Japanese Joint Statistical Meeting 2018

2018/09/09 (Sun.) – 2018/09/13 (Thu.)

Stochastic Variational Inference of Mixture Models in Phylogenetics

Tung Dang and Hirohisa Kishino

Laboratory of Biometrics and Bioinformatics,
University of Tokyo

Corresponding author:
dangthanhtung91@vn-bml.com

2018年度

統計関連学会連合大会

Japanese Joint Statistical Meeting　2018年9月9日(日)～13日(木)

■E会場（5333教室）10:00 - 12:00

【一般講演】English Session (4): Medical Statistics and Biostatistics

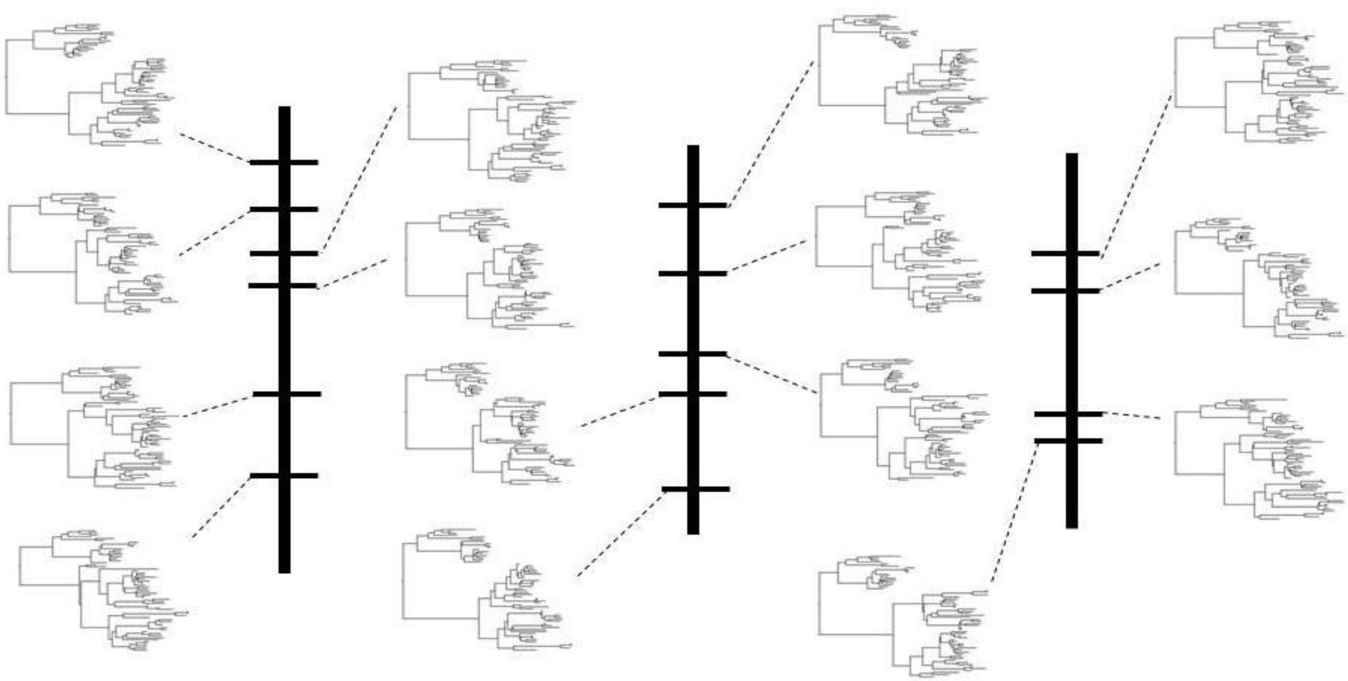
【Chair】Ryuji Uozumi (Kyoto University)

	講演タイトル	発表者（所属）	共著者（所属）	報告集
1	Use of external information for assessing efficacy equivalence in biosimilar clinical trials	Ryuji Uozumi (Kyoto University)	Shinjo Yada (A2 Healthcare)	報告集
2	Bartlett-type corrections for the confidence interval of a treatment effect of the multivariate random effects meta-analysis model via analytical approach	Masahiro Kojima (Kyowa Hakko Kirin)		報告集
3	The impact of unobserved heterogeneity and competing risks with shared frailty in radiation risk assessment	Kyoji Furukawa (Kurume University)		報告集
4	Modelling life history under varying temperature conditions	Hideyasu Shimadzu (Loughborough University)		
5	Sufficient dimension reduction via random-partitions for the large-p-small-n problem	Hung Hung (National Taiwan University)		
6	Stochastic Variational Inference of Mixture Models in Phylogenetics	Tung Dang (University of Tokyo)	Hirohisa Kishino (University of Tokyo)	

