Course guide
270206 - PIE1 - Probability and Statistics 1

Unit in charge: Barcelona School of Informatics
Teaching unit: 749 - MAT - Department of Mathematics.

Degree: BACHELOR'S DEGREE IN DATA SCIENCE AND ENGINEERING (Syllabus 2017). (Compulsory subject).
Academic year: 2022  ECTS Credits: 7.5  Languages: Catalan, Spanish, English

LECTURER

Coordinating lecturer: GUILLEM PERARNAU LLOBET

Others: Segon quadrimestre:
ANNA DE MIER VINUÉ - 11
PEDRO FRANCISCO DELICADO USEROS - 11, 12
MIQUEL ORTEGA SÁNCHEZ COLOMER - 12
GUILLEM PERARNAU LLOBET - 11, 12

PRIOR SKILLS

The contents of the previous subjects in the Degree.

DEGREE COMPETENCES TO WHICH THE SUBJECT CONTRIBUTES

Specific:
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.

General:
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.

Transversal:
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
CT6. Autonomous Learning. Detect deficiencies in one’s own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

Basic:
CB1. That students have demonstrated to possess and understand knowledge in an area of ??study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.
CB3. That students have the ability to gather and interpret relevant data (usually within their area of ??study) to make judgments that include a reflection on relevant social, scientific or ethical issues.
CB5. That the students have developed those learning skills necessary to undertake later studies with a high degree of autonomy.
TEACHING METHODOLOGY

Theory:
Lectures develop the theory and include illustrative examples.

Problems:
The students have in advance the list of problems relevant to the topic being developed in theory. They had the opportunity to try to solve problems before the problems class. They require the teacher’s help in the points where they have encountered difficulties. The teacher solves these questions in the blackboard and develops the full solution of some problems that he or she considers that are especially challenging.

Laboratory:
The teacher introduces the R language during the course, with special emphasis on random variables simulation tools, descriptive statistics and univariate statistical inference.

LEARNING OBJECTIVES OF THE SUBJECT

1. At the end of the course, students will know the definition of probability and their properties, and will apply them to solve probability calculation problems.
2. At the end of the course students will know how to use the concept of random variable to formalize and solve probability calculation problems.
3. At the end of the course students will know how to simulate complex random phenomena with the computer and deduce approximate values of amounts of interest (probabilities, characteristics of random variables) that are difficult to calculate analytically.
4. At the end of the course students will know the most common probabilistic distributions and will be able to recognize situations where they are used to model real phenomena.
5. At the end of the course, students will know how to calculate distributions and expected expectations and use them in prediction.
6. At the end of the course, students will know whether two random variables are independent, and if they are not, they will be able to measure the linear correlation coefficient.
7. At the end of the course, students will know the Law of the Great Names and the Central Limit Theorem.
8. At the end of the course, the students will understand stochastic processes and will know how to model random-flavoured problems using Markov chains.
9. At the end of the course students will know the basic tools of descriptive statistics and will know how to apply them.
10. At the end of the course students will know the concepts of population, sample, parameter and estimator, and know the basic properties.
11. At the end of the course, students will know the basics of timely estimation and will know how to calculate them in real situations.

STUDY LOAD

<table>
<thead>
<tr>
<th>Type</th>
<th>Hours</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours small group</td>
<td>30,0</td>
<td>16.00</td>
</tr>
<tr>
<td>Self study</td>
<td>112,5</td>
<td>60.00</td>
</tr>
<tr>
<td>Hours large group</td>
<td>45,0</td>
<td>24.00</td>
</tr>
</tbody>
</table>

Total learning time: 187.5 h

CONTENTS

<table>
<thead>
<tr>
<th>Probability spaces and random variables</th>
</tr>
</thead>
</table>

Description:
## Random variables

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
</table>

## Random vectors

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
</table>

## Sum of random variables

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution of the sum. Markov's, Chebyshev's and Chernoff's Inequalities. Law of Large Numbers. Central Limit Theorem.</td>
</tr>
</tbody>
</table>

## Stochastic processes

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
</table>

## Population and sample

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
</table>

## Point estimation

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
</table>
## ACTIVITIES

### Developing the Topic "Probability and random variables"

**Description:**
Developing the Topic "Probability and random variables"

**Specific objectives:**
1, 3

**Related competencies:**
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
CB5. That the students have developed those learning skills necessary to undertake later studies with a high degree of autonomy.
CB1. That students have demonstrated to possess and understand knowledge in an area of study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply knowledge from the vanguard of their field of study.
CB3. That students have the ability to gather and interpret relevant data (usually within their area of study) to make judgments that include a reflection on relevant social, scientific or ethical issues.

