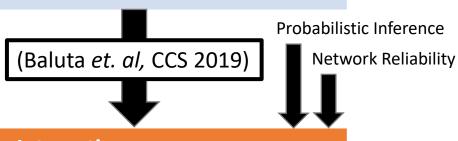
How many discrete solutions does a set of equations have?





Neural Network (Log-Linear) Robustness

What is the probability that an input sampled **from a log-linear distribution** is adversarial for a given neural network?



Discrete Integration

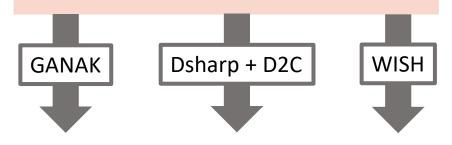
Our contribution

Add new variables

simulate weights

to **exactly**

How many **weighted** discrete solutions does a set of equations have?





Does not scale