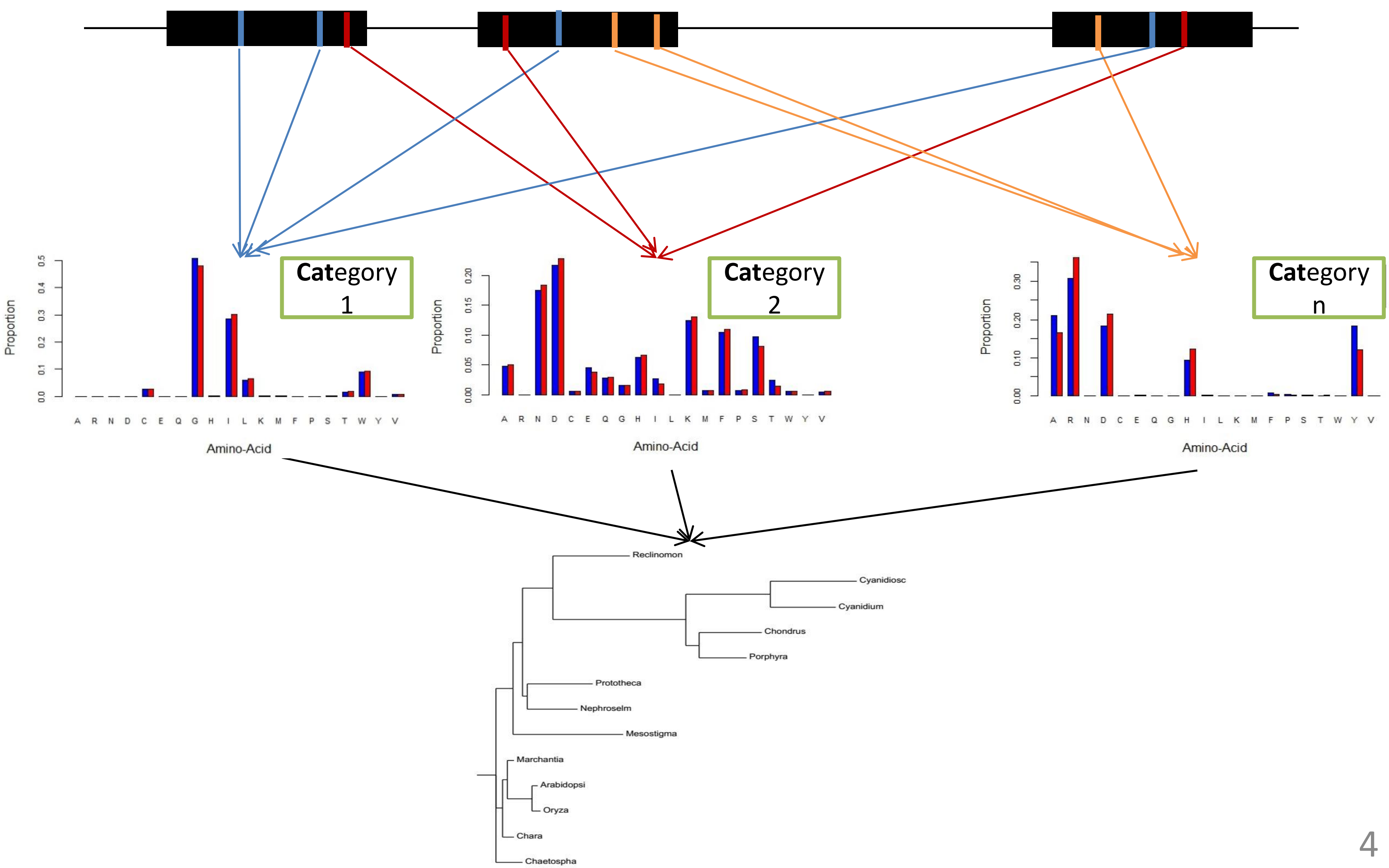
←タイムテーブルに戻る | ↑ページのTOPへ

Problem of Phylogenomics

- The pattern of molecular evolution varies among gene sites and genes in a genome.
- By taking into account the complex heterogeneity of evolutionary processes among sites in a genome, Bayesian infinite mixture models of genomic evolution enable robust phylogenetic inference.



CAT model (Lartillot et al 2004)



The likelihood of CAT model

$$p(\Xi_{ij} \mid \pi_{ka}, r_i, l_j) = \prod_{k=1}^{\infty}\prod_{a=1}^{20}\left(\pi_{ka}\, p(\mathrm{n}_{ij} \mid r_i, l_j)\right)^{I[\mathrm{Z}_i=\mathrm{k}]}$$

$$p(\Xi_{ij} \mid \pi_{ka}, r_i, l_j) = \left(\prod_{k=1}^{\infty}\prod_{a=1}^{20}\pi_{ka}{}^{w_{ka}}\right)^{\mathrm{I}[\mathrm{Z}_i=\mathrm{k}]}\left(\prod_i r_i{}^{\sum_j n_{ij}}\right)\left(\prod_j l_j{}^{\sum_i n_{ij}}\right)\left(\prod_{ij}\frac{e^{-r_i l_j}}{n_{ij}\,!}\right)$$

