A moment-matching problem with an entropy constraint

Adway Girish

EPFL

EPFL Information Processing Group

July 2, 2025

A moment-matching problem with an entropy constraint

Adway Girish

thanks to Shlomo Shamai and Emre Telatar

(Girish–Shamai–Telatar, ISIT 2025)

EPFL

EPFL Information Processing Group

July 2, 2025

Prelude: computing integrals numerically

- $\mathbb{E}[g(W)] = \int_{\mathbb{R}} g(w) f_W(w) \, \mathrm{d}w$

Prelude: computing integrals numerically

- $\mathbb{E}[g(W)] = \int_{\mathbb{R}} g(w) f_W(w) \, \mathrm{d}w \approx \sum_{i=1}^{m} p_i g(x_i)$

Prelude: computing integrals numerically

- $\mathbb{E}[g(W)] = \int_{\mathbb{R}} g(w) f_W(w) \, \mathrm{d}w \approx \sum_{i=1}^{m} p_i g(x_i) = \mathbb{E}[g(X)]$

Prelude: computing integrals numerically

- $\mathbb{E}[g(W)] = \int_{\mathbb{R}} g(w) f_W(w)\, \mathrm{d}w \approx \sum_{i=1}^{m} p_i g(x_i) = \mathbb{E}[g(X)]$
- if X has $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, \ldots, k$

Prelude: computing integrals numerically

- $\mathbb{E}[g(W)] = \int_{\mathbb{R}} g(w) f_W(w)\,\mathrm{d}w \approx \sum_{i=1}^{m} p_i g(x_i) = \mathbb{E}[g(X)]$

- if X has $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, \ldots, k$, then

$$\int_{\mathbb{R}} g(w) f_W(w)\,\mathrm{d}w = \sum_{i=1}^{m} p_i g(x_i)$$

  for polynomial g of degree $\leq k$

Prelude: computing integrals numerically

- $\mathbb{E}[g(W)] = \int_{\mathbb{R}} g(w) f_W(w)\, \mathrm{d}w \approx \sum_{i=1}^{m} p_i g(x_i) = \mathbb{E}[g(X)]$

- if X has $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, \ldots, k$, then

$$\int_{\mathbb{R}} g(w) f_W(w)\, \mathrm{d}w = \sum_{i=1}^{m} p_i g(x_i)$$

  for polynomial g of degree $\leq k$

- given f_W and constraints on X, how large can k be?

Outline

1. Moment problems
2. Entropy-constrained version
3. Application to a hypothesis testing problem

Outline

1. Moment problems
2. Entropy-constrained version
3. Application to a hypothesis testing problem

Classical moment problem

Classical moment problem

- Q: Given $s_1, s_2, s_3, \ldots$, does there exist X on $\mathbb{R}$ such that $\mathbb{E}[X^n] = s_n$ for $n = 1, 2, 3, \ldots$?

Classical moment problem

- Q: Given $s_1, s_2, s_3, \ldots$, does there exist X on $\mathbb{R}$ such that $\mathbb{E}[X^n] = s_n$ for $n = 1, 2, 3, \ldots$?

  A: iff $\mathrm{H}_n(s_1, \ldots, s_{2n}) = \begin{pmatrix} 1 & s_1 & \ldots & s_n \\ s_1 & s_2 & \ldots & s_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_n & s_{n+1} & \ldots & s_{2n} \end{pmatrix} \succeq 0$ for $n = 1, 2, 3, \ldots$

Classical moment problem

- Q: Given $s_1, s_2, s_3, \ldots$, does there exist X on $\mathbb{R}$ such that $\mathbb{E}[X^n] = s_n$ for $n = 1, 2, 3, \ldots$?

  A: iff $\mathrm{H}_n(s_1, \ldots, s_{2n}) = \begin{pmatrix} 1 & s_1 & \ldots & s_n \\ s_1 & s_2 & \ldots & s_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_n & s_{n+1} & \ldots & s_{2n} \end{pmatrix} \succeq 0$ for $n = 1, 2, 3, \ldots$;

  infinite support iff $\mathrm{H}_n \succ 0$ for all n

Truncated moment problem

- Q: Given $s_1, s_2, s_3, \ldots$, does there exist X on $\mathbb{R}$ such that $\mathbb{E}[X^n] = s_n$ for $n = 1, 2, 3, \ldots$?

  A: iff $\mathrm{H}_n(s_1, \ldots, s_{2n}) = \begin{pmatrix} 1 & s_1 & \ldots & s_n \\ s_1 & s_2 & \ldots & s_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_n & s_{n+1} & \ldots & s_{2n} \end{pmatrix} \succeq 0$ for $n = 1, 2, 3, \ldots$;

  infinite support iff $\mathrm{H}_n \succ 0$ for all n

- Q: Given $s_1, s_2, s_3, \ldots, s_k$, does there exist X on $\mathbb{R}$ such that $\mathbb{E}[X^n] = s_n$ for $n = 1, 2, 3, \ldots, k$?

