Probability and Statistics course

Introduction

Gabor Vigh

TTK Department of Probability Theory and Statistics
ELTE

February 8, 2026

Agenda for the semester

1. Combinatorial Analysis & Axioms of Probability (12th Feb)
2. Conditional Probability and Independence (19th Feb)
3. Random Variables (26th Feb)
4. Discrete distributions(5th Mar)
5. Continuous Random Variables & Jointly Distributed Random Variables (12th Mar)
6. Properties of Expectation & Limit Theorems (19th Mar)
7. Midterm (26th Mar)
8. (Autumn holiday: 1st Apr - 7th Apr)
9. Descriptive statistics (9th Apr)
10. The basics of statistics: estimation (maximum likelihood, moments) (16th Apr)
11. Confidence intervals (23rd Apr)
12. Parametric and non-parametric probes (30th Apr)
13. Linear regression (7th May)
14. Final Exam (14th May)

Rules

- There will be a midterm worth of 50 points on 26th Mar and a exam worth 50 points on 14th May 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.
- The class is on every Thursday from 8:00 to 9:30 pm.
- the participation on the class is mandatory
- I need a volunteer who can create a Teams chat for the class - who wants to volunteer?

Materials

[Baron, 2014] [Ross, 2012]

Baron, M. (2014).
Probability and statistics for computer scientists.
CRC Press.

Ross, S. (2012).
First Course in Probability.
Pearson.

Week 1

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.

Definition

A collection of all elementary results, or outcomes of an experiment, is called a sample space.

Definition

Any set of outcomes is an event. Thus, events are subsets of the sample space.

Examples

Consider a football game between the Dallas Cowboys and the New York Giants. The sample space consists of 3 outcomes,$\Omega = \{\text{Cowboys win, Giants win, they tie}\}$

Set operations

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

Consider an event that consists of some finite or countable collection of mutually exclusive outcomes,

$$E = \{\omega_1, \omega_2, \omega_3, \dots\}.$$

Summing probabilities of these outcomes, we obtain the probability of the entire event,

$$P(E) = \sum_{\omega_k \in E} P(\omega_k) = P(\omega_1) + P(\omega_2) + P(\omega_3) + \dots$$

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

Examples

There is a 1% probability for a hard drive to crash. Therefore, it has two backups, each having a 2% probability to crash, and all three components are independent of each other. The stored information is lost only in an unfortunate situation when all three devices crash. What is the probability that the information is saved?

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.

Reliability applications

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. what is the probability that both has errors?

Examples

Suppose that a shuttle's launch depends on three key devices that operate independently and malfunction with probabilities $p_1 = 0.01$, $p_2 = 0.02$, and $p_3 = 0.02$.

Solutions

Since the parts are independent:

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2) = 0.2 \times 0.3 = 0.06$$

The launch occurs on time only if all three devices function correctly. Let F_i be the event that device i functions. The probability that a device functions is:

$$P(F_i) = 1 - P(\text{malfunction}_i)$$

Given the independence of the devices, the probability of a timely launch is the product of the individual success probabilities:

$$\begin{aligned} P(\text{Launch}) &= P(F_1) \cdot P(F_2) \cdot P(F_3) \\ &= (1 - 0.01) \cdot (1 - 0.02) \cdot (1 - 0.02) \\ &= 0.99 \times 0.98 \times 0.98 \\ &= 0.950796 \end{aligned}$$

Thus, there is approximately a **95.08%** chance that the shuttle will be launched on time.

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:

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

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

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.

Additional examples

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?