Probability and Statistics course

Introduction

Gabor Vigh

TTK Department of Probability Theory and Statistics
ELTE

September 14, 2025

Rules

- There will be a midterm worth of 50 points on October 20 and a exam worth 50 points on December 8 during the class. . Thresholds for the grades:
  - 0 – 34 : 1
  - 35 – 52 : 2
  - 53 – 65: 3
  - 66 – 78: 4
  - 79 – 100: 5.
- Both exams and mid-term are mandatory. To pass them, you need to obtain at least 30% from both of them. If you do not manage, you need to participate in the retake. There is only one possibility for a retake. The only cases where you are allowed not to participate in the midterm/exam is when you are ill, and then I expect you to bring me a proper document from the doctor confirming that you were not able to appear at the faculty due to illness, or when you have another exam at the same time, and then you need to bring a confirmation from another professor.

Rules continued

- If you bring me such a document, I will organize 2 retakes for you if needed (so that everyone has 2 attempts).
- In case you cannot appear on the midterm/exam and you fail to bring the proper documentation justifying your absence, I will organize the so-called "utovizsga" at the end of the semester for you, which is a big exam consisting of the material from the whole semester.
- class is from 8:00 to 9:30 every Monday

Agenda for the semester

1. Combinatorial Analysis & Axioms of Probability (8th of Sep)
2. Conditional Probability and Independence (15th of Sep)
3. Random Variables (22nd of Sep)
4. Discrete distributions(29th of Sep)
5. Continuous Random Variables & Jointly Distributed Random Variables (6th of Oct)
6. Properties of Expectation & Limit Theorems (13th of Oct)
7. Midterm (20th of Oct)
8. (Autumn holiday: 27th of Oct - 1st of Nov)
9. Descriptive statistics (3rd of Nov)
10. The basics of statistics: estimation (maximum likelihood, moments) (10th of Nov)
11. Confidence intervals (17th of Nov)
12. Parametric and non-parametric probes (24th of Nov)
13. Linear regression (1st of Dec)
14. Final Exam (8th of Dec)

Materials

Week 1

The generalized basic principle of counting

The generalized basic principle of counting

If r experiments that are to be performed are such that the first one may result in any of n_1 possible outcomes; and if, for each of these n_1 possible outcomes, there are n_2 possible outcomes of the second experiment; and if, for each of the possible outcomes of the first two experiments, there are n_3 possible outcomes of the third experiment; and if ..., then there is a total of $n_1 \cdot n_2 \cdot \cdots \cdot n_r$ possible outcomes of the r experiments.

Examples

A small community consists of 10 women, each of whom has 3 children. If one woman and one of her children are to be chosen as mother and child of the year, how many different choices are possible?

Permutations

How many different ordered arrangements of letters a, b, and c are possible?

Permutations

How many different ordered arrangements of letters a, b, and c are possible?

Definition (Permutations without repetition)

An arrangement of n (distinct) elements is $n!$.

How many different letter arrangements can be formed from the PEPPER letters?

Permutations

How many different ordered arrangements of letters a, b, and c are possible?

Definition (Permutations without repetition)

An arrangement of n (distinct) elements is $n!$.

How many different letter arrangements can be formed from the PEPPER letters?

Definition (Permutations with repetition)

The number of possible arrangements of n elements where identical elements are not distinguished (assuming that $k_1, ..., k_r$ identical elements of one type) is calculated using the following formula:

$$P_n^{(k_1,k_2,\ldots,k_r)} = \frac{n!}{k_1! \cdot k_2! \cdot \cdots \cdot k_r!} = \binom{n}{k_1, ..., k_r}$$

where $k_1 + k_2 + \cdots + k_r = n$.

Combinations

How many different groups of 3 could be selected from the 5 items A, B, C, D, and E?

Combinations

How many different groups of 3 could be selected from the 5 items A, B, C, D, and E?

Definition (Combinations without repetition)

The selection of k items at the same time from n types of items (assuming that the order does not matter and there is no replacement). The number of such combinations is given by the formula:

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

Combinations con't

How many different groups of 3 could be selected from the 5 items A, B, C, D, and E assuming one letter can be chosen multiple times?

