Course guide
270206 - PIE1 - Probability and Statistics 1

Last modified: 31/01/2025

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: 2024 **ECTS Credits:** 7.5 **Languages:** Catalan, Spanish

LECTURER

Coordinating lecturer: GUILLEM PERARNAU LLOBET

Others: Segon quadrimestre:
RICHARD JOHANNES LANG - 11, 12
GUILLEM PERARNAU LLOBET - 11, 12
ANDREA TOLOBA LÓPEZ-EGEA - 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.

Generical:
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

Type	Hours	Percentage
Hours small group	30,0	16.00
Self study	112,5	60.00
Hours large group	45,0	24.00

Total learning time: 187.5 h

CONTENTS

Probability spaces and random variables

Description:
Random experiences. Algebra of events. Probability space. Conditional probability. Independence of events. Bayes' Theorem. Simulation of random experiments.

Random variables

Description:
Definition of random variable. Probability distribution function. Discrete random variables (probability function) and continuous variables (probability density function). Expectation and moments. Models of usual distributions. Simulation of random variables.

Random vectors

Description:
Multidimensional distributions. Independence. Conditioned distributions. Covariance and correlation. Expectation and covariance matrix. Conditioned expectation. Multinomial distribution. Multivariate normal distribution.

Sum of random variables

Description:
Distribution of the sum. Markov's, Chebyshev's and Chernoff's Inequalities. Law of Large Numbers. Central Limit Theorem.

Stochastic processes

Description:
Stochastic processes. Markov chains. Recurrence and transience. Ergodic theorem.

Population and sample

Description:
Random sample. Parametric statistical models. Parameters and estimators. Descriptive statistics.

Point estimation

Description:
Method of moments. Maximum likelihood. Properties of the estimators (bias, variance, mean square error, consistency).

ACTIVITIES

Developing the Topic "Probability and random variables"

Description:
Developing the Topic "Probability and random variables"

Specific objectives:
1, 3

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

Full-or-part-time: 17h
Self study: 8h
Theory classes: 5h
Practical classes: 2h
Laboratory classes: 2h

Developing the Chapter "Random variables"

Description:
Developing the Chapter "Random variables"

Specific objectives:
2, 3, 4, 9

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

Full-or-part-time: 17h
Self study: 8h
Theory classes: 5h
Practical classes: 2h
Laboratory classes: 2h

Developing the Chapter "Random vectors"

Description:
Developing the Chapter "Random vectors"

Specific objectives:
3, 5, 6, 9

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

Full-or-part-time: 18h 48m
Self study: 6h 18m
Theory classes: 7h 30m
Practical classes: 3h
Laboratory classes: 2h

Developing the Chapter "Sum of random variables"

Description:
Developing the Chapter "Sum of random variables"

Specific objectives:
3, 5, 7

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

Full-or-part-time: 14h 18m
Self study: 7h 48m
Theory classes: 5h
Practical classes: 1h 30m

Developing the Chapter "Stochastic Processes"

Description:
Developing the Chapter "Stochastic Processes"

Specific objectives:
8

Related competencies :
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.
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
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.
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: 25h 42m
Self study: 11h 12m
Theory classes: 8h 30m
Practical classes: 3h
Laboratory classes: 3h

Developing the Chapter "Population and sample"

Description:
Developing the Chapter "Population and sample"

Specific objectives:
3, 9, 10

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

Full-or-part-time: 12h
Self study: 4h 30m
Theory classes: 3h 30m
Practical classes: 2h
Laboratory classes: 2h

Developing the Chapter "Point estimation"

Description:
Developing the Chapter "Point estimation"

Specific objectives:
11

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

Full-or-part-time: 15h 12m
Self study: 6h 42m
Theory classes: 5h 30m
Practical classes: 1h 30m
Laboratory classes: 1h 30m

Final examination

Description:
Final examination

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

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

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

Mid-term examination

Description:
Mid-term examination

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

Related competencies :
CB5. That the students have developed those learning skills necessary to undertake later studies with a high degree of autonomy
CE3. Analyze complex phenomena through probability and statistics, and propose models of these types in specific situations. Formulate and solve mathematical optimization problems.
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.
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: 12h
Self study: 10h
Guided activities: 2h

Mid-term lab examination

Description:
Mid-term lab examination

Specific objectives:
3, 9

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

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

Final lab examination

Description:
Final lab examination

Specific objectives:
3, 9, 11

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

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

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·EP +0.45·EF +0.10·ACP,

and if the resit exam is submitted

NF =0.45·REP +0.45·EF +0.10·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 * ER + 0.10*ACP.

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

BIBLIOGRAPHY

Basic:
- Baumer, B.S.; Kaplan, D.T.; Horton, N.J. Modern data science with R. Boca Raton: Taylor & Francis CRC Press, 2017. ISBN 9781498724487.
- Bertsekas, D.P.; Tsitsiklis, J.N. Introduction to probability. 2nd ed. Belmont, Massachussets: Athena Scientific, 2008. ISBN 9781886529236.
- DeGroot, M.H.; Schervish, M.J. Probability and statistics. 4th ed. Boston: Pearson, 2012. ISBN 9780321709707.
- Evans, M.J.; Rosenthal, J.S. Probability and statistics: the science of uncertainty. 2nd ed. New York: W.H. Freeman and Company, 2010. ISBN 9781429224628.
- Baron, M. Probability and statistics for computer scientists. 3rd ed. Boca Raton, FL: CRC Press, 2019. ISBN 9781138044487.

Complementary:
- Pitman, J. Probability. New York: Springer, 1993. ISBN 0387979743.
- Mitzenmacher, M.; Upfal, E. Probability and computing: randomization and probabilistic techniques in algorithms and data analysis. 2nd ed. Cambridge, United Kingdom ; New York, NY, USA: Cambridge University Press, 2017. ISBN 9781107154889.
- Grinstead, C.M.; Snell, J.L. Introduction to probability. 2nd rev. ed. Providence (R.I.): American Mathematical Society, 1997. ISBN 0821807498.

RESOURCES

Hyperlink:
- http://www.datascienceassn.org/sites/default/files/Introduction%20to%20Probabili