

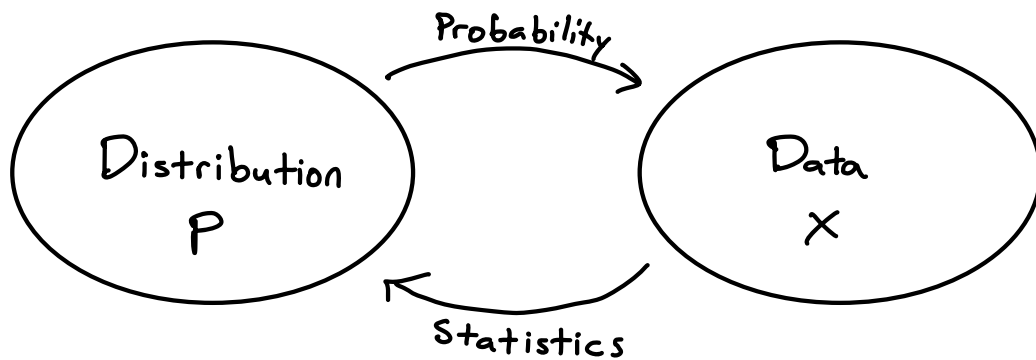
Statistical models and decisions

Outline

- 1) Statistical models
- 2) Estimation
- 3) Comparing estimators

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

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

Assume $X \sim P$ for some $P \in \mathcal{P}$ (don't know which)
 X yields evidence about which P (hopefully)

Coin flipping

Recall: 48 humans tossed coins $n = 350,757$ total times
 $X = 178,079$ landed same-side up.

Model 1:

All flips independent, with same probability $\theta \in (0,1)$

$$\Rightarrow X \sim P_{\theta} = \text{Binom}(n, \theta)$$

θ indexes P known unknown (varies over model)

Probability mass function $p_{\theta}(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$
for $x = 0, 1, \dots, n$

$$\mathcal{P} = \{P_{\theta} : \theta \in (0,1)\}$$

Model 2: Flippers have different biases

$$\Rightarrow X_i \stackrel{\text{ind.}}{\sim} \text{Binom}(n_i, \theta_i) \quad i=1, \dots, 48$$

flipper is same-side flips i 's total flips i 's same-side prob.

Parameter vector $(\theta_1, \dots, \theta_{48}) \in (0,1)^{48}$

Model 3: Biases change over time, non-increasing

$$\Rightarrow X_{i,j} \stackrel{\text{ind.}}{\sim} \text{Bernoulli}(\theta_{i,j}) \quad \begin{array}{l} i=1, \dots, 48 \\ j=1, \dots, n_i \end{array}$$

j^{th} flip by flipper i

Constrain $\theta_{i,1} \geq \theta_{i,2} \geq \dots \geq \theta_{i,n_i}$ (n parameters!)

[Question: Why does the sample space keep changing?]

Parametric vs. Nonparametric

Parametric model dist.s indexed by parameter $\theta \in \Theta$

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

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

Use $P_\theta(\cdot)$, $\mathbb{E}_\theta(\cdot)$ to denote corresp. quantities

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

Still usually makes assumptions, e.g.

- independence
- shape constraints (e.g. P has $\downarrow$ density on $\mathbb{R}_+^d$)

Example $X_1, \dots, X_n \overset{\text{iid}}{\sim} P$ P any distr. on $\mathbb{R}$
↖ (independent & ident. distr.)

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

Boundary between parametric & non-parametric models is somewhat shaggy. Which is Model 3?

We can use "parametric notation" $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$
without loss of generality (could take $\theta = P$, $\Theta = \mathcal{P}$)

Estimation

Observe $X \sim \text{Binom}(n, \theta)$ $\theta \in (0, 1)$ unknown

Ask: What is θ ?

Skeptic's answer: Could be anything

Any $X \in \{0, \dots, n\}$ is possible under any θ

Bayesian answer: Assume θ random with known prior

$\Rightarrow$ Conditional (posterior) distribution for θ given X

Frequentist answer: Inductive behavior

Find a method for using X to estimate θ , e.g. $\delta_\theta(X) = \frac{X}{n}$

Show it generally works well for any θ

Doesn't really answer question about this θ and this $\delta(X)$

General setup Model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ (or non-par. $\mathcal{P}$)

Estimand $g(\theta)$ (or $g(P)$)

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

$\delta(\cdot)$ called estimator.

