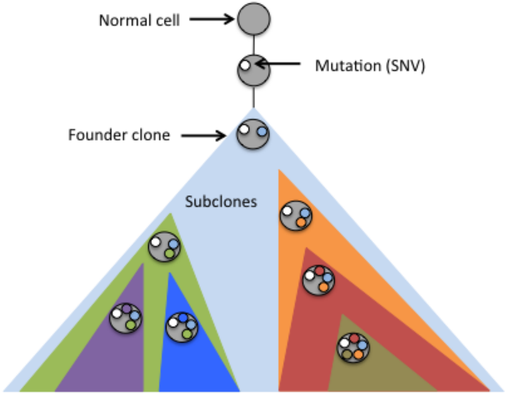
Nonparametric Clustering with Variational Inference for Tumor Heterogeneity

David Liu

Advisor: Prof Ben Raphael
Reader: Prof Erik Sudderth

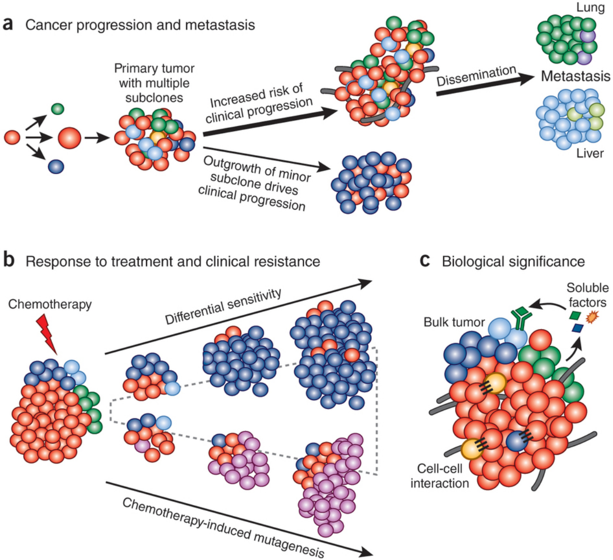
May 1, 2017

Cancer is an evolutionary disease.



[1] El-Kebir 2015

Clinical significance.


a Cancer progression and metastasis
Primary tumor with multiple subclones
Increased risk of clinical progression
Outgrowth of minor subclone drives clinical progression
Dissemination
Lung
Metastasis
Liver
b Response to treatment and clinical resistance
Chemotherapy
Differential sensitivity
Chemotherapy-induced mutagenesis
c Biological significance
Soluble factors
Bulk tumor
Cell-cell interaction

[1] Kleppe 2014

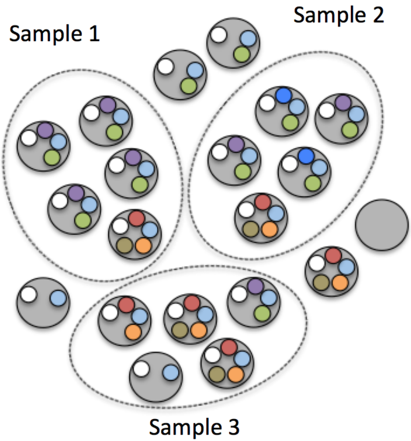
Clonal mixture from bulk-sequencing data.

- We observe DNA bulk-sequencing data.
    - Reference and variant reads.

Clonal mixture from bulk-sequencing data.

- We observe DNA bulk-sequencing data.
    - Reference and variant reads.
- Cluster mutations that occur with similar frequency.
    - Mutations from the same cluster should occur with the same frequency.

Multiple samples increase resolution.



Observed data

Given $n = 1, \ldots, N$ SNVs, and $m = 1, \ldots, M$ samples, we are given:

Observed data

Given $n = 1, \ldots, N$ SNVs, and $m = 1, \ldots, M$ samples, we are given:

- Variant reads, v_{mn}

Observed data

Given $n = 1, \ldots, N$ SNVs, and $m = 1, \ldots, M$ samples, we are given:

- Variant reads, v_{mn}
- Total reads, d_{mn}

Observed data

Given $n = 1, \ldots, N$ SNVs, and $m = 1, \ldots, M$ samples, we are given:

- Variant reads, v_{mn}
- Total reads, d_{mn}

So for each SNV, we can vectorize across samples:

Observed data

Given $n = 1, \ldots, N$ SNVs, and $m = 1, \ldots, M$ samples, we are given:

