Differential Entropy Estimation under Gaussian Convolutions

Ziv Goldfeld

MIT

Information Theory and Applications Workshop

February 14th, 2019

Collaborators: Kristjan Greenewald, Jonathan Weed and Yury Polyanskiy

New Estimation Problem

Setup: Estimate $h(X+Z)$ for d-dimensional, independent X and Z:

New Estimation Problem

Setup: Estimate $h(X+Z)$ for d-dimensional, independent X and Z:

- $X \sim P$, where $P \in \mathcal{F}_d$ is unknown (nonparametric class)

New Estimation Problem

Setup: Estimate $h(X+Z)$ for d-dimensional, independent X and Z:

- $X \sim P$, where $P \in \mathcal{F}_d$ is unknown (nonparametric class)
- $Z \sim \mathcal{N}_\sigma \triangleq \mathcal{N}(0, \sigma^2 \mathrm{I}_d)$

New Estimation Problem

Setup: Estimate $h(X+Z)$ for d-dimensional, independent X and Z:

- $X \sim P$, where $P \in \mathcal{F}_d$ is unknown (nonparametric class)
- $Z \sim \mathcal{N}_\sigma \triangleq \mathcal{N}(0, \sigma^2 \mathrm{I}_d)$

Resources: An estimator $\hat{h}$ of $h(X+Z)$ can use

New Estimation Problem

Setup: Estimate $h(X+Z)$ for d-dimensional, independent X and Z:

- $X \sim P$, where $P \in \mathcal{F}_d$ is unknown (nonparametric class)
- $Z \sim \mathcal{N}_\sigma \triangleq \mathcal{N}(0, \sigma^2 \mathrm{I}_d)$

Resources: An estimator $\hat{h}$ of $h(X+Z)$ can use

1. n i.i.d. samples $X^n \triangleq (X_i)_{i=1}^n$ from P.

New Estimation Problem

Setup: Estimate $h(X+Z)$ for d-dimensional, independent X and Z:

- $X \sim P$, where $P \in \mathcal{F}_d$ is unknown (nonparametric class)
- $Z \sim \mathcal{N}_\sigma \triangleq \mathcal{N}(0, \sigma^2 \mathrm{I}_d)$

Resources: An estimator $\hat{h}$ of $h(X+Z)$ can use

1. n i.i.d. samples $X^n \triangleq (X_i)_{i=1}^n$ from P.
2. Knowledge of $\mathcal{N}_\sigma$.

New Estimation Problem

Setup: Estimate $h(X+Z)$ for d-dimensional, independent X and Z:

- $X \sim P$, where $P \in \mathcal{F}_d$ is unknown (nonparametric class)
- $Z \sim \mathcal{N}_\sigma \triangleq \mathcal{N}(0, \sigma^2 \mathrm{I}_d)$

Resources: An estimator $\hat{h}$ of $h(X+Z)$ can use

1. n i.i.d. samples $X^n \triangleq (X_i)_{i=1}^n$ from P.
2. Knowledge of $\mathcal{N}_\sigma$.

Absolute Error Minimax Risk:

$$\mathcal{R}^\star(n, \sigma, \mathcal{F}_d) \triangleq \inf_{\hat{h}} \sup_{P \in \mathcal{F}_d} \mathbb{E}\left| h(P * \mathcal{N}_\sigma) - \hat{h}(X^n, \sigma) \right|$$

New Estimation Problem

Setup: Estimate $h(X+Z)$ for d-dimensional, independent X and Z:

- $X \sim P$, where $P \in \mathcal{F}_d$ is unknown (nonparametric class)
- $Z \sim \mathcal{N}_\sigma \triangleq \mathcal{N}(0, \sigma^2 \mathrm{I}_d)$

Resources: An estimator $\hat{h}$ of $h(X+Z)$ can use

1. n i.i.d. samples $X^n \triangleq (X_i)_{i=1}^n$ from P.
2. Knowledge of $\mathcal{N}_\sigma$.

Absolute Error Minimax Risk:

$$\mathcal{R}^\star(n, \sigma, \mathcal{F}_d) \triangleq \inf_{\hat{h}} \sup_{P \in \mathcal{F}_d} \mathbb{E} \left| h(P * \mathcal{N}_\sigma) - \hat{h}(X^n, \sigma) \right|$$

⊛ **Sample complexity** $n^\star(\eta, \sigma, \mathcal{F}_d)$: least n needed for η-gap estimation.

Motivation - Information Theory & Deep Learning

Motivation - Information Theory & Deep Learning

- **Information Bottleneck Theory** [Tishby-Zaslavsky'15, Shwartz-Tishby'17]
  Estimate mutual information between layers of a DNN

Motivation - Information Theory & Deep Learning

- **Information Bottleneck Theory** [Tishby-Zaslavsky'15, Shwartz-Tishby'17]
  Estimate mutual information between layers of a DNN
- **Unsupervised Learning** [Hjelm et al'18, Oord-Li-Vinyals'18]
  Mutual information for learning representations (Deep InfoMax, CPC)

Motivation - Information Theory & Deep Learning

- **Information Bottleneck Theory** [Tishby-Zaslavsky'15, Shwartz-Tishby'17]
  Estimate mutual information between layers of a DNN

- **Unsupervised Learning** [Hjelm et al'18, Oord-Li-Vinyals'18]
  Mutual information for learning representations (Deep InfoMax, CPC)

⊛ IT measure **degenerate** over DNNs with fixed parameters

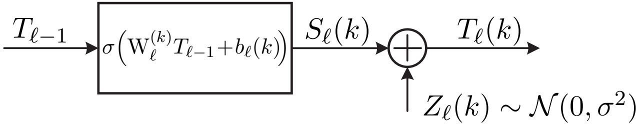
Motivation - Information Theory & Deep Learning

- **Information Bottleneck Theory** [Tishby-Zaslavsky'15, Shwartz-Tishby'17]
  Estimate mutual information between layers of a DNN

- **Unsupervised Learning** [Hjelm et al'18, Oord-Li-Vinyals'18]
  Mutual information for learning representations (Deep InfoMax, CPC)

⊛ IT measure **degenerate** over DNNs with fixed parameters

$\Longrightarrow$ **Study Info. Flow in DNNs:** Stochastic DNNs via noise injection
[G.-Berg-Greenewald-Melnyk-Nguyen-Kingsbury-Polyanskiy'18]



Motivation - Information Theory & Deep Learning

- **Information Bottleneck Theory** [Tishby-Zaslavsky'15, Shwartz-Tishby'17]

  Estimate mutual information between layers of a DNN

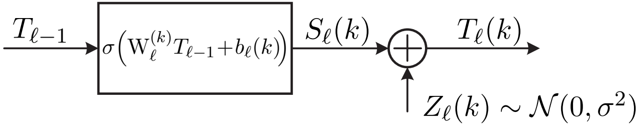
- **Unsupervised Learning** [Hjelm et al'18, Oord-Li-Vinyals'18]

  Mutual information for learning representations (Deep InfoMax, CPC)

⊛ IT measure **degenerate** over DNNs with fixed parameters

$\Longrightarrow$ **Study Info. Flow in DNNs:** Stochastic DNNs via noise injection

[G.-Berg-Greenewald-Melnyk-Nguyen-Kingsbury-Polyanskiy'18]



⊛ Can sample S_ℓ (gen. model) & want to estimate $h(T_\ell)=h(S_\ell+Z_\ell)$

Naive Approach - General-Purpose Estimators

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Naive Approach - General-Purpose Estimators

