Probability and Statistics

Kristel Van Steen, PhD[2]

Montefiore Institute - Systems and Modeling

GIGA - Bioinformatics

ULg

kristel.vansteen@ulg.ac.be

COURSE INTRODUCTION

1. Course contents

- CH1: Probability theory
- CH2: Random variables and associated functions
- CH3: Some important distributions
- CH4: It is all about data
- CH5: Parameter estimation
- CH6: Hypothesis testing

# 2. General course objectives

2.1. Introduction

- General objectives are 3-fold:
  - o Have a good notion about / understand the basic elements of probability and statistics
  - o Be able to use the concepts in practical applications (practical sessions)
  - o Be able to generalize material to a broader variety of practical problems

- Specific objectives:
  - o These are chapter/topic dependent
  - o These will be communicated to you as classes advance
  - o Meeting the specific objectives of this course ➲ Passing this course

- For example:

Compute the probability that a particular event will occur	CH1
Use the right probability distribution (normal, t, binomial, etc.) for your analysis	CH2-CH3
Retrieve relevant information by looking at your data	CH4
Estimate population means and proportions, based on sample data	CH5
Determine margin of error and confidence levels	CH5
Test hypotheses about means and proportions	CH6

2.2 An engineer's perspective
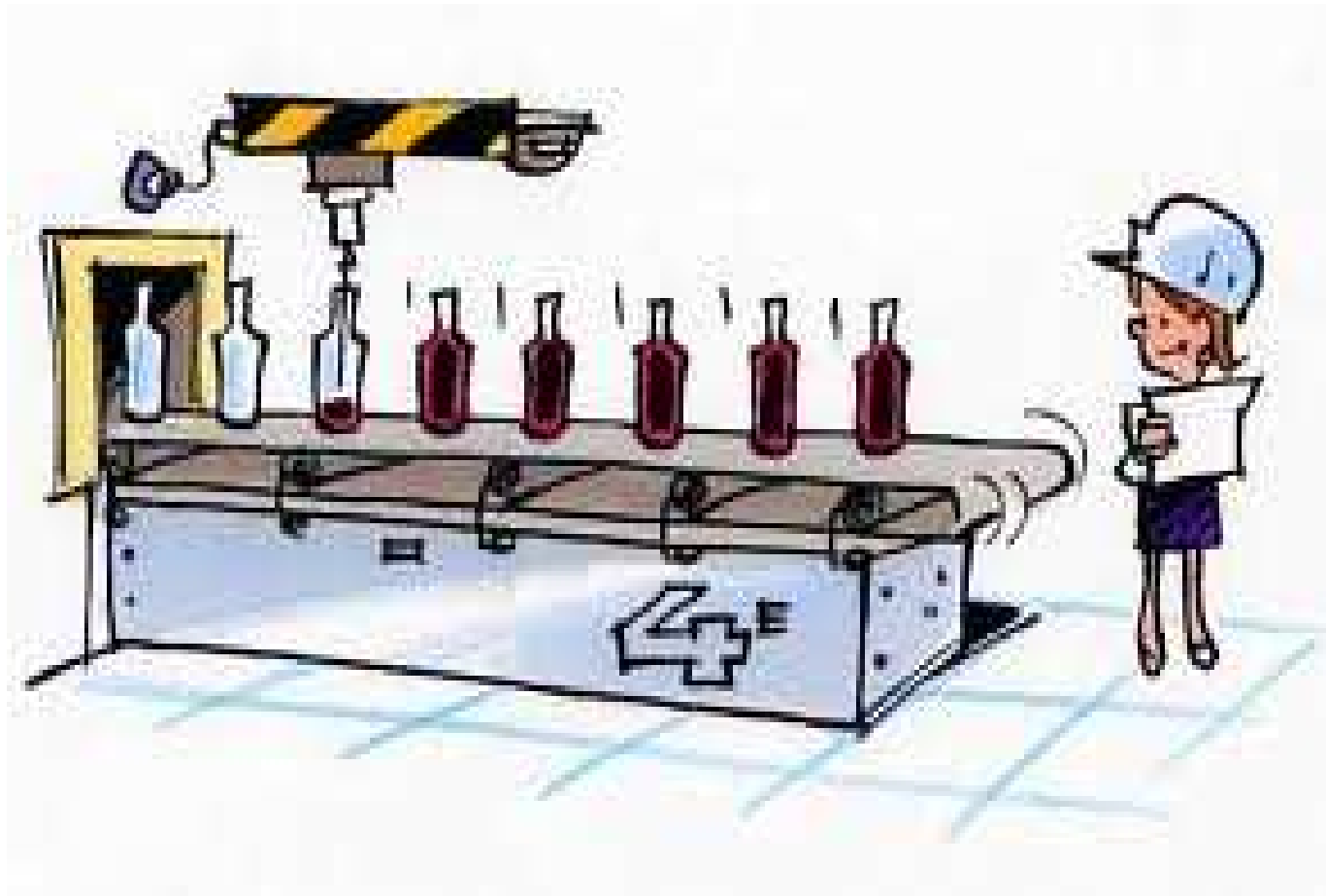
3.   Organization of the classes

3.1.   Course websites:

General information + notes theory classes:

www.montefiore.ulg.ac.be/~kvansteen

Notes practical classes:

www.montefiore.ulg.ac.be/~vanlishout

www.montefiore.ulg.ac.be/~kvansteen

Introduction to Probability and Statistics 2011-2012

MATH0008-2: Introduction to Probability and Statistics

SOME PRACTICAL ARRANGEMENTS:

- When are the classes given and which topics will be covered?
  - Complete Course Schedule