l_j : branch lengths

r_i : rate of substitution

π_{ka} : equilibrium frequency profile

A stick-breaking construction

V_k : unit length of the stick

$\mathrm{I}[\mathrm{Z}_i=\mathrm{k}]$: allocation variable

Ξ_{ij} : substitution mapping data

n_{ij} : number of substitutions

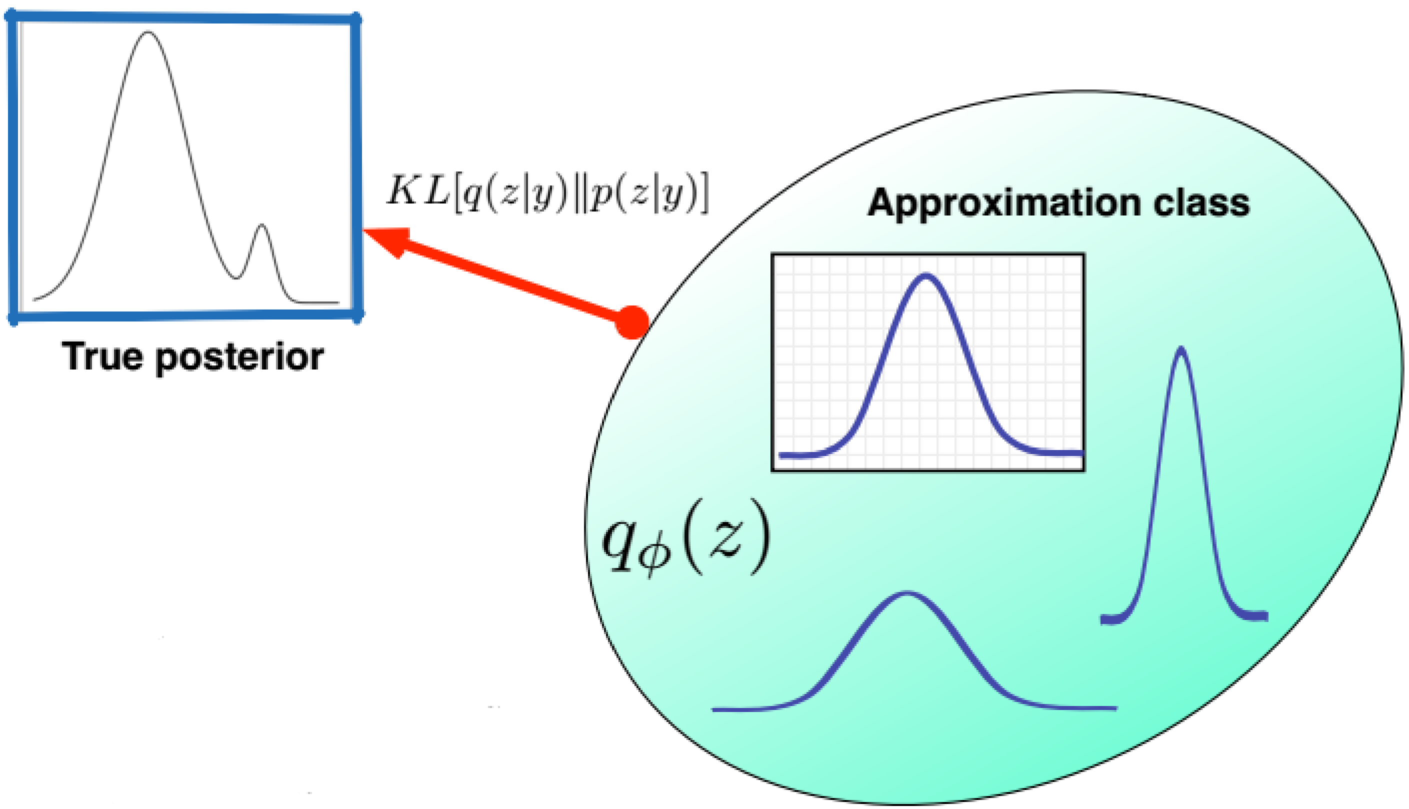
w_{ka} : number of type of substitutions

MCMC becomes infeasible for large datasets

- The computational burden of MCMC is prohibitive for large data sets.
- Even well-designed sampling schemes needs a large sample to achieve convergence.
- Diagnosis of convergence is difficult for high dimensional parameter space.

Variational Method

approximate complicated densities by a simpler class of densities, called **variational distribution**.



Evidence Lower Bound (ELBO)

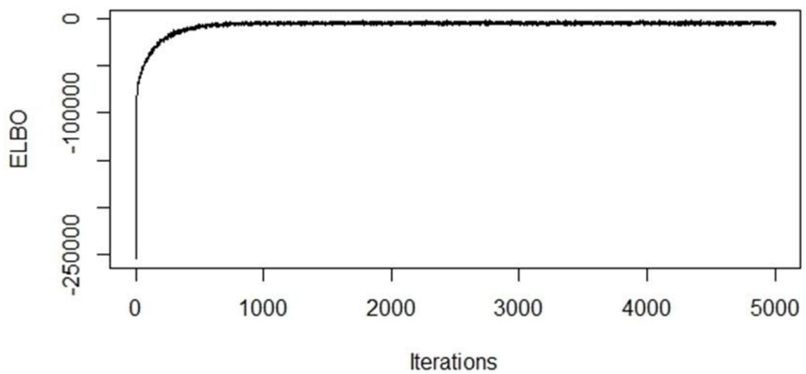
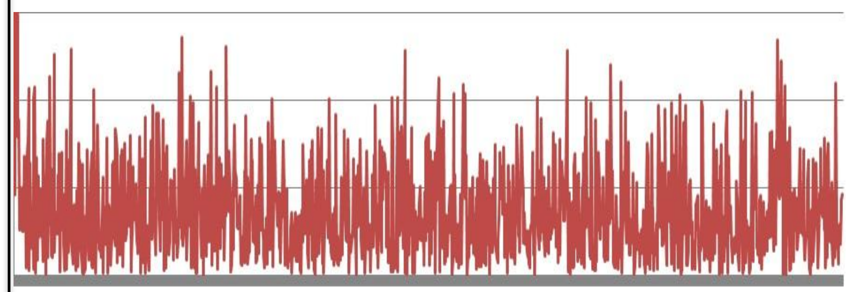
$$\mathrm{KL}[q(\boldsymbol{\theta};\boldsymbol{\Theta}) \parallel p(\boldsymbol{\theta}|\mathbf{X})]$$

$$= \mathrm{E}_q[\log q(\boldsymbol{\theta};\boldsymbol{\Theta})] - \mathrm{E}_q[\log p(\mathbf{X},\boldsymbol{\theta})] + \log p(\mathbf{X})$$

-ELBO: computable

Computational headache
But constant !!

Variational Inference and MCMC

VI	MCMC
	

Maximizing ELBO

Monte Carlo simulation with proposal and acceptance steps

Example: *Variational Inference of Bayesian Ridge Regression*

- The likelihood is

$$p(t|x,w,\beta)=\prod_{n=1}^{N}\mathcal{N}\left(t_n|w^T x_n,\beta^{-1}\right)$$

- The regression coefficient $p(w|\alpha)=\mathcal{N}\left(w|0,\alpha^{-1}I\right)$

- The extent of shrinkage $p(\alpha)=Gam\left(\alpha|a_0,b_0\right)$

- The variational mean-field representation

$$q(w,\alpha)=q(w)\,q(\alpha)$$

- The practical variational distributions

$$\begin{aligned} q(w) &= \mathcal{N}\left(w|m_N,S_N\right)\\ q(\alpha) &= Gam\left(\alpha|a_N,b_N\right) \end{aligned}$$