Truncated moment problem

- Q: Given $s_1, s_2, s_3, \ldots$, does there exist X on $\mathbb{R}$ such that $\mathbb{E}[X^n] = s_n$ for $n = 1, 2, 3, \ldots$?

  A: iff $\mathrm{H}_n(s_1, \ldots, s_{2n}) = \begin{pmatrix} 1 & s_1 & \ldots & s_n \\ s_1 & s_2 & \ldots & s_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_n & s_{n+1} & \ldots & s_{2n} \end{pmatrix} \succeq 0$ for $n = 1, 2, 3, \ldots$;

  infinite support iff $\mathrm{H}_n \succ 0$ for all n

- Q: Given $s_1, s_2, s_3, \ldots, s_k$, does there exist X on $\mathbb{R}$ such that $\mathbb{E}[X^n] = s_n$ for $n = 1, 2, 3, \ldots, k$?

  A: iff $(k$ odd$)$

  $(k$ even$)$

Truncated moment problem

- Q: Given $s_1, s_2, s_3, \ldots$, does there exist X on $\mathbb{R}$ such that $\mathbb{E}[X^n] = s_n$ for $n = 1, 2, 3, \ldots$?

  A: iff $\mathrm{H}_n(s_1, \ldots, s_{2n}) = \begin{pmatrix} 1 & s_1 & \ldots & s_n \\ s_1 & s_2 & \ldots & s_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_n & s_{n+1} & \ldots & s_{2n} \end{pmatrix} \succeq 0$ for $n = 1, 2, 3, \ldots$;

  infinite support iff $\mathrm{H}_n \succ 0$ for all n

- Q: Given $s_1, s_2, s_3, \ldots, s_k$, does there exist X on $\mathbb{R}$ such that $\mathbb{E}[X^n] = s_n$ for $n = 1, 2, 3, \ldots, k$?

  A: iff $(k \text{ odd})$ there exists $\tilde{s}_{k+1}$ such that $\mathrm{H}_{\frac{k+1}{2}}(s_1, \ldots, s_k, \tilde{s}_{k+1}) \succeq 0$

  $(k \text{ even})$

Truncated moment problem

- Q: Given $s_1, s_2, s_3, \ldots$, does there exist X on $\mathbb{R}$ such that $\mathbb{E}[X^n] = s_n$ for $n = 1, 2, 3, \ldots$?

  A: iff $\mathrm{H}_n(s_1, \ldots, s_{2n}) = \begin{pmatrix} 1 & s_1 & \ldots & s_n \\ s_1 & s_2 & \ldots & s_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_n & s_{n+1} & \ldots & s_{2n} \end{pmatrix} \succeq 0$ for $n = 1, 2, 3, \ldots$;

  infinite support iff $\mathrm{H}_n \succ 0$ for all n

- Q: Given $s_1, s_2, s_3, \ldots, s_k$, does there exist X on $\mathbb{R}$ such that $\mathbb{E}[X^n] = s_n$ for $n = 1, 2, 3, \ldots, k$?

  A: iff (k odd) there exists $\tilde{s}_{k+1}$ such that $\mathrm{H}_{\frac{k+1}{2}}(s_1, \ldots, s_k, \tilde{s}_{k+1}) \succeq 0$

  (k even) there exist $\tilde{s}_{k+1}, \tilde{s}_{k+2}$ such that $\mathrm{H}_{\frac{k}{2}+1}(s_1, \ldots, s_k, \tilde{s}_{k+1}, \tilde{s}_{k+2}) \succeq 0$

Truncated moment problem

- Q: Given $s_1, s_2, s_3, \ldots$, does there exist X on $\mathbb{R}$ such that $\mathbb{E}[X^n] = s_n$ for $n = 1, 2, 3, \ldots$?

  A: iff $\mathrm{H}_n(s_1, \ldots, s_{2n}) = \begin{pmatrix} 1 & s_1 & \ldots & s_n \\ s_1 & s_2 & \ldots & s_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_n & s_{n+1} & \ldots & s_{2n} \end{pmatrix} \succeq 0$ for $n = 1, 2, 3, \ldots$;

  infinite support iff $\mathrm{H}_n \succ 0$ for all n

- Q: Given $s_1, s_2, s_3, \ldots, s_k$, does there exist X on $\mathbb{R}$ such that $\mathbb{E}[X^n] = s_n$ for $n = 1, 2, 3, \ldots, k$?

