Random Variables and Expectations

Vladimir Podolskii

Computer Science Department, Higher School of Economics

Outline

Random Variables

Average

Expectation

Random Variables

- We have studied probability distributions

Random Variables

- We have studied probability distributions
- Studied events (subsets of outcomes) and their probabilities

Random Variables

- We have studied probability distributions
- Studied events (subsets of outcomes) and their probabilities
- Events correspond to yes or no questions

Random Variables

- We have studied probability distributions
- Studied events (subsets of outcomes) and their probabilities
- Events correspond to yes or no questions
- It is important to study numerical characteristics of random outcomes

Random Variables

- We have studied probability distributions
- Studied events (subsets of outcomes) and their probabilities
- Events correspond to yes or no questions
- It is important to study numerical characteristics of random outcomes
- So we introduce random variables

Random Variables

- Random variable f is a variable whose value is determined by a random experiment

Random Variables

- Random variable f is a variable whose value is determined by a random experiment
- We have a probability distribution on the finite set X of k outcomes

Random Variables

- Random variable f is a variable whose value is determined by a random experiment
- We have a probability distribution on the finite set X of k outcomes
- Outcomes have probabilites $p_1, \ldots, p_k$

Random Variables

- Random variable f is a variable whose value is determined by a random experiment
- We have a probability distribution on the finite set X of k outcomes
- Outcomes have probabilites $p_1, \ldots, p_k$
- To define f we assign a number a_i to each outcome

Random Variables

- Random variable f is a variable whose value is determined by a random experiment
- We have a probability distribution on the finite set X of k outcomes
- Outcomes have probabilites $p_1, \ldots, p_k$
- To define f we assign a number a_i to each outcome
- Then f has value a_i with probability p_i

Random Variables

- Looks familiar

Random Variables

- Looks familiar
- We have already done this!

Random Variables

- Looks familiar
- We have already done this!
- Outcomes of the dice throw are labeled by numbers

wikimedia.org

Random Variables

Other examples:

Random Variables

Other examples:

- Tossing a coin: heads=0, tail=1

Random Variables

Other examples:

- Tossing a coin: heads=0, tail=1
- An age of a random person in the class

Random Variables

Other examples:

- Tossing a coin: heads=0, tail=1
- An age of a random person in the class
- Grade of a random person in the class

Random Variables

Other examples:

- Tossing a coin: heads=0, tail=1
- An age of a random person in the class
- Grade of a random person in the class
- Sum of outcomes of two dice throws

Outline

Random Variables

Average

Expectation

Average

- What is an average salary in a country?

Average

- What is an average salary in a country?
- The total salary of all population divided by the number of employees

Average

- What is an average salary in a country?
- The total salary of all population divided by the number of employees
- This is the standard notion of average

Average

- What is an average salary in a country?
- The total salary of all population divided by the number of employees
- This is the standard notion of average
- It is called arithmetic mean in mathematics

Example

Problem

Student got scores 78, 72 and 87 on three tests. What is his average score?

Example

Problem

Student got scores 78, 72 and 87 on three tests. What is his average score?

- Have to add all scores and divide by their number

Example

Problem

Student got scores 78, 72 and 87 on three tests. What is his average score?

- Have to add all scores and divide by their number
- Here it is:

$$\frac{78 + 72 + 87}{3} = \frac{237}{3} = 79$$

Example

Problem

Student got scores 78, 72 and 87 on three tests. What is his average score?

- Have to add all scores and divide by their number
- Here it is:

$$\frac{78 + 72 + 87}{3} = \frac{237}{3} = 79$$

- We got lucky and the answer is integer; this is not guaranteed

HR Management Strategy

Problem

Suppose HR management in some company uses the following strategy: fire everyone who performs below average. What will be a result of such strategy?

HR Management Strategy

Problem

Suppose HR management in some company uses the following strategy: fire everyone who performs below average. What will be a result of such strategy?

- Might sound reasonable, but ...

HR Management Strategy

Problem

Suppose HR management in some company uses the following strategy: fire everyone who performs below average. What will be a result of such strategy?

