Probability Distributions
Week 11

Yunkyu Sohn
School of Political Science and Economics
Waseda University

2019 Spring Statistics II

Contents

Random Variables and Probability Distribution

- Random variable assigns a numeric value to each event of the experiment.

  - Coin flip side: head = 1; tail =0

  - #secs took for commuting: any value greater than 0.

- These values represent mutually exclusive and exhaustive events, together forming the entire sample space Ω.

- A discrete RV takes a finite or at most countably infinite #distinct values.

  - coin flip; race; number of years of education

- A continuous RV assumes an uncountably infinite number of values.

  - height, distance from earth, gross domestic product

- Probability distribution: Probability that a random variable takes a certain value or range of values.

  - P(side): P(side=1) = 0.5; P(side=0) = 0.5

  - P(#secs): P(0<#secs<1000) = 0.3; P(1000<#secs<2000) = 0.4 ….

Probability Density / Mass Function

▶ Probability mass function (PMF): f(x) for a discrete random variable

▶ Probability density function (PDF): f(x) for a continuous random variable

▶ Recall $P(\Omega) = 1$: total sum of f(x) (PMF), or the area of f(x) (PDF), must equal to 1.

▶ Cumulative mass function (CMF):

$$F(x) = P(X \leq x) = \sum_{K \leq x} f(k)$$

▶ Cumulative density function (CDF):

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

▶ What is the probability that a random variable X takes a value equal to or less than x?

▶ Area under the density curve

▶ Non-decreasing

Bernoulli Distribution

- Ω={0,1}

- PMF
$$f(x) = \begin{cases} p & \textbf{if } x = 1 \\ 1 - p & \textbf{if } x = 0 \\ 0 & \textbf{otherwise.} \end{cases}$$

- CMF
$$F(x) = \begin{cases} 0 & \textbf{if } x < 0 \\ 1 - p & \textbf{if } 0 \leq x < 1 \\ 1 & \textbf{if } x \geq 1 \end{cases}$$
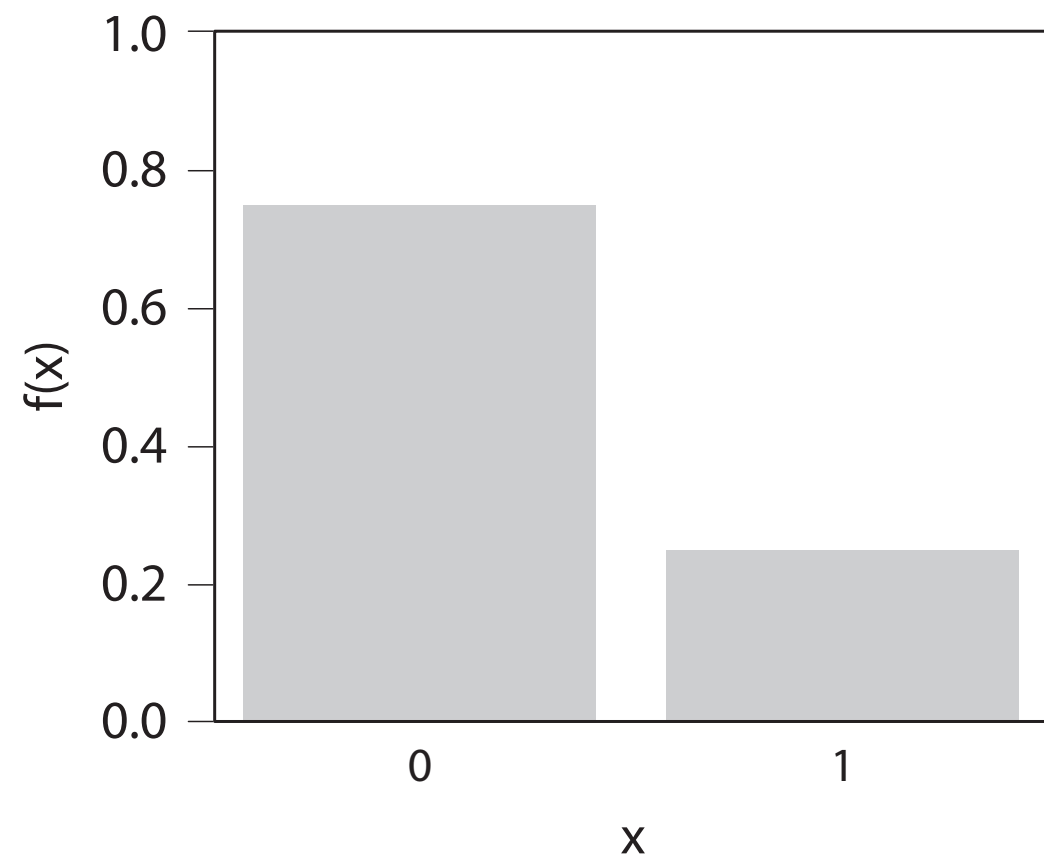
Bernoulli Distribution

▶ $\Omega=\{0,1\}$

▶ PMF
$$f(x) = \begin{cases} p & \textbf{if } x = 1 \\ 1 - p & \textbf{if } x = 0 \\ 0 & \textbf{otherwise.} \end{cases}$$
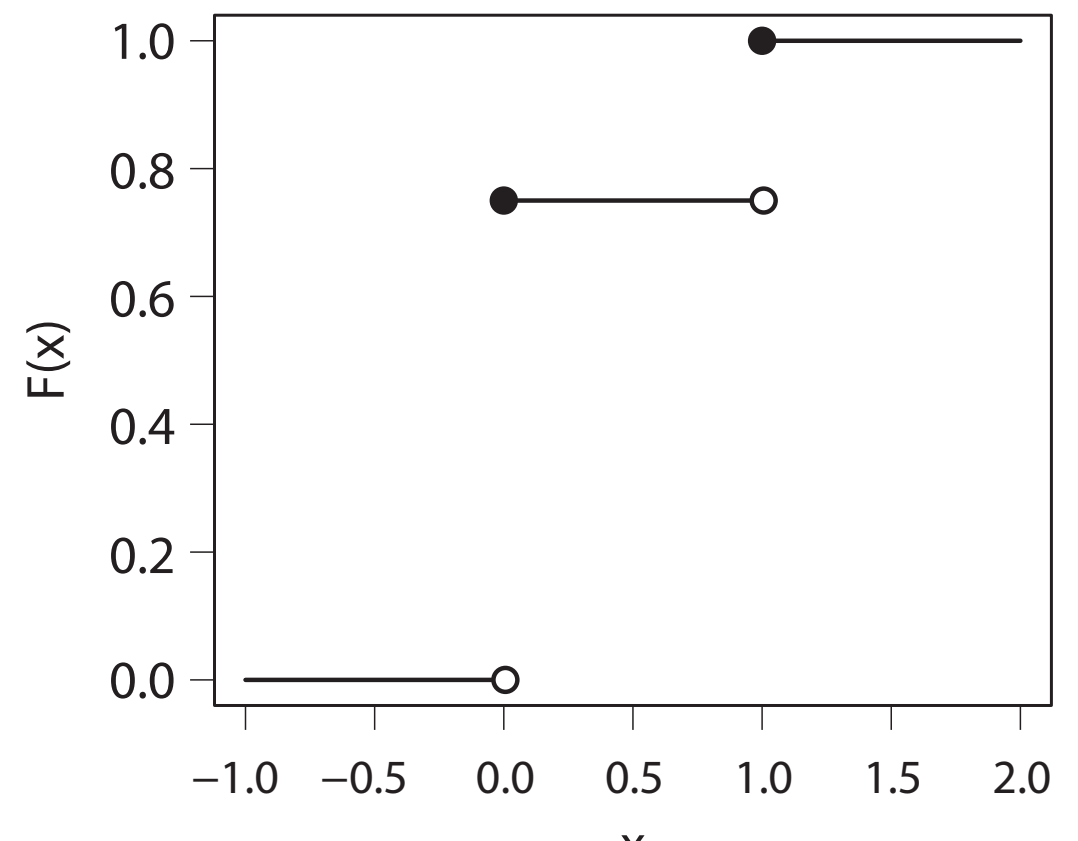
▶ PMF and CDF of Bernoulli distribution for p = 0.25

Probability mass function

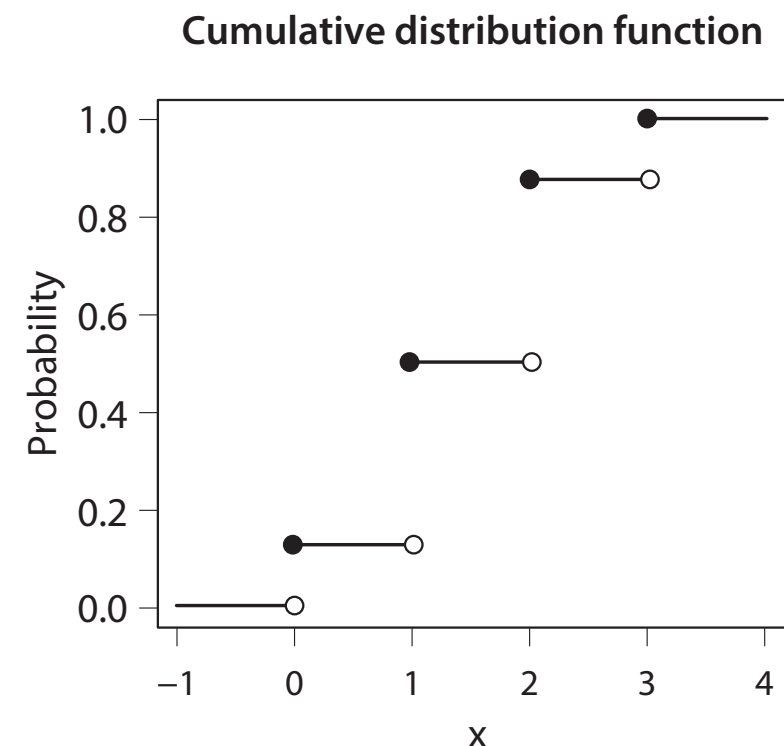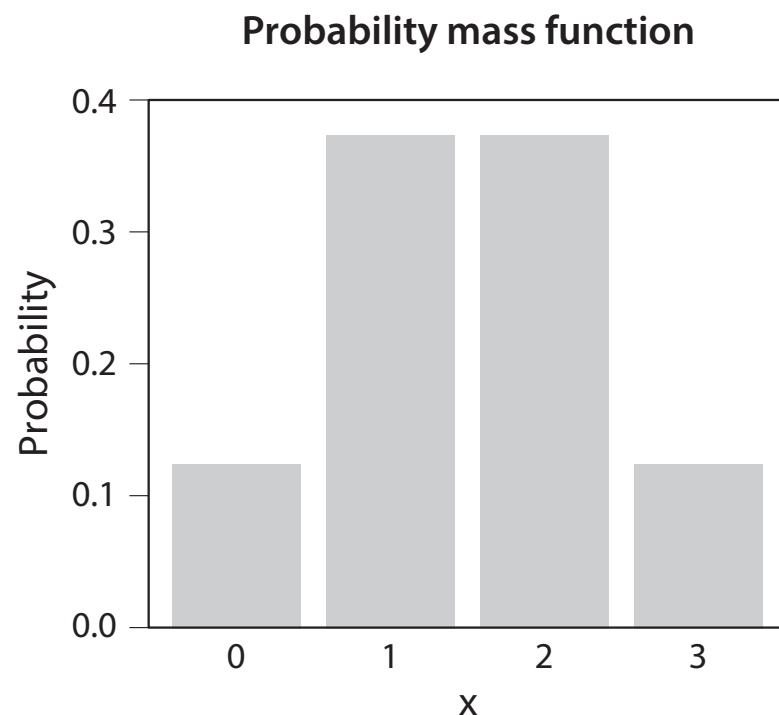**Cumulative distribution function**

Binomial Distribution

▶ The number of 1s (one of the binary outcomes) in multiple Bernoulli trials

▶ $\Omega=\{0,1,\ldots,n\text{-}1,n\}$

▶ PMF

$$f(x) = P(X = x) = \binom{n}{x} p^x(1 - p)^{n-x} \qquad \binom{n}{x} =_n C_x$$

▶ CMF

$$F(x) = P(X \le x) = \sum_{k=0}^{x} \binom{n}{k} p^k(1 - p)^{n-k}$$

▶ $p = 0.5$ and $n = 3$

Probability mass function

Cumulative distribution function

Binomial Distribution

▶ ex1) In a small department: There are exactly 10 who support candidate A, another 10 people who support candidate B for electing the chair of the department. Suppose that we expect their individual turnout probability is equal to their previous overall turnout rate which was 70%. What is the chance that exactly 7 people vote for candidate A and 7 people vote for candidate B, and the election ends in a tie?

