Statistical Mixture Modeling for Cell Subtype Identification in Flow Cytometry

Cliburn Chan[1], Feng Feng[1], Thomas B Kepler[1,2,3] & Mike West[3]

September 30, 2007

Abstract

Statistical mixture modeling provides an opportunity for automated identification and resolution of cell subtypes in flow cytometry data. The configuration of cells as represented by multiple markers simultaneously can be arbitrarily well modeled as a mixture of Gaussian distributions in the dimension of the number of markers, and fitting such models is enabled using standard Bayesian statistical approaches and Markov chain Monte Carlo computations. Cellular subtypes may be related to one or multiple components of such mixtures, and fitted mixture models can be evaluated in the full set of markers as an alternative, or adjunct, to traditional subjective gating methods that relies on choosing one or two dimensions. This paper describes studies of statistical mixture modeling of flow cytometry data, and demonstrates their utility in examples with data from human peripheral blood samples using four markers.

1 Introduction

One of the fundamental uses of flow cytometry is the identification and quantification of distinct cell subsets with phenotypes defined by the density of cell surface or intracellular markers. Ideally, such a biological classification should be *objective*, *stable* and *predictive*[1].

Objectivity and stability are problematic with the traditional approach in which samples are sequentially gated in 1- or 2-dimensional plots. In particular, the choice of which sequence of markers to gate on and where to draw the gates depends on expertise and is highly subjective. This makes it difficult to replicate the cell subset identification procedure across different laboratories. The problem is compounded with multi-color flow cytometry, since the number of possible gating sequences rises rapidly with the number of channels used.

Statistical mixture models are very widely used in scientific problems where objects represented in several or many dimensions are to be clustered or classified. One appeal of mixture models is the ability to represent essentially any observed data distribution to a high degree of accuracy [2, 3]. Some useful background on methodology and ideas underlying mixture models, as well as some specific applications, appears in [4, 5], and a range of examples in biomedical problems that provide useful insights into various applied aspects appear in [6, 7, 8], for example. In some applications the identification of underlying scientific meaning of specific mixture components is of relevance, whereas in others mixtures are of interest primarily as flexible data smoothers. Our interest here is the potential utility of multivariate mixture models for flow cytometry cell subtype identification,

[1]Center for Computational Immunology, Department of Biostatistics and Bioinformatics, Duke University, Durham, NC 27705, USA

[1]Department of Immunology, Duke University, Durham, NC 27708, USA

[2]Department of Statistical Science, Duke University, Durham, NC 27708, USA

so that the resolution of mixture components is of interest. Some of the potential utility is that of directly modeling and resolving flow cytometry data for all the markers simultaneously, so that determining a 1- or 2- dimensional marker sequence for gating is unnecessary. Further, the analysis of mixtures using current computational statistical technology is automatic, and will apply in as many dimensions as we have markers.

We describe our studies of statistical mixture modeling using Gaussian mixtures for flow cytometry data densities. We use a Bayesian mixture modeling approach that is effectively standard modern statistical methodology, and fit such models using Markov Chain Monte Carlo (MCMC) computational algorithms [9, 2]. The basic mixture model framework seems apt for modeling distinct cell subsets, each of which may be reflected in possibly more than one of the multivariate Gaussian components. The analysis is open to exploiting biological expert knowledge in the specification of priors where available, or alternatively can be run in default, objective mode. The MCMC computations use a standard, efficient and flexible algorithm that requires little tuning, and that works well with complex multi-modal distributions such as routinely arise with flow cytometry recordings.

We demonstrate the application and utility of mixture modeling of flow cytometry data in a series of examples in which multi-modality arises naturally as a result of individual or groups of mixture components that map well to biologically relevant cell subsets. The examples involve analyses of four-color flow cytometry data of human peripheral blood samples. They demonstrate the methodology and suggest that Bayesian mixture models can be used to effectively identify cell subsets, improve specificity and remove outliers in such data, and to do so automatically. We also provide supporting software for others interested in such analyses.

2 Materials and Methods

2.1 Experiments

2.1.1 Cell preparation

Human whole blood was obtained from a healthy volunteer. The heparinized blood was treated with FACS lysis solution (BD Pharmingen). The cell suspensions were washed by centrifugation (328Xg, 5 min, 4°C), and the cell pellet was re-suspended in FACS buffer.

2.1.2 Flow cytometry

Human blood cells (5 x 10^5 cells/test) re-suspended in FACS buffer were stained in 20 μl of appropriated mAb for 25 min at 4°C in the dark. Then the cells were washed by the addition of 500 μl of FACS buffer and centrifuged (328Xg, 5 min, 4°C). FITC-conjugated anti-CD3 (HIT3a), PE-conjugated anti-CD8 (HIT8a), PE-cy5-conjugated anti-CD4 (RPA-T4) and APC-conjugated anti-CD19 (HIB19) Abs were used to label the human blood cells. All the mAbs were purchased from BD Pharmingen. Flow cytometric acquisition was performed on a FACSCalibur (BD Biosciences).

