Certifiable Deep Importance Sampling for Rare-Event Simulation of Black-Box Safety-Critical Systems

Henry Lam

Industrial Engineering & Operations Research

Columbia University

Our Team

Joint work with:

- Mansur Arief (Carnegie Mellon)
- Yuanlu Bai (Columbia)
- Wenhao Ding (Carnegie Mellon)
- Shengyi He (Columbia)
- Zhiyuan Huang (Carnegie Mellon & Tongji University)
- Guru Koushik Senthil Kumar (Carnegie Mellon)
- Ding Zhao (Carnegie Mellon)

Motivation: Autonomous Vehicle Safety Testing

Unsafe deployment of AI-driven physical systems can lead to catastrophic events

Tesla Autopilot Crash, May 2016

Uber Self-Drive Crash, March 2018

Tesla Autopilot Crash, March 2018

Tesla Autopilot Crash, March 2019

Testing autonomous vehicles (AVs) is challenging because:

- Test matrix approaches cannot screen out AVs that excel in the test but not other safety-critical situations (Peng & LeBlanc ‘12)
- Naturalistic testing takes insurmountable time (Zhao et al. ‘15)

Motivation: Autonomous Vehicle Safety Testing

(NHTSA 2013) In US in 2013,

5.7 million police-reported crashes

30,057 fatal crashes

0.53 million miles for one police-reported crash

99 million miles for a fatal crash

38 years for police-reported crash

6,877 years for a fatal crash

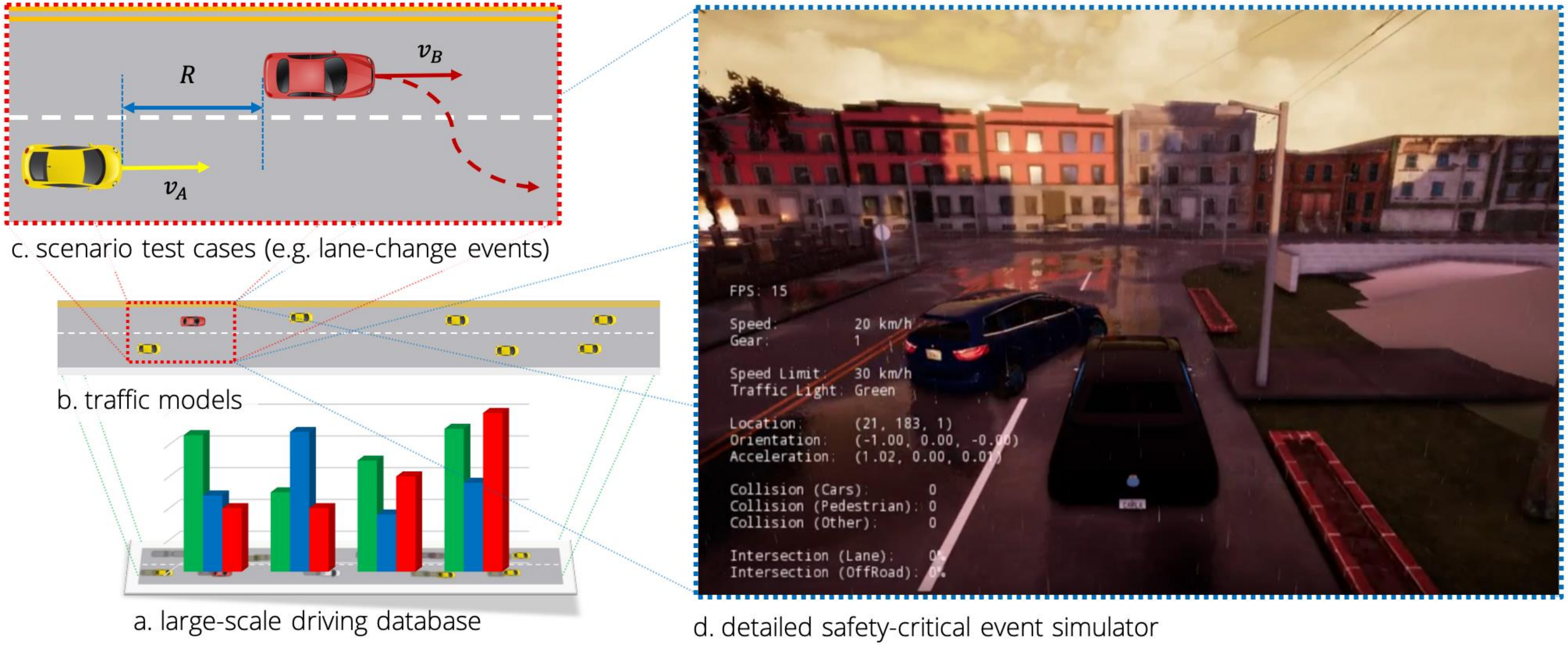
2.99 trillion miles traveled by all vehicles

14,012 annual mileage traveled by an average vehicle

- Impractical to deploy "test" AVs to observe enough crashes
- Approach: Integrate AV algorithms into high-fidelity simulated naturalistic driving environment (built from historical data)
- Rare-event simulation technique to enhance crash observations in simulation (Zhao et al. 2015, Zhao et al. 2017, Huang et al. 2017, Huang et al. 2018, O'Kelly et al. 2018).

Motivation: Autonomous Vehicle Safety Testing

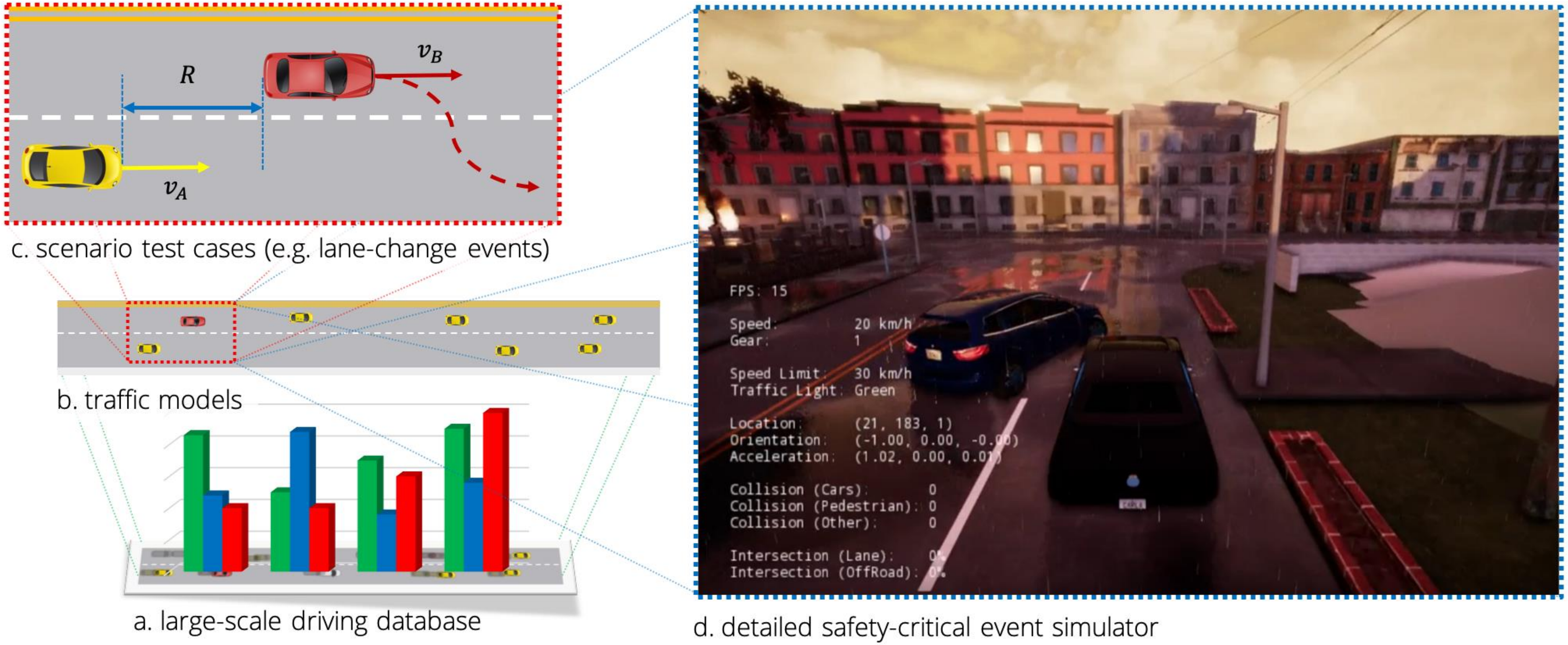
- **Task**: estimate safety measures (e.g. crash probability) of the tested vehicle under specific traffic scenario



Motivation: Autonomous Vehicle Safety Testing

- **Task**: estimate safety measures (e.g. crash probability) of the tested vehicle under specific traffic scenario

$$P(\,(AV\ algorithm, environment) \in conflict\ set\,)$$



Problem Setting

- A stochastic environment denoted $X \sim p$
- Goal: Estimate $\mu = P(X \in S)$ for a rare-event set S
- Rarity parameter γ, so $S = S_\gamma$ such that $\mu \to 0$ as $\gamma \to \infty$ E.g., $S_\gamma = \{x \in R^d: f(x) \geq \gamma\}$
- Key Challenge: Complicated or “black-box” S

Problem Setting

- A stochastic environment denoted $X \sim p$
- Goal: Estimate $\mu = P(X \in S)$ for a rare-event set S
- Rarity parameter γ, so $S = S_\gamma$ such that $\mu \to 0$ as $\gamma \to \infty$ E.g., $S_\gamma = \{x \in R^d: f(x) \geq \gamma\}$
- Key Challenge: Complicated or "black-box" S
- Proposals?
  - Mathematical analysis
  - "Black-box" methods such as cross-entropy (De Boer '05, Rubinstein & Kroese '13…), multi-level splitting / subset simulation (Au & Beck '01, Dean & Dupuis '09, Villen-Altamirano '94…)