**Full-or-part-time:** 17h

- Theory classes: 6h
- Practical classes: 2h
- Laboratory classes: 2h
- Self study: 7h
Developing the Chapter "Random variables"

Description:
Developing the Chapter "Random variables"

Specific objectives:
2, 3, 4, 9

Related competencies:
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
CB5. That the students have developed those learning skills necessary to undertake later studies with a high degree of autonomy
CB1. That students have demonstrated to possess and understand knowledge in an area of ??study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.
CB3. That students have the ability to gather and interpret relevant data (usually within their area of ??study) to make judgments that include a reflection on relevant social, scientific or ethical issues.

Full-or-part-time: 17h
Theory classes: 6h
Practical classes: 2h
Laboratory classes: 2h
Self study: 7h
Developing the Chapter "Random vectors"

Description:
Developing the Chapter "Random vectors"

Specific objectives:
3, 5, 6, 9

Related competencies:
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
CB5. That the students have developed those learning skills necessary to undertake later studies with a high degree of autonomy.
CB1. That students have demonstrated to possess and understand knowledge in an area of study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.
CB3. That students have the ability to gather and interpret relevant data (usually within their area of study) to make judgments that include a reflection on relevant social, scientific or ethical issues.

Full-or-part-time: 18h 48m
Theory classes: 7h 30m
Practical classes: 3h
Laboratory classes: 2h
Self study: 6h 18m
Developing the Chapter "Sum of random variables"

Description:
Developing the Chapter "Sum of random variables"

Specific objectives:
3, 5, 7

Related competencies:
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.

Full-or-part-time: 14h 18m
Theory classes: 6h
Practical classes: 1h 30m
Laboratory classes: 1h 30m
Self study: 5h 18m

Developing the Chapter "Stochastic Processes"

Description:
Developing the Chapter "Stochastic Processes"

Specific objectives:
8

Related competencies:
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CB1. That students have demonstrated to possess and understand knowledge in an area of study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.

Full-or-part-time: 25h 42m
Theory classes: 9h 30m
Practical classes: 3h
Laboratory classes: 4h
Self study: 9h 12m
Developing the Chapter "Population and sample"

**Description:**
Developing the Chapter "Population and sample"

**Specific objectives:**
3, 9, 10

**Related competencies :**
- CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
- CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
- CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
- CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
- CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
- CB5. That the students have developed those learning skills necessary to undertake later studies with a high degree of autonomy.
- CB3. That students have the ability to gather and interpret relevant data (usually within their area of ??study) to make judgments that include a reflection on relevant social, scientific or ethical issues.

**Full-or-part-time:** 12h
- Theory classes: 4h
- Practical classes: 2h
- Laboratory classes: 2h
- Self study: 4h

Developing the Chapter "Point estimation"

**Description:**
Developing the Chapter "Point estimation"

**Specific objectives:**
11

**Related competencies :**
- CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
- CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
- CT6. Autonomous Learning. Detect deficiencies in one's own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
- CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
- CB5. That the students have developed those learning skills necessary to undertake later studies with a high degree of autonomy.
- CB3. That students have the ability to gather and interpret relevant data (usually within their area of ??study) to make judgments that include a reflection on relevant social, scientific or ethical issues.