2.2 Modeling

Representing the measured markers on a single cell as the $d-$dimensional vector x, we model flow cytometry data as a mixture of normals

$$x \sim \sum_{i=1}^{k} \alpha_j \, \mathcal{N}(x|\mu_j, \Sigma_j) \tag{1}$$

where: k is the number of mixture components, $\alpha_1, \ldots, \alpha_k$ are the mixing weights (or mixing probabilities) that sum to 1, $\mathcal{N}$ denotes the multivariate normal distribution, μ_j is the d-dimensional mean of mixture component j and Σ_j the corresponding $d \times d$ covariance matrix. In terms of the probability density function (p.d.f.), we have the underlying population p.d.f.

$$p(x) = \sum_{i=1}^{k} \alpha_j \; (2\pi)^{-d/2} |\Sigma_j|^{-1/2} \exp\left\{-(x-\mu_j)^T \Sigma_j^{-1} (x-\mu_j)/2\right\} \tag{2}$$

where $|\Sigma_j|$ the determinant of Σ_j. In any given $k-$component mixture, a unique model specification is achieved by the identifying constraint that the mixture probabilities are in decreasing order, $\alpha_1 > \alpha_2 > \cdots > \alpha_k$, and we use this here.

We aim to fit such a model to data $X = \{x_1, \ldots, x_n\}$ representing n measured marker vectors on n cells. Standard mixture model analysis augments the model with the underlying, latent mixture component indicators, with one such indicator for each cell, to generate an equivalent but more tractable specification. For each cell i, let $z_i = j$ represent the event that x_i arises from the j^{th} component of the mixture. Then $Z = \{z_1, \ldots, z_n\}$ is the set of (latent) component indicators for the cells. As in the EM algorithm, use of Z allows us to decompose the joint p.d.f. of the n observations in the more tractable conditional form

$$p(X|Z) = \prod_{i:x_i=1} f(x_i|\mu_1, \Sigma_1) \ldots \prod_{i:x_i=k} f(x_i|\mu_k, \Sigma_k) \tag{3}$$

That is, the data break up into k groups, and within each group there is a single normal distribution for that data subset. We use Gibbs sampling[10], one specific version of the general MCMC methodology, as the basis for computations that exploit this simplified structure. The Gibbs sampler generates sequentially simulated values of the full set of model parameters $\{\alpha_j, \mu_j, \Sigma_j : \; j = 1, \ldots, k\}$ and the underlying indicators Z. Repeat simulations delivers (large) samples from the implied posterior distribution for the Z and the parameters together, and this posterior sample provides the basis for all statistical inferences (e.g., [2]), as our examples below illustrate.

The complete analysis follows the details originally described in [9] for the standard Gibbs sampler in Gaussian mixtures. Additional components needed include prior distributions that are taken as the standard conjugate priors, namely

$$\begin{array}{rcl} (p_1, \ldots, p_k) & \sim & \mathcal{D}(\alpha_1, \ldots, \alpha_k) \\ \mu_i & \sim & \mathcal{N}(\zeta_i, \tau_i \Sigma_i) \\ \Omega_i & \sim & \mathcal{W}(r_i, W_i) \end{array} \tag{4}$$

where $\mathcal{D}$ is the Dirichlet distribution, the precision matrix $\Omega_i = \Sigma_i^{-1}$ and $\mathcal{W}(r, W)$ is the Wishart distribution. The Gibbs sampler also requires initialization and we do this with initial values of p, μ and Σ set using the proportions, centroids and sample covariance matrices calculated from the

clustering given by a k-means algorithm [11]. Alternative initializations based on EM are just as effective.

One aspect of evaluation of a fitted Gaussian mixture is the identification of the modes of the p.d.f. One efficient numerical strategy to locate and compute the modes of a mixture is to use multiple restarts of the Nelder-Mead algorithm[12], starting from each of the centroids μ_j.

2.3 Practicalities of model fitting

2.3.1 MCMC convergence

MCMC methods generate sequences of successively computed values of the model parameters and latent component indicators together. As in all areas of application of MCMC methods the theory underlying the model setup ensures us that these samples eventually represent random draws from the posterior distribution that encodes all inferences. In practice, there is no fool-proof method to confirm that convergence has been achieved, although many standard and well-used graphical and numerical heuristics are available [13]. The most useful, and usual, is simple inspection of trace plots - plots of successively simulated values of the selected parameters. Based on experience with simulated data sets we routinely run the MCMC sampler for several hundred initial observations before assuming that the sampled values have settled down to represent the posterior distribution; example trace plots of mixture probabilities demonstrate the use of this visual inspection (Figure 1).