Problem Setting

- A stochastic environment denoted $X \sim p$
- Goal: Estimate $\mu = P(X \in S)$ for a rare-event set S
- Rarity parameter γ, so $S = S_\gamma$ such that $\mu \to 0$ as $\gamma \to \infty$ E.g., $S_\gamma = \{x \in R^d : f(x) \geq \gamma\}$
- Key Challenge: Complicated or "black-box" S
- Proposals?
  - Mathematical analysis
  - "Black-box" methods such as cross-entropy (De Boer '05, Rubinstein & Kroese '13…), multi-level splitting / subset simulation (Au & Beck '01, Dean & Dupuis '09, Villen-Altamirano '94…)
  - "Deep Probabilistic Accelerated Evaluation" (Deep-PrAE) (Arief et al. '21)

Efficiency Certificate

- To estimate a small probability μ using Monte Carlo estimator $\hat{\mu}_n$, we need a relative accuracy

$$P(|\hat{\mu}_n - \mu| > \epsilon\mu) \leq \delta$$

for some $0 < \delta, \epsilon < 1$

- Say $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$ where Z_i i.i.d., unbiased

- Markov inequality:

$$\frac{Var(Z_i)}{n\epsilon^2\mu^2} \leq \delta$$

Efficiency Certificate

- To estimate a small probability μ using Monte Carlo estimator $\hat{\mu}_n$, we need a relative accuracy

$$P(|\hat{\mu}_n - \mu| > \epsilon\mu) \leq \delta$$

for some $0 < \delta, \epsilon < 1$

- Say $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$ where Z_i i.i.d., unbiased

- Markov inequality:

$$n \geq \frac{Var(Z_i)}{\delta\epsilon^2\mu^2}$$

Efficiency Certificate

- To estimate a small probability μ using Monte Carlo estimator $\hat{\mu}_n$, we need a relative accuracy

$$P(|\hat{\mu}_n - \mu| > \epsilon\mu) \leq \delta$$

for some $0 < \delta, \epsilon < 1$

- Say $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$ where Z_i i.i.d., unbiased

- Markov inequality:

$$n \geq \frac{RE}{\delta\epsilon^2} \qquad Relative\ Error\ (RE) = \frac{Var(Z_i)}{\mu^2}$$

Efficiency Certificate

- To estimate a small probability μ using Monte Carlo estimator $\hat{\mu}_n$, we need a relative accuracy

$$P(|\hat{\mu}_n - \mu| > \epsilon\mu) \leq \delta$$

for some $0 < \delta, \epsilon < 1$

- Say $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$ where Z_i i.i.d., unbiased

- Markov inequality:

$$n \geq \frac{RE}{\delta\epsilon^2} \qquad Relative\ Error\ (RE) = \frac{Var(Z_i)}{\mu^2}$$

Small RE $\Rightarrow$ Small required n

Efficiency Certificate

- Suppose we use naïve Monte Carlo (NMC):

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} I(X_i \in S)$$

- Then $RE = \frac{\mu(1-\mu)}{\mu^2} = \frac{1}{\mu} \quad \Rightarrow \quad n \approx \frac{1}{\mu}$ blows up when $\mu \to 0$

- If $\mu \approx e^{-c\gamma}$, then $n \approx e^{c\gamma}$ Exponential growth in γ

Efficiency Certificate

- Suppose we use naïve Monte Carlo (NMC):

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} I(X_i \in S)$$

- Then $RE = \frac{\mu(1-\mu)}{\mu^2} = \frac{1}{\mu} \quad \Rightarrow \quad n \approx \frac{1}{\mu}$ blows up when $\mu \to 0$

- If $\mu \approx e^{-c\gamma}$, then $n \approx e^{c\gamma}$ Exponential growth in γ

Efficiency certificate: required n or RE $\approx \log(\frac{1}{\mu})$

How to Obtain Efficiency Certificate?

Importance sampling (IS):

We generate X from a new IS distribution $\tilde{p}$, and output

$$Z = I(X \in S)L(X)$$

where $L(X) = \frac{dp}{d\tilde{p}}(X)$ is the likelihood ratio

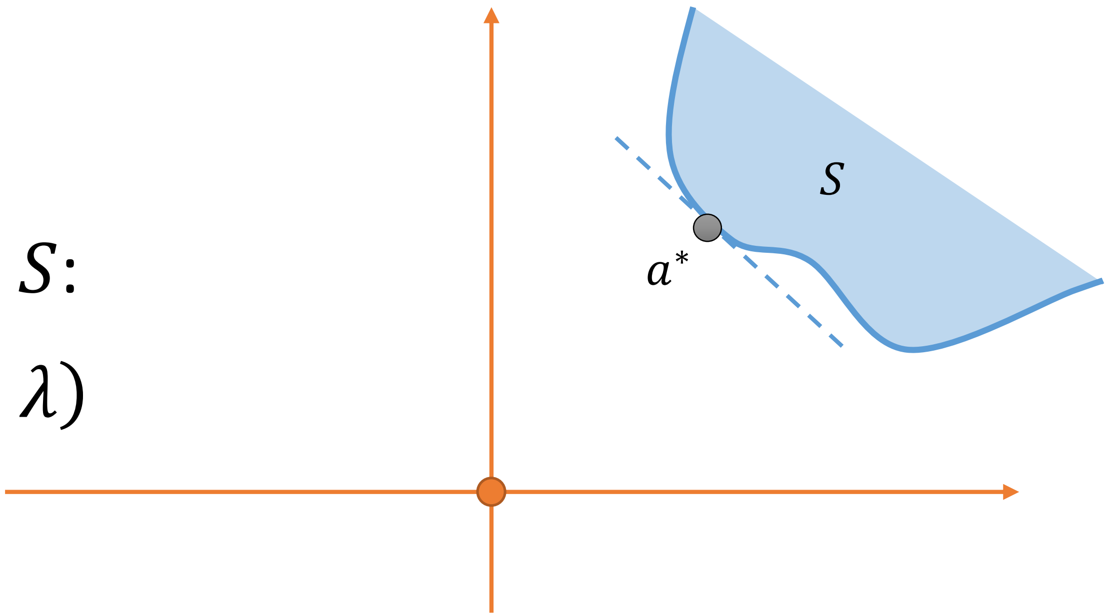
Estimator $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$ is unbiased

How to Obtain Efficiency Certificate?

How do we obtain a Z that has an efficiency certificate?

- Say $X \sim N(\lambda, \Sigma)$
- Estimate $P(X \in S)$
- Look for highest-density point inside S:

$$a^* = argmin_{x \in S} \frac{1}{2}(x - \lambda)^T \Sigma^{-1}(x - \lambda)$$

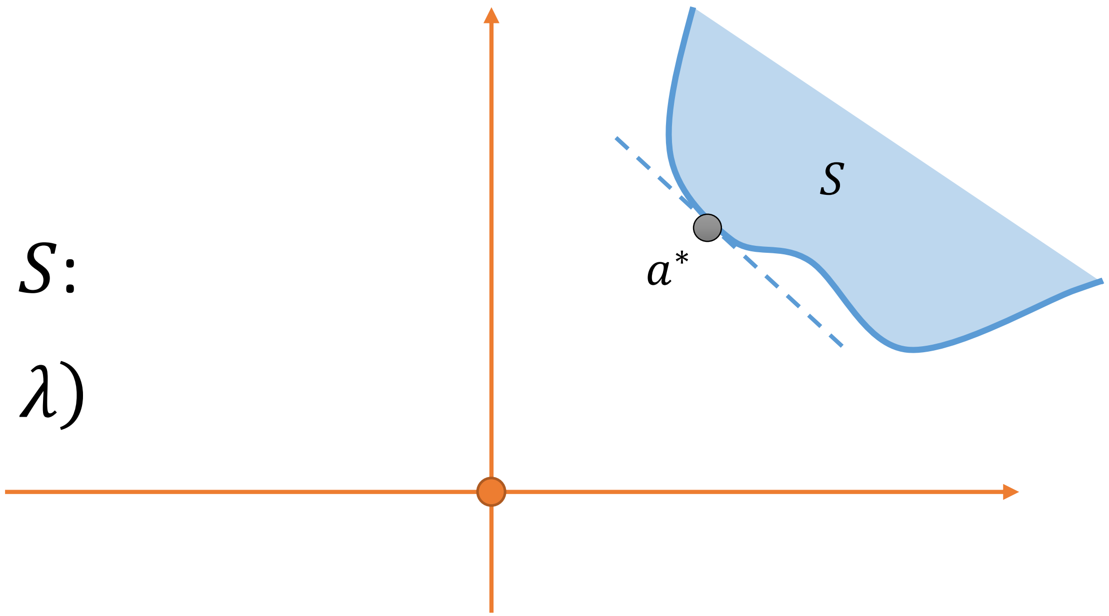


How to Obtain Efficiency Certificate?

How do we obtain a Z that has an efficiency certificate?