  - T: Theory; P: Practical

    (*Updated:* 15 September 2011)

- Where do the classes take place?
  - Room Schedule
  - T: Theory; P: Practical

    (*Updated:* 15 September 2011)

- How can I pass the exam?
  - General information about the course: Click here

    (*Updated:* 15 September 2011)

- What if I have questions on the course in general or on specifics?
  - Consult François Van Lishout (**f.vanlishout@ulg.ac.be**) or

3.2. Theoretical classes

Date	T (10.30-12.30)	P (08.15-10.15)	Changes	T Chapter	T Keywords
20/09	B7b "les petits amphis"		10.30-13.00	CH1	Probability
04/10	B7b			CH2	Random
11/10	B7b			CH2-3	Functions / discrete
18/10	B7b			CH3	Continuous distr
25/10	B7b			CH4	Explore data
08/11	B7b			CH5	Estimate
15/11	B7b			CH6	Test
20/12	*To be announced*			Repetition	Case study (guest)

Written notes

- Theory:
  - o Slides in English
    - o Downloadable from: **http://www.montefiore.ulg.ac.be/~kvansteen/**

- What if I have questions on the course in general or on specifics?
  - o Consult François Van Lishout (**f.vanlishout@ulg.ac.be**) or
  - o Kristel Van Steen (**kristel.vansteen@ulg.ac.be**)

  - o Send an e-mail first to check availability !!!

CLASS MATERIAL (use 7-Zip to unpack possibly zipped files):

T1 Class: 20 September 2011 : 10h30-12h30

- Course Introduction
- Theory course notes: Chapter 1

o The theory slides are "complete" for the purposes of this class. If you are in need for a good reference book to this course, we recommend:

o **In English:**

"Introduction to the theory of statistics –
3rd edition", Alexander M Mood; Franklin
A Graybill; and Duane C Boes, McGraw-Hill
series in probability and statistics 1974.
ISBN : 0-07-042864-6

o **In French:**

"Probabilités analyse des données et
statistique - 2e
édition révisée et augmentée", Saporta G.,
Editions
TECHNIP 2006, Paris, France
ISBN : 978-2-7108-0814-5