2.3.2 Choice and interpretation of the number of mixture components

For a given data set, repeatedly fitting the mixture model with increasing numbers of components k can generate useful insights into the suitability of particular values of k. The practical reality of mixture modeling is that, beginning with $k = 1, 2, \ldots$, we expect to see an initial range of values over which the mixture model with increasing numbers of components represents substantial non-normal aspects of the observed data configuration. Increasing k further, however, will eventually begin to add additional components with low probabilities. These components can modestly increase the apparent fit of the model to the data, catering to idiosyncratic features involving a very small number of data points, while often not representing substantively meaningful components. This is an inherent feature of mixture modeling and all attempts to automatically estimate or choose k are impacted by it. From a practical perspective a reasonable strategy utilizes larger numbers of components and bears this "over-fitting" tendency in mind; a fitted model will be inspected and may aim to cut-back to components that appear to be substantively relevant.

Methods for treating k as a parameter to be estimated include AIC, BIC and other informational approximations to likelihoods. The graph in (Figure 2) is an example of BIC (also known as the Schwartz criterion [14]) and is quite typical of what is to be expected in practice. The BIC "optimal" value of k is the maximizing value. Whereas the curve increases quickly from small values of k, indicating that there is indeed mixture structure, it flattens and stays very flat after peaking as additional, low probability components come into play. From the viewpoints of scientific parsimony and relevance, cutting back to smaller values than the absolute peak is generally recommended. In this example, we are interested in the extent to which we can identify biologically meaningful cell subsets present in a human peripheral blood population using a mixture model, and the BIC plot suggests 10–15 components as a region where the information curve dramatically flattens off after a steep rise. Though the mode is near 20 components, it is flat between 15–25 and more where multiple low probability components are added. Evaluating and interpreting a mixture model with, say, 15 components may reveal much of the relevant cellular sub-types; repeating the analysis with,

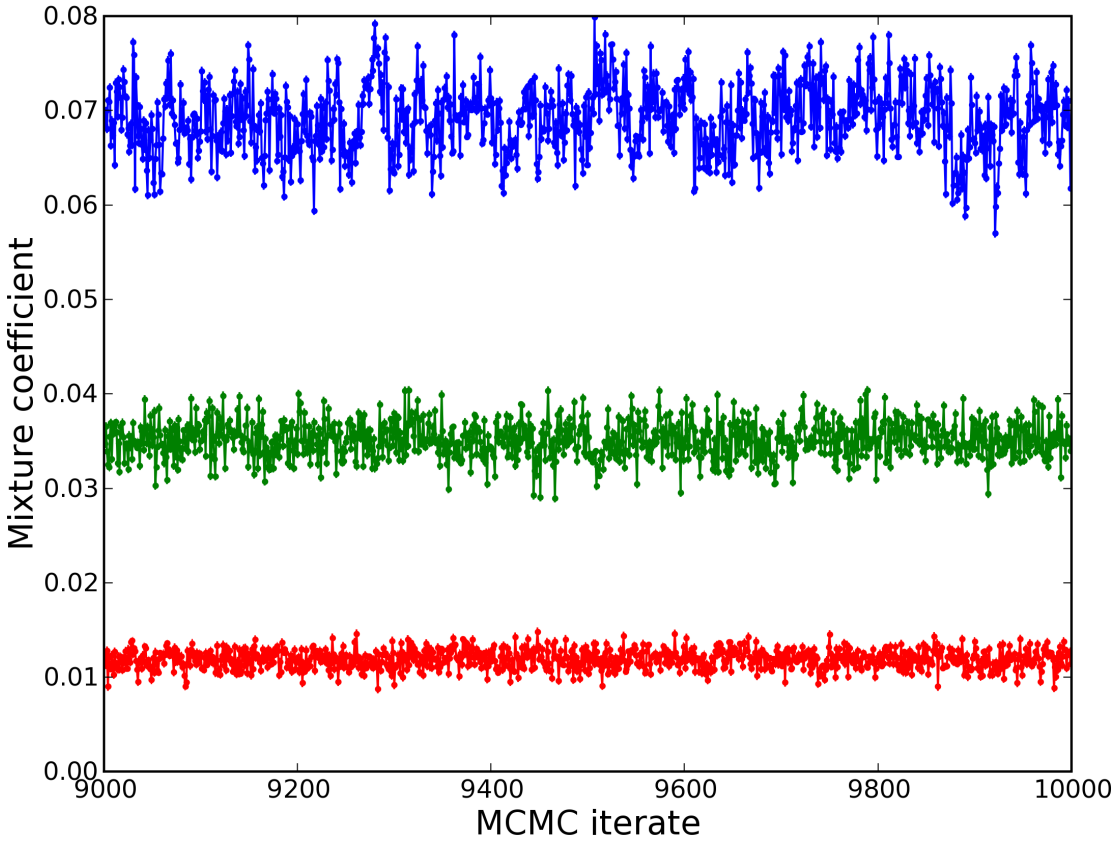

Figure 1: Trace plots for the proportion of the 5th (blue), 10th (green) and 15th (red) largest mixture components (π_i) over the last 1000 MCMC iterations suggesting convergence.

say, 20 components, and with due thought to the potential irrelevance of low probability estimated components, may then be viewed as a confirmatory step.

