Minimax Lower Bounds via f-divergences

Adityanand Guntuboyina,
Department of Statistics,
Yale University.

14 June, 2010

Slides: www.stat.yale.edu/~ag575
Paper(full version): http://arxiv.org/abs/1002.0042v1

f-divergences

$$D_f\left(P\|Q\right) = \int f\left(\frac{dP}{dQ}\right) dQ$$

P, Q probs; P<<Q

f convex on [0, ∞)
f(1)=0

KL, χ^2 (chi-squared), Total Variation, Hellinger etc.

Estimation

F finite size N

To estimate $\theta \sim \text{unif}(F)$

Observe $X \sim p(x|\theta)$ or p_θ or P_θ

$T(X)$: estimator for θ based on X

$$P_e(T) = \mathrm{P}(T(X) \neq \theta) = \frac{1}{N} \sum_{\theta \in F} \mathrm{P}_\theta(T(X) \neq \theta)$$

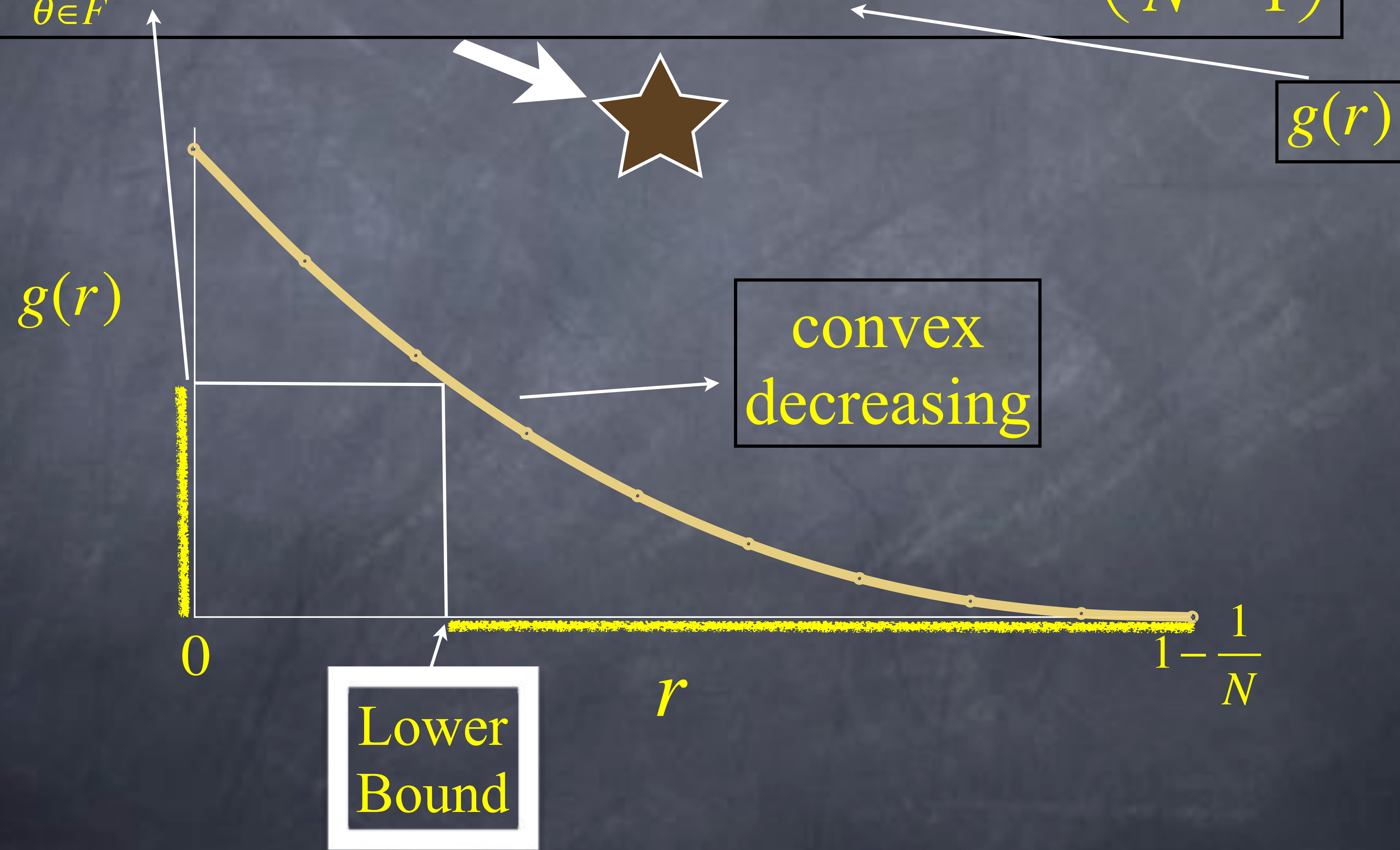
$$r = \inf_T P_e(T)$$

We'll lower bound r using f-divergences

$$r = 1 - \frac{1}{N} \int \max_{\theta \in F} p_\theta \quad \leq 1 - \frac{1}{N}$$

Implicit Lower Bound for r

$$\inf_Q \sum_{\theta \in F} D_f(P_\theta \| Q) \geq f(N(1-r)) + (N-1) f\left(\frac{Nr}{N-1}\right)$$



Proof of ★

CONVEXITY

$b > 0$

$$\sum_{\theta\in F} f(ba_\theta) \geq f(bm) + (N-1) f\left(\frac{b(1-m)}{N-1}\right)$$

$a_\theta \geq 0; \sum_{\theta\in F} a_\theta = 1$

