Minimax lower bound

이광민

May 2018

1 Notation

모수공간 : Θ

Action space : $\mathcal{A}$

Loss function : $L : \Theta \times \mathcal{A} \to [0, \infty)$

Sample space : $\mathcal{X}$

Data : $X \sim P_\theta$ (Probability measure on sample space)

Decision rule : $\mathcal{D} : \mathcal{X} \to \mathcal{A}$

Minimax Risk : $R_{minimax} := \inf_{\mathcal{D}} \sup_{\theta \in \Theta} \mathbb{E}_\theta L(\theta, \mathcal{D}(X))$

Notation for asymptotic

Sample space : $\mathcal{X}^n$

Data : $X^{(n)} \sim P_\theta^n$ (Probability measure on sample space)

Decision rule : $\mathcal{D}^{(n)} : \mathcal{X}^n \to \mathcal{A}$

Minimax Risk : $R_{minimax}^{(n)} := \inf_{\mathcal{D}} \sup_{\theta \in \Theta} \mathbb{E}_\theta L(\theta, \mathcal{D}(X^{(n)}))$

2 목표

특정 추정량($\mathcal{D}^*$)이 minimax관점에서 최적인 추정량임을 보장하기 위해 다음을 만족해야 함.

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta L(\theta, \mathcal{D}^*(X)) = R_{minimax}$$

따라서 $R_{minimax}$값을 계산할 필요가 있음. 그런데, 정확한 $R_{minimax}$를 구하기 힘들기 때문에 minimax관점에서 최적인 추정량을 다음과 같이 재정의한다.

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta L(\theta, \mathcal{D}^{*(n)}(X)) \asymp R^{(n)}_{minimax}$$

즉 Minimax risk가 최적의 수렴속도와 asymptotic하게 일치하는 것으로 만족.

따라서 $R^{(n)}_{minimax}$를 계산해야 하고, 이를 정확하게 계산하는 것이 아닌 asymptotic한 계산만으로 충분하다. 그리고 해당 발표의 주제는 일반적인 setting에서 $R_{minimax}$를 계산하기 위해 사용되는 방법들에 대해 다룬다. 즉 for some $a, b > 0$ 다음 식을 만족하는 ϵ_n을 구하면 된다.

$$R^{(n)}_{minimax} \le a\epsilon_n$$
$$R^{(n)}_{minimax} \ge b\epsilon_n$$

이를 다시 표현하면, Upper bound에 대해서는 적당한 $\mathcal{D}^{*(n)}$를 잡아서, $\sup_{\theta \in \Theta} \mathbb{E}_\theta L(\theta, \mathcal{D}^{*(n)}(X))$의 upper bound를 계산한다. (이를 $a \times u_n$이라 함.)
그리고 lower bound에 대해서는 $\inf_{\mathcal{D}} \sup_{\theta \in \Theta} \mathbb{E}_\theta L(\theta, \mathcal{D}(X^{(n)}))$의 lower bound를 계산한다. (이를 $b \times l_n$이라 함.)
이때 $u_n = l_n$이 될때까지 최대한 upper, lower bound를 tight하게 만든다.

lower bound를 계산하기 위해서 모든 조합들 다 고려해 봐야하는데, 이는 실질적으로 불가능하고, 베이즈 위험에 바탕해 둔 lower bound 계산 기술들에 대해 다룬다.

