On the relationship between Adaptive Monte Carlo and Online Learning

NordStat2021
24 June 2021

Xiaoyu Lu, Tom Rainforth, Yee Whye Teh
Department of Statistics, University of Oxford

https://arxiv.org/abs/1810.13296

UNIVERSITY OF OXFORD

Metropolis Algorithm

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

Equation of State Calculations by Fast Computing Machines

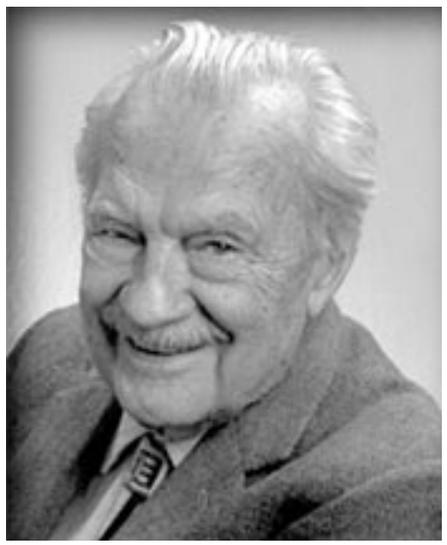
NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.








[Metropolis et al 1953, Hastings 1970]

Rosenbluth?
~~Metropolis~~ Algorithm

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 21, NUMBER 6 JUNE, 1953

Equation of State Calculations by Fast Computing Machines

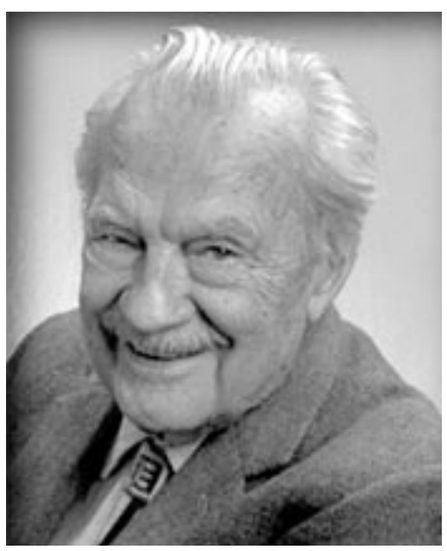
NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.








[Metropolis et al 1953, Hastings 1970, Rosenbluth 2003, Gubernatis 2005]

Monte Carlo Simulation

- Sample space $\mathcal{X}$.
- Target distribution $\pi(x) = f(x)/Z$ over $\mathcal{X}$.
- Target expectations $\mathbb{E}_\pi[\varphi(x)]$ for test functions $\varphi(x)$.

- Procedure to simulate random samples $x_1, x_2, \ldots, x_n \in \mathcal{X}$.
- Approximate target expectations

$$\mathbb{E}_\pi[\varphi(x)] \approx \frac{1}{n} \sum_{i=1}^{n} \varphi(x_i)$$

- Weighted samples, with weights $w_1, w_2, \ldots, w_n \in \mathbb{R}_+$.

$$\mathbb{E}_\pi[\varphi(x)] \approx \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i \varphi(x_i)$$

Markov Chain Monte Carlo

- Construct a Markov chain with Markov kernel K targeting π

$$X_1 \to X_2 \to X_3 \to \cdots$$

$$\mathbb{P}(X_{n+1} \in dx_{n+1} \,|\, X_n = x_n) = K(dx_{n+1}, x_n)$$

- Ergodic average after warmup period

$$\mathbb{E}_\pi[\varphi(x)] \approx \frac{1}{n} \sum_{i=m+1}^{m+n} \varphi(x_i)$$

- Efficacy of MCMC depends crucially on choice of Markov kernels, e.g.:
  - Choice of step size, number of leap-frog steps, mass matrix in Hamiltonian Monte Carlo
  - Choice of proposal kernel in ~~Metropolis-~~Rosenbluth-Hastings Algorithm

Importance Sampling

- Draw samples $x_1, x_2, \ldots, x_n \sim_{iid} q$ from a proposal distribution q.

$$\mathbb{E}_\pi[\varphi(x)] = \mathbb{E}_q\left[\frac{\pi(x)}{q(x)}\varphi(x)\right] \approx \frac{1}{n}\sum_{i=1}^{n} w_i \varphi(x_i)$$

  with weights $w_i = w(x_i) = \pi(x_i)/q(x_i)$.

- If $\pi \ll q$, estimates are unbiased, with variance

$$\frac{\sigma_q^2}{n} = \frac{1}{n}\left(\mathbb{E}_q\left[w(x)^2\varphi(x)^2\right] - \mathbb{E}_\pi[\varphi(x)]^2\right)$$

- Efficacy of IS determined by proposal distribution.
- Sample size required for generic $\varphi(x)$ is $\exp(\mathrm{KL}(\pi\|q))$.

[Owen 2013, Chatterjee & Diaconis 2018]

Self-Normalised Importance Sampling

- Often, target is known up to normalising constant, $\pi(x) = f(x)/Z$ with known $f(x)$ but unknown Z.
- Draw samples $x_1, x_2, \ldots, x_n \sim_{iid} q$ from a proposal distribution q.

$$\mathbb{E}_\pi[\varphi(x)] = \frac{\int f(x)\varphi(x)dx}{\int f(x)dx} = \frac{\mathbb{E}_q\left[\frac{f(x)}{q(x)}\varphi(x)\right]}{\mathbb{E}_q\left[\frac{f(x)}{q(x)}\right]} \approx \frac{\sum_{i=1}^n w_i\varphi(x_i)}{\sum_{i=1}^n w_i}$$

with weights $w_i = w(x_i) = f(x_i)/q(x_i)$.

