Lecture 2 (8/29/2023)

Outline

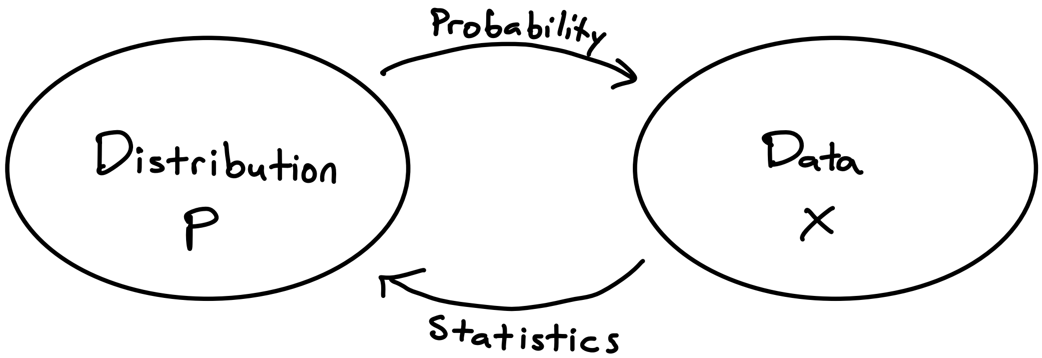
1) Statistical models

2) Estimation

3) Decision theory

Statistical Models

Probability vs statistics



Probability: Distribution P fully specified

What can we say about $X \sim P$?

Deductive

Statistics: Observe data X from unknown dist. P

What can we conclude about P?

Inductive

"the model"

Statistical model Family $\mathcal{P}$ of candidate probability distributions for data X

Assume $X \sim P$ for some $P \in \mathcal{P}$

X yields evidence about which P (hopefully)

Parametric vs. Nonparametric

Parametric model dists indexed by parameter $\theta \in \Theta$

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$$

Typically $\Theta \subseteq \mathbb{R}^d$, d called model dimension

Example $X \sim \text{Binom}(n, \theta)$ for $\theta \in [0,1]$

n "known," θ "unknown" (by analyst)

$$\mathcal{P} = \{\text{Binom}(n,\theta) : \theta \in [0,1]\}$$

Nonparametric model no natural way to index $\mathcal{P}$

Still usually makes assumptions, e.g.

- independence
- shape constraints (e.g. unimodal density)

(independent & ident. distr.)

Example $X_1, \ldots, X_n \overset{\text{iid}}{\sim} P$ P any distr. on $\mathbb{R}$

$$\mathcal{P} = \{P^n : P \text{ is a distr. on } \mathbb{R}\} \quad (\text{for } X = (X_1, \ldots, X_n))$$

We can use "parametric notation" $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ wlog

(could take $\theta = P$, $\Theta = \mathcal{P}$)

Bayesian vs. "Frequentist" Inference

Assume $X \sim P_\theta$ $\quad \theta$ unknown

<u>Bayesian assumption</u> : θ random with known dist.

Inference = calculating dist. $(\theta | X)$ (<u>posterior</u>)

Considered a strong assumption

(will consider interp., pros & cons later)

Alternate perspective : treat θ as fixed, unknown

Methods designed without knowledge of θ

Study <u>frequency properties</u> as θ varies

Estimation

Setup Model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ (wlog)

Estimand $g(\theta)$ (something we want to know)

Observe X, calculate **estimate** $\delta(X)$

$\delta(\cdot)$ called **estimator**.

We want to **evaluate** & **compare** estimators

Example Flip a biased coin n times

$\theta \in [0,1]$ probability of heads

$X = \#$ heads

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

Goal: estimate θ

Natural estimator is $\delta_0(X) = \frac{X}{n}$

How good is it?

Loss and Risk

Loss function $L(\theta, d)$

Disutility of guessing $g(\theta) = d$

Typically non-negative, with $L(\theta, d) = 0$ iff $d = g(\theta)$

[Different for every realization]

Squared error loss: $L(\theta, d) = (d - g(\theta))^2$

Risk function: expected loss of an estimator

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

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

Risk for sq. error loss is **mean squared error** (MSE)

$$MSE(\theta; \delta(\cdot)) = \mathbb{E}_\theta \left[(\delta(X) - g(\theta))^2 \right]$$