3 베이즈 위험을 이용한 bound

Definiton and notation

Bayes risk for prior w :
$R_{Bayes}(w) := \inf_{\mathcal{D}} \int_\Theta \mathbb{E}_\theta L(\theta, \mathcal{D}(X)) w(d\theta)$

$B_{(w,L)}(x) : \inf_{a \in \mathcal{A}} B^a_{(w,L)}(x)$

$B^a_{(w,L)}(x) := \int_\Theta L(\theta, a)p_\theta(x)w(d\theta)$

Since $R_{minimax} \geq R_{Bayes}(w)$, $\forall w$

Assuming the condition for interchanging of order of integration

$$\begin{aligned} R_{minimax} &\geq \int_\Theta \mathbb{E}_\theta L(\theta, \mathcal{D}(X))w(d\theta) \\ &= \int_\mathcal{X} \int_\Theta L(\theta, \mathcal{D}(x))p_\theta(x)w(d\theta)\mu(dx) \\ &\geq \int_\mathcal{X} B_{w,L}(x)\mu(dx) \end{aligned}$$

minimax에 대한 lower bound는 적당한 사전분포를 잘 잡아서 베이즈리스크 구하는 것으로 환원된다.

4 Standard Technique

앞장에서 사전분포 w를 잡을 때, 전체 모수공간에 대한 measure를 고려하는 것이 아닌, Finite set $F \subset \Theta$에 대해 사전분포를 부여하는 방식으로 베이즈 Risk를 제시한다.

4.1 Notation

- $d(\theta_1, \theta_2) := \inf\{L(\theta_1, a) + L(\theta_2, a) : a \in \mathcal{A}\}$ for $\theta_1, \theta_2 \in \Theta$

- $d(\Theta_1, \Theta_2) := \inf\{d(\theta_1, \theta_2) : \theta_1 \in \Theta_1, \theta_2 \in \Theta_2\}$ for subsets $\Theta_1, \Theta_2 \subset \Theta$

- finite set $F \subset \Theta$는 η-seperated if $d(\theta_1, \theta_2) \geq \eta$ for all $\theta_1, \theta_2 \in F$ $\theta_1 \neq \theta_2$

- For measure on $P_1, \ldots, P_N$ on $\mathcal{X}$ and weights ρ_i

$$\bar{r}_\rho(P_1, \ldots, P_N) := 1 - \int_\mathcal{X} \max_{1 \leq i \leq N} [\rho_i p_i(x)]\mu(dx), \text{ where } p_i := dP_i/d\mu$$

(이는 lower bound를 표현하는 단위로서 중요한 역할을함)

- Hamming distance on hyper cube $\{0,1\}^m$

$$\Gamma(\tau,\tau') = \sum_{i=1}^{m} \{\tau_i \neq \tau'_i\}$$

4.2 Multiple Hypothesis Testing

Assume $\Theta = \mathcal{A} = \{1,\ldots,N\}$ and $L(\theta,a) = \{\theta \neq a\}$. Then,

$$R_{minimax} \geq \bar{r}_\rho(P_1,\ldots,P_N) \text{ for every prob } \rho$$

Proof 정의에 의해

$$B^a_{\rho,L}(x) = \sum_{i=1}^{N} \{a \neq i\} p_i(x)\rho_i = \sum_{i=1}^{N} p_i(x)\rho_i - p_a(x)\rho_a$$

$$B_{\rho,L}(x) = \sum_{i=1}^{N} p_i(x)\rho_i - \max_j p_j(x)\rho_j$$

이 성립하고 베이즈 리스크에 의한 bound를 그대로 적용하면 됨.

4.3 General Testing Bound

앞의 multiple hypotesis testing은 일반적인 loss function에 적용가능한 것이 아니기에 실질적으로는 의미 없는 정리이다. 하지만 이를 적용하여 다음 정리를 유도할 수 있다.
For every η-separated finite subset F of Θ we have

$$R_{minimax} \geq \frac{\eta}{2} \bar{r}_\rho(P_\theta, \theta \in F) \text{ for all } \rho_\theta \geq 0, \theta \in F \text{ with } \sum_{\theta \in F} \rho_\theta = 1$$

Proof Since $L(\theta,a) \geq (\eta/2)\{L(\theta,a) \geq \eta/2\}$,

$$B^a_{w,L}(x) \geq \frac{\eta}{2}\Big(\sum_{\theta \in F} \rho_\theta p_\theta(x) - \sum_{\theta \in F} \rho_\theta p_\theta(x)\{L(\theta,a) < \eta/2\}\Big)$$

$$B_{w,L}(x) \geq \frac{\eta}{2}\Big(\sum_{\theta \in F} \rho_\theta p_\theta(x) - \max_{\theta \in F}[\rho_\theta p_\theta(x)]\Big) \text{ since F is } \eta - separated$$

4.4 Assouad's method

General Hypothesis를 이용하였을때의 lower bound는 여전히 일반적인 상황에서 계산하기 용이하지 않다. 좀더 계산용이한 lower bound version하나 제시함.

