Probabilistic Programming and Machine Learning

Sam Staton
Oxford Computer Science

Probabilistic programming is...

Writing statistical models by writing programs.

Ref: Intro to prob prog (van de Meent, Paige, Yang, Wood) arxiv:1809.10756

Probabilistic programming is...

Writing statistical models by writing programs.

Statistical models

Covid-19 models

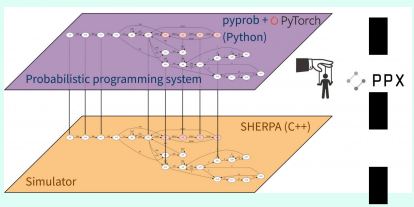
Large Hadron Collider models

Encoder / decoder

Models of cognition

...

Baydin et al, Etalumis. arxiv: 1907.03382



Chandra et al, Memo on Memo, 2025

Inference algorithms

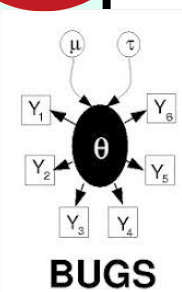
Langevin Monte Carlo

Hamiltonian Monte Carlo

Sequential Monte Carlo

Variational Inference

...



PyMC

PYRO

Ref: Intro to prob prog (van de Meent, Paige, Yang, Wood) arxiv:1809.10756

Probabilistic Programming

1. **Example and overview**
2. Programming language ideas
3. Safeguarded AI

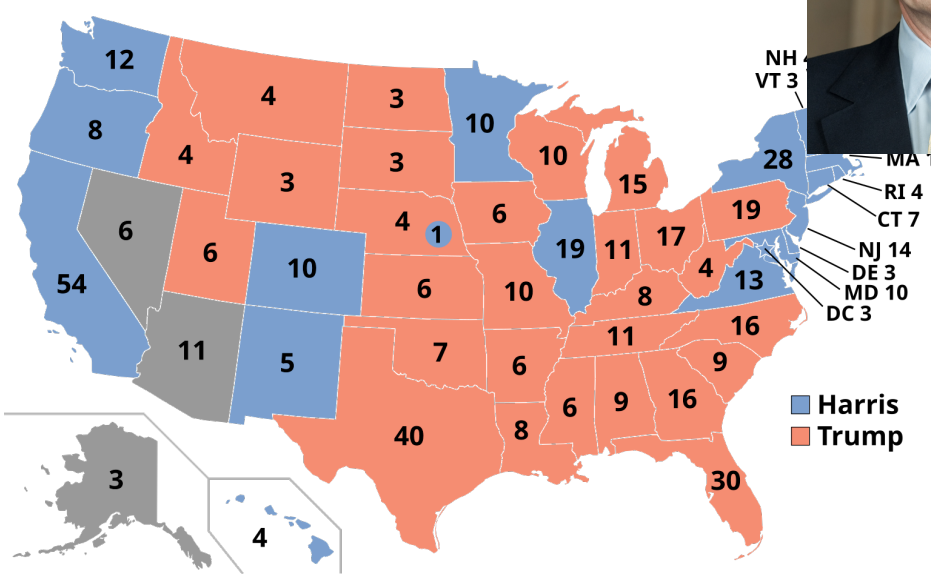
High level view: poll example

A very simple model deducing chance of win from poll.

Question:

A quick poll gives 51:49 votes. What is the chance of winning?

Roger Harris, CCBY



UK FCDO and Government, OGL v3

High level view: poll example

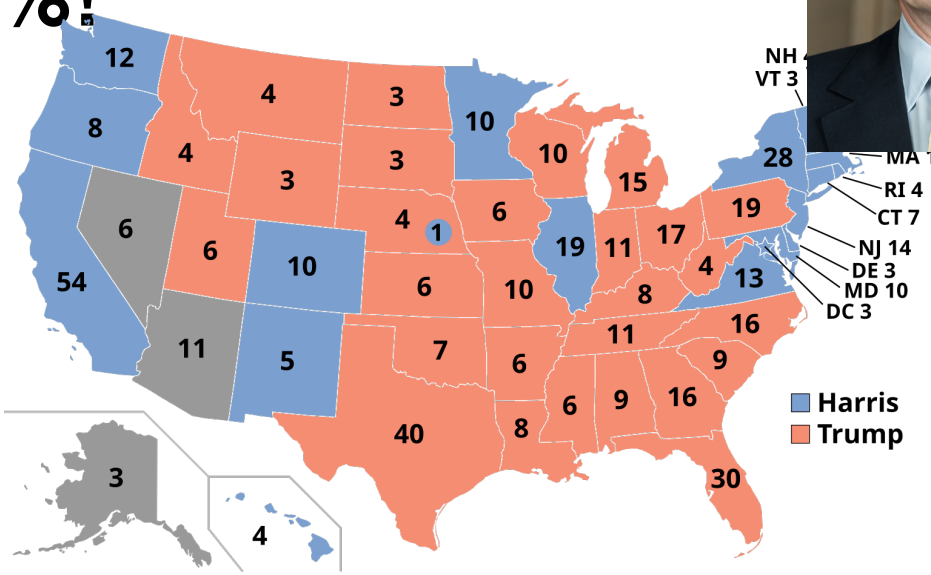
A very simple model deducing chance of win from poll.

Question:

A quick poll gives 51:49 votes. What is the chance of winning?

Clue: it's not 51%!