$m := \max_{\theta\in F} a_\theta$

$$b = \frac{\sum p_\theta}{q}; ba_\theta = \frac{p_\theta}{q}$$

Integrate with Q
Jensen's Inequality

$$r = 1 - \frac{1}{N}\int \max_{\theta\in F} p_\theta$$

★ $$\inf_Q \sum_{\theta\in F} D_f(P_\theta \| Q) \geq f(N(1-r)) + (N-1) f\left(\frac{Nr}{N-1}\right)$$

f(x) := x log x; Fano

$$\frac{1}{N}\inf_{Q}\sum_{\theta\in F} D(P_\theta \,\|\, Q) = \frac{1}{N}\sum_{\theta\in F} D(P_\theta \,\|\, \bar{P}) =: I(\theta, X)$$

KL

Mutual Information

$$I(\theta, X) \geq (1-r)\log N - H(r)$$

$\leq \log 2$

Binary Entropy

$$r \geq 1 - \frac{\log 2 + I(\theta; X)}{\log N}$$

Fano's Inequality

$f(x) = x^2$-1;chi-squared

Right Hand Side of ★ ⇒ $N^2\left(r-(1-1/N)\right)^2$

$$r \geq 1 - \frac{1}{\sqrt{N}}\sqrt{\inf_Q \frac{1}{N}\sum_{\theta\in F}\chi^2(P_\theta \| Q) - \frac{1}{N}}$$

χ^2-analogue of Mutual Information

$N = 2$

$$r = 1 - \frac{1}{2}\int \max(p_1, p_2) = \frac{1-V}{2}$$

$$V := \frac{1}{2}\int |p_1 - p_2|$$

Total Variation

$$\inf_Q \left[D_f\left(P_1 \| Q\right) + D_f\left(P_2 \| Q\right)\right] \geq f(1+V) + f(1-V)$$

$f(x) = x \log x$

$f(1+V) + f(1-V) \geq V^2$

Topsøe (2000)

