STATISTICS FOR BIOINFORMATICS(MTH g 340)

The ambitious aim of this four-credit one-semester course is to introduce students to both conceptual basics and algorithms applied in Computational Genomics and its Applications (such as Drug Discovery). It is viewed as a conceptual intro to more technical courses. Some of the material must become a working instrument in further work, while the rest is introduced for orientation in the multidisciplinary field of Bioinformatics. Basic relevant concepts from Probability, Stochastic processes and fields, Information Theory, Statistitics and Experimental Design will be introduced by examples from Molecular Biology, Genomics, Population Genetics and Drug Discovery with an outline of algorithms and software.

Textbook : "Bioinformatics" by A. Polanski and M. Kimmel, Springer, 2007.

Supplementary preliminary texts are:

`Intro to the Practice of Statistics' by D.Moore and G. McCabe, 5th edition, Freeman, N.Y., 2006,

The overview can be downloaded from:

www.math.neu.edu/Malioutov/IPS5e ch0x.ppt

x=0, 1,…,16;

`Stats Problems in Genetics and Molecular Biology' by N. Drinkwater and C. Denniston, which can be downloaded from www.math.neu.edu/Malioutov/DrinkwaterLectures.pdf

Handouts of relevant material from additional sources will be provided.

Two labs on elementary statistical simulation, estimation and testing based on the package SPSS (or EXCEL)

and two MATLAB labs on HMM and ML-based distances between sequences are part of homework.

Time and Place: Tu. and Th. , 7.30-9 p.m. , 544 Nightingale Hall

.**Instructor**: Prof. Mike Malioutov, Mathematics Dept., NU, Phone: 373 5650, office: 545 Lake Hall, e-mail: M.Malioutov@neu.edu,

Office hours TBA or by appointment via phone or e-mail..

home page: http://www.math.neu.edu/~Malioutov

Tutoring for weaker students might be organized by Vladimir Ufimtsev (Ufimtsev.V@neu.edu)

Grade: 20% weekly homework, 40% two best tests or projects, 40% final test.

Optimistic Syllabus (some parts may be skipped because of lack of time)

Descriptive Stats, one or two variables,

IA. Survey of Distributions.

IB. Conditional Probability and Bayes rules.

IC. Distribution functionals: Expectation, Variance, Correlation, Entropy,Divergence=Relative Entropy.

ID. Approximations: Law of Large Numbers, Central Limit Theorem, Poisson approximation, Large Deviations.

IE. Markov and stationary processes and fields.

IF. Stochastic Simulation.

IIABasics of Experimental Design..

IIB. Estimation: Maximum Likelihood, BIC, Jaynes principle, Minimum Description Length, Regression Models. Optimal Design for Estimation.

IIC. Testing Hypotheses or Discrimination between models: Likelihood Ratio Test, GLRT, Bootstrap and Permutation test,

Error Probabilities, Power of a test. Optimal Design for Discrimination and Screening, Sequential Design, ANOVA and Mixed Models.

IIIA. Monte Carlo, MCMC, Gibbs Sampler, Genetic Algorithms.

IIIB. The EM algorithm

Homework (Printouts from Moore&McCabe will be distributed, any ½ of problems solved is enough)

First portion:

Chapters 1: Looking at Data-Distributions

Section 1.1	page 25: 1, 2. 7, 15, 17, 18, 20, 21, 22, 24, 37
Section 1.2	page 56: 41-45, 47, 58, 59, 64
Section 1.3	page 86: 86, 89, 92-107, 110-114

Chapter 2: Looking at Data-Relationships

Section 2.1	page 112: 1, 3, 6, 11, 20
Section 2.2	page 127: 21, 25, 26, 29, 32, 37
Section 2.3	page 146: 42, 43, 49, 51, 55, 58
Section 2.4	page 164: 63, 65, 67
Section 2.5	page 179: 86, 87, 92-94

Chapter 3: Producing Data

Section 3.1	page 197: 3-7
Section 3.2	page 210: 9-11, 25, 34
Section 3.3	page 225: 36, 38, 55, 56
Section 3.4	page 240: 62, 63, 66, 67, 69, 72

Chapter 4: Probability

Section 4.1	page 257: 2
Section 4.2	page 271: 11, 13, 14, 22, 23, 24, 27, 28, 30, 31
Section 4.3	page 287: 42
Section 4.4	page 305: 59, 74, 76, 80, 82
Section 4.5	page 323: TBA

Chapter 5: Sampling Distributions

Section 5.1	page 351: 1, 2, 8, 15, 20, 23, 25
Section 5.2	page 369: 28, 30, 34, 36, 37, 39, 42, 43

Chapter 6: Introduction to Inference

Section 6.1	page 396: 1, 5, 13, 15, 22, 24, 28
Section 6.2	page 417: 34, 35, 37, 38, 39, 40-47, 51, 54-56, 62-64, 68
Section 6.3	page 428: 74, 76, 77, 82, 88
Section 6.4	page 439: TBA

Chapter 7: Inference for Distributions

Section 7.1	page 471: 1, 5, 6, 11, 12, 16-18, 21, 26, 27, 33-36
Section 7.2	page 505: 54, 55, 58, 62, 68, 69, 74, 82, 83

Chapter 8: Inference for Proportions

Section 8.1	page 549: 1, 5, 7, 11-13, 15, 21, 22, 24, 26
Section 8.2	page 566: 31, 36, 38, 43, 47, 50, 51, 53

Chapter 12: One-Way ANOVA

Section 12.1, 2	page 752: 6-10, 13, 15, 17, 19, 20, 26, 31, 39, 41