  A: iff (k odd) there exists $\tilde{s}_{k+1}$ such that $\mathrm{H}_{\frac{k+1}{2}}(s_1, \ldots, s_k, \tilde{s}_{k+1}) \succeq 0$

  (k even) there exist $\tilde{s}_{k+1}, \tilde{s}_{k+2}$ such that $\mathrm{H}_{\frac{k}{2}+1}(s_1, \ldots, s_k, \tilde{s}_{k+1}, \tilde{s}_{k+2}) \succeq 0$;

  finite support: at most $\lfloor k/2 \rfloor + 1$ atoms (if it exists)

Gauss–Hermite quadrature

Gauss–Hermite quadrature

- $W \sim \mathcal{N}(0,1)$ with density $\varphi(w) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{w^2}{2})$

Gauss–Hermite quadrature

- $W \sim \mathcal{N}(0,1)$ with density $\varphi(w) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{w^2}{2})$
- "Gauss–Hermite quadrature" X_m^Q with m atoms matches $2m-1$ moments:

Gauss–Hermite quadrature

- $W \sim \mathcal{N}(0,1)$ with density $\varphi(w) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{w^2}{2})$
- "Gauss–Hermite quadrature" X_m^Q with m atoms matches $2m-1$ moments:

$$\text{atom } x_i = \text{root of } P_m(x) = \frac{(-1)^m}{\varphi(x)} \frac{\mathrm{d}^m \varphi(x)}{\mathrm{d}x^m},$$

Gauss–Hermite quadrature

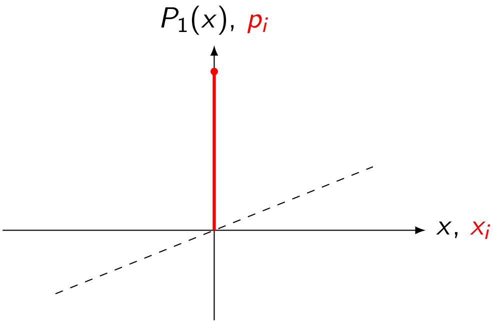
- $W \sim \mathcal{N}(0,1)$ with density $\varphi(w) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{w^2}{2})$
- "Gauss–Hermite quadrature" X_m^Q with m atoms matches $2m-1$ moments:

$$\text{atom } x_i = \text{root of } P_m(x) = \frac{(-1)^m}{\varphi(x)} \frac{\mathrm{d}^m \varphi(x)}{\mathrm{d}x^m}, \quad \text{with probability } p_i = \frac{(m-1)!}{m P_{m-1}^2(x_i)}$$

Gauss–Hermite quadrature

- $W \sim \mathcal{N}(0,1)$ with density $\varphi(w) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{w^2}{2})$
- "Gauss–Hermite quadrature" X_m^Q with m atoms matches $2m-1$ moments:

$$\text{atom } x_i = \text{root of } P_m(x) = \frac{(-1)^m}{\varphi(x)} \frac{\mathrm{d}^m \varphi(x)}{\mathrm{d}x^m}, \quad \text{with probability } p_i = \frac{(m-1)!}{m P_{m-1}^2(x_i)}$$

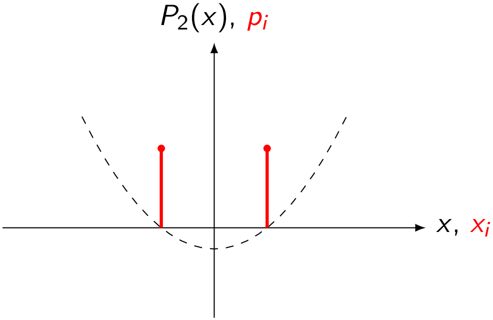


$$\mathbb{E}[X] = 0 \qquad \mathbb{E}[X^2] \neq 1$$

Gauss–Hermite quadrature

- $W \sim \mathcal{N}(0,1)$ with density $\varphi(w) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{w^2}{2})$
- "Gauss–Hermite quadrature" X_m^Q with m atoms matches $2m-1$ moments:

$$\text{atom } x_i = \text{root of } P_m(x) = \frac{(-1)^m}{\varphi(x)} \frac{\mathrm{d}^m \varphi(x)}{\mathrm{d}x^m}, \quad \text{with probability } p_i = \frac{(m-1)!}{m P_{m-1}^2(x_i)}$$

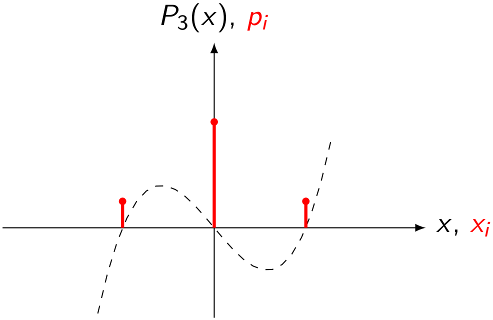


$$\mathbb{E}[X] = 0 \qquad \mathbb{E}[X^2] = 1$$
$$\mathbb{E}[X^3] = 0 \qquad \mathbb{E}[X^4] \neq 3$$

