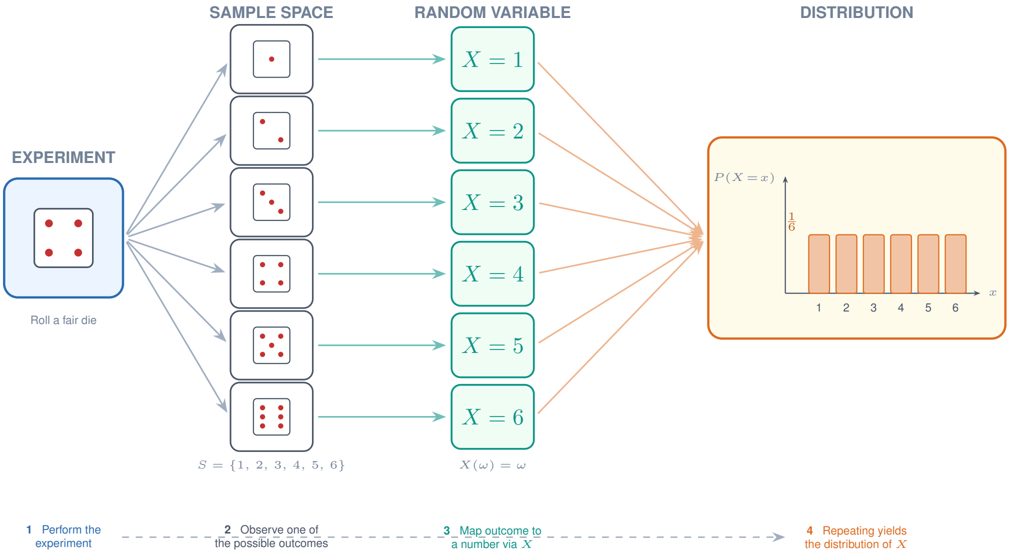
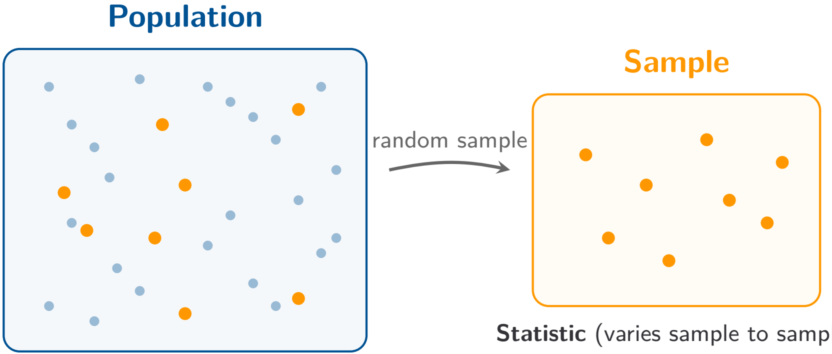
Chapter 15 – Recap

Sampling Distributions

Random Variables



Parameter vs. Statistic



Parameter (fixed, usually unknown)

$\mu, \ \sigma, \ p$

Statistic (varies sample to sample)

$\bar{x}, \ s, \ \hat{p}$

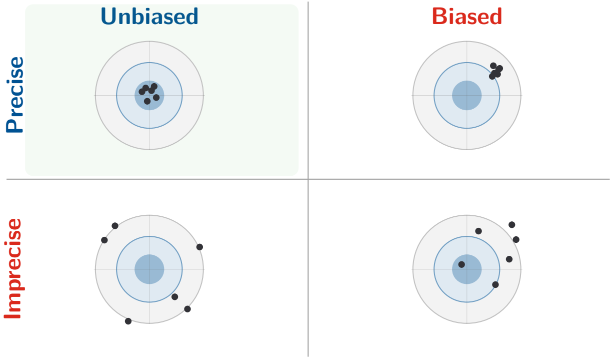
How do we study estimators?

Bias

An estimator is **unbiased** if it hits the target on average.

Precision

An estimator is **precise** if repeated estimates cluster tightly together.



What is a sampling distribution?

Sampling Distribution

The **sampling distribution** of a statistic is the distribution of that statistic over all possible random samples of size n.

