A new look at flow cytometry

Cliburn Chan
Duke University
8 July 2007
Presentation at Annual Scientific Meeting
Mt Sinai Hospital, New York

Flow cytometry today

- I/O (FCS, CSV)
- Graphics (Histogram, 2D plots)
- Statistics (Counts, summary stats)
- Gating
  - manual
  - ad-hoc, based on “expertise”
  - “fragile”
    - difficult to replicate results from lab to lab
    - difficult to compare analysis from lab to lab

Our research interests

- Each event is a d-vector ($d \sim 2$ to 20)
- Collection of n such events constitutes one data set ($n \sim 100{,}000$ to $1{,}000{,}000$)
- Treat as random samples from some unknown multivariate density
- Can we map features of this density to biological concepts like cell subsets and activation states?

Research directions

- Dimension reduction by projections
  - PCA
  - ICA
  - Projection pursuit
- Bump hunting with kernel density estimation
- Bayesian mixture models with MCMC
- Optimization approaches
  - Variational Bayes
  - Expectation propagation

What we desire

- More flexible I/O options
  - FCS, CSV, XLS, XML, HDF, RDBMS
- 1D, 2D, 3D graphics and animations
- Integration of statistics and graphics
- Multi-stage data analysis
- Extensible via plug-ins
- Easy to use!

Example

- Bayesian mixture model
  - On original data
  - On projected data
  - On subsample of projected data
- Graphics for
  - trace plots
  - confidence ellipses
  - decision boundaries
  - classification of points based on posterior

Make it simple and powerful

- Simple menu
- Hierarchical tree control
  - Can visualize history of data transformations
  - Can do sequential transformations
  - Can copy, move, delete statistical transformations
- No programming required
- Sensible defaults provided wherever possible

Make it easy to develop for

- Plug-in architecture
  - I/O
  - Visualization
  - Statistics
  - Projections
- Glue
  - Python
  - Fortran
  - C/C++
  - R

Tour of flow software (warning: may crash and burn)

- FCS
  - histogram
  - 2D plots
    - dot
    - contour
  - 3D plots
  - PCA
  - kmeans on original
  - kmeans on proj
- HDF KDE
  - vector sigs
  - 3D plots with z
- HDF Bayes
  - trace
  - confidence intervals
  - vector sigs
  - 3D plots with z

Acknowledgements

- Jacob Frelinger
- Mike West
- Tom Kepler