STA 3000 Lecture 2.

---

e.g. stat learning $\quad X = (Z_i, Y_i)_{i=1}^n \quad \mathcal{A} = \text{functions}$

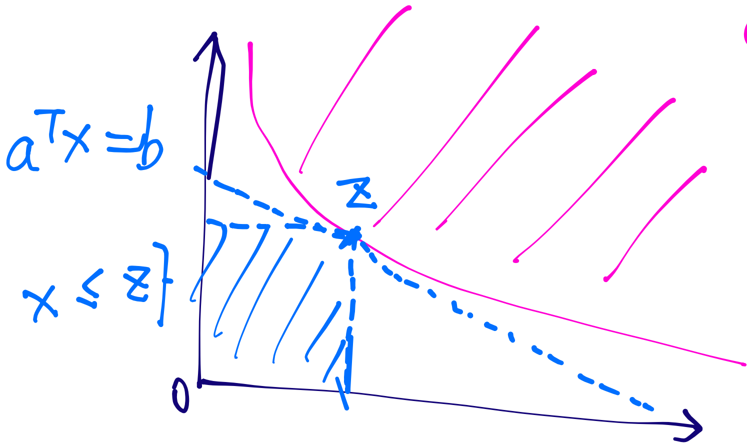
$$L(\theta, a) = \mathbb{P}_\theta\big(a(Z) \neq Y\big).$$

in ML $\quad \mathbb{1}\{a(Z) \neq Y\}$

Fact. Admissible rules are Bayes when $K = |\Theta| < +\infty$

$$C = \left\{ \big(R(\theta_j, \delta)\big)_{j=1}^K : \begin{array}{c}\text{decision}\\ \text{rule}\\ \delta\end{array} \right\}$$

$z \in C.$



Proof: C is convex.

$x_1, x_2 \in C, \quad \lambda \in (0,1). \qquad \lambda x_1 + (1-\lambda) x_2 \in C.$

$$\delta_1, \delta_2 \qquad \tilde{\delta} = \begin{cases} \delta_1 & \text{w.p. } \lambda \\ \delta_2 & \text{w.p. } (1-\lambda) \end{cases}$$

By admissibility,

$$C \cap \{x : x \leq z\} = \{z\}$$

$$\exists\ a \in \mathbb{R}^k,\ b \in \mathbb{R} \quad \text{s.t.} \quad a^T x \begin{cases} \geq b & \text{when } x \in C \\ \leq b & \text{when } x \leq z. \end{cases}$$

$$\pi = \frac{a}{\|a\|_1} \qquad (a_i \geq 0)\ \ 1 \leq i \leq k.$$

Minimax rules.

$$\inf_{\delta} \sup_{\theta} R(\theta, \delta)$$

Alice and Bob playing zero-sum game

$a \in A$ $b \in B$ Alice's payoff $R(a,b)$

$a \sim \pi_a$ $b \sim \pi_b$ Bob's payoff $-R(a,b)$

Scheme 1: Alice choose π_a, then Bob choose π_b

Scheme 2: Bob choose π_b, then Alice choose π_a.

Thm (weak duality).

$$\sup_{\pi_a} \inf_{\pi_b} \mathbb{E}[R(a,b)] \leq \inf_{\pi_b} \sup_{\pi_a} \mathbb{E}[R(a,b)].$$

Proof: For any a ~~fixed~~

$$\inf_{\pi_b} \mathbb{E}[R(a,b)] \leq \inf_{\pi_b} \sup_{\pi_a} \mathbb{E}[R(a,b)].$$

Take sup over a.

Thm (von Neumann). Finite A, B.

$$\sup_{\pi_A} \inf_{\pi_B} \mathbb{E}[R(a,b)] = \inf_{\pi_B} \sup_{\pi_A} \mathbb{E}[R(a,b)].$$

Alice: nature choose prior π_A

Bob: choose (randomized) decision rules.

Finite $\Theta, \mathcal{X}, \mathcal{A}$: $\inf \sup = \sup \inf$

(Can be extended: compact subset of $\mathbb{R}^d$)

$\Downarrow$

minimax rules are Bayes.

Prop — Constant risk Bayes rules δ are minimax.

Proof — For any other decision rule δ'

$$\sup_\theta R(\theta,\delta') \geq \int_\Theta R(\theta,\delta')\,\pi(d\theta)$$

$$\geq \int_\Theta R(\theta,\delta)\,\pi(d\theta) \quad (\delta \text{ is Bayes rule under } \pi).$$

$$= \sup_\theta R(\theta,\delta).$$

eg. $X \sim \mathrm{Binom}(n,p)$. $\pi = \mathrm{Beta}(\alpha,\beta)$.