- If $\pi \ll q$, estimates are biased but consistent, with variance

$$\frac{\sigma^2_{q,\mathsf{sn}}}{n} = \frac{1}{n}\mathbb{E}_q\left[w(x)^2(\varphi(x) - \mathbb{E}_\pi[\varphi(x)])^2\right]$$

- Weights are unbiased estimates of Z:

$$Z = \int_{\mathcal{X}} f(x)dx = \mathbb{E}_q\left[\frac{f(x)}{q(x)}\right]$$

[Owen 2013]

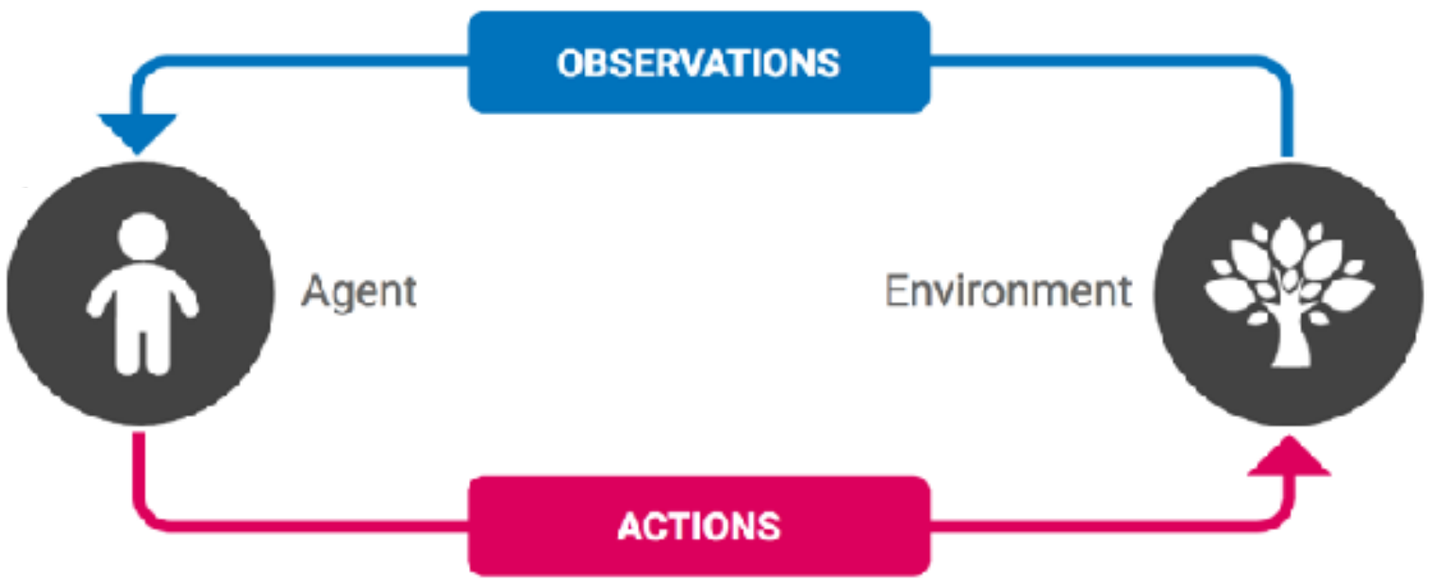
Adaptive Monte Carlo

$$\text{Algorithm}_1 \to x_1$$
$$x_1 \to \text{Algorithm}_2 \to x_2$$
$$x_{1:2} \to \text{Algorithm}_3 \to x_3$$
$$\vdots$$
$$x_{1:t} \to \text{Algorithm}_{t+1} \to x_{t+1}$$

- Use previous samples $x_{1:t}$
  - to learn about target π and
  - to improve the Monte Carlo algorithm.
- However: adaptations can go wrong.
- A form of online learning.

[Oh & Berger 1992, Tierney & Mira 1999, Atchade & Rosenthal 2005, Roberts & Rosenthal 2009, Atchade et al 2011, Vihola 2012, many others…]

Online Learning



- Reinforcement learning:
  - actions affect environment: changing evironment state
  - complex temporal planning.
- Online learning:
  - fixed environment
  - actions/observations gain rewards and information
  - changing information state.
  - exploration-exploitation trade-off.

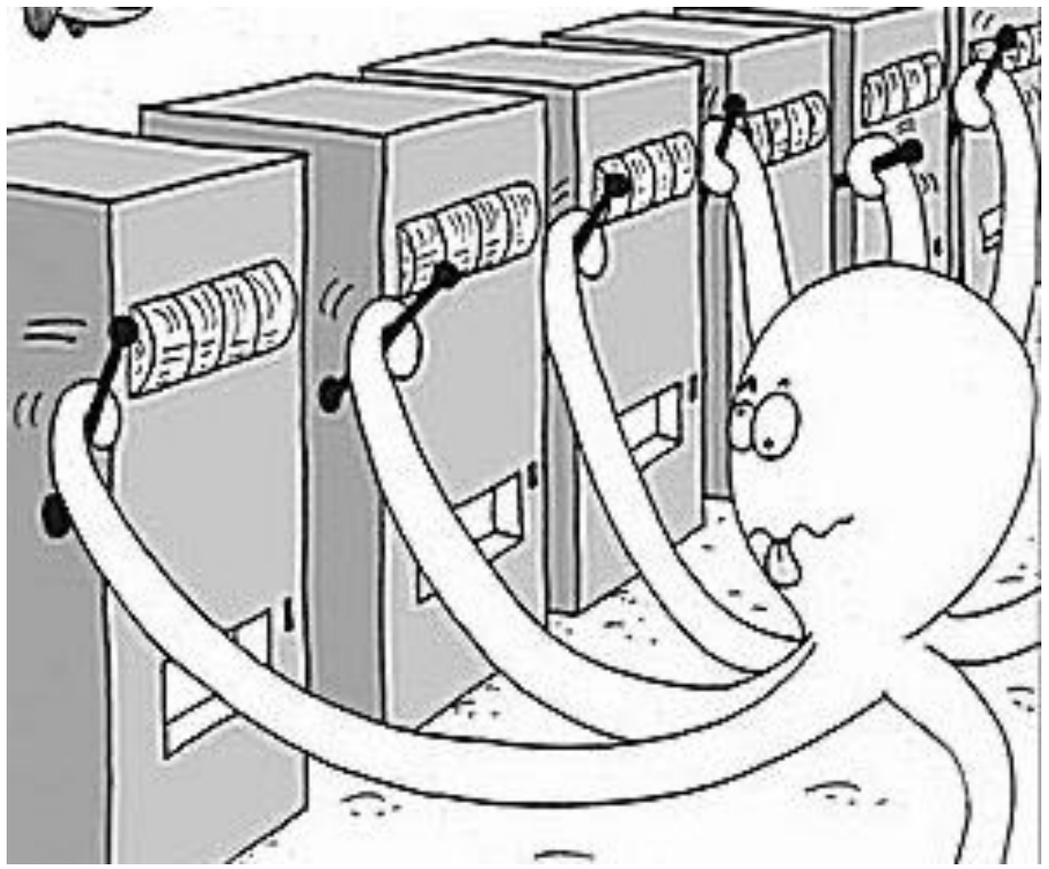
[Bubeck & Cesa-Bianchi 2012, Lattimore & Szepesvari 2020]

Multi-Armed Bandits

- Each one-armed bandit gives reward with unknown probability.

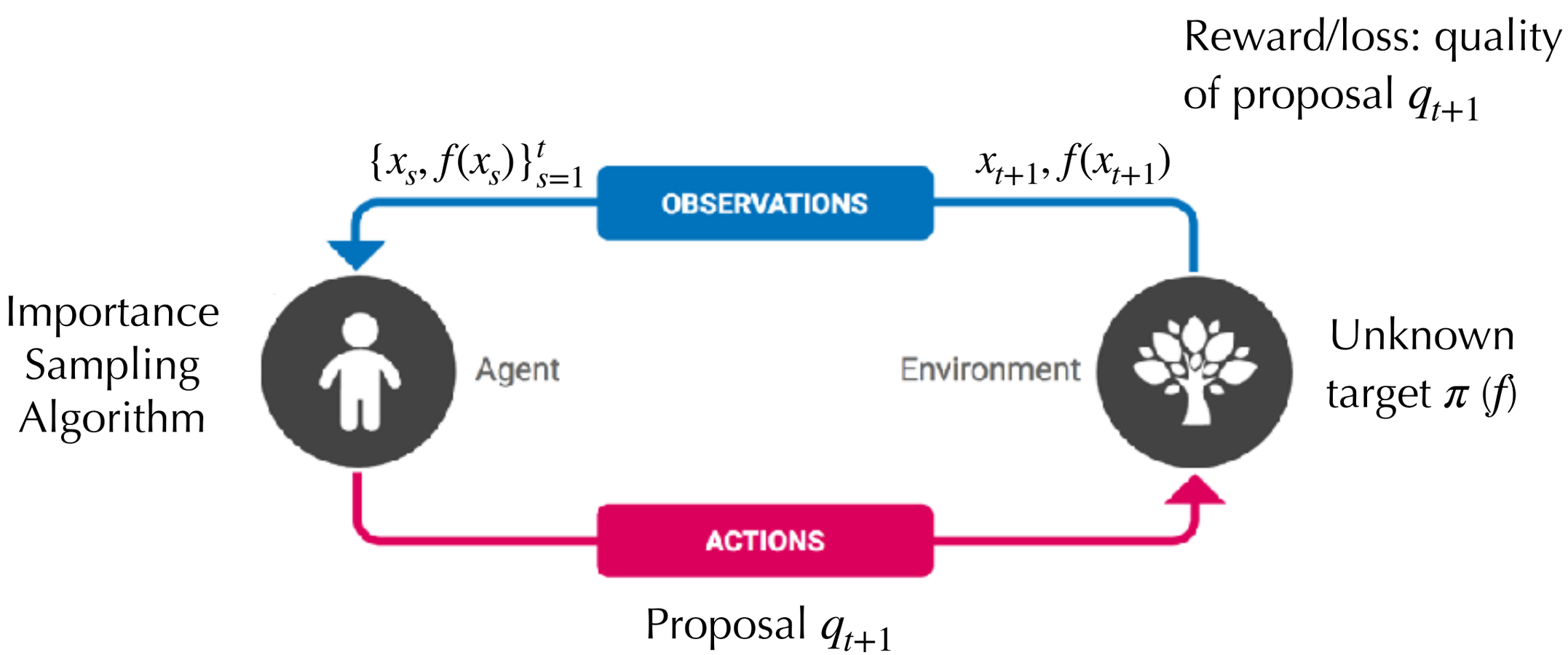
- Exploration: find out about probabilities.
- Exploitation: play arm with highest expected reward.

- How to balance exploration and exploitation optimally.