3.3. Practical classes

Date	P (08.15-10.15)	T Chapter	T Keyword
04/10	B4 (R.52) / B4 (R.53) / B4 (R.54)	CH1	Probability
11/10	B4 (R.52) / B4 (R.53) / B4 (R.54)	CH2	Random
18/10	B4 (R.52) / B4 (R.53) / B4 (R.54)	CH2-3	Functions
25/10	B4 (R.52) / B4 (R.53) / B4 (R.54)	CH3	Distributions
08/11	B4 (R.52) / B4 (R.53) / B4 (R.54)	CH4	Explore
15/11	*To be announced*	CH5	Estimate
22/11	*To be announced*	CH6	Test
20/12	*To be announced*	Loose end + your T+P questions	Case study (guest)

- Special efforts will be made to ensure that practicals are given AFTER the relevant theory has been seen
- If not, the relevant theory needed to solve the exercises will be summarized.
- Details about the practical sessions:

http://www.montefiore.ulg.ac.be/~vanlishout/

Partitioning for practical sessions

B4 (R.52)	B4 (R.53)	B4 (R.54)
Group 1	**Group 2**	**Group 3**
A-F (Van Lishout)	**G-M (Lousberg)**	**N-Z (Huaux)**

4. Course Assessment

- Exam - written:
    - o 1h15 theory (closed book) : multiple choice questions
        - ▪ French and English versions will be provided
    - o 15 minutes BREAK
    - o 2h30 exercises (open book) : 4 exercises

- Weights :
    - o 1/3 for the theoretical part and 2/3 for the exercise part
    - o Total score: 20/20

- The same system for May-June / August-September. However, there is the opportunity to orally explain solutions in August-September.
    - o Oral explanations can obviously be given in French (English when desired)

Will NOT be rated ok …. ☺
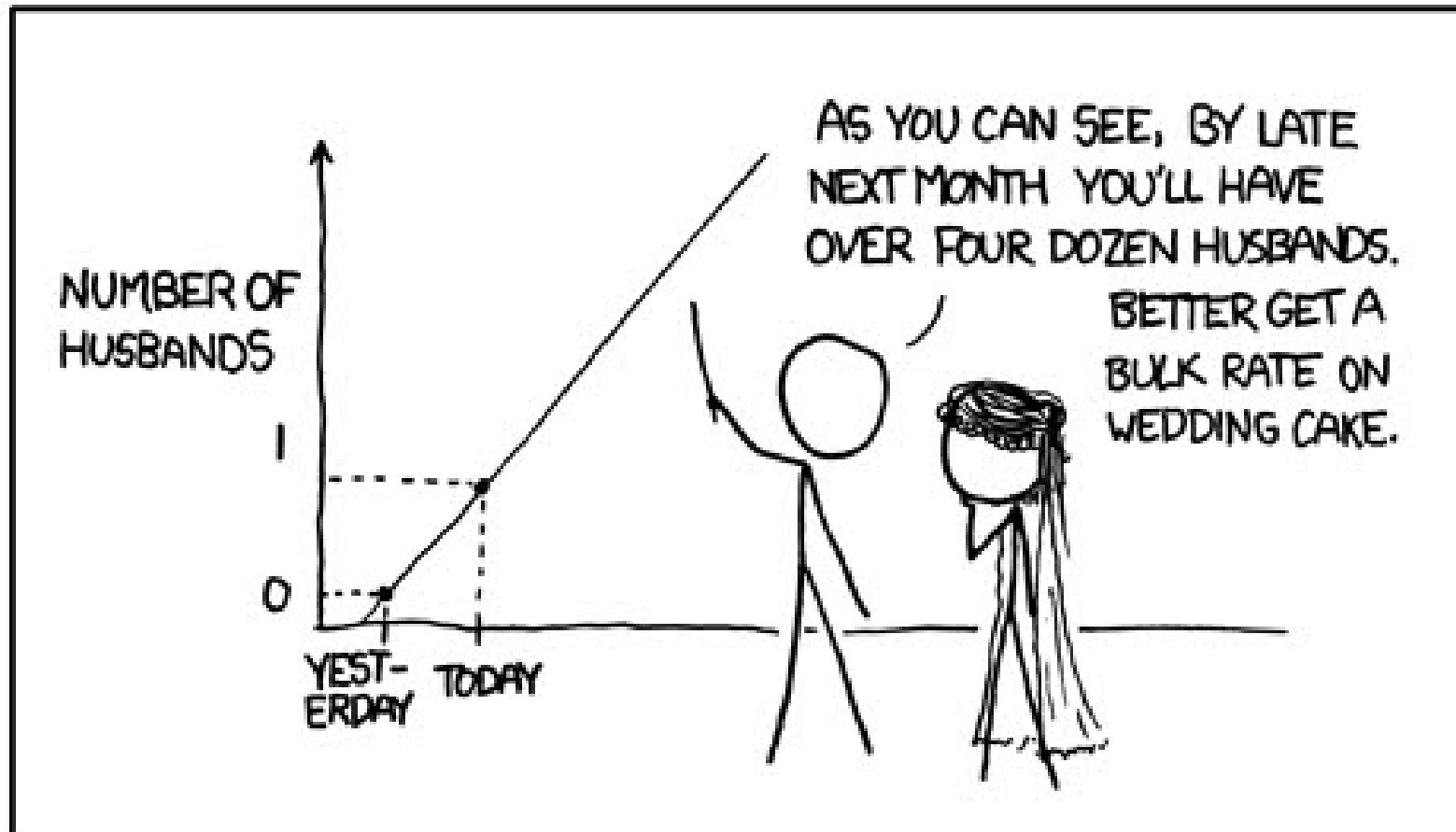
(Sunday Comics. Posted by Brad Walters)