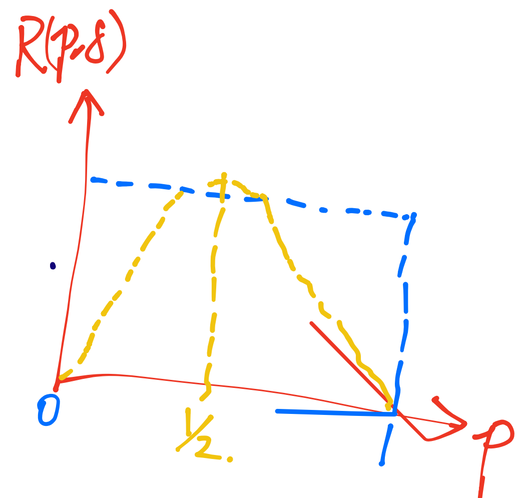
$(L(p,a) = (a-p)^2)$ $\delta(X) = \dfrac{\alpha+X}{\alpha+\beta+n}$

$$R(p,\delta) = \frac{np(1-p) + (\alpha(1-p)-\beta p)^2}{(\alpha+\beta+n)^2}$$

$$\alpha = \beta = \sqrt{n}/2.$$

$$R_{minimax} = \frac{1}{4(1+\sqrt{n})^2}.$$

$\delta'(x) = x/n$. $R(p,\delta') = \dfrac{p(1-p)}{n}$



Prop — If δ satisfies

$\exists (\pi_j, r_j)_{j=1}^{+\infty}$ s.t. $\liminf_{j\to+\infty} r_j \geq \sup_\theta R(\theta,\delta)$

then δ is minimax.

Proof:

$$\sup_\theta R(\theta,\delta') \geq \liminf_{j\to+\infty} \int_\Theta R(\theta,\delta')\,\pi_j(d\theta)$$

$$\geq \liminf_{j\to+\infty} \int_\Theta R(\theta,\delta_\pi)\,\pi_j(d\theta) \quad (= r_j).$$

$$\geq \sup_\theta R(\theta,\delta).$$

e.g. $X_1, X_2, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\theta,1)$.

$\pi_k = \mathcal{N}(0,k)$ $\quad r_{\pi_k} = \frac{k}{kn+1}$.

$k=1,2,\dots$

$R(\theta,\delta) = \frac{1}{n} \quad (\forall\theta)$.

$\cancel{\delta(X) = \frac{1}{n}\sum_1^n X_i}$

$$\lim_{k\to+\infty} r_{\pi_k} = \sup_{\theta\in\mathbb{R}} R(\theta,\delta) = \frac{1}{n}.$$

Sufficiency.

e.g. $X_1, X_2, \dots, X_n \overset{iid}{\sim} \mathrm{Ber}(\theta)$

$$T(X) = \sum_{i=1}^n X_i.$$

Conditionally on $\{T=t\}$, $\{X_i\}_{i=1}^n$ uniform distribution over all the seq s.t. $\sum_{i=1}^{n} x_i = t$.

Def. $X \sim P_\theta \in \mathcal{P}_\Theta$, we call $T = T(X)$ sufficient

if $\forall t, \theta$, cond. distribution of X under P_θ

given $T = t$ does not depend on θ.

Thm. $T(X)$ is sufficient, δ is a decision rule

$\exists \tilde{\delta}(T(X))$ s.t. $\delta(X) \overset{d}{=} \tilde{\delta}(T(X))$.

μ_t = cond. distr of X given $T = t$

$\tilde{\delta}$ = sample $X' \sim \mu_{T(X)}$

output $\delta(X')$.

Thm (Rao - Blackwellization)

If $L(\theta, \cdot)$ is a convex function.

δ estimator of $g(\theta)$

$$\eta(T(X)) = \mathbb{E}[\delta(X) \mid T].$$

$$R(\theta, \eta) \leq R(\theta; \delta).$$

Proof :- $R(\theta, \eta) = \mathbb{E}\, L(\theta, \mathbb{E}[\delta(X) \mid T])$

$\leq \mathbb{E}\, \mathbb{E}[L(\theta, \delta(X)) \mid T] = R(\theta, \delta)$.

eg. $X_1, X_2, \cdots, X_n \overset{iid}{\sim} N(\theta, 1)$

$$P_\theta(X_1, \cdots X_n) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i - \theta)^2\right)$$

$$\propto \exp\left(-\frac{n\theta^2}{2} + \theta\, T(X) - \frac{1}{2}\sum_{1}^{n} X_i^2\right).$$

$T = \sum_{i=1}^{n} X_i$. $\quad X_1, \cdots X_n \mid T(X) \sim N(T(X), 1)$.

Factorization Thm.

$\{P_\theta : \theta \in \Theta\}$, $\forall \theta$, density $p_\theta = \frac{dP_\theta}{d\mu}$ $\exists$.

Then T is sufficient for P_Θ

$\Longleftrightarrow$ $\exists\, g_\theta, h$ s.t. $p_\theta(x) = g_\theta(T(x)) \cdot h(x)$.