[Thompson 1933, Lai & Robbins 1985, Berry & Fristedt 1985, Auer et al 2002]

Adaptive Importance Sampling as Online Learning



- Exploration: put proposal mass where we do not know $f(x)$ well.
- Exploitation: put proposal mass where we think $f(x)$ is high.
- Can we design Adaptive Monte Carlo algorithms based on online learning insights? An initial step…

Setup

- Partition sample space into K subsets $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \cdots \cup \mathcal{X}_K$.
- Target distribution $\pi(x) = f(x)/Z$ can be written as

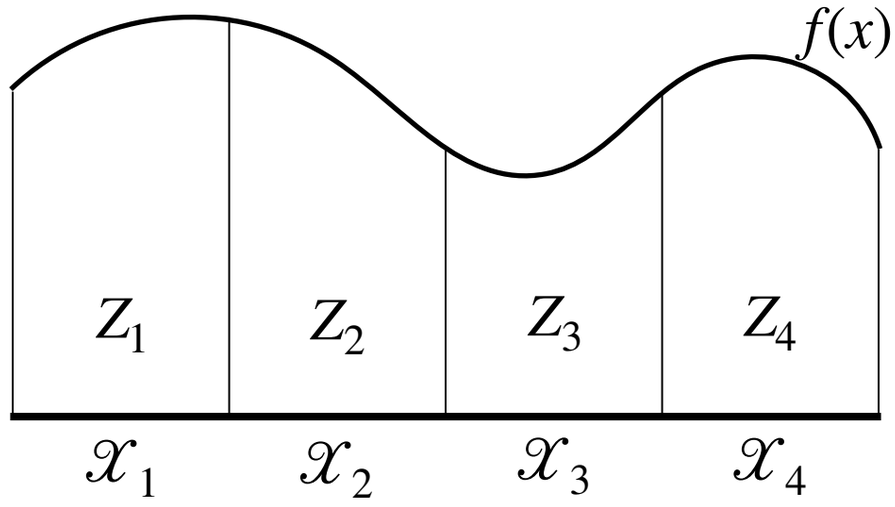
$$\pi(x) = \sum_{a=1}^{K} \pi_a p_a(x)$$

$$\pi_a = \frac{Z_a}{Z} = \frac{Z_a}{\sum_{b=1}^{K} Z_b}$$

$$p_a(x) = \frac{f(x)\mathbb{I}(x \in \mathcal{X}_a)}{Z_a}$$

$$Z_a = \int_{\mathcal{X}_a} f(x)dx$$

$$Z = \int_{\mathcal{X}} f(x)dx$$



Proposal Distribution

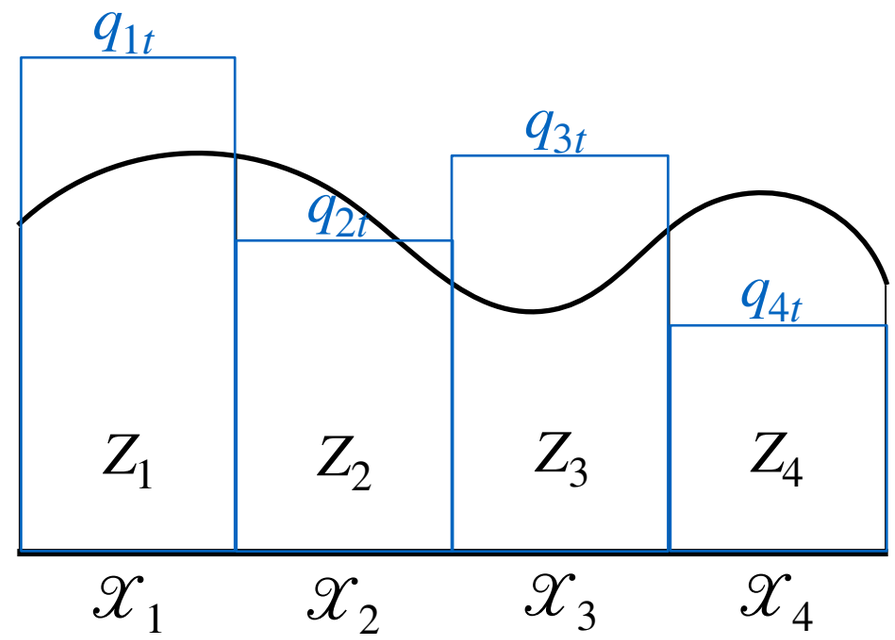
$$\pi(x) = \sum_{a=1}^{K} \pi_a p_a(x)$$

- Family of proposal distributions:

$$q_t(x) = \sum_{a=1}^{K} q_{at} g_a(x)$$

  - Parameterised by $(q_{1t}, \ldots, q_{Kt}), q_{at} \geq 0, \sum_{a=1}^{K} q_{at} = 1.$
  - $g_a(x)$ is a fixed distribution over $\mathcal{X}_a$.

- At iteration t, propose $x_t \sim q_t$:
  - Draw $A_t \sim \mathsf{Discrete}(q_{1t}, \ldots, q_{Kt})$.
  - Draw $x_t \sim g_{A_t}$.



Efficacy and Regret

- Use $\mathsf{KL}(\pi \| q_t)$ to measure (in)efficacy of IS with proposal q_t.

$$\mathsf{KL}(\pi \| q_t) = \sum_{a=1}^{K} \pi_a \log \frac{\pi_a}{q_{at}} + \sum_{a=1}^{K} \pi_a \mathsf{KL}(p_a \| g_a)$$

- Best proposal in class has $q_a^* = \pi_a$. Relative to this, regret of q_t is:

$$\mathsf{Regret}_t = \mathsf{KL}(\pi \| q_t) - \mathsf{KL}(\pi \| q^*) = \sum_{a=1}^{K} \pi_a \log \frac{\pi_a}{q_{at}}$$

- How to adapt/learn q_{at}'s to (provably) minimise regret?

