Introduction to Cross-Entropy Method (CEM)

Presented by Xi Huang @SIST, ShanghaiTech

(Thanks for the help of Prof. Shao, Xin, and Junge)

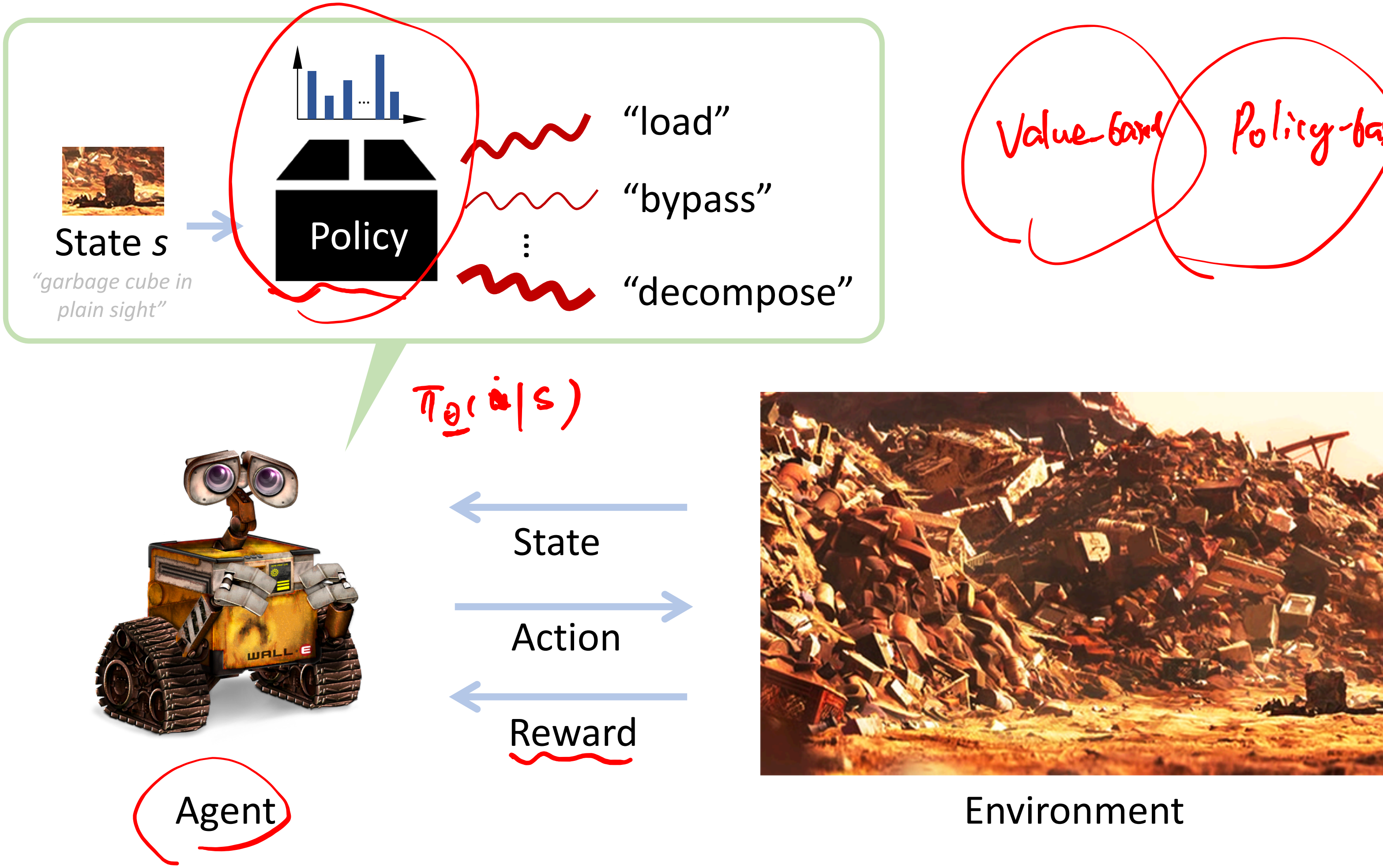
Lecture Flow

- Motivation & Key Idea
- A Close Look at Cross-Entropy Method (CEM)
- CEM for Combinatorial Optimization Problems
- CEM for Policy Optimization in RL
- Closing Remarks

Lecture Flow

- Motivation & Key Idea
- A Close Look at Cross-Entropy Method (CEM)
- CEM for Combinatorial Optimization Problems
- CEM for Policy Optimization in RL
- Closing Remarks

Motivation: Policy Optimization in RL



Value-based

Policy-based

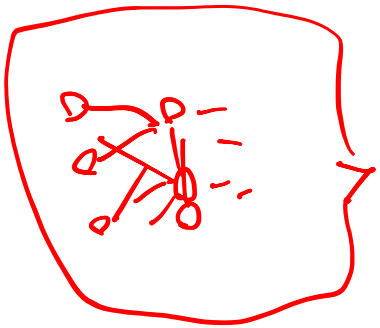
$\pi_{\theta}(a|s)$

Policy Optimization

Simulation

Optimization Objective: $\theta \in \text{argmax}_\theta \boxed{J(\theta) = \mathbb{E}_{\pi_\theta}[\sum_t r(s_t, a_t)]}$ Perf.

Tabu s, a side



$\boxed{\pi_\theta(\cdot\,|S)}$

Function Approximators

- ☆ Decision trees ✓
- ☆ Neural networks ✓ Hot!
- ☆ Linear combinations of features ✓

…

Key idea of PO

collected experiences

Conducting Rollouts

gr... θ

Improving Policy

updated policy

on-policy ✓

off-policy ✓

Policy Gradient

PPO TRPO

- Gradient-based methods
- Derivative-free methods

Concerns about Gradient-based Methods

☆ Costly overheads of gradient computation

- *e.g.*, when applying second-order methods

☆ Convergence

divergence

- Convergence guarantee

Model-free RL

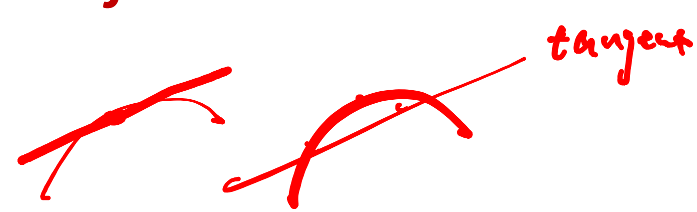
- Long convergence time in many cases
- May get stuck at local optima / saddle point
- Gradient vanishing / exploding problems

☆ Intrinsic non-differentiability of some function approximators

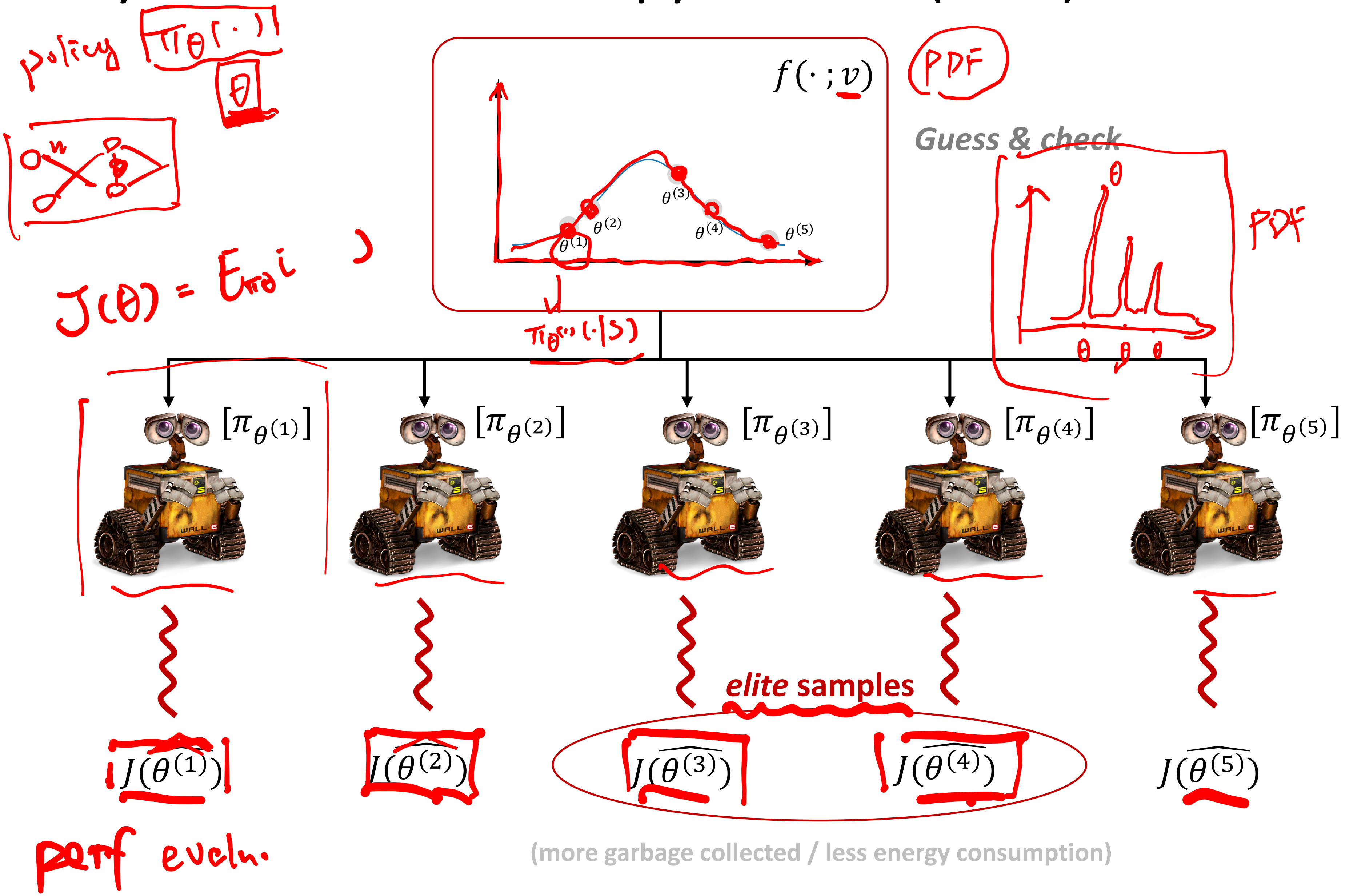
How about derivative-free methods?