- Might sound reasonable, but ...
- Unless everyone works equally (extremely rare case) ...

HR Management Strategy

Problem

Suppose HR management in some company uses the following strategy: fire everyone who performs below average. What will be a result of such strategy?

- Might sound reasonable, but ...
- Unless everyone works equally (extremely rare case) ...
- There is always someone who works below average!

HR Management Strategy

Problem

Suppose HR management in some company uses the following strategy: fire everyone who performs below average. What will be a result of such strategy?

- Might sound reasonable, but ...
- Unless everyone works equally (extremely rare case) ...
- There is always someone who works below average!
- If we fire them, the average performance will grow

HR Management Strategy

Problem

Suppose HR management in some company uses the following strategy: fire everyone who performs below average. What will be a result of such strategy?

- Might sound reasonable, but ...
- Unless everyone works equally (extremely rare case) ...
- There is always someone who works below average!
- If we fire them, the average performance will grow
- New people get below average!

HR Management Strategy

Problem

Suppose HR management in some company uses the following strategy: fire everyone who performs below average. What will be a result of such strategy?

- Might sound reasonable, but ...
- Unless everyone works equally (extremely rare case) ...
- There is always someone who works below average!
- If we fire them, the average performance will grow
- New people get below average!
- We will fire everyone except one best employee

Average Outcome of Dice Throw

Problem

Suppose we throw a dice many times. What is the average outcome?

Average Outcome of Dice Throw

Problem

Suppose we throw a dice many times. What is the average outcome?

- Can we give a precise answer?

Average Outcome of Dice Throw

Problem

Suppose we throw a dice many times. What is the average outcome?

- Can we give a precise answer?
- No, it is a random variable

Average Outcome of Dice Throw

Problem

Suppose we throw a dice many times. What is the average outcome?

- Can we give a precise answer?
- No, it is a random variable
- But we can give an approximation that is good with high probability

Average Outcome of Dice Throw

- Suppose we throw a dice n times for a very large n

Average Outcome of Dice Throw

- Suppose we throw a dice n times for a very large n
- Then among outcomes there are approximately $n/6$ ones, $n/6$ twos, and so on

Average Outcome of Dice Throw

- Suppose we throw a dice n times for a very large n
- Then among outcomes there are approximately $n/6$ ones, $n/6$ twos, and so on
- The sum of results is then approximately

$$\frac{n}{6} \times 1 + \frac{n}{6} \times 2 + \frac{n}{6} \times 3 + \frac{n}{6} \times 4 + \frac{n}{6} \times 5 + \frac{n}{6} \times 6$$

$$= \frac{n(1+2+3+4+5+6)}{6} = \frac{21n}{6} = 3.5n$$

Average Outcome of Dice Throw

- Suppose we throw a dice n times for a very large n
- Then among outcomes there are approximately $n/6$ ones, $n/6$ twos, and so on
- The sum of results is then approximately
$$\frac{n}{6} \times 1 + \frac{n}{6} \times 2 + \frac{n}{6} \times 3 + \frac{n}{6} \times 4 + \frac{n}{6} \times 5 + \frac{n}{6} \times 6$$
$$= \frac{n(1+2+3+4+5+6)}{6} = \frac{21n}{6} = 3.5n$$
- The average can be obtained by division by the number of throws n

Average Outcome of Dice Throw

- Suppose we throw a dice n times for a very large n
- Then among outcomes there are approximately $n/6$ ones, $n/6$ twos, and so on
- The sum of results is then approximately
$$\frac{n}{6} \times 1 + \frac{n}{6} \times 2 + \frac{n}{6} \times 3 + \frac{n}{6} \times 4 + \frac{n}{6} \times 5 + \frac{n}{6} \times 6$$
$$= \frac{n(1+2+3+4+5+6)}{6} = \frac{21n}{6} = 3.5n$$
- The average can be obtained by division by the number of throws n
- Thus the average is approximately 3.5