Roger Harris, CCBY



UK FCDO and Government, OGL v3

High level view: poll example

A very simple model deducing chance of win from poll.

$$v \sim Uniform(0,1) \qquad d_i \sim Bernoulli(v) \quad (i = 1...100)$$

Traditional stats model

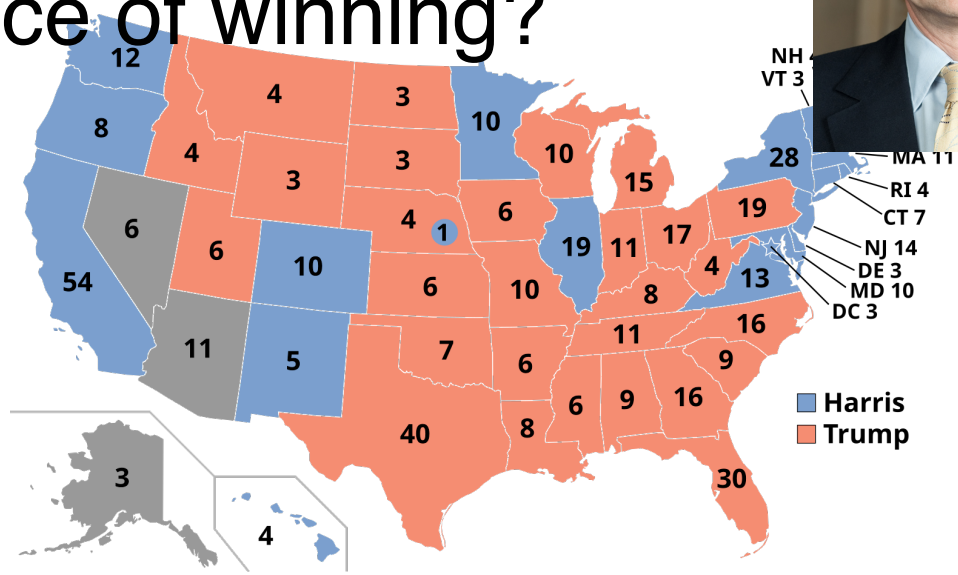
UK FCDO and Government, OGL v3

Question:

A quick poll gives 51:49 votes.

What is the chance of winning?

Roger Harris, CCBY



High level view: poll example

A very simple model deducing chance of win from poll.

$$v \sim Uniform(0,1) \qquad d_i \sim Bernoulli(v) \quad (i = 1...100)$$

What is $P(v > 0.5 \mid d_i = poll_i)$?

Traditional stats model

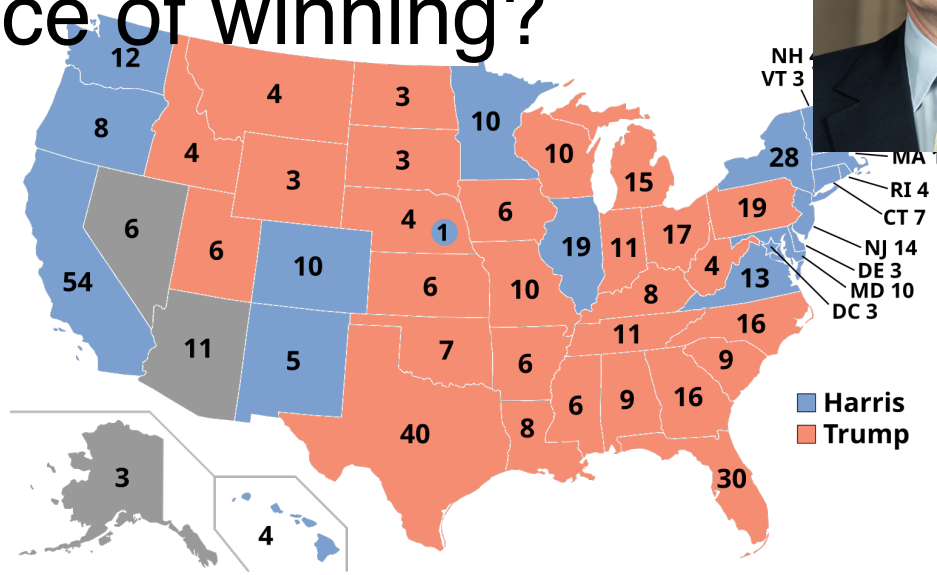
UK FCDO and Government, OGL v3

Question:

A quick poll gives 51:49 votes.

What is the chance of winning?

Roger Harris, CCBY



High level view: poll example

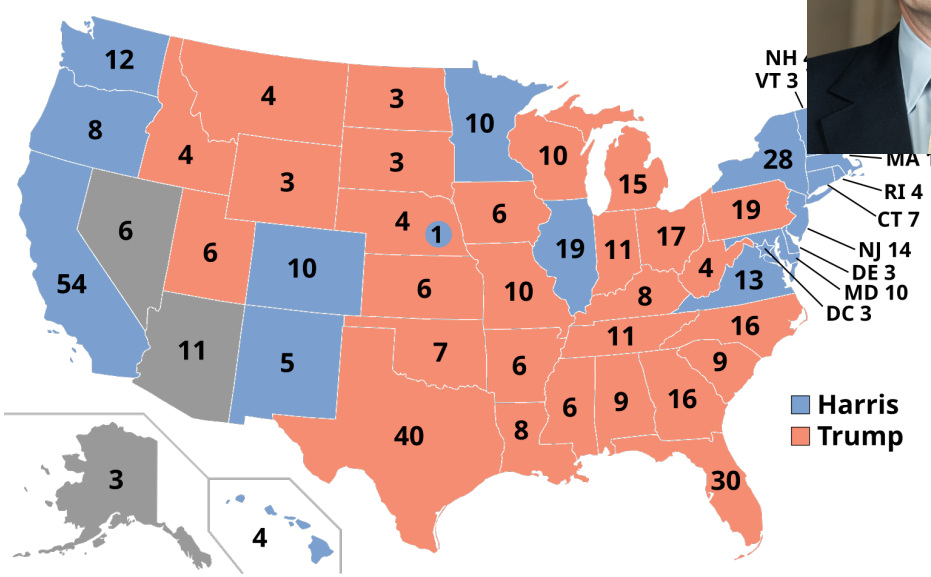
A very simple model deducing chance of win from poll.

```
model :: Prob ([Bool] , Bool)
model = do
  voteShare <- uniform 0 1
  votes <- replicateM 100 (bernoulli voteShare)
  return (votes , (voteShare > 0.5))
```