- Variant reads, v_{mn}
- Total reads, d_{mn}

So for each SNV, we can vectorize across samples:

$$\mathbf{d_n} \triangleq \begin{bmatrix} d_{1n} \\ d_{2n} \\ \vdots \\ d_{Mn} \end{bmatrix}, \quad \mathbf{v_n} \triangleq \begin{bmatrix} v_{1n} \\ v_{2n} \\ \vdots \\ v_{Mn} \end{bmatrix}.$$

Observed data

Given $n = 1, \ldots, N$ SNVs, and $m = 1, \ldots, M$ samples, we are given:

- Variant reads, v_{mn}
- Total reads, d_{mn}

So for each SNV, we can vectorize across samples:

$$\mathbf{d_n} \triangleq \begin{bmatrix} d_{1n} \\ d_{2n} \\ \vdots \\ d_{Mn} \end{bmatrix}, \quad \mathbf{v_n} \triangleq \begin{bmatrix} v_{1n} \\ v_{2n} \\ \vdots \\ v_{Mn} \end{bmatrix}.$$

Let $\mathbf{x}_n$ be general notation for $\{\mathbf{d}_n, \mathbf{v}_n\}$, where the use of the total or variant reads will be clear from context.

Assigning mutations to clusters

Suppose that each SNV $n \in \{1, ..., N\}$ belongs to a cluster $k \in \{1, \ldots, K\}$, $K \leq N$.

Assigning mutations to clusters

Suppose that each SNV $n \in \{1, ..., N\}$ belongs to a cluster $k \in \{1, \ldots, K\}$, $K \leq N$.

- Latent variables $\mathbf{z}_n$, a 1-of-K indicator vector that denotes the cluster assignment of SNV n to cluster k.

Assigning mutations to clusters

Suppose that each SNV $n \in \{1, ..., N\}$ belongs to a cluster $k \in \{1, \ldots, K\}$, $K \leq N$.

- Latent variables $\mathbf{z}_n$, a 1-of-K indicator vector that denotes the cluster assignment of SNV n to cluster k.
- We don't know the number of clusters in advance.

Problem statement

SNV clustering problem

Suppose that for SNVs $n \in \{1, ..., N\}$ in samples $m \in \{1, \ldots M\}$. Further suppose that there exists clones (clusters) $k \in \{1, \ldots, K\}$ and a true clustering $\mathbf{z}$. Given total reads $\mathbf{d}_1, \ldots, \mathbf{d}_n$ and variant reads $\mathbf{v}_1, \ldots, \mathbf{v}_n$, we seek to infer $\mathbf{z}$.

Observation model

Now suppose that the variant reads are binomially distributed.

Observation model

Now suppose that the variant reads are binomially distributed.

- Each cluster emits variant reads with *cluster frequency* ϕ_{mk}.

Observation model

Now suppose that the variant reads are binomially distributed.

- Each cluster emits variant reads with *cluster frequency* ϕ_{mk}.
- Then

$$\mathbf{v_n} \sim \begin{bmatrix} \mathrm{Bin}(v_{1n}; d_{1n}, \phi_{1k}) \\ \mathrm{Bin}(v_{2n}; d_{2n}, \phi_{2k}) \\ \vdots \\ \mathrm{Bin}(v_{Mn}; d_{Mn}, \phi_{Mk}) \end{bmatrix} \triangleq \mathbf{Bin}(\boldsymbol{\phi_k})$$

Dirichlet Process

Definition (Dirichlet Process)

For any measurable finite partition $\{B_i\}_{i=1}^n$ of a measurable set S ,

$$\text{if } X \sim \mathrm{DP}\left(H, \alpha\right)$$
$$\text{then } \left(X\left(B_1\right), \ldots, X\left(B_n\right)\right) \sim \mathrm{Dir}\left(\alpha H\left(B_1\right), \ldots, \alpha H\left(B_n\right)\right)$$

where Dir denotes the Dirichlet distribution.

- A non-parametric prior on the number of clusters.

Dirichlet Process

It is convenient to represent the Dirichlet Process in terms of its stick-breaking representation:

$$\pi_i(\mathbf{v}) = v_i \prod_{j=1}^{i-1} (1 - v_j)$$

$$DP = \sum_{i=1}^{\infty} \pi_i(\mathbf{v}) \delta_{\phi_i}$$

where ϕ_i are the parameters for the realized distribution, and v_i are iid $\mathrm{Beta}(1, \alpha)$.

We will use the stick breaking representation here.