CEM

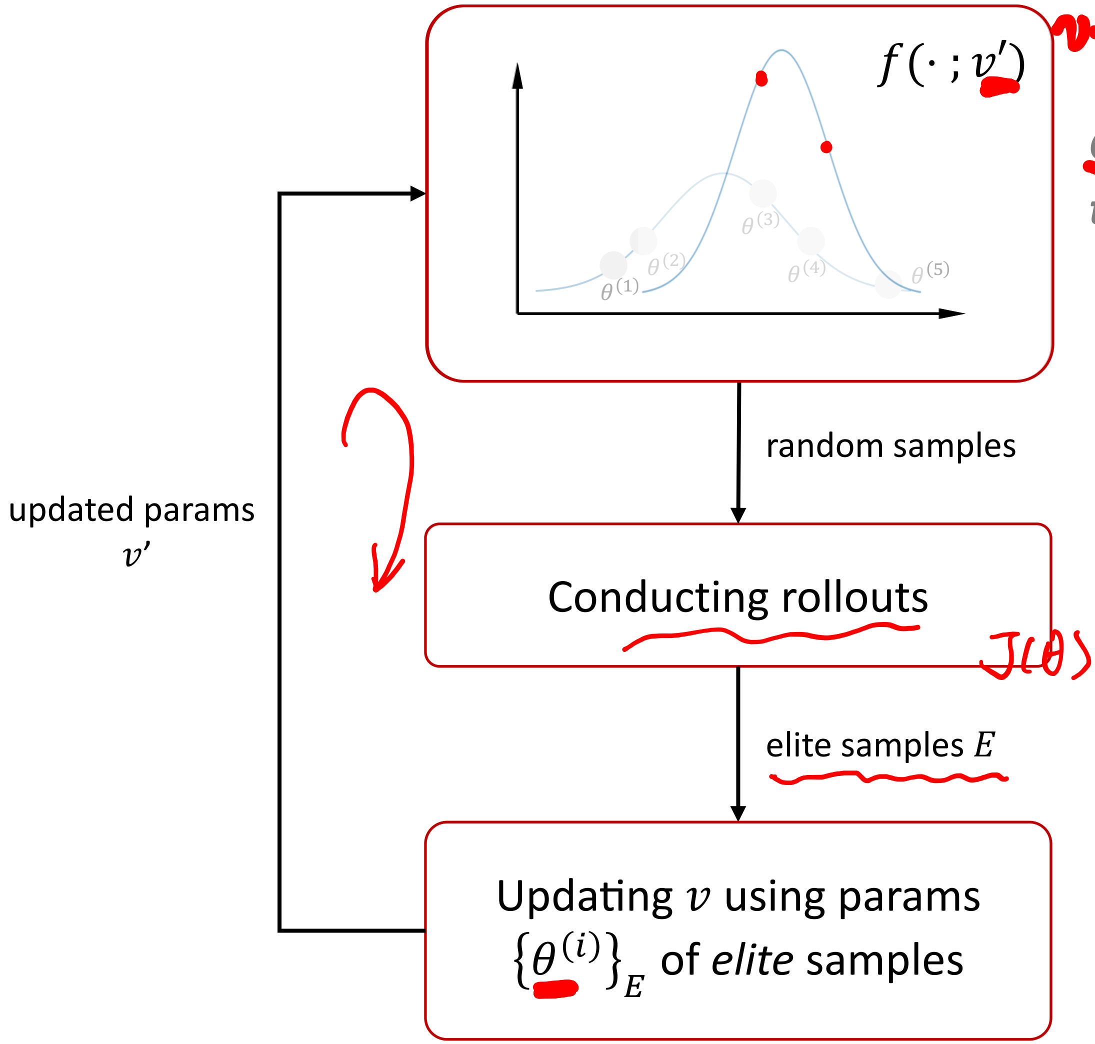
FD



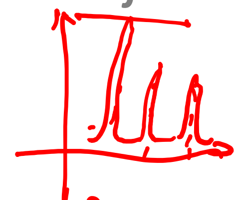
Key Idea of Cross-Entropy Method (CEM)



Key Idea of Cross-Entropy Method (CEM)



Guess & check, then adjust.



PPF

"should"

"can"

$\theta \rightarrow J(\theta)$

$\downarrow$

$\pi_\theta(\cdot|s)$

Lecture Flow

- Motivation & Key Idea
- A Close Look at Cross-Entropy Method (CEM)
- CEM for Combinatorial Optimization Problems
- CEM for Policy Optimization in RL
- Closing Remarks

Origin of CEM

☆ Proposed by Reuven Rubinstein in his paper titled

The Cross-Entropy Method for Combinatorial and Continuous Optimization

- ▶ Israeli scientist (1938 – 2012)
- ▶ Known for his contributions in MC simulation, applied probability, stochastic modeling & stochastic optimization
- ▶ Authored 100+ papers & 6 books

Objective: Rare Event Estimation

e

$Pr(e) = l$

$\downarrow$

$l < 10^{-5}$

☆ Ingredients

$J(\cdot)$

▶ $S(\cdot)$: real-valued function on Θ → performance

▶ $f(\cdot; \boldsymbol{u})$: PDF defined over Θ

☆ Problem

▶ How do we estimate the probability of *rare* event $\{S(X) \geq \boxed{\gamma}\}$?

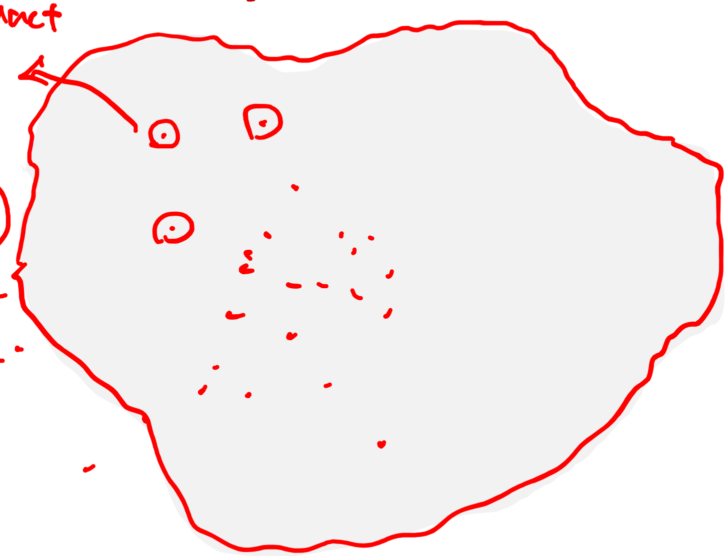
$$l = \mathbb{P}_{\boldsymbol{u}}(S(X) \geq \gamma) = \mathbb{E}_{\boldsymbol{u}}[\mathrm{I}\{S(X) \geq \gamma\}]$$

Why *rare* event?

extract

Θ

sample space.



domain Θ

Estimation via Crude Monte-Carlo Simulation

☆ Estimate l

$$l = \mathbb{P}_{\boldsymbol{u}}(S(X) \geq \gamma) = \mathbb{E}_{\boldsymbol{u}}[\mathrm{I}\{\mathrm{S}(\mathrm{X}) \geq \gamma\}]$$

▶ Monte-Carlo simulation (sample & average)

10^{-5} 10^{8} 10^{9}

Draw random samples $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N$, then

$$\hat{l}_{MC} = \frac{1}{N} \sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \gamma\}$$

Estimation via Importance Sampling

☆ Estimate l

$$l = \mathbb{P}_{\boldsymbol{u}}(S(X) \geq \gamma) = \mathbb{E}_{\boldsymbol{u}}[\mathrm{I}\{\mathrm{S}(\mathrm{X}) \geq \gamma\}]$$

▶ Importance Sampling (IS) via an *auxiliary* density g

PPO

Draw random samples $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_N$ over g, then

$$\boxed{\hat{l}_{IS} = \frac{1}{N} \sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \gamma\} \frac{f(\boldsymbol{X}_i; \boldsymbol{u})}{g(\boldsymbol{X}_i)}}$$

off-policy

estimator

great variance

policy π_θ

policy $\pi_{\theta'}$

sample.

Selection of Density g

☆ Optimal density $g^*(\cdot)$

$$g^*(\boldsymbol{x}) = \frac{I\{S(\boldsymbol{x}) \geq \gamma\} f(\boldsymbol{x}; \boldsymbol{u})}{l}$$

estimate

- *Optimal* in the sense that

> The *variance* of the IS estimator $\hat{l}_{IS}$ is minimized, *i.e.*,
>
> $$g^* \in \mathrm{argmin}_{\mathrm{g}} \mathrm{Var}_{\mathrm{g}}\left[\frac{\mathrm{I}\{\mathrm{S}(\boldsymbol{X}) \geq \gamma\}\mathrm{f}(\boldsymbol{X}; \boldsymbol{u})}{\mathrm{g}(\boldsymbol{X})}\right]$$

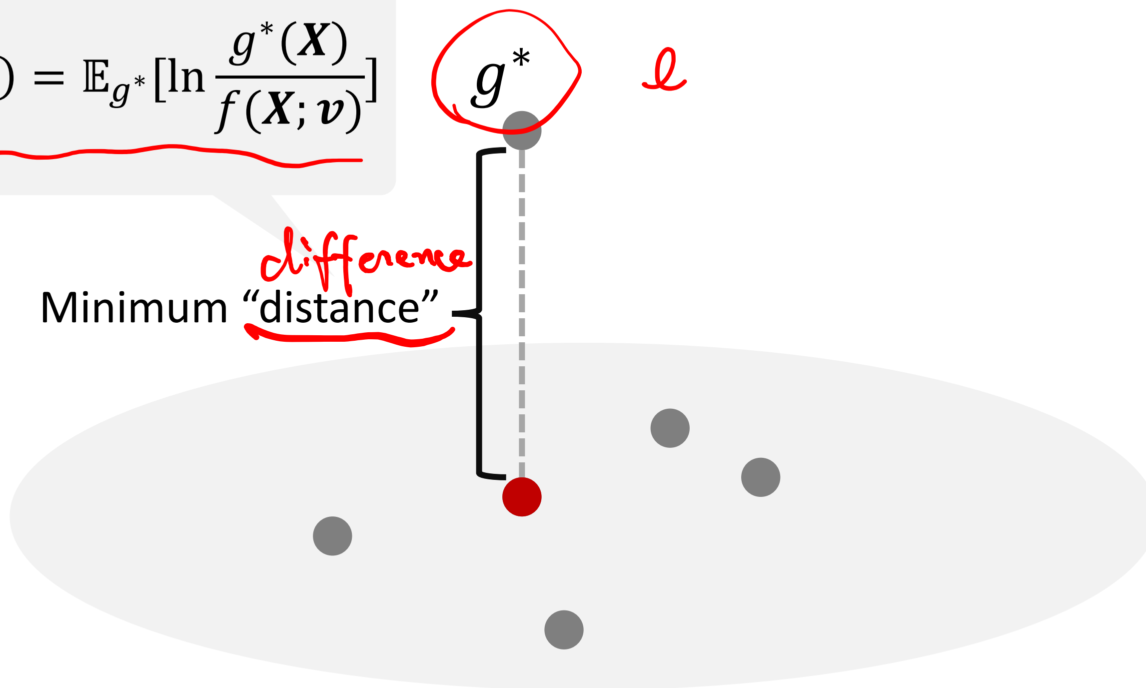
Approximating Density g^*

Kullback-Leibler (KL) divergence

$$D(g^* \,||\, f) = \mathbb{E}_{g^*}[\ln \frac{g^*(\boldsymbol{X})}{f(\boldsymbol{X};\boldsymbol{v})}]$$

difference

Minimum "distance"