Adapting/Learning the Proposal Distribution

- How to adapt/learn q_{at}'s to minimise (provably) regret?
- Optimal proposal has probabilities

$$q_a^* = \pi_a = \frac{Z_a}{\sum_{b=1}^{K} Z_b}$$

- We can estimate each Z_a using previous importance samples.
- Define:

$$N_{at} = \#\{s \ : \ s \leq t, A_s = a\} \qquad Y_{as} = \frac{f(x_s)}{g_a(x_s)} \text{ if } x_s \in \mathcal{X}_a$$

- Y_{as} is an unbiased estimate of Z_a: $\mathbb{E}_{g_a}[f(x)/g_a(x)] = Z_a$.
- Estimate:

$$Z_a \approx \hat{Z}_{at} = \frac{1}{N_{at}} \sum_{s \leq t : A_s = a} Y_{as}$$

Daisee: aDAptive IS with Exploration-Exploitation

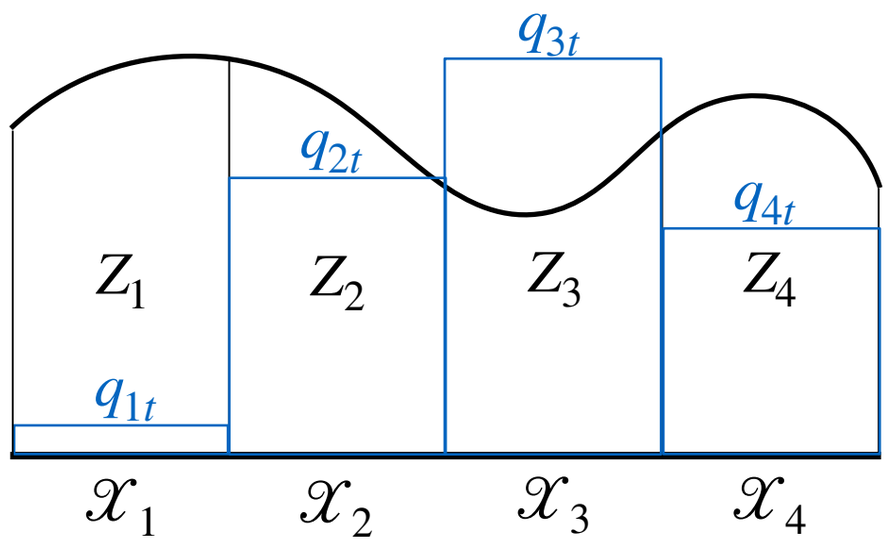
- Can we use the following adaptation to proposal?

$$Z_a \approx \hat{Z}_{at} = \frac{1}{N_{at}} \sum_{s \le t : A_s = a} Y_{as} \qquad\qquad q_{a,t+1} = \frac{\hat{Z}_{at}}{\sum_{b=1}^{K} \hat{Z}_{bt}}$$

- Issues:
  - $\hat{Z}_{at}$ is in fact a biased estimate of Z_a.
  - Cannot guarantee that q_{at} converges to π_a due to under-exploration.

- Fix: add an "optimism boost"

$$q_{a,t+1} = \frac{\hat{Z}_{at} + \sigma_{at}}{\sum_{b=1}^{K} \hat{Z}_{bt} + \sigma_{bt}}$$



Assumptions and Guarantees

- Assume that each importance weight is sub-Gaussian

$$\mathbb{E}[e^{\theta(Y_{at}-\mathbb{E}[Y_{at}])}] \leq e^{\frac{1}{2}\tau_a^2\theta^2}$$

- Define optimism boost as

$$\sigma_{at} = c\tau_a\sqrt{\log t/N_{at}}, c = \sqrt{2.07(2+\gamma+2\log_2 e)} \text{ for some } \gamma > 0$$

- Then:

$$q_{at} \to \pi_a \quad \text{almost surely}$$

$$\mathbb{E}[\text{Regret}_t] \in \mathcal{O}(t^{-\frac{1}{2}}(\log t)^{\frac{3}{4}})$$

- Basic approach: show that probabilities of "bad" events are small

$$B_t = \{\forall a, t/2 \leq s \leq t, |\hat{Z}_{as} - Z_a| < \sigma_{as}\}$$

$$C_t = \left\{ \forall a, N_{at} > t\frac{Z_a}{4\sum_b Z_b + 2c\tau_b\sqrt{logt}} \right\}$$

Generalisation to α-divergence Losses

- α-divergence generalises the KL divergence

$$D_\alpha(\pi\|q) = \frac{1}{1-\alpha} + \frac{1}{\alpha} - \frac{1}{\alpha(1-\alpha)} \int_{\mathcal{X}} \left(\frac{\pi(x)}{q(x)}\right)^\alpha q(x)dx$$

$$D_1(\pi\|q) = \mathsf{KL}(\pi\|q)$$
$$D_0(\pi\|q) = \mathsf{KL}(q\|\pi)$$
$$D_2(\pi\|q) = \frac{1}{2}\int_{\mathcal{X}} \left(\frac{\pi(x)}{q(x)} - 1)\right)^2 q(x)dx$$