Average Outcome of Dice Throw

- Suppose we throw a dice n times for a very large n
- Then among outcomes there are approximately $n/6$ ones, $n/6$ twos, and so on
- The sum of results is then approximately
$$\frac{n}{6} \times 1 + \frac{n}{6} \times 2 + \frac{n}{6} \times 3 + \frac{n}{6} \times 4 + \frac{n}{6} \times 5 + \frac{n}{6} \times 6$$
$$= \frac{n(1+2+3+4+5+6)}{6} = \frac{21n}{6} = 3.5n$$
- The average can be obtained by division by the number of throws n
- Thus the average is approximately 3.5
- This is an expected value or expectation of a dice throw

Outline

Random Variables

Average

Expectation

Expectation

- Let’s consider the general case

Expectation

- Let's consider the general case
- Suppose we have a random variable f on the distribution with 4 outcomes

Expectation

- Let's consider the general case
- Suppose we have a random variable f on the distribution with 4 outcomes
- Probabilities of outcomes are p_1, p_2, p_3, p_4

Expectation

- Let's consider the general case
- Suppose we have a random variable f on the distribution with 4 outcomes
- Probabilities of outcomes are p_1, p_2, p_3, p_4
- Values of f are a_1, a_2, a_3, a_4

Expectation

- Let's consider the general case
- Suppose we have a random variable f on the distribution with 4 outcomes
- Probabilities of outcomes are p_1, p_2, p_3, p_4
- Values of f are a_1, a_2, a_3, a_4
- Let's repeat the random experiment many times

Expectation

$$\underbrace{\qquad\quad}_{p_1} \qquad \underbrace{\qquad\quad}_{p_2} \qquad \underbrace{\qquad\quad}_{p_3} \qquad \underbrace{\qquad\quad}_{p_4}$$

Expectation

Expectation

Expectation

Expectation

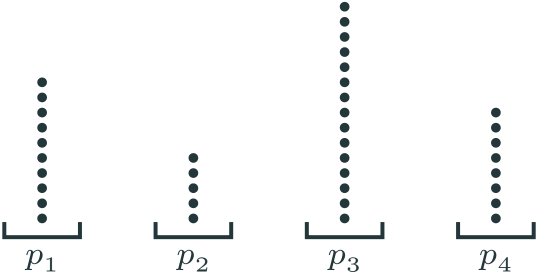
Expectation

Expectation

Expectation

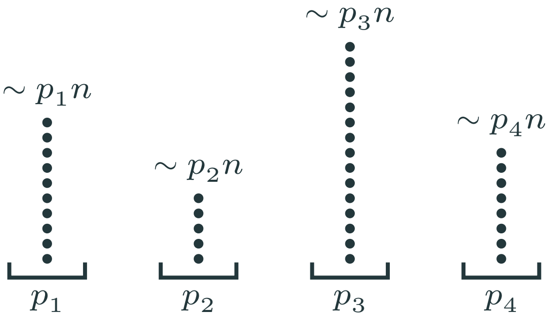
Expectation

Expectation



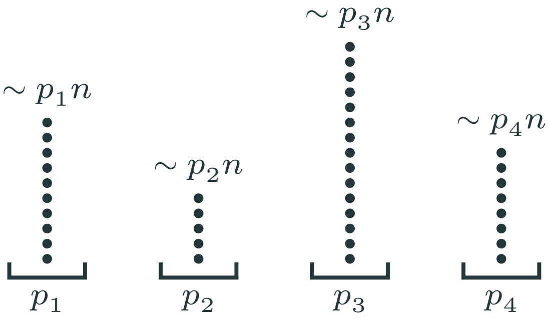
- Repeat n times, where n is large

Expectation



- Repeat n times, where n is large

Expectation



- Repeat n times, where n is large
- What is the average value of f on these outcomes?