Family of densities $\{f(\cdot\,;\boldsymbol{v})\}_{\boldsymbol{v}}$

$f(\cdot\,;\,\cdot)$

Approximating g^* *by* Minimizing *KL* Divergence

CEM

Minimize $D(g^* \,||\, f) = \mathbb{E}_g\left[\ln \frac{g(\boldsymbol{X})}{f(\boldsymbol{X};\boldsymbol{v})}\right]$

g^* f

$$= \underbrace{\boxed{\int g^*(\boldsymbol{x}) \ln g^*(\boldsymbol{x}) d\boldsymbol{x}}}_{\text{constant}} - \underbrace{\boxed{\int g^*(\boldsymbol{x}) \ln f(\boldsymbol{x};\boldsymbol{v}) d\boldsymbol{x}}}_{\text{Cross-entropy}}$$

$$\text{Maximize}_{\boldsymbol{v}\in\Theta} \boxed{\int g^*(\boldsymbol{x}) \ln f(\boldsymbol{x};\boldsymbol{v}) d\boldsymbol{x}}$$

(Minus of cross entropy w.r.t. g^* & f)

$$g^*(\boldsymbol{x}) = \frac{I\{S(\boldsymbol{x}) \geq \gamma\} f(\boldsymbol{x};\boldsymbol{u})}{l}$$

$\mathbb{E}_u$

$$\text{Maximize}_{\boldsymbol{v}\in\Theta} \ \mathbb{E}_{\boldsymbol{u}}[I\{S(\boldsymbol{X}) \geq \gamma\} \ln f(\boldsymbol{X};\boldsymbol{v})]$$

$f(\cdot;u)$

Approximating g^* *by* Minimizing *KL* Divergence

rare

$$\text{Maximize}_{v\in\Theta}\ \mathbb{E}_{\boldsymbol{u}}[I\{S(\boldsymbol{X}) \geq \gamma\} \ln f(\boldsymbol{X};\boldsymbol{v})]$$

loop

Importance sampling (variance)

$$\text{Maximize}_{v\in\Theta}\ \mathbb{E}_{\boldsymbol{w}}[I\{S(\boldsymbol{X}) \geq \gamma\} \frac{f(\boldsymbol{X};\boldsymbol{u})}{f(\boldsymbol{X};\boldsymbol{w})} \ln f(\boldsymbol{X};\boldsymbol{v})]$$

$$W(\boldsymbol{X};\boldsymbol{u},\boldsymbol{w}) = \frac{f(\boldsymbol{X};\boldsymbol{u})}{f(\boldsymbol{X};\boldsymbol{w})}$$

(Likelihood ratio)

$$\boldsymbol{v}^* \in \text{argmax}_{v}\ \mathbb{E}_{\boldsymbol{w}}[I\{S(\boldsymbol{X}) \geq \gamma\} W(\boldsymbol{X};\boldsymbol{u},\boldsymbol{w}) \ln f(\boldsymbol{X};\boldsymbol{v})]$$

$g^* \sim f(\cdot\,; v^*)$

Approximating g^* *by* Minimizing *KL* Divergence

☆ Finding $\boldsymbol{v}^*$

$$\boldsymbol{v}^* \in \operatorname{argmax}_{\boldsymbol{v}} \mathbb{E}_{\boldsymbol{w}}[I\{S(\boldsymbol{X}) \geq \gamma\} W(\boldsymbol{X}; \boldsymbol{u}, \boldsymbol{w}) \ln f(\boldsymbol{X}; \boldsymbol{v})]$$

Simulation for approximation

$$\boldsymbol{v}^* \in \operatorname{argmax}_{\boldsymbol{v}} \frac{1}{N} \sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \gamma\} W(\boldsymbol{X}_i; \boldsymbol{u}, \boldsymbol{w}) \ln f(\boldsymbol{X}_i; \boldsymbol{v})$$

~~Interesting connection to *MLE*~~

Solving *w.r.t.* $\boldsymbol{v}$

$$\frac{1}{N} \sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \gamma\} W(\boldsymbol{X}_i; \boldsymbol{u}, \boldsymbol{w}) \nabla \ln f(\boldsymbol{X}_i; \boldsymbol{v}) = \boldsymbol{0}$$

Are We Done?

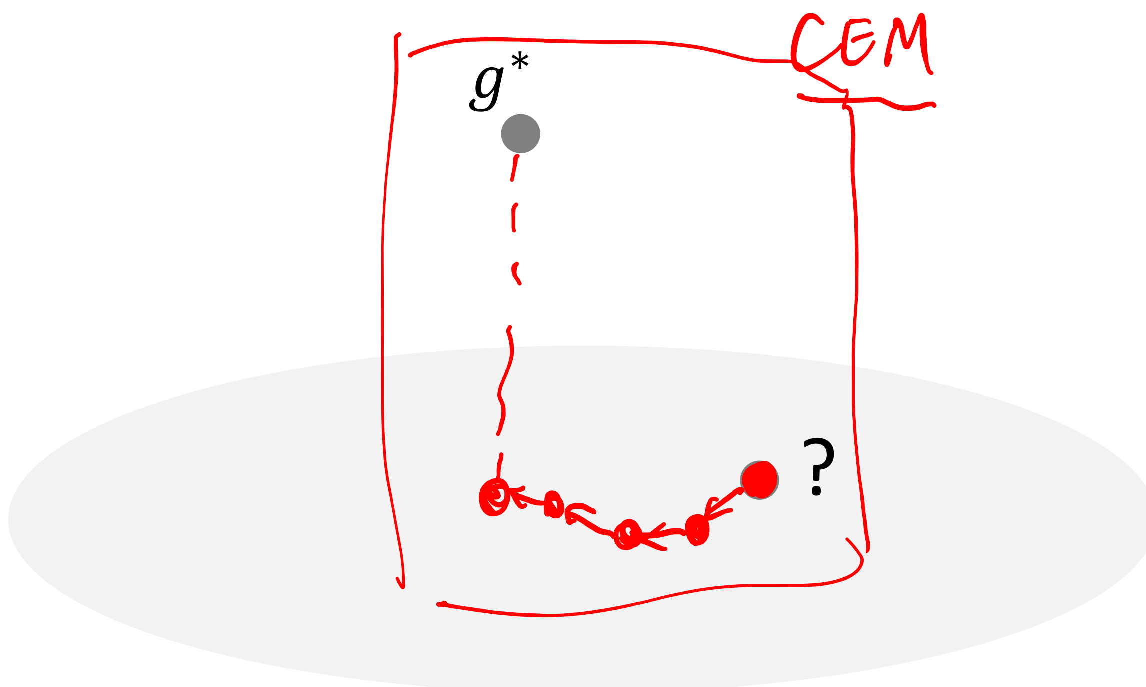
f(·; u)

$\mathbb{E}_u$

☆ Sampling over $f(\boldsymbol{X}; \boldsymbol{w})$

f(·; w)

$$\boldsymbol{v}^* \in \text{argmax}_{\boldsymbol{v}} \frac{1}{N}\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \gamma\} W(\boldsymbol{X}_i; \boldsymbol{u}, \boxed{\boldsymbol{w}}) \ln f(\boldsymbol{X}_i; \boldsymbol{v})$$



Are We Done?

☆ Sampling over $f(\boldsymbol{X}; \boldsymbol{w})$

rareness metric

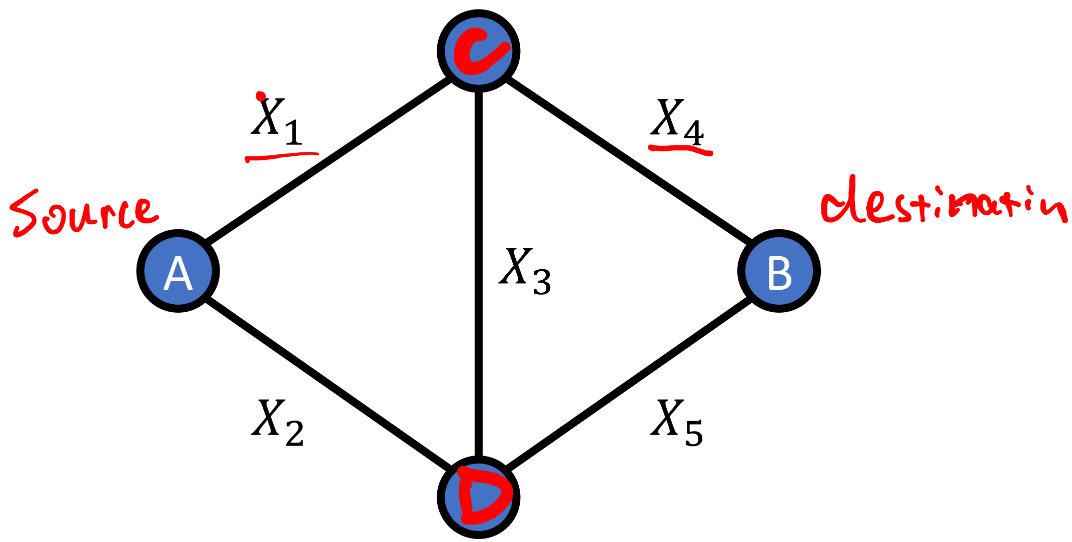
$$\boldsymbol{v}^* \in \operatorname{argmax}_{\boldsymbol{v}} \frac{1}{N} \sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \boxed{\gamma}\} W(\boldsymbol{X}_i; \boldsymbol{u}, \boldsymbol{w}) \ln f(\boldsymbol{X}_i; \boldsymbol{v})$$

☆ Key idea of CEM $f(\cdot\,; v)$

1. Draw random guess, keep the *elite* (samples of interest) guesses
2. Adjust the threshold for rareness (not directly γ) $\boxed{\text{elite}} \rightarrow \boxed{\gamma_t} \rightarrow \gamma$
3. Update param $\boldsymbol{v}$ to make the *elite* guesses more likely to be drawn
4. Repeat $v' \leftarrow$ update v. $\boxed{f(\cdot\,; v')}$ $f(\cdot\,; v^*)$

$\boxed{f(\cdot\,; v^*)} \rightarrow \hat{\ell}$

A Walk-through Example



Shortest path: $S(\boldsymbol{X}) = \min(X_1 + X_4,\ \ X_2 + X_5,\ \ X_1 + X_3 + X_5,\ X_2 + X_3 + X_4)$

