Lecture 2 (9/1/2020)

Outline

1) Review
2) Estimation
3) Loss & Risk Functions
4) Comparing estimators

Estimation

A <u>statistical model</u> is a family of candidate probability distributions $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ for some random variable $X \sim P_\theta$

$X \in \mathcal{X}$ called <u>data</u>, observed

θ called <u>parameter</u>, unobserved

[could be infinite dimensional, e.g. density fcn.]

For now, θ fixed and unknown

<u>Goal of estimation</u>: Observe $X \sim P_\theta$, guess value of some <u>estimand</u> $g(\theta)$

<u>Ex. 3.1</u> Flip a biased coin n times

$\theta \in [0,1]$ prob. of heads

X = # heads after n flips

$\sim \text{Binom}(n, \theta)$

$$p_\theta(x) = \theta^x (1-\theta)^{n-x} \binom{n}{x}$$

density wrt counting measure on $\mathcal{X} = \{0, .., n\}$

A **statistic** is any function $T(X)$ of data X
[NOT of X and θ]

An **estimator** $\delta(X)$ of $g(\theta)$ is a statistic which is intended to guess $g(\theta)$

Ex 3.1 (Cont'd) Natural estimator: $\delta_0(X) = \hat{\lambda}$

Question: is it a good estimator? Is another better?

Loss & Risk

Loss function $L(\theta, d)$ measures how bad an estimate is

Ex $L(\theta, d) = (d - g(\theta))^2$ **squared-error loss**

Typical properties:

$L(\theta, d) \geq 0 \quad \forall \theta, d$

$L(\theta, g(\theta)) = 0 \quad \forall \theta$ [(no loss from a perfect guess)]

[Loss is random, reflects both whether we choose a good estimator **and** whether we are lucky]

__Risk function__ is expected loss (risk) as a function of θ for an estimator $\delta(\cdot)$

$$R(\theta; \delta(\cdot)) = \mathbb{E}_\theta\left[L(\theta, \delta(X))\right]$$

tells us which parameter value is in effect, **NOT** "what randomness to integrate over"

__Ex 3.1__ (Cont'd)

$$\delta_0(X) = \frac{X}{n}$$

$$\mathbb{E}_\theta\left[\frac{X}{n}\right] = \theta \qquad \text{(unbiased)}$$

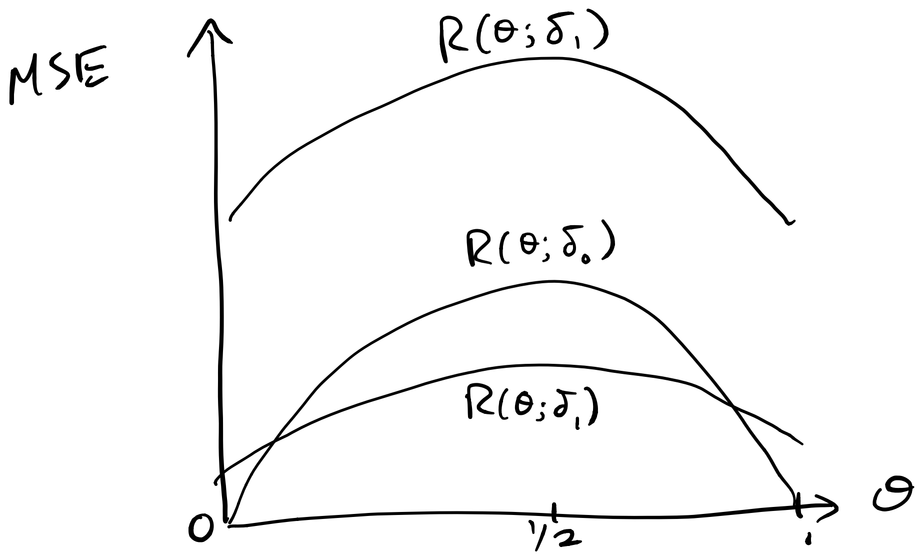
$$\Rightarrow \text{MSE}(\theta; \delta_0) = \text{Var}_\theta\left(\frac{X}{n}\right)$$

$$= \frac{1}{n}\theta(1-\theta)$$

__Other choices__:

$$\delta_1(X) = \frac{X+3}{n} \qquad \text{(stupid)}$$

$$\delta_2(X) = \frac{X+3}{n+6} \qquad \text{(6 "pseudo-flips," 3 heads)}$$



Comparing Estimators

We know $\delta_1(X)$ is bad, δ_0 vs. δ_2 more ambiguous

An estimator δ is <u>inadmissible</u> if $\exists\, \delta^*$ with

a) $R(\theta;\delta^*) \le R(\theta;\delta) \quad \forall\, \theta \in \Theta$

b) $R(\theta;\delta^*) < R(\theta;\delta) \quad$ some $\theta \in \Theta$

δ_1 is inadmissible

[We can rule out very bad estimators like δ_1, but it is virtually always impossible to find a single uniformly best estimator.

Thought experiment: $\delta_3(X) \equiv 2/3$ best if $\theta = \frac{2}{3}$]

Strategies to resolve ambiguity:

1) Summarize risk function by a scalar

a) <u>Average-case risk</u>: minimize

$$\int_{\Theta} R(\theta;\delta)\, d\Lambda(\theta) \quad \text{w.r.t. some measure } \Lambda$$

$\rightsquigarrow$ Bayes estimator, Λ called prior

- Λ "improper" if not normalizable
- This gives a frequentist motivation for Bayes methods
- δ_2 is a Bayes estimator

b) <u>Worst-case risk</u>: minimize

$$\sup_{\theta \in \Theta} R(\theta; \delta)$$

$\rightsquigarrow$ Minimax estimator, closely related to Bayes

Why not best-case risk? Again consider δ_3

2) Constrain choice of estimator

a) Only consider unbiased δ:

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

Rules out $\delta_1, \delta_2, \delta_3$.

δ_1 is best unbiased estimator.