Binomial Mixture Model with DP prior

Likelihood of data given its cluster membership:

$$\Pr(\mathbf{x_n}|\phi_k) = \prod_{m=1}^{M} \mathrm{Bin}(v_{mn}; d_{mn}, \phi_k)$$

Binomial Mixture Model with DP prior

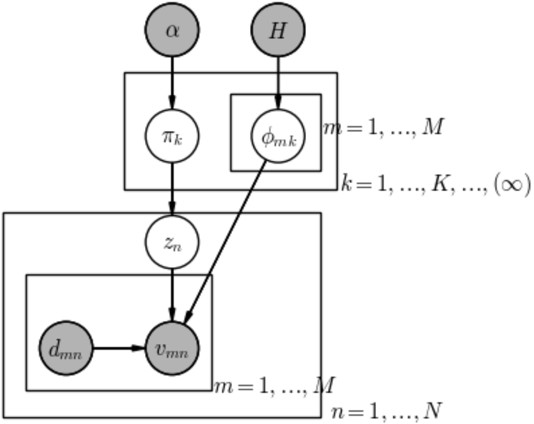
Likelihood of data given its cluster membership:

$$\Pr(\mathbf{x_n}|\phi_k) = \prod_{m=1}^{M} \mathrm{Bin}(v_{mn}; d_{mn}, \phi_k)$$

Joint likelihood of observed data and cluster memberships:

$$\Pr(\mathbf{x_n}, \mathbf{z}|\boldsymbol{\pi}, \phi) = \prod_{k=1}^{K} \prod_{m=1}^{M} (\pi_k \mathrm{Bin}(v_{mn}; d_{mn}, \phi_{mk}))^{\mathbf{z}_{nk}}$$

Binomial Mixture Model with DP prior



Binomial Mixture Model with DP prior

The full cluster assignment posterior

$$p(\mathbf{z}|\mathbf{x}, \alpha, H) = \int p(\mathbf{x}|\phi)p(\phi|\mathbf{x}, \alpha, H)\, d\phi$$

involves a Dirichlet Process and is thus analytically intractable. We must use some sort of computational technique, such as variational inference, to perform inference on this posterior.

Variational Inference

Variational Inference (VI) is an alternative to MCMC sampling-based inference methods.

Variational Inference

Variational Inference (VI) is an alternative to MCMC sampling-based inference methods.

- Generalized version of EM; deterministic given an initialization.

Variational Inference

Variational Inference (VI) is an alternative to MCMC sampling-based inference methods.

- Generalized version of EM; deterministic given an initialization.
- Faster than MCMC.

Variational Inference

Variational Inference (VI) is an alternative to MCMC sampling-based inference methods.

- Generalized version of EM; deterministic given an initialization.
- Faster than MCMC.
- Scales well on large datasets.

Overview of Variational Inference: I

Let $\mathbf{z}$ denote the latent variables, and $\mathbf{x}$ denote the data. We seek to approximate the posterior $p(\mathbf{z}|\mathbf{x})$ from a family of distributions $\mathcal{D}$ by solving the following optimization problem:

$$q^*(z) = \arg\min_{q(\mathbf{z})\in\mathcal{D}} \mathrm{KL}\left((q(\mathbf{z})||p(\mathbf{z}|\mathbf{x}))\right).$$

where KL is the KL-divergence, which measures the "distance" between two distributions.

[1] Blei 2006

Overview of Variational Inference: II

However, the KL-divergence requires us to compute the log evidence (which is intractable over the space of all $\mathbf{z}$), since

$$\mathrm{KL}\left(q(\mathbf{z}) \| p(\mathbf{z} \mid \mathbf{x})\right) = \mathrm{E}\left[\log q(\mathbf{z})\right] - \mathrm{E}\left[\log p(\mathbf{z}, \mathbf{x})\right] + \log p(\mathbf{x}).$$

Overview of Variational Inference: II

However, the KL-divergence requires us to compute the log evidence (which is intractable over the space of all $\mathbf{z}$), since

$$\mathrm{KL}\left(q(\mathbf{z}) \| p(\mathbf{z} \mid \mathbf{x})\right) = \mathrm{E}\left[\log q(\mathbf{z})\right] - \mathrm{E}\left[\log p(\mathbf{z}, \mathbf{x})\right] + \log p(\mathbf{x}).$$

Instead, we optimize the an objective function which is not dependent on $\log p(\mathbf{x})$, called the evidence lower bound (ELBO):

$$\mathrm{ELBO}(q) = \mathrm{E}[\log p(\mathbf{z}, \mathbf{x})] - \mathrm{E}[\log q(\mathbf{z})].$$

The ELBO is a lower bound for the log evidence.