$f(\cdot\,; u)$

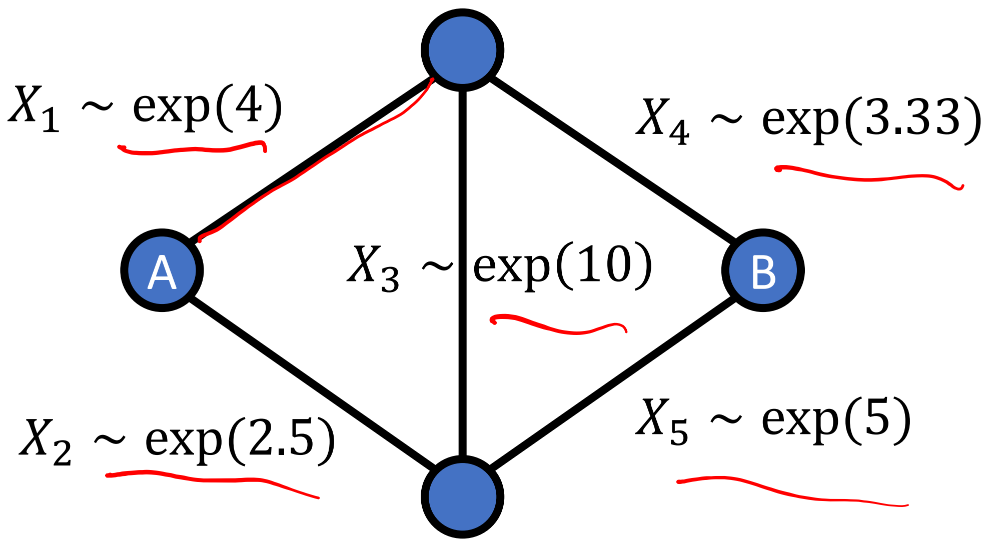
- Random weights: $X_i \sim \exp(u_i)$; $\boldsymbol{X} = (X_1, \cdots, X_5)$; $\boldsymbol{u} = (u_1, \cdots u_5)$

☆ Goal: estimating the probability of rare event $\{S(\boldsymbol{X}) \geq \gamma\}$

$$\ell = \mathbb{P}_{\boldsymbol{u}}(S(\boldsymbol{X}) \geq \gamma) = \mathbb{E}_{\boldsymbol{u}}[I\{S(\boldsymbol{X}) \geq \gamma\}]$$

Estimating Probability of Event $\{S(\boldsymbol{X}) \geq 2\}$

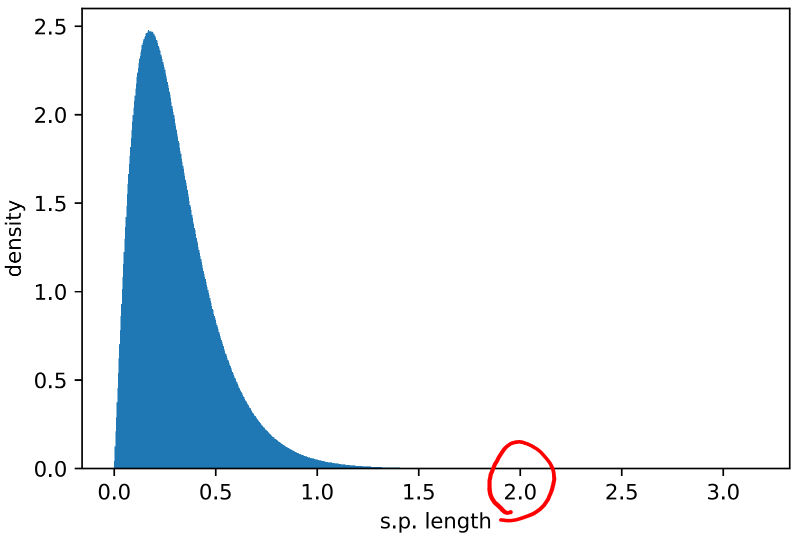
How rare? γ



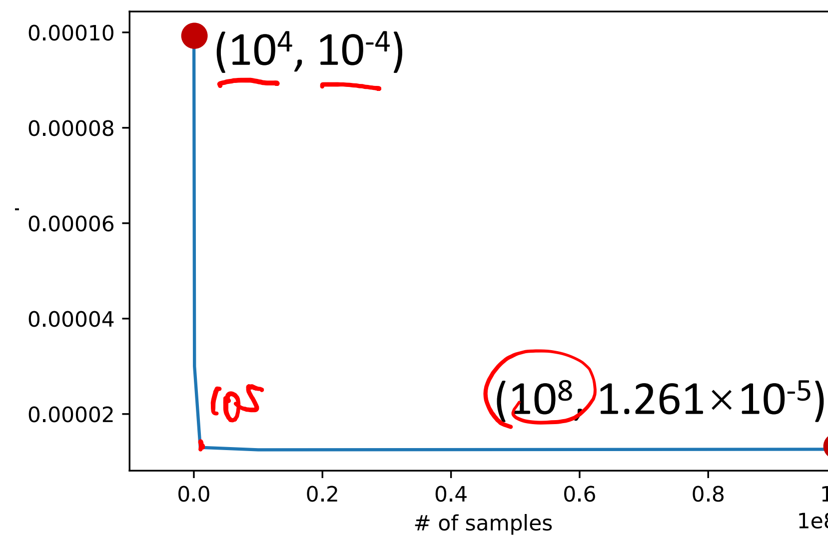
Means: $(0.25, 0.4, 0.1, 0.3, 0.2)$

$X_1 \quad \cdots \quad X_5$

Estimation via Crude Monte Carlo Simulation



MC w/ 10^8 samples



# of samples vs. estimated prob.

Estimation via CEM

Basic Settings

10%

Elite proportion $\boldsymbol{\rho}$	0.1
# of samples per iteration $\mathbf{N}$	1000
Pdf parameter $\widehat{\boldsymbol{v}}_0$	$\boldsymbol{\mu} = (4, 2.5, 10, 3.33, 5)$

elite size : 100

$f(\cdot\,; (4, 2.5, 10, 3.33, 5)) \Rightarrow f(\cdot\,; \mu)$

Estimation via CEM (One-round Walk-through)

Basic Settings

Elite proportion $\boldsymbol{\rho}$	0.1
# of samples per iteration $\mathbf{N}$	1000
Pdf parameter $\widehat{\boldsymbol{v}}_0$	$\boldsymbol{\mu} = (4, 2.5, 10, 3.33, 5)$

[Round-1] Step 1: sampling ("guess")

X_1 X_2 X_3 X_4 X_5

Sample 1	(0.193, 0.597, 0.042, 0.044, 0.239)
Sample 2	(0.164, 1.236, 0.107, 0.574, 0.163)
…	… …
Sample N	(0.124, 0.065, 0.091, 0.180, 0.224)

Estimation via CEM (One-round Walk-through)

Basic Settings

Elite proportion $\boldsymbol{\rho}$	0.1
# of samples per iteration $\mathbf{N}$	1000
Pdf parameter $\widehat{\boldsymbol{v}}_0$	$\boldsymbol{\mu} = (4, 2.5, 10, 3.33, 5)$

[Round-1] Step 2: evaluation ("check")

	Shortest Path Length
Sample 1	0.237
Sample 2	0.434
…	… …
Sample N	0.289

Estimation via CEM (One-round Walk-through)

$\{S(x) \geq \gamma_i\}$

Basic Settings

Elite proportion $\boldsymbol{\rho}$	0.1
# of samples per iteration $\mathbf{N}$	1000
Pdf parameter $\widehat{\boldsymbol{v}}_0$	$\boldsymbol{\mu} = (4, 2.5, 10, 3.33, 5)$

[Round-1] Step 3: elite picking ("filter")

- Take the top $\boxed{\mathbf{100}}$ with the longest shortest path
- Set the rareness threshold γ_1 as the shortest among elites

$\{S(x) \geq \gamma\}$

an estimate of sample quality in terms of *rareness*

$\{S(x) \geq \gamma_1\}$

$\boxed{\gamma_1 = S_{(100)}}$

Estimation via CEM (One-round Walk-through)

$\boxed{v = u.}$

[Round-1] Step 4: update pdf param $\boldsymbol{v_t}$ ("adjust")

$f(\cdot\,; v)$

param of pdf *w.r.t.* edge X_j

$$\boxed{\hat{v}_{t,j} = \frac{\sum_{i=1}^{N} W(\boldsymbol{X}_i; \boldsymbol{u}, \hat{\boldsymbol{v}}_{t-1}) X_{ij}}{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\} W(\boldsymbol{X}_i; \boldsymbol{u}, \hat{\boldsymbol{v}}_{t-1})},} \quad j \in \{1, \cdots, 5\}$$

$$\max_{\boldsymbol{v}} \widehat{D}(\boldsymbol{v}) = \max_{\boldsymbol{v}} \frac{1}{N} \sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\} W(\boldsymbol{X}_i; \boldsymbol{u}, \hat{\boldsymbol{v}}_{t-1}) \ln f(\boldsymbol{X}_i; \boldsymbol{v})$$

w

$\hat{v}_{t-1}$

$$f(\boldsymbol{x}; \boldsymbol{v}) = \exp\left(-\sum_{j=1}^{5} \frac{x_j}{v_i}\right) \prod_{j=1}^{5} \frac{1}{v_j}$$

$$\Longrightarrow \quad \frac{\partial}{\partial v_j} \ln f(\boldsymbol{x}; \boldsymbol{v}) = \frac{x_j}{v_j^2} - \frac{1}{v_j}$$

$$\Longrightarrow \quad \frac{\partial}{\partial v_j} \widehat{D}(\boldsymbol{v}) = \frac{1}{N} \sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\} W(\boldsymbol{X}_i; \boldsymbol{u}, \hat{\boldsymbol{v}}_{t-1}) \left(\frac{X_{ij}}{v_j^2} - \frac{1}{v_j}\right) = 0$$

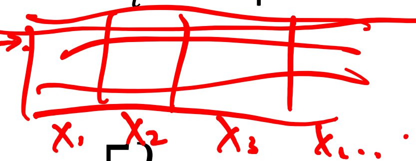
$$\Longrightarrow \quad \boldsymbol{v}_{t,j} = \frac{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\} W(\boldsymbol{X}_i; \boldsymbol{u}, \hat{\boldsymbol{v}}_{t-1}) X_{ij}}{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\} W(\boldsymbol{X}_i; \boldsymbol{u}, \hat{\boldsymbol{v}}_{t-1})}, \quad j \in \{1, \cdots, 5\}$$

More About Param Update

[Round-1] Step 4: update pdf param $\boldsymbol{v}_t$ ("adjust")

Credit of sampling elite $\boldsymbol{X}_i$ with param $\hat{\boldsymbol{v}}_{t-1}$

i-th →



elite.

$$\boxed{\hat{v}_{t,j}} = \frac{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i)\geq\hat{\gamma}_t\} W(\boldsymbol{X}_i;\boldsymbol{u},\hat{\boldsymbol{v}}_{t-1}) X_{ij}}{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i)\geq\hat{\gamma}_t\} W(\boldsymbol{X}_i;\boldsymbol{u},\hat{\boldsymbol{v}}_{t-1})}, \; j \in \{1,\cdots,5\}$$

Likelihood of sampling elite $\boldsymbol{X}_i$ with param $\hat{\boldsymbol{v}}_{t-1}$

$$W(X_i;u;\hat{v}_{t-1}) = \frac{f(X_i;u)}{f(X_i;\hat{v}_{t-1})}$$

likelihood.

