Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

SLUkurs

Syllabus

PFS0008 Statistical Methods in Bioinformatics 1, 7.5 credits

Syllabus approved

2003-01-08

Subjects

Statistics, Computer Science and Systems Science

Education cycle

Third cycle

Grading scale

Pass / Failed
The requirements for attaining different grades are described in the course assessment criteria which are contained in a supplement to the course syllabus. Current information on assessment criteria shall be made available at the start of the course.

Language

English

Prior knowledge

Basic courses in mathematical statistics with applications for students in biology (5 credits)

Objective, including learning outcomes

Bioinformatics does not aim to lay down fundamental mathematical laws that govern biological systems parallel to those laid down in physics. Such laws, if they

exist, are a long way from being determined for biological systems. Instead, at this stage the main utility of mathematics in the field is in the creation of tools that investigators can use to analyze data. For example, biologists need tools for the statistical assessment of the similarity between two or more DNA or protein sequences, for finding genes in genomic DNA, and for estimating differences in how genes are expressed in different tissues. Such tools involve statistical modeling of biological systems, and it is the goal of the course to introduce probability, statistics, and stochastic processes in the context of bioinformatics.

Why are probability and statistics so important in bioinformatics? Bioinformatics involves the analysis of biological data. Many chance mechanisms are involved in the creation of these data, most importantly the many random processes inherent in biological evolution and the randomness inherent in many sampling processes. Stochastic process theory involves the description of the evolution of random processes occurring over time or space.

The aim is to give an introductory account of some of the probability theory, statistics, and stochastic process theory appropriate to computational biology and bioinformatics. The focus is, as stated above, on probability, statistics, and stochastic processes. We do, however, discuss aspects of the important algorithmic side of bioinformatics. In particular, the dynamic programming algorithm is introduced.

Content

· Probability theory Discrete random variables, some discrete probability distributions, continuous random variables, mean and variance, some continuous distributions, multivariate random variables, marginal and conditional distributions, covariance and correlation, maximum of random variables.
· Statistical inference Estimation, Hypothesis testing (parametric and non-parametric)

· Stochastic processes Poisson process, Markov chains
· Analysis of DNA and protein sequences

Requirements for examination

Exercises and written examination.

Additional information

Responsible department

Department of Forest Economics