Stat 301

Decision Theory (A. Wald)

$(P_\theta : \theta \in \Theta)$ $X_1, \dots, X_n \overset{iid}{\sim} P_\theta$

$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$.

loss function $L(\hat{\theta}, \theta)$. e.g. $\|\hat{\theta} - \theta\|^2$

risk $R(\hat{\theta}, \theta) = \mathbb{E}_\theta L(\hat{\theta}, \theta) = \int L(\hat{\theta}(x), \theta) P_\theta(x) dx$.

Theorem (Rao - Blackwell). Assume $L(\hat{\theta}, \theta)$ is convex in $\hat{\theta}$, for any $\hat{\theta}$ and any sufficient T, define $\tilde{\theta} = \mathbb{E}_\theta(\hat{\theta} \mid T)$, then $R(\tilde{\theta}, \theta) \le R(\hat{\theta}, \theta)$.

Proof:
$$L(\tilde{\theta},\theta) = L\big(E_\theta(\hat{\theta}\,|\,T),\theta\big)$$
$$\le E_\theta\big(L(\hat{\theta},\theta)\,|\,T\big) \quad \text{Jensen inequality.}$$

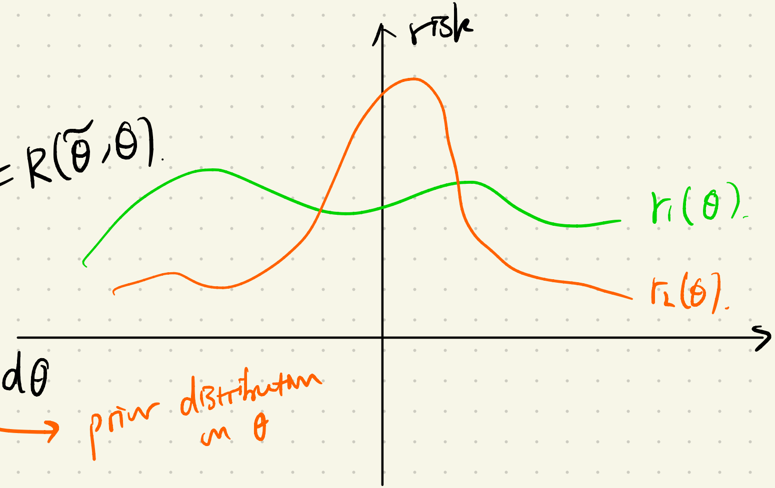
$$\Rightarrow \quad E_\theta L(\tilde{\theta},\theta) \le E_\theta\, E_\theta\big(L(\hat{\theta},\theta)\,|\,T\big) = E_\theta L(\hat{\theta},\theta) \qquad \square$$

---

two estimators $\hat{\theta}$, $\tilde{\theta}$.

$r_1(\theta) = R(\hat{\theta},\theta)$

$r_2(\theta) = R(\tilde{\theta},\theta)$



average risk. $\int R(\hat{\theta},\theta)\,\pi(\theta)\,d\theta$ — prior distribution on θ

maximum risk. $\sup_{\theta\in\Theta} R(\hat{\theta},\theta)$.

def: ① $\hat{\theta}$ is a Bayes estimator if $\hat{\theta} = \underset{\hat{\theta}}{\operatorname{argmin}} \int R(\hat{\theta}, \theta)\, \pi(\theta)\, d\theta$

$\Leftrightarrow \forall \tilde{\theta}, \quad \int R(\hat{\theta}, \theta)\pi(\theta)\, d\theta \le \int R(\tilde{\theta}, \theta)\pi(\theta)\, d\theta.$

②. $\hat{\theta}$ is a minimax estimator if $\hat{\theta} = \underset{\hat{\theta}}{\operatorname{argmin}} \sup_{\theta \in \Theta} R(\hat{\theta}, \theta).$

$\Leftrightarrow \forall \tilde{\theta}, \quad \sup_{\theta \in \Theta} R(\hat{\theta}, \theta) \le \sup_{\theta \in \Theta} R(\tilde{\theta}, \theta).$

Bayes estimator.

$$\int R(\hat{\theta}, \theta)\, \pi(\theta)\, d\theta = \iint L(\hat{\theta}(x), \theta)\, \underbrace{P_\theta(x)}_{= p(x|\theta)}\, \pi(\theta)\, dx\, d\theta.$$

$\rightarrow$ joint distribution of (x, θ).

$$p(x|\theta)\pi(\theta) = \frac{p(x|\theta)\pi(\theta)}{\int p(x|\theta)\pi(\theta)\, d\theta} \cdot \int p(x|\theta)\pi(\theta)\, d\theta$$

$$= \underbrace{\pi(\theta|x)}_{\text{posterior}} \cdot \underbrace{m(x)}_{\text{marginal of } x}.$$

$$\int R(\hat\theta, \theta)\,\pi(\theta)\,d\theta = \iint \underbrace{L(\hat\theta(x), \theta)\,\pi(\theta|X)\,d\theta}_{\text{a function of } X}\, m(x)\,dx. \quad (\text{Fubini}).$$

Claim: $\hat\theta_\pi(x) = \arg\min_a \int L(a, \theta)\,\pi(\theta|X)\,d\theta$ is Bayes.

Proof: want to show for any $\hat\theta$

$$\int R(\hat\theta_\pi, \theta)\,\pi(\theta)\,d\theta \le \int R(\hat\theta, \theta)\,\pi(\theta)\,d\theta.$$

$$\begin{aligned}\int R(\hat\theta_\pi, \theta)\,\pi(\theta)\,d\theta &= \iint L(\hat\theta_\pi(x), \theta)\,\pi(\theta|X)\,d\theta\, m(x)\,dx \\ &\le \iint L(\hat\theta(x), \theta)\,\pi(\theta|X)\,d\theta\, m(x)\,dx. \\ &= \int R(\hat\theta, \theta)\,\pi(\theta)\,d\theta \end{aligned}$$

□.

an important example. $\Theta \subseteq \mathbb{R}$. $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$.

$$\hat{\theta}_\pi(X) = \underset{a}{\operatorname{argmin}} \int (a-\theta)^2 \pi(\theta|X) d\theta.$$

$$= \underset{a}{\operatorname{argmin}} \ \mathbb{E}\left((a-\theta)^2 \mid X\right)$$

$$= \mathbb{E}(\theta|X).$$

a r.v. $Y \in \mathbb{R}$, $\mu \in \mathbb{R}$. $\mathbb{E}(Y-\mu)^2 = \mathrm{Var}(Y) + (\mathbb{E}Y - \mu)^2$.

$$\mathbb{E}\left((a-\theta)^2 | X\right) = \mathrm{Var}(\theta|X) + \left(\mathbb{E}(\theta|X) - a\right)^2.$$

e.g. $X_1, \cdots, X_n \overset{iid}{\sim} \mathrm{Bern}(p)$ loss $(\hat{p}-p)^2$.

prior $\pi = \mathrm{Beta}(\alpha, \beta)$., $\pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$.

$$p | X_1, \cdots, X_n \sim \mathrm{Beta}\left(\sum_{i=1}^n X_i + \alpha, \sum_{i=1}^n (1-X_i) + \beta\right).$$

Bayes estimator $\hat{p} = \mathbb{E}(p \mid X_1, \dots, X_n) = \dfrac{\sum_{i=1}^{n} X_i + \alpha}{n+\alpha+\beta}$

$$R(\hat{p}, p) = \mathbb{E}_p(\hat{p}-p)^2 = \mathrm{Var}(\hat{p}) + (\mathbb{E}\hat{p} - p)^2.$$

$$= \left(\frac{n}{n+\alpha+\beta}\right)^2 \frac{p(1-p)}{n} + \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)^2 \left(\frac{\alpha}{\alpha+\beta} - p\right)^2.$$

minimax estimator: $\hat{\theta}_{minimax} = \underset{\hat{\theta}}{\operatorname{argmin}} \sup_{\theta \in \Theta} R(\hat{\theta}, \theta).$

Theorem: If for some π, $\hat{\theta}$ satisfies.

$$\sup_{\theta \in \Theta} R(\hat{\theta}, \theta) = \inf_{\tilde{\theta}} \int R(\tilde{\theta}, \theta)\, \pi(\theta)\, d\theta.$$

then $\hat{\theta}$ is minimax.

Proof: $\forall \tilde{\theta}$, $\sup_{\theta \in \Theta} R(\tilde{\theta}, \theta) \geq \int R(\tilde{\theta}, \theta) \pi(\theta) d\theta$

$$\geq \inf_{\tilde{\theta}} \int R(\tilde{\theta}, \theta) \pi(\theta) d\theta$$

$$= \sup_{\theta \in \Theta} R(\hat{\theta}, \theta) \qquad \square$$

Corollary: If $\hat{\theta} = \hat{\theta}_\pi$ for some π and $R(\hat{\theta}_\pi, \theta)$ is constant over $\theta \in \Theta$.

then $\hat{\theta}$ is minimax.

