6.s077 — INTRODUCTION TO DATA SCIENCE
EECS, MIT, Spring 2018

Lecture 3

Properties and assessment of parameter estimators

Today's agenda

- Estimators (incl. review)

- Sampling distributions
  - Bias, variance, standard error, MSE
  - for the sample mean
  - for ML

- Sampling distributions, in general
  - analytical
  - parametric bootstrap
  - general bootstrap

- Confidence intervals
  - definition and interpretation
  - based on normal (and other) approximations
  - via bootstrap

Review

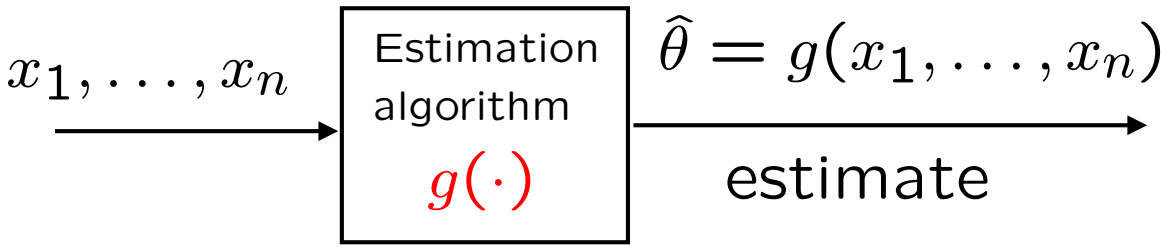
- data $x_1, \ldots, x_n$
  - observed values of i.i.d. $X_1, \ldots, X_n$, distributed according to $\mathbb{P}$ unknown
- determine empirical distribution $\widehat{\mathbb{P}}$
  - probability $1/n$ to each data point x_i

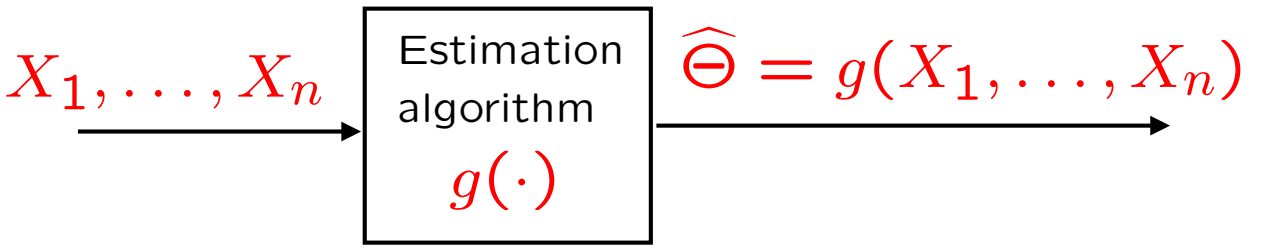
- Three types (so far) of estimators
  - **Plugin:** $a = h(\mathbb{P}) \longrightarrow \widehat{A} = h(\widehat{\mathbb{P}})$
  - **Feature matching** (and method of moments): solve $h(\mathbb{P}^\theta) = h(\widehat{\mathbb{P}})$
  - **Maximum likelihood:** maximize $\mathbb{P}^\theta(X = x)$