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Method: Estimate $h(P * \mathcal{N}_\sigma)$ via i.i.d. (**noisy**) samples from $P * \mathcal{N}_\sigma$

Naive Approach - General-Purpose Estimators

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Method: Estimate $h(P * \mathcal{N}_\sigma)$ via i.i.d. (**noisy**) samples from $P * \mathcal{N}_\sigma$

Theoretical Guarantees:

Naive Approach - General-Purpose Estimators

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Method: Estimate $h(P * \mathcal{N}_\sigma)$ via i.i.d. (**noisy**) samples from $P * \mathcal{N}_\sigma$

Theoretical Guarantees:

- Most results assume lower bounded density

Naive Approach - General-Purpose Estimators

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Method: Estimate $h(P * \mathcal{N}_\sigma)$ via i.i.d. (**noisy**) samples from $P * \mathcal{N}_\sigma$

Theoretical Guarantees:

- Most results assume lower bounded density $\Longrightarrow$ **Inapplicable**

Naive Approach - General-Purpose Estimators

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Method: Estimate $h(P * \mathcal{N}_\sigma)$ via i.i.d. (**noisy**) samples from $P * \mathcal{N}_\sigma$

Theoretical Guarantees:

- Most results assume lower bounded density $\implies$ **Inapplicable**
- **Applicable Here:**

Naive Approach - General-Purpose Estimators

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Method: Estimate $h(P * \mathcal{N}_\sigma)$ via i.i.d. (**noisy**) samples from $P * \mathcal{N}_\sigma$

Theoretical Guarantees:

- Most results assume lower bounded density $\Longrightarrow$ **Inapplicable**
- **Applicable Here:**
  1. **[Han-Jiao-Weissman-Wu'17]:** KDE + Best poly. approximation

Naive Approach - General-Purpose Estimators

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Method: Estimate $h(P * \mathcal{N}_\sigma)$ via i.i.d. (**noisy**) samples from $P * \mathcal{N}_\sigma$

Theoretical Guarantees:

- Most results assume lower bounded density $\implies$ **Inapplicable**
- **Applicable Here:**
  1. **[Han-Jiao-Weissman-Wu'17]:** KDE + Best poly. approximation

$\implies$ P subgaussian, $\mathsf{Risk}_{\mathsf{KDE}} \leq O\left(n^{-\frac{2}{2+d}}\right)$ (Analysis: restricted smoothness)*

$\star$ Omitting multiplicative polylogarithmic factors.

Naive Approach - General-Purpose Estimators

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Method: Estimate $h(P * \mathcal{N}_\sigma)$ via i.i.d. (**noisy**) samples from $P * \mathcal{N}_\sigma$

Theoretical Guarantees:

- Most results assume lower bounded density $\Longrightarrow$ **Inapplicable**
- **Applicable Here:**
  1. **[Han-Jiao-Weissman-Wu'17]:** KDE + Best poly. approximation

$\Longrightarrow$ P subgaussian, $\text{Risk}_{\text{KDE}} \leq O\left(n^{-\frac{2}{2+d}}\right)$ (Analysis: restricted smoothness)*

  2. **[Berrett-Samworth-Yuan'19]:** Weighted kNN (Kozachenko-Leonenko)

Naive Approach - General-Purpose Estimators

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Method: Estimate $h(P * \mathcal{N}_\sigma)$ via i.i.d. (**noisy**) samples from $P * \mathcal{N}_\sigma$

Theoretical Guarantees:

- Most results assume lower bounded density $\implies$ **Inapplicable**
- **Applicable Here:**
  1. **[Han-Jiao-Weissman-Wu'17]:** KDE + Best poly. approximation

     $\implies$ P subgaussian, $\mathsf{Risk}_{\mathsf{KDE}} \leq O\left(n^{-\frac{2}{2+d}}\right)$ (Analysis: restricted smoothness)*
  2. **[Berrett-Samworth-Yuan'19]:** Weighted kNN (Kozachenko-Leonenko)

     $\implies$ P compactly supported, $\mathsf{Risk}_{\mathsf{w-kNN}} \leq O\left(1/\sqrt{n}\right)$ (dependence on d?)

Structured Estimator

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Structured Estimator

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P*\mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Our Estimator: $\hat{h}(X^n,\sigma) \triangleq h(\hat{P}_{X^n} * \mathcal{N}_\sigma)$, where $\hat{P}_{X^n} \triangleq \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$

Structured Estimator

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P*\mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Our Estimator: $\hat{h}(X^n,\sigma) \triangleq h(\hat{P}_{X^n} * \mathcal{N}_\sigma)$, where $\hat{P}_{X^n} \triangleq \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$

Comment: $\hat{h}$ is **plug-in** est. for the functional $\mathsf{T}_\sigma(P) \triangleq h(P*\mathcal{N}_\sigma)$

Structured Estimator

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Our Estimator: $\hat{h}(X^n, \sigma) \triangleq h(\hat{P}_{X^n} * \mathcal{N}_\sigma)$, where $\hat{P}_{X^n} \triangleq \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$

Comment: $\hat{h}$ is **plug-in** est. for the functional $\mathsf{T}_\sigma(P) \triangleq h(P * \mathcal{N}_\sigma)$

Nonparametric Classes:

Structured Estimator

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Our Estimator: $\hat{h}(X^n, \sigma) \triangleq h(\hat{P}_{X^n} * \mathcal{N}_\sigma)$, where $\hat{P}_{X^n} \triangleq \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$

Comment: $\hat{h}$ is **plug-in** est. for the functional $\mathsf{T}_\sigma(P) \triangleq h(P * \mathcal{N}_\sigma)$

Nonparametric Classes:

1. **Compact Support:** $\mathcal{F}_d \triangleq \{P \,|\, \mathrm{supp}(P) \subseteq [-1,1]^d\}$

Structured Estimator

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P*\mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$*.*

Our Estimator: $\hat{h}(X^n,\sigma) \triangleq h(\hat{P}_{X^n} * \mathcal{N}_\sigma)$, where $\hat{P}_{X^n} \triangleq \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$

Comment: $\hat{h}$ is **plug-in** est. for the functional $\mathsf{T}_\sigma(P) \triangleq h(P*\mathcal{N}_\sigma)$

Nonparametric Classes:

1. **Compact Support:** $\mathcal{F}_d \triangleq \left\{P \middle| \operatorname{supp}(P) \subseteq [-1,1]^d\right\}$
2. **Subgaussian:** $\mathcal{F}^{(\mathsf{SG})}_{d,\mu,K} \triangleq \left\{P \middle| X \sim P \text{ is } K\text{-subgaussian}, \ \|\mathbb{E}X\| \leq \mu\right\}$

   where X is K-SG if $\qquad \mathbb{E}e^{\alpha^{\mathsf{T}}(X-\mathbb{E}X)} \leq e^{\frac{1}{2}K^2\|\alpha\|^2}, \qquad \forall \alpha \in \mathbb{R}^d.$

Structured Estimator

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P*\mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Our Estimator: $\hat{h}(X^n,\sigma) \triangleq h(\hat{P}_{X^n}*\mathcal{N}_\sigma)$, where $\hat{P}_{X^n} \triangleq \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$

Comment: $\hat{h}$ is **plug-in** est. for the functional $\mathsf{T}_\sigma(P) \triangleq h(P*\mathcal{N}_\sigma)$

Nonparametric Classes:

1. **Compact Support:** $\mathcal{F}_d \triangleq \{P \,|\, \mathrm{supp}(P) \subseteq [-1,1]^d\}$
2. **Subgaussian:** $\mathcal{F}^{(\mathsf{SG})}_{d,\mu,K} \triangleq \{P \,|\, X\sim P \text{ is } K\text{-subgaussian}, \|\mathbb{E}X\| \leq \mu\}$

   where X is K-SG if $\quad \mathbb{E}e^{\alpha^\mathsf{T}(X-\mathbb{E}X)} \leq e^{\frac{1}{2}K^2\|\alpha\|^2}, \quad \forall \alpha\in\mathbb{R}^d$.