Example: *Variational Inference of Bayesian Ridge Regression*

- The ELBO for Bayesian ridge regression model.

$$\mathcal{L}\left[q(w,\alpha|m_N,S_N,a_N,b_N)\right] = \mathrm{E}_q\left[\log p(t|w)\right] + \mathrm{E}_q\left[\log p(w|\alpha)\right] + \mathrm{E}_q\left[\log p(\alpha)\right]$$
$$-\mathrm{E}_q\left[\log q(w|m_N,S_N)\right] - \mathrm{E}_q\left[\log q(\alpha|a_N,b_N)\right]$$

- Update m_N, S_N : Given $a_N = a_N^0, b_N = b_N^0, m_N = m_N^0, S_N = S_N^0$

$$m_N = \beta S_N^0 X^T t$$
$$S_N = \left(\mathrm{E}_q[\alpha] + \beta X^T X\right)^{-1} I = \left(\frac{a_N^0}{b_N^0} + \beta X^T X\right)^{-1} I$$

- Update a_N, b_N : Given $a_N = a_N^0, b_N = b_N^0, m_N = m_N^0, S_N = S_N^0$

$$a_N = a_0 + \frac{M}{2}$$
$$b_N = b_0 + \frac{1}{2}\mathrm{E}_q\left[w^T w\right] = b_0 + \frac{1}{2}\left({m_N}^{0^T} {m_N}^0 + {S_N}^0\right)$$

Example: *Variational Inference of Dirichlet mixture model*

- The PDF of a DMM can be represented

$$f(\mathbf{X};\mathbf{\Pi},\mathbf{U}) = \prod_{n=1}^{N}\sum_{i=1}^{I}\pi_i\,\mathrm{Dir}(\mathbf{x}_n;\mathbf{u}_i), \quad \pi_i > 0, \quad \sum_{i=1}^{I}\pi_i = 1$$

- The mixture weights $\mathbf{\Pi} = [\pi_1, \ldots, \pi_I]^{\mathrm{T}}$
- The parameter matrix $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_I]$
- The joint distribution of the observation X and all the latent variables

$$\begin{aligned}
f(\mathbf{X},\mathcal{Z}) &= f(\mathbf{X},\mathbf{U},\mathbf{\Pi},\mathbf{Z}) \\
&= f(\mathbf{X}|\mathbf{Z},\mathbf{U})f(\mathbf{Z}|\mathbf{\Pi})f(\mathbf{\Pi})f(\mathbf{U}) \\
&= \prod_{n=1}^{N}\prod_{i=1}^{I}\left[\pi_i\frac{\Gamma(\sum_{k=1}^{K+1}u_{ki})}{\prod_{k=1}^{K+1}\Gamma(u_{ki})}\prod_{k=1}^{K+1}x_{kn}^{u_{ki}-1}\right]^{z_{ni}} \times \frac{\Gamma(\sum_{i=1}^{I}c_{i_0})}{\prod_{i=1}^{I}\Gamma(c_{i_0})}\prod_{i=1}^{I}\pi_i^{c_{i_0}-1} \\
&\quad \times \prod_{i=1}^{I}\prod_{k=1}^{K+1}\frac{\alpha_{ki_0}^{\mu_{ki_0}}}{\Gamma(\mu_{ki_0})}u_{ki}^{\mu_{ki_0}-1}e^{-\alpha_{ki_0}u_{ki}}.
\end{aligned}$$

- The ELBO for Dirchlet Mixture model

$$\mathcal{L} = \mathbb{E}_{\mathcal{Z}}[\ln f(\mathbf{X},\mathcal{Z})] - \mathbb{E}_{\mathcal{Z}}[\ln f(\mathcal{Z})]$$

Example: *Variational Inference of Dirichlet mixture model*

- The optimal solution to the posterior distribution of u_{ki} is

$$\begin{aligned}\ln f^*(u_{ki};\mu_{ki},\alpha_{ki}) \\ &= \sum_{n=1}^{N} \mathbb{E}[z_{ni}] \times \boxed{\mathbb{E}_{\mathcal{Z}\setminus u_{ki}}\left[\ln \frac{\Gamma(\sum_{k=1}^{K+1} u_{ki})}{\prod_{k=1}^{K+1}\Gamma(u_{ki})}\right]} \\ &\quad + u_{ki}\sum_{n=1}^{N}\mathbb{E}[z_{ni}]\ln x_{kn} + (\mu_{ki_0}-1)\ln u_{ki} - \alpha_{ki_0}u_{ki} + \text{const.}\end{aligned}$$

- The optimal solution to the posterior distribution

$$\ln f^*(\pi_i;c_i) = \mathbb{E}_{\mathcal{Z}\setminus\pi_i}[\ln f(\mathbf{X},\mathcal{Z})] = \ln \pi_i \times \sum_{n=1}^{N}\mathbb{E}[z_{ni}] + \ln\pi_i(c_{i_0}-1) + \text{const.}$$

- For the variable z_{ni}, the optimal approximation to the posterior distribution is

$$\begin{aligned}\ln f^*(z_{ni}) &= \mathbb{E}_{\mathcal{Z}\setminus z_{ni}}[\ln f(\mathbf{X},\mathcal{Z})] \\ &= z_{ni} \times \left[\mathbb{E}[\ln \pi_i] + \sum_{k=1}^{K+1}(u_{ki}-1)\ln x_{kn}\right] \\ &\quad + z_{ni} \times \boxed{\mathbb{E}\left[\ln\frac{\Gamma(\sum_{k=1}^{K+1}u_{ki})}{\prod_{k=1}^{K+1}\Gamma(u_{ki})}\right]} + \text{const.}\end{aligned}$$