It is easy to get lost in misconceptions ...

The Monty Hall problem

http://www.youtube.com/watch?v=mhlc7peGlGg

CHAPTER 1: PROBABILITY THEORY

1 What's in a name

1.1 Relevant questions in a probabilistic context

1.2 Relevant questions in a statistics context

2 Probability and statistics: two related disciplines

2.1 Probability

3 Different flavors of probability

3.1 Classical or a priori probability

3.2 Set theory

3.3 Sample space and probability measures

3.4 A posteriori or frequency probability

4 Statistical independence and conditional probability

4.1 Independence

4.2 Conditional probability

Law of total probability

Bayes' theorem

Bayesian odds

Principle of proportionality

5 In conclusion

5.1 Take-home messages

5.2 The birthday paradox

CHAPTER 2: RANDOM VARIABLES AND ASSOCIATED FUNCTIONS

1 Random variables

1.1 Formal definition

2 Functions of one variable

2.1 Probability distribution functions

2.2 The discrete case: probability mass functions

2.3 The binomial distribution

2.4 The continuous case: density functions

2.5 The normal distribution

2.6 The inverse cumulative distribution function

2.7 Mixed type distributions

2.8 Comparing cumulative distribution functions

3   Two or more random variables

3.1 Joint probability distribution function

3.2 The discrete case: Joint probability mass function

A two-dimensional random walk

3.3 The continuous case: Joint probability density function

Meeting times

4   Conditional distribution and independence

5   Expectations and moments

5.1 Mean, median and mode

A one-dimensional random walk

5.2 Central moments, variance and standard deviation

5.3 Moment generating functions

6 Functions of random variables

6.1 Functions of one random variable

6.2 Functions of two or more random variables

6.3 Two or more random variables: multivariate moments

7 Inequalities

7.1 Jensen inequality

7.2 Markov's inequality

7.3 Chebyshev's inequality

7.4 Cantelli's inequality

7.5 The law of large numbers

CHAPTER 3: SOME IMPORTANT DISTRIBUTIONS

1   Discrete case

1.1 Bernoulli trials

Binomial distribution – sums of binomial random variables

Hypergeometric distribution

Geometric distribution

Memoryless distributions

Negative binomial distribution

1.2 Multinomial distribution

1.3 Poisson distribution

Sums of Poisson random variables

1.4 Summary

2 Continuous case

2.1 Uniform distribution

2.2 Normal distribution

Probability tabulations

Multivariate normality

Sums of normal random variables

2.3 Lognormal distribution

Probability tabulations

2.4 Gamma and related distributions

Exponential distribution

Chi-squared distribution

2.5 Where discrete and continuous distributions meet

2.6 Summary

CHAPTER 4: IT IS ALL ABOUT DATA

1 An introduction to statistics

1.1 Different flavors of statistics

1.2 Trying to understand the true state of affairs

Parameters and statistics