⊛ **Relation:** Exists $K'>0$ such that $\mathcal{F}_d \subseteq \mathcal{F}^{(\mathsf{SG})}_{d,0,K'}$

Structured Estimator

Differential Entropy Estimation under Gaussian Convolutions

Estimate $h(P * \mathcal{N}_\sigma)$ *based on* $X^n \overset{iid}{\sim} P \in \mathcal{F}_d$ *and knowledge of* $\mathcal{N}_\sigma$.

Our Estimator: $\hat{h}(X^n, \sigma) \triangleq h(\hat{P}_{X^n} * \mathcal{N}_\sigma)$, where $\hat{P}_{X^n} \triangleq \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$

Comment: $\hat{h}$ is **plug-in** est. for the functional $\mathsf{T}_\sigma(P) \triangleq h(P * \mathcal{N}_\sigma)$

Nonparametric Classes:

1. **Compact Support:** $\mathcal{F}_d \triangleq \{P \,|\, \mathrm{supp}(P) \subseteq [-1,1]^d\}$
2. **Subgaussian:** $\mathcal{F}^{(\mathsf{SG})}_{d,\mu,K} \triangleq \{P \,|\, X \sim P$ is K-subgaussian, $\|\mathbb{E}X\| \leq \mu\}$

   where X is K-SG if $\quad \mathbb{E}e^{\alpha^{\mathsf{T}}(X-\mathbb{E}X)} \leq e^{\frac{1}{2}K^2\|\alpha\|^2}, \quad \forall \alpha \in \mathbb{R}^d$.

⊛ **Relation:** Exists $K' > 0$ such that $\mathcal{F}_d \subseteq \mathcal{F}^{(\mathsf{SG})}_{d,0,K'}$

$\Longrightarrow$ Use $\mathcal{F}_d$ for **Lower Bounds** & $\mathcal{F}^{(\mathsf{SG})}_{d,\mu,K}$ for **Upper Bounds**

Structured Estimator - Convergence Rate

Theorem (G.-Greenewald-Weed-Polyanskiy'19)

For any $\sigma > 0,\ d \geq 1$, *we have*

$$\sup_{P \in \mathcal{F}^{(\mathsf{SG})}_{d,\mu,K}} \mathbb{E}\left| h(P * \mathcal{N}_\sigma) - h(\hat{P}_{X^n} * \mathcal{N}_\sigma) \right| \leq C_{\sigma,d,\mu,K} \frac{1}{\sqrt{n}}$$

where $C_{\sigma,d,\mu,K} = O_{\sigma,\mu,K}(c^d)$ *for a constant* c.

Structured Estimator - Convergence Rate

Theorem (G.-Greenewald-Weed-Polyanskiy'19)

For any $\sigma > 0,\ d \geq 1$, *we have*

$$\sup_{P \in \mathcal{F}^{(\mathsf{SG})}_{d,\mu,K}} \mathbb{E}\left| h(P * \mathcal{N}_\sigma) - h(\hat{P}_{X^n} * \mathcal{N}_\sigma) \right| \leq C_{\sigma,d,\mu,K} \frac{1}{\sqrt{n}}$$

where $C_{\sigma,d,\mu,K} = O_{\sigma,\mu,K}(c^d)$ *for a constant* c.

Comments:

Structured Estimator - Convergence Rate

Theorem (G.-Greenewald-Weed-Polyanskiy'19)

For any $\sigma > 0,\ d \geq 1$, *we have*

$$\sup_{P \in \mathcal{F}^{(\mathsf{SG})}_{d,\mu,K}} \mathbb{E}\left| h(P * \mathcal{N}_\sigma) - h(\hat{P}_{X^n} * \mathcal{N}_\sigma) \right| \leq C_{\sigma,d,\mu,K} \frac{1}{\sqrt{n}}$$

where $C_{\sigma,d,\mu,K} = O_{\sigma,\mu,K}(c^d)$ *for a constant* c.

Comments:

- **Explicit Expression:** Enables concrete error bounds in simulations

Structured Estimator - Convergence Rate

Theorem (G.-Greenewald-Weed-Polyanskiy'19)

For any $\sigma > 0,\ d \geq 1$, *we have*

$$\sup_{P \in \mathcal{F}_{d,\mu,K}^{(\mathsf{SG})}} \mathbb{E}\left| h(P * \mathcal{N}_\sigma) - h(\hat{P}_{X^n} * \mathcal{N}_\sigma) \right| \leq C_{\sigma,d,\mu,K} \frac{1}{\sqrt{n}}$$

where $C_{\sigma,d,\mu,K} = O_{\sigma,\mu,K}(c^d)$ *for a constant* c.

Comments:

- **Explicit Expression:** Enables concrete error bounds in simulations

$$C_{\sigma,d,\mu,K} = \left(\frac{1}{\sqrt{2}} + \frac{K}{\sigma}\right)^{\frac{d}{2}} \sqrt{\frac{16}{\sigma^4}\left(2\mu^4 + 32d^2K^4 + d(d+2)\left(\frac{\sigma}{\sqrt{2}} + K\right)^4\right)} \times e^{\frac{3d}{16} + \frac{\mu^2}{4\left(K+\sigma/\sqrt{2}\right)^2}}$$

Structured Estimator - Convergence Rate

Theorem (G.-Greenewald-Weed-Polyanskiy'19)

For any $\sigma > 0,\ d \geq 1$, *we have*

$$\sup_{P \in \mathcal{F}^{(\mathsf{SG})}_{d,\mu,K}} \mathbb{E}\left| h(P * \mathcal{N}_\sigma) - h(\hat{P}_{X^n} * \mathcal{N}_\sigma) \right| \leq C_{\sigma,d,\mu,K} \frac{1}{\sqrt{n}}$$

where $C_{\sigma,d,\mu,K} = O_{\sigma,\mu,K}(c^d)$ *for a constant* c.

Comments:

- **Explicit Expression:** Enables concrete error bounds in simulations
- **Minimax Rate Optimal:** Attains parametric rate of estimation

Structured Estimator - Convergence Rate

Theorem (G.-Greenewald-Weed-Polyanskiy'19)

For any $\sigma > 0,\ d \geq 1$, *we have*

$$\sup_{P \in \mathcal{F}^{(\mathsf{SG})}_{d,\mu,K}} \mathbb{E}\left| h(P * \mathcal{N}_\sigma) - h(\hat{P}_{X^n} * \mathcal{N}_\sigma) \right| \leq C_{\sigma,d,\mu,K} \frac{1}{\sqrt{n}}$$

where $C_{\sigma,d,\mu,K} = O_{\sigma,\mu,K}(c^d)$ *for a constant* c.

