MS&E 125: Intro to Applied Statistics

Inference and confidence intervals

Professor Udell

Management Science and Engineering
Stanford

April 28, 2023

Announcements

- Quizzes will be every other week (weeks 2,4,6,8,10) either in-class on W or take-home on F
- Section today will involve candy...
- Info posted to website on projects

Confusions from last time

- z-score
- θ vs $\hat{\theta}$
- estimator
- standard error
- confidence interval
- CLT
- analytic vs computational solution

Pearson vs Fisher

- what is a distribution?
- parameters
- statistics aka estimators

source: David Salzburg, “The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century”

Models and samples

a **statistical model** says how data is generated

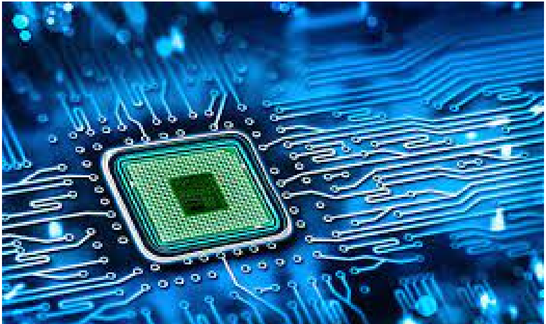
example: we model a coin flip as a Bernoulli random variable with parameter θ

we can **sample** from that model to create a dataset

example: $X_1, \ldots, X_n \sim \text{Bernoulli}(\theta)$

Application: process control

- Intel produces microprocessors and integrated circuits with varying performance using a complex process.
- Chips are binned based on performance, with higher-performing chips assigned to higher grades.
- Lower-performing chips are assigned to lower grades and sold as lower-end models.
- Binning chips allows manufacturers to maximize their process and offer customers a range of performance options.



Outline

Models and inference

Normal approximation

Confidence intervals

Inference

inference goes backwards: we use the data to make statements about the model

- also called **learning** the model or distribution

example: we can learn the parameter θ from the data

one important kind of inference is **estimation**: we use the data to estimate some parameter of the model

- e.g., a mean or variance
- **point estimate**: a single value
- **confidence interval**: a range of values likely to contain the parameter

Inference

inference goes backwards: we use the data to make statements about the model

- also called **learning** the model or distribution

example: we can learn the parameter θ from the data

one important kind of inference is **estimation**: we use the data to estimate some parameter of the model

- e.g., a mean or variance
- **point estimate**: a single value
- **confidence interval**: a range of values likely to contain the parameter

Q: how to estimate θ from $X_1, \ldots, X_n$?

Inference

inference goes backwards: we use the data to make statements about the model

- also called **learning** the model or distribution

example: we can learn the parameter θ from the data

one important kind of inference is **estimation**: we use the data to estimate some parameter of the model

- e.g., a mean or variance
- **point estimate**: a single value
- **confidence interval**: a range of values likely to contain the parameter

Q: how to estimate θ from $X_1, \ldots, X_n$?
A: $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i$

Bias of an estimator

Definition

the **bias** of estimator $\hat{\theta}$ is $\mathbb{E}[\hat{\theta}] - \theta$

the estimator is **unbiased** if $\mathbb{E}[\hat{\theta}] = \theta$

Bias of an estimator

Definition

the **bias** of estimator $\hat{\theta}$ is $\mathbb{E}[\hat{\theta}] - \theta$

the estimator is **unbiased** if $\mathbb{E}[\hat{\theta}] = \theta$

Q: What is the bias of $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$?

Bias of an estimator

Definition

the **bias** of estimator $\hat{\theta}$ is $\mathbb{E}[\hat{\theta}] - \theta$

the estimator is **unbiased** if $\mathbb{E}[\hat{\theta}] = \theta$

Q: What is the bias of $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$?
A:

$$\begin{aligned} \hat{\theta} - \theta &= \frac{1}{n}\sum_{i=1}^{n} X_i - \theta = \frac{1}{n}\sum_{i=1}^{n}(X_i - \theta) \\ \mathbb{E}[\hat{\theta}] - \theta &= \frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[X_i] - \theta) = 0 \end{aligned}$$

Poll: what is another example of an unbiased estimator for θ?