$\boxed{\hat{v}_t}$

$f(\cdot;\hat{v}_t)$

Estimation via CEM (One-round Walk-through)

$\boxed{N \cdot \rho}$ $\rho \gg \ell$

[Round-1] Step 5: whether to continue?

$\hat{\gamma}_t \geq \gamma$? (is the rareness *enough*)

$\boxed{f(\cdot\,; \hat{v}_t)}$

No

Go to 2nd Round

Yes

(the resulting pdf is good enough)

Calculate estimate $\hat{l}$ via IS

$$\boxed{\hat{\ell} = \frac{1}{N_1} \sum_{i=1}^{N_1} I\{S(\boldsymbol{X}_i) \geq \gamma\}\, W(\boldsymbol{X}_i; \boldsymbol{u}, \hat{\boldsymbol{v}}_T)}$$

Procedure of CEM

$f(\cdot;\hat{v}_0)$

Initialize $\widehat{\boldsymbol{v}}_0 = \boldsymbol{u}$

$f(\cdot;u)$

Draw N samples $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N$ over $f(\cdot\,;\widehat{\boldsymbol{v}}_{\boldsymbol{t-1}})$

Conduct rollouts to acquire $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_N$

Sort them in an ascending order $\boldsymbol{S}_{(1)}, \ldots, \boldsymbol{S}_{(N)}$

Set $\hat{\gamma}_t = S_{(\lceil(1-\rho)N\rceil)}$

$\hat{\gamma}_t \geq \gamma$?

No

Update $\widehat{\boldsymbol{v}}_{\boldsymbol{t}}$ as

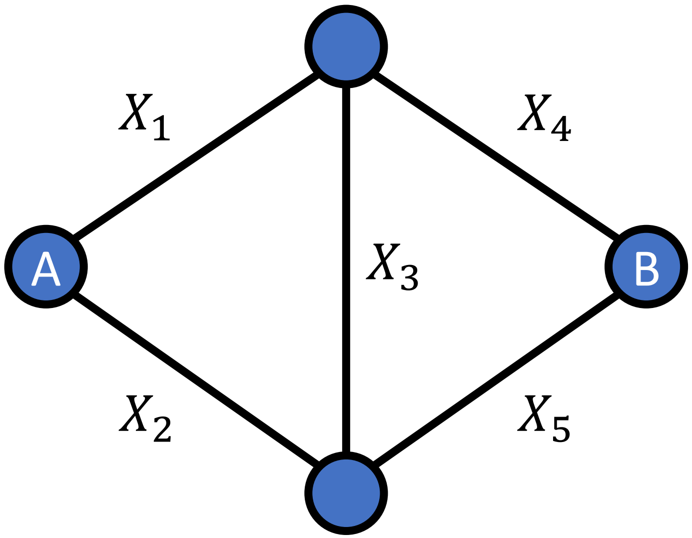
$$\hat{v}_{t,j} = \frac{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i)\geq\hat{\gamma}_t\}\, W(\boldsymbol{X}_i;\boldsymbol{u},\widehat{\boldsymbol{v}}_{t-1}) X_{ij}}{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i)\geq\hat{\gamma}_t\}\, W(\boldsymbol{X}_i;\boldsymbol{u},\widehat{\boldsymbol{v}}_{t-1})},\ j \in \{1,\cdots,5\}$$

Yes

Calculate estimate $\hat{l}$ via IS

$$\hat{\ell} = \frac{1}{N_1}\sum_{i=1}^{N_1} I\{S(\boldsymbol{X}_i) \geq \gamma\}\, W(\boldsymbol{X}_i;\boldsymbol{u},\widehat{\boldsymbol{v}}_T)$$

Approximation via Iterative Procedure



How many steps does it take?

$$\widehat{\boldsymbol{v}}_0 \to \widehat{\boldsymbol{v}}_1 \to \widehat{\boldsymbol{v}}_2 \to \cdots \to \boxed{\widehat{\boldsymbol{v}}_t} \to \cdots \to v^*$$

$$\hat{\gamma}_1 \to \hat{\gamma}_2 \to \cdots \to \boxed{\hat{\gamma}_t} \to \cdots \to \gamma$$

Evaluation ($\gamma = 2$)

$10^7 - 10^8$

Elite proportion $\boldsymbol{\rho}$	0.1
# of samples per iteration $\mathbf{N}$	1000
Pdf parameter $\widehat{\boldsymbol{v}}_0$	$\boldsymbol{\mu} = (4, 2.5, 10, 3.33, 5)$

5×10^3

t	$\widehat{\gamma}_t$	$\widehat{\boldsymbol{v}}_t$				
0		0.250	0.400	0.100	0.300	0.200
1	0.575	0.513	0.718	0.122	0.474	0.335
2	1.032	0.873	1.057	0.120	0.550	0.436
3	1.502	1.221	1.419	0.121	0.707	0.533
4	1.917	1.681	1.803	0.132	0.638	0.523
5	2.000	1.692	1.901	0.129	0.712	0.564

$$\widehat{\boldsymbol{v}}_5 = (1.692, 1.901, 0.129, 0.712, 0.564), \;\; N_1 = 10^5$$

$$\widehat{\ell} = \frac{1}{N_1} \sum_{i=1}^{N_1} I\{S(\boldsymbol{X}_i) \geq \gamma\}\, W(\boldsymbol{X}_i; \boldsymbol{u}, \widehat{\boldsymbol{v}}_5) = \boxed{1.34 \cdot 10^{-5}}$$

General Procedure of CEM

Initialize $\hat{\boldsymbol{v}}_0 = \boldsymbol{u}$

Draw N samples $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N$ over $f(\cdot\,; \hat{\boldsymbol{v}}_{\boldsymbol{t-1}})$

Conduct rollouts to acquire $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_N$

Sort them in an ascending order $\boldsymbol{S}_{(1)}, \ldots, \boldsymbol{S}_{(N)}$

Set $\hat{\gamma}_t = S_{(\lceil (1-\rho)N \rceil)}$

$\hat{\gamma}_t \geq \gamma$?

No

Smoothed update can be adopted

Update $\hat{\boldsymbol{v}}_{\boldsymbol{t}}$ as

$$\hat{w}_t \leftarrow \operatorname{argmax}_{\boldsymbol{v}} \frac{1}{N} \sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \gamma\} W(\boldsymbol{X}_i; \boldsymbol{u}, \hat{\boldsymbol{v}}_{\boldsymbol{t-1}}) \ln f(\boldsymbol{X}_i; \boldsymbol{v})$$

$$\hat{v}_t = \alpha \hat{w}_t + (1-\alpha) \hat{v}_{t-1}$$

Yes

Calculate estimate $\hat{l}$ via IS over $f(\cdot\,; \hat{\boldsymbol{v}}_{\boldsymbol{T}})$

Remarks

☆ CEM enjoys asymptotic convergence properties

☆ Rollout procedure can be parallelized

☆ Often works with fast convergence in practice

☆ Widely adopted in various applications

- Combinatorial optimization
- Stochastic optimization
- Reinforcement learning
- Reliability analysis
- ......

$x_1 \quad S(x_1) \cdots$

$\vdots \quad \vdots$

$x_N \quad S(x_N)$.

CEM baseline

Lecture Flow

- Motivation & Key Idea
- A Close Look at Cross-Entropy Method (CEM)
- CEM for Combinatorial Optimization Problems
- CEM for Policy Optimization in RL
- Closing Remarks

Vector Decoding Problem

$$\boldsymbol{y} = (y_1, y_2, \dots, y_n)$$?

$y_j \in \{0, 1\}$

$x \rightarrow$

$$S(\boldsymbol{x}) = n - \sum_{j=1}^{n} |x_j - y_j|$$

?

$\rightarrow S(\boldsymbol{x})$

Black box

$$\boldsymbol{y} = \arg\max_{\boldsymbol{x}\in\Theta} S(\boldsymbol{x})$$

Set of n-dimensional binary vectors

Combinatorial optimization problem

Vector Decoding Problem

$\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ **?**

$(1, 0, 1)$

$\boldsymbol{x} \longrightarrow$ $S(\boldsymbol{x}) = n - \sum_{j=1}^{n} |x_j - y_j|$ $\longrightarrow S(\boldsymbol{x})$

?

Black box

$\boldsymbol{x}$		$S(x)$
$\cancel{(0, 1, 0)}$	$\dashrightarrow$	0
$(1, 1, 0)$	$\dashrightarrow$	1
$(1, 0, 0)$	$\dashrightarrow$	2
✓ $(1, 0, 1)$	$\dashrightarrow$	3

$\boldsymbol{y} = \arg\max_{x \in \Theta} S(\boldsymbol{x})$

Vector Decoding Problem

$$\boldsymbol{y} = (y_1, y_2, \dots, y_n)$$?

$y_j \in \{0, 1\}$

$\boldsymbol{x}$

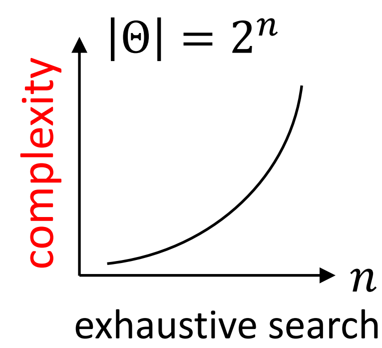
$$S(\boldsymbol{x}) = n - \sum_{j=1}^{n} \left| x_j - y_j \right|$$

?

$S(\boldsymbol{x})$

Black box

$$\boldsymbol{y} = \arg\max_{\boldsymbol{x} \in \Theta} S(\boldsymbol{x})$$



Associated Stochastic Problem

☆ Key idea: sampling ~~instead of~~ CEM ~~searching~~ pdf

$$\boxed{\boldsymbol{y} = (y_1, y_2, \dots, y_n)} \; ?$$

$$y_j \in \{0, 1\}$$

$$\xrightarrow{\boldsymbol{x}} \boxed{S(\boldsymbol{x}) = n - \sum_{j=1}^{n} \left| x_j - \underset{?}{y_j} \right|} \xrightarrow{S(\boldsymbol{x})}$$

Black box

$$\boldsymbol{X} \sim \mathrm{Bern}(\boxed{\boldsymbol{p}})$$

$$\boldsymbol{X} = (X_1, X_2, \dots, X_n)$$

$$\boldsymbol{p} = (p_1, p_2, \dots, p_n)$$

$X_i \sim \mathrm{Bern}(p_i)$

Associated Stochastic Problem

$\boldsymbol{y} = (y_1, y_2, \dots, y_n)$?