Probabilistic program

UK FCDO and Government, OGL v3

Roger Harris, CCBY



High level view: poll example

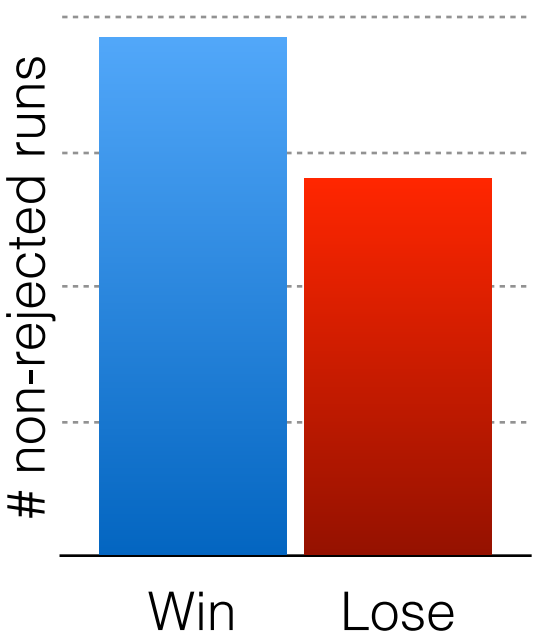
A very simple model deducing chance of win from poll.

```
model :: Prob ([Bool] , Bool)
model = do
  voteShare <- uniform 0 1
  votes <- replicateM 100 (bernoulli voteShare)
  return (votes , (voteShare > 0.5))
```

Probabilistic program

Crude rejection sampling Monte Carlo:

- Run 1000000s of times, each time getting (poll result, win?)
- Reject the runs that mis-predict poll
- What proportion of the remainder are winners?



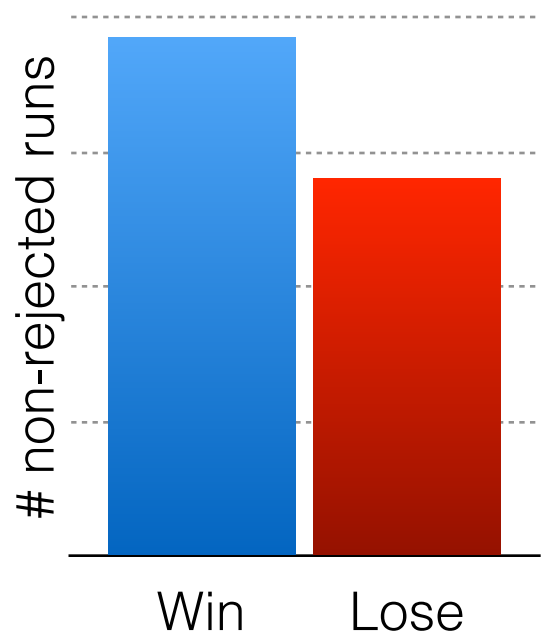
High level view: poll example

A very simple model deducing chance of win from poll.

Question:
A quick poll gives 51:49 votes. What is the chance of winning?
Answer: 0.58.

Crude rejection sampling Monte Carlo:

- Run 1000000s of times, each time getting (poll result, win?)
- Reject the runs that mis-predict poll
- What proportion of the remainder are winners?



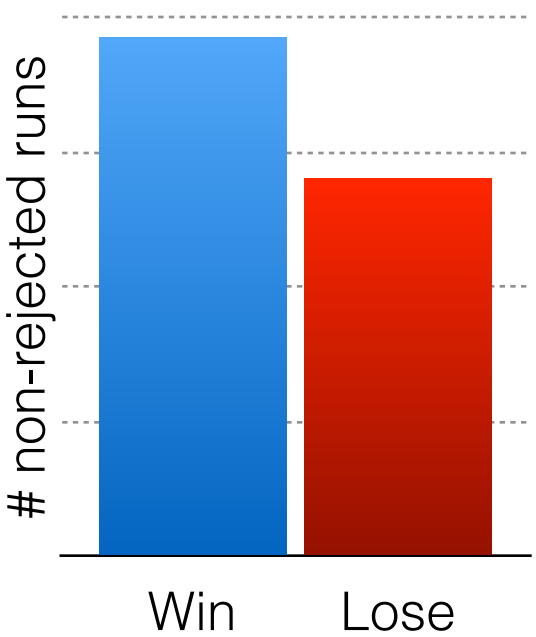
High level view: poll example

A very simple model deducing chance of win from poll.

```
model :: Prob ([Bool] , Bool)
model = do
  voteShare <- uniform 0 1
  votes <- replicateM 100 (bernoulli voteShare)
  return (votes , (voteShare > 0.5))
```

Probabilistic program

- What sort of fancy code can we write here?
- How fast/accurately will it be modelled?



Probabilistic programming is...

Writing statistical models by writing programs.

Statistical models

Covid-19 models

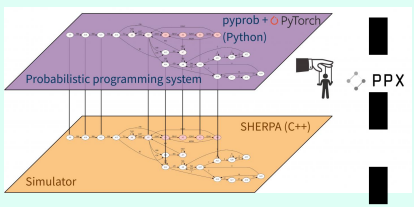
Large Hadron Collider models

Encoder / decoder

Models of cognition

...

Baydin et al, Etalumis. arxiv: 1907.03382



Chandra et al, Memo on Memo, 2025

Inference algorithms

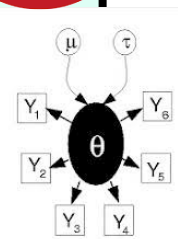
Langevin Monte Carlo

Hamiltonian Monte Carlo

Sequential Monte Carlo

Variational Inference

...



PyMC

PYRO

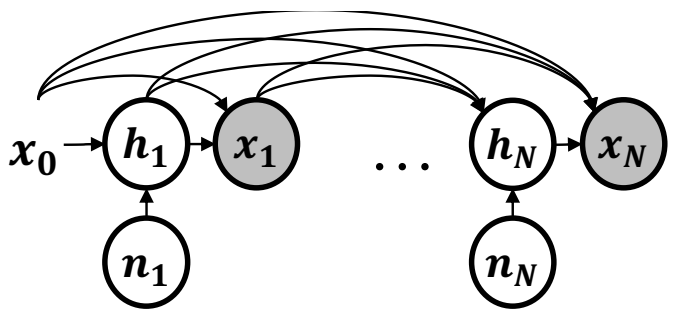
Ref: Intro to prob prog (van de Meent, Paige, Yang, Wood) arxiv:1809.10756

High level example: infilling LLM

Idea: use LLM as a distribution (e.g. LLaMPPL)

See also e.g. Bengio et al, Flow Network based Generative Models

Prompt:
"To tell the truth, every**[BLANK]** he**[BLANK]** to**[BLANK]** another**[BLANK]**."




```
def infilling(prompt):
  # Initialize
  parts = prompt.split("[BLANK]")
  s = parts[0]
  x = new_context(s)
  # Generate for each blank
  for part in parts[1:]:
    n = sample(geom(0.5)) + 1
    for _ in range(n):
      s += sample(llm(x))
    for t in tokenize(part):
      s += observe(llm(x), t)
  return s
```

"To tell the truth, every day I heave a sigh of relief to myself that another night has gone without incident."