Gauss–Hermite quadrature

- $W \sim \mathcal{N}(0,1)$ with density $\varphi(w) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{w^2}{2})$
- "Gauss–Hermite quadrature" X_m^Q with m atoms matches $2m-1$ moments:

$$\text{atom } x_i = \text{root of } P_m(x) = \frac{(-1)^m}{\varphi(x)} \frac{\mathrm{d}^m \varphi(x)}{\mathrm{d}x^m}, \quad \text{with probability } p_i = \frac{(m-1)!}{m P_{m-1}^2(x_i)}$$



$$\mathbb{E}[X] = 0 \quad \mathbb{E}[X^2] = 1$$

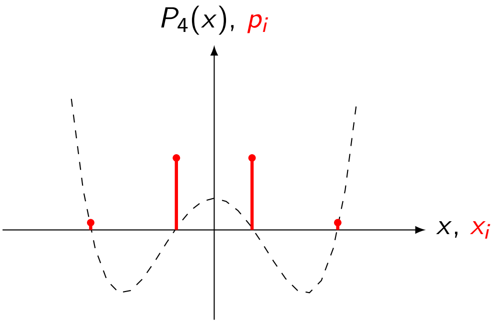
$$\mathbb{E}[X^3] = 0 \quad \mathbb{E}[X^4] = 3$$

$$\mathbb{E}[X^5] = 0 \quad \mathbb{E}[X^6] \neq 15$$

Gauss–Hermite quadrature

- $W \sim \mathcal{N}(0,1)$ with density $\varphi(w) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{w^2}{2})$
- "Gauss–Hermite quadrature" X_m^Q with m atoms matches $2m-1$ moments:

$$\text{atom } x_i = \text{root of } P_m(x) = \frac{(-1)^m}{\varphi(x)} \frac{\mathrm{d}^m \varphi(x)}{\mathrm{d}x^m}, \quad \text{with probability } p_i = \frac{(m-1)!}{m P_{m-1}^2(x_i)}$$

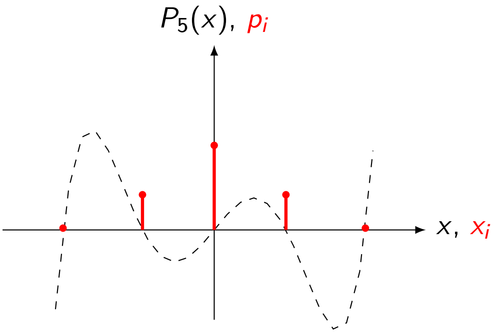


$$\mathbb{E}[X] = 0 \quad \mathbb{E}[X^2] = 1$$
$$\mathbb{E}[X^3] = 0 \quad \mathbb{E}[X^4] = 3$$
$$\mathbb{E}[X^5] = 0 \quad \mathbb{E}[X^6] = 15$$
$$\mathbb{E}[X^7] = 0 \quad \mathbb{E}[X^8] \neq 105$$

Gauss–Hermite quadrature

- $W \sim \mathcal{N}(0,1)$ with density $\varphi(w) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{w^2}{2})$
- "Gauss–Hermite quadrature" X_m^Q with m atoms matches $2m-1$ moments:

$$\text{atom } x_i = \text{root of } P_m(x) = \frac{(-1)^m}{\varphi(x)} \frac{\mathrm{d}^m \varphi(x)}{\mathrm{d}x^m}, \quad \text{with probability } p_i = \frac{(m-1)!}{m P_{m-1}^2(x_i)}$$

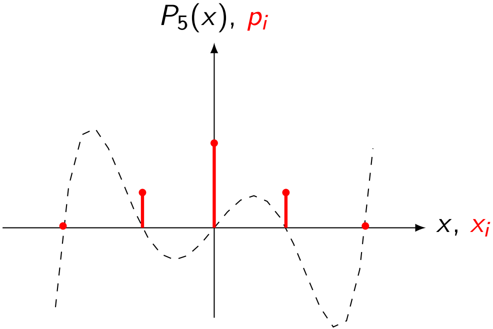


$$\begin{array}{ll} \mathbb{E}[X] = 0 & \mathbb{E}[X^2] = 1 \\ \mathbb{E}[X^3] = 0 & \mathbb{E}[X^4] = 3 \\ \mathbb{E}[X^5] = 0 & \mathbb{E}[X^6] = 15 \\ \mathbb{E}[X^7] = 0 & \mathbb{E}[X^8] = 105 \\ \mathbb{E}[X^9] = 0 & \mathbb{E}[X^{10}] \neq 1155 \end{array}$$

Gauss–Hermite quadrature

- $W \sim \mathcal{N}(0,1)$ with density $\varphi(w) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{w^2}{2})$
- "Gauss–Hermite quadrature" X_m^Q with m atoms matches $2m-1$ moments:

$$\text{atom } x_i = \text{root of } P_m(x) = \frac{(-1)^m}{\varphi(x)} \frac{\mathrm{d}^m \varphi(x)}{\mathrm{d}x^m}, \quad \text{with probability } p_i = \frac{(m-1)!}{m P_{m-1}^2(x_i)}$$



$$\begin{array}{ll} \mathbb{E}[X] = 0 & \mathbb{E}[X^2] = 1 \\ \mathbb{E}[X^3] = 0 & \mathbb{E}[X^4] = 3 \\ \mathbb{E}[X^5] = 0 & \mathbb{E}[X^6] = 15 \\ \mathbb{E}[X^7] = 0 & \mathbb{E}[X^8] = 105 \\ \mathbb{E}[X^9] = 0 & \mathbb{E}[X^{10}] \neq 1155 \end{array}$$

