Probability Distributions and Estimation

Week 12

Yunkyu Sohn
School of Political Science and Economics
Waseda University

2019 Spring Statistics II

Contents

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?
  - <u>Area</u> under the density curve
  - Non-decreasing

Binomial Distribution

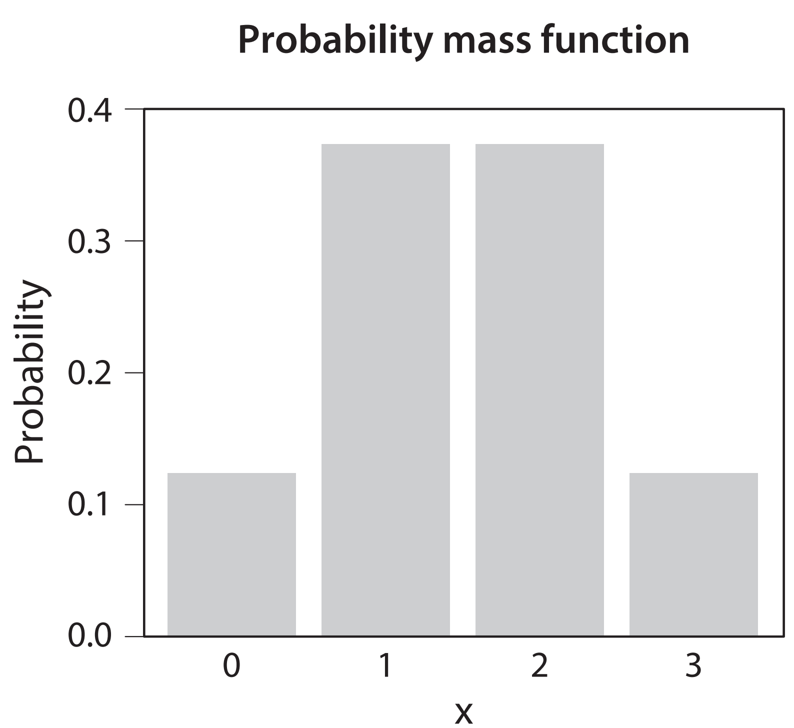
- The number of 1s (one of the binary outcomes) in multiple Bernoulli trials
- $\Omega=\{0,1,\ldots.,n-1,n\}$
- PMF

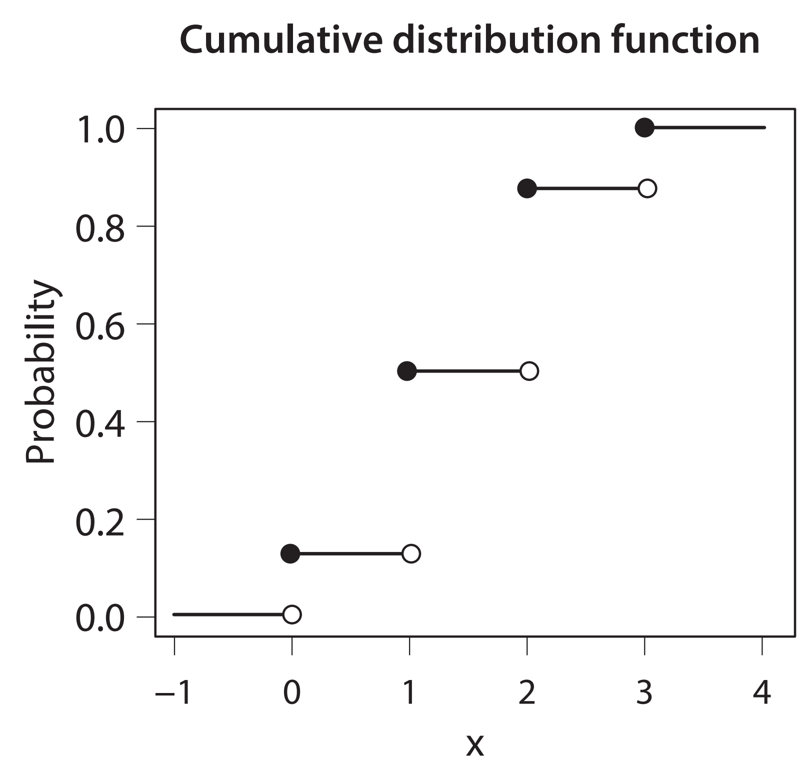
$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \binom{n}{x} = {}_nC_x$$

- CMF

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

- $p = 0.5$ and $n = 3$





Uniform Distribution

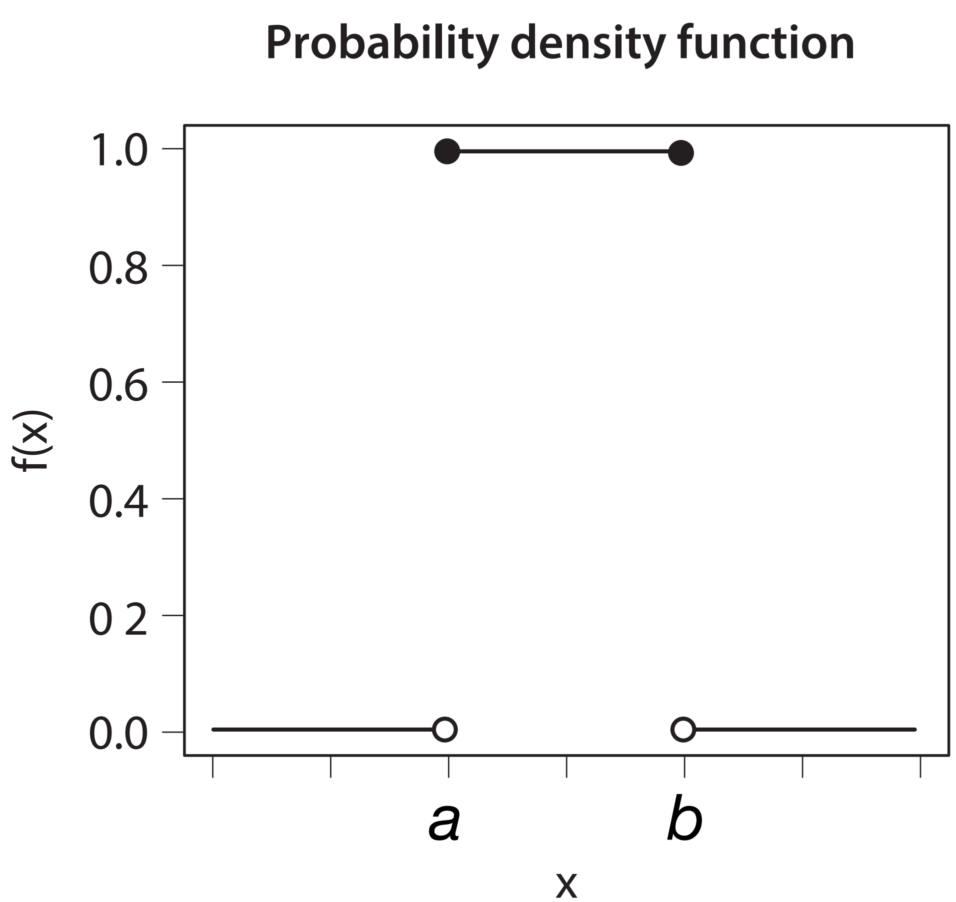
- Every number in an interval has an equal chance of appearance
- $\Omega =$ **set of real numbers in a range** $[a, b]$
- PDF

