



# Role of the Mathematical Sciences in Bioinformatics Education

## My own Perspective!

Susmita Datta  
Professor, Graduate Director &  
Distinguished University Scholar  
Dept. of Bioinformatics & Biostatistics  
University of Louisville



[www.susmitadatta.org](http://www.susmitadatta.org)  
email: [susmita.datta@louisville.edu](mailto:susmita.datta@louisville.edu)

- \* Developed Interdisciplinary Bioinformatics programs and Bioinformatics concentrations within the Biostatistics PhD Program.
- \* Bioinformatics – Finding Information from high dimensional bio-molecular data with the help of various “informatics” techniques available through various fields such as applied mathematics, computer sciences, physics and statistics.

\*Mathematical sciences are involved from the very beginning (design of experiment, wishful thinking?), representing the data in terms of numbers. For example,

1) Image data available from microarray gene expression data or protein array data, we do image analysis and find expression values in terms of numbers. We use statistical techniques to properly normalize the data, predict the missing values of the data and also, filtering of the data.

- 2) We then try to find patterns in the high dimensional data. We can use fine graphics.
  
- 3) However, due to massive nature of the data it is better to use data mining techniques e. g. Cluster analysis and classification analysis. We try to find appropriate data mining methods to represent the data in a meaningful pattern. These techniques belong to the area of mathematical sciences mentioned before. The results of such exploratory analysis often provide feedback to the scientists to explore more.

- 4) Next, we try to build deterministic (differential equation type models) or stochastic models (regression models) to explain the systems at work. Statistical models can provide predictive performances of the molecular profiles.
  
- 5) We also find important features of the data which can explain certain phenotypes or disease statuses etc. We develop tests of hypotheses for statistical inference which leads to deeper biological consequences.

\*Big data (biological in case of bioinformatics)

Complex molecular data along with other  
electronic medical records!

Analysis poses big challenges and opportunities

\*Different disciplines within mathematical  
sciences will have to work even more closely  
than ever before.

## Essential Skills

- \* Compute cluster or cloud Computing
- \* Demands for much improved statistical computational skills (including parallelization) which can make the program run efficiently and much faster. Multiple programming languages C++, R, Python and Java are also helpful especially for students pursuing bioinformatics emphasis.

\*Should we learn statistical inference well enough?

“Yes”

\*However, the syllabus must include some chapters on

“High Dimensional Data analysis”.

\*Some molecular biology and biochemistry

\* Duration of graduate work?????



Are the differences thinning between different fields of mathematical sciences through bioinformatics?

- \* “Yes”. Data Science???
- \* Statisticians are getting better with algorithms. Computer scientists/applied mathematicians are using a lot of statistical methods

\*We are being better in maintaining our data security

\*We are getting better to maintain lab-book so that we know the details regarding the data analysis so that we can reproduce the results.

## Get better in **hard skills**:

- \* Technical ability
- \* Computational proficiency

## Get better in **soft skills**:

- \* Communication
- \* Patience
- \* Teamwork
- \* Time management

**Thank you for your attention!**