▶ ex2) In a small department: There are exactly 10 who support candidate A, another 10 people who support candidate B for electing the chair of the department. Suppose that we expect their individual turnout probability is equal to their previous overall turnout rate which was 70%. What is the chance that the election ends in a tie?
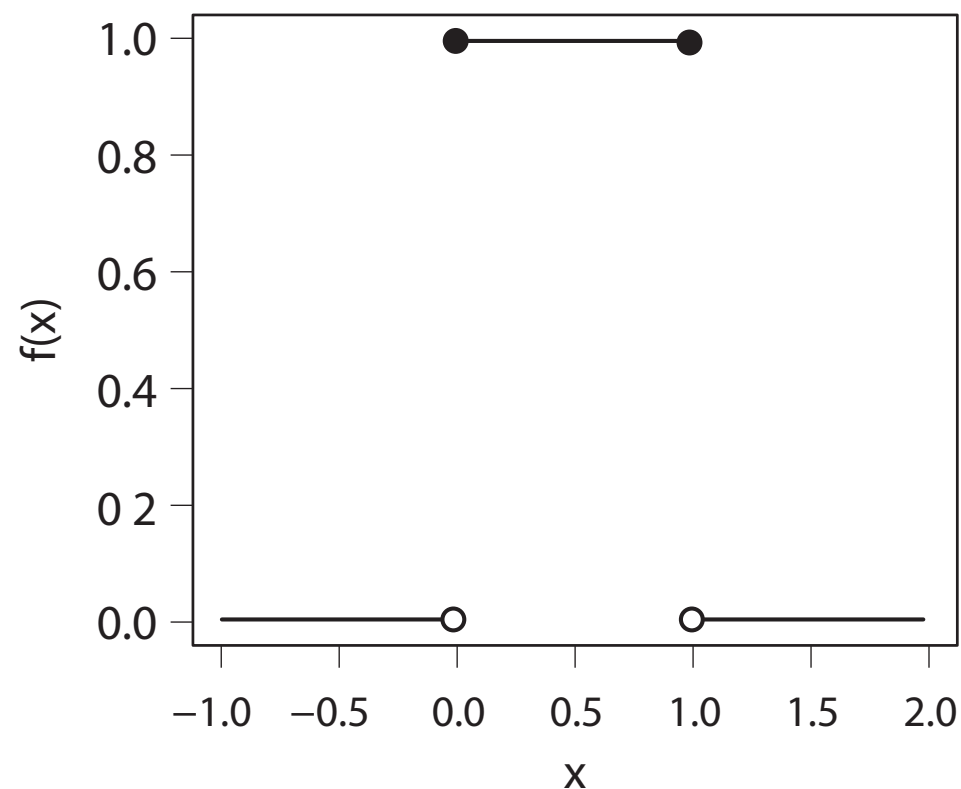
Uniform Distribution

▶ Every number in an interval has an equal chance of appearance
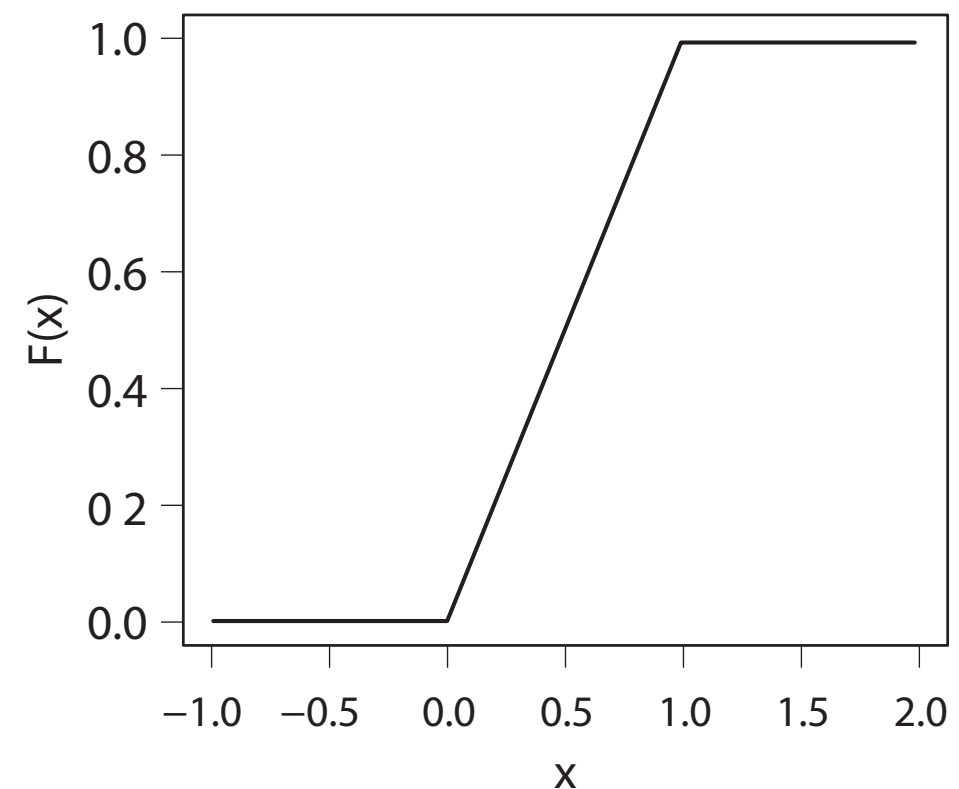
▶ $\Omega =$ **set of real numbers in a range** $[a, b]$

▶ PDF
$$f(x) = \begin{cases} \frac{1}{b-a} & \textbf{if } a \leq x \leq b \\ 0 & \textbf{otherwise.} \end{cases}$$

▶ CDF
$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & \textbf{if } a \leq x \leq b \\ 1 & x \geq b \end{cases}$$

▶ Uniform distribution for the interval $[0,1]$

Probability density function

Cumulative distribution function

Uniform Distribution

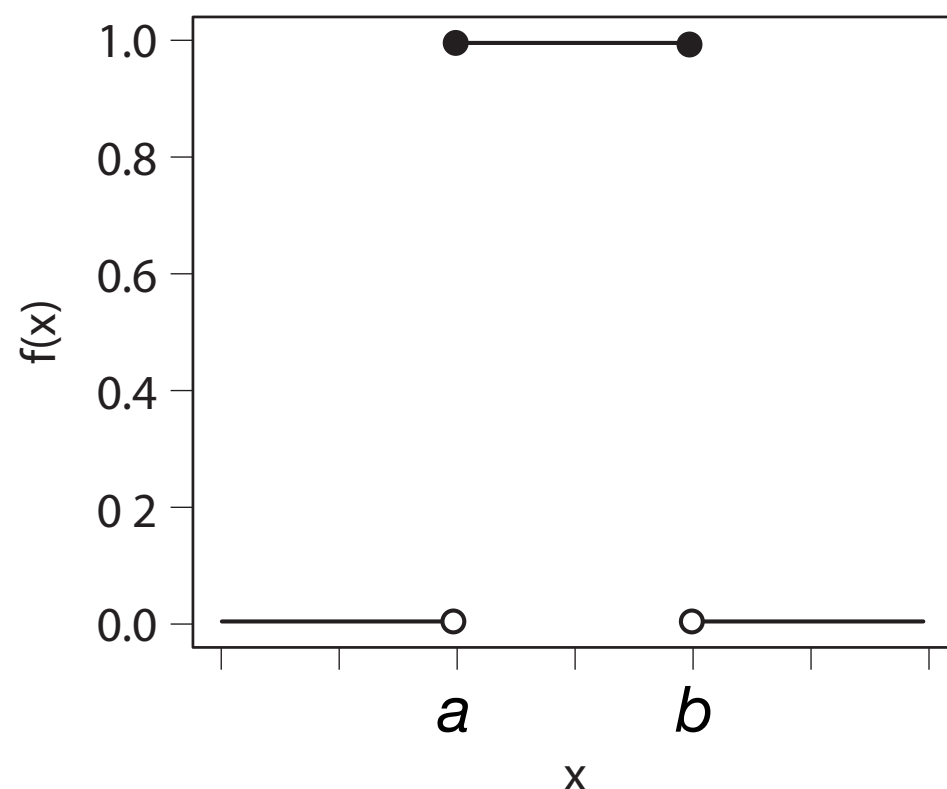▶ Every number in an interval has an equal chance of appearance

▶ $\Omega =$ **set of real numbers**

▶ PDF
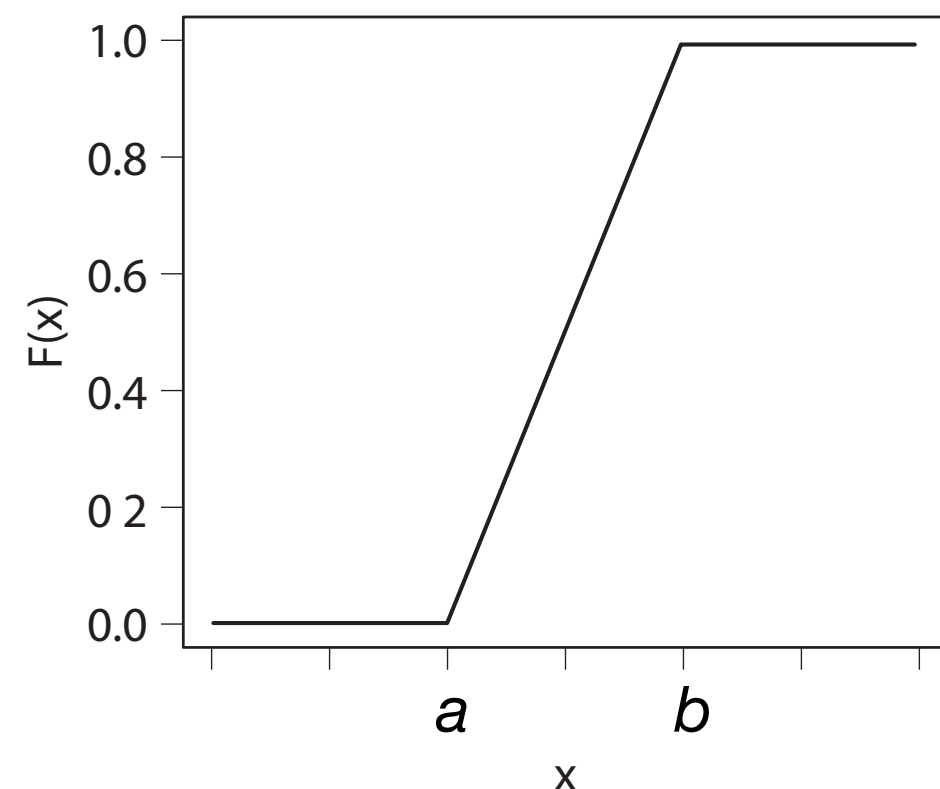$$f(x) = \begin{cases} \dfrac{1}{b-a} & \textbf{if } a \le x \le b \\ 0 & \textbf{otherwise.} \end{cases}$$