Data and estimators

Data: $x_1, \ldots, x_n$

realized values of r.v.s:
$X_1, \ldots, X_n$, i.i.d., $\sim \mathbb{P}^\theta$



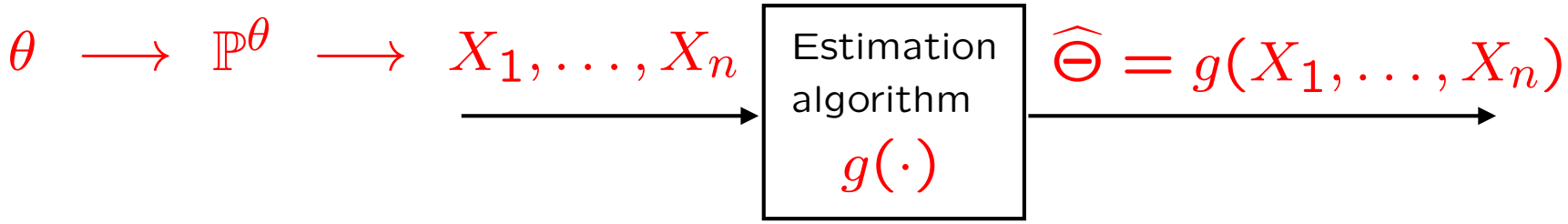


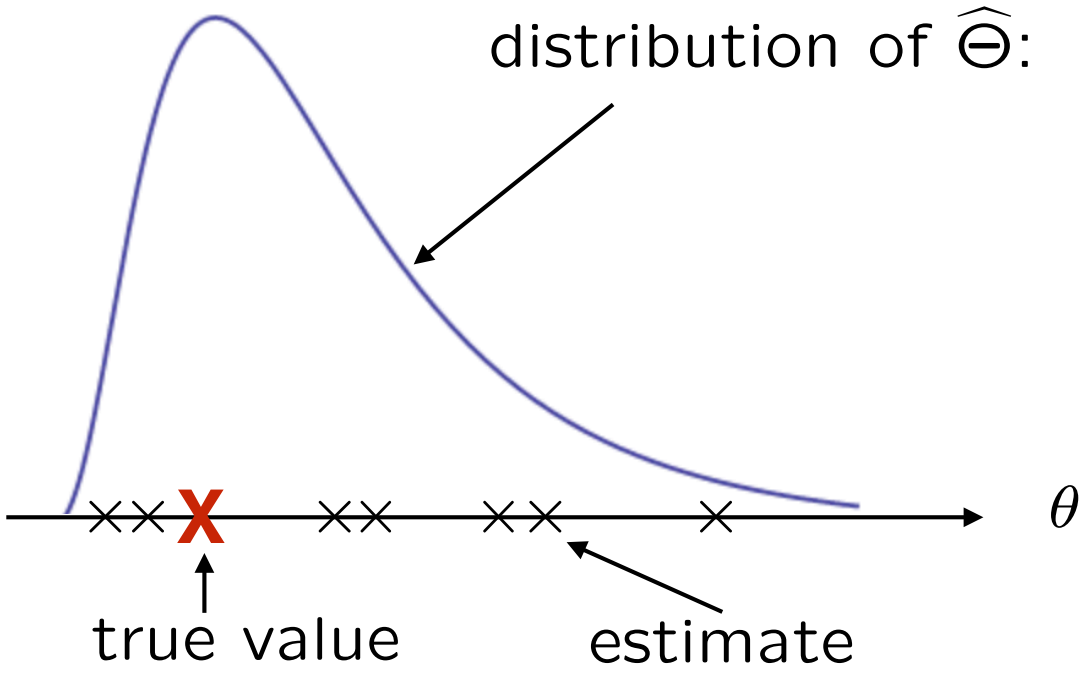
estimator: $\widehat{\Theta}$ $g(\cdot)$

- Data scientist: design "good" estimator $g(\cdot)$
  - before looking at the data
  - e.g., plugin, feature matching, ML, etc.

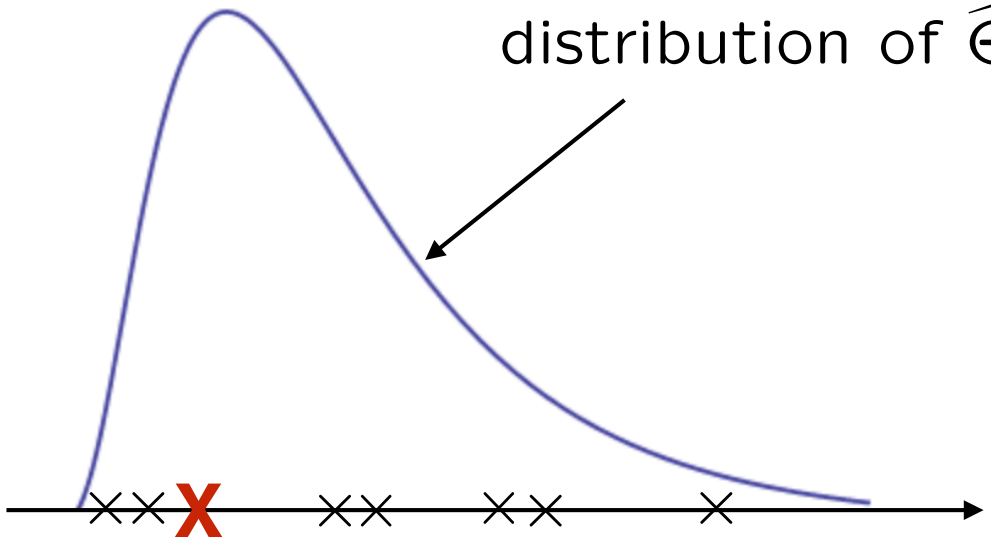
- What does "good" mean?