From Lew et al., Sequential Monte Carlo Steering of Large Language Models using Probabilistic Programs. arxiv:2306.03081.

Probabilistic Programming

1. Example and overview
2. **Programming language ideas**
3. Safeguarded AI

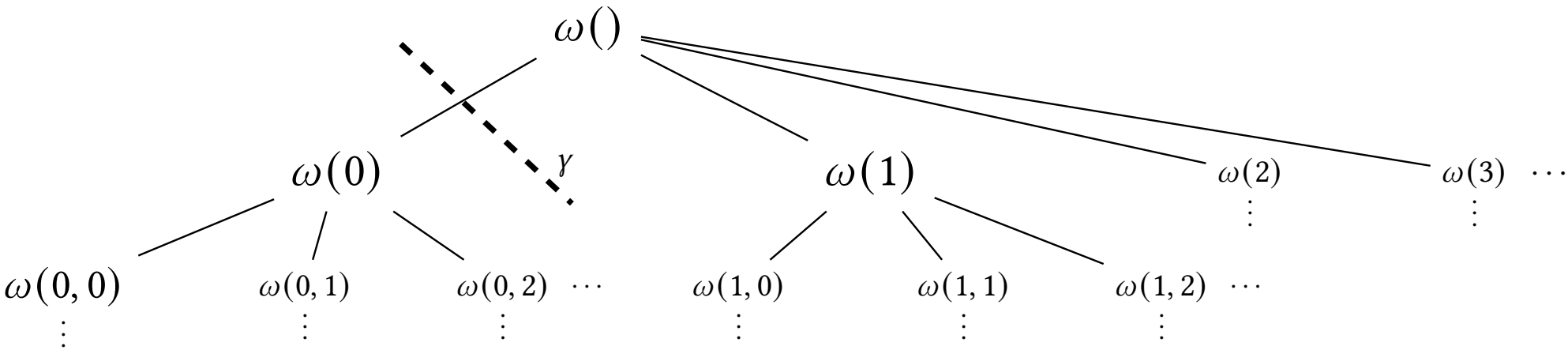
Programming language ideas

Often want to say *"let X be an independent uniform random variable"*

Idea in LazyPPL / quasi-Borel spaces:

Ω = **infinite lazy rose trees** labelled by uniform random draws.

So $\Omega \cong \Omega \times \Omega$

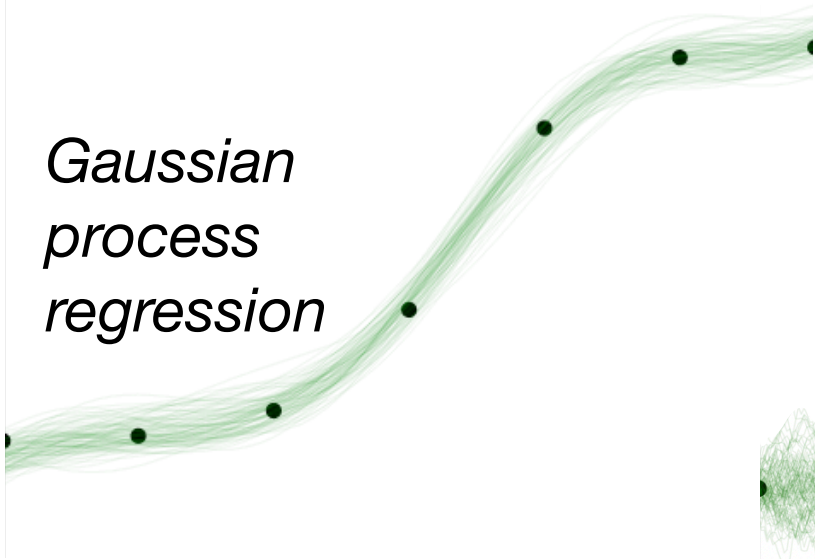


From Dash et al, Affine Monads and Lazy Structures for Bayesian Programming. arxiv:2212.07250.