▶ CDF
$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & \textbf{if } a \le x \le b \\ 1 & x \ge b \end{cases}$$

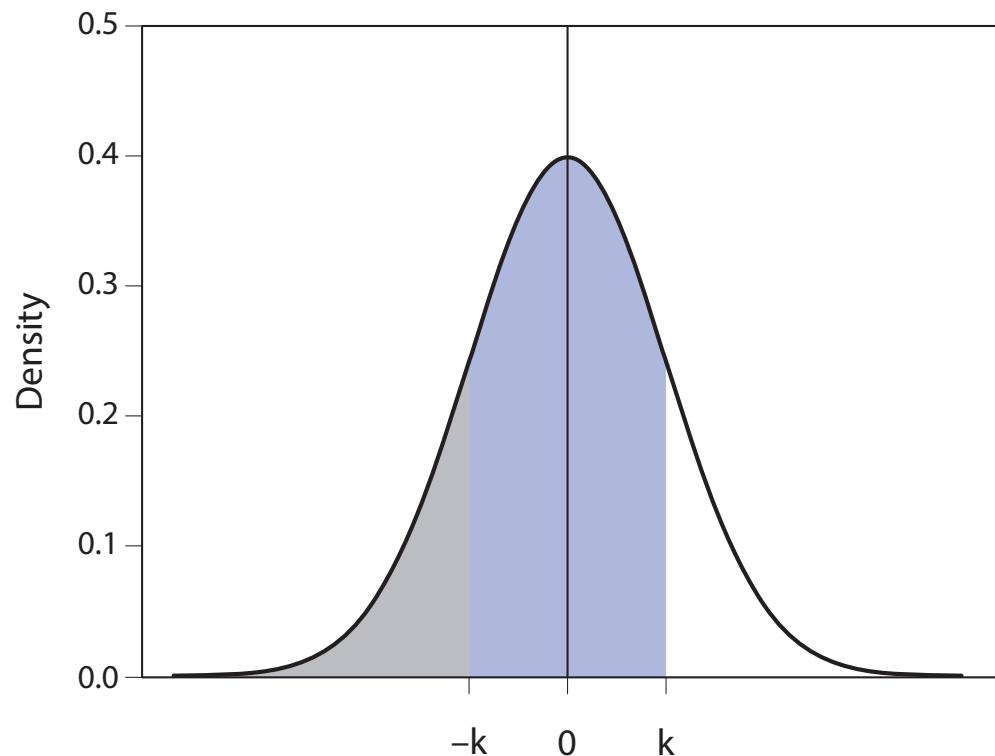▶ Uniform distribution for the interval $[a,b]$

Probability density function

Cumulative distribution function

Normal Distribution

▶ Most famous and frequently observed distribution (Why?)

▶ Ω=real numbers (continuous number)

▶ X is normal RV with mean **μ** and standard deviation **σ**: $X \sim \mathcal{N}(\mu, \sigma^2)$

▶ PDF
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\}$$

▶ e.g. $X \sim \mathcal{N}(0,1)$

▶ Singled peaked, symmetric

▶ about 2/3 are within 1 standard deviation (**σ**) from the mean

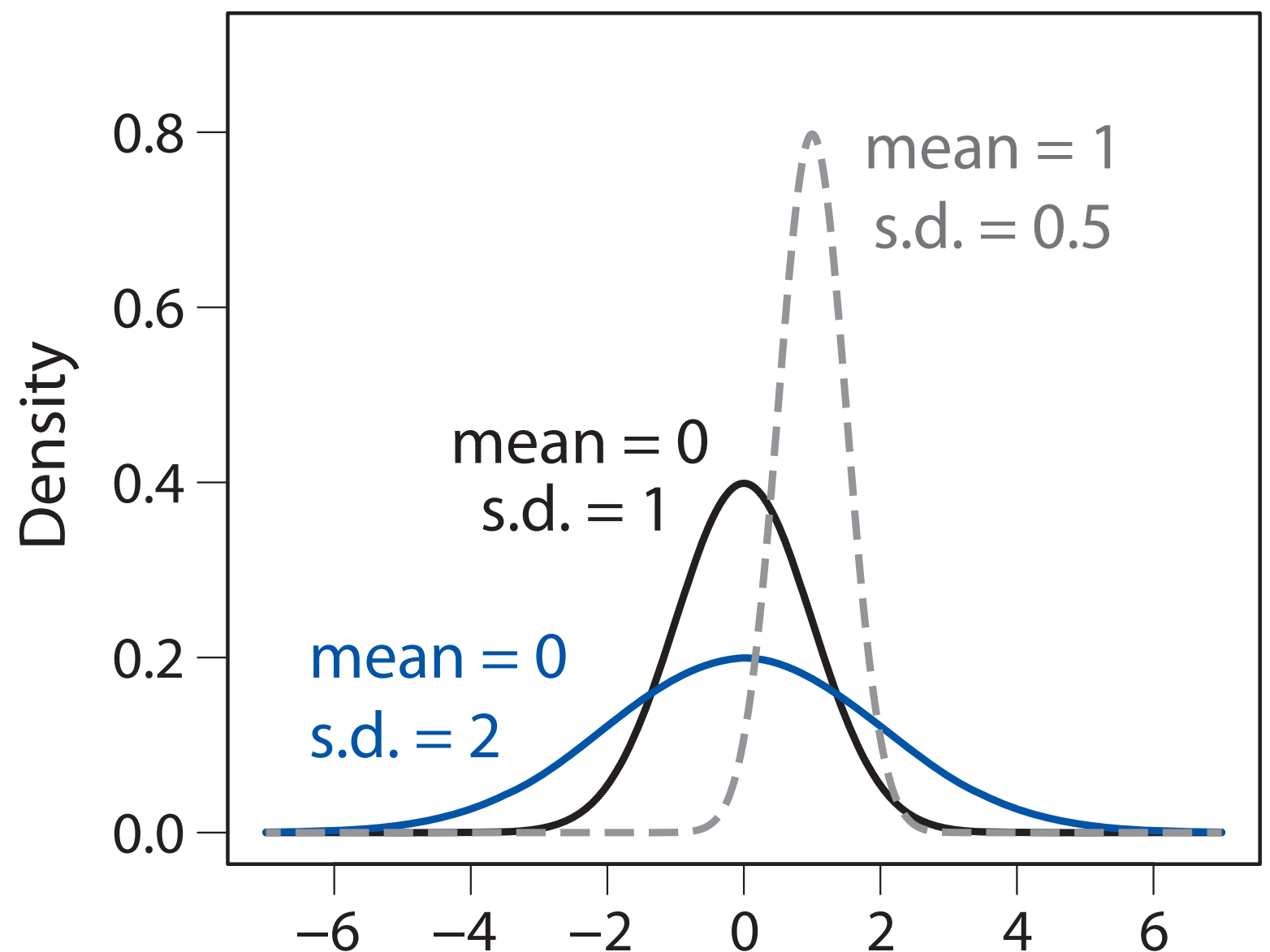▶ about 95% are within 2 standard deviations (2**σ**) from the mean

▶ CDF $\quad F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2}(t-\mu)^2 \right\} dt$

Normal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

▶ $X \sim \mathcal{N}(\mu, \sigma^2)$

▶ $S = X + c \rightarrow S \sim \mathcal{N}(\mu + c, \sigma^2)$

▶ $Y = cX \rightarrow Y \sim \mathcal{N}(c\mu, (c\sigma)^2)$

▶ $T = aX + b \rightarrow T \sim \mathcal{N}(a\mu + b, (a\sigma)^2)$
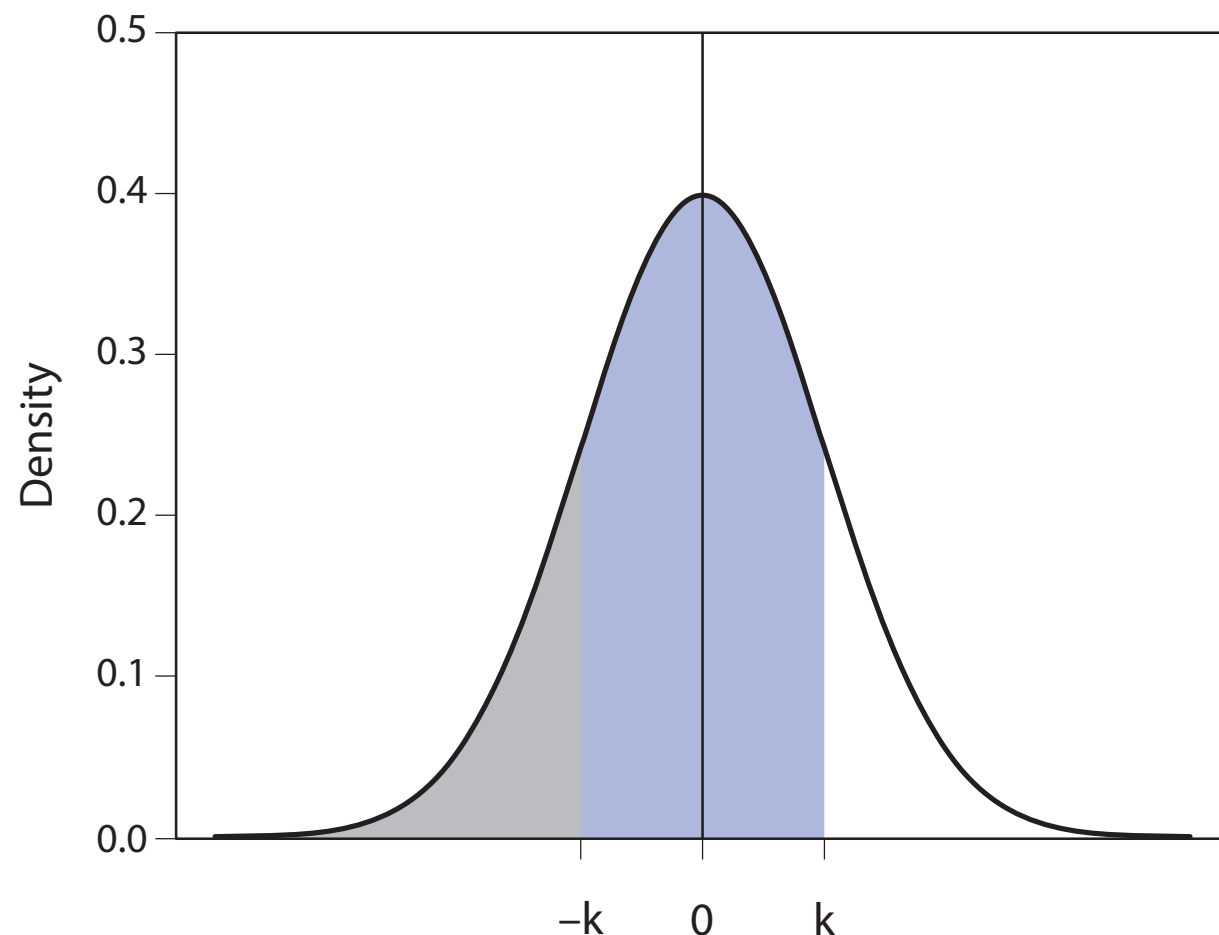
Probability density function