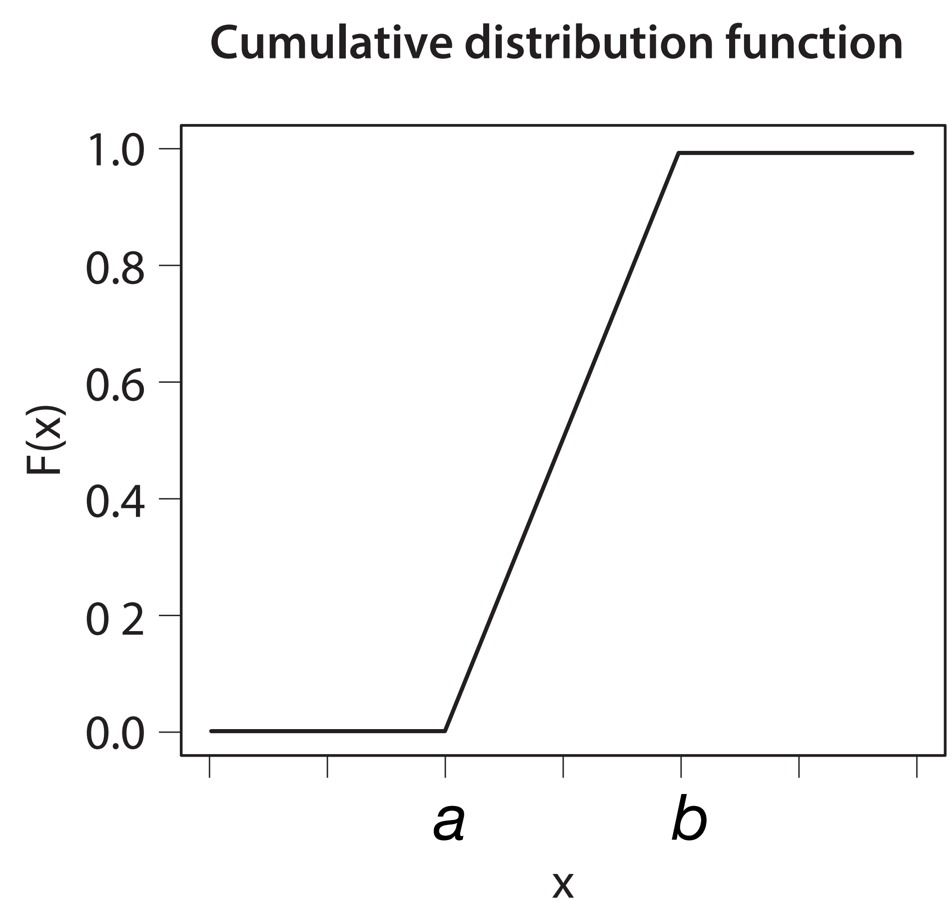
$$f(x) = \begin{cases} \frac{1}{b-a} & \textbf{if } a \le x \le b \\ 0 & \textbf{otherwise.} \end{cases}$$

- CDF

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

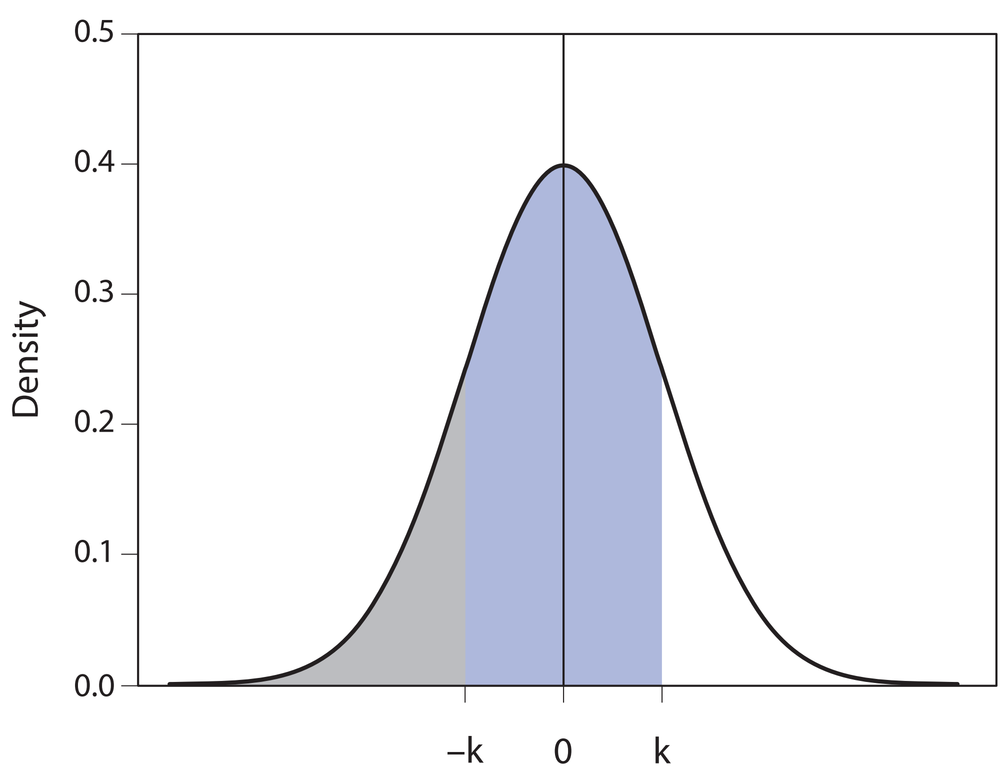
- Uniform distribution for the interval $[a,b]$





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 (2σ) from the mean

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

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_i 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)

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

- 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

$$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2 = \int_a^b \frac{x^2}{b-a}\,dx - \left(\frac{a+b}{2}\right)^2 \\ &= \left.\frac{x^3}{3(b-a)}\right|_a^b - \left(\frac{a+b}{2}\right)^2 = \frac{1}{12}(b-a)^2.\end{aligned}$$

- Square root of population variance is population standard deviation

https://towardsdatascience.com/why-sample-variance-is-divided-by-n-1-89821b83ef6d

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, $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.

Normal Distribution

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

- If X is a normal RV, any transformation of form aX + b will follows normal
- $X \sim \mathcal{N}(\mu, \sigma^2)$
- $S = X + c \rightarrow S \sim \mathcal{N}(\mu + c, \sigma^2)$
- $Y = cX \rightarrow Y \sim \mathcal{N}(c\mu, (c\sigma)^2)$
- $T = aX + b \rightarrow T \sim \mathcal{N}(a\mu + b, (a\sigma)^2)$
- **z-score:** $Z = (X - \mu)/\sigma \rightarrow Z \sim \mathcal{N}(0,1)$

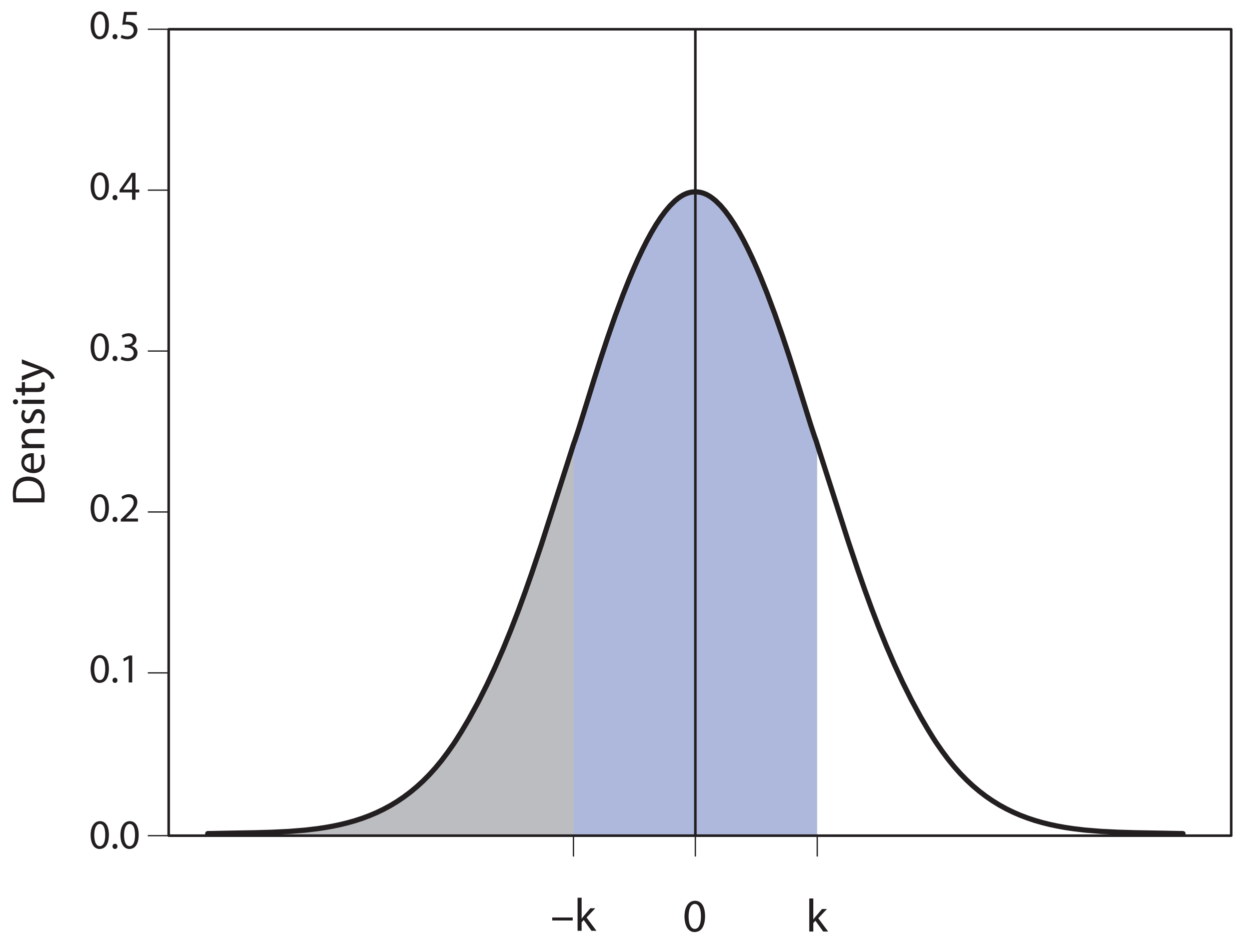
1. $\mathbb{E}(a) = a$.
2. $\mathbb{E}(aX) = a\mathbb{E}(X)$.
3. $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.
4. $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.