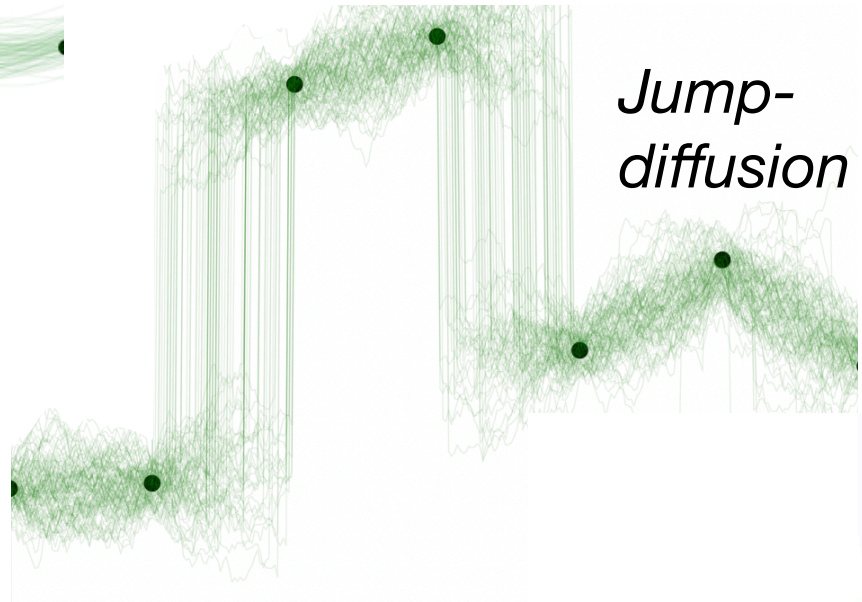
Infinite rose trees in LazyPPL

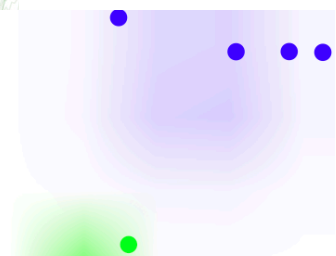
lazyppl-team.github.io

Gaussian process regression



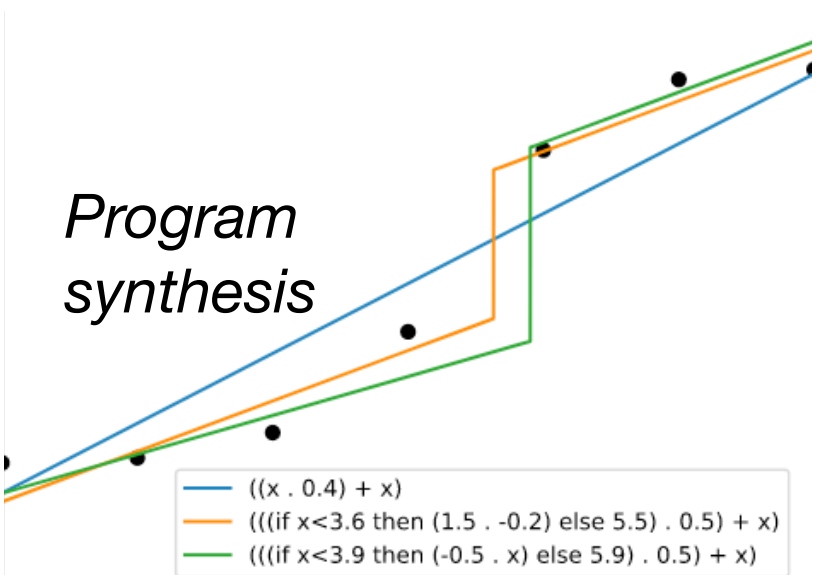
Jump-diffusion

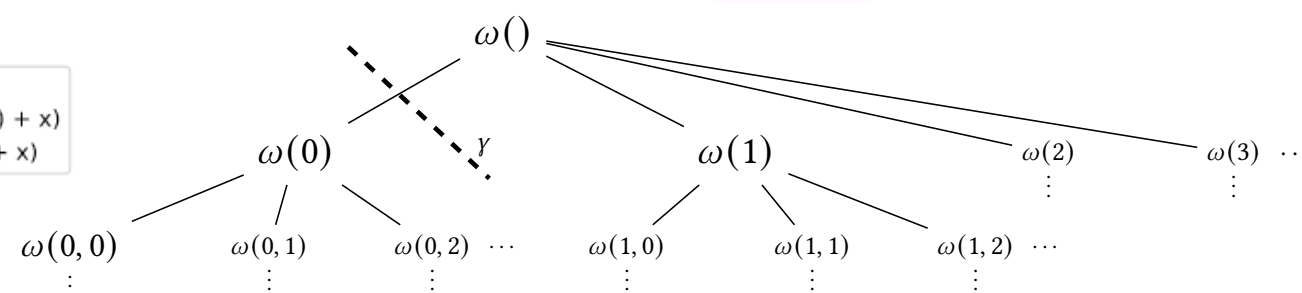




Dirichlet process clustering

Program synthesis

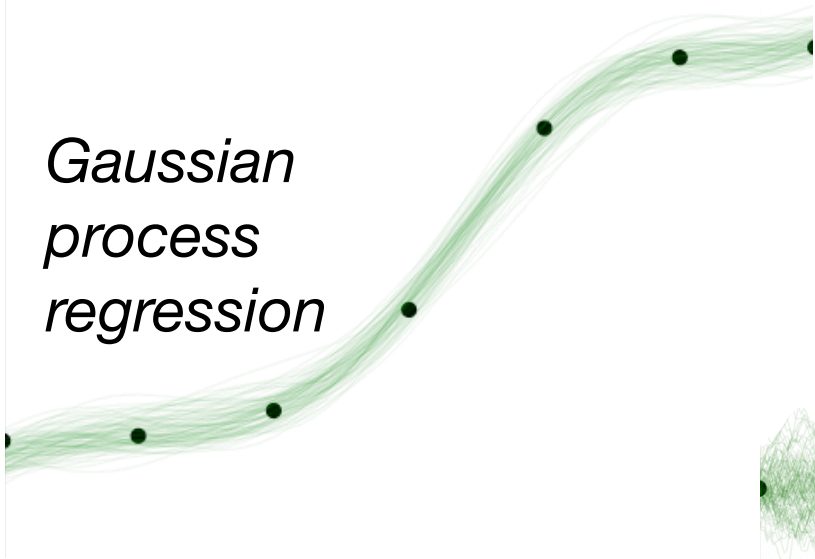