Normal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

▶ $X \sim \mathcal{N}(\mu, \sigma^2)$

▶ $S = X + c \rightarrow S \sim \mathcal{N}(\mu + c, \sigma^2)$

▶ $Y = cX \rightarrow Y \sim \mathcal{N}(c\mu, (c\sigma)^2)$

▶ $T = aX + b \rightarrow T \sim \mathcal{N}(a\mu + b, (a\sigma)^2)$

▶ **z-score:** $Z = (X - \mu)/\sigma \rightarrow Z \sim \mathcal{N}(0,1)$

Normal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}$$

▶ Probability that a normal random variable with mean **μ** and sdv **σ** lies within *k* standard deviations from the mean for a positive constant $k > 0$

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) = P(-k\sigma \leq X - \mu \leq k\sigma)$$

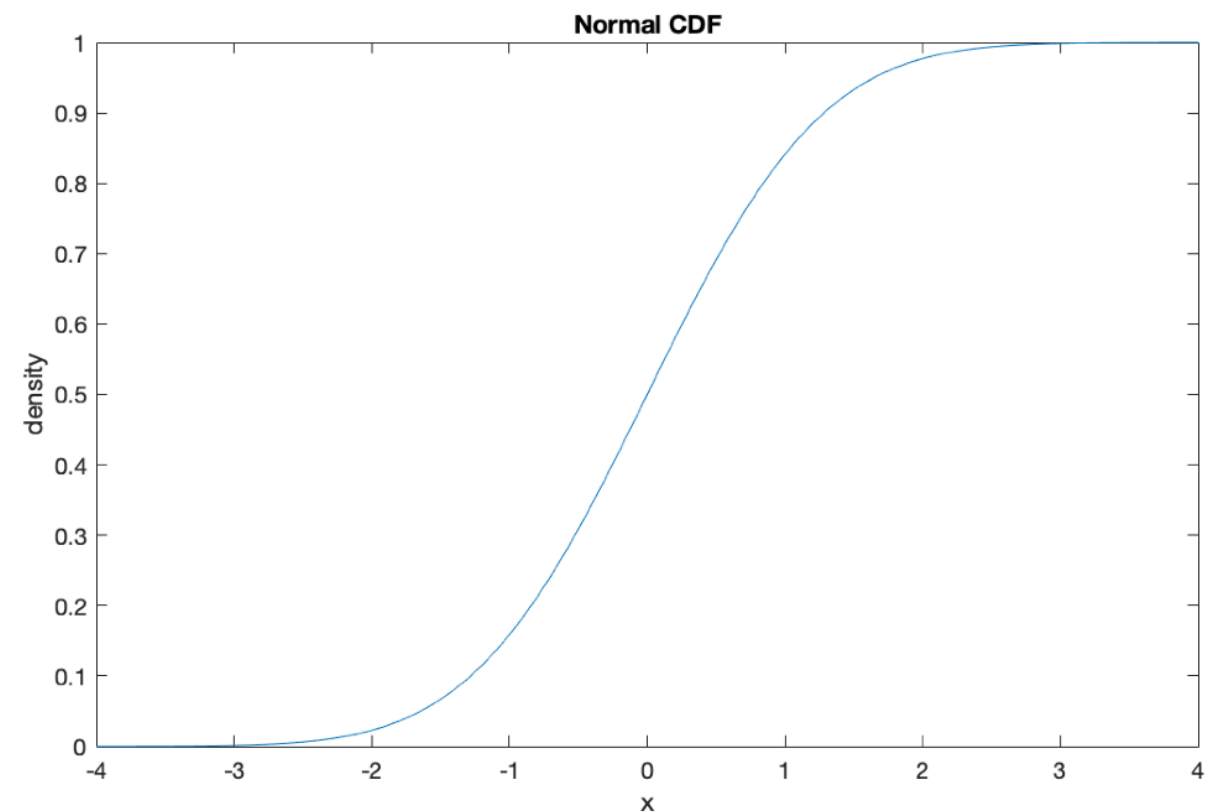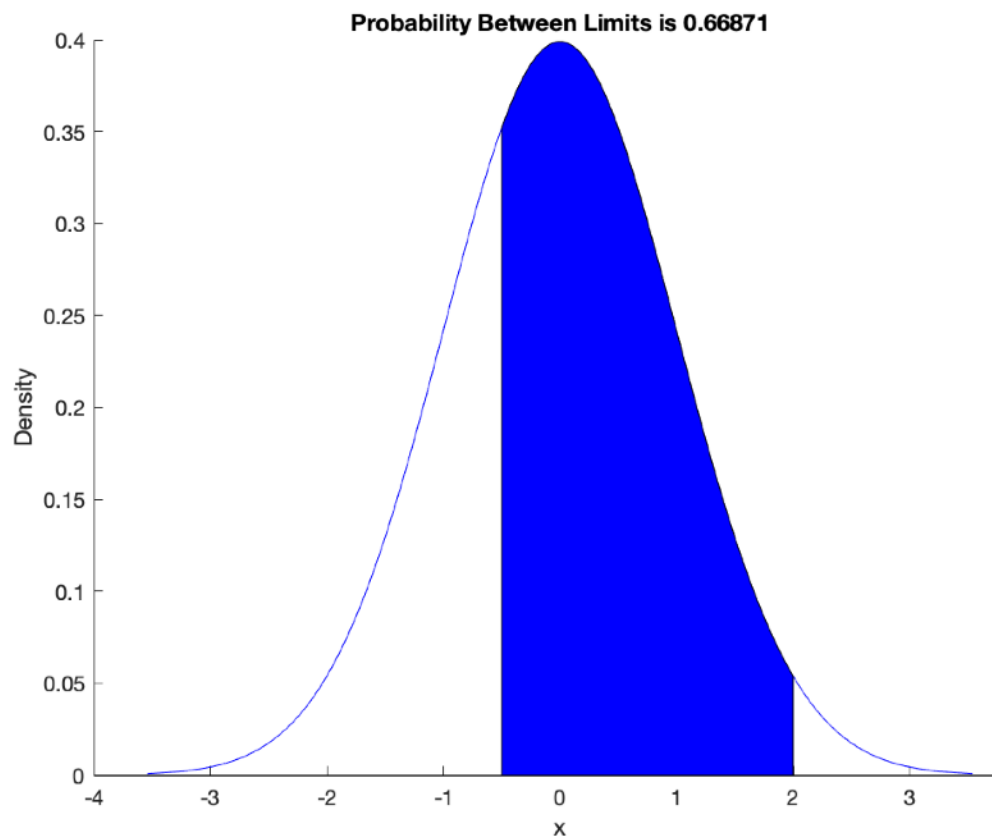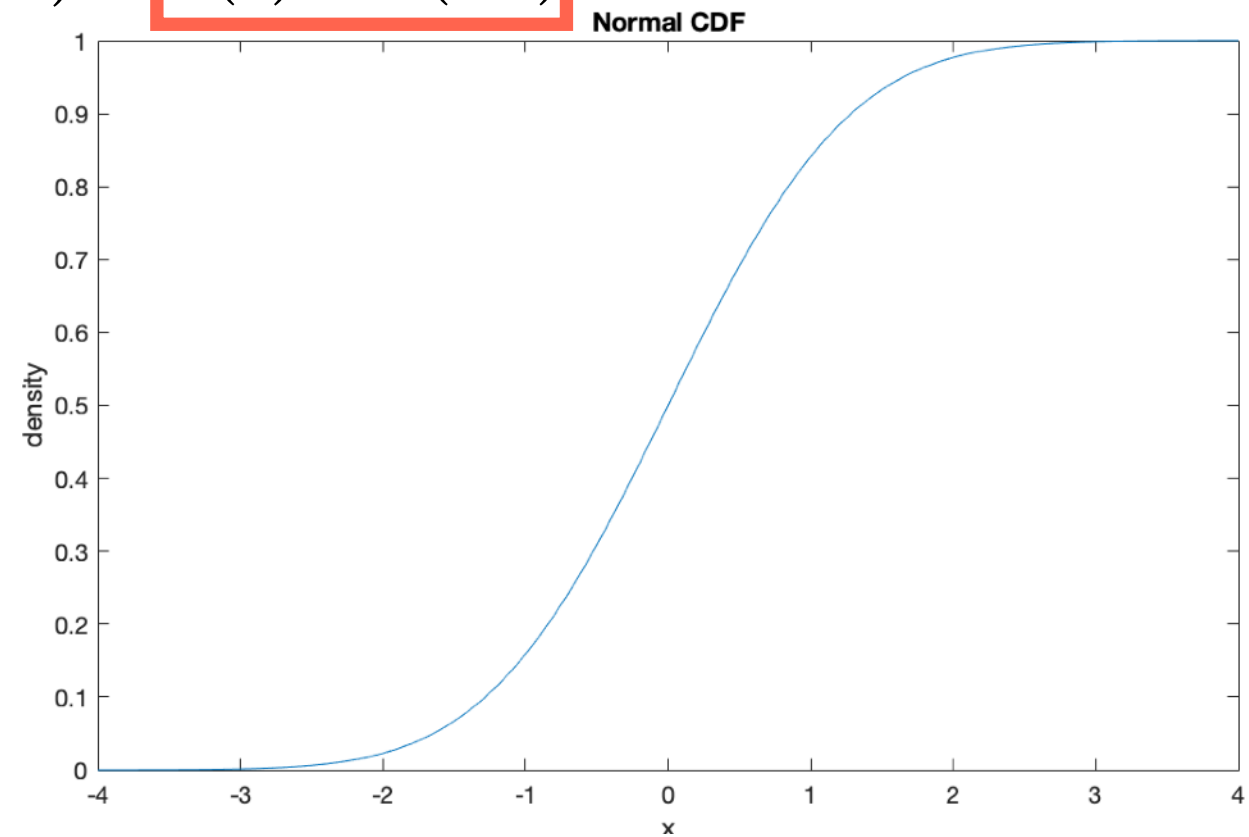$$= P\left(-k \leq \frac{X - \mu}{\sigma} \leq k\right)$$
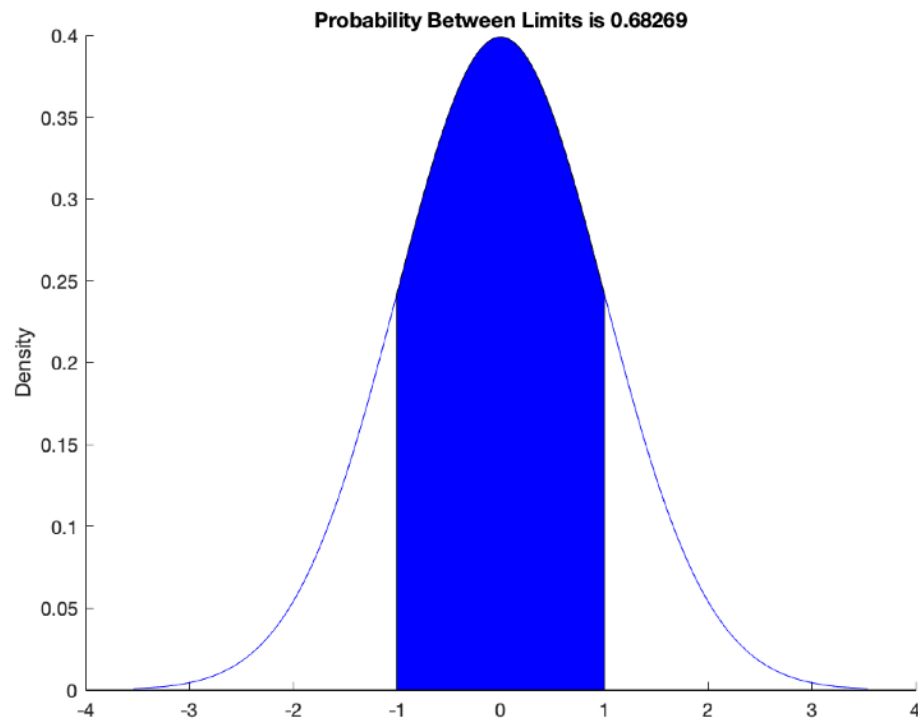
$$= P(-k \leq Z \leq k),$$

$$P(-k \leq Z \leq k) = P(Z \leq k) - P(Z \leq -k) = \boxed{F(k) - F(-k)}$$

Normal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$$P(-k \leq Z \leq k) = P(Z \leq k) - P(Z \leq -k) = \boxed{F(k) - F(-k)}$$