Example: *Variational Inference of Dirichlet mixture model*

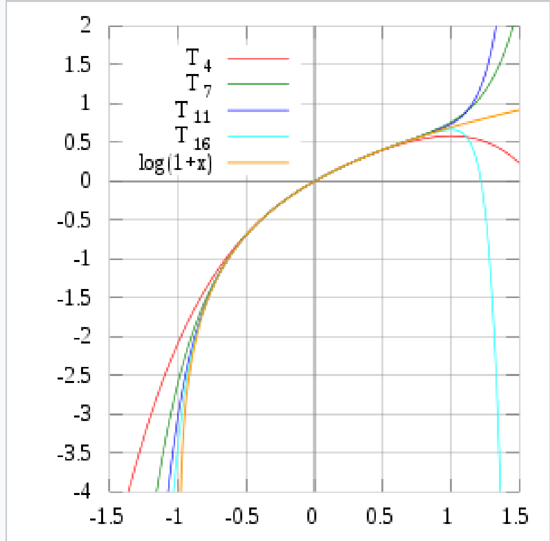
Definition [edit]

The Taylor series of a real or complex-valued function $f(x)$ that is infinitely differentiable at a real or complex number a is the power series

$$f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots,$$

which can be written in the more compact sigma notation as

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n,$$



The Taylor polynomials for $\log(1+x)$ only provide accurate approximations in the range $-1 < x \leq 1$. For $x > 1$, Taylor polynomials of higher degree provide worse approximations.

- A lower-bound approximation can be obtained as

$$\ln \frac{\Gamma(\sum_{k=1}^{K+1} u_{ki})}{\prod_{k=1}^{K+1} \Gamma(u_{ki})} \geq \ln \frac{\Gamma(\sum_{k=1}^{K+1} \overline{u}_{ki})}{\prod_{k=1}^{K+1} \Gamma(\overline{u}_{ki})} + \sum_{k=1}^{K+1} \left[\psi\left(\sum_{m=1}^{k} \overline{u}_{mi} + \sum_{l=k+1}^{K+1} u_{li} \right) - \psi(\overline{u}_{ki}) \right] \overline{u}_{ki}(\ln u_{ki} - \ln \overline{u}_{ki}).$$

$\psi(\cdot)$ is the digamma function defined as $\psi(x) = \partial \ln \Gamma(x)/\partial x$.

Example: *Variational Inference of Dirichlet mixture model*

- The optimal solution to the posterior distribution of uki is

$$
\begin{aligned}
\ln f^*(u_{ki};\mu_{ki},\alpha_{ki}) &= \sum_{n=1}^{N} \mathbb{E}[z_{ni}] \times \boxed{\mathbb{E}_{\mathcal{Z}\setminus u_{ki}}\left[\ln \frac{\Gamma(\sum_{k=1}^{K+1} u_{ki})}{\prod_{k=1}^{K+1}\Gamma(u_{ki})}\right]} \\
&\quad + u_{ki}\sum_{n=1}^{N} \mathbb{E}[z_{ni}] \ln x_{kn} + (\mu_{ki_0}-1)\ln u_{ki} - \alpha_{ki_0}u_{ki} + \text{const.}
\end{aligned}
$$

$$
\begin{aligned}
&\approx \mathbb{E}_{\mathcal{Z}\setminus u_{ki}}[\ln \tilde{f}(\mathbf{X},\mathcal{Z})] \\
&= \sum_{n=1}^{N} \mathbb{E}[z_{ni}] \times \left[\psi\left(\sum_{k=1}^{K+1} \bar{u}_{ki}\right) - \psi(\bar{u}_{ki})\right] \bar{u}_{ki} \boxed{\ln u_{ki}} - \ln \bar{u}_{ki}) \\
&\quad + u_{ki}\sum_{n=1}^{N} \mathbb{E}[z_{ni}] \ln x_{kn} + (\mu_{ki_0}-1)\ln u_{ki} - \alpha_{ki_0}u_{ki} + \text{const.},
\end{aligned}
$$

- For the variable z_{ni}, the optimal approximation to the posterior distribution is

$$
\begin{aligned}
\ln f^*(z_{ni}) &= \mathbb{E}_{\mathcal{Z}\setminus z_{ni}}[\ln f(\mathbf{X},\mathcal{Z})] \\
&= z_{ni} \times \left[\mathbb{E}[\ln \pi_i] + \sum_{k=1}^{K+1} (u_{ki}-1)\ln x_{kn}\right] \\
&\quad + z_{ni} \times \boxed{\mathbb{E}\left[\ln \frac{\Gamma(\sum_{k=1}^{K+1} u_{ki})}{\prod_{k=1}^{K+1}\Gamma(u_{ki})}\right]} + \text{const.}
\end{aligned}
$$

$$
\begin{aligned}
&\approx \mathbb{E}_{\mathcal{Z}\setminus z_{ni}}[\ln f(\mathbf{X},\mathcal{Z})] \\
&= z_{ni} \times \left[\mathbb{E}[\ln \pi_i] + \sum_{k=1}^{K+1} (u_{ki}-1) \ln x_{kn}\right] \\
&\quad + z_{ni} \times \underbrace{\sum_{k=1}^{K+1} \left[\psi\left(\sum_{k=1}^{K+1} \bar{u}_{ki}\right) - \psi(\bar{u}_{ki})\right] \bar{u}_{ki} \boxed{\mathbb{E}_{u_{ki}}[\ln u_{ki}]} - \ln \bar{u}_{ki})}_{P_i} + \text{const.}
\end{aligned}
$$

Example: *Variational Inference of Dirichlet mixture model*

Algorithm 1. Variational DMM.