1. $\mathbb{V}(a) = 0$.
2. $\mathbb{V}(aX) = a^2\mathbb{V}(X)$.
3. $\mathbb{V}(X + b) = \mathbb{V}(X)$.
4. $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$.

https://www.youtube.com/watch?v=nqNzKeVCYBU

Normal Distribution

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



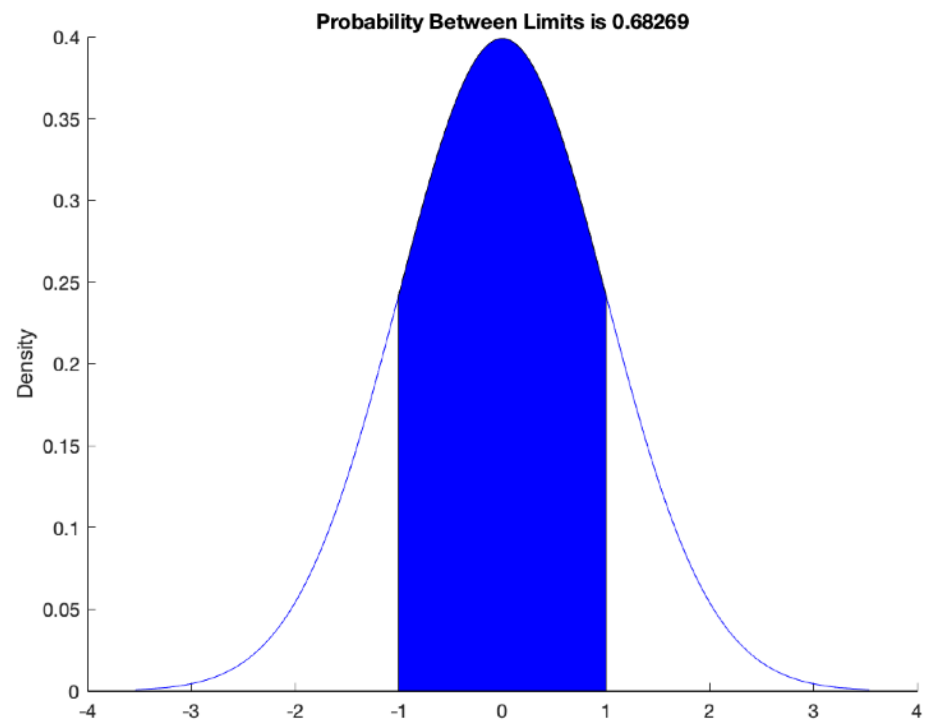
- Probability that a normal random variable with mean **μ** and sdv **σ** lies within k standard deviations from the mean for a positive constant $k > 0$

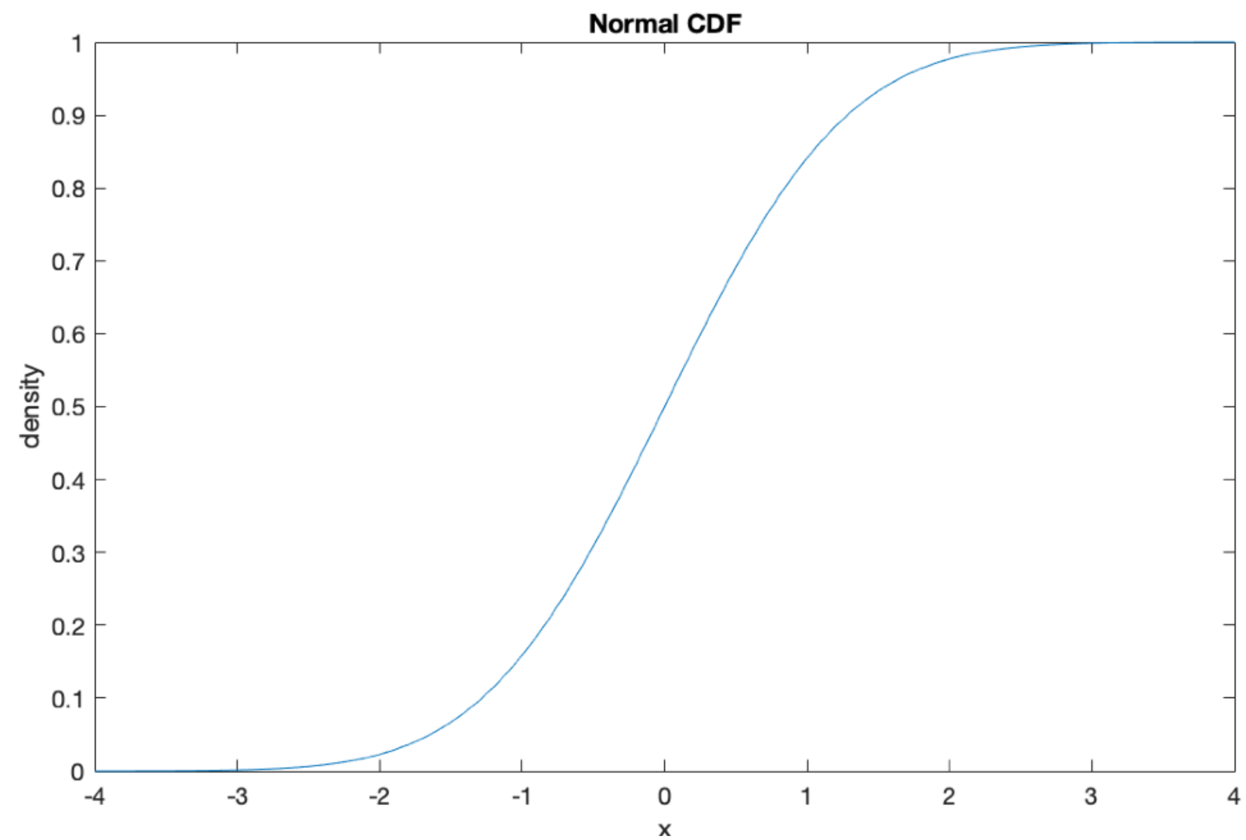
$$\begin{aligned} P(\mu - k\sigma \le X \le \mu + k\sigma) &= P(-k\sigma \le X - \mu \le k\sigma) \\ &= P\left(-k \le \frac{X-\mu}{\sigma} \le k\right) \\ &= P(-k \le Z \le k), \end{aligned}$$