Rare Event: $\{S(\boldsymbol{X}) \geq n\}$

$y_j \in \{0, 1\}$

$y_j = 1 \; ; \; 0$

$X_j \sim \text{Bern}(1)$ (0)

$\boldsymbol{x} \longrightarrow \boxed{S(\boldsymbol{x}) = n - \sum_{j=1}^{n} |x_j - y_j|} \longrightarrow S(\boldsymbol{x})$

?

Black box

Rare event: $\boxed{\{S(X) \geq n\}}$

$\boxed{\text{pdf}}$

$\boldsymbol{X} \sim \text{Bern}(\underline{\boldsymbol{p}})$

$\boldsymbol{X} = (X_1, X_2, \dots, X_n)$

$\boldsymbol{p} = (p_1, p_2, \dots, p_n)$

$\boxed{\boldsymbol{p}^* = (y_1, y_2, \dots, y_n)}$

$\boldsymbol{X} \sim \text{Bern}(\underline{\boldsymbol{p}^*}) \Rightarrow \underline{\mathbb{P}_{\boldsymbol{p}^*}(\boldsymbol{X} = \boldsymbol{y}) = 1}$

Decoding via CEM

$\boxed{f(\cdot\,;\underline{u})}$

v

Basic Settings (with $y = (1,0,1,0,1)$**)**

Elite proportion $\boldsymbol{\rho}$	0.1
# of samples per iteration $\mathbf{N}$	50
Pdf parameter $\widehat{\boldsymbol{v}}_0$	$(0.5, 0.5, 0.5, 0.5, 0.5)$

Decoding via CEM (One-round Walk-through)

Basic Settings (with $y = (1,0,1,0,1)$**)**

Elite proportion $\boldsymbol{\rho}$	0.1
# of samples per iteration $\mathbf{N}$	50
Pdf parameter $\widehat{\boldsymbol{v}}_0$	$(0.5, 0.5, 0.5, 0.5, 0.5)$

5

[Round-1] sampling & evaluation ("guess & check")

	x	$S(x)$
Sample 1	(1, 0, 0, 0, 0)	3
Sample 2	(0, 1, 0, 0, 1)	2
…	… …	…
Sample N	(1, 1, 0, 1, 1)	2

Decoding via CEM (One-round Walk-through)

[Round-1] Elite picking ("filter")

Elite 1	(1, 0, 0, 0, 1)	4
Elite 2	(0, 0, 1, 0, 1)	4
Elite 3	(1, 1, 1, 0, 1)	4
Elite 4	(1, 0, 1, 1, 0)	3
Elite 5	(0, 1, 1, 0, 1)	3

Elite List (sorted) with $\mathrm{y} = (1,0,1,0,1)$

$\boxed{\gamma_1 = 3}$

Decoding via CEM (One-round Walk-through)

[Round-1] Update pdf param $\boldsymbol{p}_t$ (“adjust”):

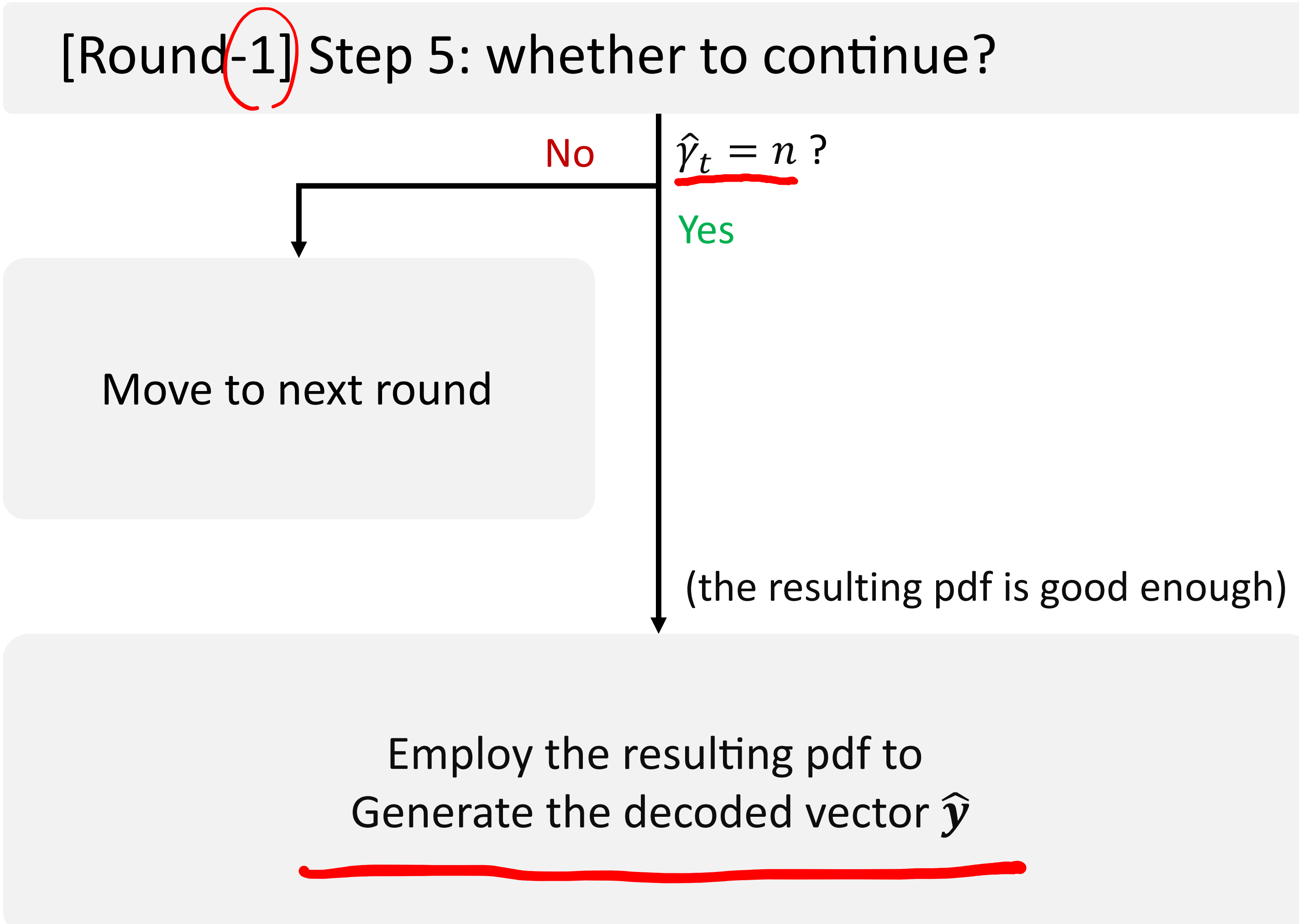
$$\hat{p}_{1,j} = \frac{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\} I\{X_{ij} = 1\}}{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\}}$$

Elite 1	(1, 0, 0, 0, 1)	4
Elite 2	(0, 0, 1, 0, 1)	4
Elite 3	(1, 1, 1, 0, 1)	4
Elite 4	(1, 0, 1, 1, 0)	3
Elite 5	(0, 1, 1, 0, 1)	3

0.5 0.5 0.5 0.5 0.5

$\hat{p}_{1,1}$	$\hat{p}_{1,2}$	$\hat{p}_{1,3}$	$\hat{p}_{1,4}$	$\hat{p}_{1,5}$
0.6	0.4	0.8	0.2	0.8

Decoding via CEM (One-round Walk-through)



Decoding with CEM

$$\boldsymbol{y} = (y_1, y_2, \dots, y_n) \quad ?$$

$$y_j \in \{0, 1\}$$

$$\boldsymbol{x} \longrightarrow \boxed{S(\boldsymbol{x}) = n - \sum_{j=1}^{n} \left| x_j - \underset{?}{y_j} \right|} \longrightarrow S(\boldsymbol{x})$$

Black box

$$\widehat{\boldsymbol{p}}_0 \to \widehat{\boldsymbol{p}}_1 \to \cdots \to \widehat{\boldsymbol{p}}_t \to \cdots \to \boldsymbol{y}$$

$$\hat{\gamma}_1 \to \cdots \to \hat{\gamma}_t \to \cdots \to n$$

Procedure of CEM

Initialize $\widehat{\boldsymbol{p}}_0 = \left(\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}\right)$

Draw N samples $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_N$ over $\mathrm{Bern}(\widehat{\boldsymbol{p}}_{t-1})$

Conduct rollouts to acquire $\boldsymbol{S}_1, \dots, \boldsymbol{S}_N$

Sort them in an ascending order $\boldsymbol{S}_{(1)}, \dots, \boldsymbol{S}_{(N)}$

Set $\hat{\gamma}_t = S_{(\lceil (1-\rho)N \rceil)}$

Stopping criterion is met?