Combinations con't

How many different groups of 3 could be selected from the 5 items A, B, C, D, and E assuming one letter can be chosen multiple times?

Definition (Combinations with repetition)

The selection of k items at the same time from n types of items, where the order does not matter and a single item can be selected multiple times. The number of such combinations is given by the formula:

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k! \cdot (n-1)!}$$

Variations

10 people are competing in a running race. In how many different ways can the top three places (first, second, and third) be filled?

Variations

10 people are competing in a running race. In how many different ways can the top three places (first, second, and third) be filled?

Definition (Variation without repetition)

An ordered arrangement of k elements chosen from n (distinct) elements, where $k \leq n$ and there is no replacement. The number of such arrangements is given by the formula: $\frac{n!}{(n-k)!}$

How many different 4-digit PIN codes can be created if each digit from 0 to 9 can be used multiple times?

Variations

10 people are competing in a running race. In how many different ways can the top three places (first, second, and third) be filled?

Definition (Variation without repetition)

An ordered arrangement of k elements chosen from n (distinct) elements, where $k \leq n$ and there is no replacement. The number of such arrangements is given by the formula: $\frac{n!}{(n-k)!}$

How many different 4-digit PIN codes can be created if each digit from 0 to 9 can be used multiple times?

Definition (Variation with repetition)

An ordered arrangement of k elements chosen from n (distinct) elements, where $k \leq n$ and there is replacement. The number of such arrangements is given by the formula is n^k.

Sample Space and Events

We introduce the concept of the probability of an event and then show how probabilities can be computed in certain situations. As a preliminary, however, we need the concept of the sample space and the events of an experiment. If the experiment consists of flipping two coins, then the sample space consists of the following four points:

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

Any subset E of the sample space is known as an event. If $E = \{(H, H), (H, T)\}$, then E is the event that a head appears on the first coin.
Let $F = \{(T, H), (H, T)\}$, then $E \cup F = \{(H, H), (H, T), (T, H)\}$ and $E \cap F = \{(H, T)\}$
If $E \cap F = \emptyset$, then E and F are said to be mutually exclusive.
Finally, for any event E, we define the new event E^c, referred to as the complement of E, to consist of all outcomes in the sample space S that are not in E.

Axioms of probability

Consider an experiment whose sample space is S. For each event E of the sample space S, we assume that a number $P(E)$ is defined and satisfies the following three axioms:

1. $0 \leq P(E) \leq 1$
2. $P(S) = 1$
3. For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (that is, events for which $E_i \cap E_j = \emptyset\,, when\, i \neq j$),
$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$
We refer to $P(E)$ as the probability of the event E.

From this, we can show: $P(\emptyset) = 0$ and

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$

Some simple Propositions

- $P(E^c) = 1 - P(E)$
- If $E \subset F$, then $P(E) \leq P(F)$.
- $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Examples

If n people are present in a room, what is the probability that no two of them celebrate their birthday on the same day of the year? How large need n be so that this probability is less than $1/2$?

Homework

1. How many ways can 8 rooks be placed on a chessboard so that they do not attack each other?
2. What is the probability that a randomly selected 6-digit number has all different digits?
3. If we draw three cards with replacement from a 32-card Hungarian deck (8 cards of each suit: red, green, acorn, pumpkin), what is the probability that:
   a) We draw exactly one red card?
   b) We draw at least one red card?
4. In a store, out of 10 seemingly identical computers, 3 are refurbished and the rest are new. What is the probability that if we buy 5 computers for a lab, exactly 2 of them will be refurbished?
5. If we randomly choose our 6-character password from the 10 digits and 26 letters of the alphabet, what is the probability that it contains exactly 3 digits?
6. What is the probability of winning the "Ötöslottó", and also the probability of getting at least a four-match? Additionally, what is the probability that all five drawn numbers are even? How does this compare to a scenario with replacement?

Week 2

Conditional probability

Examples

Suppose that we toss 2 dice, and suppose that each of the 36 possible outcomes is equally likely to occur and hence has probability $1/36$. Suppose further that we observe that the first die is a 3. Then, given this information, what is the probability that the sum of the 2 dice equals 8?