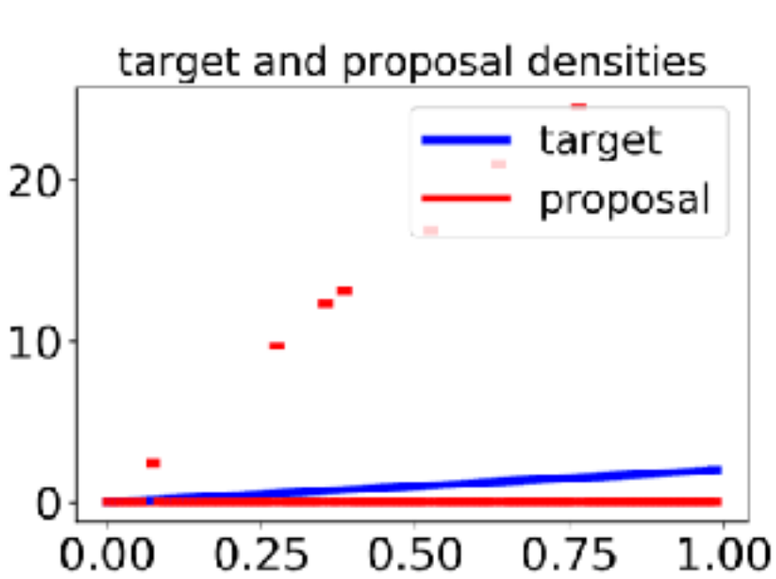
- Optimal proposal now is

$$q_a^* = \frac{(\pi_a^{(\alpha)})^{\frac{1}{\alpha}}}{\sum_{b=1}^K (\pi_b^{(\alpha)})^{\frac{1}{\alpha}}} \qquad \pi_a^{(\alpha)} = \mathbb{E}_{g_a}\left[\left(\frac{f(x)}{g_a(x)}\right)^\alpha\right]$$

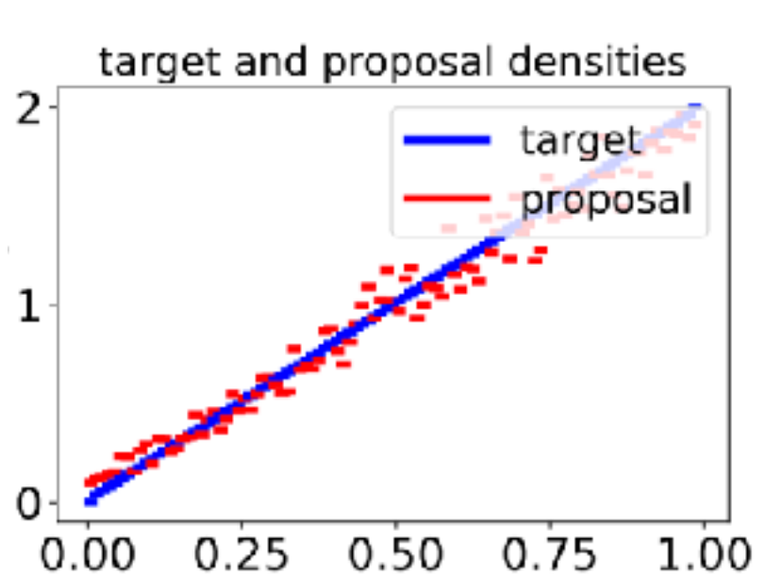
- Can still show that $\mathbb{E}[\mathsf{Regret}_t^{(\alpha)}] = O(t^{-\frac{1}{2}}(\log t)^{\frac{1+2\alpha}{4\alpha}})$.

Experimental Validation

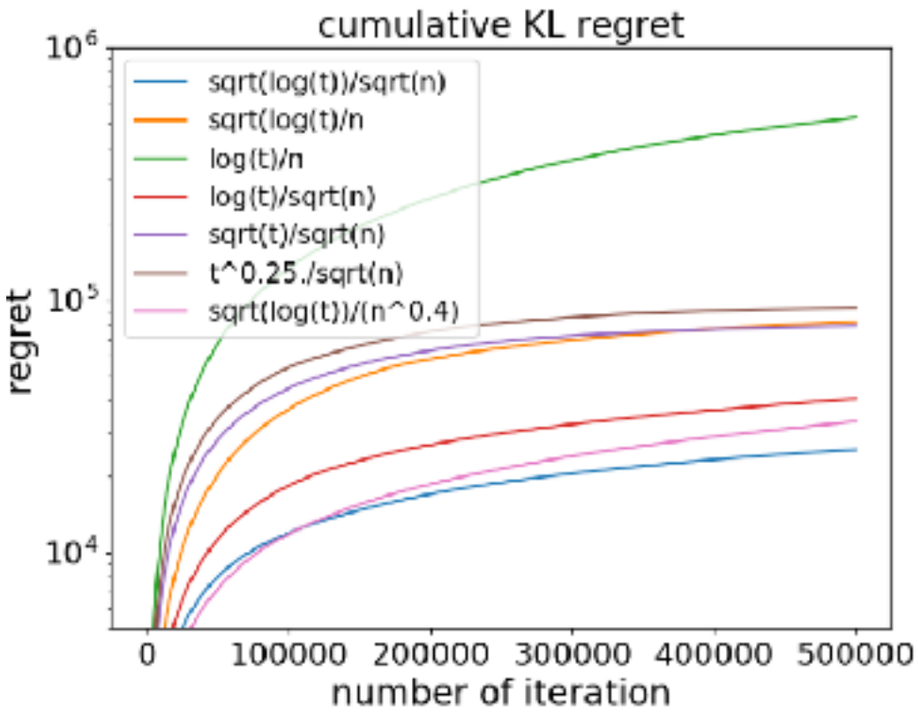
- Sample space [0,1] partitioned into $K = 100$ equal intervals.
- Target and interval proposals g_a chosen so that $Y_{at} \sim 2a\text{Bernoulli}(1/100)/101$.