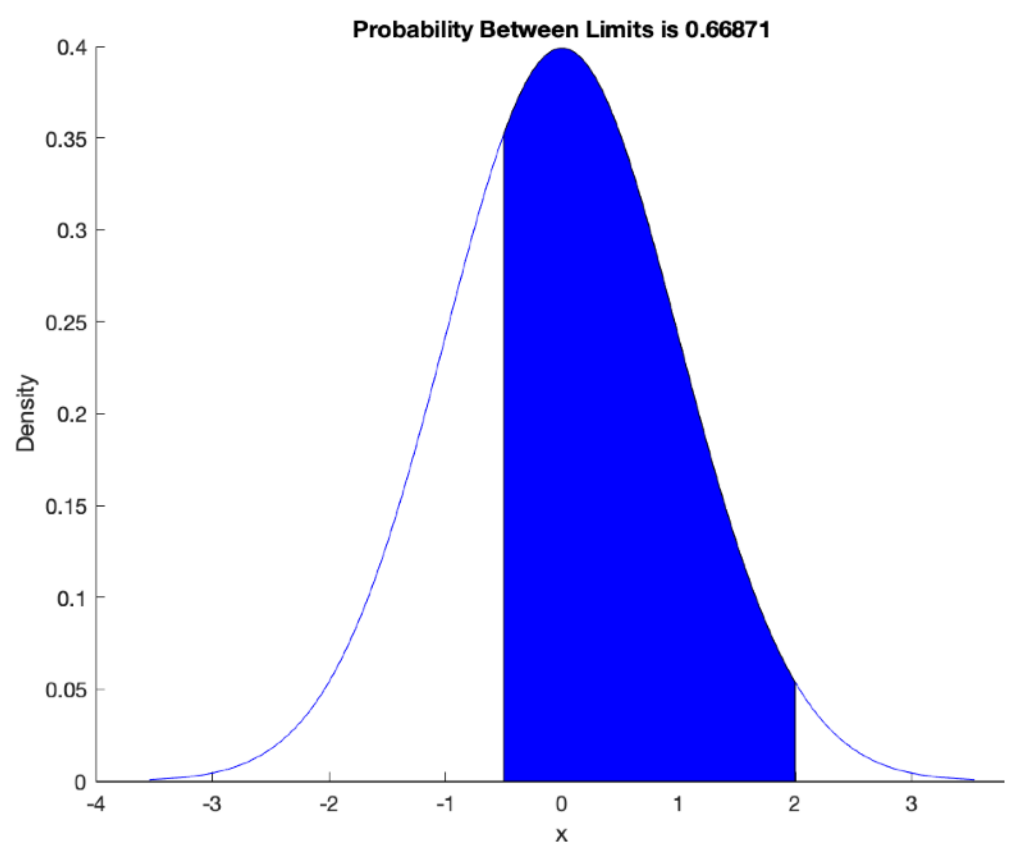
$$P(-k \le Z \le k) = P(Z \le k) - P(Z \le -k) = \boxed{F(k) - F(-k)}$$