Normal Distribution

- Singled peaked, symmetric

- about 2/3 are within 1 standard deviation (σ) from the mean

- about 95% are within 2 standard deviations (2σ) from the mean

```
## plus minus 1 standard deviation from the mean
pnorm(1) - pnorm(-1)

## [1] 0.6826895

## plus minus 2 standard deviations from the mean
pnorm(2) - pnorm(-2)

## [1] 0.9544997
```

```
mu <- 5
sigma <- 2
## plus minus 1 standard deviation from the mean
pnorm(mu + sigma, mean = mu, sd = sigma) - pnorm(mu - sigma, mean = mu, sd = sigma)

## [1] 0.6826895

## plus minus 2 standard deviations from the mean
pnorm(mu + 2*sigma, mean = mu, sd = sigma) - pnorm(mu - 2*sigma, mean = mu, sd = sigma)

## [1] 0.9544997
```

Expectation: Definition and General Properties

▶ Expectation (population mean) of a random variable X

  ▶ Fixed value given a probability distribution (different from sample means)

$$\mathbb{E}(X) = \begin{cases} \sum_x x \times f(x) & \text{if } X \text{ is discrete,} \\ \int x \times f(x)\, dx & \text{if } X \text{ is continuous} \end{cases}$$

▶ Properties of expectation (a, b: constant values; X,Y: independent RVs)

1. $\mathbb{E}(a) = a$.
2. $\mathbb{E}(aX) = a\mathbb{E}(X)$.
3. $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.
4. $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.
5. If X and Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. But generally, $\mathbb{E}(XY) \neq \mathbb{E}(X)\mathbb{E}(Y)$.

Expectation: Examples

$$\mathbb{E}(X) = \begin{cases} \sum_x x \times f(x) & \text{if } X \text{ is discrete,} \\ \int x \times f(x)\,dx & \text{if } X \text{ is continuous} \end{cases}$$

- ▶ Expectation (population mean)

  - ▶ Expected value of a random variable

    - ▶ e.g. PMF: Bernoulli random variable

    $$\mathbb{E}(X) = 0 \times P(X=0) + 1 \times P(X=1) = 0 \times f(0) + 1 \times f(1) = 0 \times (1-p) + 1 \times p = p$$

    - ▶ e.g. PMF: Binomial random variable

    $$\mathbb{E}(X) = 0 \times f(0) + 1 \times f(1) + \cdots + n \times f(n) = \sum_{x=0}^{n} x \times f(x)$$

    - ▶ e.g. PMF: Binomial random variable (Y is a Bernoulli RV with p)

    $$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \mathbb{E}(Y_i) = np$$

    - ▶ e.g. PDF: uniform random variable defined in the interval [a,b]

    $$\mathbb{E}(X) = \int_a^b x \times f(x)\,dx = \int_a^b \frac{x}{b-a}\,dx = \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{a+b}{2}$$

Variance: Definition and General Properties

▶ Population variance (different from sample variance)

$$\mathbb{V}(X) = \mathbb{E}[\{X - \mathbb{E}(X)\}^2]$$
$$= \mathbb{E}[X^2 - 2X\mathbb{E}(X) + \{\mathbb{E}(X)\}^2]$$
$$= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \{\mathbb{E}(X)\}^2$$
$$= \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2.$$

▶ e.g. PMF: Bernoulli random variable

$$\mathbb{V}(X) = \mathbb{E}(X) - \{\mathbb{E}(X)\}^2 = p(1 - p)$$

▶ e.g. PDF: Uniform random variable

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2 = \int_a^b \frac{x^2}{b-a}\,dx - \left(\frac{a+b}{2}\right)^2$$

$$= \left.\frac{x^3}{3(b-a)}\right|_a^b - \left(\frac{a+b}{2}\right)^2 = \frac{1}{12}(b-a)^2.$$

▶ Square root of population variance is population standard deviation

# Variance: Definition and General Properties

1. $\mathbb{V}(a) = 0$.

2. $\mathbb{V}(aX) = a^2\mathbb{V}(X)$.

3. $\mathbb{V}(X + b) = \mathbb{V}(X)$.

4. $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$.

5. If X and Y are independent, $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.

Contents

- ▶ Random Variables and Probability Distributions (Chapter 6.3)

    - ▶ Overview

    - ▶ Bernoulli / Binomial distribution

    - ▶ Uniform / Normal distribution

    - ▶ Expectation and Variance

- ▶ Election prediction (6.3.6)
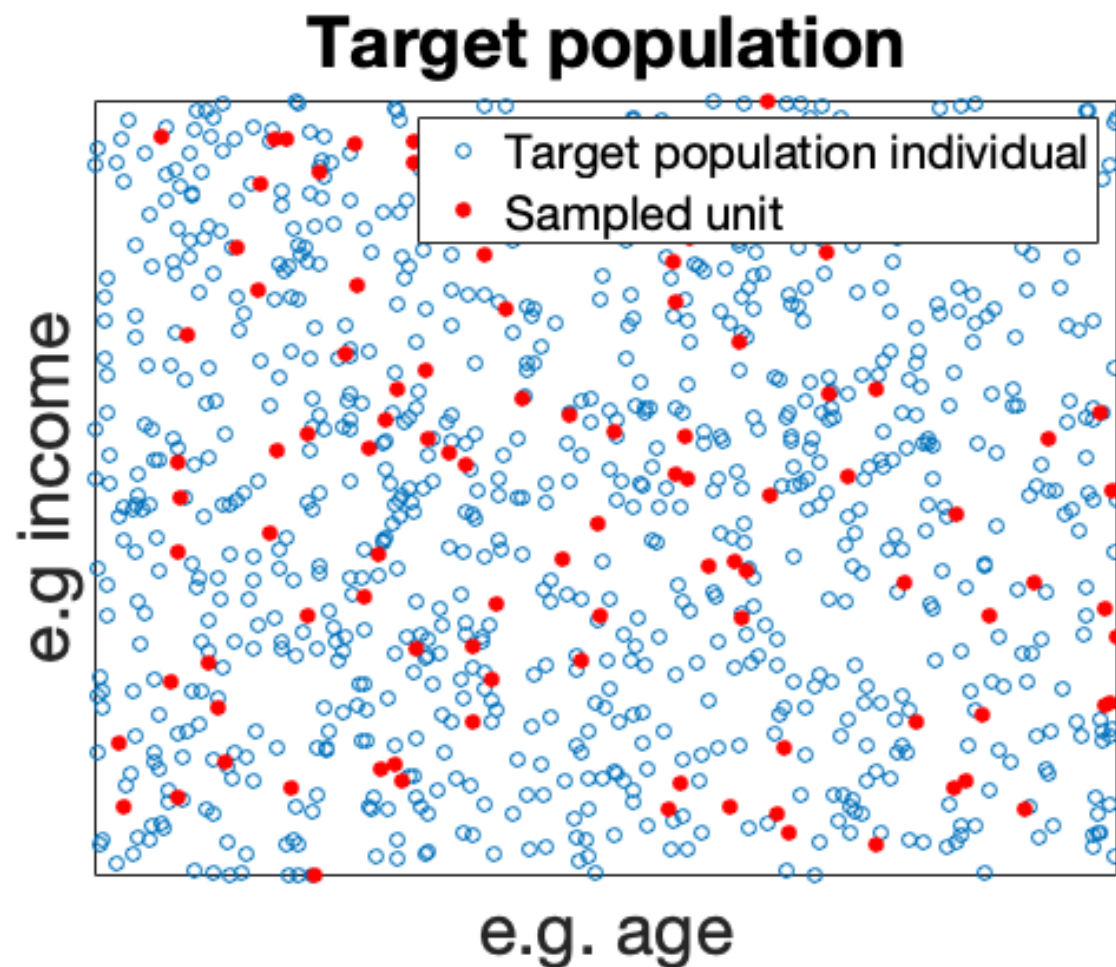
- ▶ The Law of large numbers (6.4.1)

Election Prediction (6.3.6)

► Binomial outcomes in 2 candidate elections

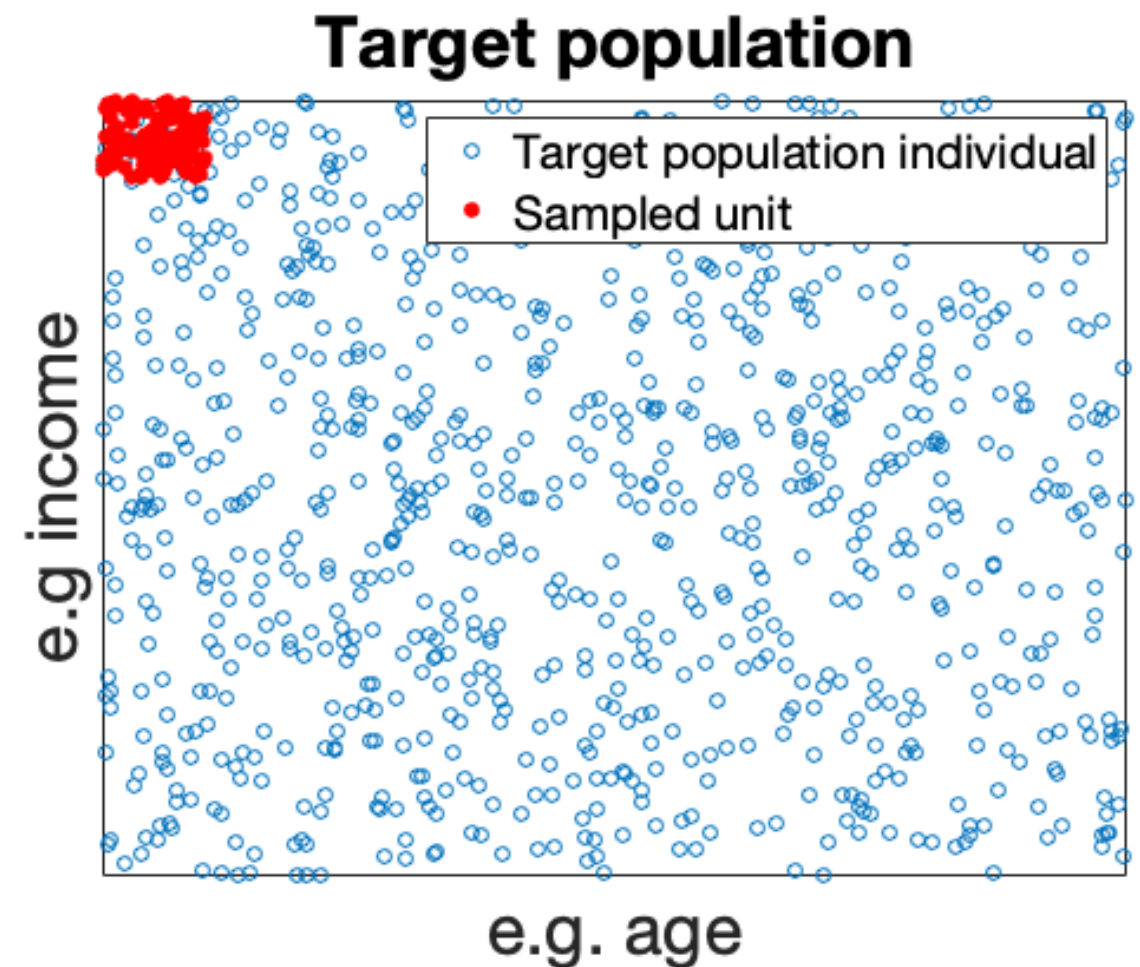   ► Vote share = sum of individual binary choices
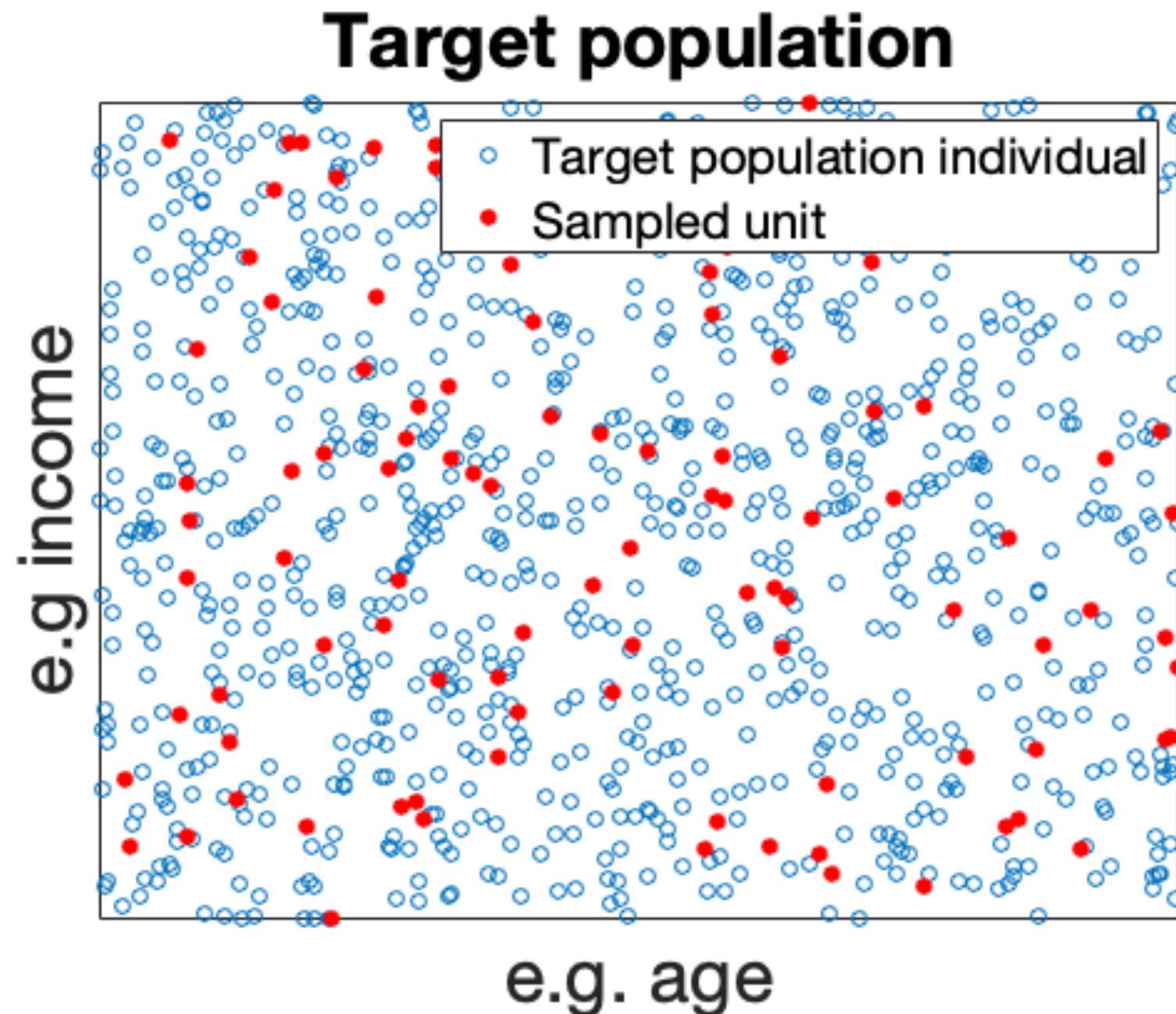
Election Prediction (6.3.6)

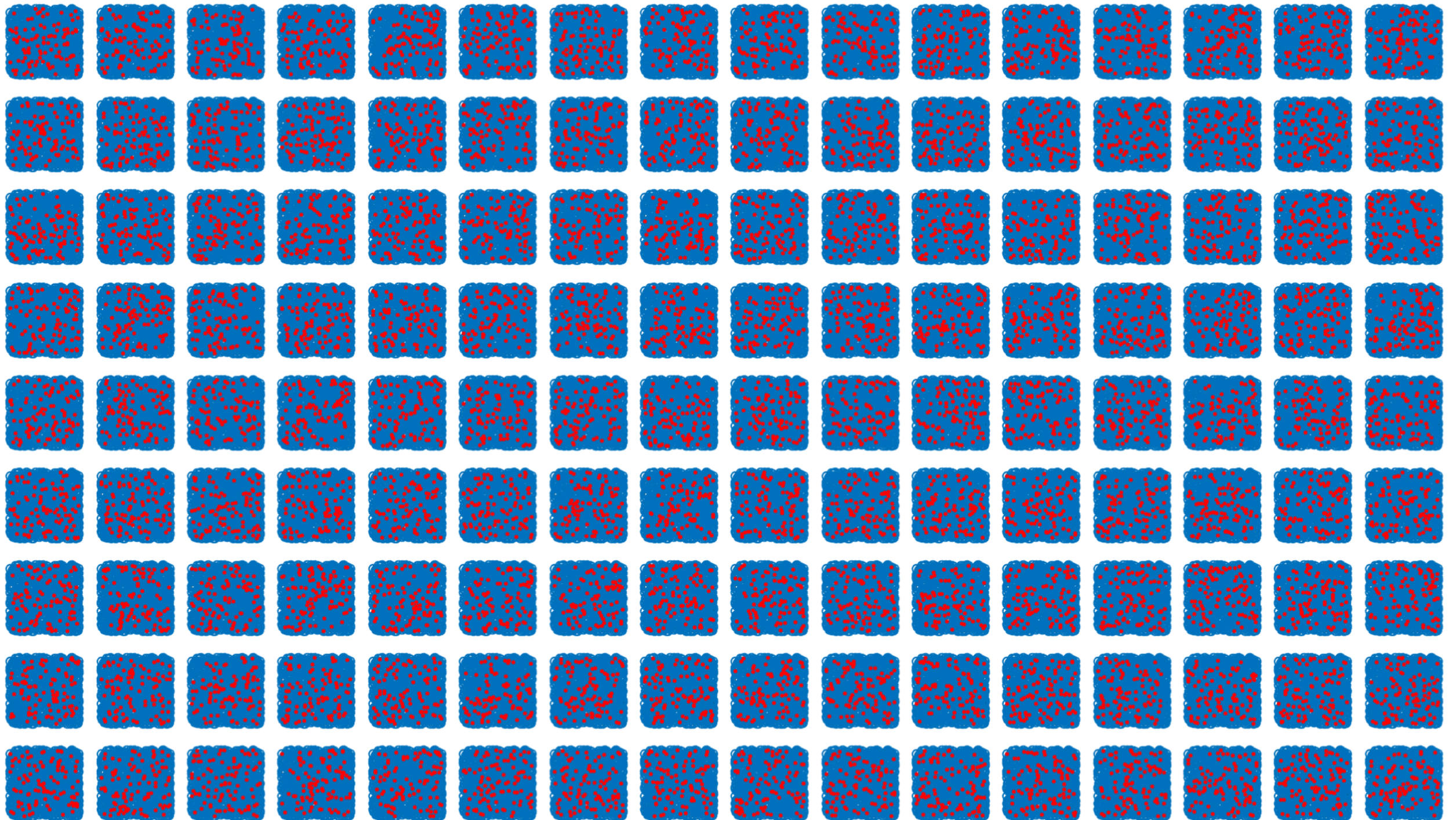► e.g. even for the case of random sampling: **sampling variability**
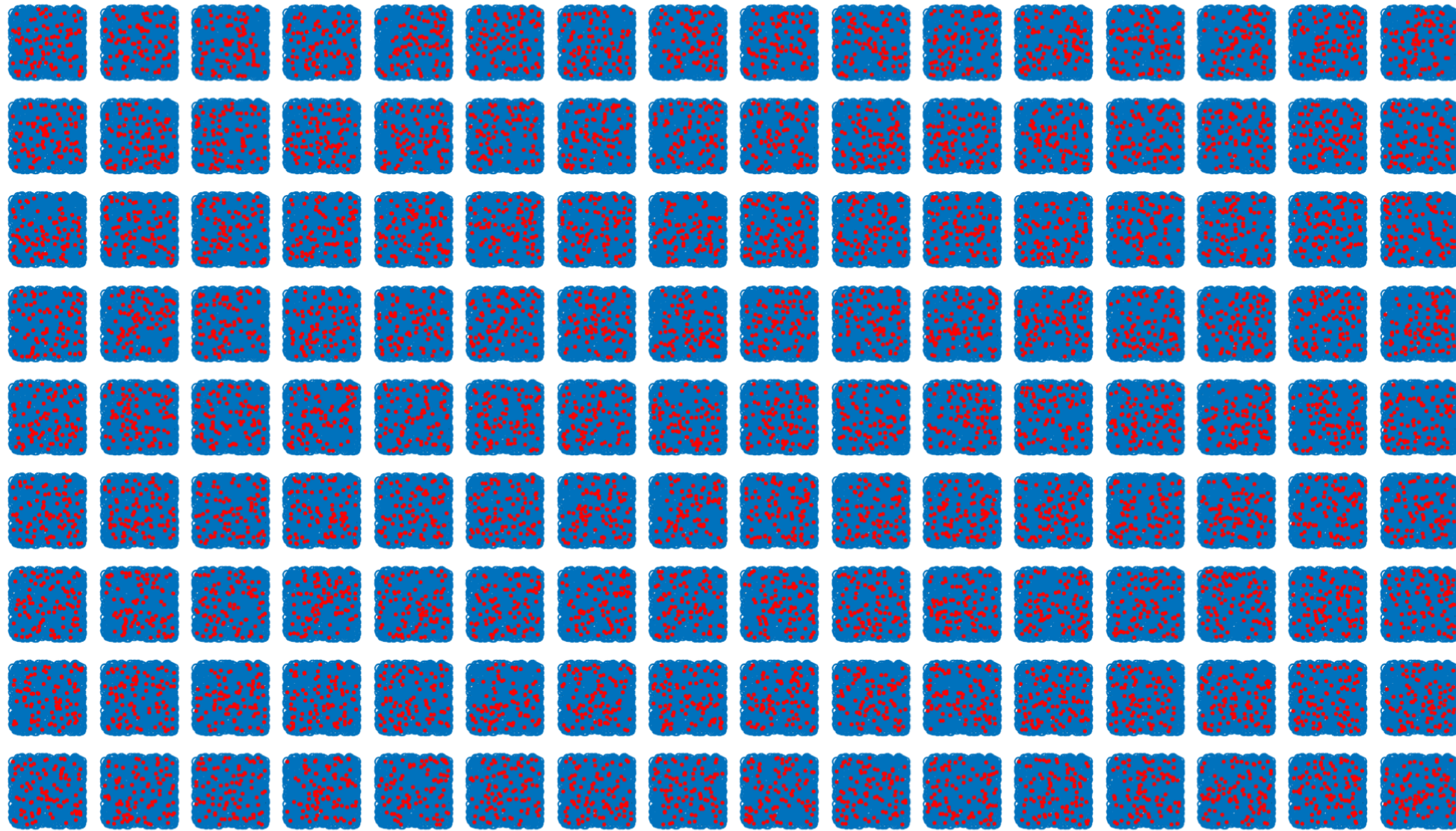

VS

► We cannot ignore such possibility!