without optimism boost

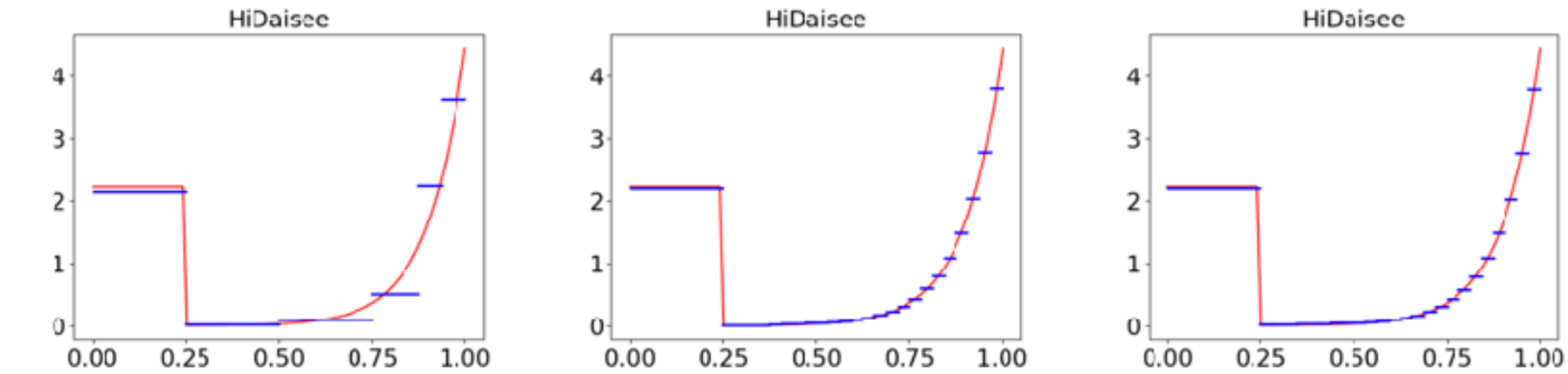


with optimism boost

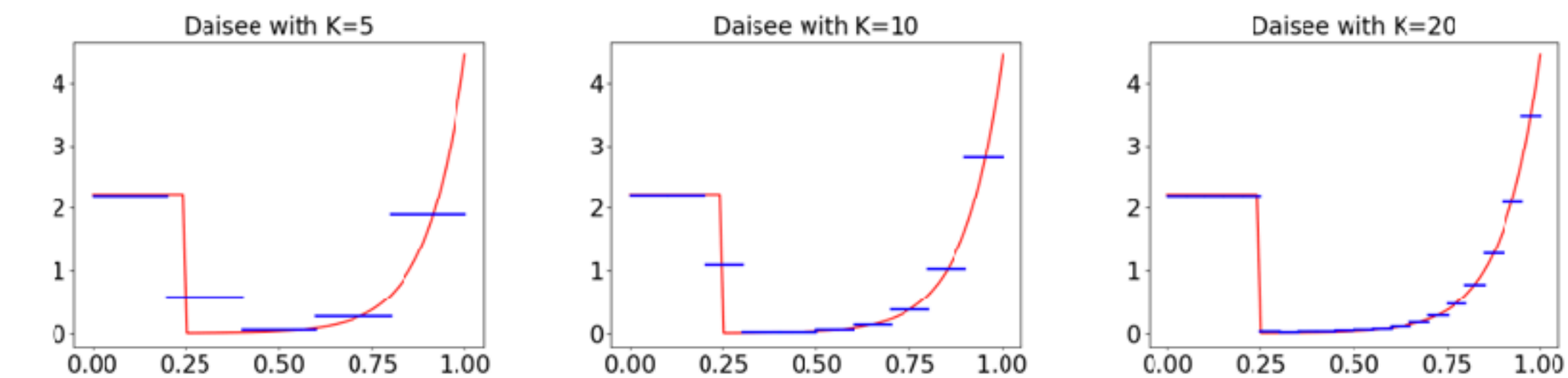


Hierarchical Daisee

- Heuristic extension to Daisee: refine partition by splitting subset if effective sample size too small.

