Seminar 5: Expectation and Variance of Discrete Random Variables

Carlos Buitrago & Anastasiya Chikalova

Table of contents

In previous lectures we introduced discrete random variables and their distributions. We now discuss the most important numerical characteristics of a random variable: its **expectation** and **variance**.

Roughly speaking, the expectation describes the center of a distribution, while the variance measures the spread of the values around that center.

1 Expectation of a Discrete Random Variable

Let ξ be a discrete random variable taking values

$$X_\xi = \{x_1, x_2, ...\}.$$

The **expectation** (or **mean**) of ξ is defined by

$$\mathbb{E}[\xi] = \sum_i x_i \, \mathbb{P}(\xi = x_i),$$

provided that the series $\sum_i |x_i| \, \mathbb{P}(\xi = x_i)$ converges.

1.1 Properties of Expectation

The expectation enjoys several important properties.

1.1.1 Linearity

For any constants $a, b \in \mathbb{R}$,

$$\mathbb{E}[a\xi + \eta b] = a \, \mathbb{E}[\xi] + b \, \mathbb{E}[\eta].$$

A remarkable fact is that **independence is not required** for this property.

1.1.2 Expectation of a Function

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a function. Then

$$\mathbb{E}[\varphi(\xi)] = \sum_i \varphi(x_i)\, \mathbb{P}(\xi = x_i).$$

This formula allows us to compute expectations of transformed random variables directly from the distribution of ξ.

1.1.3 Monotonicity

If $\xi \leq \eta$ almost surely, then $\mathbb{E}[\xi] \leq \mathbb{E}[\eta]$.

1.1.4 Absolute Value Inequality

For every random variable,

$$\big|\mathbb{E}[\xi]\big| \leq \mathbb{E}[|\xi|].$$

This inequality is often useful when estimating expectations.

1.1.5 Product of Independent Random Variables

If ξ and η are independent, then

$$\mathbb{E}[\xi\eta] = \mathbb{E}[\xi]\, \mathbb{E}[\eta].$$

This property generally fails if the random variables are not independent.

1.2 Simulations on Expectation

In this part of the seminar, we use simulations to understand expectation empirically. The goal is not only to compute formulas, but to see what these quantities mean when we generate data.

Throughout the notebook, we use the empirical mean

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

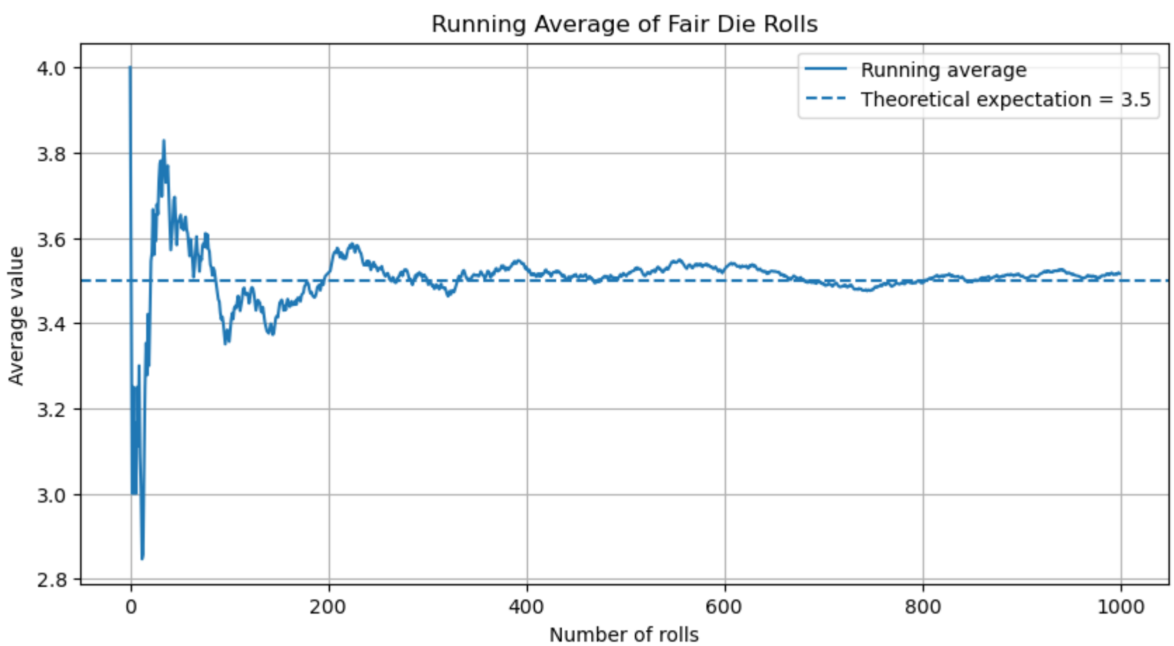
as an approximation to the theoretical expectation $\mathbb{E}[X]$.

1.2.1 Simulation 1: Expectation of a Fair Die

Let ξ be the result of rolling a fair die. Then ξ takes values $1, 2, 3, 4, 5, 6$ with equal probabilities. Therefore,

$$\mathbb{E}[\xi] = \frac{1+2+3+4+5+6}{6} = 3.5.$$

Notice that 3.5 is not a possible outcome of a die roll. This is an important point: the expectation is not necessarily the most likely value, nor even a possible value. It is a long-run average.

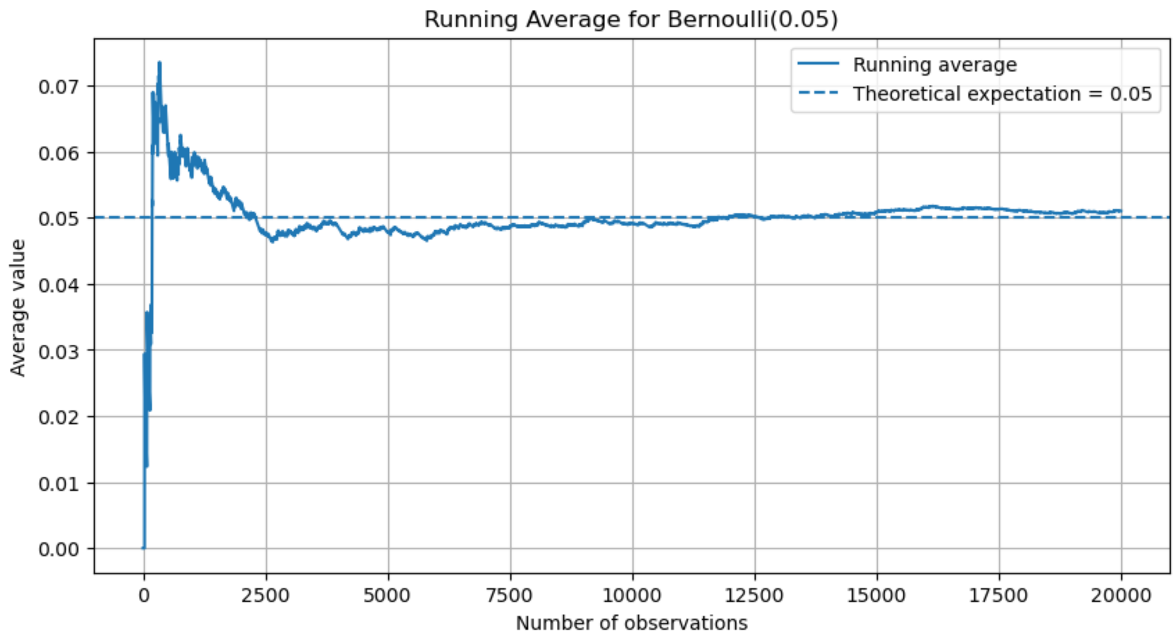


1.2.2 Simulation 2: Rare Event Bernoulli Random Variable

Let $\xi \sim \mathrm{Ber}(0.05)$. We may interpret $\xi = 1$ as a user clicking an advertisement and $\xi = 0$ as not clicking it. Then

$$\mathbb{E}[\xi] = 0.05.$$

Most observations are equal to 0, but the long-run average still approaches 0.05.




```
Empirical mean: 0.051
Theoretical mean: 0.05
```

2 Variance

While the expectation tells us where a distribution is centered, it does not tell us how spread out the values are.