Populations and samples

1.3 True state of affairs + Chance = Sample data

Random and independent samples

1.4 Sampling distributions

Formal definition of a statistics

Sample moments

Sampling from a finite population

Strategies for variance estimation - The Delta method

1.5 The Standard Error of the Mean: A Measure of Sampling Error

1.6 Making formal inferences about populations: a preview to hypothesis testing

2   Exploring data

2.1 Looking at data

2.2 Outlier detection and influential observations

2.3 Exploratory Data Analysis (EDA)

2.4 Box plots and violin plots

2.5 QQ plots

CHAPTER 5: PARAMETER ESTIMATION

1 Estimation Methods

1.1 Estimation by the Method of Moments

1.2 Estimation by the Method of Maximum Likelihood

2 Properties of Estimators

2.1 Unbiasedness

2.2 Consistency

2.3 Efficiency

2.34 Limiting distributions

2.5 Examples

Sample mean

Sample variance

Pooling variances

3 Confidence Intervals

3.1 Definitions

3.2 Method of finding confidence intervals in practice: Pivotal quantity

3.3 One-sample problems

Confidence Intervals for σ^2

Derivation of the chi-square distribution Properties of the chi-square distribution

Distribution of S^2

Independence of $\overline{X}$ and S^2

Known mean versus unknown mean

Confidence Intervals for μ

Derivation of the student t distribution Properties of the student t distribution

Known variance versus unknown variance

3.4 Two-sample problems

Confidence Interval for σ_1^2/σ_2^2

Derivation of the F-distribution

Properties of the F distribution

Distribution of S_1^2/S_2^2

Confidence Interval for $\mu_1 - \mu_2$

4 Bayesian estimation

4.1 Bayes' theorem for random variables

4.2 Post is prior × likelihood

4.3 Likelihood

4.4 Prior

4.5 Posterior

4.6 Normal Prior and Likelihood

5 In conclusion

CHAPTER 6: HYPOTHESIS TESTING

1 Terminology and Notation

1.1 Tests of Hypotheses

1.2 Size and Power of Tests

1.3 Examples

2 One-sided and Two-sided Tests

2.1 Case(a) Alternative is one-sided

2.2 Case (b) Two-sided Alternative

2.3 Two Approaches to Hypothesis Testing

3 Connection between Hypothesis testing and CI's

4 One-sample problems

4.1 Testing hypotheses about σ^2 when mean is known

4.2 Testing hypotheses about σ^2 when mean is unknown

4.3 Testing hypotheses about μ when σ^2 is known

4.3 Testing hypotheses about μ when σ^2 is unknown

5 Two-Sample Problems

5.1 Testing equality of normal means

5.2 Testing equality of binomial proportions

5.3 Testing equality of sample variances

6 Selecting an appropriate test statistic: some guidelines

7 In conclusion

8 Course summary

Probability

Statistics

RULES $\longrightarrow$ DATA $\longrightarrow$ RULES

Given the rules,

Given only the data, guess what the rules were

describe the likelihood of events occurring

Uncertainty involved

Looking forward

Looking backward

Predict

Inferences

POPULATION