Consistency of an estimator

estimator $\hat{\theta}$ is **consistent** if $\hat{\theta} \to \theta$ as $n \to \infty$

Consistency of an estimator

estimator $\hat{\theta}$ is **consistent** if $\hat{\theta} \to \theta$ as $n \to \infty$

Q: Is the estimator $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i$ consistent?

Consistency of an estimator

estimator $\hat{\theta}$ is **consistent** if $\hat{\theta} \to \theta$ as $n \to \infty$

Q: Is the estimator $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i$ consistent?
A: The sample mean is consistent because $\frac{1}{n} n\theta \to \theta$.

Consistency of an estimator

estimator $\hat{\theta}$ is **consistent** if $\hat{\theta} \to \theta$ as $n \to \infty$

Q: Is the estimator $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i$ consistent?
A: The sample mean is consistent because $\frac{1}{n} n\theta \to \theta$.

Poll: which of these estimators is consistent?

Standard error

the standard deviation of $\hat{\theta}$ is called the **standard error**

$$\mathsf{se}(\hat{\theta}) = \sqrt{\mathbf{Var}[\hat{\theta}]}$$

we call the estimated standard error $\widehat{\mathsf{se}}$

Standard error

the standard deviation of $\hat{\theta}$ is called the **standard error**

$$\mathsf{se}(\hat{\theta}) = \sqrt{\mathbf{Var}[\hat{\theta}]}$$

we call the estimated standard error $\hat{\mathsf{se}}$

example:

- if $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$, then $se = \frac{\mathbf{Var}\, X}{n}$

Standard error

the standard deviation of $\hat{\theta}$ is called the **standard error**

$$\text{se}(\hat{\theta}) = \sqrt{\mathbf{Var}[\hat{\theta}]}$$

we call the estimated standard error $\hat{\text{se}}$

example:

- if $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$, then $se = \frac{\mathbf{Var}\, X}{n}$
  why?

$$\mathbf{Var}[\hat{\theta}] = \mathbf{Var}[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{Var}[X_i] = \frac{1}{n^2} n \,\mathbf{Var}\, X = \frac{\mathbf{Var}\, X}{n}$$

Standard error

the standard deviation of $\hat{\theta}$ is called the **standard error**

$$\text{se}(\hat{\theta}) = \sqrt{\textbf{Var}[\hat{\theta}]}$$

we call the estimated standard error $\hat{\text{se}}$

example:

- if $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$, then $se = \frac{\textbf{Var}\,X}{n}$
  why?

$$\textbf{Var}[\hat{\theta}] = \textbf{Var}[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\textbf{Var}[X_i] = \frac{1}{n^2}n\,\textbf{Var}\,X = \frac{\textbf{Var}\,X}{n}$$

- for our coin flip model, $\hat{\text{se}} = \sqrt{\frac{\theta(1-\theta)}{n}}$
  why? $\textbf{Var}\,X = \theta(1-\theta)$, so $\textbf{Var}[\hat{\theta}] = \frac{\theta(1-\theta)}{n}$

Demo

https://colab.research.google.com/github/stanford-mse-125/demos/blob/main/inference.ipynb

Outline

Models and inference

Normal approximation

Confidence intervals

Central limit theorem

the **central limit theorem** says that the distribution of $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is approximately normal with mean θ and variance $\mathbf{Var}\, X/n = \mathrm{se}(\theta)^2$

$$\frac{\hat{\theta} - \theta}{\mathrm{se}} \to \mathcal{N}(0, 1)$$

- the distribution of $\hat{\theta}$ is approximately normal with mean θ and standard deviation se
- also true if the standard error se is replaced by the estimated standard error $\widehat{\mathrm{se}}$

assumptions:

- $\mathbb{E}X = \theta$
- $\mathbf{Var}\, X$ is finite

Central limit theorem

the **central limit theorem** says that the distribution of $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is approximately normal with mean θ and variance $\mathbf{Var}\, X/n = \text{se}(\theta)^2$

$$\frac{\hat{\theta} - \theta}{\text{se}} \to \mathcal{N}(0, 1)$$

- the distribution of $\hat{\theta}$ is approximately normal with mean θ and standard deviation se
- also true if the standard error se is replaced by the estimated standard error $\widehat{\text{se}}$

assumptions:

- $\mathbb{E}X = \theta$
- $\mathbf{Var}\, X$ is finite

example: for our coin flip model, $\hat{\theta} \sim \mathcal{N}(\theta, \frac{\theta(1-\theta)}{n})$

Why use a normal approximation?