Overview of Variational Inference: III

The standard technique is to select a simple family of distributions for $\mathcal{D}$, the mean-field variational family. In this family, the latent variables $\mathbf{z}$ are mutually independent so that the joint distribution factorizes:

$$q(\mathbf{z}) = \prod_{j=1}^{m} q_j(z_j).$$

where q_j is a bounded variation dependent only on z_j. The structure of the model will dictate the optimal form of q_j.

Overview of Variational Inference: IV

The mean-field assumption gave us independence between variables. This suggests a coordinate ascent algorithm.

[1] Bishop 2006

Overview of Variational Inference: IV

The mean-field assumption gave us independence between variables. This suggests a coordinate ascent algorithm.

Let $\mathbf{z}_{-j}$ denote the set of latent variables $\mathbf{z}_l$ such that $l \neq j$. Then we can show that

$$\begin{aligned} \mathrm{ELBO}(q) &= \int \prod q_i(\mathbf{z}_i) \left(\log p(\mathbf{z}, \mathbf{x}) - \sum_i \log q_i(\mathbf{z}_i) \right) \, d\mathbf{z} \\ &\propto \int q_j(z_j) \mathrm{E}_{-j} \left[\log p(\mathbf{x}, \mathbf{z}) \right] \, d\mathbf{z}_j - \int q_j(z_j) \log q_j(\mathbf{z}_j) \, d\mathbf{z}_j \end{aligned}$$

[1] Bishop 2006

Overview of Variational Inference: IV

The mean-field assumption gave us independence between variables. This suggests a coordinate ascent algorithm.

Let $\mathbf{z}_{-j}$ denote the set of latent variables $\mathbf{z}_l$ such that $l \neq j$. Then we can show that

$$\begin{aligned}\mathrm{ELBO}(q) &= \int \prod q_i(\mathbf{z}_i) \left(\log p(\mathbf{z}, \mathbf{x}) - \sum_i \log q_i(\mathbf{z}_i) \right) \, d\mathbf{z} \\ &\propto \int q_j(z_j) \mathrm{E}_{-j} \left[\log p(\mathbf{x}, \mathbf{z}) \right] \, d\mathbf{z}_j - \int q_j(z_j) \log q_j(\mathbf{z}_j) \, d\mathbf{z}_j\end{aligned}$$

Now suppose that we fix z_{-j} and maximize the ELBO. It can be shown that

$$q_j^*(\mathbf{z}_j) \propto \exp\left(\mathrm{E}_{-j}\left[\log p(\mathbf{x}, \mathbf{z})\right]\right)$$

[1] Bishop 2006

Overview of Variational Inference: IV

The mean-field assumption gave us independence between variables. This suggests a coordinate ascent algorithm.

Let $\mathbf{z}_{-j}$ denote the set of latent variables $\mathbf{z}_l$ such that $l \neq j$. Then we can show that

$$\begin{aligned} \mathrm{ELBO}(q) &= \int \prod q_i(\mathbf{z}_i) \left(\log p(\mathbf{z}, \mathbf{x}) - \sum_i \log q_i(\mathbf{z}_i) \right) \, d\mathbf{z} \\ &\propto \int q_j(z_j) \mathrm{E}_{-j} \left[\log p(\mathbf{x}, \mathbf{z}) \right] \, d\mathbf{z}_j - \int q_j(z_j) \log q_j(\mathbf{z}_j) \, d\mathbf{z}_j \end{aligned}$$

Now suppose that we fix z_{-j} and maximize the ELBO. It can be shown that

$$q_j^*(\mathbf{z}_j) \propto \exp \left(\mathrm{E}_{-j} \left[\log p(\mathbf{x}, \mathbf{z}) \right] \right)$$

By iterating through these coordinate updates, we reach a local optimum of the ELBO.

[1] Bishop 2006

Overview of Variational Inference: V

If the posterior is in the exponential family, then the computation of coordinate ascent and ELBO can be generalized.

[1] Hughes 2015

Overview of Variational Inference: V

If the posterior is in the exponential family, then the computation of coordinate ascent and ELBO can be generalized.