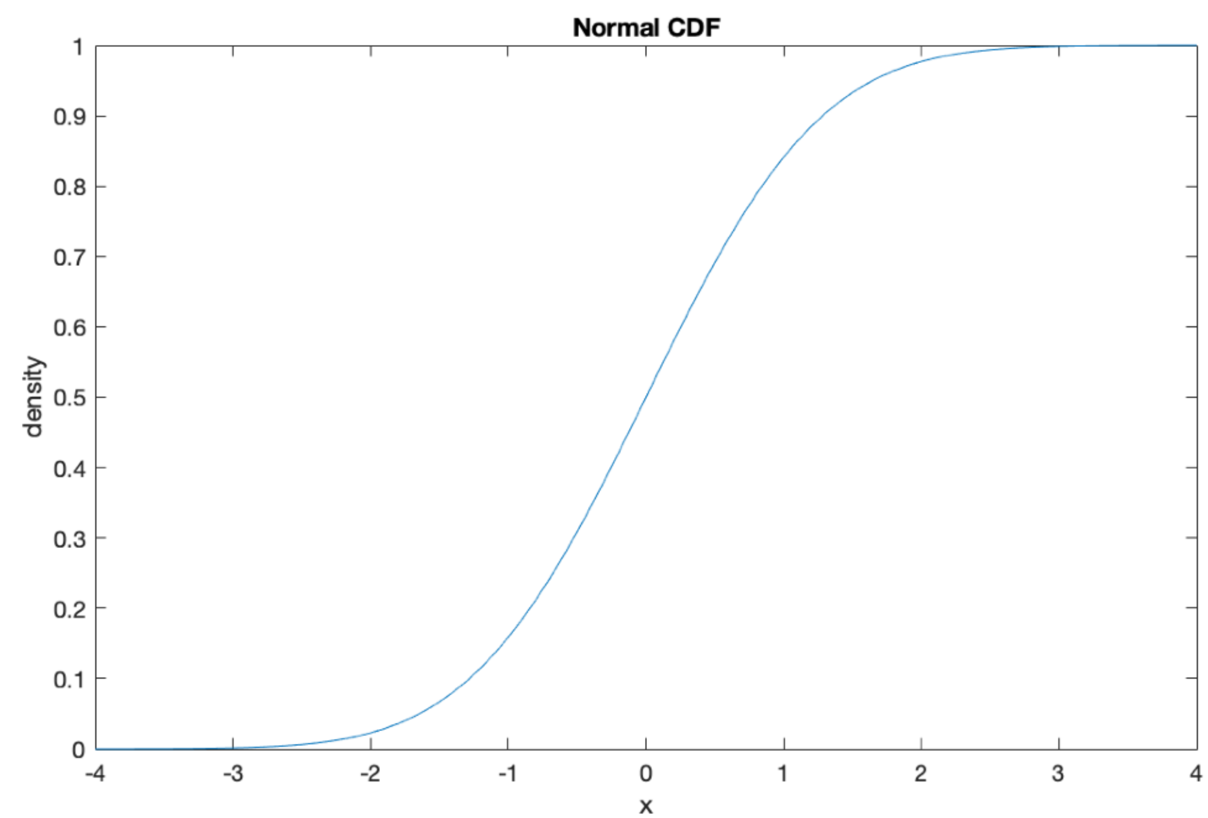
Normal Distribution

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









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
```

Variance: Definition and General Properties

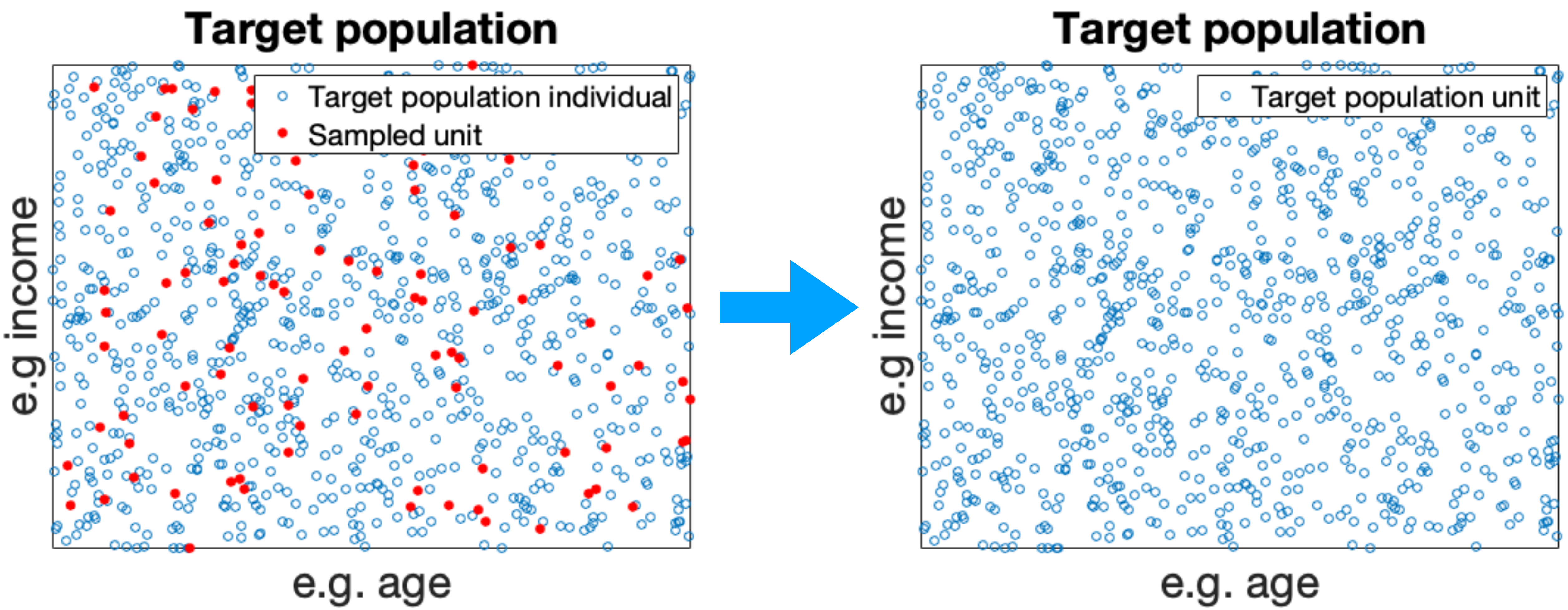
1. $\mathbb{V}(a) = 0$.
2. $\mathbb{V}(aX) = a^2\mathbb{V}(X)$.
3. $\mathbb{V}(X + b) = \mathbb{V}(X)$.
4. $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$.
5. If X and Y are independent, $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$.

Contents

The Law of Large Numbers (6.4.1)

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

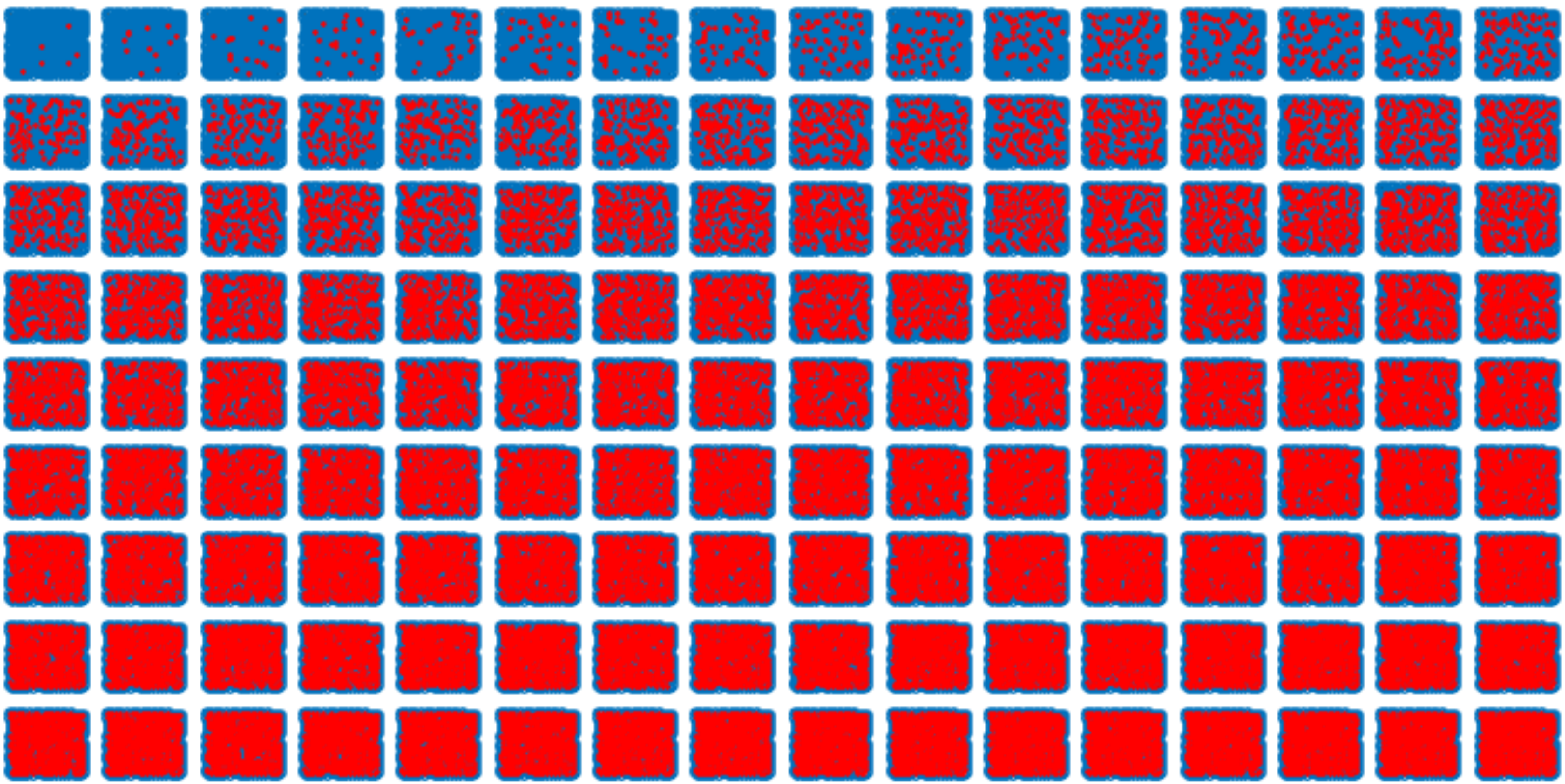
$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}(X)$$



The Law of Large Numbers (6.4.1)

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

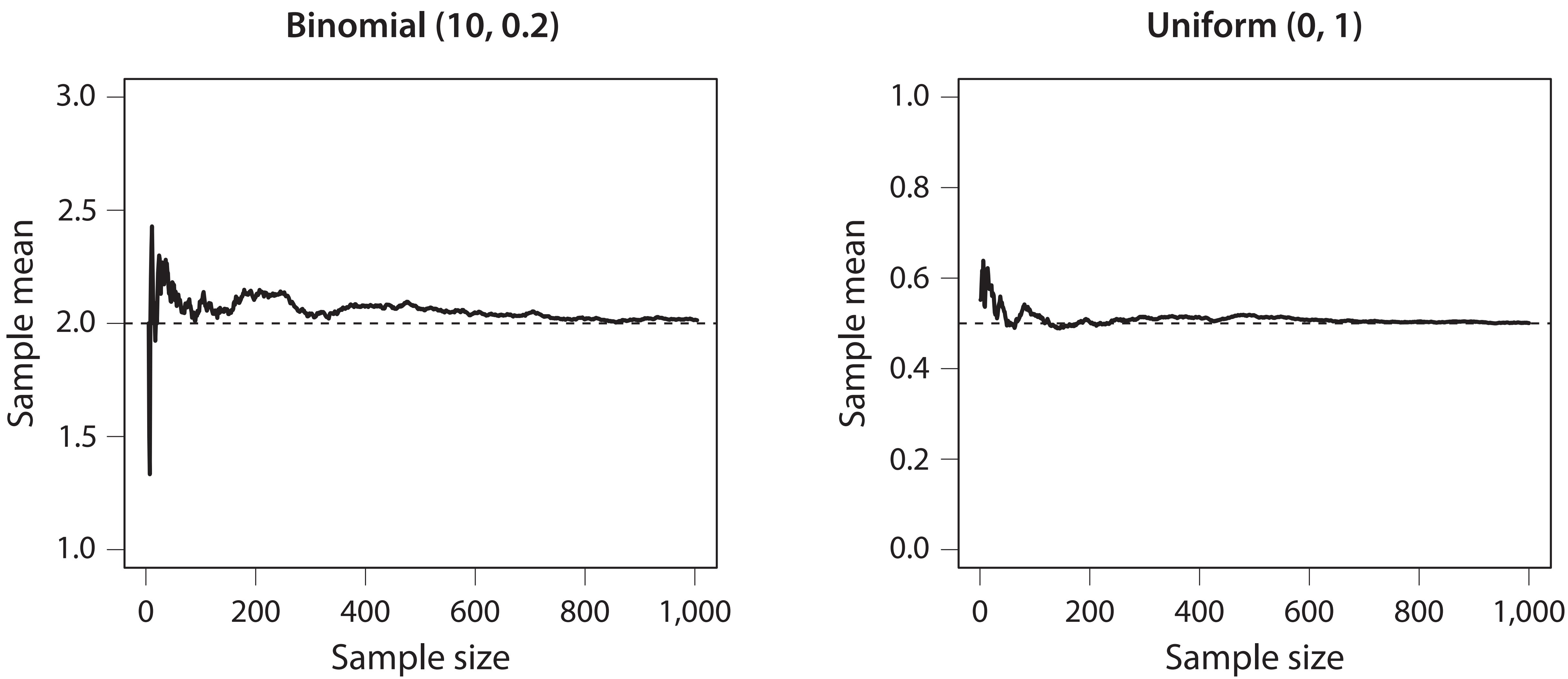
$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}(X)$$



The Law of Large Numbers (6.4.1)

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

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



The Central Limit Theorem (6.4.2)

- The law of large numbers:

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

  - relationship between the size of sample, its mean, and population mean
    - What about its distribution?

- The central limit theorem