Input: observation $\mathbf{X}$, number of mixture components I
Initialize α_{ki_0}, μ_{ki_0}, c_{i_0}, for $i=1,\dots,I$, $k=1,\dots,K+1$[3];
repeat
for each k, i
$\alpha^*_{ki}=\alpha_{ki_0}-\sum_{n=1}^{N}\mathbb{E}[z_{ni}]\ln x_{kn}$
$\mu^*_{ki}=\mu_{ki_0}+\sum_{n=1}^{N}\mathbb{E}[z_{ni}]\overline{u}_{ki}[\psi(\sum_{k=1}^{K+1}\overline{u}_{ki})-\psi(\overline{u}_{ki})]$
$c^*_i=c_{i_0}+\sum_{n=1}^{N}\mathbb{E}[z_{ni}]$
until stop criteria are reached.
Output: the optimal hyperparameters α^*_{ki}, μ^*_{ki}, c^*_i.
(The quantities $\overline{u}_{ki}$ and $\mathbb{E}[z_{ni}]$ are calculated)

$$\overline{u}_{ki}=\frac{\mu_{ki}}{\alpha_{ki}},\quad \mathbb{E}[z_{ni}]=\frac{\rho_{ni}}{\sum_{i=1}^{I}\rho_{ni}},$$

$$\ln\rho_{ni}=\psi(c_i)-\psi(\mathbf{c}^{\mathrm{T}}\mathbf{1}_I)+P_i+(\mathbf{u}_i-1)^{\mathrm{T}}\ln\mathbf{x}_n$$

Variational distribution for CAT model

- The mean-field variational representation

$$
\begin{aligned}
& q(\Xi, z, V, \pi, l, r \mid \Theta) \\
= & \prod_j q\left(l_j \mid \gamma_j, \gamma_j^{'}\right) \times \prod_i q\left(r_i \mid \zeta_i, \zeta_i^{'}\right) \\
& \times \prod_{k=1}^{K_{\max}} \prod_{a=1}^{20} q\left(\pi_a^k \mid \lambda_a^k\right) \times \prod_{k=1}^{K_{\max}} q\left(V_k \mid \vartheta_k, \vartheta_k^{'}\right) \\
& \times \prod_i \prod_{k=1}^{K_{\max}} q\left(z_i^k \mid \phi_i^k\right) \times \prod_{ij} q\left(n_{ij} \mid \omega_{ij}\right) \\
& \times \prod_{k=1}^{K_{max}} \prod_{a=1}^{20} q\left(w_a^k \mid \iota_a^k\right)
\end{aligned}
$$

$$
\begin{aligned}
q\left(l_j \mid \gamma_j, \gamma_j^{'}\right) &= Gamma\left(l_j \mid \gamma_j, \gamma_j^{'}\right) \\
q\left(r_i \mid \zeta_i, \zeta_i^{'}\right) &= Gamma\left(r_i \mid \zeta_i, \zeta_i^{'}\right) \\
q\left(\pi_a^k \mid \lambda_a^k\right) &= Dirichlet\left(\pi_a^k \mid \lambda_a^k\right) \\
q\left(V_k \mid \vartheta_k, \vartheta_k^{'}\right) &= Beta\left(V_k \mid \vartheta_k, \vartheta_k^{'}\right) \\
q\left(z_i^k \mid \phi_i^k\right) &= Multinomial\left(z_i^k \mid \phi_i^k\right) \\
q\left(n_{ij} \mid \omega_{ij}\right) &= Poisson\left(n_{ij} \mid \omega_{ij}\right) \\
q\left(w_a^k \mid \iota_a^k\right) &= Multinominal\left(w_a^k \mid \iota_a^k\right)
\end{aligned}
$$

Stochastic Variational Inference

- These distributions are in the exponential family

$$p(\pi \mid \Xi, \mathrm{z}, \mathrm{l}, \mathrm{r}) = h(\pi)\exp\Big(\underbrace{\eta(\Xi, \mathrm{z}, \mathrm{l}, \mathrm{r})}_{\textit{Natural parameter}}^{T}\,\underbrace{t(\pi)}_{\textit{Sufficient statistics}} - \underbrace{a}_{\textit{Log-normalizer}}\big(\eta(\Xi, \mathrm{z}, \mathrm{l}, \mathrm{r})\big)\Big)$$

- The variational parameters $q(\pi \mid \lambda) = h(\pi)\exp\big(\lambda^{T} t(\pi) - a(\lambda)\big)$

- The natural gradient of the ELBO [Amari, 1998]

$$\nabla_{\lambda}^{nat} ELBO(\lambda) = \big(prior + NE_{q}\big[t(\Xi, z, l, r)\big]\big) - \lambda$$

- It is good for stochastic optimization.
  - Its expectation is the exact gradient (*unbiased*).
  - It only depends on optimized parameters of one data point (*cheap*).

Variational approximation for model of nucleotide substitution

- **Evidence Lower Bound (ELBO)**

$$
\begin{aligned}
\mathcal{L} = & \sum_{k=1}^{K_{max}} \mathrm{E}_q \left[\log\left(p\left(V_k|1,\kappa\right)\right)\right] - \sum_{k=1}^{K_{\max}} \mathrm{E}_q \left[\log\left(q\left(V_k|\vartheta_k,\vartheta_k^{'}\right)\right)\right] \\
& + \sum_i \sum_{k=1}^{K_{max}} \mathrm{E}_q \left[\log p(Z_i^k|V_1,V_2,...,V_{K_{max}})\right] \\
& - \sum_i \sum_{k=1}^{K_{\max}} \mathrm{E}_q \left[\log\left(q\left(Z_i^k|\phi_i^k\right)\right)\right] \\
& + \sum_{k=1}^{K_{max}} \sum_{a=1}^{20} \mathrm{E}_q \left[\log p(\pi_a^k|\upsilon_a)\right] - \sum_{k=1}^{K_{\max}} \sum_{a=1}^{20} \mathrm{E}_q \left[\log\left(q\left(\pi_a^k|\lambda_a^k\right)\right)\right] \\
& + \sum_j \mathrm{E}_q \left[\log\left(p\left(l_j|1,\beta\right)\right)\right] - \sum_j \mathrm{E}_q \left[\log\left(q\left(l_j|\gamma_j,\gamma_j^{'}\right)\right)\right] \\
& + \sum_i \mathrm{E}_q \left[\log\left(p\left(r_i|\alpha,\alpha\right)\right)\right] - \sum_i \mathrm{E}_q \left[\log\left(q\left(r_i|\zeta_i,\zeta_i^{'}\right)\right)\right] \\
& + \int_\pi \int_r \int_l \sum_z \sum_n q(z)q(n)q(\pi)q(r)q(l) \log\left(p(n|D,z,\pi,r,l)\right) d(\pi)d(r)d(l) \\
& - \sum_n q(n) \log\left(q(n)\right) \\
& + \int_\pi \int_r \int_l \sum_z \sum_w q(z)q(w)q(\pi)q(r)q(l) \log\left(p(w|D,z,\pi,r,l)\right) d(\pi)d(r)d(l) \\
& - \sum_w q(w) \log\left(q(w)\right)
\end{aligned}
$$