Recall that a distribution is in exponential form if it can parameterized by

$$f_X(x \mid \theta) = h(x) \exp\left(\theta^T \cdot T(x) - A(\theta)\right)$$

where $T(x)$ are the sufficient statistics, θ are the natural parameters,$A(\theta)$ is the cumulant, and $h(x)$ is the base measure.

[1] Hughes 2015

Overview of Variational Inference: V

If the posterior is in the exponential family, then the computation of coordinate ascent and ELBO can be generalized.

Recall that a distribution is in exponential form if it can parameterized by

$$f_X(x \mid \theta) = h(x) \exp\left(\theta^T \cdot T(x) - A(\theta)\right)$$

where $T(x)$ are the sufficient statistics, θ are the natural parameters,$A(\theta)$ is the cumulant, and $h(x)$ is the base measure.

The intuition is that because $q^* \propto \exp(\mathrm{E}[\log(.)])$, then writing the distribution in exponential form reveals dependencies that hold for all exponential family members.

[1] Hughes 2015

Overview of Variational Inference: VI

So why not always use variational inference?

Overview of Variational Inference: VI

So why not always use variational inference?

- Accuracy depends on quality of approximation.

Overview of Variational Inference: VI

So why not always use variational inference?

- Accuracy depends on quality of approximation.
  - q^* might be far off from the true posterior.

Overview of Variational Inference: VI

So why not always use variational inference?

- Accuracy depends on quality of approximation.
  - q^* might be far off from the true posterior.
  - MCMC is potentially more accurate...but it might get stuck in local optima or take forever to converge.

Overview of Variational Inference: VI

So why not always use variational inference?

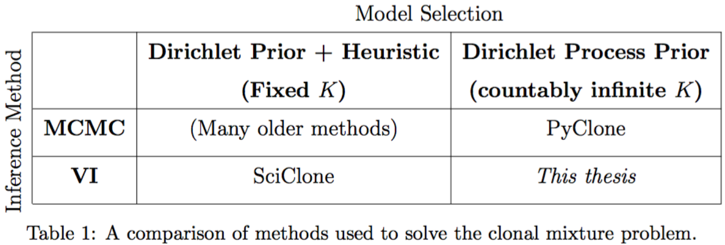
- Accuracy depends on quality of approximation.
    - q^* might be far off from the true posterior.
    - MCMC is potentially more accurate...but it might get stuck in local optima or take forever to converge.
- MCMC, like Gibbs sampling, is typically easier to implement.

Overview of Variational Inference: VI

So why not always use variational inference?

- Accuracy depends on quality of approximation.
    - q^* might be far off from the true posterior.
    - MCMC is potentially more accurate...but it might get stuck in local optima or take forever to converge.
- MCMC, like Gibbs sampling, is typically easier to implement.
    - No need for painful derivations.

Existing methods



Model Selection

Inference Method		**Dirichlet Prior + Heuristic (Fixed K)**	**Dirichlet Process Prior (countably infinite K)**
	MCMC	(Many older methods)	PyClone
	VI	SciClone	*This thesis*

Table 1: A comparison of methods used to solve the clonal mixture problem.

VI for the DP/Binomial mixture model

By beta-binomial conjugacy,

$$\begin{aligned} q(\phi_k) &= \prod_{m=1}^{M} q(\phi_{mk}) \\ &= \prod_{m=1}^{M} \mathrm{Beta}(\phi_k | \alpha_{mk}, \beta_{mk}) \end{aligned}$$

VI for the DP/Binomial mixture model

By beta-binomial conjugacy,

$$\begin{aligned} q(\phi_k) &= \prod_{m=1}^{M} q(\phi_{mk}) \\ &= \prod_{m=1}^{M} \mathrm{Beta}(\phi_k | \alpha_{mk}, \beta_{mk}) \end{aligned}$$

Thus, combined with the DP variational parameters,

$$q(\mathbf{z}, \mathbf{v}, \phi) = \underbrace{\prod_{k=1}^{K} q(\phi_k)}_{\substack{\text{Observation: likelihoods} \\ \text{Product of betas} \\ 2MK \text{ variational parameters} \\ \{\alpha_{mk},\beta_{mk}\}_{m=1,k=1}^{M,K}}} \times \underbrace{\prod_{k=1}^{K} q(\mathbf{v}_k)}_{\substack{\text{Allocation: cluster proportions} \\ \text{Product of betas} \\ 2K \text{ variational parameters} \\ \{\eta_{k0},\eta_{k1}\}_{k=1}^{K}}} \times \underbrace{\prod_{n=1}^{N} q(z_n)}_{\substack{\text{Allocation: cluster responsibilities} \\ \text{Product of categoricals} \\ 2NK \text{ variational parameters} \\ \{\hat{r}_{nk}\}_{n=1,k=1}^{N,K}}}$$

Most of the details are in the thesis appendices.