Beyond BIC and related methods, formal Bayesian analysis that treats k as a parameter and includes it within the MCMC analysis is available and quite widely used. The main class of models is the class of Dirichlet process mixtures [4, 5], nowadays very widely used in clustering and mixture deconvolution in machine learning and statistics. Though this inevitably faces the same issue of over-fitting and over-estimation of k, it allows and requires the specification of a prior parameter that can weigh against the data and cut-back to smaller values of k. We do not develop this particular approach here, but it can provide a useful exploratory analysis as an initial step that suggests relevant ranges of k.

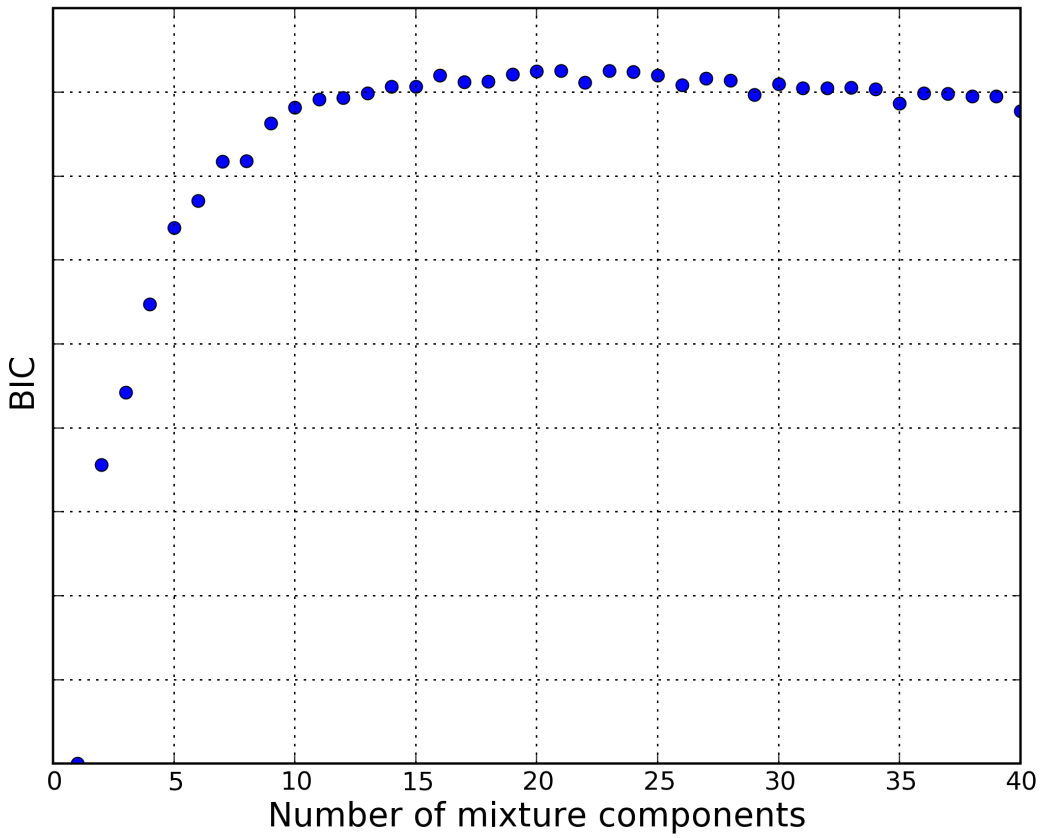

Figure 2: Plot of the BIC against number of mixture components.

3 Results

3.1 Identifying cell subsets with mixture components

3.1.1 Filtering irrelevant mixture components

As we have discussed above, in a model fitted with a larger value of k and for purely statistical reasons, very low probability mixture components may be of no interest to an experimentalist. In addition, components that model background noise are typically of no interest. The challenge is then how to filter out such irrelevant mixture components.

A simple definition of a component that models background noise is that the *density* of that component is below some factor α of the density of a single component model (i.e., using a single multivariate Gaussian to fit the entire data set.) A measure of the *volume* of a component is given by the square root of the covariance matrix determinant. Using a conservative value of $\alpha = 1$, if the density of a mixture component is below that of the single component model, that component is rejected as modeling background noise.

3.1.2 Grouping components to form cell subsets

Since the hypersurfaces of constant density for a normal distributions are ellipsoidal, cell subsets with an asymmetrical density may require more than one normal component for a good fit. In other words, a cell subset may be identified with a group of mixture components, rather than a single component.