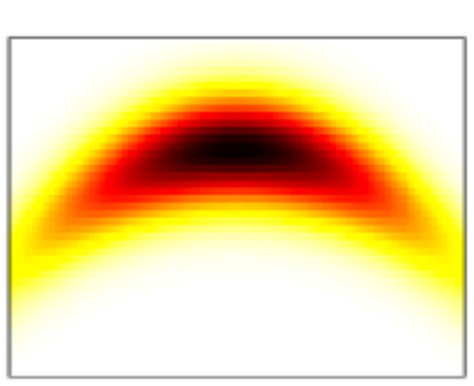


(a) Iteration 10^2 (b) Iteration 10^4 (c) Iteration 10^5

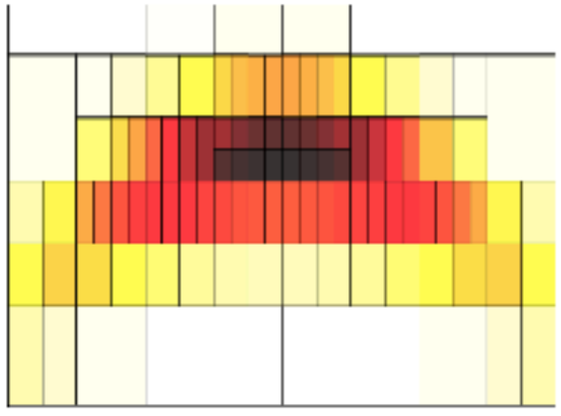


(d) Iteration 10^5 (e) Iteration 10^5 (f) Iteration 10^5

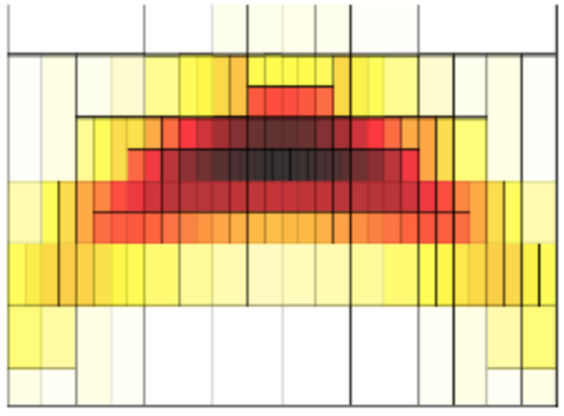
Hierarchical Daisee



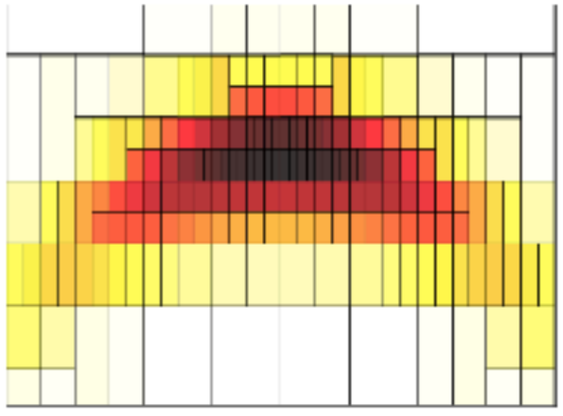
(a) Target



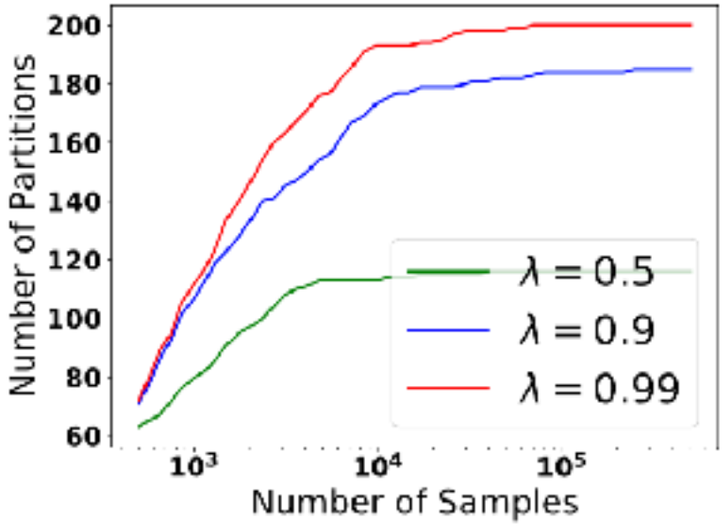
(b) λ = 50%



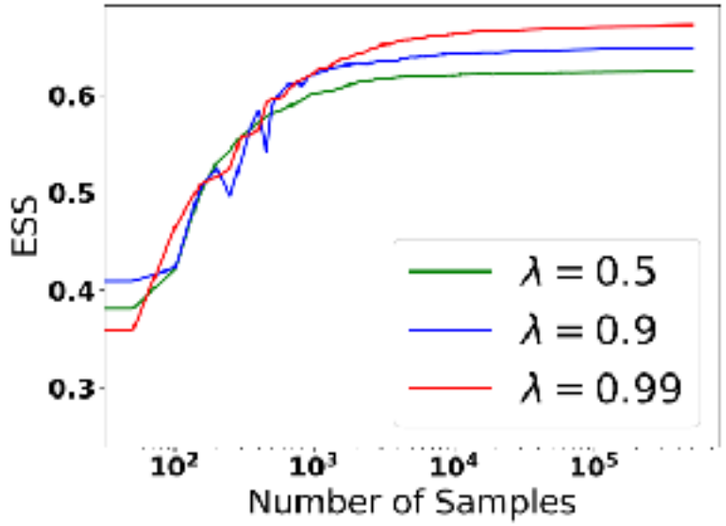
(c) λ = 70%



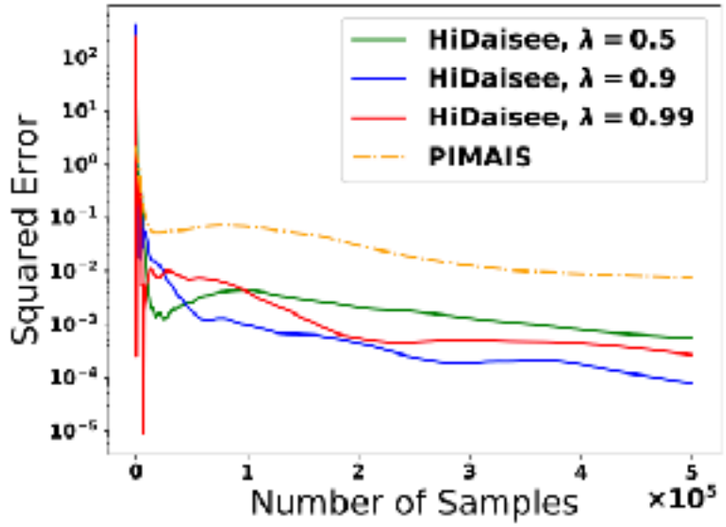
(d) λ = 95%



(e) Number of Partitions



(f) Effective Sample Size



(g) Squared Error

Discussion

- Adaptive Monte Carlo methods can be viewed from the perspective of online learning of the target distribution.
- Daisee: A first step developing Adaptive Importance Sampling methods by borrowing ideas from the online learning.

- Future directions:
  - Tightening analysis, relaxing sub-Gaussian assumption.
  - Analysis and alternatives for hierarchical partitioning AIS.
  - Other adaptive Monte Carlo methods.

Thank you!