Comments:

- **Explicit Expression:** Enables concrete error bounds in simulations
- **Minimax Rate Optimal:** Attains parametric rate of estimation
- **General-Purpose Methods:**

Structured Estimator - Convergence Rate

Theorem (G.-Greenewald-Weed-Polyanskiy'19)

For any $\sigma > 0,\ d \geq 1$, *we have*

$$\sup_{P \in \mathcal{F}^{(\mathsf{SG})}_{d,\mu,K}} \mathbb{E}\left| h(P * \mathcal{N}_\sigma) - h(\hat{P}_{X^n} * \mathcal{N}_\sigma) \right| \leq C_{\sigma,d,\mu,K} \frac{1}{\sqrt{n}}$$

where $C_{\sigma,d,\mu,K} = O_{\sigma,\mu,K}(c^d)$ *for a constant* c.

Comments:

- **Explicit Expression:** Enables concrete error bounds in simulations
- **Minimax Rate Optimal:** Attains parametric rate of estimation
- **General-Purpose Methods:**
  - Faster than $O\left(n^{-\frac{2}{2+d}}\right)$ of [Han-Jiao-Weissman-Wu'17]

Structured Estimator - Convergence Rate

Theorem (G.-Greenewald-Weed-Polyanskiy'19)

For any $\sigma > 0,\ d \geq 1$, *we have*

$$\sup_{P \in \mathcal{F}^{(\mathsf{SG})}_{d,\mu,K}} \mathbb{E}\left| h(P * \mathcal{N}_\sigma) - h(\hat{P}_{X^n} * \mathcal{N}_\sigma) \right| \leq C_{\sigma,d,\mu,K} \frac{1}{\sqrt{n}}$$

where $C_{\sigma,d,\mu,K} = O_{\sigma,\mu,K}(c^d)$ *for a constant* c.

Comments:

- **Explicit Expression:** Enables concrete error bounds in simulations
- **Minimax Rate Optimal:** Attains parametric rate of estimation
- **General-Purpose Methods:**
  - Faster than $O\left(n^{-\frac{2}{2+d}}\right)$ of [Han-Jiao-Weissman-Wu'17]
  - Characterized dependence on d compared to [Berrett-Samworth-Yuan'19]

Proof Outline

Lemma 1 (G.-Greenewald-Weed-Polyanskiy'19)

For any continuous RVs $U \sim p_U$ *and* $V \sim p_V$ *with* $|h(U)|, |h(V)| < \infty$*:*

$$|h(U) - h(V)| \leq \max\left\{ \int |\log p_U(z)| \cdot |p_U(z) - p_V(z)| \mathsf{d}z, \int |\log p_V(z)| \cdot |p_U(z) - p_V(z)| \mathsf{d}z \right\}.$$

Proof Outline

Lemma 1 (G.-Greenewald-Weed-Polyanskiy'19)

For any continuous RVs $U \sim p_U$ *and* $V \sim p_V$ *with* $|h(U)|, |h(V)| < \infty$*:*

$$|h(U) - h(V)| \leq \max\left\{ \int |\log p_U(z)| \cdot |p_U(z) - p_V(z)| \mathsf{d}z, \int |\log p_V(z)| \cdot |p_U(z) - p_V(z)| \mathsf{d}z \right\}.$$

Main Ideas:

Proof Outline

Lemma 1 (G.-Greenewald-Weed-Polyanskiy'19)

For any continuous RVs $U \sim p_U$ *and* $V \sim p_V$ *with* $|h(U)|, |h(V)| < \infty$*:*

$$|h(U) - h(V)| \leq \max\left\{\int |\log p_U(z)| \cdot |p_U(z) - p_V(z)| \mathsf{d}z, \int |\log p_V(z)| \cdot |p_U(z) - p_V(z)| \mathsf{d}z\right\}.$$

Main Ideas:

1. **Identity:** $h(U) - h(V) + D(p_U \| p_V) = \mathbb{E} \log \frac{p_V(V)}{p_V(U)} \leq \left|\mathbb{E} \log \frac{p_V(V)}{p_V(U)}\right|$

Proof Outline

Lemma 1 (G.-Greenewald-Weed-Polyanskiy'19)

For any continuous RVs $U \sim p_U$ *and* $V \sim p_V$ *with* $|h(U)|, |h(V)| < \infty$:

$$|h(U) - h(V)| \leq \max\left\{ \int |\log p_U(z)| \cdot |p_U(z) - p_V(z)| \mathsf{d}z, \int |\log p_V(z)| \cdot |p_U(z) - p_V(z)| \mathsf{d}z \right\}.$$

Main Ideas:

1. **Identity:** $h(U) - h(V) + D(p_U \| p_V) = \mathbb{E} \log \frac{p_V(V)}{p_V(U)} \leq \left| \mathbb{E} \log \frac{p_V(V)}{p_V(U)} \right|$
2. **Coupling:** For any coupling ν we have $\left| \mathbb{E} \log \frac{p_V(V)}{p_V(U)} \right| \leq \mathbb{E}_\nu \left| \log \frac{p_V(V)}{p_V(U)} \right|$

Proof Outline

Lemma 1 (G.-Greenewald-Weed-Polyanskiy'19)

For any continuous RVs $U \sim p_U$ and $V \sim p_V$ with $|h(U)|, |h(V)| < \infty$:

$$|h(U) - h(V)| \leq \max\left\{ \int |\log p_U(z)| \cdot |p_U(z) - p_V(z)| \mathrm{d}z, \int |\log p_V(z)| \cdot |p_U(z) - p_V(z)| \mathrm{d}z \right\}.$$

Main Ideas:

1. **Identity:** $h(U) - h(V) + D(p_U \| p_V) = \mathbb{E} \log \frac{p_V(V)}{p_V(U)} \leq \left| \mathbb{E} \log \frac{p_V(V)}{p_V(U)} \right|$
2. **Coupling:** For any coupling ν we have $\left| \mathbb{E} \log \frac{p_V(V)}{p_V(U)} \right| \leq \mathbb{E}_\nu \left| \log \frac{p_V(V)}{p_V(U)} \right|$
3. **Maximal TV Coupling:** Choose $\nu = \pi_{\mathsf{TV}}$ and analyze integral

Proof Outline

Lemma 1 (G.-Greenewald-Weed-Polyanskiy'19)

For any continuous RVs $U \sim p_U$ and $V \sim p_V$ with $|h(U)|, |h(V)| < \infty$:

$$|h(U) - h(V)| \leq \max\left\{ \int |\log p_U(z)| \cdot |p_U(z) - p_V(z)| \mathsf{d}z, \int |\log p_V(z)| \cdot |p_U(z) - p_V(z)| \mathsf{d}z \right\}.$$

Main Ideas:

1. **Identity:** $h(U) - h(V) + D(p_U \| p_V) = \mathbb{E} \log \frac{p_V(V)}{p_V(U)} \leq \left| \mathbb{E} \log \frac{p_V(V)}{p_V(U)} \right|$
2. **Coupling:** For any coupling ν we have $\left| \mathbb{E} \log \frac{p_V(V)}{p_V(U)} \right| \leq \mathbb{E}_\nu \left| \log \frac{p_V(V)}{p_V(U)} \right|$
3. **Maximal TV Coupling:** Choose $\nu = \pi_{\mathsf{TV}}$ and analyze integral

$$\left(\forall \text{ measurable } g: \quad |\mathbb{E}g(U) - \mathbb{E}g(V)| \leq \int |g(z)| \cdot |p_U(z) - p_V(z)| \mathsf{d}z \right)$$

Proof Outline (Cntd.)