The **variance** of a random variable ξ is defined by

$$\mathrm{Var}(\xi) = \mathbb{E}\left[(\xi - \mathbb{E}[\xi])^2\right].$$

The quantity

$$\sigma_\xi = \sqrt{\mathrm{Var}(\xi)}$$

is called the **standard deviation**.

2.1 Alternative Formula for Variance

Expanding the square gives

$$\mathrm{Var}(\xi) = \mathbb{E}[\xi^2] - \big(\mathbb{E}[\xi]\big)^2.$$

This formula is often much easier to use in computations.

2.2 Properties of Variance

1. $\mathrm{Var}(\xi) \geq 0$. Moreover, $\mathrm{Var}(\xi) = 0 \iff \mathbb{P}(\xi = c) = 1$, for some constant c.
2. For any constants $a, b \in \mathbb{R}$,

$$\mathrm{Var}(a\xi + b) = a^2\,\mathrm{Var}(\xi).$$

3. If ξ and η are independent, then

$$\mathrm{Var}(\xi + \eta) = \mathrm{Var}(\xi) + \mathrm{Var}(\eta).$$

3 Important Examples

The following expectations and variances are useful enough to remember.

Distribution

Expectation

Variance

$\mathrm{Ber}(p)$

p

$p(1-p)$

$\mathrm{Bin}(n,p)$

np

$np(1-p)$

Geom(p)

$\frac{1}{p}$

$\frac{1-p}{p^2}$

Pois(λ)

λ

λ

$U(\{1, \ldots, n\})$

$\frac{n+1}{2}$

$\frac{n^2-1}{12}$

3.1 Comparing Different Distributions

The following figure compares a Binomial, Poisson, and Geometric distribution.

For each distribution:

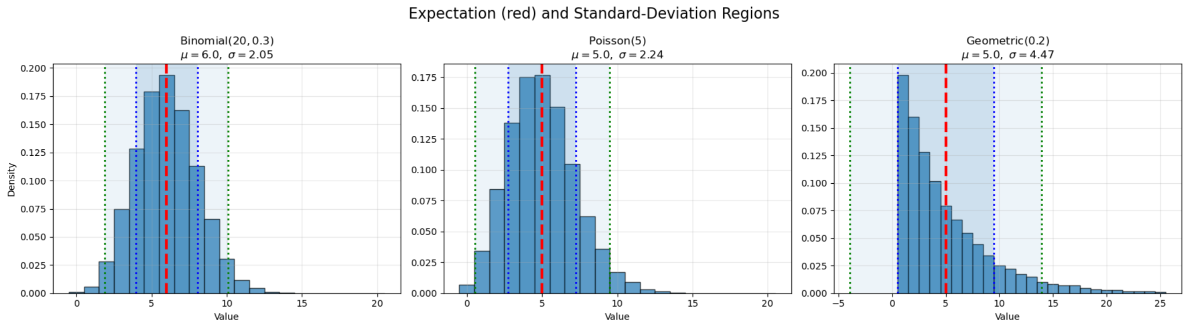
- the **red dashed line** shows the expectation μ;
- the **blue dotted lines** show $\mu \pm \sigma$;
- the **green dotted lines** show $\mu \pm 2\sigma$.

Although expectation and variance provide useful summaries of a distribution, they do not completely describe its shape. Notice how the three distributions differ in symmetry, spread, and skewness despite having well-defined means and variances.

Throughout the notebook, we use the empirical variance

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

as an approximation to the theoretical variance $\mathrm{Var}(X)$.



3.2 Example: Sample Mean Becomes Less Variable

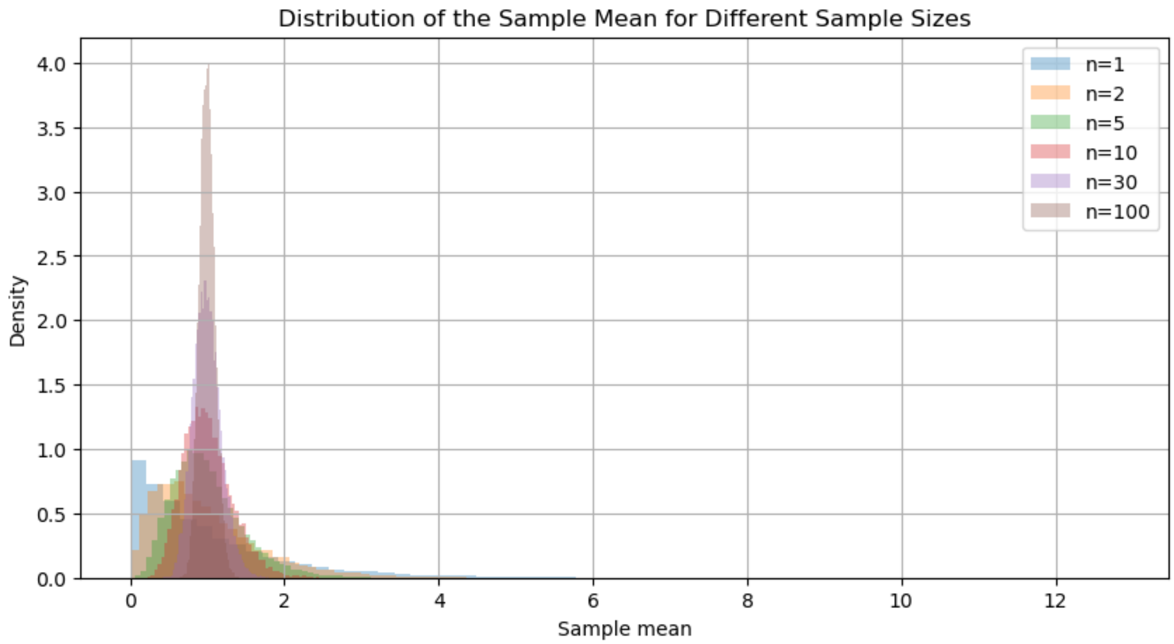
Let $X_1, \dots, X_n$ be independent observations with expectation μ and variance σ^2. The sample mean is

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Its variance is

$$\mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}.$$

Thus, as n grows, the sample mean becomes more stable.



3.3 Example: Apples and Oranges

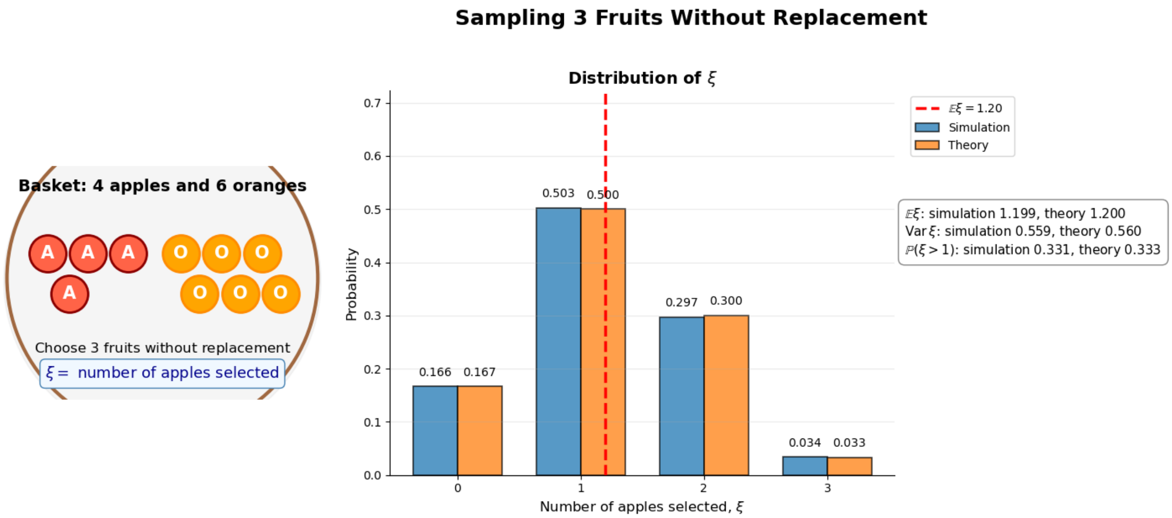
A basket contains 4 apples and 6 oranges. We randomly choose 3 fruits without replacement. Let ξ be the number of apples among the selected fruits. Let's compute the expectation and variance of ξ.

Then ξ can take the values $0, 1, 2, 3$. Since we are sampling without replacement, ξ has a hypergeometric distribution:

$$\xi \sim \text{Hypergeom}(N = 10, K = 4, n = 3),$$

where $N = 10$ is the total number of fruits, $K = 4$ is the number of apples, and $n = 3$ is the number of selected fruits.