- Say $X \sim N(\lambda, \Sigma)$
- Estimate $P(X \in S)$
- Look for highest-density point inside S:

$$a^* = argmin_{x \in S} \frac{1}{2}(x - \lambda)^T \Sigma^{-1} (x - \lambda)$$

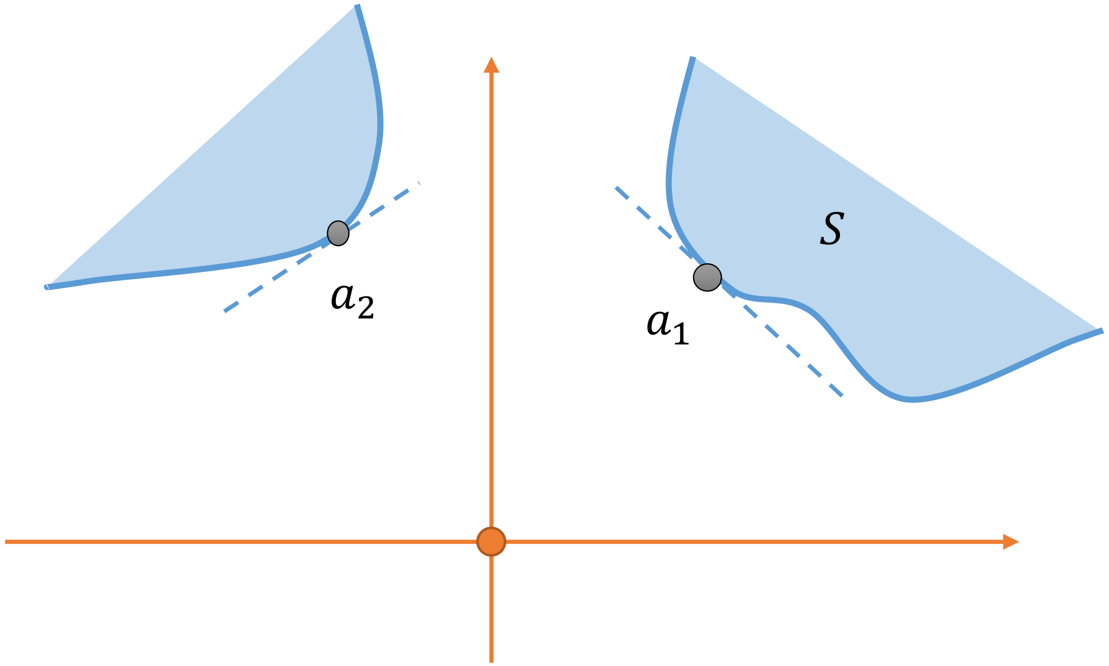


Not always good..

How to Obtain Efficiency Certificate?

How do we obtain a Z that has an efficiency certificate?

Dominant set = $\{a_j$: any point in S is on the "far-side" half-space of at least one $a_j\}$

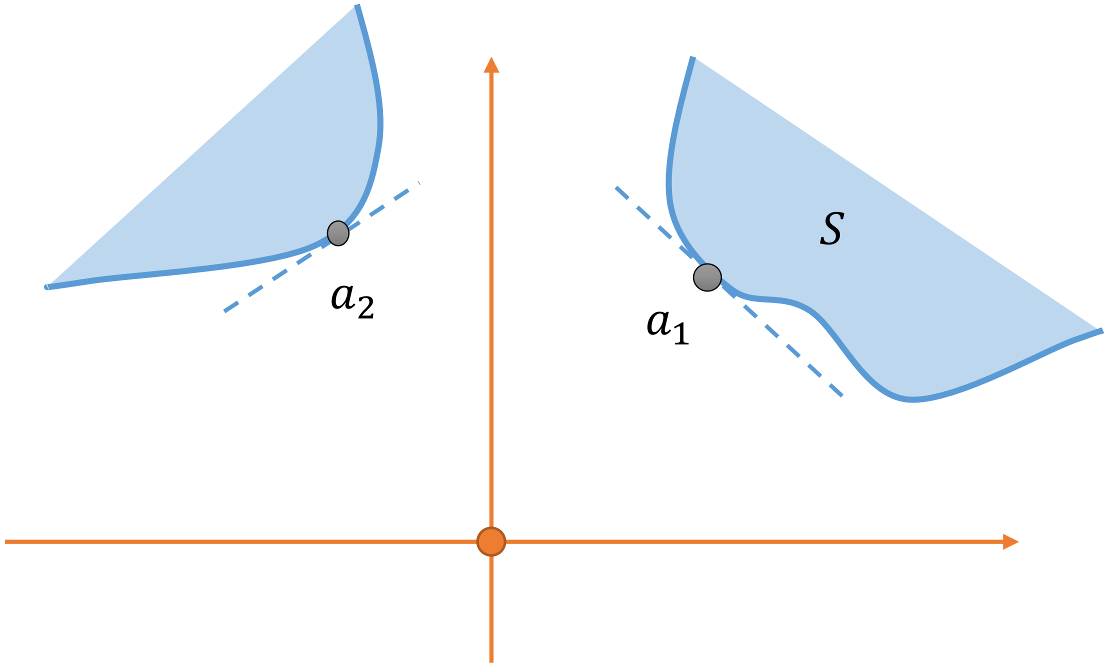


How to Obtain Efficiency Certificate?

How do we obtain a Z that has an efficiency certificate?

An efficient IS needs to “take care” of all dominant points, by using a mixture distribution:

$$\tilde{p} = \sum_{j} q_j N(a_j, \Sigma)$$



References: Sadowsky & Bucklew 1990, Asmussen & Glynn 2007, Blanchet & L. 2012, Juneja & Shahabuddin 2006, Rubino & Tuffin 2009, Owen ‘13, L’Ecuyer et al. 2009, Honnappa et al. 2018, Nakayama 2012, Botev et al. 2007, Rubinstein & Kroese 2016, Rhee et al. 2019…

Perils of Black-Box Variance Reduction Algorithms

Suppose we estimate $\mu = P(X \geq \gamma \; or \; X \leq -k\gamma)$ where $X \sim \; p = N(0,1)$ and $0 < k < 3$. We choose $\tilde{p} = N(\gamma, 1)$ as the IS distribution to obtain $\hat{\mu}_n$.

Perils of Black-Box Variance Reduction Algorithms

Suppose we estimate $\mu = P(X \geq \gamma \; or \; X \leq -k\gamma)$ where $X \sim \; p = N(0,1)$ and $0 < k < 3$. We choose $\tilde{p} = N(\gamma, 1)$ as the IS distribution to obtain $\hat{\mu}_n$. Then

- 1) The relative error of $\hat{\mu}_n$ grows exponentially in γ.
- 2) If n is polynomial in γ, we have $P(|\hat{\mu}_n - \overline{\Phi}(\gamma)| > \varepsilon\overline{\Phi}(\gamma)) = O(\frac{\gamma}{n\varepsilon^2})$ for any $\varepsilon > 0$ where $\overline{\Phi}(\gamma) = P(X \geq \gamma) < \mu$, and the empirical relative error $= O(n^2)$ with probability higher than $1 - 1/2^n$.

Perils of Black-Box Variance Reduction Algorithms

Suppose we estimate $\mu = P(X \geq \gamma \; or \; X \leq -k\gamma)$ where $X \sim \; p = N(0,1)$ and $0 < k < 3$. We choose $\tilde{p} = N(\gamma, 1)$ as the IS distribution to obtain $\hat{\mu}_n$. Then

- 1) RE blows up
- 2) If n is polynomial in γ, we have $P(|\hat{\mu}_n - \bar{\Phi}(\gamma)| > \varepsilon\bar{\Phi}(\gamma)) = O(\frac{\gamma}{n\varepsilon^2})$ for any $\varepsilon > 0$ where $\bar{\Phi}(\gamma) = P(X \geq \gamma) < \mu$, and the empirical relative error $= O(n^2)$ with probability higher than $1 - 1/2^n$.

Perils of Black-Box Variance Reduction Algorithms

Suppose we estimate $\mu = P(X \geq \gamma \ or \ X \leq -k\gamma)$ where $X \sim \ p = N(0,1)$ and $0 < k < 3$. We choose $\tilde{p} = N(\gamma, 1)$ as the IS distribution to obtain $\hat{\mu}_n$. Then

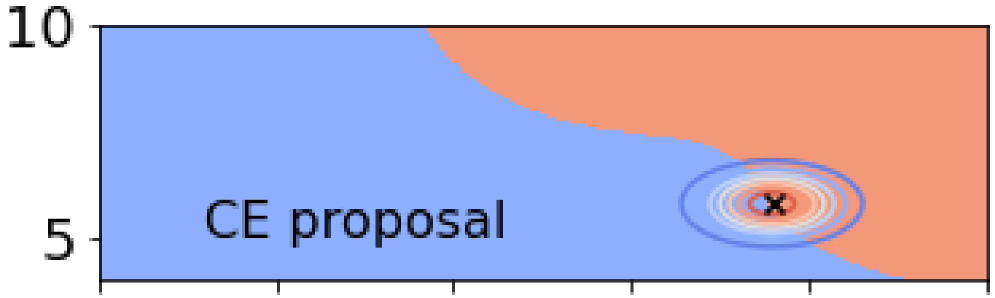
- 1) RE blows up
- 2) But you don't know..