**Full-or-part-time:** 15h 12m
- Theory classes: 6h
- Practical classes: 1h 30m
- Laboratory classes: 1h 30m
- Self study: 6h 12m
Final examination

Description:
Final examination

Specific objectives:
1, 2, 4, 5, 6, 7, 10, 11

Related competencies:
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CT6. Autonomous Learning. Detect deficiencies in one’s own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
CB5. That the students have developed those learning skills necessary to undertake later studies with a high degree of autonomy.
CB1. That students have demonstrated to possess and understand knowledge in an area of study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply knowledge from the vanguard of their field of study.
CB3. That students have the ability to gather and interpret relevant data (usually within their area of study) to make judgments that include a reflection on relevant social, scientific or ethical issues.

Full-or-part-time: 33h
Guided activities: 3h
Self study: 30h

Mid-term examination

Description:
Mid-term examination

Specific objectives:
1, 2, 4, 5, 6

Related competencies:
CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
CT6. Autonomous Learning. Detect deficiencies in one’s own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
CB5. That the students have developed those learning skills necessary to undertake later studies with a high degree of autonomy.

Full-or-part-time: 12h
Guided activities: 2h
Self study: 10h
**Mid-term lab examination**

**Description:**
Mid-term lab examination

**Specific objectives:**
3, 9

**Related competencies:**
- CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
- CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
- CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
- CT6. Autonomous Learning. Detect deficiencies in one’s own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
- CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
- CB5. That the students have developed those learning skills necessary to undertake later studies with a high degree of autonomy
- CB1. That students have demonstrated to possess and understand knowledge in an area of study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.
- CB3. That students have the ability to gather and interpret relevant data (usually within their area of study) to make judgments that include a reflection on relevant social, scientific or ethical issues.

**Full-or-part-time:** 11h
Guided activities: 1h
Self study: 10h

---

**Final lab examination**

**Description:**
Final lab examination

**Specific objectives:**
3, 9, 11

**Related competencies:**
- CG1. To design computer systems that integrate data of provenances and very diverse forms, create with them mathematical models, reason on these models and act accordingly, learning from experience.
- CG2. Choose and apply the most appropriate methods and techniques to a problem defined by data that represents a challenge for its volume, speed, variety or heterogeneity, including computer, mathematical, statistical and signal processing methods.
- CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
- CT6. Autonomous Learning. Detect deficiencies in one’s own knowledge and overcome them through critical reflection and the choice of the best action to extend this knowledge.
- CT5. Solvent use of information resources. Manage the acquisition, structuring, analysis and visualization of data and information in the field of specialty and critically evaluate the results of such management.
- CB5. That the students have developed those learning skills necessary to undertake later studies with a high degree of autonomy
- CB1. That students have demonstrated to possess and understand knowledge in an area of study that starts from the base of general secondary education, and is usually found at a level that, although supported by advanced textbooks, also includes some aspects that imply Knowledge from the vanguard of their field of study.
- CB3. That students have the ability to gather and interpret relevant data (usually within their area of study) to make judgments that include a reflection on relevant social, scientific or ethical issues.

**Full-or-part-time:** 11h 30m
Guided activities: 1h 30m
Self study: 10h
GRADING SYSTEM

A midterm exam (EP) and a final exam (EF). The midterm exam will assess the first part of the course, and the final exam, the second one. Each of them has a part of theory and problems, and may contain a part of laboratory. Optionally, the day of the final exam it will be possible to resit the midterm exam (REP), if the exam is submitted, its mark will replace the mark of the midterm exam.

During the course short activities or assignments (ACP) will be proposed.

The final grade (NF) is computed as follows: if the resit is not submitted

\[
NF = 0.45 \cdot EP + 0.45 \cdot EF + 0.10 \cdot ACP,
\]

and if the resit exam is submitted

\[
NF = 0.45 \cdot REP + 0.45 \cdot EF + 0.10 \cdot ACP.
\]

Only students with NF smaller than 5 can opt to re-evaluation. The re-evaluation exam grade (ER) replaces the 100% of the midterm and final exams grade. So the final grade after re-evaluation (NFreav) will be

\[
NFreav = 0.90 \cdot ER + 0.10 \cdot ACP.
\]

In case that NFreav is smaller than 5, the final mark will be the maximum between NF and NFreav.

BIBLIOGRAPHY

Basic:

Complementary:

RESOURCES

Hyperlink: