Probability and Statistics

Kristel Van Steen, PhD

Montefiore Institute - Systems and Modeling
GIGA - Bioinformatics
ULg

kristel.vansteen@ulg.ac.be
INTRODUCTION

Course contents

• CH1: Probability theory
• CH2: Random variables
• CH3: Parametric families of univariate distributions
• CH4: Distributions in the presence of multi-dimensionality and functions of random variables
• CH5: Sampling to approximate the true world
• CH6: A gentle introduction to everyday statistics
• CH7: Introduction to statistical inference
• CH8: Relationships
Course objective

• To understand present-day applications of Probability and Statistics.
• To provide pointers to more advanced literature so as to be able to find material to deal with more sophisticated problems
Workshops

• As the instructors are shared with the Calculus professor, practical sessions are given before the theory sessions:
  o Recapitulation of previous class – hence, this is your chance to ask for clarifications!
  o Exercises about last week’s course chapters
Organization

- 15h of Theory lessons by the Professor (in English) +
- 15h of Practical works by Instructors (in French) in smaller groups for better follow-up
- Contact addresses:
  - Theory: Prof Dr Dr K Van Steen
    - kristel.vansteen@ulg.ac.be
    - www.montefiore.ulg.ac.be/~kvansteen
  - Practicals + Theory:
    - Mr Darmont (Marcel): Marcel.Darmont@ulg.ac.be
    - Mrs Huaux (Hélène): H.Huaux@ulg.ac.be
    - Mr Lousberg (Pierre): plousberg@ulg.ac.be
## Partition of the workshop groups

<table>
<thead>
<tr>
<th>Class</th>
<th>Group</th>
<th>Representative</th>
<th>Room</th>
</tr>
</thead>
<tbody>
<tr>
<td>3ème B. Ing Cvl [ A → G ]</td>
<td></td>
<td>Mr Darmont</td>
<td>R52(B4 - Europe)</td>
</tr>
<tr>
<td>3ème B. Ing Cvl [ H → R ]</td>
<td></td>
<td>Mr Lousberg</td>
<td>R53(B4 - Europe)</td>
</tr>
<tr>
<td>3ème B. Ing Cvl [ S → Z ]</td>
<td>3ème B. Ing Archi.</td>
<td>2ème B. en Sc. Info.</td>
<td>R54(B4 - Europe)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>le 16/11 : S74 (B4)</td>
</tr>
</tbody>
</table>
## Organization

<table>
<thead>
<tr>
<th>PROB and STATS</th>
<th>Theory</th>
<th># slides (*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tu 21/09</td>
<td>2hrs intro + CH1</td>
<td>CH1 92</td>
</tr>
<tr>
<td>Tu 28/09</td>
<td>2hrs CH2 + CH3</td>
<td>CH2 90</td>
</tr>
<tr>
<td>Tu 05/10</td>
<td>2hrs CH3 + CH4</td>
<td>CH3 127</td>
</tr>
<tr>
<td>Tu 12/10</td>
<td></td>
<td>CH4 151</td>
</tr>
<tr>
<td>Tu 19/10</td>
<td>2hrs CH5</td>
<td>CH5 119</td>
</tr>
<tr>
<td>Tu 26/10</td>
<td>2hrs CH6</td>
<td>CH6 145</td>
</tr>
<tr>
<td>Tu 02/11</td>
<td>2hrs CH7</td>
<td>CH7 280</td>
</tr>
<tr>
<td>Tu 09/11</td>
<td>2hrs CH8</td>
<td>CH8 172</td>
</tr>
</tbody>
</table>

- The slides are comprehensive and provide sufficient background info
- Some of the course notes are part of self-study time: ask questions during workshops as well!
- Practicals start on Tu 28 September
- Ch8 exercises + course: 16 or 17 December?
Course website  
(www.montefiore.ulg.ac.be/~kvansteen)

Kristel Van Steen, PhD

Contact Information
Département d'Électricité, Électronique et Informatique (Inst. Montefiore) / Bioinformatique
Mail Adresse: BAT. 528 Bioinformatique
Grande Traverse, 10
4000 Liége 1
Belgium
Office: 0.15 (BAT 37)
Tel: +32 4 360 2692
Email: Kristel.VanSteen@ulg.ac.be

Research Interests

Statistical Genetics
- Components analysis
- FEAT testing
- Gene-environment interactions
- Gene-gene interactions and interaction graphs
- Genetic heterogeneity
- Genetic imprinting
- Genome-wide association analysis
- Kinship and genomic background
- Multifactor dimensionality reduction strategies
- Multi-locus or combined group approaches
- Noisy or erroneous data handling
- Omics integrated analysis
- Phenotypes
- Population stratification

Links to affiliations
- ULG homepage
- Institut Montefiore
- Center for Medical Genetics (UZ Leuven)
- Center for Statistics (UZ Leuven)
- Center for Human Genetics (UZ Leuven)
- Global Atopy and Asthma European Network

Teaching 2010-2011
- MATH2002: Introduction to Probability and Statistics
- GBIO0008-1: Bioinformatique

Teaching 2009-2010
- GBIO0001-1: Introduction to Biomedical Engineering
- GBIO0001-1: Genetic Epidemiology for筋性 /Soma
Course website (www.montefiore.ulg.ac.be/~kvansteen)

Introduction to Probability and Statistics 2010-2011 (Preliminary Version - Check for Final Version)

MATH3008-2: Introduction to Probability and Statistics

SOME PRACTICAL ARRANGEMENTS

- Where are the classes given?
  - Complete Lecture Schedule (Updated)

- Where do the classes take place?
  - Theory B 1st - Room 142

- How can I pass the exam?
  - More Information? Click here

- What if I have questions?
  - Consult Karel Van Steen (karel.vansteen@ulg.ac.be) for any questions about theory or homeworks, probably on
    - Tuesdays before after class
    - Wednesdays or Fridays (check availability)

CLASS MATERIAL (use .tex to unpack ziped files):

- Preview on the course notes chapters (versions in progress):
  - Chapter 1
    - Chapter 1
    - Chapter 2
    - Chapter 3
    - Chapter 4
    - Chapter 5
    - Chapter 6
    - Chapter 7
    - Chapter 8
  - Corresponding references list for chapters: to be completed

First class: 21 September 2010:
10h30-12h00

- [Introduction/Notes]
  - Course notes [Chapter 1]
    - Downloading instructions: R and Bioconductor: [Follow this link]
    - Information on using R:
      - Part 1: R basics
      - Part 2: BioConductor
Written notes

• Slides in English = course notes, exercise book in French
• Reference books and publications - Recommended:

  o **In English:**
    "Probability and Statistical Inference - 8th edition",
    Hogg R.V., Tanis E.A. Pearson
    Education International,
    Prentice Hall 2010.

  o **In French:**
    "Probabilités analyse des données et statistique - 2e édition révisée et augmentée",
    Saporta G., Editions TECHNIP 2006, Paris, France
• Other references:


Assessment

• June:
  o Open book written exam
  o Involves solving several exercises based on the theory and practical classes
  o Total score: 20

• August:
  o Closed book oral
  o Involves explaining several important concepts, while completing your story by making references to as many course note sections that are applicable
  o Total score: 20
Assessment

• Note:
  o From next year onwards, also the workshop part will be re-organized, and hence the exam system will change
  o e.g.,
    ▪ June: open book multiple choice questions with some exercises/
    ▪ August: open book oral examination

UNLESS you would like to have this system already implemented from this year onwards ...
Questions?