Perils of Black-Box Variance Reduction Algorithms

Suppose we estimate $\mu = P(X \geq \gamma \text{ } or \text{ } X \leq -k\gamma)$ where $X \sim \; p = N(0,1)$ and $0 < k < 3$. We choose $\tilde{p} = N(\gamma, 1)$ as the IS distribution to obtain $\hat{\mu}_n$. Then

- 1) RE blows up
- 2) But you don't know..
- In experiments, you see a nice empirical RE that gives you high "confidence" on your result, but in fact you run into systematic under-estimation

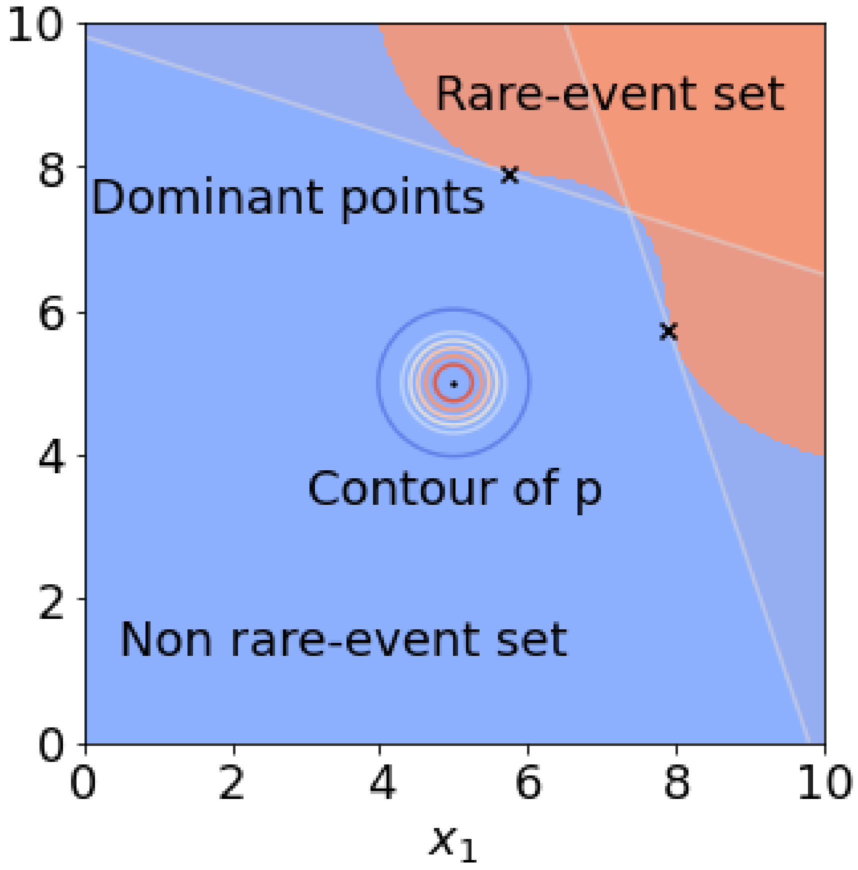
Perils of Black-Box Variance Reduction Algorithms



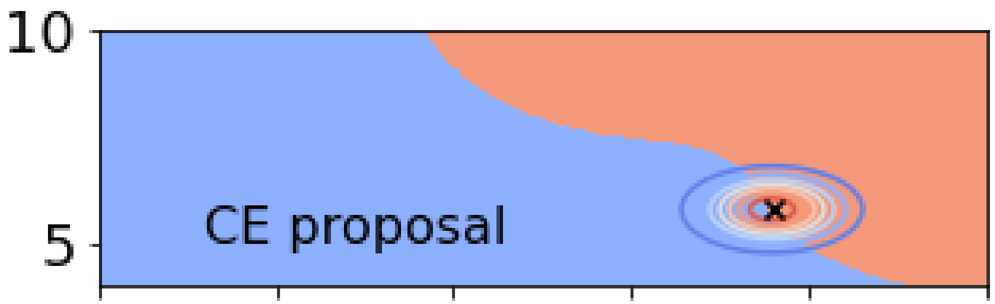
Cross-entropy method:

- Choose a parametric IS class $\tilde{p}$
- Empirically minimize Kullback-Leibler divergence between $\tilde{p}$ and a "zero-variance" distribution

May not converge to an IS that has efficiency certificate..



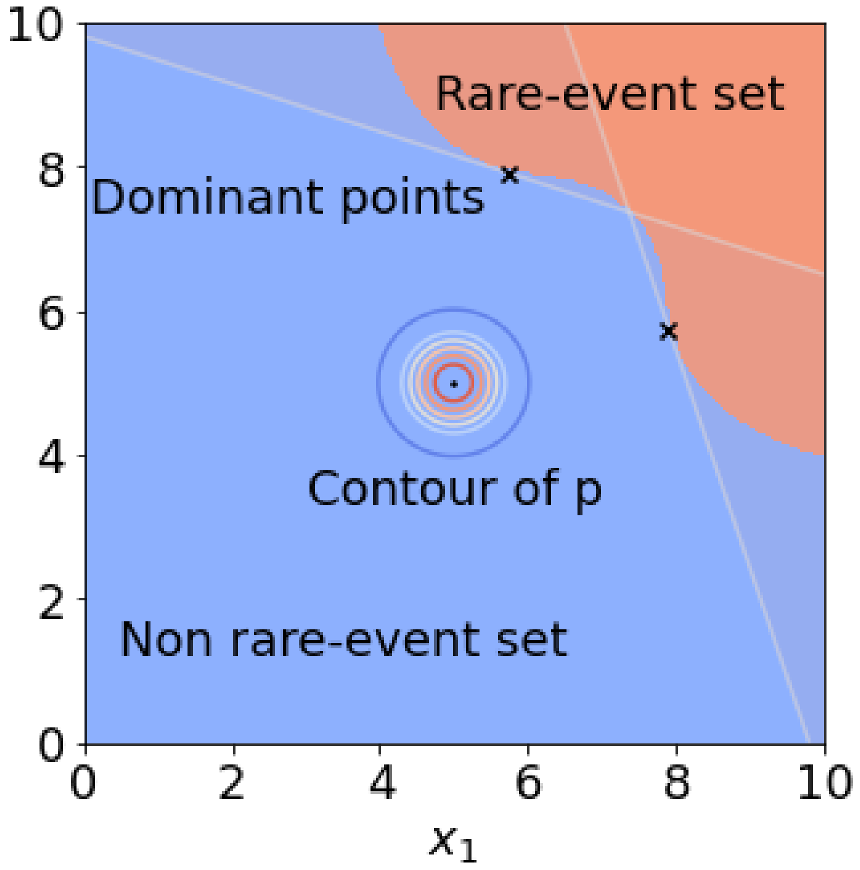
Perils of Black-Box Variance Reduction Algorithms



Cross-entropy method:

- Choose a parametric IS class $\tilde{p}$
- Empirically minimize Kullback-Leibler divergence between $\tilde{p}$ and a "zero-variance" distribution

May not converge to an IS that has efficiency certificate..
Hence possibly under-estimate, but you don't know..



Deep Probabilistic Accelerated Evaluation

- Relaxed Efficiency Certificate: Relax the rare-event estimation problem to estimating upper and lower bounds

Deep Probabilistic Accelerated Evaluation

- Relaxed Efficiency Certificate: Relax the rare-event estimation problem to estimating upper and lower bounds

- A two-stage procedure:

  - Stage 1: Rare-event-set learning

  - Stage 2: Search dominant points of the learned set and run mixture IS

Deep Probabilistic Accelerated Evaluation

- Relaxed Efficiency Certificate: Relax the rare-event estimation problem to estimating upper and lower bounds

- A two-stage procedure:

  - Stage 1: Rare-event-set learning

  - Stage 2: Search dominant points of the learned set and run mixture IS

- A suitably constructed ReLU neural network classifier to learn set (Stage 1) achieves the relaxed efficiency certificate

Relaxed Efficiency Certificate

We achieve a relative accuracy

$$P(|\hat{\mu}_n - \mu| > \epsilon\mu) \leq \delta$$

with sample size $n \approx \log\left(\frac{1}{\mu}\right)$

Relaxed Efficiency Certificate

We achieve a relative accuracy

$$P(\hat{\mu}_n - \mu < -\epsilon\mu) \leq \delta$$

with sample size $n \approx \log\left(\frac{1}{\mu}\right)$

Relaxed Efficiency Certificate

We achieve a relative accuracy

$$P(\hat{\mu}_n - \mu < -\epsilon\mu) \leq \delta$$

with sample size $n \approx \log\left(\frac{1}{\mu}\right)$

Want to get a good upper bound for μ with $n \approx \log\left(\frac{1}{\mu}\right)$

Analogously for lower bound

Relaxed Efficiency Certificate

We achieve a relative accuracy

$$P(\hat{\mu}_n - \mu < -\epsilon\mu) \leq \delta$$

with sample size $n \approx \log\left(\frac{1}{\mu}\right)$

A sufficient condition:

- $\hat{\mu}_n$ is upward biased, i.e., $\bar{\mu} = E\hat{\mu}_n \geq \mu$, and
- $\hat{\mu}_n$ has efficiency certificate to estimate $\bar{\mu}$

Relaxed Efficiency Certificate

We achieve a relative accuracy

$$P(\hat{\mu}_n - \mu < -\epsilon\mu) \leq \delta$$

with sample size $n \approx \log\left(\frac{1}{\mu}\right)$

A sufficient condition:

- $\hat{\mu}_n$ is upward biased, i.e., $\bar{\mu} = E\hat{\mu}_n \geq \mu$, and
- $\hat{\mu}_n$ has efficiency certificate to estimate $\bar{\mu}$ ← Use mixture IS, if we know the rare-event set

Relaxed Efficiency Certificate

We achieve a relative accuracy

$$P(\hat{\mu}_n - \mu < -\epsilon\mu) \leq \delta$$

with sample size $n \approx \log\left(\frac{1}{\mu}\right)$

A sufficient condition:

- $\hat{\mu}_n$ is upward biased, i.e., $\bar{\mu} = E\hat{\mu}_n \geq \mu$, and
- $\hat{\mu}_n$ has efficiency certificate to estimate $\bar{\mu}$

Outer Rare-Event Set Approximation

If we can learn $\bar{S}$ such that $\bar{S} \supset S$, then $\bar{\mu} = P(X \in \bar{S}) \geq P(X \in S) = \mu$

Goal: Obtain a good outer approximation of S

Outer Rare-Event Set Approximation

If we can learn $\bar{S}$ such that $\bar{S} \supset S$, then $\bar{\mu} = P(X \in \bar{S}) \geq P(X \in S) = \mu$

Goal: Obtain a good outer approximation of S

How to learn a set?