MAP estimates

For each cluster we pool reads by cluster membership:

$$v_{mk}^{\text{pooled}} = \sum_n (v_{mn})^{\mathbf{z}_n}$$

$$d_{mk}^{\text{pooled}} = \sum_n (d_{mn})^{\mathbf{z}_n}$$

MAP estimates

For each cluster we pool reads by cluster membership:

$$v_{mk}^{\text{pooled}} = \sum_n (v_{mn})^{\mathbf{z}_n}$$

$$d_{mk}^{\text{pooled}} = \sum_n (d_{mn})^{\mathbf{z}_n}$$

and we can make MAP estimates by converting from the variational parameters back to the original parameters of the posterior:

$$\mathbf{z}_n^{\text{MAP}} = \arg\max_k \hat{r}_{nk}$$

$$\phi_{mk}^{\text{MAP}} = \frac{v_{mk}^{\text{pooled}} + \alpha_{mk} - 1}{d_{mk}^{\text{pooled}} + \alpha_{mk} + \beta_{mk} - 2}.$$

Implementation

Implementation

- Responsibilities initialized with k-means++ on N clusters.

Implementation

- Responsibilities initialized with k-means++ on N clusters.
- Uniform priors for the betas.

Implementation

- Responsibilities initialized with k-means++ on N clusters.
- Uniform priors for the betas.
- DP hyperparameters chosen empirically.

Implementation

- Responsibilities initialized with k-means++ on N clusters.
- Uniform priors for the betas.
- DP hyperparameters chosen empirically.

We declared convergence when the difference in ELBO between two iterations was less than 0.01.

Implementation

- Responsibilities initialized with k-means++ on N clusters.
- Uniform priors for the betas.
- DP hyperparameters chosen empirically.

We declared convergence when the difference in ELBO between two iterations was less than 0.01.

Everything was implemented in Python.

Coordinate ascent algorithm

```
Algorithm 1: CAVI FOR THE DP BINOMIAL MIXTURE MODEL
Input: Data $\mathbf{x}_n$, where each $x_i$ is an integer vector with $M$ entries.
       $\gamma_0, \gamma_1, \alpha_0, \beta_0$, hyperparameters
Output: Converged variational parameters
       $\{\alpha_{mk}, \beta_{mk}\}_{m=1,k=1}^{M,K}, \{\eta_{k0}, \eta_{k1}\}_{k=1}^{K}, \{\hat{r}_{nk}\}_{n=1,k=1}^{N,K}$
Initialize: $\alpha_0 = \beta_0 = \alpha_{mk} = \beta_{mk} = 1, \forall m, k$
       $\gamma_1 = \eta_1 = 1.0, \gamma_0 = \eta_0 = 1.5$
       $\hat{r}_{nk} \leftarrow \texttt{kmeans++}(\mathbf{x})$
while the ELBO has not converged do
    ▷ Compute data-specific (local) parameters
        $\mathrm{E}_q[\log p(x_n|\alpha_{mk}, \beta_{mk})] \leftarrow \mathrm{E}_q[\log \left( \binom{d_{mn}+v_{mn}}{v_{mn}} (\phi_k)^{v_{mn}} (1-\phi_k)^{d_{mn}} \right)]$
        $\hat{r}_{nk} \leftarrow \exp(S_k)$
    ▷ Compute sufficient statistics
        $S_k = \sum_{n=1}^{N} \hat{r}_{nk} s(x_n) = \sum_{n=1}^{N} \hat{r}_{nk} \left[ \begin{bmatrix} v_{1n} & d_{1n} \end{bmatrix} \quad \cdots \quad \begin{bmatrix} v_{Mn} & d_{Mn} \end{bmatrix} \right]$
        $N_k = \sum_{n=1}^{N} \hat{r}_{nk}$
        $N_k^{>} = \sum_{k+1}^{K} N_k$
    ▷ Compute cluster-specific (global) parameters
        $\eta_{k1} \leftarrow 1 + \sum_n \hat{r}_{nk} = 1 + N_k$
        $\eta_{k0} \leftarrow \gamma + \sum_n \sum_{j=k+1}^{K} \hat{r}_{nj} = N_k^{>}$
        $\alpha_{mk} \leftarrow (\alpha_0 - 1) + S_{km}$
        $\beta_{mk} \leftarrow (\beta_0 - 1) + S_{km}$
    Compute $\mathrm{ELBO}(q) = \mathbb{E}[\log p(\mathbf{z}, \mathbf{x})] - \mathbb{E}[\log q(\mathbf{z})]$
end
return Converged variational parameters
```