Suppose that we obtain a random sample of n independently and identically distributed (i.i.d.) observations, $X_1, X_2, \ldots, X_n$, from a probability distribution with mean $\mathbb{E}(X)$ and variance $\mathbb{V}(X)$. Let us denote the sample average by $\overline{X}_n = \sum_{i=1}^{n} X_i/n$. Then, the **central limit theorem** states

$$\frac{\overline{X}_n - \mathbb{E}(X)}{\sqrt{\mathbb{V}(X)/n}} \rightsquigarrow \mathcal{N}(0, 1). \tag{6.41}$$

In the theorem, $\rightsquigarrow$ indicates "convergence in distribution" as the sample size n increases.

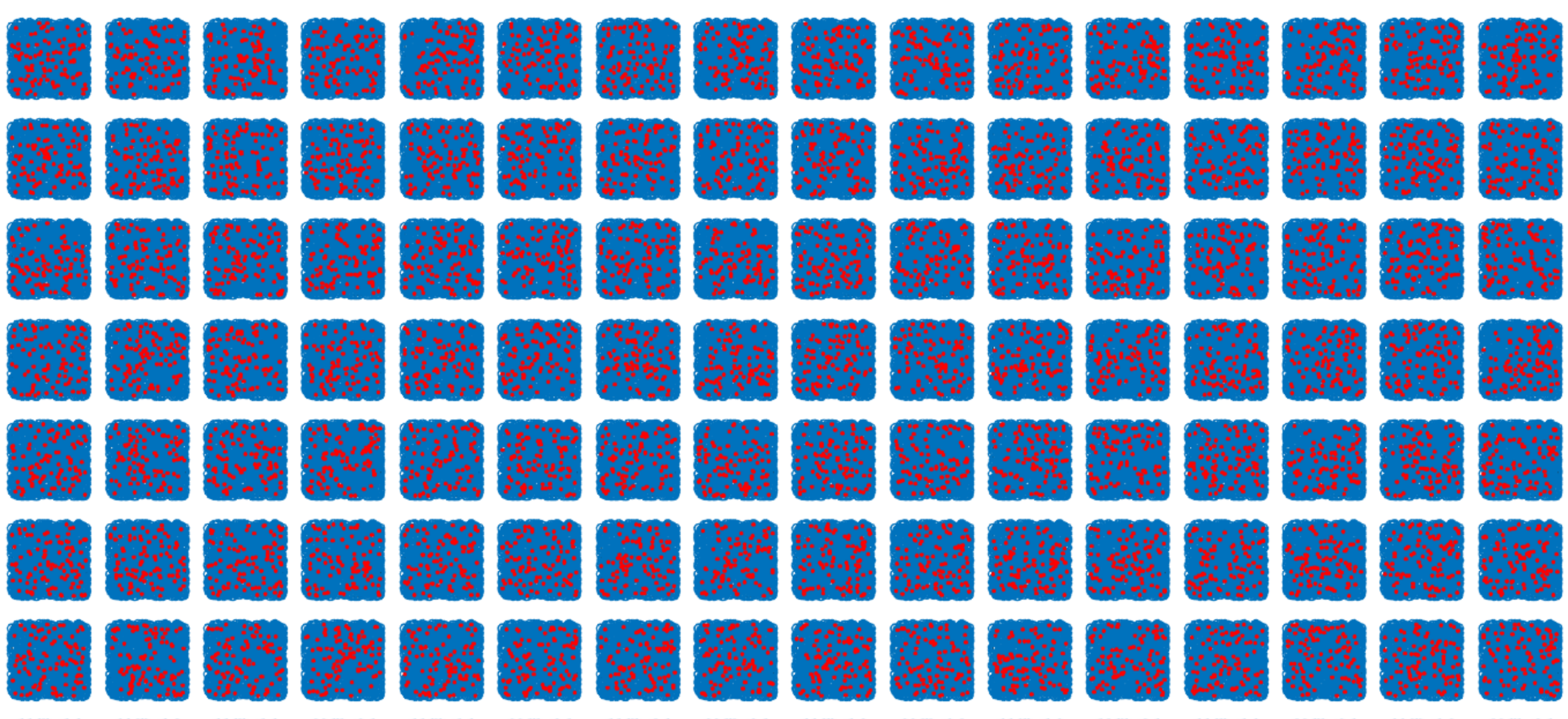
The Central Limit Theorem (6.4.2)

- The central limit theorem

Suppose that we obtain a random sample of n independently and identically distributed (i.i.d.) observations, $X_1, X_2, \ldots, X_n$, from a probability distribution with mean $\mathbb{E}(X)$ and variance $\mathbb{V}(X)$. Let us denote the sample average by $\overline{X}_n = \sum_{i=1}^{n} X_i / n$. Then, the **central limit theorem** states

$$\frac{\overline{X}_n - \mathbb{E}(X)}{\sqrt{\mathbb{V}(X)/n}} \rightsquigarrow \mathcal{N}(0, 1). \tag{6.41}$$

In the theorem, $\rightsquigarrow$ indicates "convergence in distribution" as the sample size n increases.



The Central Limit Theorem (6.4.2)

- The central limit theorem