Classification task:

- Sample $\tilde{X}_i$ and label $Y_i = I(\tilde{X}_i \in S)$
- Then "train" a classifier using the data $(\tilde{X}_i, Y_i)$

Outer Rare-Event Set Approximation

If we can learn $\bar{S}$ such that $\bar{S} \supset S$, then $\bar{\mu} = P(X \in \bar{S}) \geq P(X \in S) = \mu$

Goal: Obtain a good outer approximation of S

How to learn a set?

Classification task:

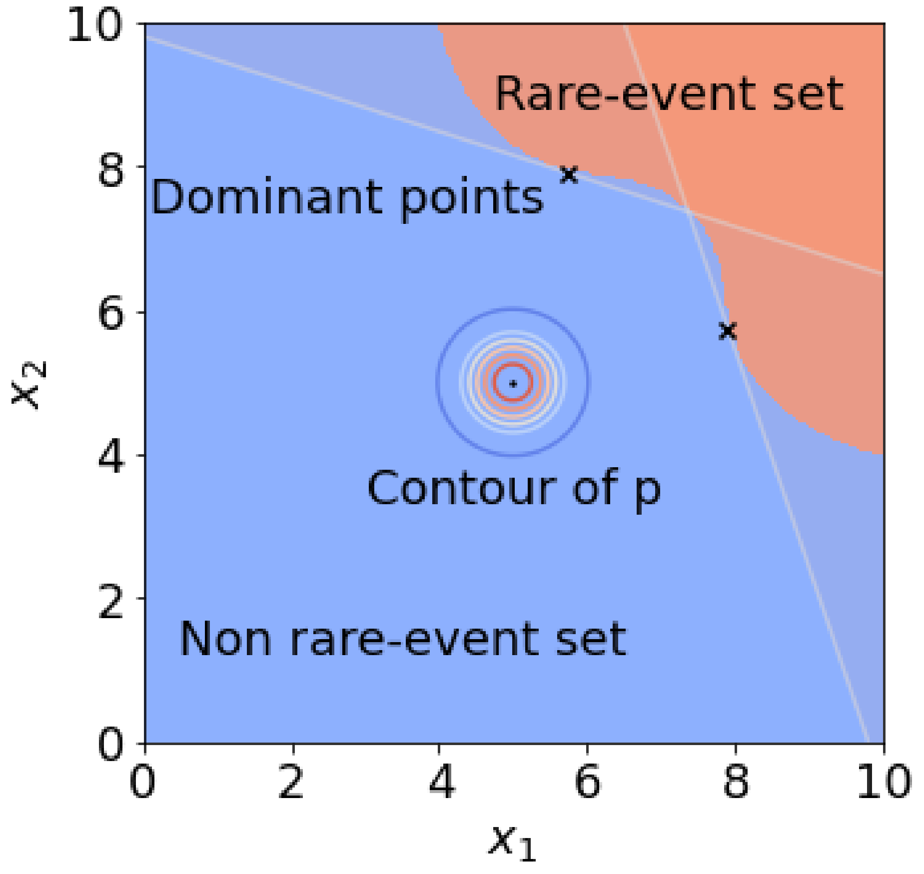
- Sample $\tilde{X}_i$ and label $Y_i = I(\tilde{X}_i \in S)$
- Then "train" a classifier using the data $(\tilde{X}_i, Y_i)$

Outer approximation means that the false negative rate is zero…

Outer Rare-Event Set Approximation

Orthogonal monotone set:

If $x \in S$, then any $x' \geq x$ also $\in S$

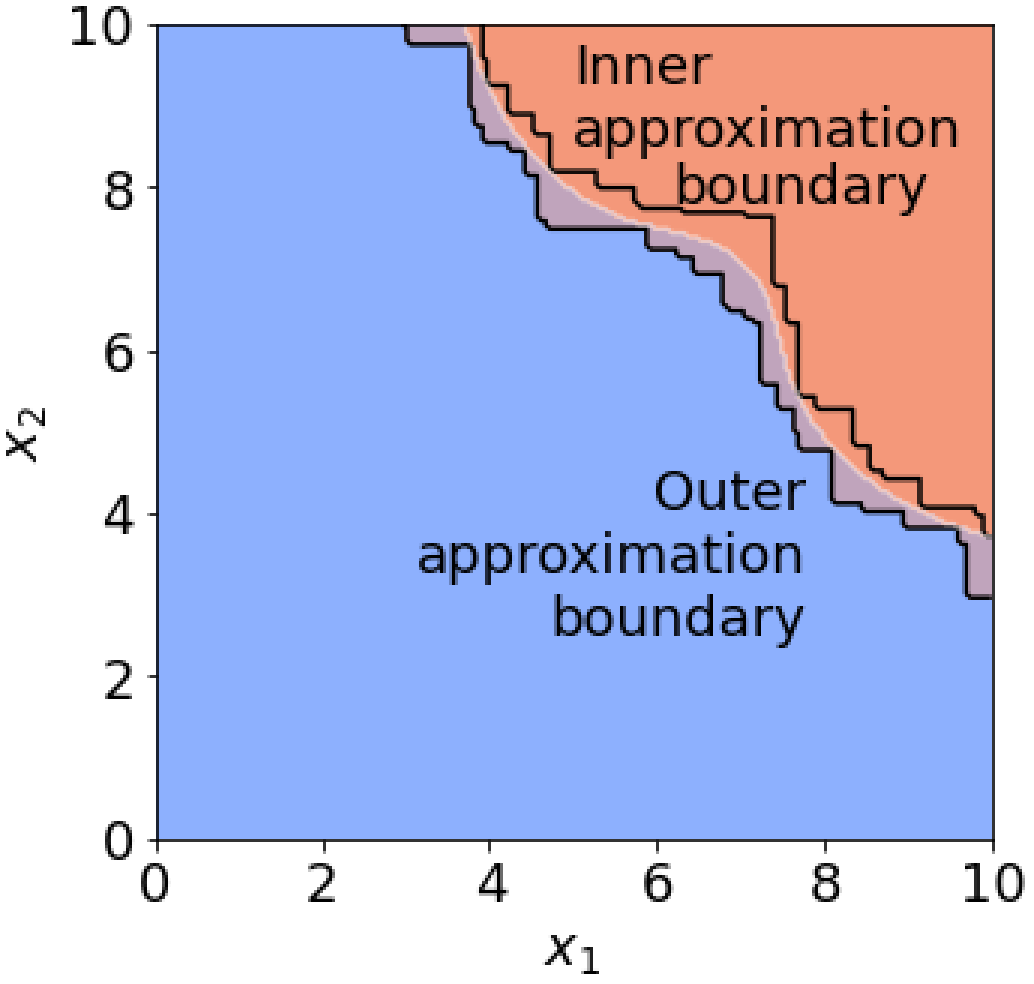


Outer Rare-Event Set Approximation

Orthogonal monotone set:

If $x \in S$, then any $x' \geq x$ also $\in S$

An easy outer approximation is to take the complement of the "step-boundary" formed by all 0-labeled $\tilde{X}_i$



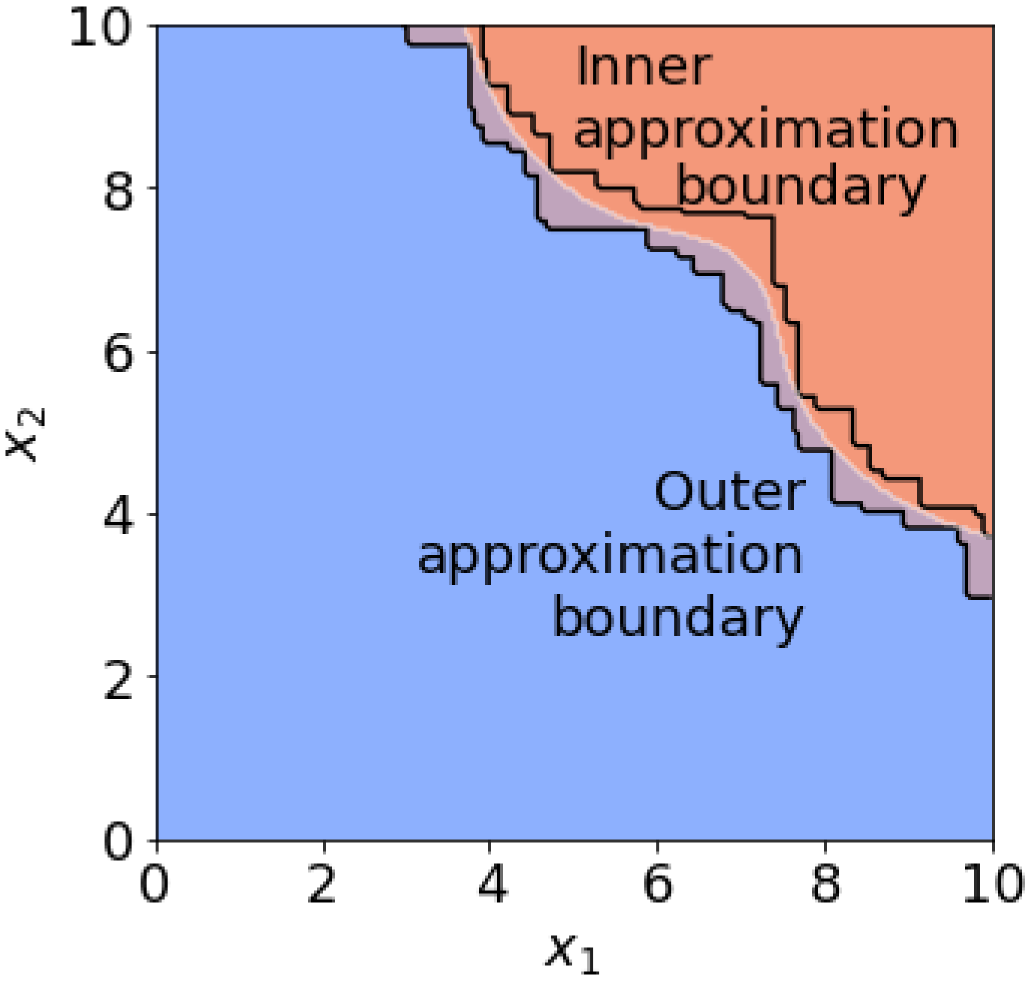
Outer Rare-Event Set Approximation

Orthogonal monotone set:

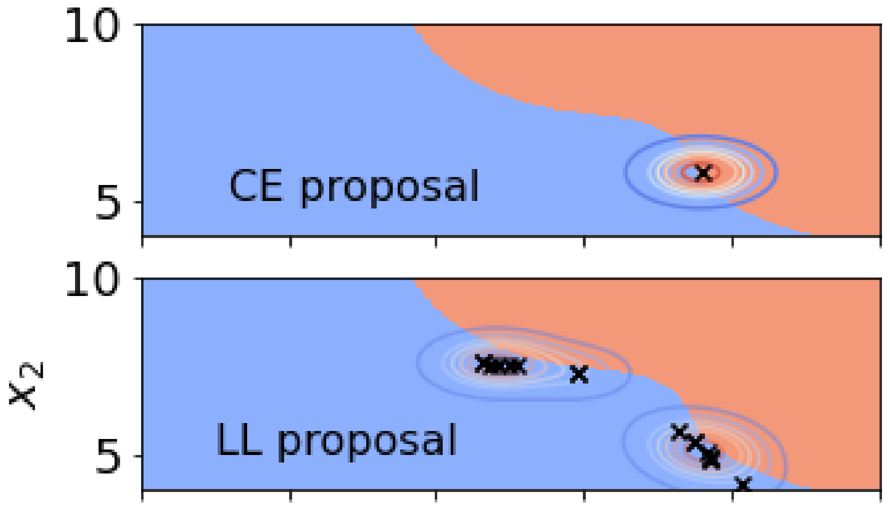
If $x \in S$, then any $x' \geq x$ also $\in S$

An easy outer approximation is to take the complement of the “step-boundary” formed by all 0-labeled $\tilde{X}_i$

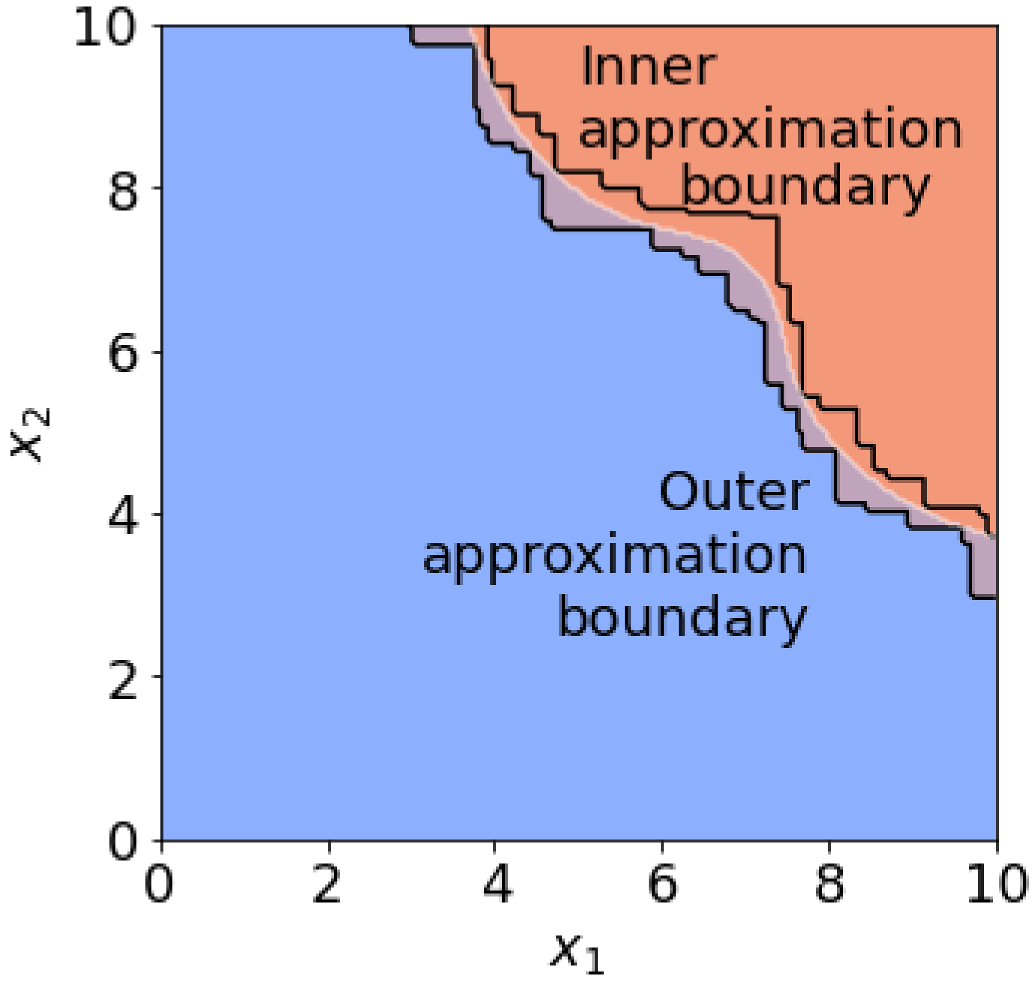
Lazy-learner IS: Simulate the probability of falling into this “step-boundary” set



Outer Rare-Event Set Approximation



Lazy-learner IS:
Too many dominant points..



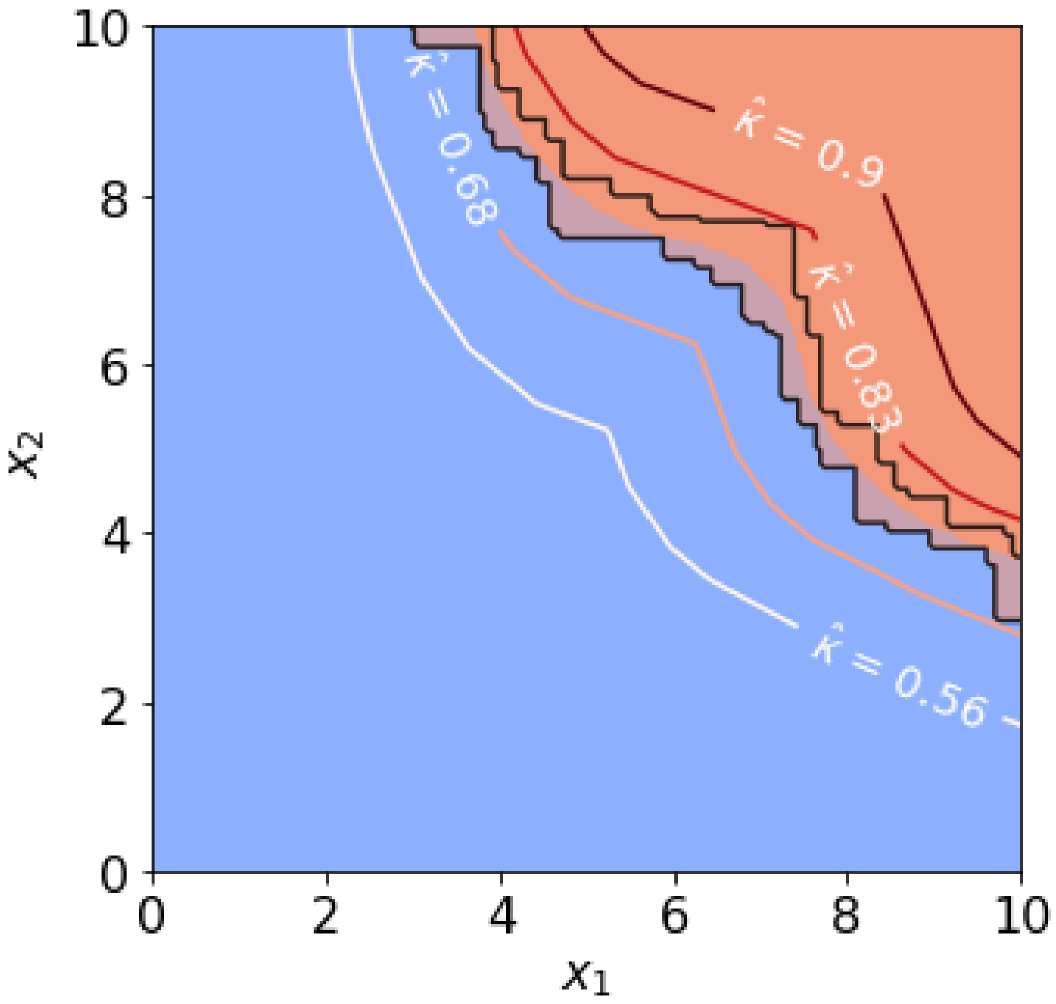
Outer Rare-Event Set Approximation

Deep-learning-based IS:

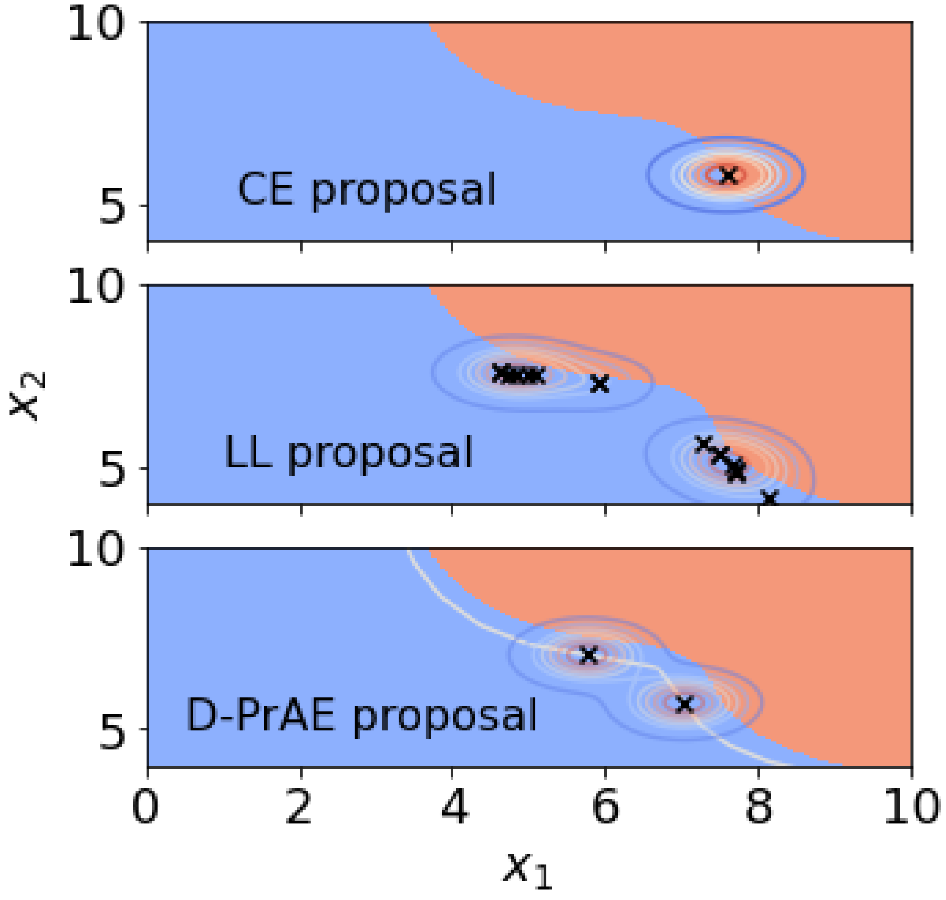
Train a ReLU-activated classifier to obtain an outer approximation of S

First train $\{\hat{g}(x) \geq \kappa\}$, then tune κ to satisfy $\bar{S} \supset S$

Stage 2: Learn the dominant points and construct mixture IS



Deep Probabilistic Accelerated Evaluation



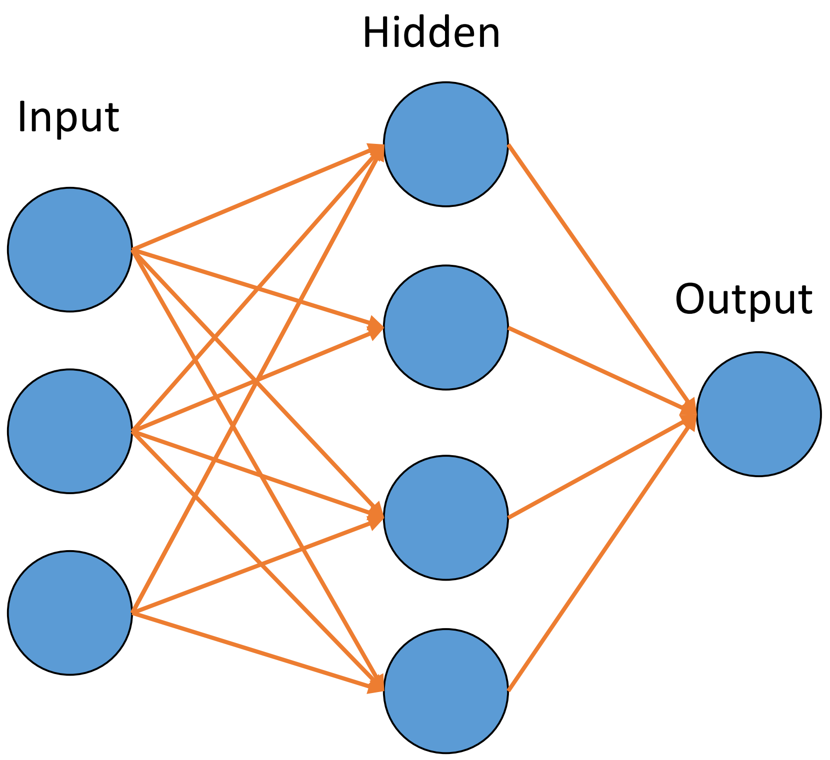
Deep-PrAE IS:

Right # of dominant points, and satisfy relaxed efficiency certificate

Behind-the-Scene 1: MIP to Search "Best" Dominant Points in ReLU Network

ReLU network:

- Input s_0.
- In layer k with input s_k, $y_k = W_k s_k + b_k$, $s_{k+1} = \sigma(y_k) = \max\{y_k, 0\}$.



$$\min_{x, s_0, \ldots, s_L, z_1 \ldots, z_L} \|x\|^2$$

$$\begin{aligned}
s.t.\quad & s_L \geq \gamma \\
& s_i \leq W_i{}^T s_{i-1} + b_i - l(1 - z_i),\ i = 1, \ldots, L \\
& s_i \geq W_i{}^T s_{i-1} + b_i,\ i = 1, \ldots, L \\
& s_i \leq u z_i,\ i = 1, \ldots, L \\
& s_i \geq 0,\ i = 1, \ldots, L \\
& z_i \in \{0, 1\}^{n_i},\ i = 1, \ldots, L \\
& s_0 = x
\end{aligned}$$

$g(x) \geq \gamma$

x: input variables
s_i's: input/output in each layer
z_i's: integer variables

$s_i = \max\{W_i^T s_{i-1} + b_i, 0\}$

- The number of integer variables is the number of neurons.
- Each neuron has four corresponding constraints.

Behind-the-Scene 2: Sequential "Cutting Plane" to Search All Dominant Points in ReLU Network

Input: Prediction model $g(x)$, threshold γ.

Output: Dominating-points set A.

Start with $A = \emptyset$;

While $\{x : g(x) \geq \gamma, a_i'(x - a_i) < 0,\ \forall a_i \in A\} \neq \emptyset$ **do**

Find a dominating point a by solving the optimization problem

$$\begin{aligned} a = \arg\min_{x} \quad & \|x\|^2 \\ s.t. \quad & g(x) \geq \gamma \\ & a_i'(x - a_i) < 0, \text{ for } \forall a_i \in A \end{aligned}$$

and update $A \leftarrow A \cup \{a\}$;

End

Behind-the-Scene 3: Conservativeness of Upper Bounds

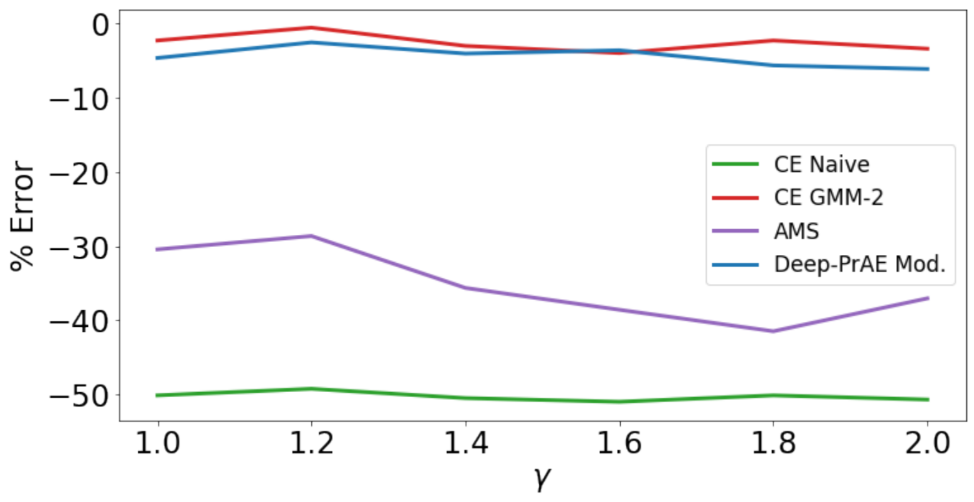
With probability at least $1-\delta$,

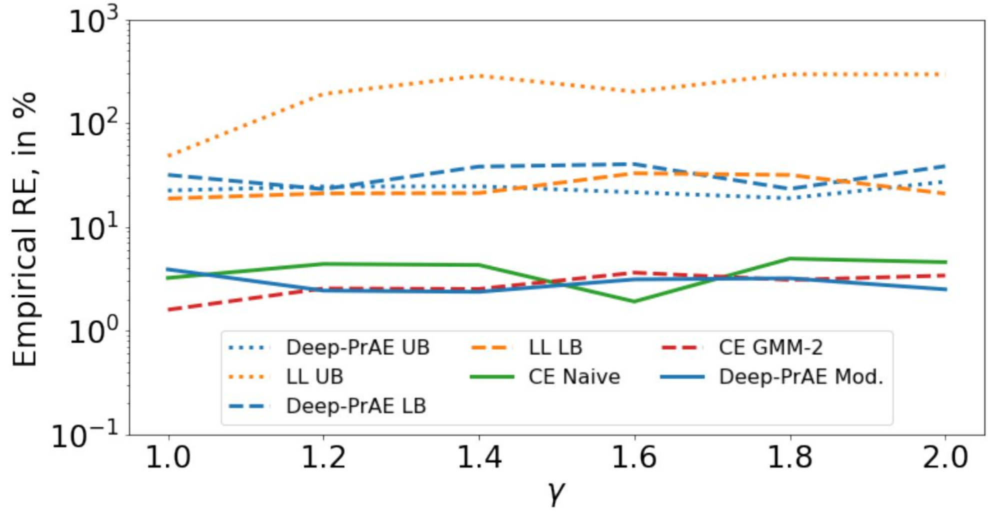
$$P(false+ve) \leq \frac{R(g^*) + 2\sup_{g\in\mathcal{G}}\left|R_{n_1}(g) - R(g)\right|}{h\left(\kappa^* - t(\delta, n_1)\sqrt{d}\,\mathrm{Lip}(g^*) - \|\hat{g} - g^*\|_\infty\right)}.$$

Here, $\mathrm{Lip}(g^*)$ is the Lipschitz parameter of g^*, and $t(\delta, n_1) = 3\left(\frac{\log(n_1 q_l) + \mathrm{d}\log M + \log\frac{1}{\delta}}{n_1 q_l}\right)^{\frac{1}{\delta}}$.