Proof "$\Longleftarrow$": $\cancel{P_\theta}(x \mid T(X) = t) = \dfrac{\cancel{g_\theta}(t) \cdot h(x) \cdot \mathbb{1}\{T(x)=t\}}{\int_{T(z)=t} \cancel{g_\theta}(t) \cdot h(z) \, d\mu(z)}$.

$$= \frac{h(x) \cdot \mathbb{1}\{T(x)=t\}}{\int_{T(z)=t} h(z)\, d\mu(z)} \quad \text{indp of } \theta.$$

e.g. Exponential family.

$$P_\theta(x) = \underbrace{\exp\left(\eta(\theta)^T T(x) - B(\theta)\right)}_{g_\theta(T(x))} \cdot \underbrace{h(x)}_{h(x)}.$$

Covering.

— normal. $\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$

$$\propto \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x + \mu^T\Sigma^{-1}x - \frac{1}{2}\mu^T\Sigma^{-1}\mu\right).$$

$$T(x) = \begin{bmatrix} x \\ vec(xx^T) \end{bmatrix} \in \mathbb{R}^{d+d^2}$$

$$\eta = \begin{bmatrix} \Sigma^{-1}\mu \\ vec(\Sigma^{-1}) \end{bmatrix} \in \mathbb{R}^{d+d^2}.$$

— Binomial $\binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$

$$= \binom{n}{x} \cdot \exp\left(\left(\log\frac{p}{1-p}\right)\cdot x + n\cdot\log(1-p)\right).$$

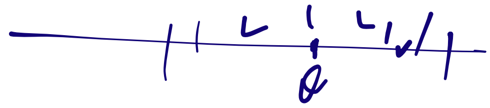
$$T(x) = x \qquad \eta = \log\frac{p}{1-p}.$$

Property. $x_i \overset{iid}{\sim} P_\theta$ exp family $i=1,2,\dots n.$

$$P_\theta(x_1, \dots x_n) = \prod_{i=1}^n h(x_i)\cdot \exp\left(\eta(\theta)^T\cdot \sum_{i=1}^n T(x_i) - n\,B(\theta)\right).$$

$$T(X_1, \dots, X_n) = \sum_{i=1}^{n} T(X_i).$$

e.g. $\text{Unif}([\theta-1, \theta+1])$.



$$T(X) = (X_{(1)}, X_{(n)}).$$

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}.$$

$$P_\theta(X_1, \dots, X_n) = 2^{-n} \cdot \mathbb{1}\{\theta-1 \le X_1 \cdots X_n \le \theta+1\}.$$

$$= 2^{-n} \mathbb{1}\{\theta-1 \le X_{(1)} \le X_{(n)} \le \theta+1\}.$$

e.g. $X_1, X_2, \dots, X_n \overset{iid}{\sim} P$.

$$T(X) = (X_{(1)}, X_{(2)}, \dots, X_{(n)}).$$

Unbiased estimation.

Def — δ unbiased $g(\theta)$.

$$\mathbb{E}_\theta[\delta(X)] = g(\theta) \quad (\forall \theta \in \Theta).$$

Def — (UMVU) $\forall \delta'$ unbiased,

we have $\text{var}_\theta(\delta) \le \text{var}_\theta(\delta')$ $(\forall \theta \in \Theta)$.

Uniformly Minimum Variance Unbiased

Def – "Complete statistics"

$T(X)$ is complete for $\mathcal{P} = (P_\theta : \theta \in \Theta)$ if

$\mathbb{E}_\theta[f(T)] = 0 \ (\forall \theta)$ implies $f(T) = 0$ a.s.

(for any f).

e.g. $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathrm{Unif}([0, \theta])$.

$$T(X) = \max_{1 \le i \le n} X_i.$$

$$\mathbb{P}_\theta(T \le t) = (t/\theta)^n \qquad \forall t \in [0, \theta].$$

$$\mathbb{E}_\theta[f(T)] = \frac{n}{\theta^n} \cdot \int_0^\theta f(t) \cdot t^{n-1} dt = 0 \qquad (\forall \theta \in \mathbb{R}^+).$$

$$\int_0^\theta f(t)\, t^{n-1} dt = 0 \implies f(\theta) = 0 \quad (\forall \theta).$$

Def – $p_\theta(x) = \exp(\eta(\theta)^T T(x) - B(\theta)) \cdot h(x)$. $(\theta \in \Theta)$.

is called full-rank if $\eta(\Theta)$ has an interior point.

$(T_1(x), T_2(x), \ldots, T_d(x))$ linearly indep.

(If x is an int pt of A

$\exists \delta > 0$ $B(x, \delta) \subseteq A$).

Thm – Full-rank exp family $\implies$ T is sufficient & complete