Pinsker's Inequality

SHARP

$$P_1\{1\} = P_2\{2\} = \frac{1+V}{2}$$

$$P_1\{2\} = P_2\{1\} = \frac{1-V}{2}$$

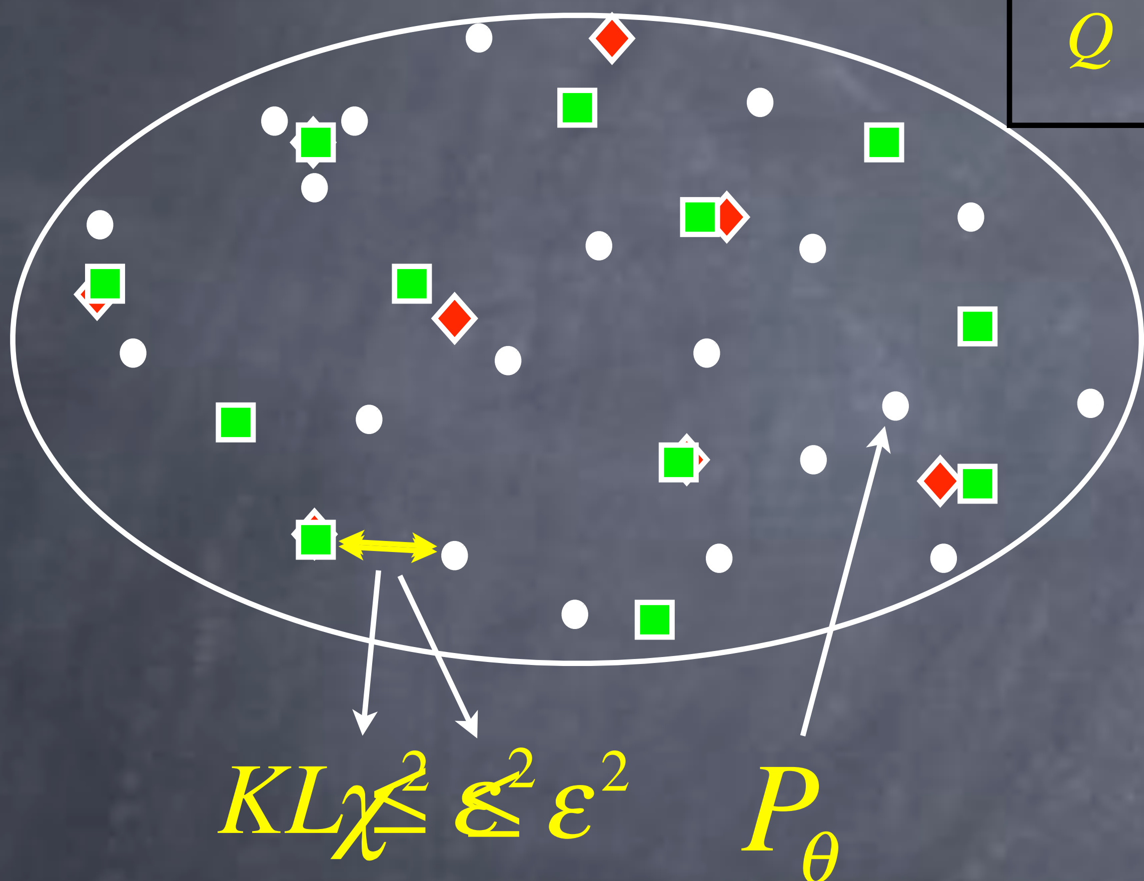
$$Q\{1\} = Q\{2\} = \frac{1}{2}$$

Upper Bound for $\inf_Q \sum_{\theta \in F} D_f(P_\theta \| Q)$

$I(\theta, X)$

\# ◆

$$\inf_Q \frac{1}{N} \sum_{\theta \in F} D(P_\theta \| Q) \le \log M_{KL}(\varepsilon) + \varepsilon^2$$



Yang and Barron (1999)

Covering Number Bounds

\# ■

$$\inf_Q \frac{1}{N} \sum_{\theta \in F} \chi^2(P_\theta \| Q) \le M_C(\varepsilon)(1 + \varepsilon^2)$$