Expectation

- We have n experiments

Expectation

- We have n experiments
- Value a_i occurs about $p_i n$ times

Expectation

- We have n experiments
- Value a_i occurs about $p_i n$ times
- On average we have

$$\sim \frac{a_1 p_1 n + a_2 p_2 n + a_3 p_3 n + a_4 p_4 n}{n}$$

$$= a_1 p_1 + a_2 p_2 + a_3 p_3 + a_4 p_4$$

Expectation

- We have n experiments
- Value a_i occurs about $p_i n$ times
- On average we have

$$\sim \frac{a_1 p_1 n + a_2 p_2 n + a_3 p_3 n + a_4 p_4 n}{n}$$

$$= a_1 p_1 + a_2 p_2 + a_3 p_3 + a_4 p_4$$

- This is denoted by $\mathsf{E} f$ and called the expectation of f

Expectation

- We have n experiments
- Value a_i occurs about $p_i n$ times
- On average we have

$$\sim \frac{a_1 p_1 n + a_2 p_2 n + a_3 p_3 n + a_4 p_4 n}{n}$$

$$= a_1 p_1 + a_2 p_2 + a_3 p_3 + a_4 p_4$$

- This is denoted by $\mathrm{E} f$ and called the expectation of f
- Does not depend on n

Expectation

- We have n experiments
- Value a_i occurs about $p_i n$ times
- On average we have

$$\sim \frac{a_1 p_1 n + a_2 p_2 n + a_3 p_3 n + a_4 p_4 n}{n}$$

$$= a_1 p_1 + a_2 p_2 + a_3 p_3 + a_4 p_4$$

- This is denoted by $\mathrm{E} f$ and called the expectation of f
- Does not depend on n
- An approximation to what we would expect as an average outcome of an experiment repeated many times

Expectation

- In the general case, outcomes $a_1, \ldots, a_k$ with probabilities $p_1, \ldots, p_k$

Expectation

- In the general case, outcomes $a_1, \dots, a_k$ with probabilities $p_1, \dots, p_k$
- To compute the expectation multiply $a_i \times p_i$ over all i

Expectation

- In the general case, outcomes $a_1, \ldots, a_k$ with probabilities $p_1, \ldots, p_k$
- To compute the expectation multiply $a_i \times p_i$ over all i
- And add up through 1 to k

Expectation

- In the general case, outcomes $a_1, \ldots, a_k$ with probabilities $p_1, \ldots, p_k$
- To compute the expectation multiply $a_i \times p_i$ over all i
- And add up through 1 to k
- Expectation is a number!

Expectation

- In the general case, outcomes $a_1, \ldots, a_k$ with probabilities $p_1, \ldots, p_k$
- To compute the expectation multiply $a_i \times p_i$ over all i
- And add up through 1 to k
- Expectation is a number!
- An important characteristic of a random variable

Geometric Interpretation of Expectation

Suppose f obtains values a_1, a_2, a_3, a_4 with probabilities p_1, p_2, p_3, p_4

Geometric Interpretation of Expectation

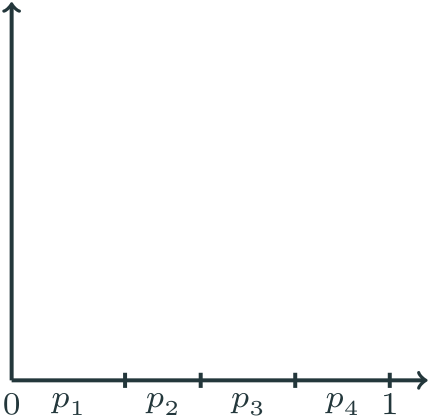
Suppose f obtains values a_1, a_2, a_3, a_4 with probabilities p_1, p_2, p_3, p_4

$\mathsf{E}f = a_1p_1 + a_2p_2 + a_3p_3 + a_4p_4$

Geometric Interpretation of Expectation

Suppose f obtains values a_1, a_2, a_3, a_4 with probabilities p_1, p_2, p_3, p_4