We assume that a subset consists, by definition, of a single cell type and that its density is unimodal. The density may not be Gaussian, however, and thus may be fit with multiple mixture components as described above. We therefore find the modes in the density by numerical optimization of the mixture density, starting from the means of the mixture components found. Modes that fall outside the range of the data are discarded. While this strictly only detects modes within the convex hull of the component means, and modes that arise from summing overlaps between two or more mixture components will not be detected, such modes do not plausibly correspond to distinct cell subsets and are not considered. Mixture components with a common mode are therefore merged for the purposes of assigning a cell type label.

There is another complication for log-transformed data in that low and negative value fluorescence intensities in any channel will result in the data piling up against an axis. As a consequence, the resulting distribution is markedly non-normal, and typically an extra mixture component is necessary to model this. We therefore also merge mixture components that are smeared against an axis with its nearest contiguous neigboring cluster. Doing this, we find that the number of cell subsets is typically smaller than the number of mixture components.

3.1.3 Identifying cell subsets in human peripheral blood

As an example, we will describe the procedure for 15 mixture components, which is in the early plateau of the BIC plot shown in Figure 2. After filtering out very low density components (8,13–15), we are left with 11 mixture components (Figure 3. Of these, it was clear from the FSC/SSC plot that component 5 was debris and 11 consisted of doublets. Components 1 and 4 share a mode, and hence are considered as a single subset. Component 2 clusters together with 1 and 4 but required a separate component because it was smeared against the CD19 axis. Together 1, 2 and 4 form the granulocyte subset. Component 10 was identified as the monocyte subset. Components 3, 6, 9 and 12 appear to be lymphocytes, with 3 being $CD3^{+}CD4^{+}$ (CD4 lymphocytes), 6 being $CD3^{+}CD8^{+}$ (CD8 lymphocytes) and 9 being $CD19^{+}$ (B lymphocytes). Component 12 was $CD3^{+}$ and $CD8^{-}$ to $CD8^{\mathrm{dim}}$, and may the NK cell population. We were unable to identify component 7 (dead cells?), which was negative for CD3, CD4, CD8 and CD19. The final set of putative cell subsets with their frequencies is shown in Table 1). The analysis for 20, 25 and 30 mixture components was very similar, as the additional components mostly represented either noise, events piled up against some axis or multiple components with a common mode. The only essential difference was that with a larger number of components, the component corresponding to 12 was split into at least 2 components, $CD8^{\mathrm{dim}}$ and $CD8^{-}$.

3.2 Improving classification accuracy

3.2.1 Thresholding to reduce false positives

The allocation of flow cytometry events to one of the six identified groupings was done by assigning each event to the component with the highest posterior probability. While this procedure minimizes mis-classification, we can improve specificity by setting a threshold for the posterior probability. If the highest posterior probability for an event falls below this threshold, the event is not assigned