General Estimation Problem

To estimate $\theta \in \Theta$ 　 (Θ, ρ) metric space

Observe $X \sim p(x|\theta)$ or p_θ or P_θ

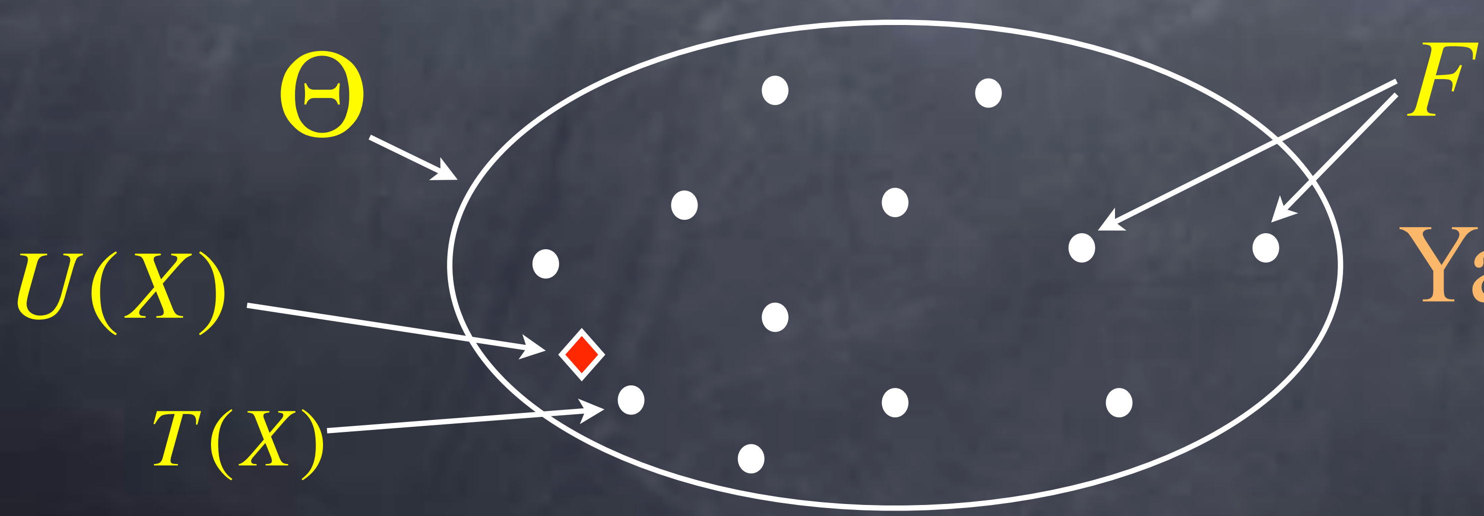
U(X) : estimator for θ based on X

$$P_e(U) = \sup_{\theta\in\Theta} E_\theta\left(\rho\left(U(X),\theta\right)\right)^2$$

$$R = \inf_U P_e(U)$$

Minimax Risk

F⊆Θ finite & η-separated ⇒ $R \geq \left(\frac{\eta}{2}\right)^2 r$



Yang and Barron (1999)

Global Entropy Bounds

$$R \geq \left(\frac{\eta}{2}\right)^2 \left[1 - \frac{\log M_{KL}(\varepsilon) + \varepsilon^2 + \log 2}{\log N(\eta)}\right]$$

Yang and Barron (1999)

Size of any η-separated subset of Θ

$$R \geq \left(\frac{\eta}{2}\right)^2 \left[1 - \sqrt{\frac{M_C(\varepsilon)(1+\varepsilon^2)}{N(\eta)}} - \frac{1}{N(\eta)}\right]$$

Optimize over η and ε

Convenient when global covering/packing numbers are available

Example: Parametric Estimation

$\Theta = [1,2]$

$\rho(\theta,\theta') = |\theta - \theta'|$

$N(\eta) = \frac{c_1}{\eta}$

$M_{KL}(\varepsilon) = c_2 \frac{\sqrt{n}}{\varepsilon}$

$M_C(\varepsilon) = c_3 \frac{\sqrt{n}}{\sqrt{\log(1+\varepsilon^2)}}$

eg. $X_1, X_2, \ldots, X_n \sim \exp(\theta)$ or $N(\theta,1)$

$R \geq o\left(\frac{1}{n}\right)$ ← KL

χ^2 → $R \geq \frac{c}{n}$

Conclusion

- New Lower Bound for Estimation using f-divergences.
- Special cases include Fano's inequality and Pinsker's inequality.
- The lower bound involves a version of Mutual Information for f-divergences, which can be bounded from above using global covering numbers.
- Global entropy numbers determine the parametric minimax lower bound as well.

References

- Y.Yang and A. Barron, "Information-theoretic determination of minimax rates of convergence,"*Annals of Statistics*, vol. 27, pp. 1564-1599, 1999.
- F. Topsøe, "Some inequalities for information divergence and related measures of discrimination," *IEEE Trans. Inform. Theory,* vol. 46, pp. 1602-1609, 2000.
- A. Guntuboyina, "Lower bounds for the minimax risk using f-divergences and applications," 2010, [Online]. Available: http://arxiv.org/abs/1002.0042v1.