- optimal trade-off of atoms to moments matched

Outline

1. Moment problems
2. Entropy-constrained version
3. Application to a hypothesis testing problem

Entropy

Entropy

- for discrete X with probability masses $p_i > 0$, define entropy as

$$\mathbb{H}[X] = \sum_i p_i \log \frac{1}{p_i} = -\mathbb{E}[\log \mu(\{X\})]$$

Entropy

- for discrete X with probability masses $p_i > 0$, define entropy as

$$\mathbb{H}[X] = \sum_i p_i \log \frac{1}{p_i} = -\mathbb{E}[\log \mu(\{X\})]$$

- measure of information/randomness/complexity

Entropy

- for discrete X with probability masses $p_i > 0$, define entropy as

$$\mathbb{H}[X] = \sum_i p_i \log \frac{1}{p_i} = -\mathbb{E}[\log \mu(\{X\})]$$

- measure of information/randomness/complexity; fundamental compressibility limit

Entropy

- for discrete X with probability masses $p_i > 0$, define entropy as

$$\mathbb{H}[X] = \sum_i p_i \log \frac{1}{p_i} = -\mathbb{E}[\log \mu(\{X\})]$$

- measure of information/randomness/complexity; fundamental compressibility limit
- back to Gaussian moment matching: $\mathbb{H}[X_m^Q] \approx \frac{1}{2} \log m \to \infty$ as $m \to \infty$

Entropy-constrained moment problem

- for discrete X with probability masses $p_i > 0$, define entropy as

$$\mathbb{H}[X] = \sum_i p_i \log \frac{1}{p_i} = -\mathbb{E}[\log \mu(\{X\})]$$

- measure of information/randomness/complexity; fundamental compressibility limit
- back to Gaussian moment matching: $\mathbb{H}[X_m^Q] \approx \frac{1}{2} \log m \to \infty$ as $m \to \infty$
- can we find $(X_k)_k$ such that $\mathbb{H}[X_k] \leq h < \infty$ and $\mathbb{E}[X_k^n] = \mathbb{E}[W^n]$ for $n = 1, \ldots, k$?

Entropy-constrained moment problem

- for discrete X with probability masses $p_i > 0$, define entropy as

$$\mathbb{H}[X] = \sum_i p_i \log \frac{1}{p_i} = -\mathbb{E}[\log \mu(\{X\})]$$

- measure of information/randomness/complexity; fundamental compressibility limit
- back to Gaussian moment matching: $\mathbb{H}[X_m^Q] \approx \frac{1}{2} \log m \to \infty$ as $m \to \infty$
- can we find $(X_k)_k$ such that $\mathbb{H}[X_k] \leq h < \infty$ and $\mathbb{E}[X_k^n] = \mathbb{E}[W^n]$ for $n = 1, \ldots, k$?
- no (would imply weak convergence to W + lower semi-continuity of $\mathbb{H}$)

Entropy-constrained moment problem

- for discrete X with probability masses $p_i > 0$, define entropy as

$$\mathbb{H}[X] = \sum_i p_i \log \frac{1}{p_i} = -\mathbb{E}[\log \mu(\{X\})]$$

- measure of information/randomness/complexity; fundamental compressibility limit
- back to Gaussian moment matching: $\mathbb{H}[X_m^Q] \approx \frac{1}{2} \log m \to \infty$ as $m \to \infty$
- can we find $(X_k)_k$ such that $\mathbb{H}[X_k] \leq h < \infty$ and $\mathbb{E}[X_k^n] = \mathbb{E}[W^n]$ for $n = 1, \ldots, k$?
- no (would imply weak convergence to W + lower semi-continuity of $\mathbb{H}$)
- how many moments can we match if entropy is at most h then?

Main result

Theorem

Main result

Theorem

For any continuous W, there exists $\eta_W \in (0, \frac{1}{2})$ such that

Main result

Theorem

For any continuous W, there exists $\eta_W \in (0, \frac{1}{2})$ such that

(i) *if X has $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3, 4$, then $\mathbb{H}[X] \geq h_2(\eta_W)$.*

Main result

Theorem

For any continuous W, there exists $\eta_W \in (0, \frac{1}{2})$ such that

(i) *if X has $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3, 4$, then $\mathbb{H}[X] \geq h_2(\eta_W)$.*

(ii) *for any $h > 0$, there is X with $\mathbb{H}[X] \leq h$ and $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3$.*

Main result

Theorem

For any continuous W, there exists $\eta_W \in (0, \frac{1}{2})$ such that

(i) *if X has $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3, 4$, then $\mathbb{H}[X] \geq h_2(\eta_W)$.*

(ii) *for any $h > 0$, there is X with $\mathbb{H}[X] \leq h$ and $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3$.*

