Statistical models and decisions

Outline

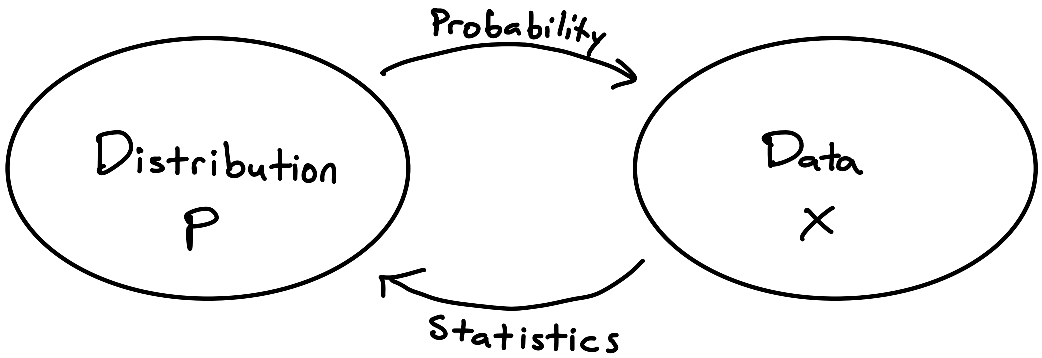
1) Statistical models

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

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 $\theta \in (0,1)$

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

known unknown (varies over model)

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

for $x = 0, 1, \ldots, n$

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

Model 2: Flippers have different biases

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

flipper i's same-side flips $\quad$ i's total flips $\quad$ i's same-side prob.

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

Model 3: Biases change over time, non-increasing

$$\Rightarrow \quad X_{i,t} \overset{\text{ind.}}{\sim} \text{Bernoulli}(\theta_{i,t}) \qquad \begin{array}{l} i = 1, \ldots, 48 \\ t = 1, \ldots, n_i \end{array}$$

t^{th} flip by flipper i

Constrain $\theta_{i,1} \geq \theta_{i,2} \geq \cdots \geq \theta_{i,n_i}$ (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

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

Still usually makes assumptions, e.g.

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

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

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

Boundary between parametric & non-parametric models is somewhat shaggy.

We can use "parametric notation" $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ wlog (could take $\theta = P$, $\Theta = \mathcal{P}$)

Estimation

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

Ask: What is θ?

Skeptic's answer: Could be anything

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

Bayesian answer: Assume θ random with known prior

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

Frequentist answer: Inductive behavior

Find a method for using X to estimate θ, e.g. $\delta_0(X) = X/n$

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

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

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

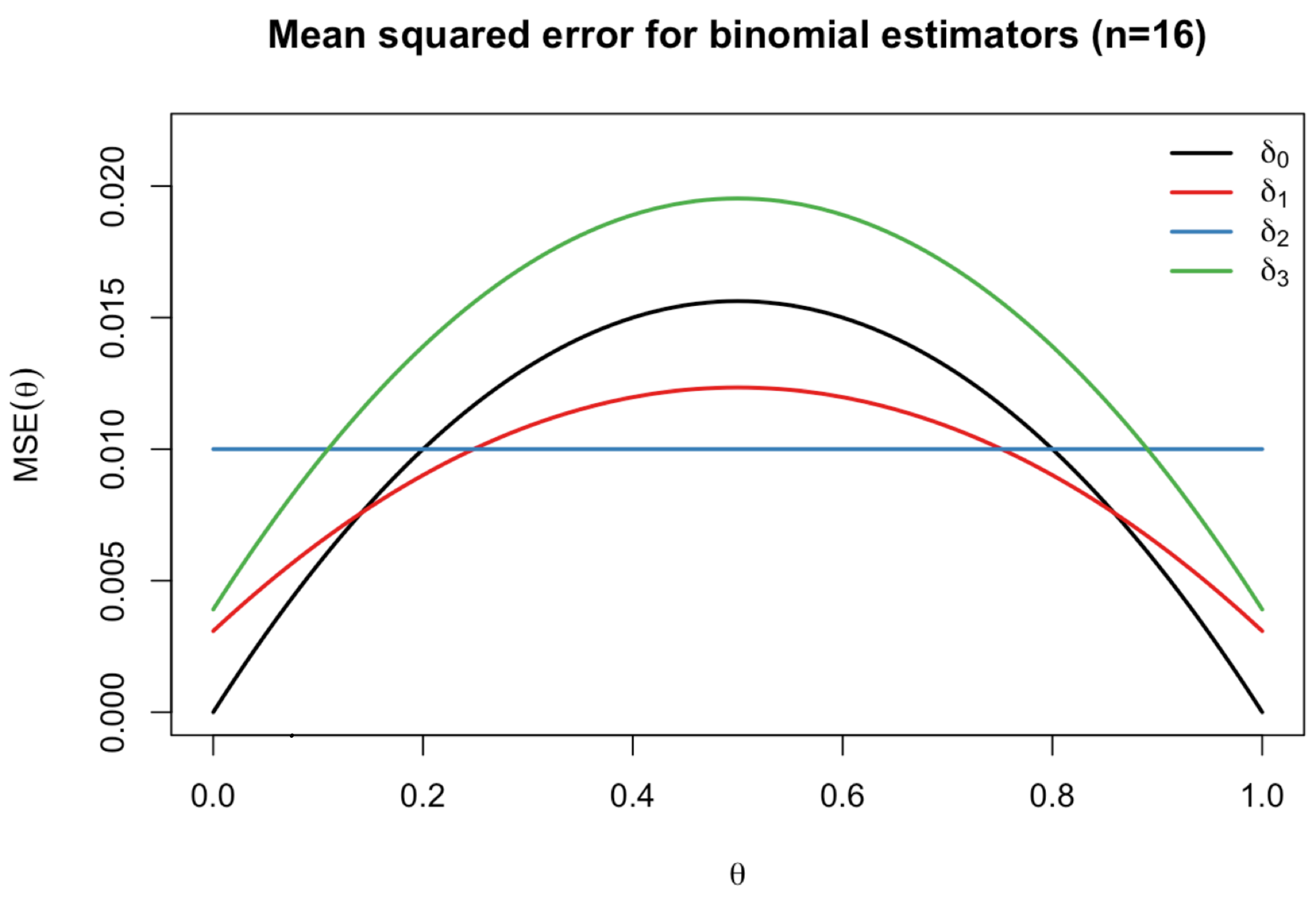
What is $MSE(\theta; \delta_0)$? $\quad \left(\delta_0(X) = \frac{X}{n} \right)$

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

$$\Rightarrow MSE(\theta; \delta_0) = \mathbb{E}_\theta\left[\left(\frac{X}{n} - \theta\right)^2\right]$$
$$= Var_\theta\left(\frac{X}{n}\right)$$
$$= \frac{1}{n}\theta(1-\theta)$$

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) $R(\theta; \delta^*) \le R(\theta, \delta)$ 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) Average-case risk

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

   for some measure π, called prior

   If π is probability measure,

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

   $\rightsquigarrow$ Bayes estimator

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

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

   b) Worst-case risk

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

   $\rightsquigarrow$ Minimax estimator

   Closely related to Bayes

   Binomial: 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