Calculations of these integrals are analytically intractable

Variational approximation for model of nucleotide substitution

- We consider a first-order Taylor expansion to preserve a bound, intractable expectations are avoided

The product of $r_i l_j$ is denoted t_{ij}, firstly, we consider a first-order Taylor expansion of $\log\left(p\left(a,b|\pi,t_{ij}\right)\right)$ for π_a and t_{ij} at $\pi_a^{'} = \frac{\upsilon_a}{\sum\limits_{a^{'}=1}^{20} \upsilon_{a^{'}}}$ and $t_{ij}^{'} = \frac{1}{\beta}\frac{\alpha}{\alpha} = \frac{1}{\beta}$

$$
\begin{aligned}
&\log\left[e^{-t_{ij}} + \left(1-e^{-t_{ij}}\right)\pi_a\right] \approx \log\left[e^{-t_{ij}^{'}} + \left(1-e^{-t_{ij}^{'}}\right)\pi_a^{'}\right] \\
&+ \frac{\partial \log\left[e^{-t_{ij}} + \left(1-e^{-t_{ij}}\right)\pi_a^{'}\right]}{\partial t_{ij}}\left(t_{ij}-t_{ij}^{'}\right) \\
&+ \frac{\partial \log\left[e^{-t_{ij}^{'}} + \left(1-e^{-t_{ij}^{'}}\right)\pi_a\right]}{\partial \pi_a}\left(\pi_a-\pi_a^{'}\right)
\end{aligned}
\tag{40}
$$

Stochastic Variational Inference for CAT model

- Iterates t times to update local variational parameters $\Theta_l = \{\vartheta, \vartheta', \phi\}$ of local variables $\Phi_l = \{V, z\}$ based on mapping data

$$\Theta_l^* = \mathrm{E}_{\Theta_g} \{\eta[\Phi_g, \Xi]\}$$

- By using the natural gradient, iterates t times to update global variational parameters $\Theta_g = \{\gamma, \gamma', \zeta, \zeta', \lambda, \omega, \iota\}$ of global variables $\Phi_g = \{\Xi, \pi, l, r\}$ based on mapping data with step size

$$\widehat{\nabla_{\Theta_g}} \mathcal{L} = prior + N\{E_{\Theta_l}[t(\Phi_n, \Xi_n), 1]\} - \Theta_g$$

$$\Theta_g^{(t)} = \Theta_g^{(t-1)} + \rho_t \widehat{\nabla_{\Theta_g}} \mathcal{L}$$