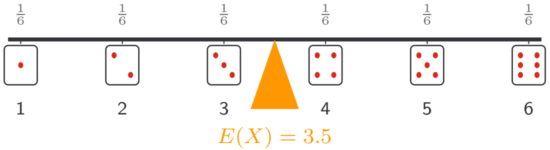
Expected Value

Expected Value

The **expected value** $E(X)$ is the long-run average of a random variable:

$$E(X) = \sum_x x \cdot P(X = x)$$

It is the **balance point** of the distribution.



Spread of a Random Variable

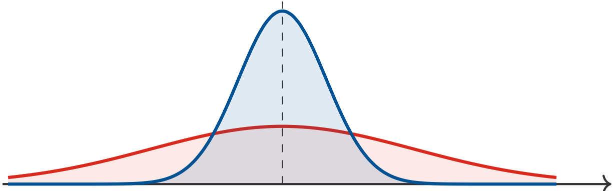
Spread

$Var(X) = SD(X)^2$ is the variance, which measures the expected squared distance from the mean:

$$Var(X) = E[(X - E(X))^2] = \sum_x (x - E(X))^2 \cdot P(X = x)$$

$SD(X)$ measures the typical distance of an outcome from the expected value:

$$SD(X) = \sqrt{\sum_x (x - E(X))^2 \cdot P(X = x)}$$

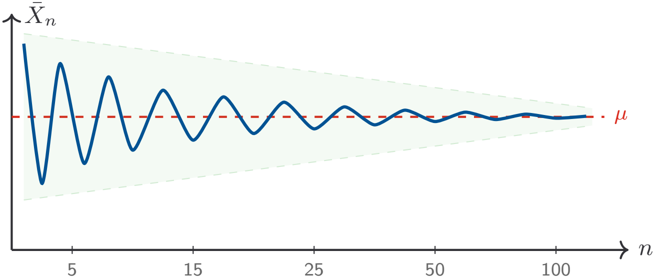


Law of Large Numbers

Law of Large Numbers

As the sample size n grows, the **sample mean** $\bar{X}_n$ converges to the **population mean** μ:

$$\bar{X}_n \xrightarrow{n\to\infty} \mu$$



Standard Error

Standard Error

The **standard error** of $\bar{X}_n$ is the standard deviation of the sampling distribution:

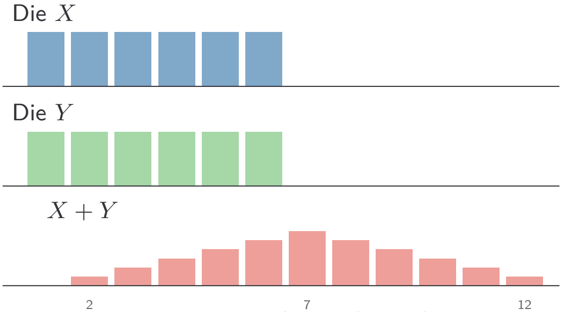
$$\mathrm{SE}(\bar{X}_n) = \mathrm{SD}(\bar{X}_n)$$

Adding Independent Random Variables

Variance Additivity

If X and Y are **independent**, then:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$



$$\mathrm{Var}(X+Y) = \underbrace{\frac{35}{12}}_{\text{Die } X} + \underbrace{\frac{35}{12}}_{\text{Die } Y} = \frac{35}{6}$$

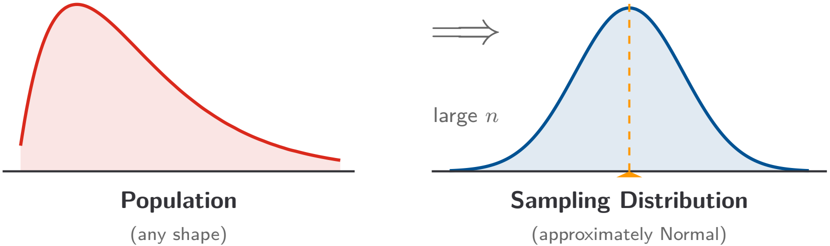
Sum of n Independent Random Variables

General Variance Rule

If $X_1, X_2, \ldots, X_n$ are **independent** and have the same distribution, then

$$\mathrm{Var}(X_1 + \cdots + X_n) = n\sigma^2$$

Central Limit Theorem



Principle: No matter what the population looks like, the sampling distribution of $\bar{X}_n$ becomes approximately $N\left(\mu, \dfrac{\sigma^2}{n}\right)$ for large n.