Yes

Output $\widehat{\boldsymbol{p}}_t$

No

Update $\widehat{\boldsymbol{p}}_{\boldsymbol{t}}$ as

$$\hat{p}_{t,j} = \frac{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\} I\{X_{ij} = 1\}}{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\}}, j \in \{1, \dots, n\}$$

Evaluation Result

$n = 10, \boldsymbol{y} = (\mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{0}), N = 50, \rho = 0.1$:

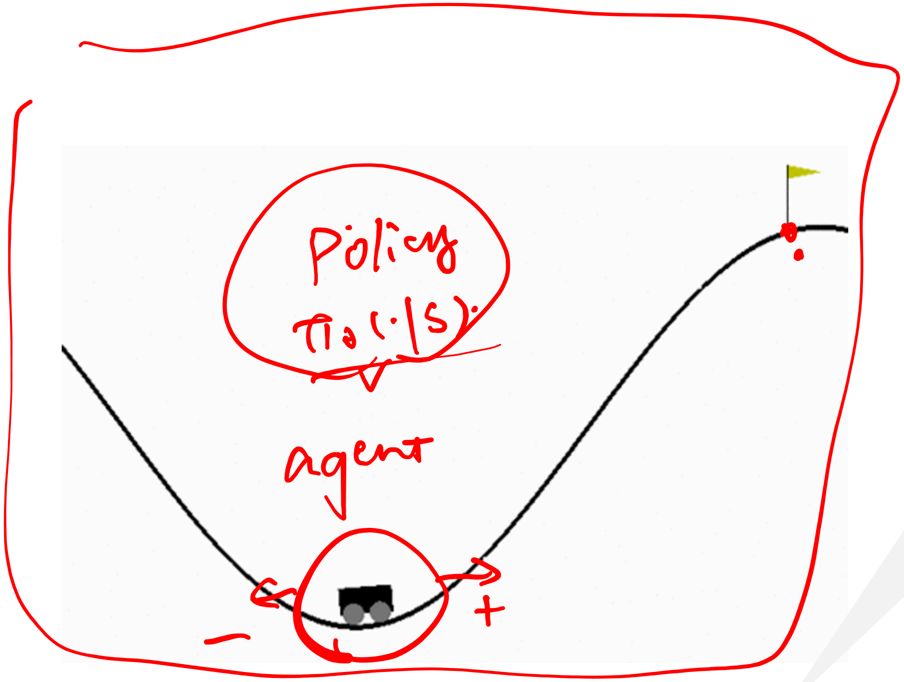
t	$\widehat{\gamma}_t$	$\widehat{\boldsymbol{p}}_t$									
0		0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
1	7	0.60	0.40	0.80	0.40	1.00	0.00	0.20	0.40	0.00	0.00
2	9	0.80	0.80	1.00	0.80	1.00	0.00	0.00	0.40	0.00	0.00
3	10	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00
4	10	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00

Convergence ($\widehat{\gamma}_t \to n, \widehat{\boldsymbol{p}}_t \to \boldsymbol{y}$)

Lecture Flow

- Motivation & Key Idea
- A Close Look at Cross-Entropy Method (CEM)
- CEM for Combinatorial Optimization Problems
- CEM for Policy Optimization in RL
- Closing Remarks

Car Control Problem in RL



Continuous Mountain Car

State: (position, velocity)

height

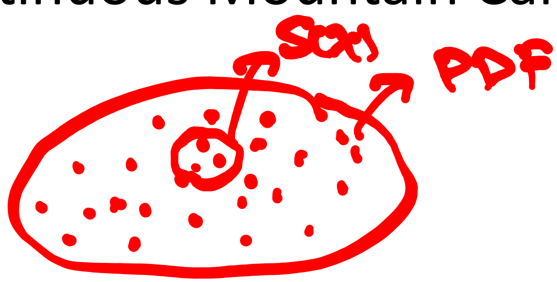
Action: force $f \in$ [-1, 1]

Reward:

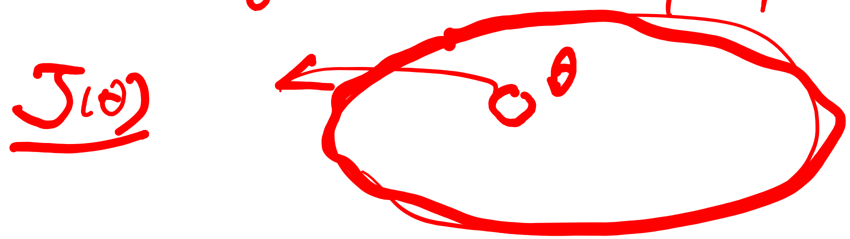
- 100 if win
- $-0.1f^2$ otherwise

Policy approximator:

FNN w/ 2 hid. Layers
with params $\boldsymbol{\theta}$



Policy π_θ



CEM for Policy Optimization

CEM	RL
Domain Θ	$\theta \in \Theta$ Domain of all weights in policy network
Rare event pdf $f(\cdot;\boldsymbol{u})$	Unspecified → PDF
Evaluation function $S(\cdot)$	$J(\cdot) = \mathbb{E}_{\pi}[\sum_t r(s_t, a_t)]$
γ & γ_t	Pre-specified γ (γ_t not used)
ρ	Usually between 0.01 & 0.1 [0.01, 0.1]
# of samples N	# of rollouts to take

Procedure of CEM for Policy Optimization

Initialize arbitrary pdf f w/ params $\widehat{\boldsymbol{\theta}}_{\mathbf{0}}$
($e.g.$, normal distr.)

$f(\cdot\,;u)$

Draw N samples $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N$ over $f(\cdot\,;\widehat{\boldsymbol{\theta}}_{\boldsymbol{t-1}})$

guess

$\arg\max_{\theta} \frac{1}{N}\sum_{i=1}^{N} \ln f(X_i;\theta)$

Conduct rollouts to acquire $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_N$

check

noise

Update $\boxed{\widehat{\boldsymbol{\theta}}_{\boldsymbol{t}}}$ as $W(X_i;u,\theta_{t-1})$

$\text{argmax}_{\boldsymbol{\theta}}\ \frac{1}{N}\sum_{i=1}^{N} I\{\boldsymbol{X}_i \in \mathcal{E}\} \ln f(\boldsymbol{X}_i;\boldsymbol{\theta})$

1. Missing $W(\cdot;\boldsymbol{\theta})$
2. Slightly diff. from MLE

Sort them in an ascending order $\boldsymbol{S}_{(1)}, \ldots, \boldsymbol{S}_{(N)}$

No

Pick the best $\lceil \rho N \rceil$ samples into elite set $\mathcal{E}$

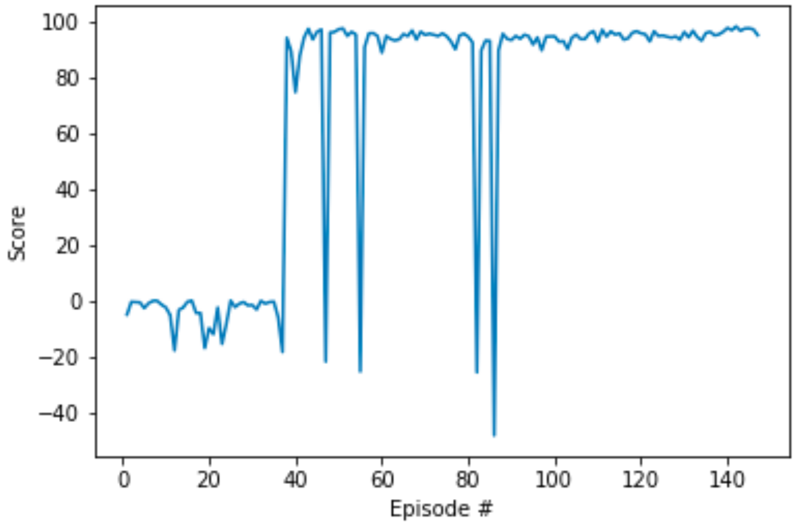
Converged or terminated?

Yes

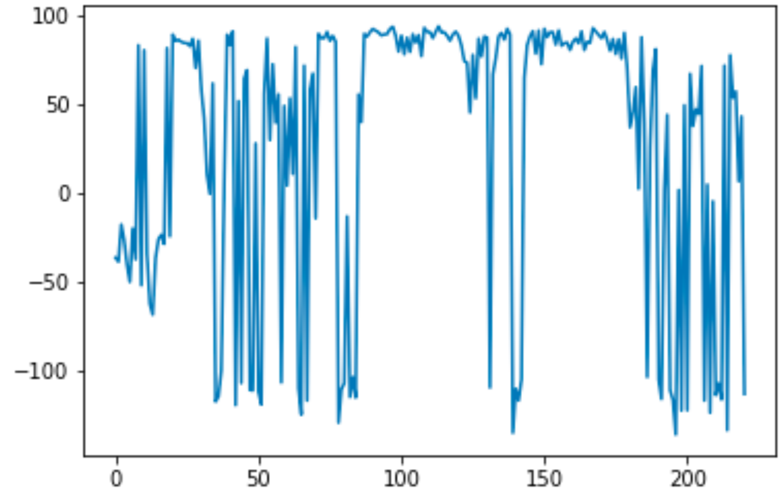
Generate params $\hat{\theta}$ via IS over $f(\cdot\,;\boldsymbol{\theta}_{\boldsymbol{T}})$

$\theta \to \pi_\theta$

Evaluation



CEM

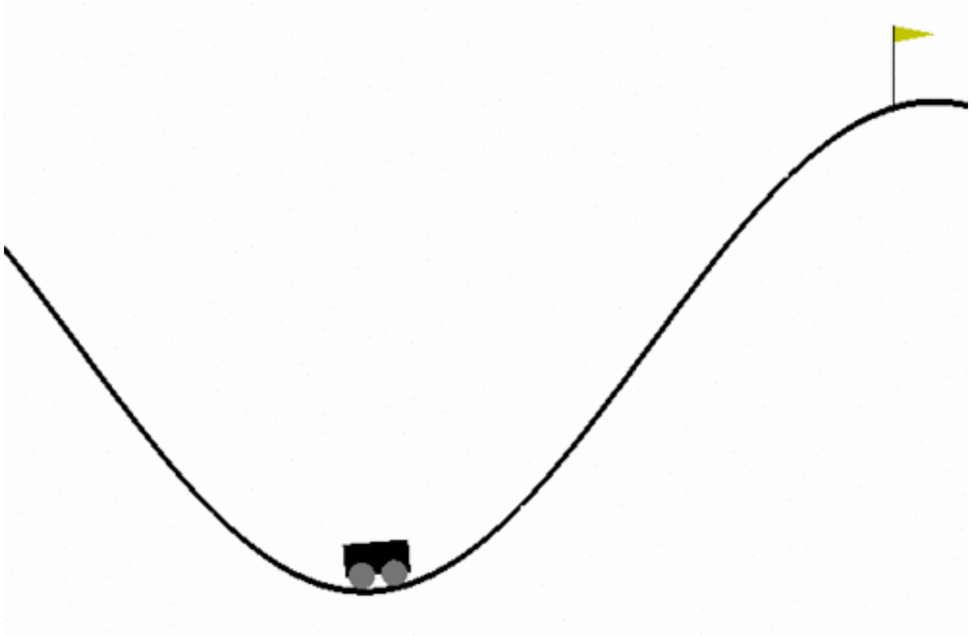


Deep Deterministic PG

DDPG.