Lemma 1: For any $P \in \mathcal{F}^{(\mathsf{SG})}_{d,\mu,K}$

$$\mathbb{E}\Big|h(P*\mathcal{N}_\sigma) - h(\hat{P}_{X^n}*\mathcal{N}_\sigma)\Big| \leq \mathbb{E}\int \max\Big\{\big|\log\tilde{q}(z)\big|, \big|\log\tilde{r}_{X^n}(z)\big|\Big\}\big|q(z)-r_{X^n}(z)\big|\mathsf{d}z$$

Proof Outline (Cntd.)

Lemma 1: For any $P \in \mathcal{F}_{d,\mu,K}^{(\mathsf{SG})}$

$$\mathbb{E}\Big|h(P * \mathcal{N}_\sigma) - h(\hat{P}_{X^n} * \mathcal{N}_\sigma)\Big| \leq \mathbb{E}\int \max\Big\{\big|\log \tilde{q}(z)\big|, \big|\log \tilde{r}_{X^n}(z)\big|\Big\}\big|q(z) - r_{X^n}(z)\big|\mathsf{d}z$$

where: q is PDF of $P * \mathcal{N}_\sigma$; r_{X^n} is PDF of $\hat{P}_{X^n} * \mathcal{N}_\sigma$

Proof Outline (Cntd.)

Lemma 1: For any $P \in \mathcal{F}^{(\mathsf{SG})}_{d,\mu,K}$

$$\mathbb{E}\left|h(P*\mathcal{N}_\sigma) - h(\hat{P}_{X^n}*\mathcal{N}_\sigma)\right| \leq \mathbb{E}\int \max\left\{\left|\log \tilde{q}(z)\right|, \left|\log \tilde{r}_{X^n}(z)\right|\right\}\left|q(z) - r_{X^n}(z)\right| \mathsf{d}z$$

where: q is PDF of $P*\mathcal{N}_\sigma$; r_{X^n} is PDF of $\hat{P}_{X^n}*\mathcal{N}_\sigma$

$$\tilde{q} \triangleq \frac{q}{\|q\|_\infty} = \frac{q}{(2\pi\sigma)^{d/2}} \quad ; \quad \tilde{r}_{X^n} \triangleq \frac{r_{X^n}}{\|r_{X^n}\|_\infty} = \frac{r_{X^n}}{(2\pi\sigma)^{d/2}}$$

Proof Outline (Cntd.)

Lemma 1: For any $P \in \mathcal{F}^{(\mathsf{SG})}_{d,\mu,K}$

$$\mathbb{E}\Big|h(P*\mathcal{N}_\sigma) - h(\hat{P}_{X^n}*\mathcal{N}_\sigma)\Big| \leq \mathbb{E}\int \max\Big\{\big|\log\tilde{q}(z)\big|,\big|\log\tilde{r}_{X^n}(z)\big|\Big\}\big|q(z)-r_{X^n}(z)\big|\mathsf{d}z$$

where: q is PDF of $P*\mathcal{N}_\sigma$; r_{X^n} is PDF of $\hat{P}_{X^n}*\mathcal{N}_\sigma$

$$\tilde{q} \triangleq \frac{q}{\|q\|_\infty} = \frac{q}{(2\pi\sigma)^{d/2}} \quad ; \quad \tilde{r}_{X^n} \triangleq \frac{r_{X^n}}{\|r_{X^n}\|_\infty} = \frac{r_{X^n}}{(2\pi\sigma)^{d/2}}$$

Lemma 2 (G.-Greenewald-Weed-Polyanskiy'19)

Let $X\sim P$. For all $z\in\mathbb{R}^d$ it holds that

$$\mathbb{E}\left[\max\left\{\big(\log\tilde{q}(z)\big)^2,\big(\log\tilde{r}_{X^n}(z)\big)^2\right\}\right] \leq \frac{1}{2\sigma^4}\mathbb{E}\|z-X\|^4$$

Proof Outline (Cntd. 2)

Finalization: Let φ_a be the PDF of $\mathcal{N}\left(0, \frac{1}{2a}\mathrm{I}_d\right)$, for $a = \frac{1}{4(K+\sigma/\sqrt{2})^2}$

Proof Outline (Cntd. 2)

Finalization: Let φ_a be the PDF of $\mathcal{N}\left(0, \frac{1}{2a}\mathrm{I}_d\right)$, for $a = \frac{1}{4(K+\sigma/\sqrt{2})^2}$

Cauchy-Schwarz:

$$\left(\mathbb{E}\int \max\left\{\left|\log \tilde{q}(z)\right|, \left|\log \tilde{r}_{X^n}(z)\right|\right\}\left|q(z) - r_{X^n}(z)\right|\mathsf{d}z\right)^2$$

$$\leq \int \mathbb{E}\frac{\left(q(z) - r_{X^n}(z)\right)^2}{\varphi_a(z)}\mathsf{d}z \cdot \int \mathbb{E}\left[\max\left\{\left(\log \tilde{q}(z)\right)^2, \left(\log \tilde{r}_{X^n}(z)\right)^2\right\}\right]\varphi_a(z)\mathsf{d}z$$

Proof Outline (Cntd. 2)

Finalization: Let φ_a be the PDF of $\mathcal{N}\left(0, \frac{1}{2a}\mathrm{I}_d\right)$, for $a = \frac{1}{4(K+\sigma/\sqrt{2})^2}$

Cauchy-Schwarz:

$$\left(\mathbb{E}\int \max\left\{\left|\log\tilde{q}(z)\right|, \left|\log\tilde{r}_{X^n}(z)\right|\right\}\left|q(z)-r_{X^n}(z)\right|\mathsf{d}z\right)^2$$

$$\leq \int\mathbb{E}\frac{\big(q(z)-r_{X^n}(z)\big)^2}{\varphi_a(z)}\mathsf{d}z \cdot \underbrace{\int\mathbb{E}\Big[\max\Big\{\big(\log\tilde{q}(z)\big)^2, \big(\log\tilde{r}_{X^n}(z)\big)^2\Big\}\Big]\varphi_a(z)\mathsf{d}z}_{\text{Lemma 2} + K\text{-subgaussianity of } X + \text{Gaussian moments}}$$

Proof Outline (Cntd. 2)

Finalization: Let φ_a be the PDF of $\mathcal{N}\left(0, \frac{1}{2a}\mathrm{I}_d\right)$, for $a = \frac{1}{4(K+\sigma/\sqrt{2})^2}$

Cauchy-Schwarz:

$$\left(\mathbb{E}\int \max\left\{\left|\log\tilde{q}(z)\right|, \left|\log\tilde{r}_{X^n}(z)\right|\right\}\left|q(z) - r_{X^n}(z)\right|\mathsf{d}z\right)^2$$

$$\leq \int \mathbb{E}\frac{\big(q(z) - r_{X^n}(z)\big)^2}{\varphi_a(z)}\mathsf{d}z \cdot \underbrace{\int \mathbb{E}\Big[\max\Big\{\big(\log\tilde{q}(z)\big)^2, \big(\log\tilde{r}_{X^n}(z)\big)^2\Big\}\Big]\varphi_a(z)\mathsf{d}z}_{\substack{\text{Lemma 2} \\ + \\ K\text{-subgaussianity of } X \\ + \\ \text{Gaussian moments}}}$$

$$\underbrace{\text{Lemma 2} + K\text{-subgaussianity of } X + \text{Gaussian moments}}_{\lesssim \frac{d^2}{\sigma^4}}$$