In the simulation below, we repeat the experiment many times and compare the empirical distribution with the theoretical one.

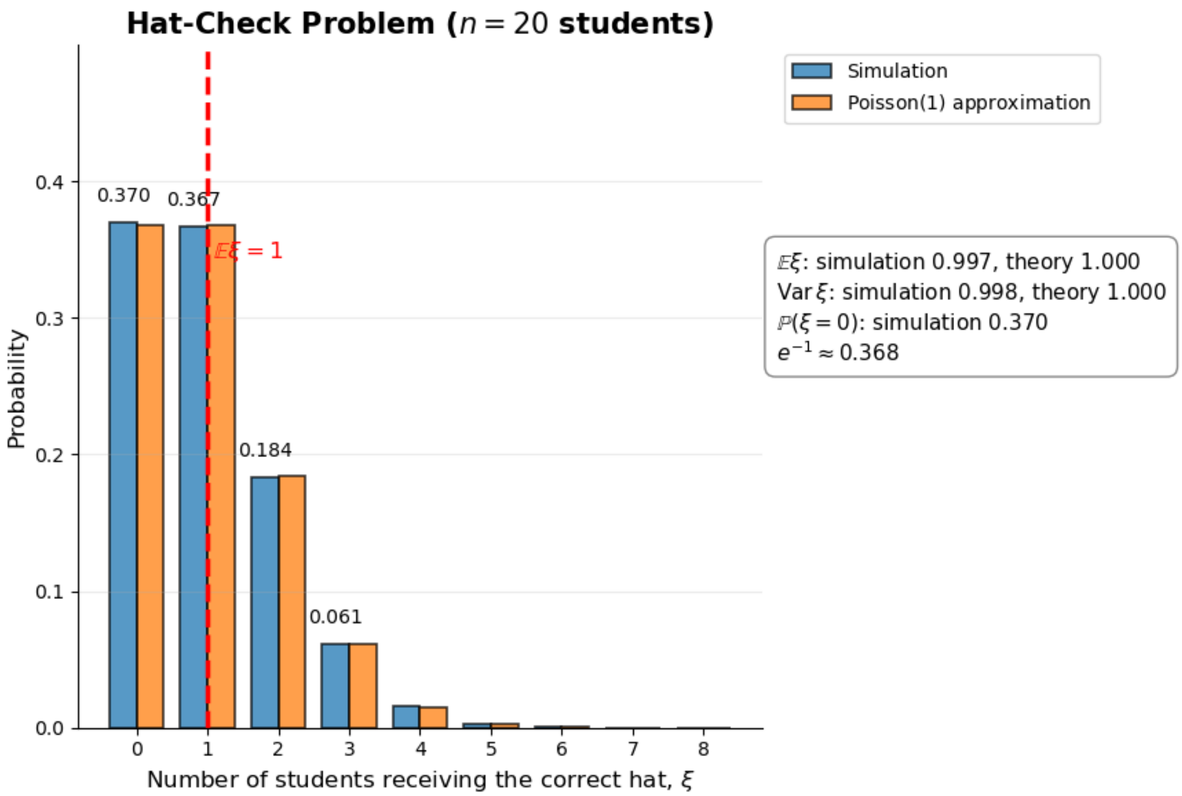



```
Theoretical probabilities:
P(xi = 0) = 0.1667
P(xi = 1) = 0.5000
P(xi = 2) = 0.3000
P(xi = 3) = 0.0333

Empirical mean         = 1.1985
Theoretical mean       = 1.2000

Empirical variance     = 0.5591
Theoretical variance   = 0.5600

Empirical P(xi > 1)    = 0.3311
Theoretical P(xi > 1) = 0.3333
```

3.4 Example: The Hat-Check Problem

A group of n students throw their hats at graduation. After that, each student randomly picks one hat from the floor. Let ξ be the number of students who get their own hat back.