Simulated Data

Number of clusters (K)	10
Number of SNVs (N)	100
Number of samples (M)	4, 5, 6
Coverage	50, 100, 1000

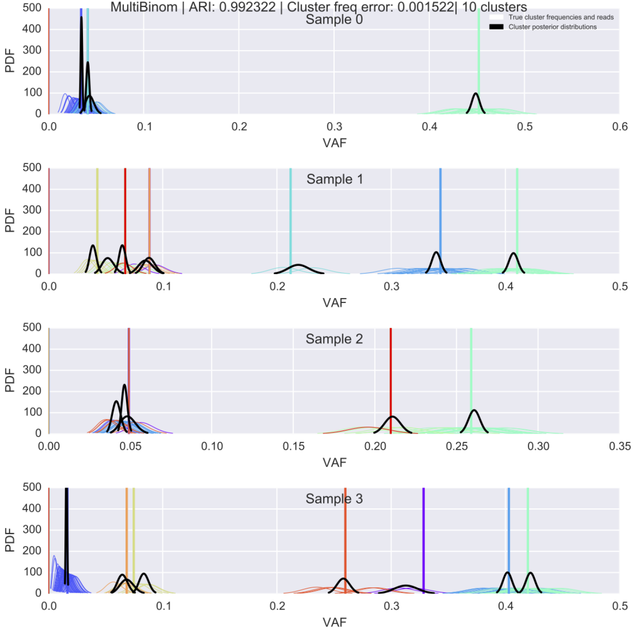
Parameters for the simulated datasets.

Evaluating results

The DP/VI coordinate ascent algorithm was benchmarked against SciClone (VI) and PyClone (DP/MCMC).

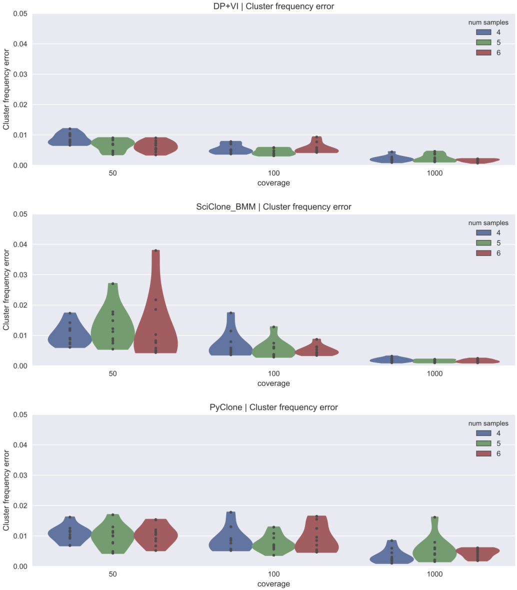
- Evaluate clustering: Adjusted Rand Index (ARI)
- Evaluate parameters: Cluster frequency error (CFE)

$$\mathrm{CFE}(\phi^{\mathrm{MAP}}) \triangleq \frac{1}{C} \sum_{c=1}^{C} \min_{k \in \{1,\dots,K\}} \left\| \phi_c^{\mathrm{MAP}} - \phi_k \right\|$$

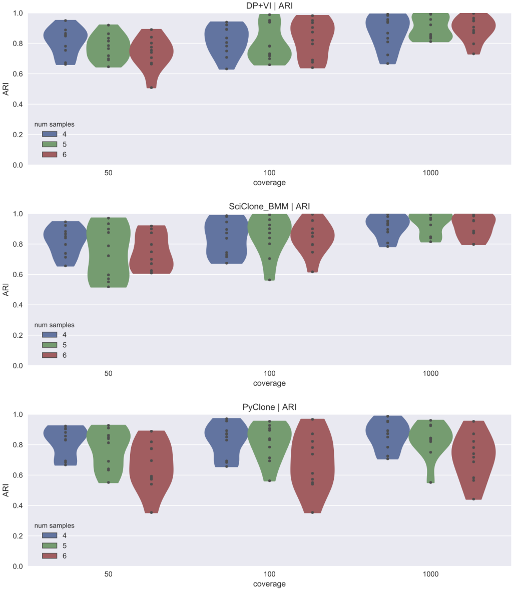


Cluster posteriors (black) and true clusters (colored, vertical lines) with $\mathbf{x}_n$ *(colored beta curves).*