Suppose $\Theta, \mathcal{A} = \{0,1\}^m$, $L(\theta, a) = \Gamma(\theta, a) = \sum_{i=1}^m \{\theta_i \neq a_i\}$ Then,

$$R_{minimax} \geq \frac{m}{2} \min_{\Gamma(\theta,\theta')=1} ||P_\theta \wedge P_{\theta'}||_1$$

Proof Let prior w as uniform then

$$B^a_{w,\Gamma} = 2^{-m} \sum_{i=1}^m \sum_{\theta \in \{0,1\}^m} \{\theta_i \neq a_i\} p_\theta(x)$$

$$B_{w,\Gamma}(x) = \frac{1}{2} \sum_{i=1}^m \min\left(\frac{\sum_{\theta:\theta_i=0} p_\theta(x)}{2^{m-1}}, \frac{\sum_{\theta:\theta_i=1} p_\theta(x)}{2^{m-1}} \right)$$

따라서 그대로 대입하게되면,

$$\begin{aligned} R_{minimax} &\geq \frac{1}{2} \sum_{i=1}^m ||\left(2^{-(m-1)} \sum_{\theta:\theta_i=0} P_\theta\right) \wedge \left(2^{-(m-1)} \sum_{\theta:\theta_i=0} P_\theta\right)||_1 \\ &\geq \frac{m}{2} \min_{\Gamma(\theta,\theta')=1} ||P_\theta \wedge P_{\theta'}||_1 \end{aligned}$$

마지막 부등식은 다음 이유때문에 성립한다.

$|| \sum P_i/d \wedge \sum Q_i/d||_1 = 1 - || \sum (P_i - Q_i)/d||_{TV} \geq 1 - \sum ||P_i - Q_i||/d \geq 1 - \max ||P_i - Q_i|| = \min ||P_i \wedge Q_i||$

4.5 General Assouad

마찬가지로 Assoud에서 정의한 loss function은 일반적으로 사용되기 힘든 loss이다.

Consider $\psi : \{0,1\}^m \to \Theta$ and suppose ξ is a positive real number such that $d(\psi(\tau), \psi(\tau')) \geq \xi \Gamma(\tau, \tau')$ Then,

$$R_{minimax} \geq \frac{m\xi}{4} \min_{\Gamma(\theta,\theta')=1} ||P_{\psi(\tau)} \wedge P_{\psi(\tau')}||_1$$

Proof Let $\tau_a := \mathrm{argmin}_\tau L(\psi(\tau), a)$,

$$L(\psi(\tau), a) \geq \frac{L(\psi(\tau), a) + L(\psi(\tau_a), a)}{2} \geq \frac{\xi}{2} \Gamma(\tau, \tau_a)$$

이 되므로 , 다음 이 성립한다.

$$B^a_{w,L}(x) \geq \frac{\xi}{2}\frac{1}{2^m}\sum_{\theta\in\{0,1\}^m}\Gamma(\tau,\tau_a)p_{\psi(\tau)}(x)$$

assouad에서와 같은 방법으로 증명 할수 있음

Finite한 공간에서의 probability measure간의 거리는 비교적 쉽게 계산할 수있다.