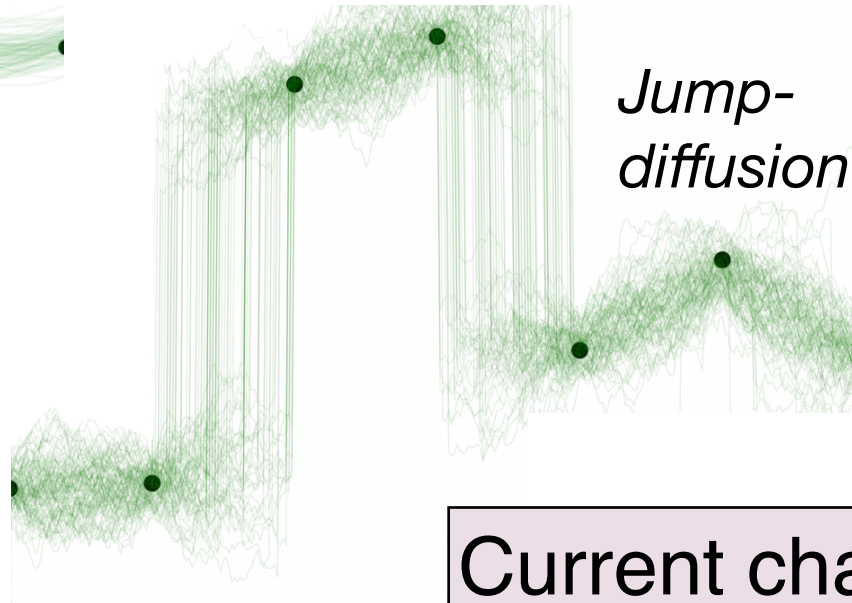
Infinite rose trees in LazyPPL

lazyppl-team.github.io

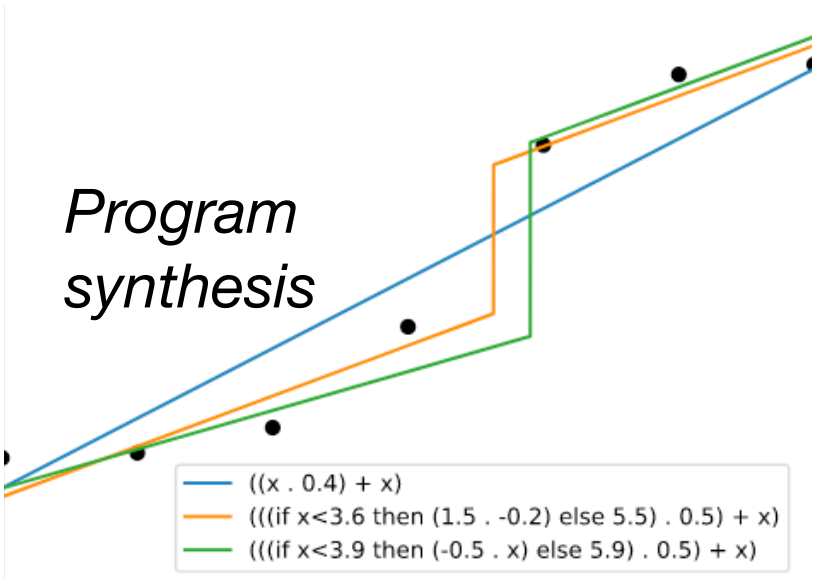
Gaussian process regression



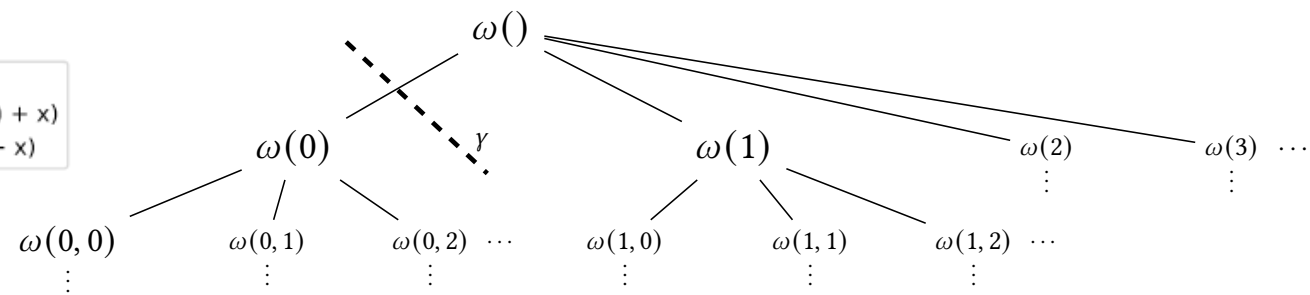
Jump-diffusion



Program synthesis



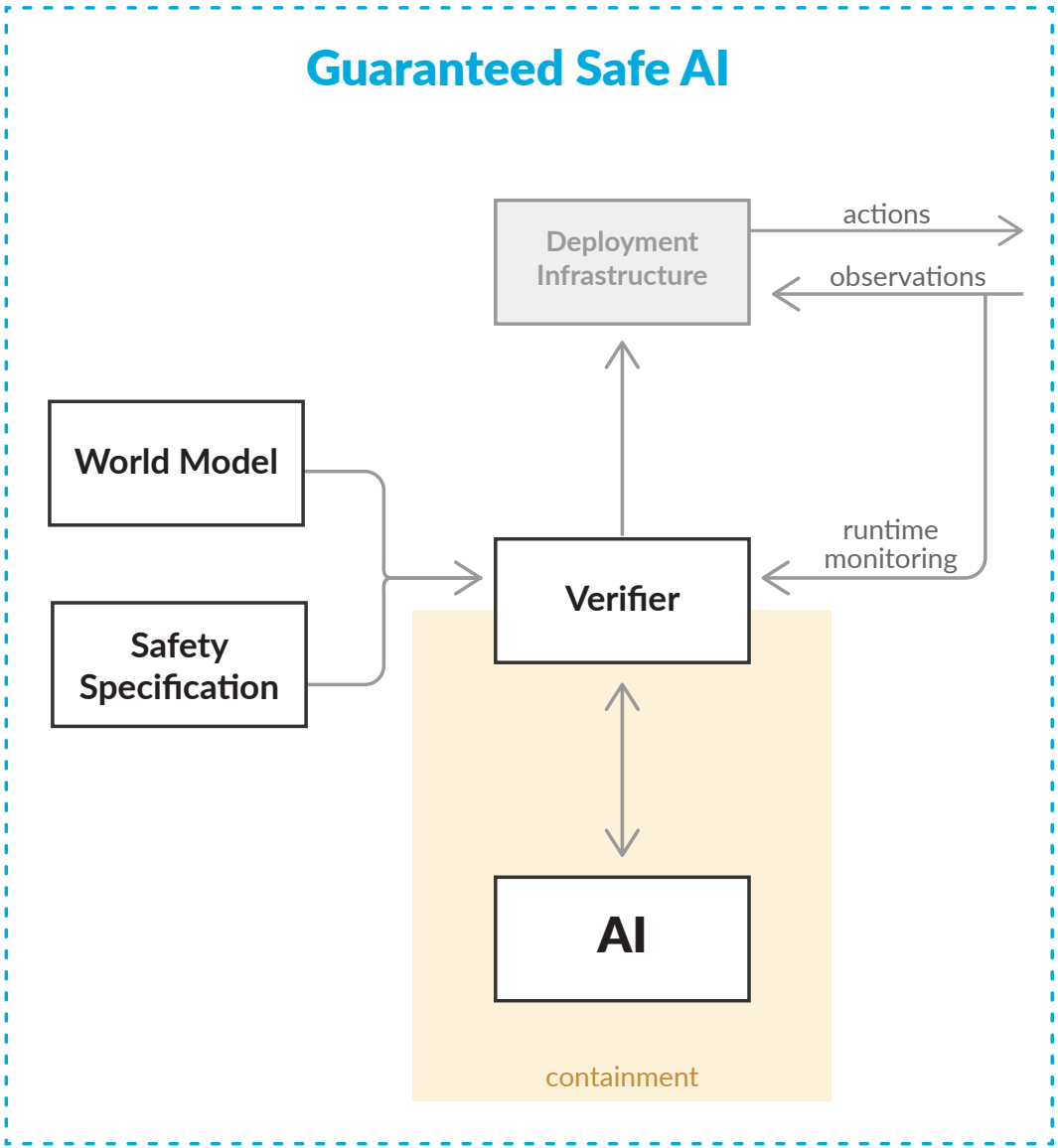
Current challenge: What is Hamiltonian Monte Carlo over this infinite dimensional space?