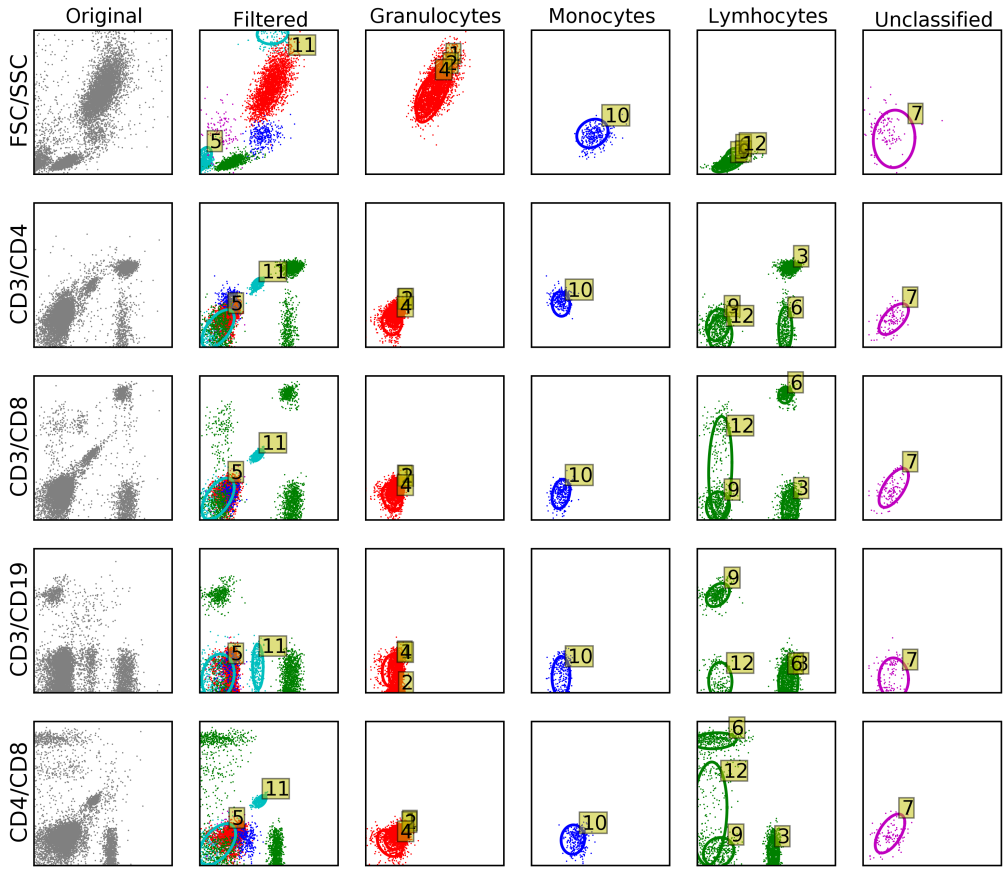

Figure 3: Column 1 shows the original flow cytometry data. Column 2 shows the color-coded mixture components filtering out noise and out-of-range mixture components, with red for granulocytes, blue for mononuclear cells, green for lymphocytes and yellow for aggregates, dead cells and debris. Columns 3–6 show the sub-components of the granulocyte, monocyte, lymphocyte and unclassified groups. Ellipses and numbered yellow labels show the 67% coverage set for each component. Every plot in a row is on the same scale.

Group	Percent	Classification
1	0.00	Granulocytes
2	0.00	Mononuclear cells
3	0.00	$CD3^{+}CD4^{+}$ lymphocytes
4	0.00	$CD3^{+}CD8^{+}$ lymphocytes
5	0.00	B lymphoyctes
6	0.00	$CD3^{-}CD8^{dim}$ NK cells

Table 1: Identification of mixture groups with putative cell subsets of interest showing proportion of cells in each subset.

to any component and treated as 'uncertain'(Figure 4). This is particularly useful if the purity of the cell subset identified is critical, as for example, in *cell sorting* applications of flow cytometry.

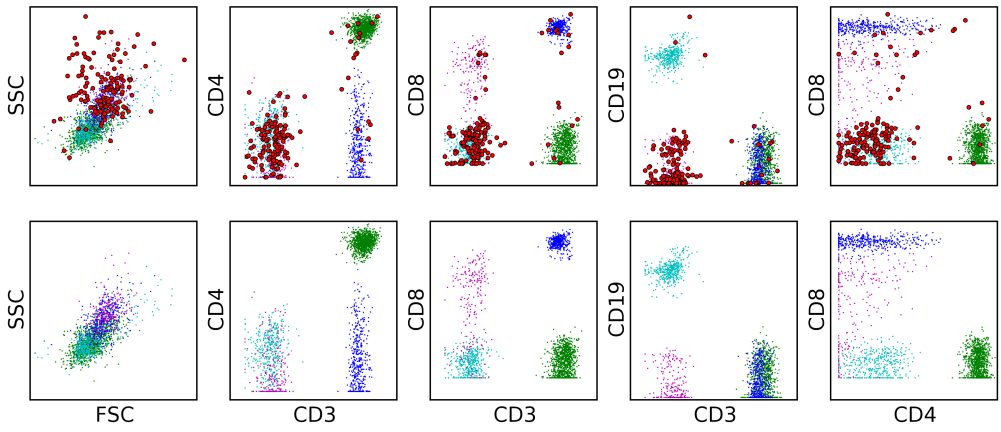

Figure 4: Using thresholds to increase specificity. The top panel shows the lymphocyte subsets in which events where the posterior probability of belonging to any lymphocyte component falls below 0.95 have been enlarged and colored red. The bottom panel shows the result after filtering out the red components. The following colors were assigned to the lymphocyte mixture components – $CD3^{+}CD4^{+}$ (green), $CD3^{+}CD8^{+}$ (blue), $CD3^{-}CD19^{+}$ (cyan) and $CD3^{-}CD8^{dim/-}$ (magenta). Most of the uncertainty is with the CD8-negative sub-population of the $CD3^{-}CD8^{dim/-}$ component, which overlaps the unclassified component.

3.2.2 Excluding outliers with coverage sets

Another advantage of Bayesian analysis for flow cytometry is the ability to exclude outliers, events that lie far away from the mean and hence may not be representative of any group. Such outliers can be rejected by filtering out events that fall outside some specified coverage level (or confidence region), which can be easily calculated from the mixture component's covariance matrix. For example, the ellipses corresponding to the 67% confidence regions are shown in the last 3 columns of Figure 3, and any other coverage level desired can just as easily be specified. Unlike the thresholding described above that filters out events which do not obviously belong to a particular group, use of coverage sets filters out events that are in some sense anomalous for the group they belong to.

3.3 Visual gating and mixture modeling

When cell subsets are identified by visual gating, events that do not belong to the cell subset but are fortuitously included are unavoidable since gates are specified only for a one- or two-dimensional slice of the data. Since mixture modeling is intrinsically multivariate, such events are naturally excluded as only events that are 'close' in *all* dimensions will belong to the same component. A simple example of this effect for the lymphocyte sub-population is illustrated in Figure 5.