Binomial example

What is $MSE(\theta; \delta_0)$? $\quad \left(\delta_0(X) = \frac{X}{n} \right)$

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

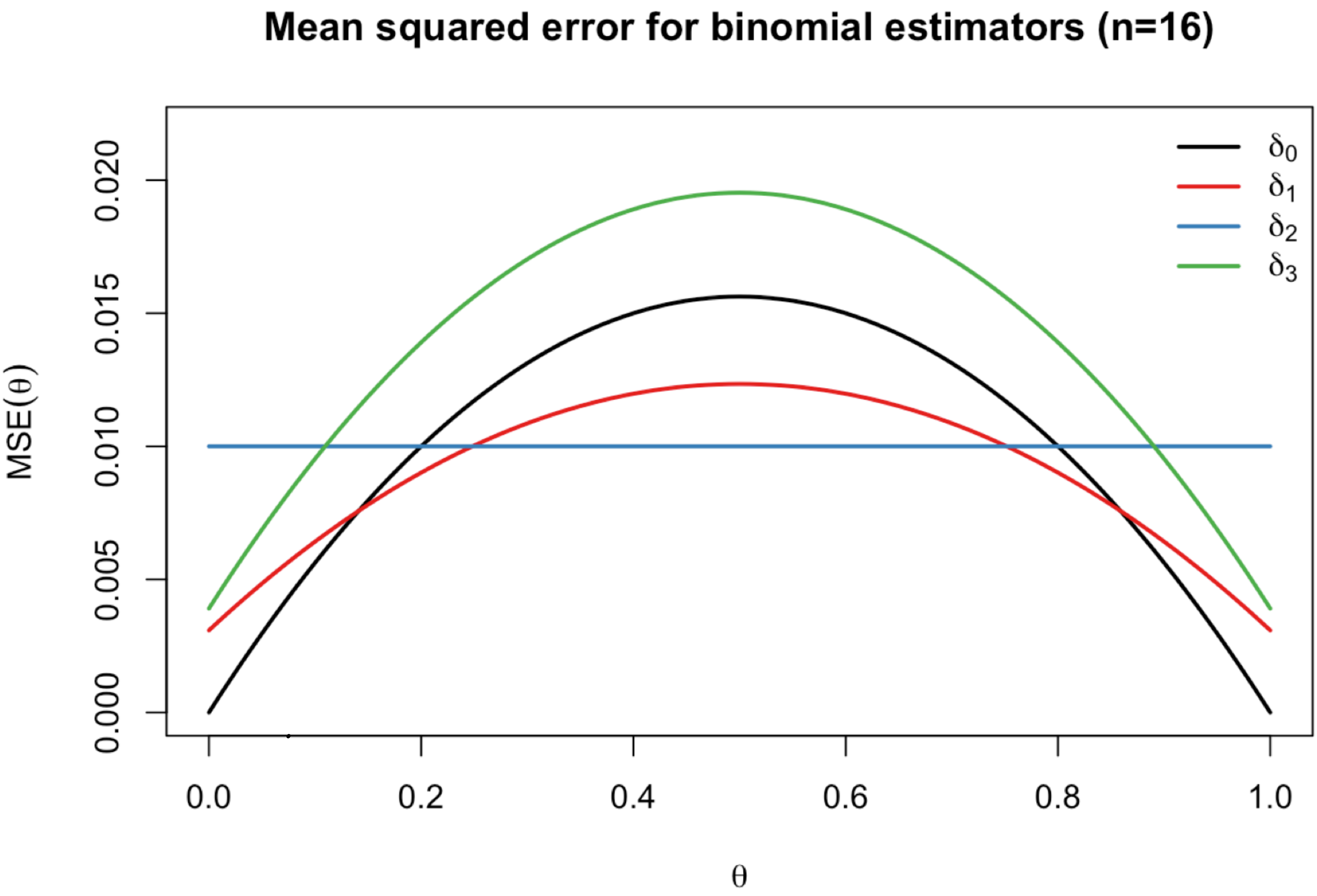
$$\Rightarrow MSE(\theta; \delta_0) = \mathbb{E}_\theta\left[\left(\frac{X}{n} - \theta\right)^2\right]$$

$$= \mathrm{Var}_\theta\left(\frac{X}{n}\right)$$

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

Other possibilities (based on adding "pseudo-flips")

$\delta_1(X) = \frac{X+1}{n+2}$ $\qquad \delta_2(X) = \frac{X+2}{n+4}$ $\qquad \delta_3(X) = \frac{X+1}{n}$



Comparing estimators

We want to choose δ to minimize R

... but this is generally not possible

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

a) $R(\theta; \delta^*) \leq R(\theta, \delta)$ for all θ

b) $R(\theta, \delta^*) < R(\theta, \delta)$ for some θ

We say δ^* <u>strictly dominates</u> δ

δ_1 is inadmissible because δ_0 dominates it

Is there any <u>uniformly</u> (all θ) best estimator for the binomial example?

Resolving ambiguity

Main strategies to resolve ambiguity:

1) Summarize risk function by a scalar:

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

   Minimize $\int_{\Theta} R(\theta;\delta)\, d\pi(\theta)$

   for some measure π, called <u>prior</u>

   If π is probability measure,

   same as $\mathbb{E}_{\theta\sim\pi}\left[R(\theta;\delta)\right]$

   $\rightsquigarrow$ Bayes estimator

   <u>Binomial</u>: δ_1 is Bayes wrt $\pi=\lambda$ on $[0,1]$

   δ_2 also Bayes wrt $\pi=\mathrm{Beta}(2,2)$

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

   Minimize $\sup_{\theta} R(\theta;\delta)$

   $\rightsquigarrow$ Minimax estimator

   Closely related to Bayes

   <u>Binomial</u>: δ_2 is minimax (for $n=16$)

2) Restrict choices of estimatos

a) Restrict to unbiased estimators:

$$\mathbb{E}_\theta[\delta(X)] = g(\theta) \quad \text{for all } \theta$$

<u>Binomial</u>: δ_0 is best unbiased estimator