Proof Outline (Cntd. 2)

Finalization: Let φ_a be the PDF of $\mathcal{N}\left(0, \frac{1}{2a}\mathrm{I}_d\right)$, for $a = \frac{1}{4(K+\sigma/\sqrt{2})^2}$

Cauchy-Schwarz:

$$\left(\mathbb{E}\int \max\left\{\left|\log\tilde{q}(z)\right|, \left|\log\tilde{r}_{X^n}(z)\right|\right\}\left|q(z)-r_{X^n}(z)\right|\mathsf{d}z\right)^2$$

$$\leq \underbrace{\int \mathbb{E}\frac{\left(q(z)-r_{X^n}(z)\right)^2}{\varphi_a(z)}\mathsf{d}z}_{} \cdot \underbrace{\int \mathbb{E}\left[\max\left\{\left(\log\tilde{q}(z)\right)^2, \left(\log\tilde{r}_{X^n}(z)\right)^2\right\}\right]\varphi_a(z)\mathsf{d}z}_{}$$

$r_{X^n}(z)$ sum of iid: $\mathbb{E}r_{X^n}(z) = q(z)$

Lemma 2

$+$

K-subgaussianity of X

$+$

Gaussian moments

$\lesssim \frac{d^2}{\sigma^4}$

Proof Outline (Cntd. 2)

Finalization: Let φ_a be the PDF of $\mathcal{N}\left(0, \frac{1}{2a}\mathrm{I}_d\right)$, for $a = \frac{1}{4(K+\sigma/\sqrt{2})^2}$

Cauchy-Schwarz:

$$\left(\mathbb{E}\int \max\left\{\left|\log\tilde{q}(z)\right|, \left|\log\tilde{r}_{X^n}(z)\right|\right\}\left|q(z) - r_{X^n}(z)\right|\mathrm{d}z\right)^2$$

$$\leq \underbrace{\int \mathbb{E}\frac{\left(q(z) - r_{X^n}(z)\right)^2}{\varphi_a(z)}\mathrm{d}z}\cdot\underbrace{\int \mathbb{E}\left[\max\left\{\left(\log\tilde{q}(z)\right)^2, \left(\log\tilde{r}_{X^n}(z)\right)^2\right\}\right]\varphi_a(z)\mathrm{d}z}$$

$r_{X^n}(z)$ sum of iid: $\mathbb{E}r_{X^n}(z) = q(z)$

$\Downarrow$

$\mathrm{var}\left(r_{X^n}(z)\right) \leq \frac{c}{n}\mathbb{E}e^{-\frac{1}{\sigma^2}\|z-X\|^2}$

Lemma 2

$+$

K-subgaussianity of X

$+$

Gaussian moments

$\underbrace{\qquad}_{\lesssim \frac{d^2}{\sigma^4}}$

Proof Outline (Cntd. 2)

Finalization: Let φ_a be the PDF of $\mathcal{N}\left(0, \frac{1}{2a}\mathrm{I}_d\right)$, for $a = \frac{1}{4(K+\sigma/\sqrt{2})^2}$

Cauchy-Schwarz:

$$\left(\mathbb{E}\int \max\left\{\left|\log\tilde{q}(z)\right|, \left|\log\tilde{r}_{X^n}(z)\right|\right\}\left|q(z)-r_{X^n}(z)\right|\mathsf{d}z\right)^2$$

$$\leq \underbrace{\int \mathbb{E}\frac{\left(q(z)-r_{X^n}(z)\right)^2}{\varphi_a(z)}\mathsf{d}z}\cdot\underbrace{\int \mathbb{E}\left[\max\left\{\left(\log\tilde{q}(z)\right)^2, \left(\log\tilde{r}_{X^n}(z)\right)^2\right\}\right]\varphi_a(z)\mathsf{d}z}$$

$r_{X^n}(z)$ sum of iid: $\mathbb{E}r_{X^n}(z) = q(z)$

$\Downarrow$

$\mathsf{var}\big(r_{X^n}(z)\big) \leq \frac{c}{n}\mathbb{E}e^{-\frac{1}{\sigma^2}\|z-X\|^2}$

$\Downarrow$

Insert back + subgaussianity

Lemma 2

$+$

K-subgaussianity of X

$+$

Gaussian moments

$\lesssim \frac{d^2}{\sigma^4}$

Proof Outline (Cntd. 2)

Finalization: Let φ_a be the PDF of $\mathcal{N}\left(0, \frac{1}{2a}\mathrm{I}_d\right)$, for $a = \frac{1}{4(K+\sigma/\sqrt{2})^2}$

Cauchy-Schwarz:

$$\left(\mathbb{E}\int \max\left\{\left|\log \tilde{q}(z)\right|, \left|\log \tilde{r}_{X^n}(z)\right|\right\}\left|q(z) - r_{X^n}(z)\right|\mathrm{d}z\right)^2$$

$$\leq \underbrace{\int \mathbb{E}\frac{\left(q(z) - r_{X^n}(z)\right)^2}{\varphi_a(z)}\mathrm{d}z}\ \cdot\ \underbrace{\int \mathbb{E}\left[\max\left\{\left(\log \tilde{q}(z)\right)^2, \left(\log \tilde{r}_{X^n}(z)\right)^2\right\}\right]\varphi_a(z)\mathrm{d}z}$$

$r_{X^n}(z)$ sum of iid: $\mathbb{E}r_{X^n}(z) = q(z)$

$\Downarrow$

$\mathrm{var}\left(r_{X^n}(z)\right) \leq \frac{c}{n}\mathbb{E}e^{-\frac{1}{\sigma^2}\|z-X\|^2}$

$\Downarrow$

$\underbrace{\text{Insert back + subgaussianity}}$

$\lesssim \frac{c^d}{n}$

Lemma 2

$+$

K-subgaussianity of X

$+$

$\underbrace{\text{Gaussian moments}}$

$\lesssim \frac{d^2}{\sigma^4}$

Is Exponentiality in Dimension Necessary?

Is Exponentiality in Dimension Necessary?

Theorem (G.-Greenewald-Polyanskiy'18)

For any $\sigma > 0$, sufficiently large d and sufficiently small $\eta > 0$, we have $n^\star(\eta, \sigma, \mathcal{F}_d) = \Omega\left(\frac{2^{\gamma(\sigma)d}}{\eta d}\right)$, where $\gamma(\sigma) > 0$ is monotonically decreasing in σ.

Is Exponentiality in Dimension Necessary?

Theorem (G.-Greenewald-Polyanskiy'18)

For any $\sigma > 0$, sufficiently large d and sufficiently small $\eta > 0$, we have $n^\star(\eta, \sigma, \mathcal{F}_d) = \Omega\left(\frac{2^{\gamma(\sigma)d}}{\eta d}\right)$, where $\gamma(\sigma) > 0$ is monotonically decreasing in σ.

Comments:

Is Exponentiality in Dimension Necessary?

Theorem (G.-Greenewald-Polyanskiy'18)

For any $\sigma > 0$, sufficiently large d and sufficiently small $\eta > 0$, we have $n^\star(\eta, \sigma, \mathcal{F}_d) = \Omega\left(\frac{2^{\gamma(\sigma)d}}{\eta d}\right)$, where $\gamma(\sigma) > 0$ is monotonically decreasing in σ.