Sampling distribution (of an estimator)





Bias, variance, MSE, standard error



error: $\widehat{\Theta} - \theta$

bias: $b^\theta = \mathbb{E}^\theta[\widehat{\Theta} - \theta]$

variance: $v^\theta = \mathbb{E}^\theta\Big[(\widehat{\Theta} - \mathbb{E}^\theta[\widehat{\Theta}])^2\Big]$

standard error: $\text{se}^\theta = \sqrt{v^\theta}$

$\text{MSE}^\theta = \mathbb{E}^\theta\Big[(\widehat{\Theta} - \theta)^2\Big] = (b^\theta)^2 + v^\theta$

- Usually: $(b^\theta)^2$ is much smaller than v^θ

  there is a tradeoff

Sampling distribution of the sample mean

X_i: i.i.d, μ, σ^2 $\qquad$ $\mu = \mathbb{E}[X_i]$ $\qquad$ $\widehat{M} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$\mathbb{E}[\widehat{M}] = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu \qquad \text{bias} = 0 \qquad \text{(unbiased)}$$

$$v = \text{var}(\widehat{M}) = \frac{1}{n^2}\text{var}\Big(\sum_{i=1}^{n} X_i\Big) = \frac{\sigma^2}{n} \underset{n\to\infty}{\longrightarrow} 0$$

$$\text{se} = \frac{\sigma}{\sqrt{n}}$$

- CLT: $$\frac{\sqrt{n}(\widehat{M}-\mu)}{\sigma} \underset{n\to\infty}{\sim} N(0,1)$$

Loosely speaking:

$$\widehat{M} \sim N(\mu, \sigma^2/n)$$

everything extends to estimation of $\mathbb{E}\big[h(X)\big]$

Sampling distribution of the sample mean

Loosely speaking:

$\widehat{M} \sim N(\mu, \sigma^2/n)$

Loosely speaking:

$\widehat{M} \sim N(\mu, \hat{\sigma}^2/n)$

- Don't know σ? Estimate it!

  plugin estimate: $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$

  $n - 1$ $\qquad$ $\widehat{M}$

- $\widehat{M}$ not quite normal for moderate n,
  especially if you are interested
  in the tails of the sampling distribution

The wonderful properties of ML estimators

- Sampling distribution?
  In general, can't avoid (difficult) problem-specific analysis or simulation

- **But,** under mild technical conditions, as $n \to \infty$, it is:
  - **consistent:** $\widehat{\Theta} \to \theta$
  - **asymptotically normal:** $\dfrac{\widehat{\Theta} - \theta}{\text{se}} \sim N(0,1)$

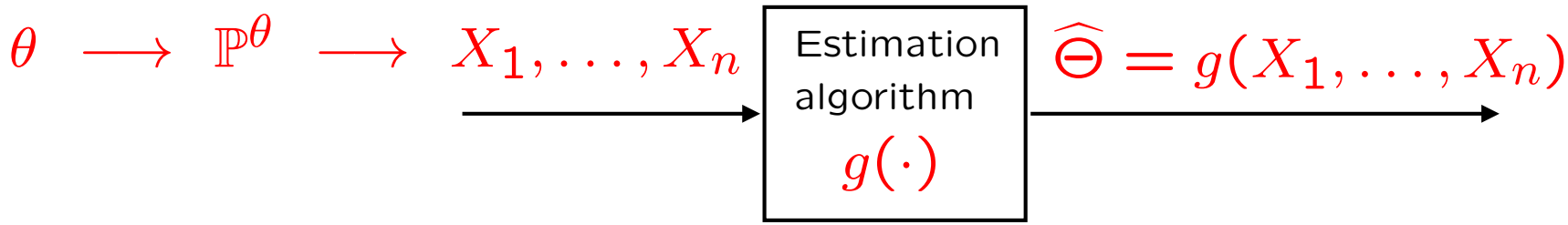
    $\widehat{\text{se}}$: estimate of se; there are good methods for obtaing it
  - **efficient:** smallest possible se (asymptotically)
    among "reasonable" estimators

Moving beyond the "friendly" cases

- For a general estimator $\widehat{\Theta} = g(X_1, \ldots, X_n)$, how do we:
  - calculate/approximate the sampling distribution?
  - calculate/approximate the bias and the standard error?
  - use the above to obtain "confidence intervals"?

- Approaches:
  - analytically
  - via the CLT
  - parametric bootstrap
  - general bootstrap

Finding the sampling distribution: analytically or via the CLT



- We have the "model" structure $\mathbb{P}^\theta$
- Pretend θ were known
- A function of r.v's $X_1, \ldots, X_n$ is an r.v.
  - knowing $\mathbb{P}^\theta$, the distribution $\mathbb{Q}^\theta$ of $\widehat{\Theta}$ is completely determined