Election Prediction (6.3.6)

▶ e.g. even for the case of random sampling: **sampling variability**

    ▶ e.g. selecting 100 individuals from 1,000 (each exhibits different share)

Election Prediction (6.3.6)
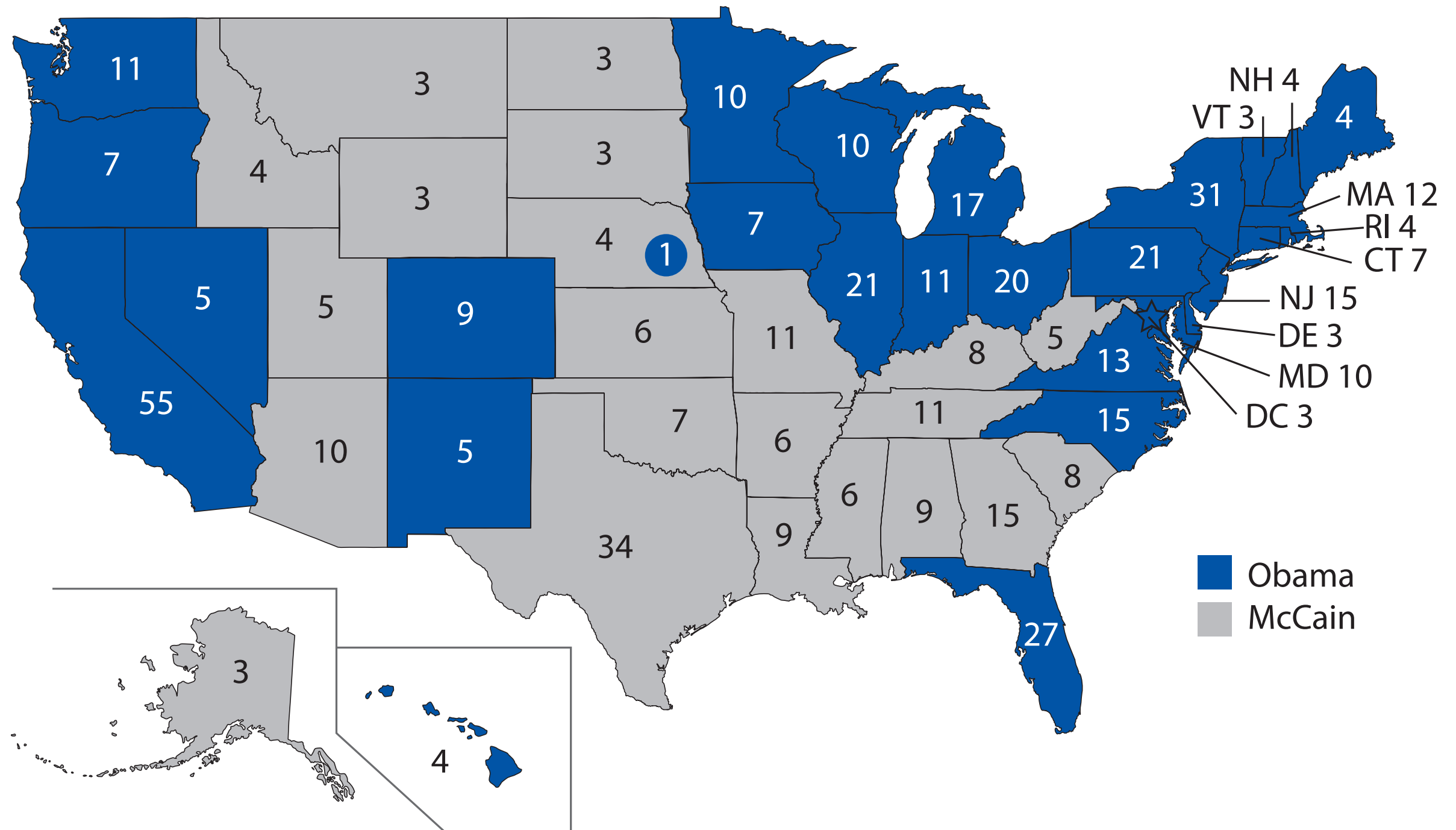
▶ e.g. even for the case of random sampling: **sampling variability**

▶ From candidate preference of individuals in your sample:

    ▶ Infer **target population preference** -> **Uncertainty** must be involved!

    ▶ Time to use random variables (binomial distribution)

Election Prediction (4.1.3)

► Electoral college system: winner-take-all for Electoral College votes /state

Legend:

- Obama (blue)
- McCain (gray)

Election Prediction (6.3.6)

▶ Given the population-level voting propensity (ignore 3rd candidate)

  ▶ Possible outcomes of 1,000 individual / state samples:

```
pres08 <- read.csv("pres08.csv")
## two-party vote share
pres08$p <- pres08$Obama / (pres08$Obama + pres08$McCain)
```

$$\mathbb{E}(\text{Obama's votes}) = \sum_{j=1}^{51} v_j \times P(\text{Obama wins state } j) = \sum_{j=1}^{51} v_j \times P(S_j > 500)$$

$$\mathbb{V}(\text{Obama's predicted votes}) = \sum_{j=1}^{51} \mathbb{V}(v_j \mathbf{1}\{S_j > 500\})$$

$$= \sum_{j=1}^{51} v_j^2 P(S_j > 500)\big(1 - P(S_j > 500)\big)$$
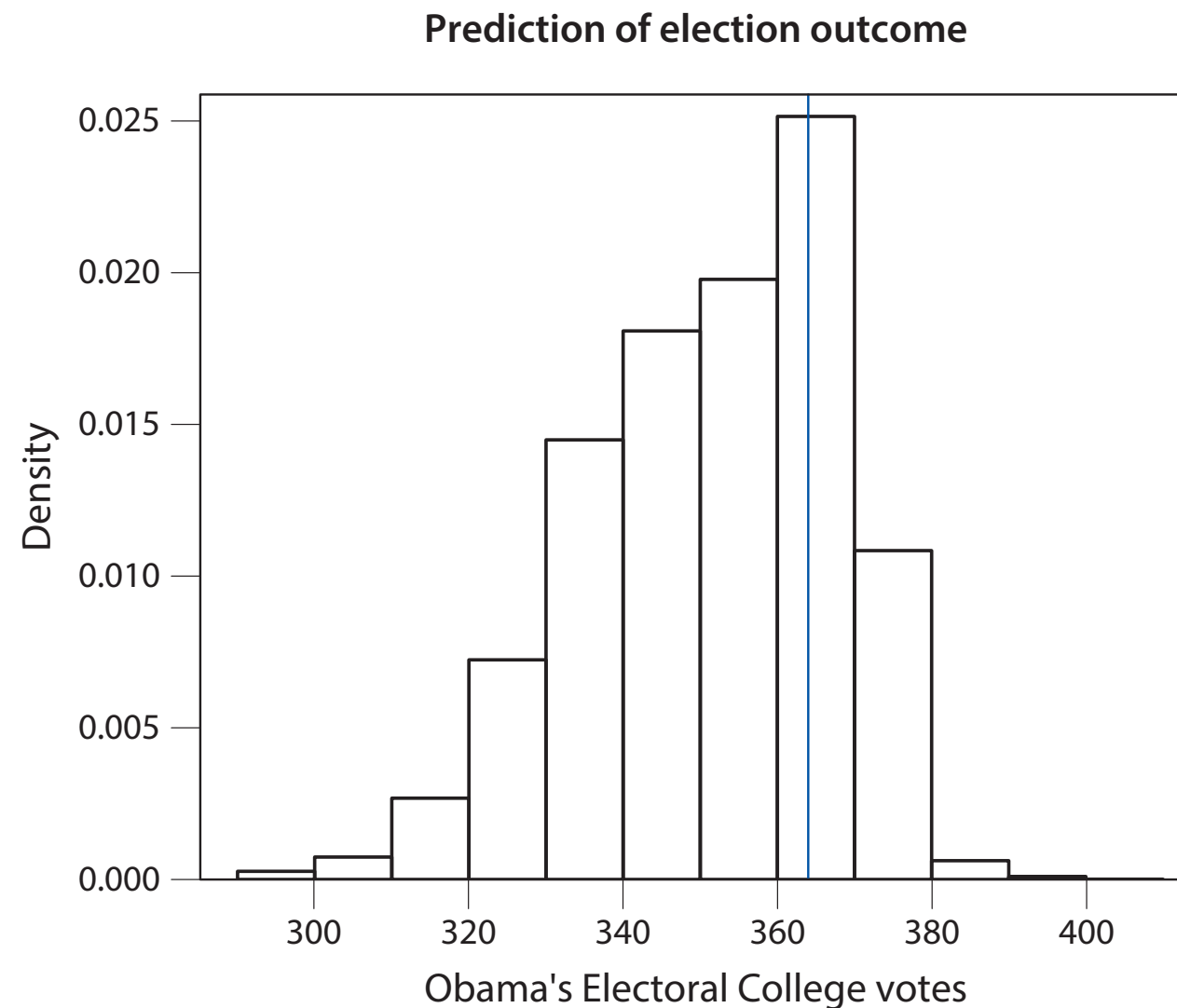
Election Prediction (6.3.6)

▶ Given the population-level voting propensity (ignore 3rd candidate)