Proof: $\sup_{\theta \in \Theta} R(\hat{\theta}, \theta) = \int R(\hat{\theta}, \theta) \pi(\theta) d\theta$

$$= \inf_{\tilde{\theta}} \int R(\tilde{\theta}, \theta) \pi(\theta) d\theta$$

by theorem, $\hat{\theta}$ is minimax.

e.g. $X_1, \cdots, X_n \overset{iid}{\sim}$ Bernoulli(p) loss $(\hat{p}-p)^2$.

$$\hat{p} = E(p \mid X_1, \cdots, X_n) = \frac{\sum_{i=1}^{n} X_i + \alpha}{n+\alpha+\beta}$$

$$R(\hat{p}, p) = E_\theta(\hat{p}-p)^2$$

$$= \left(\frac{n}{n+\alpha+\beta}\right)^2 \frac{p(1-p)}{n} + \left(\frac{\alpha+\beta}{n+\alpha+\beta}\right)^2 \left(\frac{\alpha}{\alpha+\beta} - p\right)^2$$

$$= \left[\left(\frac{\alpha+\beta}{n+\alpha+\beta}\right)^2 - \frac{1}{n}\left(\frac{n}{n+\alpha+\beta}\right)^2\right] p^2 \rightarrow 0$$

$$+ \left[\frac{1}{n}\left(\frac{n}{n+\alpha+\beta}\right)^2 - \left(\frac{\alpha+\beta}{n+\alpha+\beta}\right)^2 \frac{2\alpha}{\alpha+\beta}\right] p \rightarrow 0$$

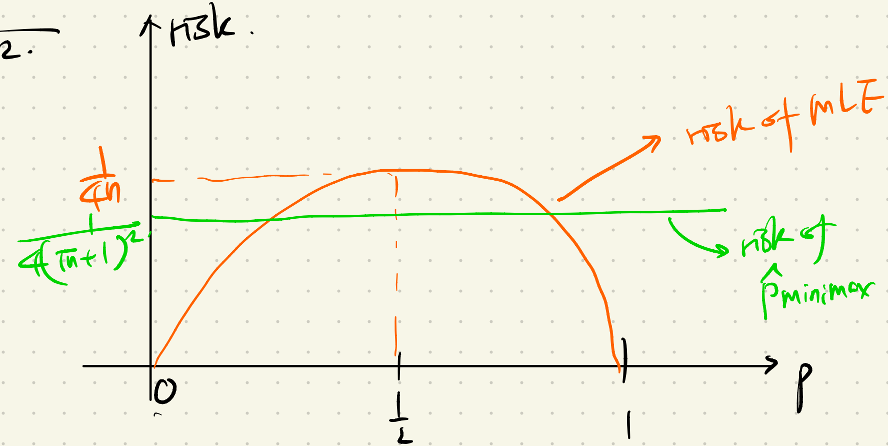
$$+ \left(\frac{\alpha+\beta}{n+\alpha+\beta}\right)^2 \left(\frac{\alpha}{\alpha+\beta}\right)^2.$$

$$\begin{cases} (\alpha+\beta)^2 = n \\ 2\alpha(\alpha+\beta) = n \end{cases} \Rightarrow \alpha = \beta = \frac{\sqrt{n}}{2}.$$

$$\hat{p}_{minimax} = \frac{\sum_{i=1}^{n} X_i + \frac{\sqrt{n}}{2}}{n + \sqrt{n}}, \qquad \hat{p}_{MLE} = \bar{X}$$

$$R(\hat{p}_{MLE}, p) = \mathbb{E}_p(\hat{p}-p)^2 = \frac{p(1-p)}{n}, \quad \max_p R(\hat{p}_{MLE}, p) = \frac{1}{4n}.$$

$$R(\hat{p}_{minimax}, p) = \frac{1}{4(\sqrt{n}+1)^2}.$$



e.g. $X_1, \cdots, X_n \overset{iid}{\sim}$ Bernoulli(p) loss $L(\hat{p}, p) = \dfrac{(\hat{p}-p)^2}{p(1-p)}$

$\pi(p) = 1.$

$$\hat{p}(x) = \underset{a}{\operatorname{argmin}} \int \frac{(a-p)^2}{p(1-p)} \pi(p|X)\, dp. = \underset{a}{\operatorname{argmin}} \int (a-p)^2 \frac{\pi(p|x)}{p(1-p)}\, dp.$$

$$\frac{\pi(p|X)}{p(1-p)} \propto p^{\sum_{i=1}^n X_i - 1} (1-p)^{\sum_{i=1}^n (1-X_i) - 1} = \text{Beta}\left(\sum_{i=1}^n X_i, \sum_{i=1}^n (1-X_i)\right).$$

$$\hat{p}(x) = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i + \sum_{i=1}^n (1-X_i)} = \bar{X} = \hat{P}_{MLE}$$

$$R(\hat{p}, p) = E_p \frac{(\bar{X}-p)^2}{p(1-p)} = \frac{1}{n} \quad \text{constant}.$$

$\hat{p} = \bar{X}$ is minimax.

e.g. $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. $\mu \in \mathbb{R}$ loss $(\hat{\mu} - \mu)^2$.

Q: is $\bar{X}$ minimax? $R(\bar{X}, \mu) = \mathbb{E}_\mu(\bar{X} - \mu)^2 = \frac{\sigma^2}{n}$.