Comments:

- The $O\left(\frac{c^d}{\sqrt{n}}\right)$ rate attained by the plugin estimator is sharp

Is Exponentiality in Dimension Necessary?

Theorem (G.-Greenewald-Polyanskiy'18)

For any $\sigma > 0$, sufficiently large d and sufficiently small $\eta > 0$, we have $n^\star(\eta, \sigma, \mathcal{F}_d) = \Omega\left(\frac{2^{\gamma(\sigma)d}}{\eta d}\right)$, where $\gamma(\sigma) > 0$ is monotonically decreasing in σ.

Comments:

- The $O\left(\frac{c^d}{\sqrt{n}}\right)$ rate attained by the plugin estimator is sharp
- Estimation is harder for smaller σ

Is Exponentiality in Dimension Necessary?

Theorem (G.-Greenewald-Polyanskiy'18)

For any $\sigma > 0$, sufficiently large d and sufficiently small $\eta > 0$, we have $n^\star(\eta, \sigma, \mathcal{F}_d) = \Omega\left(\frac{2^{\gamma(\sigma)d}}{\eta d}\right)$, where $\gamma(\sigma) > 0$ is monotonically decreasing in σ.

Comments:

- The $O\left(\frac{c^d}{\sqrt{n}}\right)$ rate attained by the plugin estimator is sharp
- Estimation is harder for smaller σ
- **Proof Idea:**

Is Exponentiality in Dimension Necessary?

Theorem (G.-Greenewald-Polyanskiy'18)

For any $\sigma > 0$, sufficiently large d and sufficiently small $\eta > 0$, we have $n^\star(\eta,\sigma,\mathcal{F}_d) = \Omega\left(\frac{2^{\gamma(\sigma)d}}{\eta d}\right)$, where $\gamma(\sigma) > 0$ is monotonically decreasing in σ.

Comments:

- The $O\left(\frac{c^d}{\sqrt{n}}\right)$ rate attained by the plugin estimator is sharp
- Estimation is harder for smaller σ
- **Proof Idea:**
  - Relate $h(P * \mathcal{N}_\sigma)$ to Shannon entropy $H(Q)$

    $\mathrm{supp}(Q) =$ peak-constrained AWGN capacity achieving codebook $\mathcal{C}_d$

Is Exponentiality in Dimension Necessary?

Theorem (G.-Greenewald-Polyanskiy'18)

For any $\sigma > 0$, sufficiently large d and sufficiently small $\eta > 0$, we have $n^\star(\eta,\sigma,\mathcal{F}_d) = \Omega\left(\frac{2^{\gamma(\sigma)d}}{\eta d}\right)$, where $\gamma(\sigma) > 0$ is monotonically decreasing in σ.

Comments:

- The $O\left(\frac{c^d}{\sqrt{n}}\right)$ rate attained by the plugin estimator is sharp
- Estimation is harder for smaller σ
- **Proof Idea:**
  - Relate $h(P * \mathcal{N}_\sigma)$ to Shannon entropy $H(Q)$

    $\mathrm{supp}(Q) =$ peak-constrained AWGN capacity achieving codebook $\mathcal{C}_d$
  - Shannon entropy est. Sample complexity $\Omega\left(\frac{|\mathcal{C}_d|}{\eta \log |\mathcal{C}_d|}\right)$ [Wu-Yang'16]

Simulations - Synthetic Experiments

Comparison: General-purpose est. accessing sample of $X + Z \sim P * \mathcal{N}_\sigma$

Simulations - Synthetic Experiments

Comparison: General-purpose est. accessing sample of $X+Z \sim P * \mathcal{N}_\sigma$

1. LOO KDE Estimator from [Kandasamy et al.'15]

Simulations - Synthetic Experiments

Comparison: General-purpose est. accessing sample of $X + Z \sim P * \mathcal{N}_\sigma$

1. LOO KDE Estimator from [Kandasamy et al.'15]
2. Kozachenko-Leonenko (KL) kNN Estimator [Kozachenko-Leonenko'87]

Simulations - Synthetic Experiments

Comparison: General-purpose est. accessing sample of $X + Z \sim P * \mathcal{N}_\sigma$

1. LOO KDE Estimator from [Kandasamy et al.'15]
2. Kozachenko-Leonenko (KL) kNN Estimator [Kozachenko-Leonenko'87]
3. Weighted KL (wKL) Estimator from [Berrett-Samworth-Yuan'19]

Simulations - Synthetic Experiments

Comparison: General-purpose est. accessing sample of $X+Z \sim P * \mathcal{N}_\sigma$

1. LOO KDE Estimator from [Kandasamy et al.'15]
2. Kozachenko-Leonenko (KL) kNN Estimator [Kozachenko-Leonenko'87]
3. Weighted KL (wKL) Estimator from [Berrett-Samworth-Yuan'19]

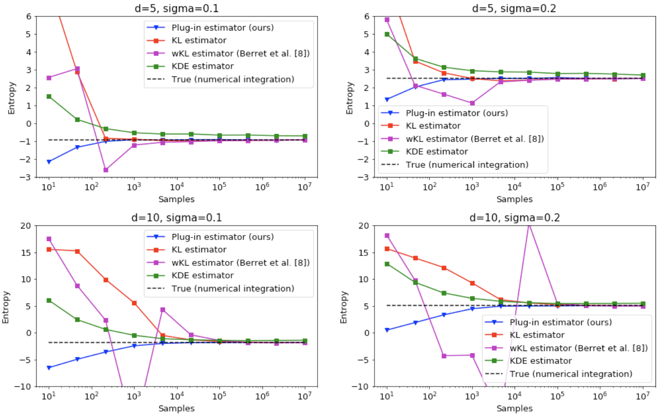
Bdd. Support: $P =$ truncated d-dim. Gaussian mixture (centers $\{\pm 1\}^d$)

Simulations - Synthetic Experiments

Comparison: General-purpose est. accessing sample of $X + Z \sim P * \mathcal{N}_\sigma$

1. LOO KDE Estimator from [Kandasamy et al.'15]
2. Kozachenko-Leonenko (KL) kNN Estimator [Kozachenko-Leonenko'87]
3. Weighted KL (wKL) Estimator from [Berrett-Samworth-Yuan'19]

Bdd. Support: $P =$ truncated d-dim. Gaussian mixture (centers $\{\pm 1\}^d$)

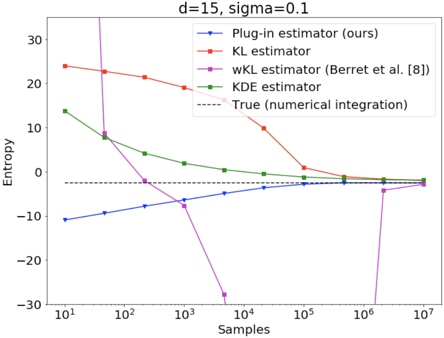


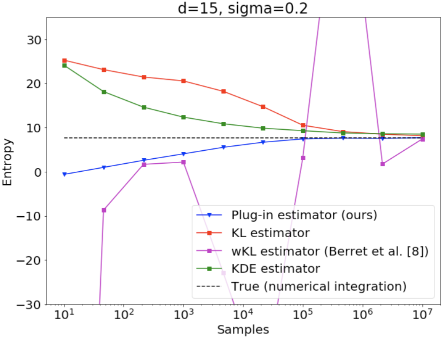
Simulations - Synthetic Experiments

Unbounded Support: $P =$ (untrunc.) d-dim. 2^d-modal Gaussian mixture

Simulations - Synthetic Experiments

Unbounded Support: $P =$ (untrunc.) d-dim. 2^d-modal Gaussian mixture





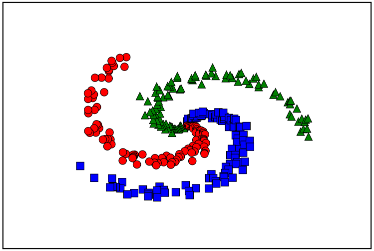
Simulations - Noisy Deep Neural Network Example