4.6 Le Cam

Θ_1, Θ_2를 Θ의 subset이라 하고, 각각에 대해 prior w_1, w_2라 한다.
그리고 $m_i(x) := \int_{\Theta_i} p_\theta(x)w_i(d\theta)$라 할때,

$$R_{minimax} \geq \frac{1}{2}d(\Theta_1,\Theta_2)||m_1 \wedge m_2||_1$$

Proof Let $w = (w_1 + w_2)/2$

$$\begin{aligned} B^a_{w,L}(x) &= \frac{1}{2}B^a_{w_1,L}(x) + \frac{1}{2}B^a_{w_2,L}(x) \\ &\geq \frac{1}{2}m_1(x)\inf_{\theta_1\in\Theta_1}L(\theta_1,a) + \frac{1}{2}m_2(x)\inf_{\theta_2\in\Theta_2}L(\theta_2,a) \\ &\geq \frac{1}{2}\min(m_1(x),m_2(x))d(\Theta_1,\Theta_2) \end{aligned}$$

마지막 부등식은 $\inf_{\theta_1\in\Theta_1}L(\theta_1,a) + \inf_{\theta_2\in\Theta_2}L(\theta_2,a) \geq \inf_{\theta_i\in\Theta_i}[L(\theta_1,a) + L(\theta_2,a)] \geq \inf_{\theta_i\in\Theta_i}d(\theta_1,\theta_2) = d(\Theta_1,\Theta_2)$이므로 성립.

4.7 Fano inequality

$F \subset \Theta$ 의 cardinality가 N이고 η-separated 일 경우에 다음이 성립한다.

$$R_{minimax} \geq \frac{\eta}{2}\Big(1 - \frac{\log 2 + \frac{1}{N}\sum_{\theta\in F}D_1(P_\theta||\bar{P})}{\log N}\Big)$$

where $\bar{P} := \sum_{\theta\in F}P_\theta/N$.

Lemma 4.1 *non negative number $a_1,\ldots,a_N$에 대해 다음 부등식이 성립한다.*

$$(\log N)\max_{1\leq i\leq N}a_i \leq \sum_{i=1}^N a_i\log\left(\frac{2a_i}{\bar{a}}\right)$$

where $\bar{a} := (a_1 + \cdots + a_N)/N$

Proof WLOG $\sum a_i = 1$, $a_1 = \max_{1 \le i \le N} a_i$ then,

$$\begin{aligned}(\log N) \max_{1 \le i \le N} a_i - \sum_{i=1}^{N} a_i \log\left(\frac{2a_i}{\bar{a}}\right) &= \sum_{i=2}^{N} a_i \log\left(\frac{\bar{a}}{2a_i}\right) + a_1 \log\left(\frac{N\bar{a}}{2a_1}\right) \\ &= \sum_{i=2}^{N} a_i \log\left(\frac{1}{2Na_i}\right) + a_1 \log\left(\frac{1}{2a_1}\right) \\ &\le 0\end{aligned}$$

Proof general testing bound에 의해

$$\begin{aligned}R_{minimax} &\ge (\eta/2)\bar{r}(P_\theta, \theta \in F) \\ &\ge 1 - \frac{\int \sum p_i/N \log \frac{2p_i}{\bar{p}}}{\log N} \\ &\ge 1 - \frac{\log 2 + \frac{1}{N}\sum_{\theta \in F} D_1(P_\theta || \bar{P})}{\log N}\end{aligned}$$

5 Bounds via f-divergences

5.1 f-divergence의 정의

Definition 5.1 *Let* $f : (0, \infty) \to \mathbb{R}$ *convex function with* $f(1) = 0$, $f(0) := \lim_{x \to 0+} f(x)$, $f(\infty) := \lim_{x \to \infty} f(x)/x$

$$D_f(P||Q) := Qf(p/q) + f'(\infty)P\{q = 0\}$$

Example 5.1
- $f(x) = |x-1|/2$*인 경우* $D_f(P||Q) = ||P - Q||_{TV}$
- *power divergence* $D_\alpha(P||Q) := D_{f_\alpha}(P||Q)$

$$f_\alpha(x) = \begin{cases} x^\alpha - 1 & \text{for } \alpha \notin [0,1] \\ 1 - x^\alpha & \text{for } \alpha \in (0,1) \\ x \log x & \text{for } \alpha = 1 \\ -\log x & \text{for } \alpha = 0 \end{cases}$$