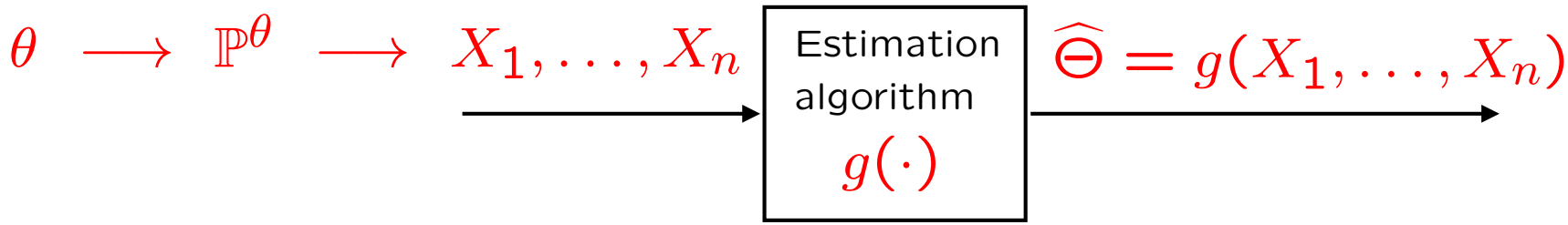
"derived distribution" exercise sometimes doable analytically

- Gives $\mathbb{Q}^\theta$
- But!!! True θ is not known

- Use $\mathbb{Q}^{\hat{\theta}}$

- **CLT-based approach:** If theory tells us that $\mathbb{Q}^\theta \sim N(\theta, \text{se}^\theta)$, approximate: $\hat{\mathbb{Q}}^\theta \approx N(\hat{\theta}, \widehat{\text{se}}^{\hat{\theta}})$

Finding the sampling distribution: Parametric bootstrap



- We have the "model" structure $\mathbb{P}^\theta$
- Pretend θ were known
- A function of r.v's $X_1, \ldots, X_n$ is an r.v.
  - knowing $\mathbb{P}^\theta$, the distribution $\mathbb{Q}^\theta$ of $\widehat{\Theta}$ is completely determined

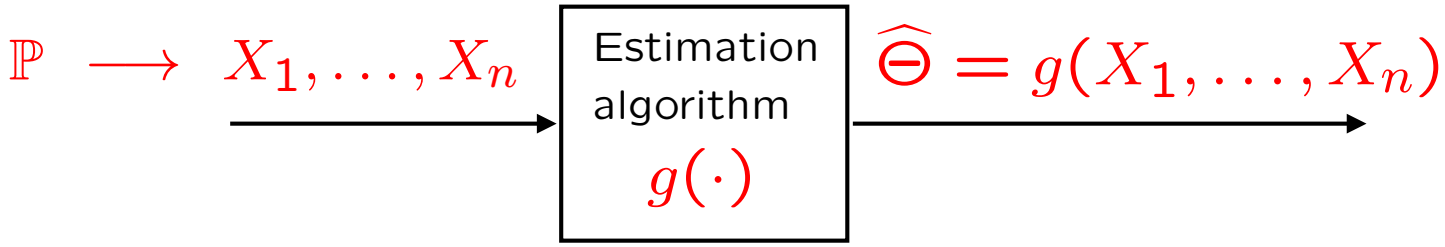
and can be obtained by simulation

- But!!! True θ is not known Rely on $\hat{\theta} \approx \theta$

Parametric bootstrap:

1. use actual data to compute $\hat{\theta}$
2. generate synthetic data $X_1, \ldots, X_n$, according to $\mathbb{P}^{\hat{\theta}}$
3. compute corresponding (new/synthetic) $\hat{\theta}$
4. repeat steps 2-3 many times to generate a histogram/empirical distribution $\hat{\mathbb{Q}}$
5. Calculate bias, standard deviation of $\hat{\mathbb{Q}}$

Finding the sampling distribution: General bootstrap



- Know little or nothing about the structure of $\mathbb{P}$
  - θ need not be a parameter that determines the distribution, but just a feature/functional; e.g., $\theta =$ median
- Want new samples from $\mathbb{P}$; but cannot simulate $\mathbb{P}$
  - but we know $\widehat{\mathbb{P}}$ (the actual data); and $\widehat{\mathbb{P}} \approx \mathbb{P}$
    use the samples that we have: draw n independent samples from $\widehat{\mathbb{P}}$

General bootstrap:

1. use actual data $x_1, \ldots, x_n$ to compute actual $\widehat{\theta}$
2. draw n samples from the set $\{x_1, \ldots, x_n\}$, uniformly, i.i.d. (with replacement)
3. use new n samples to compute corresponding (new) $\widehat{\theta}$
4. repeat steps 2-3 many times $\longrightarrow$ histogram/empirical distribution $\widehat{\mathbb{Q}}$
5. Calculate bias, standard deviation of $\widehat{\mathbb{Q}}$