Probabilistic Programming

1. Example and overview
2. Programming language ideas
3. **Safeguarded AI**

Safeguarded AI proposal



Idea:
World model should

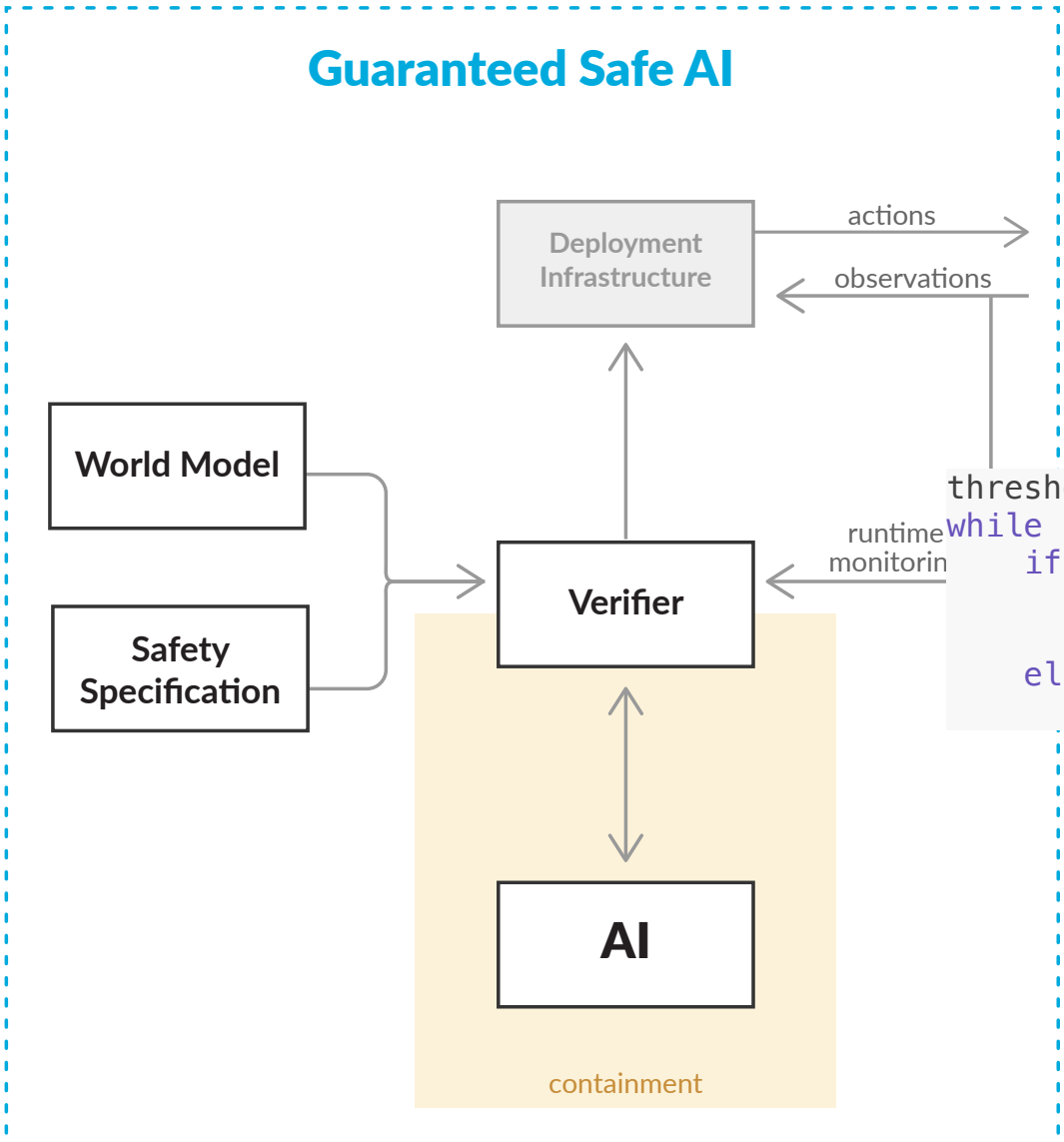
- be interpretable
- include physics, psychology, ...

Probabilistic programming languages are one good candidate.

Who writes it?

From David 'davidad' Dalrymple, Joar Skalse, et al. Towards Guaranteed Safe AI. arxiv:2405.06624.