However, visual gating and mixture modeling are not necessarily incompatible. For complex data sets, it may be beneficial to exploit available expertise by first performing visual gating to reduce the size and complexity of the data set, and then doing mixture modeling to identify and clean the remaining subsets.

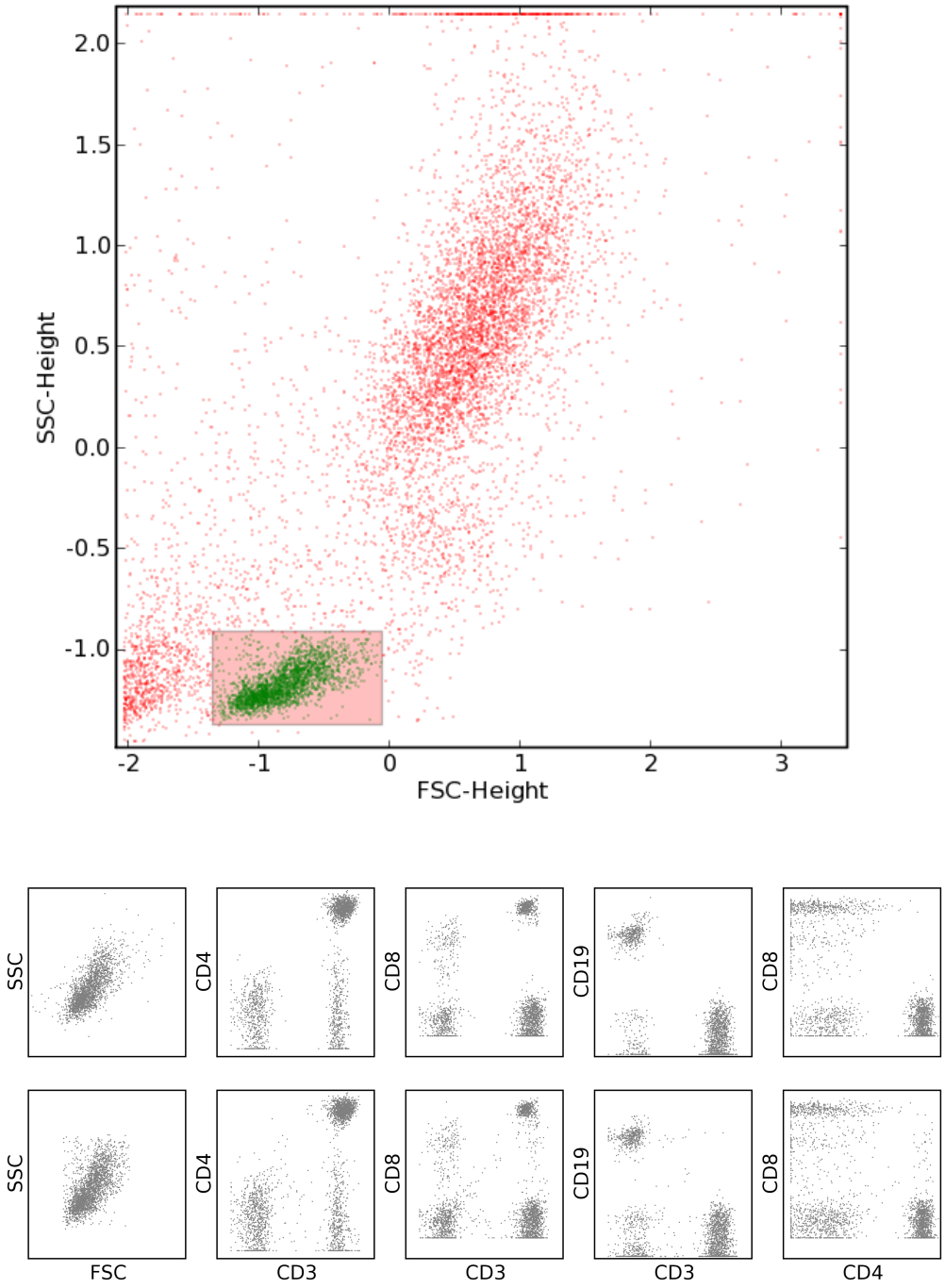

Figure 5: Comparing visual gating with mixture modeling. The top subplot shows a rectangular gate (green) on the lymphocyte sub-population on a dot plot of FSC against SSC. The bottom panel shows two rows of subplots. The top row shows lymphocytes gated visually, while the bottom shows the mixture components identified as lymphoyctes. Note that the model component events are slightly less noisy than the visually gated ones, which is most obvious in the CD4 vs CD8 subplot.

4 Discussion

We have shown in this manuscript that Bayesian mixture models can extract biologically meaningful components (cell subsets) from flow cytometry, by defining putative cell subsets as groups of mixture components fitting the following criteria:

1. each component must have a density greater than that of a single component model
2. each component must have a well-conditioned covariance matrix
3. the components share a common mode

It may also be necessary to merge components that result from events piling up against an axis, typically resulting from log transformation of the data. This is an artifact of the log transform, and disappears with linear FCS 3.0 data using the hyperlog [15] or logicle [16] transforms (data not shown).