Example

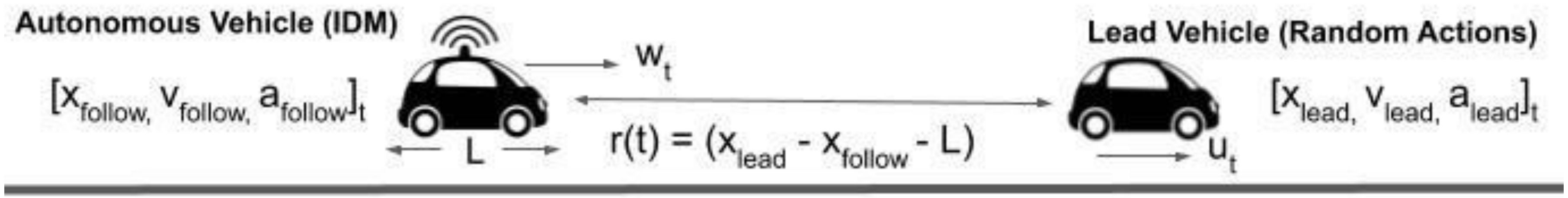
- CE GMM-2 captures enough dominant points ⇒ efficient and gives accurate estimate
- CE Naive is “confident” but under-estimate
- Deep-PrAE (modified) locates enough dominant points ⇒ efficient and gives accurate estimate





Example of Intelligent Driver Model (IDM)

- A car-following scenario involving a human-driven lead vehicle followed by an autonomous vehicle
- Time horizon $T = 60s$ with a sequence of 15 Gaussian random actions at a 4 second epoch
- 10,000 sample budget.

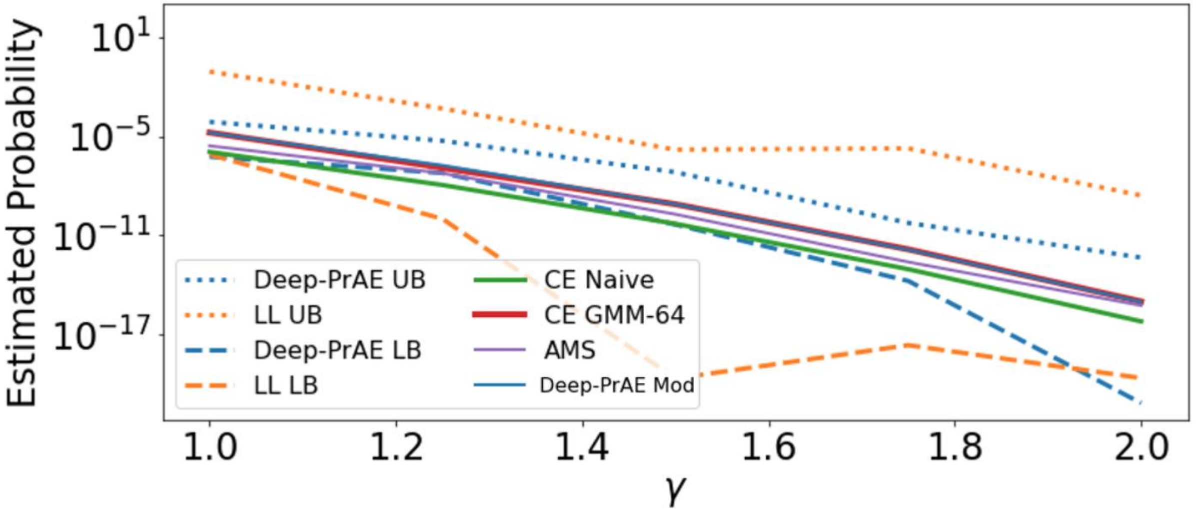


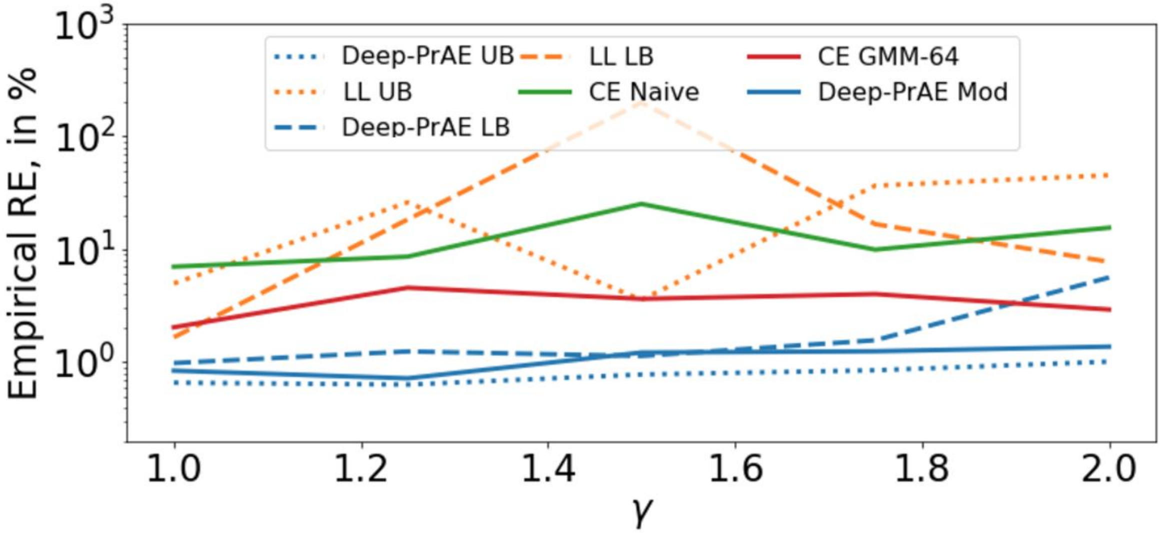
$$
\begin{aligned}
\dot{x}_f &= v_f \\
\dot{x}_l &= v_l \\
\dot{v}_l &= u_t \\
\dot{v}_f &= a\left(1 - \left(\frac{v_f}{v0}\right)^{\delta} - \left(\frac{s^*(v_f, \Delta v_f)}{s_f}\right)^2\right) \\
s^*(v_f, \Delta v_f) &= s0 + v_f T + \frac{v_f \Delta v_f}{2\sqrt{ab}} \\
s_f &= x_l - x_f - L \\
\Delta v_f &= v_f - v_l
\end{aligned}
$$

Parameters	Value
Safety Distance (s0)	2 m
Speed of AV in free traffic (v0)	30 m/s
Maximum Acceleration of AV (a)	2 m/s^2
Comfortable Deceleration of AV (b)	1.67 m/s^2
Maximum Deceleration of AV (d)	4 m/s^2
Safe Time Headway (T)	1.5 s
Acceleration Exponent Parameter (\delta)	4
Car Length (L)	4 m

Self-Driving Example

- Deep-PrAE produces tighter bounds than LL
- When $\gamma = 1$, LL UB has 5,644 dominant points vs 42 in Deep-PrAE
- Most methods are "confident" about their estimation, and some of them must under-estimate





Summary

- Motivated by safety-testing of intelligent physical systems
- Motivated from the perils of black-box variance reduction algorithms
- Deep Probabilistic Accelerated Evaluation (Deep-PrAE):

  Combine ReLU-activated neural net classifiers for set learning with dominant point methodology to design IS with relaxed efficiency certificate

+ve Thoughts:

- Towards "model-free" importance sampling
- Towards "high-dimensional" importance sampling

-ve Thoughts:

- Tail model error
- How to interpret rare-event probability

References

Application of Importance Sampling in Self-Driving Vehicle Safety Testing:

- Zhao, D., Lam, H., Peng, H., Bao, S., LeBlanc, D.J., Nobukawa, K. and Pan, C.S., 2016. Accelerated evaluation of automated vehicles safety in lane-change scenarios based on importance sampling techniques. *IEEE Transactions on Intelligent Transportation Systems*, *18*(3), pp.595-607.
- Zhao, D., Huang, X., Peng, H., Lam, H. and LeBlanc, D.J., 2017. Accelerated evaluation of automated vehicles in car-following maneuvers. *IEEE Transactions on Intelligent Transportation Systems*, *19*(3), pp.733-744.
- Huang, Z., Lam, H., LeBlanc, D.J. and Zhao, D., 2017. Accelerated evaluation of automated vehicles using piecewise mixture models. *IEEE Transactions on Intelligent Transportation Systems*, *19*(9), pp.2845-2855.

Deep Importance Sampling:

- Arief, M., Huang, Z., Kumar, G.K.S., Bai, Y., He, S., Ding, W., Lam, H. and Zhao, D., 2021, March. Deep Probabilistic Accelerated Evaluation: A robust certifiable rare-event simulation methodology for black-box safety-critical systems. In *International Conference on Artificial Intelligence and Statistics (AISTATS)* (pp. 595-603). PMLR.
- Bai, Y., Huang, Z., Lam, H. and Zhao, D., 2020. Rare-event simulation for neural network and random forest predictors. *arXiv preprint arXiv:2010.04890*. Preliminary version in *2018 Winter Simulation Conference (WSC)* (pp. 1730-1741). IEEE.