Suppose that we obtain a random sample of n independently and identically distributed (i.i.d.) observations, $X_1, X_2, \ldots, X_n$, from a probability distribution with mean $\mathbb{E}(X)$ and variance $\mathbb{V}(X)$. Let us denote the sample average by $\overline{X}_n = \sum_{i=1}^{n} X_i/n$. Then, the **central limit theorem** states

$$\frac{\overline{X}_n - \mathbb{E}(X)}{\sqrt{\mathbb{V}(X)/n}} \rightsquigarrow \mathcal{N}(0, 1). \tag{6.41}$$

In the theorem, $\rightsquigarrow$ indicates "convergence in distribution" as the sample size n increases.

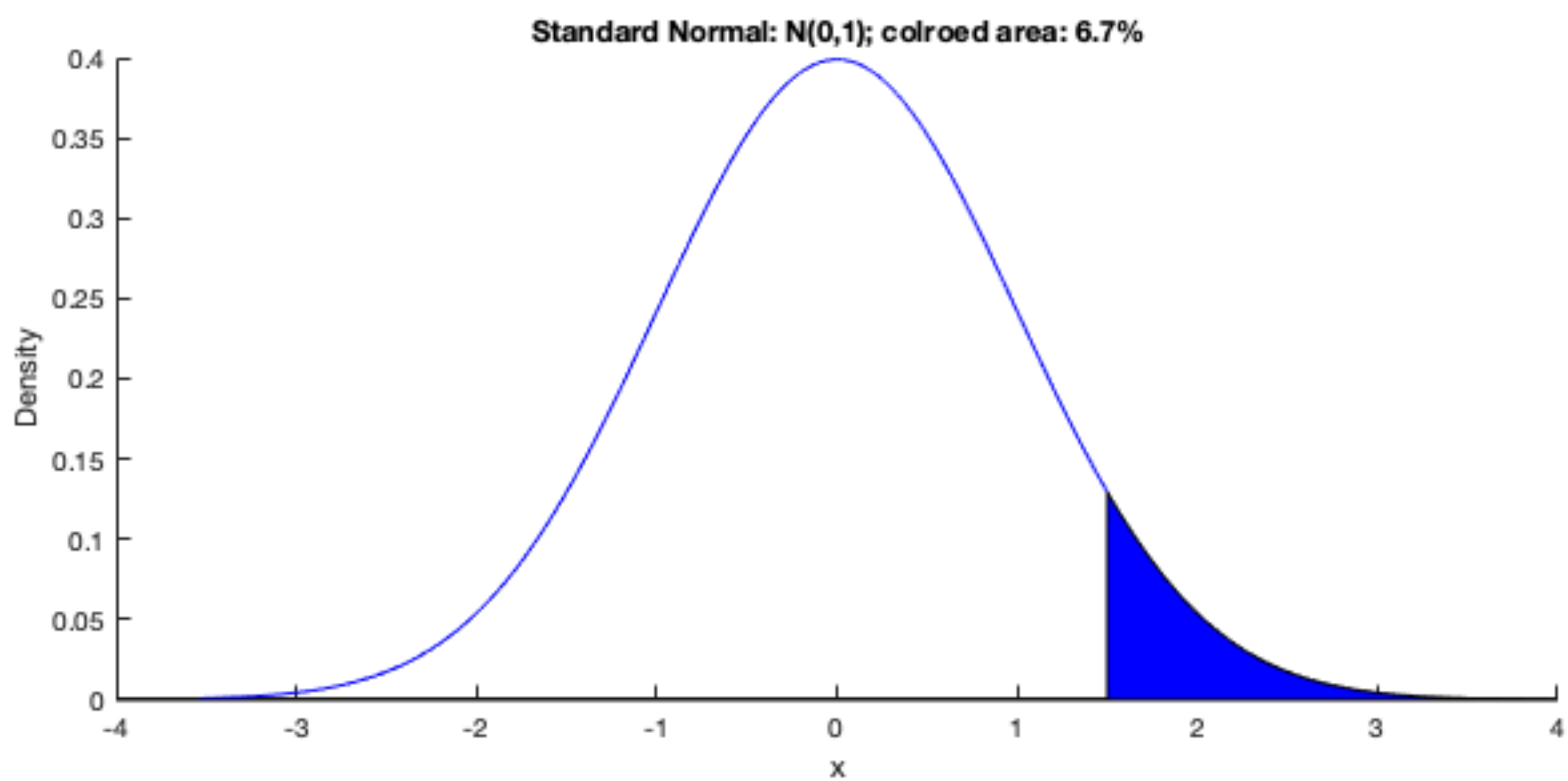
- z-score! Recall $X \sim \mathcal{N}(\mu, \sigma^2)$ **z-score:** $Z = (X - \mu)/\sigma \to Z \sim \mathcal{N}(0,1)$
  - Central limit theorem holds for ANY probability distribution

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

$$\mathbb{V}(\overline{X}_n) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) = \frac{1}{n}\mathbb{V}(X)$$

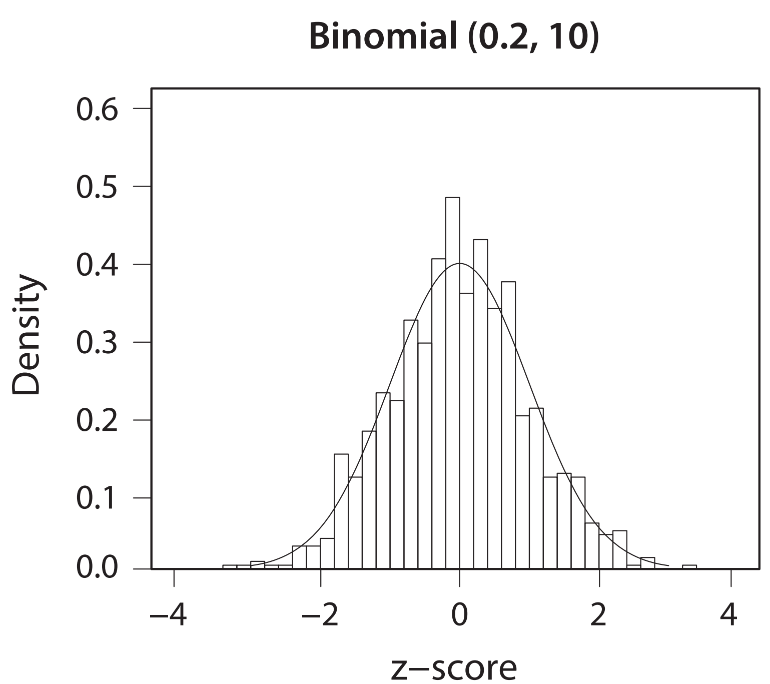
The Central Limit Theorem (6.4.2)

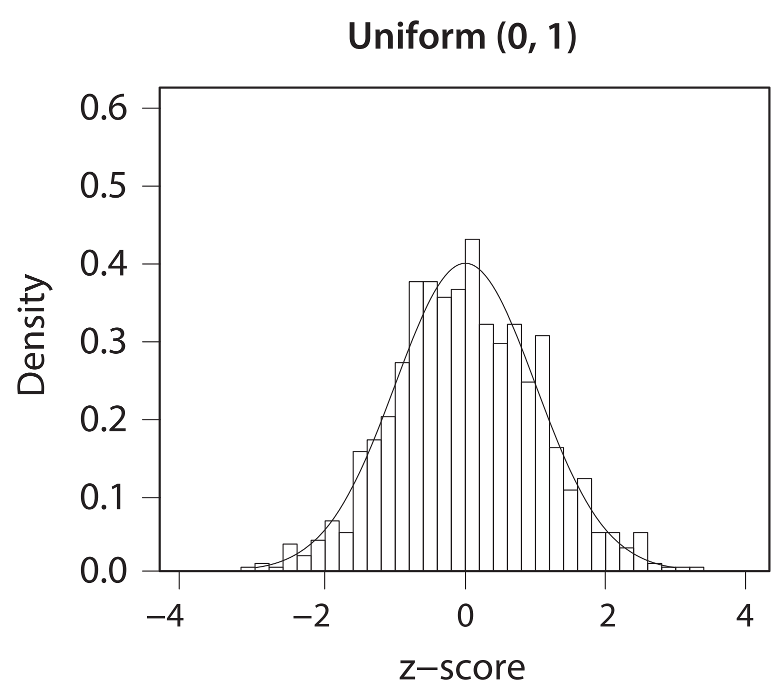
- (hyptothetical values) # Youtube videos: very large! Exceeding 10^{11}
  - Mean number of views: 40,000
  - Standard deviation for the number of views: 40,000
  - Is the distribution normal? No!
  - Question: If you select 100 videos randomly, what is the chance that their views exceeds 46,000?



The Central Limit Theorem (6.4.2)