- normal distribution has just two parameters
- can estimate those parameters from data
- we can use those parameters to reason about tails of distribution

define the **z-score**: the number of standard deviations away from the mean

$$z = \frac{\hat{\theta} - \theta}{\mathrm{se}}$$

Demo

https://colab.research.google.com/github/
stanford-mse-125/demos/blob/main/inference.ipynb

Outline

Models and inference

Normal approximation

Confidence intervals

Confidence interval

a **confidence interval** is an interval C likely to contain the parameter e.g. the $(1-\alpha)$ confidence interval satisfies

$$\mathbb{P}[\theta \in C] \geq 1-\alpha$$

- C is a random variable: it depends on the data $X_1, \ldots, X_n$
- θ is fixed

Confidence interval

a **confidence interval** is an interval C likely to contain the parameter e.g. the $(1 - \alpha)$ confidence interval satisfies

$$\mathbb{P}[\theta \in C] \geq 1 - \alpha$$

- C is a random variable: it depends on the data $X_1, \ldots, X_n$
- θ is fixed

two interpretations (e.g., for 95% confidence interval C):

- if we repeat the experiment, we expect C to contain θ $100(1 - \alpha)\%$ of the time
- if we do a bunch of different experiments, we expect the 95% confidence interval to contain the true value of θ for 95% of the experiments

Confidence intervals: examples

opinion polls:

- 49% $\pm$ 3% think U.S. should lift Cuba embargo.
- 38% $\pm$ 3% think U.S. should build more nuclear power plants.
- 16% $\pm$ 4% think St. Louis Cardinals will win the World Series.

demographic surveys:

- The average height of adult males in the United States is between 5 feet 7 inches and 5 feet 10 inches
- The average salary of software engineers in San Francisco is between $120,000 and $140,000

Confidence intervals: examples

medical research:

- average weight loss of participants in a weight loss program is between 10 and 15 pounds

operations management:

- the mean response time of the website is between 2 and 3 seconds
- the mean time to check out at the grocery store is between 2 and 3 minutes

How to construct confidence interval?

- (today) use a normal approximation with analytic formula for standard error
- (later) use a normal approximation with bootstrap estimate for standard error
- (later) use bootstrap quantiles

Normal approximation for confidence interval

Suppose $\hat{\theta} \approx N(\theta, \mathsf{se}^2)$. Then

$$C = \left[\hat{\theta} - z_{\alpha/2}\hat{\mathsf{se}}, \hat{\theta} - z_{\alpha/2}\hat{\mathsf{se}}\right]$$

is an approximate $(1-\alpha)$ confidence interval for θ, where $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution.

Confidence interval for coin flip

example: for our coin flip model, we can construct a $100(1-\alpha)\%$ confidence interval for θ as

$$\hat{\theta} \pm z_{\alpha/2}\hat{\mathsf{se}}$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution

e.g., for $\alpha = 0.05$, we use $z_{0.025} = 1.96$

Calibration

A $(1-\alpha)$ confidence interval is called **calibrated** if

$$\mathbb{P}[\theta \in C] \approx 1-\alpha$$

- if confidence interval is too large, it's useless
- if confidence interval is too small, it's wrong

Proof that normal confidence interval is calibrated

Proof:

$$\begin{aligned}
\Pr(\theta \in C_n) &= \Pr(\hat{\theta}_n - z_{\alpha/2}\hat{\text{se}} \le \theta \le \hat{\theta}_n + z_{\alpha/2}\hat{\text{se}}) \\
&= \Pr(-z_{\alpha/2}\hat{\text{se}} \le \theta - \hat{\theta}_n \le z_{\alpha/2}\hat{\text{se}}) \\
&= \Pr\left(-z_{\alpha/2} \le \frac{\theta - \hat{\theta}_n}{\hat{\text{se}}} \le z_{\alpha/2}\right) \\
&\approx \Pr\left(-z_{\alpha/2} \le Z \le z_{\alpha/2}\right) \\
&= 1 - \alpha
\end{aligned}$$

- $\hat{\text{se}}$ approximates the standard deviation of $\hat{\theta}$
- the central limit theorem says that $\hat{\theta}$ is approximately normal, so the standard deviation controls the tails of the distribution

$\Longrightarrow$ CI is calibrated if number of samples n is large enough to justify approximations