In the simulation below, we randomly permute the hats many times and compare the empirical mean and variance with the theoretical values.




```
Number of students      = 20

Empirical mean          = 0.9968
Theoretical mean        = 1.0000

Empirical variance      = 0.9976
Theoretical variance    = 1.0000

Empirical P(xi = 0)     = 0.3695
Poisson approximation   = 0.3679
```

3.5 Example: A Casino Game with Positive Expectation

Consider the following game. In one round, a player wins \$10 with probability 0.1 and loses \$1 with probability 0.9. Let X be the profit from one round.

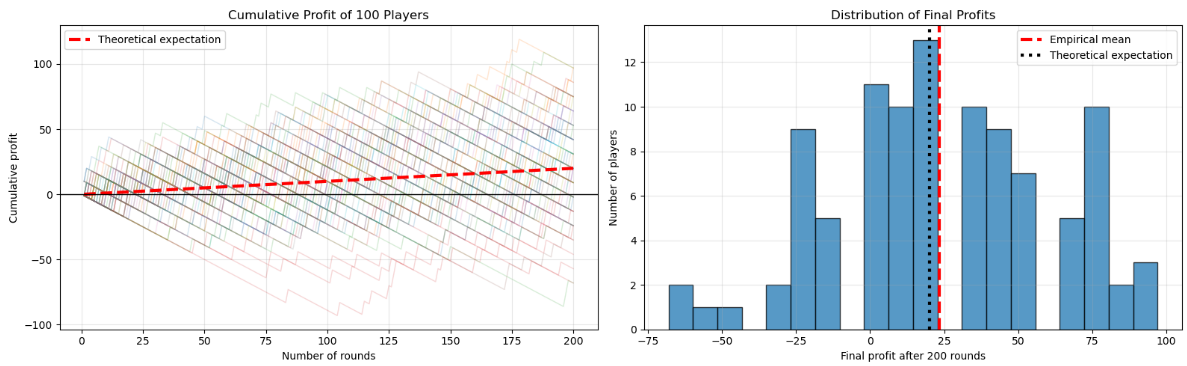
Then

$$X = \begin{cases} 10, & \text{with probability } 0.1, \\ -1, & \text{with probability } 0.9. \end{cases}$$

The expected profit is

$$\mathbb{E}[X] = 10 \cdot 0.1 + (-1) \cdot 0.9 = 0.1.$$

So the game has positive expectation. However, this does not mean that a player is guaranteed to win. In the simulation below, we see that even a favorable game can produce losing streaks and negative outcomes over short periods.




```
Expected profit per round       = 0.10
Expected profit after 200 rounds = 20.00

Empirical mean final profit     = 23.30
Empirical variance              = 1397.55

Fraction of players who lost money = 0.310
```

3.6 Why Do We Care About Expectation and Variance?

At first glance, expectation and variance may seem like purely mathematical quantities. However, they appear in almost every area of statistics, data science, machine learning, finance, engineering, and scientific research.

The expectation describes what we should expect **on average** if an experiment is repeated many times, while the variance measures how much uncertainty or variability is present around that average.

In practice, good decisions often require understanding **both** quantities.

4 Discrete Conditional Expectation

4.1 Conditional Distribution

Let X and Y be discrete random variables. For every value y satisfying $\mathbb{P}(Y = y) > 0$, the conditional probability mass function of X is

$$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}.$$

For each fixed value of y, these probabilities define a valid probability distribution. Indeed,

$$\sum_x \mathbb{P}(X = x \mid Y = y) = 1.$$

Therefore, once $Y = y$ is observed, we may compute probabilities and expectations exactly as we do for any ordinary discrete distribution.

4.2 Definition

The **conditional expectation** of X given $Y = y$ is defined by

$$\boxed{\mathbb{E}[X \mid Y = y] = \sum_x x\, \mathbb{P}(X = x \mid Y = y).}$$

4.2.1 Conditional Expectation is a Random Variable

Define the function

$$g(y) = \mathbb{E}[X \mid Y = y].$$

Since this is a function of y, we may replace y by the random variable Y itself.

This gives

$$\boxed{\mathbb{E}[X \mid Y] = g(Y).}$$

Therefore, conditional expectation is itself a random variable.

4.2.2 Interpretation

Before observing Y, our best prediction of X is $\mathbb{E}[X]$.

After observing Y, we have additional information and our prediction becomes

$$\mathbb{E}[X \mid Y].$$

The conditional expectation updates our prediction using the information contained in Y.