Proof idea

$\square$

Main result

Theorem

For any continuous W, there exists $\eta_W \in (0, \frac{1}{2})$ such that

(i) *if X has $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3, 4$, then $\mathbb{H}[X] \geq h_2(\eta_W)$.*

(ii) *for any $h > 0$, there is X with $\mathbb{H}[X] \leq h$ and $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3$.*

Proof idea

- $\mathbb{H}[X] \leq h < \log 2 \iff X = \begin{cases} \tilde{X} & \text{w.p. } \epsilon < 1/2 \\ x_0 & \text{w.p. } 1 - \epsilon > 1/2 \end{cases}$

$\square$

Main result

Theorem

For any continuous W, there exists $\eta_W \in (0, \frac{1}{2})$ such that

(i) *if X has $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3, 4$, then $\mathbb{H}[X] \geq h_2(\eta_W)$.*

(ii) *for any $h > 0$, there is X with $\mathbb{H}[X] \leq h$ and $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3$.*

Proof idea

- $\mathbb{H}[X] \leq h < \log 2 \iff X = \begin{cases} \tilde{X} & \text{w.p. } \epsilon < 1/2 \\ x_0 & \text{w.p. } 1 - \epsilon > 1/2 \end{cases}$
- need $\mathbb{E}[\tilde{X}^n] = s_n := \frac{1}{\epsilon}(\mathbb{E}[W^n] - (1 - \epsilon)x_0^n)$

$\square$

Main result

Theorem

For any continuous W, there exists $\eta_W \in (0, \frac{1}{2})$ such that

(i) *if X has $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3, 4$, then $\mathbb{H}[X] \geq h_2(\eta_W)$.*

(ii) *for any $h > 0$, there is X with $\mathbb{H}[X] \leq h$ and $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3$.*

Proof idea

- $\mathbb{H}[X] \leq h < \log 2 \iff X = \begin{cases} \tilde{X} & \text{w.p. } \epsilon < 1/2 \\ x_0 & \text{w.p. } 1 - \epsilon > 1/2 \end{cases}$
- need $\mathbb{E}[\tilde{X}^n] = s_n := \frac{1}{\epsilon}(\mathbb{E}[W^n] - (1-\epsilon)x_0^n)$
- check: $s_1, \ldots, s_4$ "valid" iff $\epsilon > \eta_W$, but s_1, s_2, s_3 always "valid" $\square$

Main result

Theorem

For any continuous W, there exists $\eta_W \in (0, \frac{1}{2})$ such that

(i) *if X has $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3, 4$, then $\mathbb{H}[X] \geq h_2(\eta_W)$.*

(ii) *for any $h > 0$, there is X with $\mathbb{H}[X] \leq h$ and $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3$.*

Main result

Theorem

For any continuous W, there exists $\eta_W \in (0, \frac{1}{2})$ such that

(i) *if X has $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3, 4$, then $\mathbb{H}[X] \geq h_2(\eta_W)$.*

(ii) *for any $h > 0$, there is X with $\mathbb{H}[X] \leq h$ and $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3$.*

For symmetric W,

$$\eta_W = \begin{cases} \dfrac{\mathbb{E}[W^2]^2}{\mathbb{E}[X^4]} & \text{if } \frac{\mathbb{E}[W^2]^2}{\mathbb{E}[X^4]} \leq \frac{1}{3}, \\[2ex] \dfrac{5\mathbb{E}[W^2]^2 - \mathbb{E}[W^4]}{9\mathbb{E}[W^2]^2 - \mathbb{E}[W^4]} & \text{if } \frac{\mathbb{E}[W^2]^2}{\mathbb{E}[X^4]} > \frac{1}{3}. \end{cases}$$

Main result

Theorem

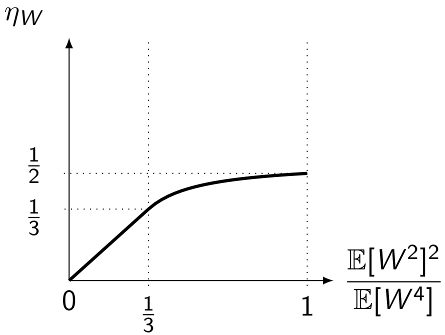
For any continuous W, there exists $\eta_W \in (0, \frac{1}{2})$ such that

(i) *if X has $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3, 4$, then $\mathbb{H}[X] \geq h_2(\eta_W)$.*

(ii) *for any $h > 0$, there is X with $\mathbb{H}[X] \leq h$ and $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3$.*

For symmetric W,

$$\eta_W = \begin{cases} \dfrac{\mathbb{E}[W^2]^2}{\mathbb{E}[X^4]} & \text{if } \frac{\mathbb{E}[W^2]^2}{\mathbb{E}[X^4]} \leq \frac{1}{3}, \\ \dfrac{5\mathbb{E}[W^2]^2 - \mathbb{E}[W^4]}{9\mathbb{E}[W^2]^2 - \mathbb{E}[W^4]} & \text{if } \frac{\mathbb{E}[W^2]^2}{\mathbb{E}[X^4]} > \frac{1}{3}. \end{cases}$$