   ▶ Possible outcomes of 1,000 individual / state samples:

```r
pres08 <- read.csv("pres08.csv")
## two-party vote share
pres08$p <- pres08$Obama / (pres08$Obama + pres08$McCain)
```

```r
n.states <- nrow(pres08) # number of states
n <- 1000 # number of respondents
sims <- 10000 # number of simulations
## Obama's electoral votes
Obama.ev <- rep(NA, sims)
for (i in 1:sims) {
    ## samples number of votes for Obama in each state
    draws <- rbinom(n.states, size = n, prob = pres08$p)
    ## sums state's Electoral College votes if Obama wins the majority
    Obama.ev[i] <- sum(pres08$EV[draws > n / 2])
}
hist(Obama.ev, freq = FALSE, main = "Prediction of election outcome",
      xlab = "Obama's Electoral College votes")
abline(v = 364, col = "blue") # actual result
```
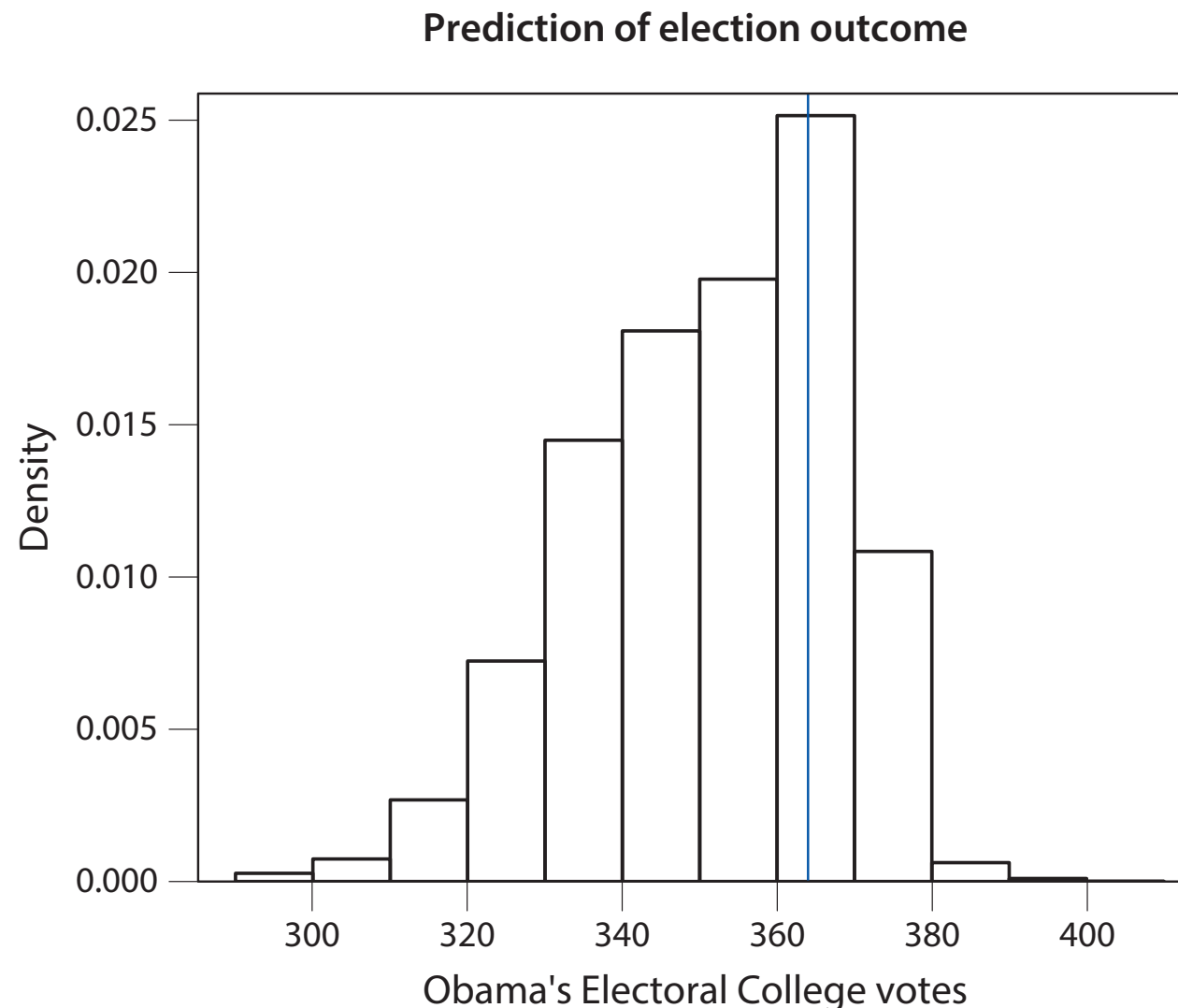
Election Prediction (6.3.6)

▶ Given the population-level voting propensity (ignore 3rd candidate)

    ▶ Possible outcomes of 1,000 individual / state samples:

Prediction of election outcome

Obama's Electoral College votes

```
summary(Obama.ev)

##     Min. 1st Qu.   Median     Mean 3rd Qu.     Max.
##    291.0    340.0    353.0    352.2    364.0    401.0
```

Election Prediction (4.1.3)

Prediction of election outcome

- Prediction error (ε) = actual outcome - predicted outcome

- Average predictor error is called bias

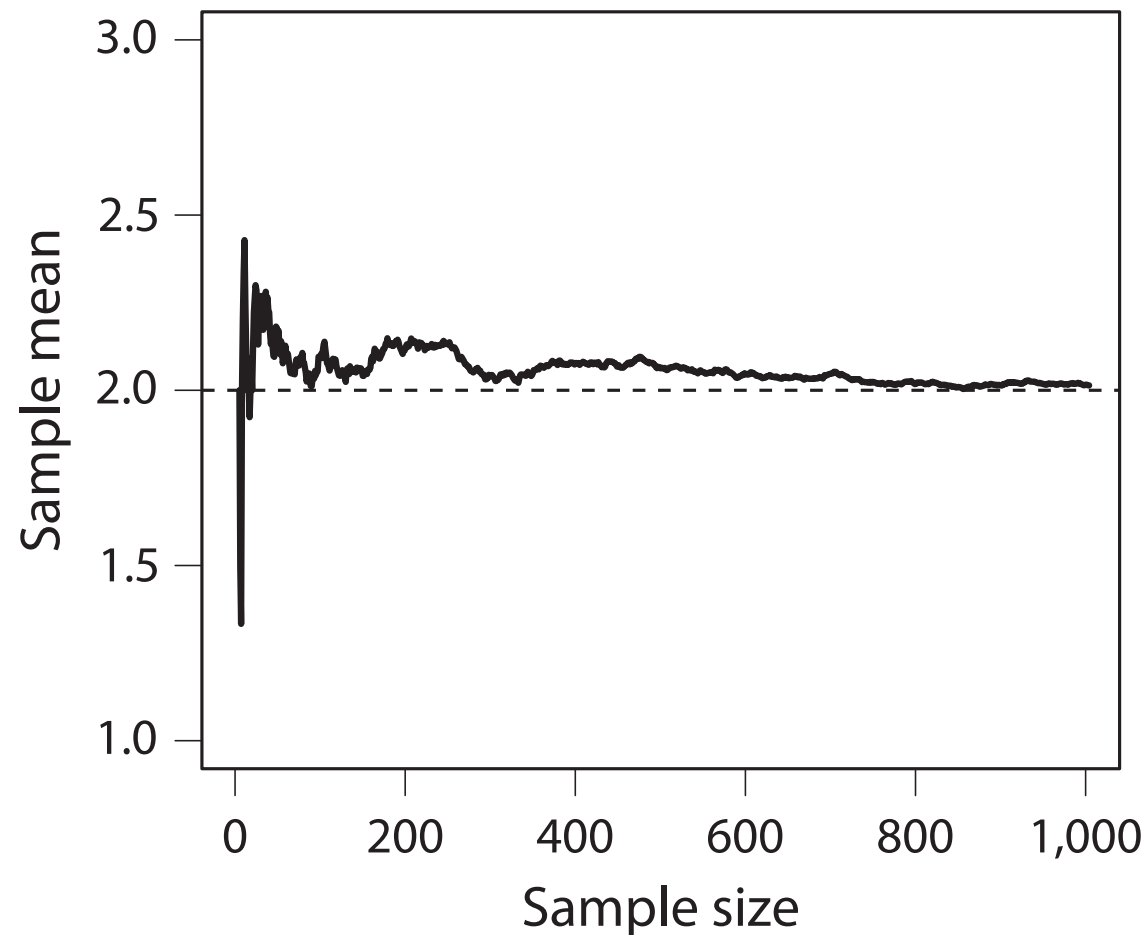- Root-mean-squared error: $\sqrt{\sum\limits_{i=1}^{n} \epsilon_i^2 / n}$

Contents

- Random Variables and Probability Distributions (Chapter 6.3)

    - Overview

    - Bernoulli / Binomial distribution

    - Uniform / Normal distribution

    - Expectation and Variance

- Election prediction (6.3.6)

- The Law of large numbers (6.4.1)

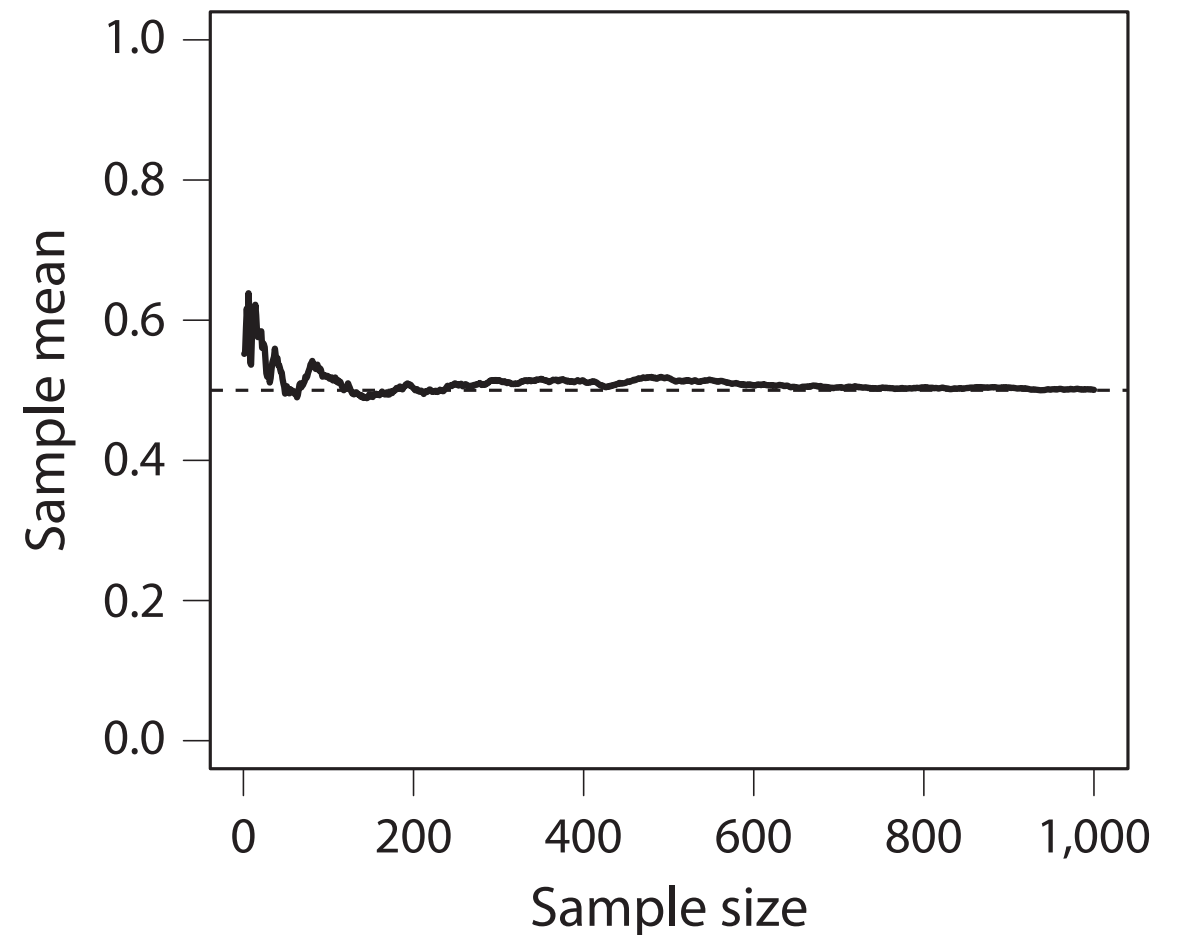- Central limit theorem (6.4.2)

The Law of Large Numbers (6.4.1)

▶ As sample size grows, sample mean approaches the population mean

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \to \mathbb{E}(X)$$

Binomial (10, 0.2)

Uniform (0, 1)

The Law of Large Numbers (6.4.1)

▶ e.g. Randomized control trials (Randomized experiments)

  ▶ Difference in the means estimator (sample average treatment effect)

    ▶ Sample average difference $\longrightarrow$??

      ▶ Recall

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \to \mathbb{E}(X)$$

      ▶ Sample ATE converges to Population average treatment effect

$$\overline{Y(1)} - \overline{Y(0)} = \frac{1}{n_1} \sum_{i=1}^{n_1} Y(1)_i - \frac{1}{n_0} \sum_{i=1}^{n_0} Y(0)_i \to \mathbb{E}(Y(1)) - \mathbb{E}(Y(0))$$

Summary

- ▶ Random Variables and Probability Distributions (Chapter 6.3)

  - ▶ Overview

  - ▶ Bernoulli / Binomial distribution

  - ▶ Uniform / Normal distribution

  - ▶ Expectation and Variance

- ▶ Election prediction (6.3.6)

- ▶ The Law of large numbers (6.4.1)

Next Week Preview

- ▶ Why is normal the most important and popular distribution?
  - ▶ https://youtu.be/6YDHBFVlvls

See you next week.