Statistical Methods for Flow Data

Joseph B. Kadane
Carnegie Mellon University
Department of Statistics and
Software Engineering Institute
kadane@stat.cmu.edu
Outline

1. The Issue
2. Bayesian Techniques
3. Advantages of Bayesian Techniques
4. Conclusion
1. The Issue

- Existing logistic regression is described in Marc Kellner’s presentation

- With 200 odd observations in a ten-dimensional space of explanatory variables, the data can be sparse. There are three sorts of responses to this:
  
  (a) reduce the space by deleting explanatory variables
  
  (b) collect an order of magnitude more data
  
  (c) use Bayesian methods to smooth the estimates
2. Bayesian Techniques

The logistic regression defines a likelihood, that is a probability distribution of the data (which is 1’s and 0’s, scans and non-scans) given the explanatory variables and the (uncertain) weights. The remaining ingredient is a prior distribution on the weights, found by interrogating one or more experts, in a process known as elicitation.

(For more on this see Elicitation, by Garthwaite, Kadane and O’Hagan www.stat.cmu.edu/tr/tr808/tr808.html).

With these two ingredients, one can compute the posterior distribution on the weights. The posterior is proportional to the prior times the likelihood.
3. Advantages of Bayesian Techniques

- pulls in discrepant and unreasonable estimates of weights
- posterior variances guide sophisticated statistical design of experiments in deciding what additional data to gather or elicitation to ask about
- conceptually easy to extend from binomial to multinomial data
4. Conclusion

Bayesian methods offer a reasonable way forward to make the logistic regression approach to scan data stable and operationally feasible.