Main result

Theorem

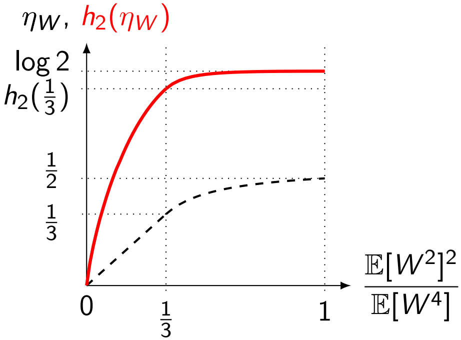
For any continuous W, there exists $\eta_W \in (0, \frac{1}{2})$ such that

(i) *if X has $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3, 4$, then $\mathbb{H}[X] \geq h_2(\eta_W)$.*

(ii) *for any $h > 0$, there is X with $\mathbb{H}[X] \leq h$ and $\mathbb{E}[X^n] = \mathbb{E}[W^n]$ for $n = 1, 2, 3$.*

For symmetric W,

$$\eta_W = \begin{cases} \dfrac{\mathbb{E}[W^2]^2}{\mathbb{E}[X^4]} & \text{if } \frac{\mathbb{E}[W^2]^2}{\mathbb{E}[X^4]} \leq \frac{1}{3}, \\[2ex] \dfrac{5\mathbb{E}[W^2]^2 - \mathbb{E}[W^4]}{9\mathbb{E}[W^2]^2 - \mathbb{E}[W^4]} & \text{if } \frac{\mathbb{E}[W^2]^2}{\mathbb{E}[X^4]} > \frac{1}{3}. \end{cases}$$



Outline

Testing against Gaussian in large noise

- observe n i.i.d. samples $Y_i = X_i + Z_i$

Testing against Gaussian in large noise

- observe n i.i.d. samples $Y_i = X_i + Z_i$ with $Z_i \sim Z \sim \mathcal{N}(0, \sigma^2)$ independent of X_i, $\sigma \to \infty$

Testing against Gaussian in large noise

- observe n i.i.d. samples $Y_i = X_i + Z_i$ with $Z_i \sim Z \sim \mathcal{N}(0, \sigma^2)$ independent of X_i, $\sigma \to \infty$

- $$\left.\begin{array}{ll} H = 0: & X_i \sim X \text{ discrete with } \mathbb{H}[X] \leq h \\ H = 1: & X_i \sim G \sim \mathcal{N}(0,1) \end{array}\right\}$$

Testing against Gaussian in large noise

- observe n i.i.d. samples $Y_i = X_i + Z_i$ with $Z_i \sim Z \sim \mathcal{N}(0, \sigma^2)$ independent of X_i, $\sigma \to \infty$

- $$\left.\begin{aligned} H = 0: &\quad X_i \sim X \text{ discrete with } \mathbb{H}[X] \leq h \\ H = 1: &\quad X_i \sim G \sim \mathcal{N}(0,1) \end{aligned}\right\} \text{decide } \hat{H} \in \{0,1\} \text{ from } Y^n$$

Testing against Gaussian in large noise

- observe n i.i.d. samples $Y_i = X_i + Z_i$ with $Z_i \sim Z \sim \mathcal{N}(0, \sigma^2)$ independent of X_i, $\sigma \to \infty$
- $$\left.\begin{array}{ll} H = 0: & X_i \sim X \text{ discrete with } \mathbb{H}[X] \le h \\ H = 1: & X_i \sim G \sim \mathcal{N}(0,1) \end{array}\right\} \text{decide } \hat{H} \in \{0,1\} \text{ from } Y^n$$
- type-I error: $\Pr\{\hat{H} = 1 \mid H = 0\}$ and type-II error: $\Pr\{\hat{H} = 0 \mid H = 1\}$

Testing against Gaussian in large noise

- observe n i.i.d. samples $Y_i = X_i + Z_i$ with $Z_i \sim Z \sim \mathcal{N}(0, \sigma^2)$ independent of X_i, $\sigma \to \infty$

- $$\left.\begin{array}{ll} H = 0 : & X_i \sim X \text{ discrete with } \mathbb{H}[X] \leq h \\ H = 1 : & X_i \sim G \sim \mathcal{N}(0,1) \end{array}\right\} \text{ decide } \hat{H} \in \{0,1\} \text{ from } Y^n$$

- type-I error: $\Pr\{\hat{H} = 1 \mid H = 0\}$ and type-II error: $\Pr\{\hat{H} = 0 \mid H = 1\}$

- Stein regime:
  type-I error $\leq \epsilon < 1$

Testing against Gaussian in large noise

- observe n i.i.d. samples $Y_i = X_i + Z_i$ with $Z_i \sim Z \sim \mathcal{N}(0, \sigma^2)$ independent of X_i, $\sigma \to \infty$