Using these putative cell subsets, we show that the accuracy of event classification can be improved by thresholding on the posterior density of each event, or by selecting events from a smaller coverage set. Critically, this analysis also provides us with a robust statistical model of flow cytometry data that can potentially be used for the rigorous statistical comparison of two or more flow cytometry data sets. If flow cytometry data can be normalized in a standardized fashion, it should also be possible to build up a training set of flow cytometry samples, allowing future automated classification of cell subsets in a sample, as well as classification of entire data samples (e.g. as *normal* or *abnormal*).

Bayesian mixture models can be extended to an arbitrary number of dimensions, and we are actively researching its utility for the analysis of polychromatic flow cytometry. In practice, however, several challenges have to be overcome for this to be practical. We have described a simple but adequate strategy for determining the number of components, by simply adding components till the contribution of the last added component is negligible. With a larger number of dimensions and a correspondingly increase in the number of mixture components necessary, such an approach may be too inefficient. More advanced sampling methods can estimate the number of components directly[17, 18, 19, 20]. In high dimensions, it is also critical to develop more efficient samplers, and advances in this direction include gradient optimization [21] and the combination of variational optimization with MCMC [22].

We believe that Bayesian models are a highly promising approach to the automated or semi-automated analysis of flow cytometry data. With the increasing dimensionality and volume of flow cytometry data being generated, such an approach is likely to prove ever more useful and necessary.

Acknowledgments

Human blood cell data was kindly provided by Ms Jennifer Lonon and Dr Mary Louise Markert, Department of Pediatrics, Duke University Medical Center. We also thank Ms Janet Ottinger for many helpful discussions on flow cyometry analysis.

References

[1] R. M. Cormack. A review of classification. *Journal of the Royal Statistical Society. Series A (General)*, 134:321–367, 1971.

[2] Christian P. Robert. Mixtures of distributions: Inference and estimation. In *Markov Chain Monte Carlo in Practice*, pages 441–464. Chapman & Hall, London, 1996.

[3] D. Titterington, A.F.M. Smith, and U. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons, 1985.

[4] M. West. Modelling with mixtures (with discussion). In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors, *Bayesian Statistics 4*, pages 503–524. Oxford University Press, 1992.

[5] M.D. Escobar and M. West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90:577–588, 1995.

[6] D.A. Turner and M. West. Statistical analysis of mixtures applied to postsynaptic potential fluctuations. *Journal of Neuroscience Methods*, 47:1–23, 1993.

[7] D.A. Turner, Y. Chen, J. Isaac, M. West, and H.V. Wheal. Heterogeneity between non-NMDA synaptic sites in paired-pulse plasticity of CA1 pyramidal neurons in the hippocampus. *Journal of Physiology*, 500:441–461, 1997.

[8] M. West. Hierarchical mixture models in neurological transmission analysis. *Journal of the American Statistical Association*, 92:587–606, 1997.

[9] M. Lavine and M. West. A Bayesian method for classification and discrimination. *Canadian Journal of Statistics*, 20:451–461, 1992.

[10] Alan E. Gelfand and Adrian F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409, 1990.

[11] M. J. de Hoon, S. Imoto, J. Nolan, and S. Miyano. Open source clustering software. *Bioinformatics*, 20(9):1453–1454, June 2004.

[12] J.A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7:308–313, 1964.

[13] Mary Kathryn Cowles and Bradley P. Carlin. Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91(434):883–904, 1996.

[14] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[15] C. Bruce Bagwell. Hyperlog-a flexible log-like transform for negative, zero, and positive valued data. *Cytometry A*, 64(1):34–42, Mar 2005.

[16] David R Parks, Mario Roederer, and Wayne A Moore. A new 'Logicle' display method avoids deceptive effects of logarithmic scaling for low signals and compensated data. *Cytometry A*, 69(6):541–551, Jun 2006.

[17] Michael D. Escobar and Mike West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430):577–588, 1995.

[18] Sylvia Richardson and Peter J. Green. On bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B*, 59:731–92., 1997.

[19] Matthew Stephens. Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods. *Ann. Statist.*, 28(1):40–74, 2000.

[20] Peter J. Green and Sylvia Richardson. Modelling heterogeneity with and without the dirichlet process. *Scandinavian Journal of Statistics*, 28:355–375, 2000.

[21] Zhaohui Qin and Jun S. Liu. Exploring hybrid monte carlo in bayesian computation. Proceedings of ISBA 2000 - The Sixth World Meeting of the International Society for Bayesian Analysis, 2001.

[22] Nando De Freitas and Stuart Russell. Variational MCMC. Technical report, UC Berkeley, 2001.