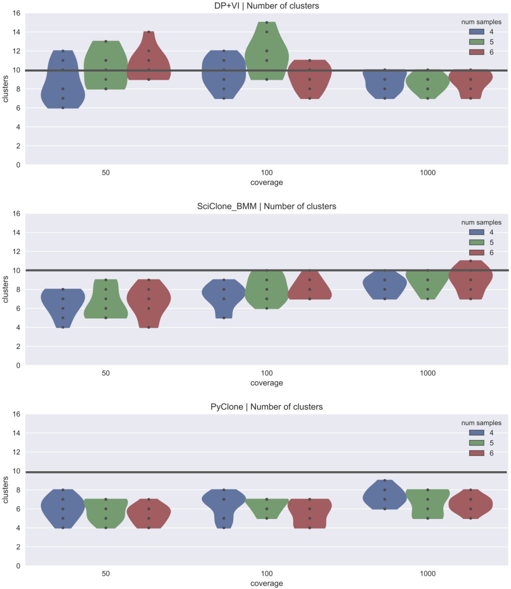
Cluster Frequency Error



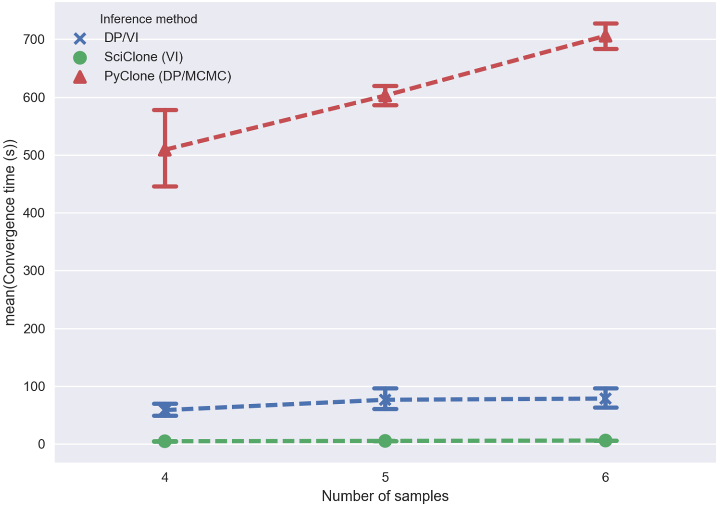
Adjusted Rand Index



Number of clusters



Runtime


Inference method
DP/VI
SciClone (VI)
PyClone (DP/MCMC)
mean(Convergence time (s))
0
100
200
300
400
500
600
700
4
5
6
Number of samples

Discussion

The DP/VI method is:

Discussion

The DP/VI method is:

- Faster and more accurate than PyClone (MCMC method)

Discussion

The DP/VI method is:

- Faster and more accurate than PyClone (MCMC method)
- Comparable to SciClone (other variational method)

Discussion

The DP/VI method is:

- Faster and more accurate than PyClone (MCMC method)
- Comparable to SciClone (other variational method)
    - Better at lower coverages.

Discussion

The DP/VI method is:

- Faster and more accurate than PyClone (MCMC method)
- Comparable to SciClone (other variational method)
    - Better at lower coverages.
    - Added benefit: nonparametric prior.

Future Work

Future Work

- Other models, such as negative binomial

Future Work

- Other models, such as negative binomial
- More advanced variational inference techniques

Future Work

- Other models, such as negative binomial
- More advanced variational inference techniques
- Optimize code, write in a faster language

Future Work

- Other models, such as negative binomial
- More advanced variational inference techniques
- Optimize code, write in a faster language
- Try as part of a phylogeny inference pipeline

Future Work

- Other models, such as negative binomial
- More advanced variational inference techniques
- Optimize code, write in a faster language
- Try as part of a phylogeny inference pipeline
- Try on real data

Acknowledgments

Thanks to:

- Prof Ben Raphael
- Mohammed El-Kebir, Gryte Satas, and the rest of the Raphael Lab
- Prof Erik Sudderth and Mike Hughes
- My friends and family
  - Especially the support of the others doing theses:
    Uthsav Chitra, Vaki Nikitopoulos, Matt Chin, and Keshav Vemuri