```r
## sims = number of simulations
n.samp <- 1000
z.binom <- z.unif <- rep(NA, sims)
for (i in 1:sims) {
    x <- rbinom(n.samp, p = 0.2, size = 10)
    z.binom[i] <- (mean(x) - 2) / sqrt(1.6 / n.samp)
    x <- runif(n.samp, min = 0, max = 1)
    z.unif[i] <- (mean(x) - 0.5) / sqrt(1 / (12 * n.samp))
}
## histograms; nclass specifies the number of bins
hist(z.binom, freq = FALSE, nclass = 40, xlim = c(-4, 4), ylim = c(0, 0.6),
     xlab = "z-score", main = "Binomial(0.2, 10)")
x <- seq(from = -3, to = 3, by = 0.01)
lines(x, dnorm(x)) # overlay the standard normal PDF
hist(z.unif, freq = FALSE, nclass = 40, xlim = c(-4, 4), ylim = c(0, 0.6),
     xlab = "z-score", main = "Uniform(0, 1)")
lines(x, dnorm(x))
```

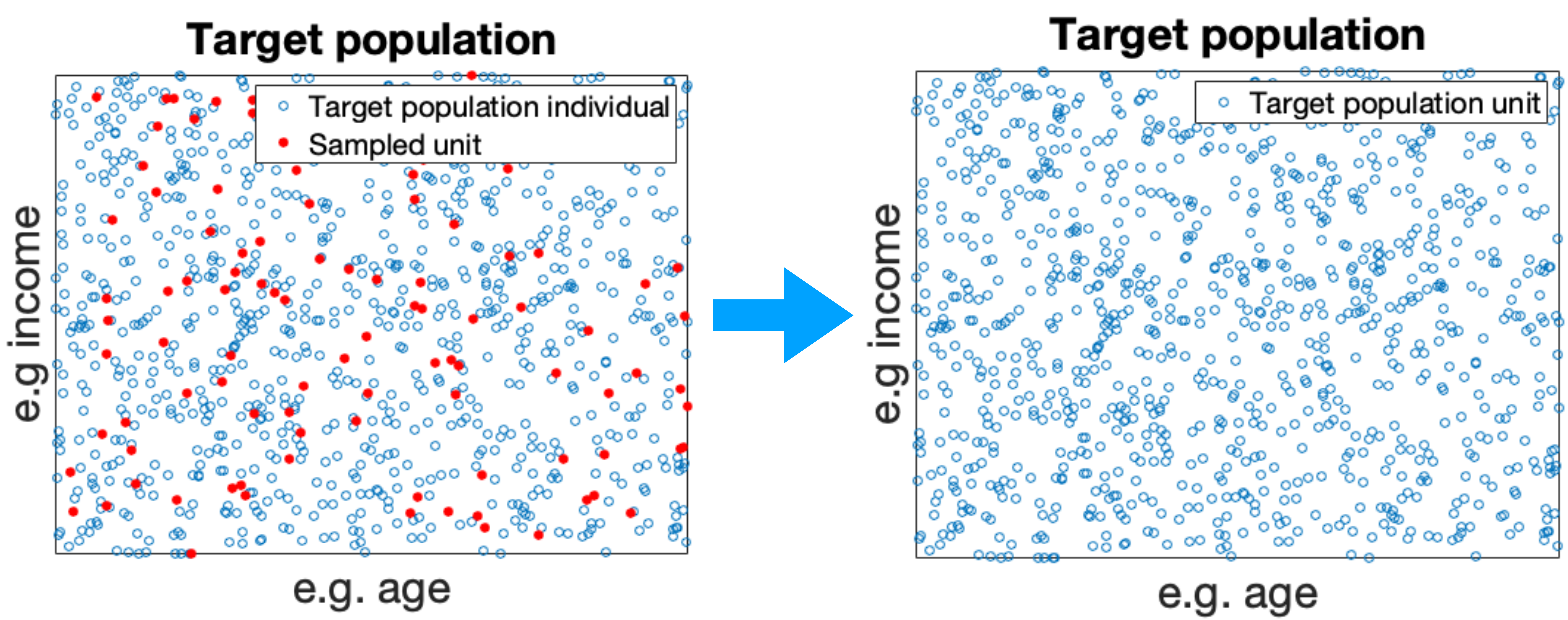




Contents

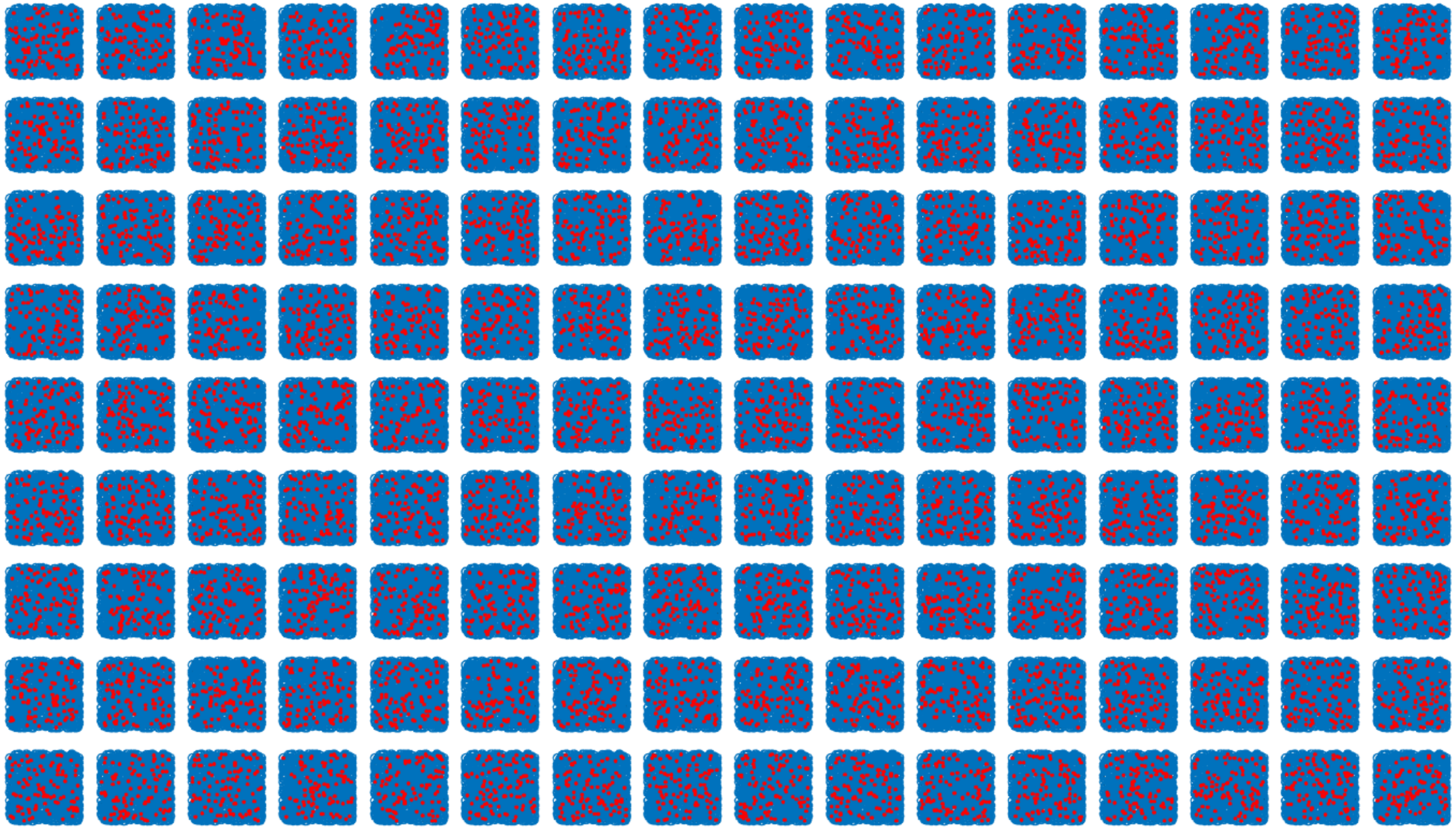
Unbiasedness

- Our ultimate quantities of interest are about populations
  - e.g. Average treatment effect for the entire population
  - e.g. Population fraction of candidate A supporters
- Challenge: In practice, we can only learn from samples
  - e.g. Sample average treatment effect estimates
  - e.g. Sample estimates of fraction of candidate A supporters



Unbiasedness

- What is formal language to describe quality of sample estimates??
  - We should relate sample estimates with population estimates
- An estimator is unbiased if its expectation equal the parameter.
  - e.g. sample average $\bar{X}_n : \mathbb{E}(\bar{X}_n) = \mathbb{E}(X)$
    - e.g. averaging an infinite number of survey samples of same size



Unbiasedness

- e.g. Sample Average Treatment Effect (SATE) for randomized control trials
  - $\text{SATE} = \frac{1}{n}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\}$
  - Difference in the means estimator
    - $$\begin{aligned}\widehat{\text{SATE}} &= \text{average of the treated} - \text{average of the untreated} \\ &= \frac{1}{n_1}\sum_{i=1}^{n} T_i Y_i - \frac{1}{n-n_1}\sum_{i=1}^{n}(1-T_i)Y_i.\end{aligned}$$
  - Unbiasedness
    - $\mathbb{E}(\widehat{\text{SATE}}) = \mathbb{E}\left(\frac{1}{n_1}\sum_{i=1}^{n} T_i Y_i(1) - \frac{1}{n-n_1}\sum_{i=1}^{n}(1-T_i)Y_i(0)\right)$

Summary

- Random Variables and Probability Distributions (Chapter 6.3)
  - Overview
  - Bernoulli / Binomial distribution
  - Uniform / Normal distribution
  - Expectation and Variance
- The Law of large numbers (6.4.1)
- Central limit theorem (6.4.2)
- Unbiasedness (7.1.1)
- Next week:
  - SATE consistency, Population ATE, bias and interval estimates

See you next week.