$\alpha = 1$*인 경우*, $D_1(P||Q) = KL(P||Q)$

$\alpha = 1/2$*인 경우*, $D_1/2(P||Q) = 1 - \int \sqrt{pq}d\mu$ *(Hellinger distance)*

Theorem 5.2 *f가 convex on $(0,\infty)$이고, $f(1)=0$인 경우 다음 부등식이 성립한다.*

$$\sum_{i=1}^{N} D_f(P_i||Q) \geq g(\bar{r})$$

where $g(a) := f(N(1-a)) + (N-1)f(\frac{Na}{N-1})$

Proof 먼저 임의의 non negative $a_1,\dots,a_N$에 대해 다음 부등식이 성립한다.

$$\sum_{i=1}^{N} f(a_i) \geq f(\max_i a_i) + (N-1)f\Big(\frac{\sum_{i=1}^N a_i - \max_i a_i}{N-1}\Big)$$

왜냐하면 $\sum_{i=1}^N f(a_i) = f(a_1)+(N-1)\sum_{i\geq 2}(f(a_i)/(N-1)) \geq f(a_1)+(N-1)f\Big(\frac{\sum_{i\geq 2} p_i(x)}{N-1}\Big)$ 이 성립(by Jensen)

그리고 $P_1,\dots,P_n << Q$인 경우 가정(아닌 경우는 생략, 비슷한 아이디어)
$p_i := dP_i/dQ$라 할때, 다음 부등식이 성립한다.

$$\sum_{i=1}^{N} f(p_i) \geq f(\max_i p_i) + (N-1)f\Big(\frac{\sum_{i=1}^N p_i - \max_i p_i}{N-1}\Big)$$

이때 각 변을 Q로 적분한다면,

$$\int \sum_{i=1}^{N} f(p_i)dQ \geq \int f(\max_i p_i)dQ + (N-1)\int f\Big(\frac{\sum_{i=1}^N p_i - \max_i p_i}{N-1}\Big)dQ$$

좌변은 정의에 의해 $\sum_{i=1}^N D_f(P_i||Q)$가 되고,
$\int f(\max_i p_i)dQ \geq f(\int \max_i p_i dQ) = f(N(1-\bar{r}))$
$\int f\Big(\frac{\sum_{i=1}^N p_i - \max_i p_i}{N-1}\Big)dQ \geq f(\int \frac{\sum_{i=1}^N p_i - \max_i p_i}{N-1}dQ) = f(\frac{N\bar{r}}{N-1})$

Remark 위 정리의 사용
$\bar{r}$은 $[0,1-1/N]$사이의 값을 갖고 g는 $[0,1-1/N]$에서 감소하기 때문에 g에 역함수통해 $\bar{r}$의 lower bound를 구할 수 있고, 이에 따라 minimax lower bound계산가능

Example 5.2 • *Total variation의 경우*

$f(x) = |x-1|/2$이고 $g(x) = N-1-N\bar{r}$이 된다. 따라서

$$\bar{r} \geq 1 - \frac{1}{N} - \frac{\sum_{i=1}^N ||P_i - Q||_{TV}}{N}$$

- *Linear approximation for g*

  g는 *convex*이고 *non-increasing*이기 때문에 $a \in (0, 1 - 1/N]$에 대해,

$$g(\bar{r}) \geq g(a) + g'_L(a)(\bar{r} - a)$$

  따라서

$$\sum_{i=1}^{N} D_f(P_i || Q) \geq g(a) + g'_L(a)(\bar{r} - a)$$

$$\bar{r} \geq a + \frac{\sum D_f(P_i || Q) - g(a)}{g'_L(a)}$$