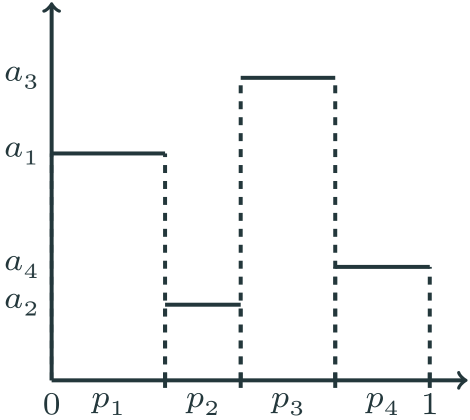
$\mathsf{E}f = a_1 p_1 + a_2 p_2 + a_3 p_3 + a_4 p_4$



Geometric Interpretation of Expectation

Suppose f obtains values a_1, a_2, a_3, a_4 with probabilities p_1, p_2, p_3, p_4

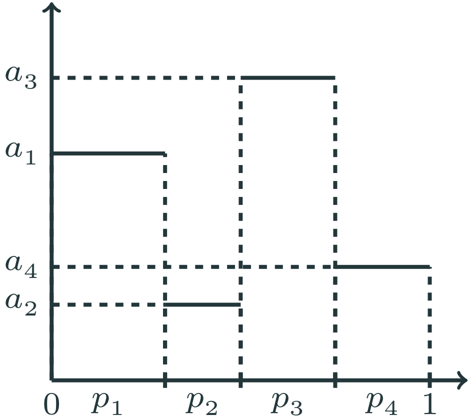
$\mathsf{E}f = a_1 p_1 + a_2 p_2 + a_3 p_3 + a_4 p_4$



Geometric Interpretation of Expectation

Suppose f obtains values a_1, a_2, a_3, a_4 with probabilities p_1, p_2, p_3, p_4

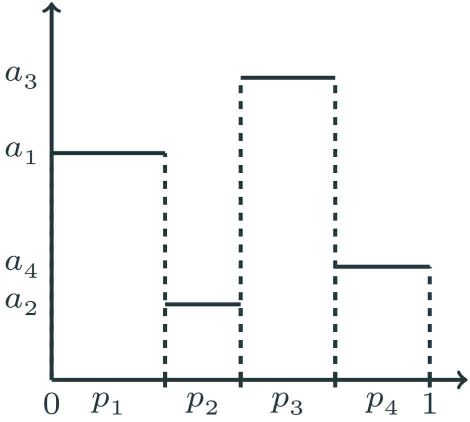
$\mathsf{E}f = a_1p_1 + a_2p_2 + a_3p_3 + a_4p_4$



Geometric Interpretation of Expectation

Suppose f obtains values a_1, a_2, a_3, a_4 with probabilities p_1, p_2, p_3, p_4

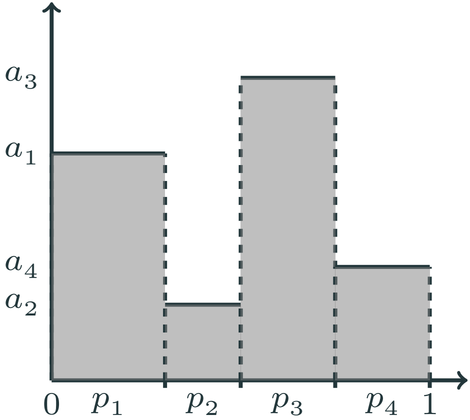
$\mathsf{E}f = a_1p_1 + a_2p_2 + a_3p_3 + a_4p_4$



Geometric Interpretation of Expectation

Suppose f obtains values a_1, a_2, a_3, a_4 with probabilities p_1, p_2, p_3, p_4

$\mathsf{E}f = a_1 p_1 + a_2 p_2 + a_3 p_3 + a_4 p_4$



- $\mathsf{E}f$ is the area of the gray region

Real Life Examples

- Expectations occur everywhere in statistics and sociology

Real Life Examples

- Expectations occur everywhere in statistics and sociology
- Average age

Real Life Examples

- Expectations occur everywhere in statistics and sociology
- Average age
- Life expectancy

Real Life Examples

- Expectations occur everywhere in statistics and sociology
- Average age
- Life expectancy
- Average grades and evaluations