Safeguarded AI proposal

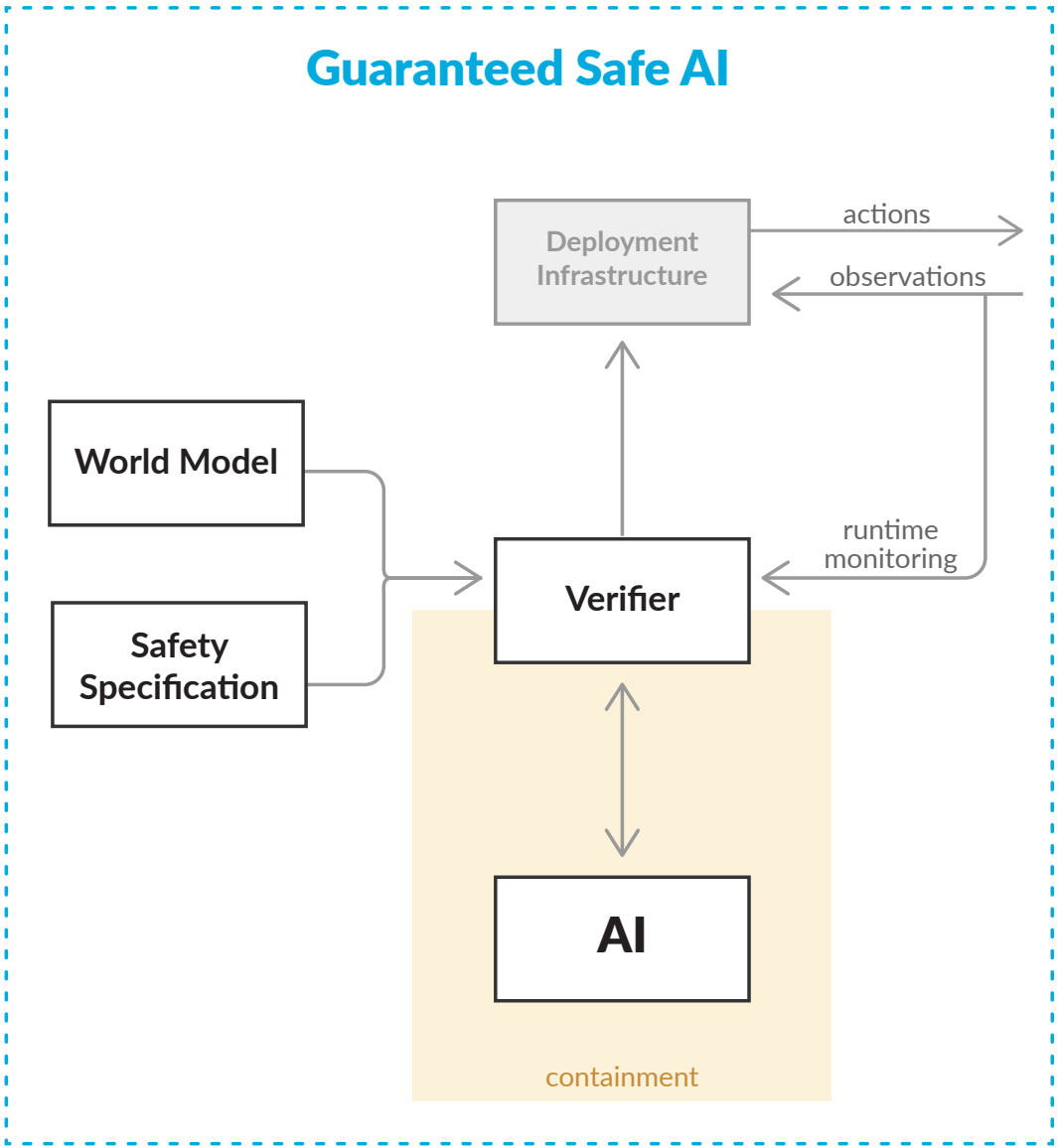


e.g.
autonomous vehicles
world model in Scenic

```
threshold = Range(4, 7)
while True:
    if self.distanceToClosest(Pedestrian) < threshold:
        strength = TruncatedNormal(0.8, 0.02, 0.5, 1)
        take SetBrakeAction(strength), SetThrottleAction(0)
    else:
        take SetThrottleAction(0.5), SetBrakeAction(0)
```

From Fremont et al, Scenic, a language for scenario specification and scene generation. arxiv:1809.09310.

Safeguarded AI proposal



Idea:
World model should

- be interpretable
- include physics, psychology, ...

Probabilistic programming languages are one good candidate.

Who writes it?

From David 'davidad' Dalrymple, Joar Skalse, et al. Towards Guaranteed Safe AI. arxiv:2405.06624.

Probabilistic programming is...

Writing statistical models by writing programs.

Statistical models

Covid-19 models

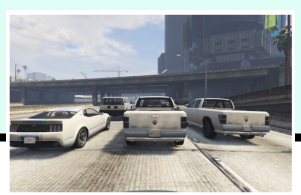
LHC models

Encoder / decoder

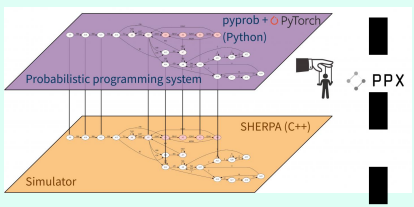
Models of cognition

Vehicle simulator

...



Baydin et al, Etalumis. arxiv: 1907.03382



Chandra et al, Memo on Memo, 2025

Inference algorithms

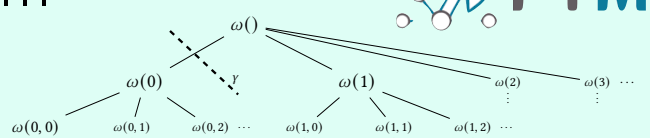
Langevin Monte Carlo

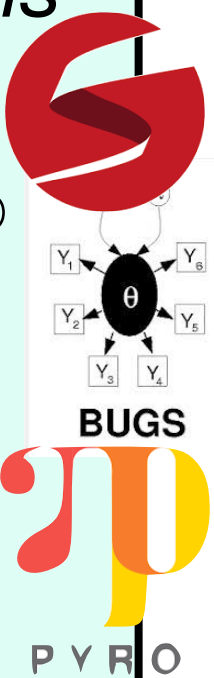
Hamiltonian Monte Carlo

Sequential Monte Carlo

Variational Inference

...





Ref: Intro to prob prog (van de Meent, Paige, Yang, Wood) arxiv:1809.10756