P. ..

Resulting Policy in Action

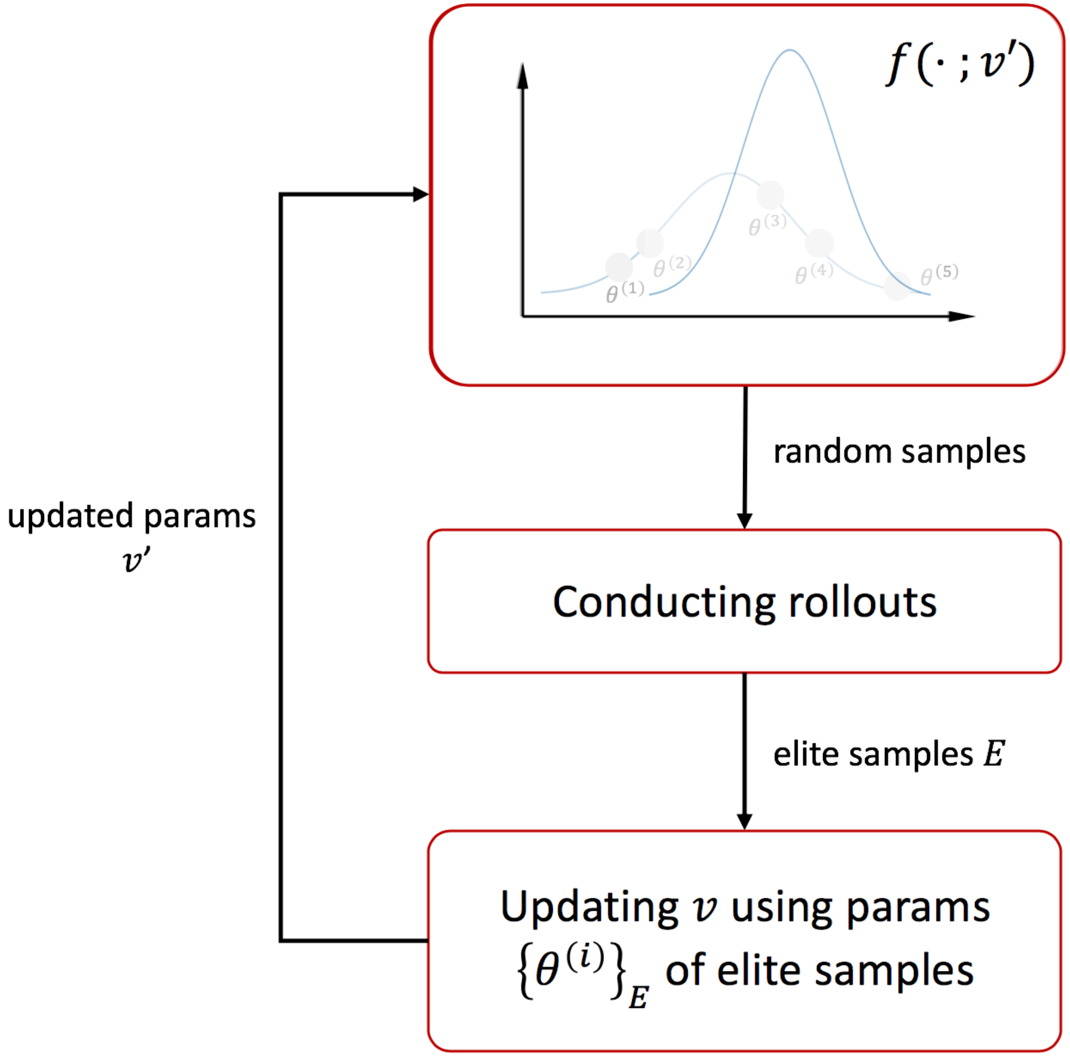


Refer to: https://github.com/udacity/deep-reinforcement-learning/ for more info.

Lecture Flow

- Motivation & Key Idea
- A Close Look at Cross-Entropy Method (CEM)
- CEM for Combinatorial Optimization Problems
- CEM for Policy Optimization in RL
- Closing Remarks

Closing Remarks: Key Idea of CEM



Closing Remarks

$S(x) = \boxed{J(\cancel{x})} \rightarrow e$

Advantages of CEM

☆ Parallelizable

Agent 1 → Nolim
⋮ →
N →

☆ Derivative-free

☆ Stable across rand. seeds

☆ Simple for implementation

☆ Sample efficient owing to self-tuning

Limitations of CEM

☆ Evaluation only based on episodic simulations

☆ Long-horizon (w/ limited samples)

☆ Open-loop planning (may fall short in env. w/ non-stationarity)

$S(x)$

References

☆ R. Y. Rubinstein, *The Cross-entropy Method for Combinatorial and Continuous Optimization*, Methodology and Computing in Applied Probability, vol. 1, no. 2, pp. 127 – 190, 1999.

☆ P.-T. de Boer, D. P. Kroese, S. Mannor, R. Y. Rubinstein, *A Tutorial on the Cross-Entropy Method*, Annals of Operations Research, vol. 134, no. 1, pp. 19 – 67, 2005.

☆ Z. I. Botev, D. P. Kroese, R. Y. Rubinstein, P. L'Ecuyer, *The Cross-Entropy Method for Optimization*, Handbook of Statistics, vol. 31, pp. 35 – 59, 2013.

☆ P. Shvechikov, A. Panin, DRL course on Udacity [with the ~~following GitHub link~~]
https://github.com/udacity/deep-reinforcement-learning/tree/master/cross-entropy

☆ J. Shulman, *Deep Reinforcement Learning via Policy Optimization*, Tutorial at RLDM, 2017.

Thank you!

Any questions?

Estimation for Vector Decoding Problem

$$\widehat{\boldsymbol{p}}_t = \mathrm{argmax}_{\boldsymbol{p}} \frac{1}{N}\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\} \ln f(\boldsymbol{X}_i; \boldsymbol{p})$$

$$\Longrightarrow \hat{p}_{t,j} = \frac{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\} I\{X_{ij} = 1\}}{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\}}$$

$$f(\boldsymbol{x}; \boldsymbol{p}) = \prod_{j=1}^{n} p_j^{x_j} (1 - p_j)^{1-x_j}$$

$$\Longrightarrow \frac{\partial}{\partial p_j} \ln f(\boldsymbol{x}; \boldsymbol{p}) = \frac{x_j}{p_j} - \frac{1 - x_j}{1 - p_j} = \frac{1}{p_j(1 - p_j)}(x_j - p_j)$$

$$\Longrightarrow \frac{\partial}{\partial p_j} \sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\} \ln f(\boldsymbol{X}_i; \boldsymbol{p}) = \frac{1}{p_j(1 - p_j)} \sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\}(X_{ij} - p_j) = 0$$

$$\Longrightarrow p_j = \frac{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\} X_{ij}}{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\}}$$

Procedure of CEM-based Decoding

Algorithm

Initialize $\widehat{\boldsymbol{p}}_0 = \left(\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}\right)$

Loop for each episode $t \geq 1$:

Draw samples $\boldsymbol{X}_1, \dots, \boldsymbol{X}_N \sim \mathrm{Bern}(\widehat{\boldsymbol{p}}_{t-1})$

Calculate $S(\boldsymbol{X}_i)$ for all $i \in \{1, \dots, N\}$ and order them from smallest to biggest as $S_{(1)} \leq \cdots \leq S_{(N)}$

$\widehat{\gamma}_t \leftarrow S_{(\lceil (1-\rho)N \rceil)}$

$\hat{p}_{t,j} \leftarrow \frac{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \widehat{\gamma}_t\} I\{X_{ij} = 1\}}{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \widehat{\gamma}_t\}}$ for all $j \in \{1, \dots, n\}$

$t \leftarrow t + 1$

until $\widehat{\gamma}_t$ does not change for a num. of subsequent iterations (or $\widehat{\boldsymbol{p}}_t$ converges to a binary vector)

Estimation for Shortest Path Problem

$$\hat{v}_{t,j} = \frac{\sum_{i=1}^{N} W(\boldsymbol{X}_i;\boldsymbol{u},\hat{\boldsymbol{v}}_{t-1})X_{ij}}{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i)\geq\hat{\gamma}_t\}W(\boldsymbol{X}_i;\boldsymbol{u},\hat{\boldsymbol{v}}_{t-1})}, \quad j \in \{1,\cdots,5\}$$

$$\max_{\boldsymbol{v}} \widehat{D}(\boldsymbol{v}) = \max_{\boldsymbol{v}} \frac{1}{N}\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\}W(\boldsymbol{X}_i;\boldsymbol{u},\hat{\boldsymbol{v}}_{t-1}) \ln f(\boldsymbol{X}_i;\boldsymbol{v})$$

$$f(\boldsymbol{x};\boldsymbol{v}) = \exp\left(-\sum_{j=1}^{5}\frac{x_j}{v_i}\right)\prod_{j=1}^{5}\frac{1}{v_j}$$

$$\Longrightarrow \frac{\partial}{\partial v_j}\ln f(\boldsymbol{x};\boldsymbol{v}) = \frac{x_j}{v_j^2} - \frac{1}{v_j}$$

$$\Longrightarrow \frac{\partial}{\partial v_j}\widehat{D}(\boldsymbol{v}) = \frac{1}{N}\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i) \geq \hat{\gamma}_t\}W(\boldsymbol{X}_i;\boldsymbol{u},\hat{\boldsymbol{v}}_{t-1})\left(\frac{X_{ij}}{v_j^2} - \frac{1}{v_j}\right) = 0$$

$$\Longrightarrow \boldsymbol{v}_{t,j} = \frac{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i)\geq\hat{\gamma}_t\}W(\boldsymbol{X}_i;\boldsymbol{u},\hat{\boldsymbol{v}}_{t-1})X_{ij}}{\sum_{i=1}^{N} I\{S(\boldsymbol{X}_i)\geq\hat{\gamma}_t\}W(\boldsymbol{X}_i;\boldsymbol{u},\hat{\boldsymbol{v}}_{t-1})}, \quad j \in \{1,\cdots,5\}$$

Procedure of CEM-based Rare Event Simulation

Algorithm

Initialize $\widehat{\boldsymbol{v}}_0 = \boldsymbol{u}$

Loop for each episode $t \geq 1$:

Draw samples $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N \sim f(\boldsymbol{X};\ \widehat{\boldsymbol{v}}_{t-1})$

Calculate $S(\boldsymbol{X}_i)$ for all $i \in \{1, \ldots, N\}$ and order them from smallest to biggest as $S_{(1)} \leq \cdots \leq S_{(N)}$

$\hat{\gamma}_t \leftarrow S_{(\lceil(1-\rho)N\rceil)}$

$$\hat{v}_{t,j} \leftarrow \frac{\sum_{i=1}^{N} I_{\{S(X_i)\geq\hat{\gamma}_t\}} W(\boldsymbol{X}_i;\boldsymbol{u},\widehat{\boldsymbol{v}}_{t-1}) X_{ij}}{\sum_{i=1}^{N} I_{\{S(X_i)\geq\hat{\gamma}_t\}} W(\boldsymbol{X}_i;\boldsymbol{u},\widehat{\boldsymbol{v}}_{t-1})}, \ j \in \{1, \cdots, 5\}$$

$t \leftarrow t + 1$

until $\hat{\gamma}_t \geq \gamma$.

Let T be the final iteration. Draw samples $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N \sim f(\boldsymbol{X};\ \widehat{\boldsymbol{v}}_T)$,

$$\hat{\ell} = \frac{1}{N_1} \sum_{i=1}^{N_1} I_{\{S(\boldsymbol{X}_i)\geq\gamma\}} W(\boldsymbol{X}_i; \boldsymbol{u}, \widehat{\boldsymbol{v}}_T)$$