We want to evaluate & compare estimators

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

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

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

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

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

Binomial example

What is $MSE(\theta; \delta_0)$? ($\delta_0(x) = \frac{x}{n}$)

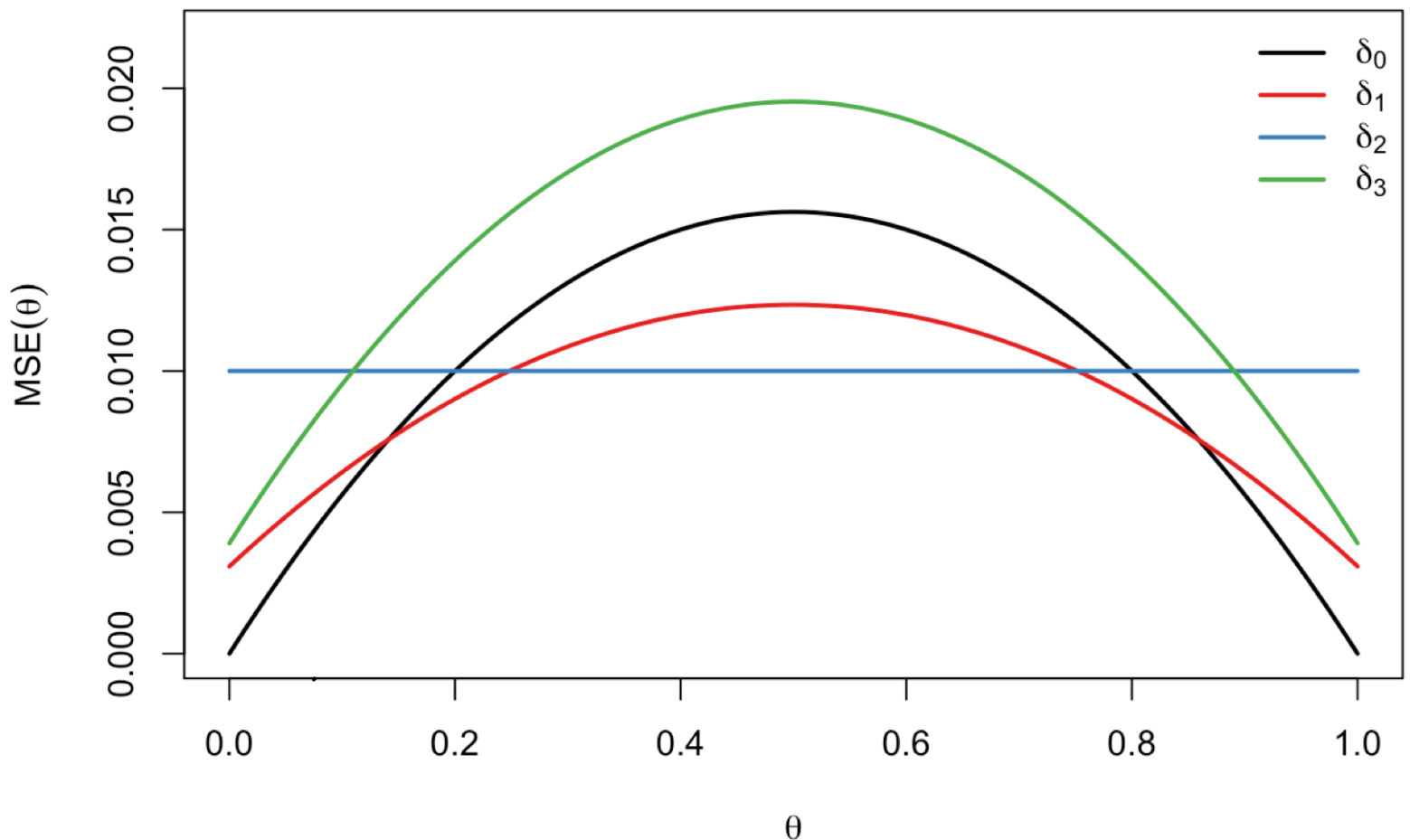
$$\mathbb{E}_\theta \left[\frac{x}{n} \right] = \theta \quad (\text{unbiased})$$

$$\begin{aligned} \Rightarrow MSE(\theta; \delta_0) &= \mathbb{E}_\theta \left[\left(\frac{x}{n} - \theta \right)^2 \right] \\ &= \text{Var}_\theta \left(\frac{x}{n} \right) \\ &= \frac{1}{n} \theta(1-\theta) \end{aligned}$$

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

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

Mean squared error for binomial estimators (n=16)



Comparing estimators

We want to choose δ to minimize R
... but this is generally not possible

An estimator δ is inadmissible if $\exists \delta^*$ with

$$a) R(\theta; \delta^*) \leq R(\theta, \delta) \quad \text{for all } \theta$$

$$b) R(\theta, \delta^*) < R(\theta, \delta) \quad \text{for some } \theta$$

We say δ^* strictly dominates δ

δ_1 is inadmissible because δ_0 dominates it

Is there any uniformly best estimator
for the binomial example?
(all θ)

Resolving ambiguity

Main strategies to resolve ambiguity:

1) Summarize risk function by a scalar:

a) Average-case risk

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

for some measure π , called prior

If π is probability measure,
same as $\mathbb{E}_{\theta \sim \pi} [R(\theta; \delta)]$

$\leadsto$ Bayes estimator

Binomial: δ_1 is Bayes wrt $\pi = \lambda$ on $[0, 1]$
 δ_2 also Bayes wrt $\pi = \text{Beta}(2, 2)$

b) Worst-case risk

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

$\leadsto$ Minimax estimator

Closely related to Bayes

Binomial: δ_2 is minimax (for $n=16$)

2) Restrict choices of estimator

a) Restrict to unbiased estimators:

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

Binomial: δ_0 is best unbiased estimator