Setup: Noisy DNN for spiral dataset classification

Simulations - Noisy Deep Neural Network Example

Setup: Noisy DNN for spiral dataset classification

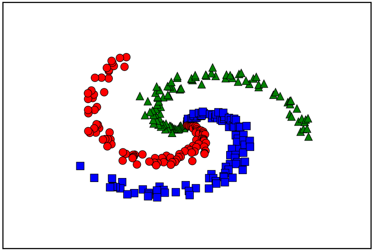
- **Dataset:** 2-dimensional 3-class spiral dataset



Simulations - Noisy Deep Neural Network Example

Setup: Noisy DNN for spiral dataset classification

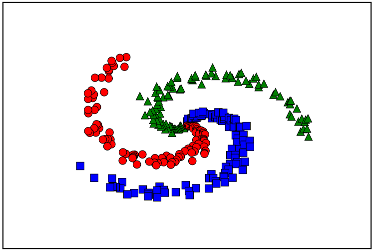
- **Dataset:** 2-dimensional 3-class spiral dataset
- **Network:** 2–8–9–10–3 fully connected noisy ($\sigma = 0.2$) tanh DNN



Simulations - Noisy Deep Neural Network Example

Setup: Noisy DNN for spiral dataset classification

- **Dataset:** 2-dimensional 3-class spiral dataset
- **Network:** 2–8–9–10–3 fully connected noisy ($\sigma = 0.2$) tanh DNN
- **Classification:** Trained to 98% test accuracy

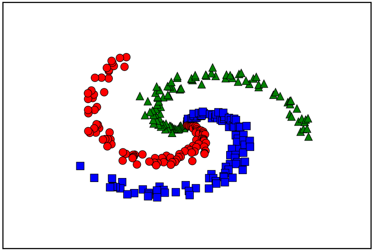


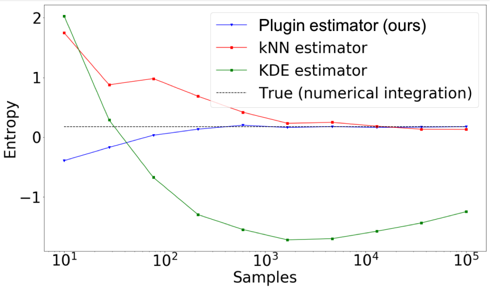
Simulations - Noisy Deep Neural Network Example

Setup: Noisy DNN for spiral dataset classification

- **Dataset:** 2-dimensional 3-class spiral dataset
- **Network:** 2–8–9–10–3 fully connected noisy ($\sigma = 0.2$) tanh DNN
- **Classification:** Trained to 98% test accuracy

⊛ Estimating the entropy of 10-dimensional layer





Summary and Concluding Remarks

Paper available at arXiv:1810.11589

- **Differential Entropy Estimation under Gaussian Convolutions:**

Summary and Concluding Remarks

Paper available at arXiv:1810.11589

- **Differential Entropy Estimation under Gaussian Convolutions:**
  - New high-dimensional & nonparametric functional estimation problem

Summary and Concluding Remarks

Paper available at arXiv:1810.11589

- **Differential Entropy Estimation under Gaussian Convolutions:**
  - New high-dimensional & nonparametric functional estimation problem
- **Intrinsically Difficult Problem:**

Summary and Concluding Remarks

Paper available at arXiv:1810.11589

- **Differential Entropy Estimation under Gaussian Convolutions:**
  - New high-dimensional & nonparametric functional estimation problem
- **Intrinsically Difficult Problem:**
  - Sample complexity is exponential in dimension

Summary and Concluding Remarks

Paper available at arXiv:1810.11589

- **Differential Entropy Estimation under Gaussian Convolutions:**
  - New high-dimensional & nonparametric functional estimation problem
- **Intrinsically Difficult Problem:**
  - Sample complexity is exponential in dimension
- **Plug-in Estimator:**

Summary and Concluding Remarks

Paper available at arXiv:1810.11589

- **Differential Entropy Estimation under Gaussian Convolutions:**
  - New high-dimensional & nonparametric functional estimation problem
- **Intrinsically Difficult Problem:**
  - Sample complexity is exponential in dimension
- **Plug-in Estimator:**
  - Attains parametric estimation rate $O\left(\frac{c^d}{\sqrt{n}}\right)$

Summary and Concluding Remarks

Paper available at arXiv:1810.11589

- **Differential Entropy Estimation under Gaussian Convolutions:**
  - New high-dimensional & nonparametric functional estimation problem
- **Intrinsically Difficult Problem:**
  - Sample complexity is exponential in dimension
- **Plug-in Estimator:**
  - Attains parametric estimation rate $O\left(\frac{c^d}{\sqrt{n}}\right)$
  - Empirically outperforms general-purpose estimation via 'noisy' samples

Summary and Concluding Remarks

Paper available at arXiv:1810.11589

- **Differential Entropy Estimation under Gaussian Convolutions:**
  - New high-dimensional & nonparametric functional estimation problem
- **Intrinsically Difficult Problem:**
  - Sample complexity is exponential in dimension
- **Plug-in Estimator:**
  - Attains parametric estimation rate $O\left(\frac{c^d}{\sqrt{n}}\right)$
  - Empirically outperforms general-purpose estimation via 'noisy' samples
- **arXiv:1810.05728**: Study MI trends during DNN training (estimation)

Summary and Concluding Remarks

Paper available at arXiv:1810.11589

- **Differential Entropy Estimation under Gaussian Convolutions:**
  - New high-dimensional & nonparametric functional estimation problem
- **Intrinsically Difficult Problem:**
  - Sample complexity is exponential in dimension
- **Plug-in Estimator:**
  - Attains parametric estimation rate $O\left(\frac{c^d}{\sqrt{n}}\right)$
  - Empirically outperforms general-purpose estimation via 'noisy' samples
- **arXiv:1810.05728**: Study MI trends during DNN training (estimation)
- **Future Work:** Non-Gaussian conv.? Multiplicative noise (Dropout)?

Summary and Concluding Remarks

Paper available at arXiv:1810.11589

- **Differential Entropy Estimation under Gaussian Convolutions:**
  - New high-dimensional & nonparametric functional estimation problem
- **Intrinsically Difficult Problem:**
  - Sample complexity is exponential in dimension
- **Plug-in Estimator:**
  - Attains parametric estimation rate $O\left(\frac{c^d}{\sqrt{n}}\right)$
  - Empirically outperforms general-purpose estimation via 'noisy' samples
- **arXiv:1810.05728**: Study MI trends during DNN training (estimation)
- **Future Work:** Non-Gaussian conv.? Multiplicative noise (Dropout)?

Thank you!