- $$\left.\begin{array}{ll} H = 0: & X_i \sim X \text{ discrete with } \mathbb{H}[X] \leq h \\ H = 1: & X_i \sim G \sim \mathcal{N}(0,1) \end{array}\right\} \text{decide } \hat{H} \in \{0,1\} \text{ from } Y^n$$

- type-I error: $\Pr\{\hat{H} = 1 \mid H = 0\}$ and type-II error: $\Pr\{\hat{H} = 0 \mid H = 1\}$

- Stein regime:
  type-I error $\leq \epsilon < 1$ and type-II error $\leq \exp\left[-n\,\mathrm{D_{KL}}(X + Z \,\|\, G + Z) + \mathrm{o}(n)\right]$

Testing against Gaussian in large noise

- observe n i.i.d. samples $Y_i = X_i + Z_i$ with $Z_i \sim Z \sim \mathcal{N}(0, \sigma^2)$ independent of X_i, $\sigma \to \infty$

- $$\left.\begin{array}{ll} H = 0: & X_i \sim X \text{ discrete with } \mathbb{H}[X] \leq h \\ H = 1: & X_i \sim G \sim \mathcal{N}(0,1) \end{array}\right\} \text{decide } \hat{H} \in \{0,1\} \text{ from } Y^n$$

- type-I error: $\Pr\{\hat{H} = 1 \mid H = 0\}$ and type-II error: $\Pr\{\hat{H} = 0 \mid H = 1\}$

- Stein regime:
  type-I error $\leq \epsilon < 1$ and type-II error $\leq \exp\left[-n \, \mathrm{D}_{\mathrm{KL}}(X + Z \,\|\, G + Z) + \mathrm{o}(n)\right]$

- worst-case exponent $= \min\limits_{X \,:\, \mathbb{H}[X] \leq h} \mathrm{D}_{\mathrm{KL}}(X + Z \,\|\, G + Z)$

Testing against Gaussian in large noise

- observe n i.i.d. samples $Y_i = X_i + Z_i$ with $Z_i \sim Z \sim \mathcal{N}(0, \sigma^2)$ independent of X_i, $\sigma \to \infty$

- $$\left.\begin{array}{ll} H = 0: & X_i \sim X \text{ discrete with } \mathbb{H}[X] \leq h \\ H = 1: & X_i \sim G \sim \mathcal{N}(0,1) \end{array}\right\} \text{decide } \hat{H} \in \{0,1\} \text{ from } Y^n$$

- type-I error: $\Pr\{\hat{H} = 1 \mid H = 0\}$ and type-II error: $\Pr\{\hat{H} = 0 \mid H = 1\}$

- Stein regime:
  type-I error $\leq \epsilon < 1$ and type-II error $\leq \exp\left[-n\, \mathrm{D_{KL}}(X + Z \,\|\, G + Z) + o(n)\right]$

- worst-case exponent $= \min\limits_{X : \mathbb{H}[X] \leq h} \mathrm{D_{KL}}(X + Z \,\|\, G + Z) = \Theta\left(\frac{1}{\sigma^8}\right)$ for $h < h_2\left(\frac{1}{3}\right)$

Testing against Gaussian in large noise

- observe n i.i.d. samples $Y_i = X_i + Z_i$ with $Z_i \sim Z \sim \mathcal{N}(0, \sigma^2)$ independent of X_i, $\sigma \to \infty$
- $$\left.\begin{array}{ll} H = 0: & X_i \sim X \text{ discrete with } \mathbb{H}[X] \le h \\ H = 1: & X_i \sim G \sim \mathcal{N}(0,1) \end{array}\right\} \text{decide } \hat{H} \in \{0,1\} \text{ from } Y^n$$
- type-I error: $\Pr\{\hat{H} = 1 \mid H = 0\}$ and type-II error: $\Pr\{\hat{H} = 0 \mid H = 1\}$
- Stein regime:
  type-I error $\le \epsilon < 1$ and type-II error $\le \exp\left[-n\,\mathrm{D_{KL}}(X + Z \,\|\, G + Z) + o(n)\right]$
- worst-case exponent $= \min\limits_{X : \mathbb{H}[X] \le h} \mathrm{D_{KL}}(X + Z \,\|\, G + Z) = \Theta\left(\frac{1}{\sigma^8}\right)$ for $h < h_2\left(\frac{1}{3}\right)$
- proof: Taylor expansion, k^{th} term depends on first k moments, can match only 3 moments

Summary

- entropy-constrained moment problem

Summary

- entropy-constrained moment problem
- can match only three moments if the entropy is “small”

Summary

- entropy-constrained moment problem
- can match only three moments if the entropy is “small”
- many applications

Summary

- entropy-constrained moment problem
- can match only three moments if the entropy is “small”
- many applications, example: hypothesis testing

Summary

- entropy-constrained moment problem
- can match only three moments if the entropy is “small”
- many applications, example: hypothesis testing

Thank you!