

INTRODUCTION TO BIOINFORMATICS (BIOMATH 01)

Bioinformatics 1 for undergraduates at US-VNU- Spring 2009– A Syllabus

Copyright by Author

Lecturer	Nguyễn V. M. Mãn, Ph.D. in Statistics
Working area	<i>Algebraic Statistics, Experiment Designs & Computational Biology</i>
Ehome:	www.cse.hcmut.edu.vn/~mnguyen
Email:	mnguyen@designcomputing.net

Prerequisites. Linear Algebra and Calculus in the first two years of any basic university program. It is good to know concepts of Biology, but it is not a compulsory need.

Audiences. Senior (4-th year) students in Math, Computing, Biology, Microbiology, Environmental Science, Pharmaceuticals or relevant subjects.

Description. This introductory Bioinformatics course covers topics:

- (1) Concepts and motivations for Bioinformatics and Life Sciences
- (2) DNA, RNA and Proteins: brief introduction
- (3) Mathematical and Statistical Tools for Bioinformatics, include: Sequence Alignment, Markov Chains and Hidden Markov Models (HMMs) for Sequence Alignment, and Parameter Estimation for HMMs.

This introductory course spans 4 credits or 60 lecture hours, brief durations is distributed as:
Theory- Exercises- Presentation= 40-10-10

Goals. After completing this course, students should demonstrate the following skills:

- Understand and use basic concepts of Stochastic processes and Computing
- To be able to employ modeling software as *OpenModelica* (2) in Computational Biology.
- Appreciate that Bioinformatics and more broader, Computational Biology are built up on Mathematics, Computing and Statistics.

Textbook.

- (1) *Introduction to Mathematical Methods in Bioinformatics*, Alexander Isaev, **Springer**, Germany 2004 [main book]
- (2) *Basic Bioinformatics*, S. Ignacimuthu, **Alpha Science**, UK, 2005 [main book]
- (3) *Statistics for Biology and Health- An Introduction*, Warren J. Evens, Gregory R. Grant, **Springer**, USA 2005 [extra reference book]

Homework and Exams.

- There will be 3 problem sets (Homework assignments A, B, C).
- There will be 2 in-class tests during the lecture hours (on weeks 6 and 13).
- There will also be a two hours final exam during final weeks.

The in-class tests and the final exam are open book and calculators are allowed.

- *Easy problems* are mostly from the textbook and reinforce the basic techniques.
- *More interesting problems.* These problems are not taken from the textbook. They are more difficult and are worth more points. When you are asked to show some fact, you are not expected to write a rigorous solution in the mathematicians sense, nor a textbook solution. However, you should write a clear argument, using English words and complete sentences, that would convince a typical Calculus student. Run your argument by a classmate; this is a good way to see if your argument is reasonable.
- The more interesting problems are indicated by symbol * in front.

Grading. To pass two midterm tests (weeks 6 and 13 of the semester), contributing 50% of the final evaluation you should try your best to solve and to gain knowledge of Problem Sets A, B, C. In particular, Problem Set A for Test 1 and Problem Set B for Test 2.

* The final exam accounts for the remaining 50%, mostly from Problem Sets B and C.

ACTIVITIES	POINTS	Due week
Two In-class 1 Hour Tests	25 each	Weeks 7 and 13
Final Exam (2 hour)	50	Week 17 or 18
Total	100	

Calendar.

Week (Dates) TOPICS KEY DATES (Theory-Exercises-Presentation= 40-10-10)

Review of basic facts.

1 (Sept 8–13) Brief Introduction to Bioinformatics and Life Sciences– Softwares (4-0-0)

2 (Sept 15–20) Review of Logic, Matrix Theory, Set and Graph Theory (2-2-0)

I: Overview of Life Sciences- History- Scope and Importance.

3 (Sept 22–27) [History] Aims and Tasks- Applications of Bioinformatics (4-0-0)

4 (Sept 29– Oct 3) [Bio 1A] Key components of *Life Sciences*: DNA, RNA & Proteins (2-0-2)

5 (Oct 6–11) [Bio 1B] Introduction to Biological Sequences (cont.) (2-0-2)

II: Fundamental Mathematical and Statistical Tools for Bioinformatics.

6 (Oct 13–18) [Math 1A] Sequence Alignment (3-1-0)

7 (Oct 20–25) [Math 1B] Multiple Sequence Alignment (2-2-0)

One hour Test 1

8 (Oct 27– 31) [Probability] Random Variables and derived concepts– Solution of Test 1 (2-2-0)

9 (Nov 3– 8) [Stochastic processes] Poisson process and Markov Chains (4-0-0)

10 (Nov 10– 15) [Computing 1A] Markov Chains and Hidden Markov Models (HMMs) (2-2-0)

11 (Nov 17– 22) [Computing 1B] The Viterbi Algorithm in HMMs (4-0-0)

12 (Nov 24– 29) [Statistics] Parameter Estimation for Hidden Markov Models (3-1-0)

III: Presentation section: Biology and Ecology- is there a close connection?

13 (Dec 8– 13) **Presentation 1** – *OpenModelica* modeling soft for **Predator-Prey** model

One hour Test 2 (0-1-3)

14 (Dec 15– 20) **Presentation 2+ 3** *OpenModelica* modeling soft for HMMs (0-0-4)

15 (Dec 22– 27) Final Review– Student Presentation 3 (1-0-3)

Note: The grader is expected to return his grading evaluation of students on weeks 9 and 15 for two in-class tests.

REFERENCES

- [1] A.K. Basu, *Introduction to Stochastic Processes*, Alpha Science 2005
- [2] OpenModelica project, Sweden 2006, www.ida.liu.se/~pelab/modelica/OpenModelica.html
(document)
- [3] Bernd Sturmfels, *Solving Polynomial Systems*, AMS, 2002
- [4] Lior Pachter and Bernd Sturmfels, *Algebraic Statistics for Computational Biology*,
<http://bio.math.berkeley.edu/ascb/>
- [5] Zhang, Ma and Wu, *A compartmental model for analysis of SARS transmission patterns and outbreak control measures in China*, Applied Mathematics and Computation, 162, 2005
- [6] Sudhir Gupta. *Balanced Factorial Designs for cDNA Microarray Experiments*, **Communications in Statistics: Theory and Methods**, Volume 35, Number 8 (2006) , pp. 1469-1476
- [7] Computer Algebra System for polynomial computations, Germany 2006 www.singular.uni-kl.de/
- [8] Glonek G.F.V. and Solomon P.J., *Factorial and time course designs for cDNA microarray experiments*, Biostatistics 5, 89-111, 2004