BIOE 310 – Computational Tools for Biological Data

Credit Hours: 3 credit hours, 3 lecture hours, 0 lab hours

Instructor: Sergei Maslov

Required Text/Resources: None. Recommended textbooks:

- Montgomery, DC and Runger, GC: Applied Statistics and Probability for Engineers. 4th to 6th Editions .(2006-2014), John Wiley & Sons, Inc.
- Vidakovic, B: Statistics for Bioengineering Sciences with MATLAB and WinBUGS Support (2011) Springer, New York
- Pevsner, J: Bioinformatics and functional genomics, 3rd edition (2015), Wiley-Blackwell
- WJ Ewens and Grant, GR: Statistical Methods in Bioinformatics, 2nd ed (2005), Springer.

Specific Course Information:

- **Course Catalog Description:** Fundamental and applied statistics, including probability distributions, parameter estimation, descriptive statistics, hypothesis testing, and linear regression. Statistical methods in genomics including sequence analysis, gene expression data analysis, human genomic variation, regulatory genomics, and cancer genomics**.**
- **Prerequisites and/or Co-requisites:** BIOE 210 or MATH 257. Students in the BS-BIOE must take BIOE 210 as the prerequisite.
- Required Course

Course Learning Goals/Objectives: At the end of the course, the student should:

- Apply core concepts of probability, including conditional probability and Bayes' theorem, to model biological variability
- Select and interpret appropriate probability distributions for biological data
- Estimate parameters and quantify uncertainty using confidence intervals and hypothesis testing
- Use maximum-likelihood estimation and goodness-of-fit tests to assess model validity
- Perform correlation and regression analyses to identify relationships among variables

Brief list of topics covered:

- Week 1 – Introduction to probability and statistics in bioengineering; sample spaces and events
- Week 2 – Definitions of probability (statistical, logical); Venn diagrams; combinatorics
- Week 3 – Conditional probability; independence; circuit problems
- Week 4 – Bayes' theorem; Simpson's paradox
- Week 5 – Discrete random variables; PMF, CDF, mean and variance; binomial distribution
- Week 6 – Poisson and geometric distributions; application to genome assembly
- Week 7 – Continuous random variables; exponential and uniform distributions
- Week 8 – Gamma and Erlang distributions; normal distribution and z-scores
- Week 9 – Central Limit Theorem; sample means and variances

- Week 10 – Parameter estimation: method of moments, maximum likelihood
- Week 11 – Confidence intervals; t-distribution; hypothesis testing (one-/two-sample)
- Week 12 – Goodness-of-fit tests; chi-square tests; multiple testing corrections
- Week 13 – Linear and multiple regression; logistic regression
- Week 14 – Dimensionality reduction: principal component analysis (PCA)
- Week 15 – Applications of statistical analysis to expression data
- Week 16 – Network analysis. Course review for the final exam