Definition (Conditional Probability)

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad \text{if } P(B) \neq 0$$

Examples

What is the probability that in two dice rolls, both rolls are 6, given that we know at least one of the rolls is a 6?

Law of Total Probability

Definition (Complete System of Events)

The events $B_1, B_2, \ldots$ form a complete system of events if: 1) $B_i \cap B_j = \emptyset$ for all $i \neq j$; 2) $\bigcup_{i=1}^{\infty} B_i = \Omega$

Theorem (Law of Total Probability)

Let $B_1, B_2, \ldots$ be a complete system of events, and A be any event, with $P(B_j) > 0$ for all j. Then $P(A) = \sum_{j=1}^{\infty} P(A|B_j)P(B_j)$.

Examples

An company believes that people can be divided into 2 classes: those who are accident-prone and those who are not. The statistics show that an accident-prone person will have an accident at some time within a fixed 1-year period with probability 0.4, whereas this probability decreases to 0.2 for a person who is not accident prone. If we assume that 30 percent of the population is accident prone, what is the probability that a new policyholder will have an accident within a year of purchasing a policy?

Bayes Theorem

Theorem (Bayes Theorem)

Let $B_1, \ldots, B_n, \ldots$ be a complete system of events, and A be any event, with $P(B_j) > 0$ for all j. Then

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{j=1}^{\infty} P(A|B_j)P(B_j)}$$

Examples

In answering a question on a multiplechoice test, a student either knows the answer or guesses. Let p be the probability that the student knows the answer and $1 - p$ be the probability that the student guesses. Assume that a student who guesses at the answer will be correct with probability $1/m$, where m is the number of multiplechoice alternatives. What is the conditional probability that a student knew the answer to a question given that he or she answered it correctly?

Independence

Definition (Independence)

Events A and B are independent if

$$P(A \cap B) = P(A) \cdot P(B)$$

This means that the occurrence of event A does not influence the occurrence of event B, and vice versa.

Examples

A computer program consists of two independent parts. One part has a probability of error of 0.2, and the other has a probability of error of 0.3. If the program indicates an error, what is the probability that both parts are faulty?

Homework

1. 41 million "Ötöslottó" tickets are filled out independently. What is the probability that there will be at least one five-match?
2. Out of 100 coins, one is fake (it has heads on both sides). We randomly select one coin and flip it 10 times, getting 10 heads. Given this outcome, what is the probability that we were flipping the fake coin?
3. A computer processor is produced in 3 factories. There is a 20% chance it was made in the first, a 30% chance in the second, and a 50% chance in the third. The probabilities of warranty failures in these factories are 10%, 4%, and 1%, respectively. If our computer's processor has broken down, what is the probability that it was made in the first factory?
4. (*Monty Hall problem*) In a television show there are 3 doors, such that there is a present behind one of them, and there is nothing behind the other two. The participant of the game chooses one of the doors. Then, the presenter opens one of the other doors, and shows that there is nothing behind it. Then he asks the participant whether he would like to stay with his original choice, or he would like to chose the other door which has not been opened yet. Is it worth changing?

Homework 2

1. There is a disease which affects 2% of the population. There is a blood test which is positive with probability 95% for people who have the disease, and it is positive with probability 1% for healthy people.
   a What is the probability that Peter's test will be positive? (Peter is a randomly chosen person.)
   b Given that the test of Peter is positive, what is the conditional probability that he has the disease?
   c Now suppose that the test was repeated k times independently for a randomly chosen person, and all results were positive. What is the conditional probability that this person has the disease? Calculate this probability for $k = 2$ and $k = 3$
2. What is the probability that maximum of the numbers that we get is equal to 5 if we throw 2 (generally n) regular dice? (Regular dice: 1, 2, 3, 4, 5, 6 with equal probabilities.)
3. Odesseus arrives at a junction while wandering on the road. One leads to Athens, one to Sparta and one to Mycenae. The athenians tell the truth every third time when asked, the mycanaeans every second time, while the spartans never lie. Odesseus does not know which way leads to which city so he chooses randomly. Arriving in the city he asks a person the question: How much is 2×2. The answer is 4. What is the probability that he is in Athens?