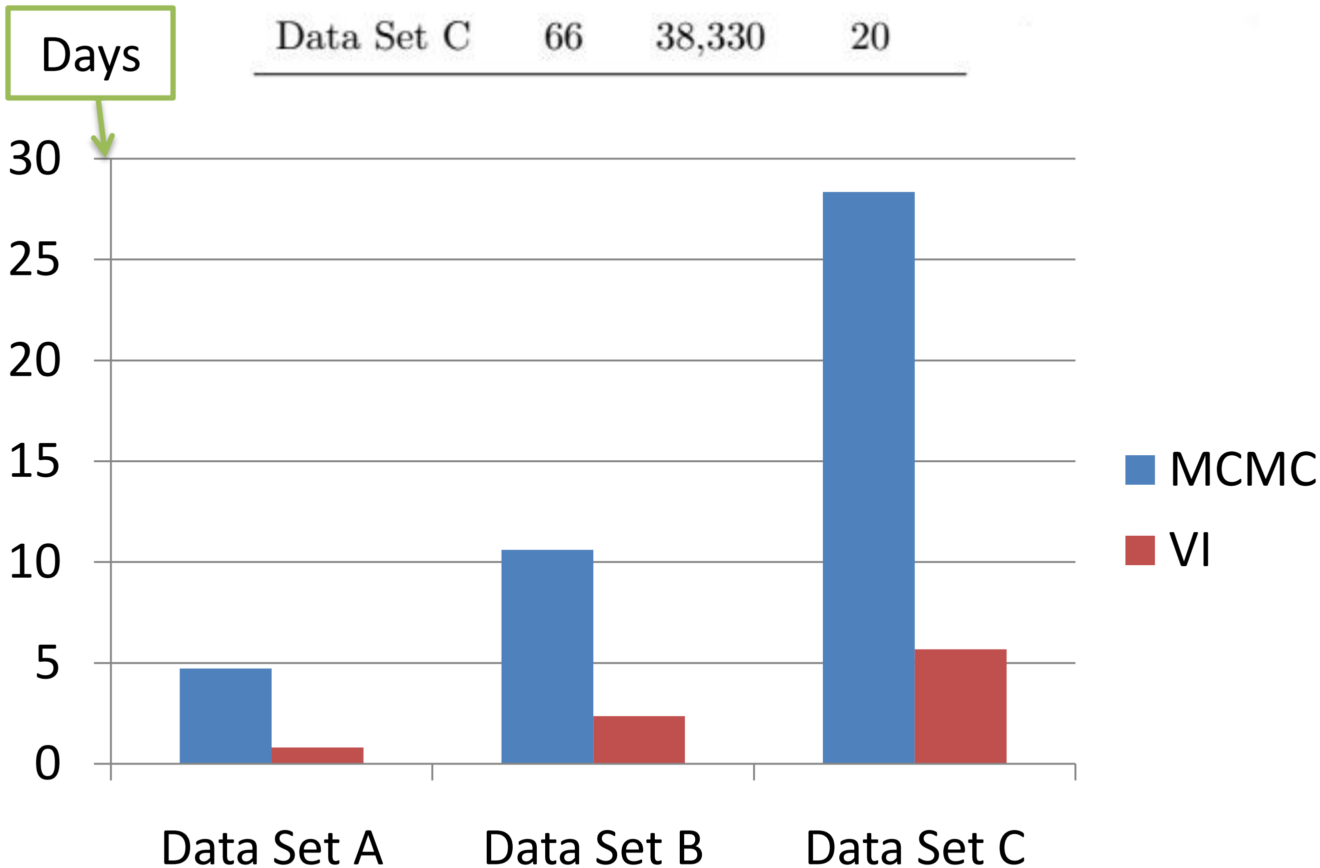
Checking the performance

- the mitochondrial data set which consisted of 33 proteins, a total of 6,622 amino acid positions, from 13 species (data set A).
- the plastid data set which consisted of 50 plastidencoded proteins, a total of 10,137 amino acid positions, from 28 species (data set B).
- mitochondrial protein sequences, a large alignment from EST and genome data, which consists of 197 genes, a total of 38,330 amino-acid positions from 66 species (data set C).

CPU time

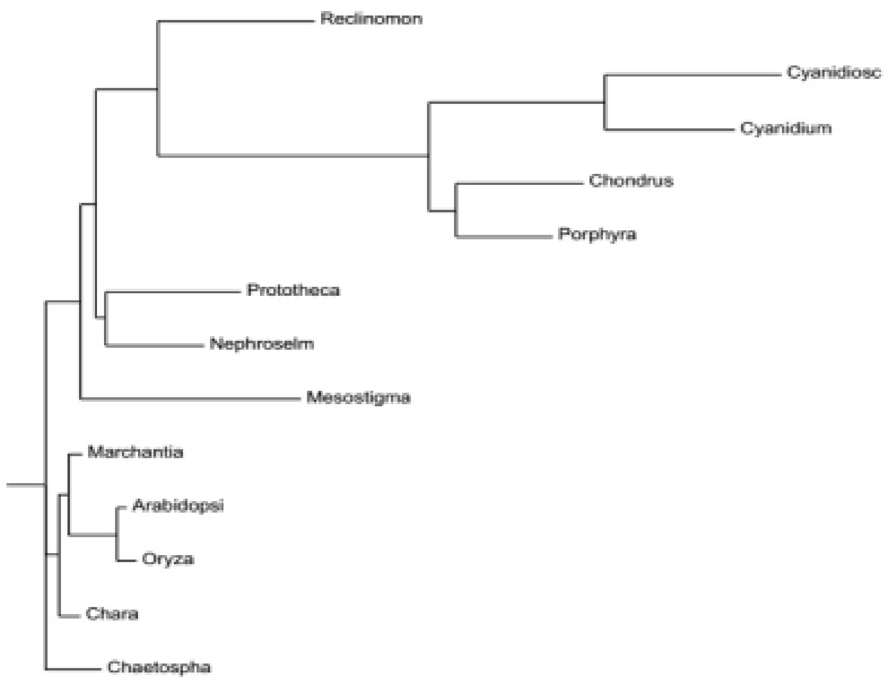
Run times of variational inference and MCMC algorithms on real data

Data set	Taxa	Sites	States
Data Set A	13	6,622	20
Data Set B	28	10,137	20
Data Set C	66	38,330	20

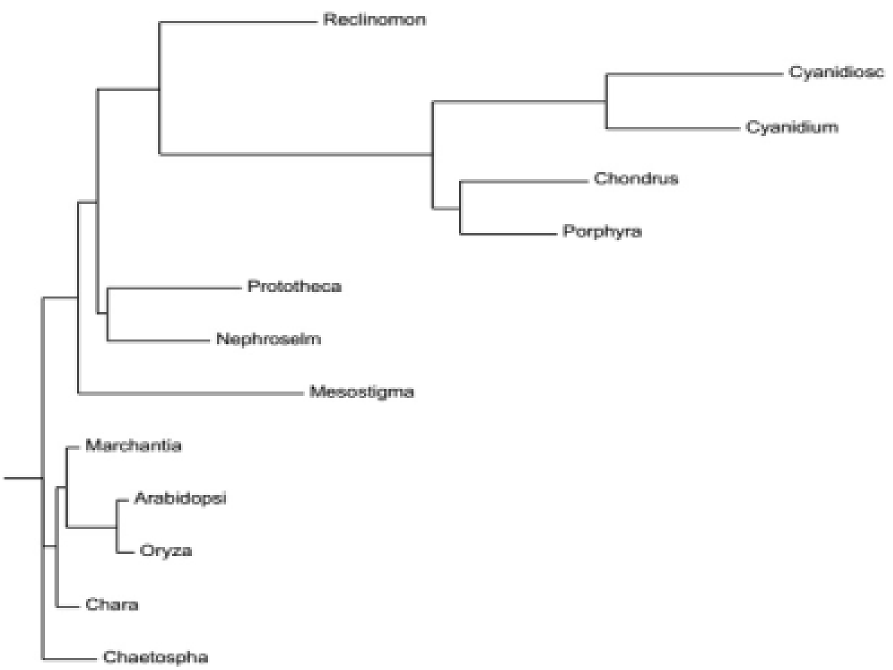


Reconstructed tree

MCMC

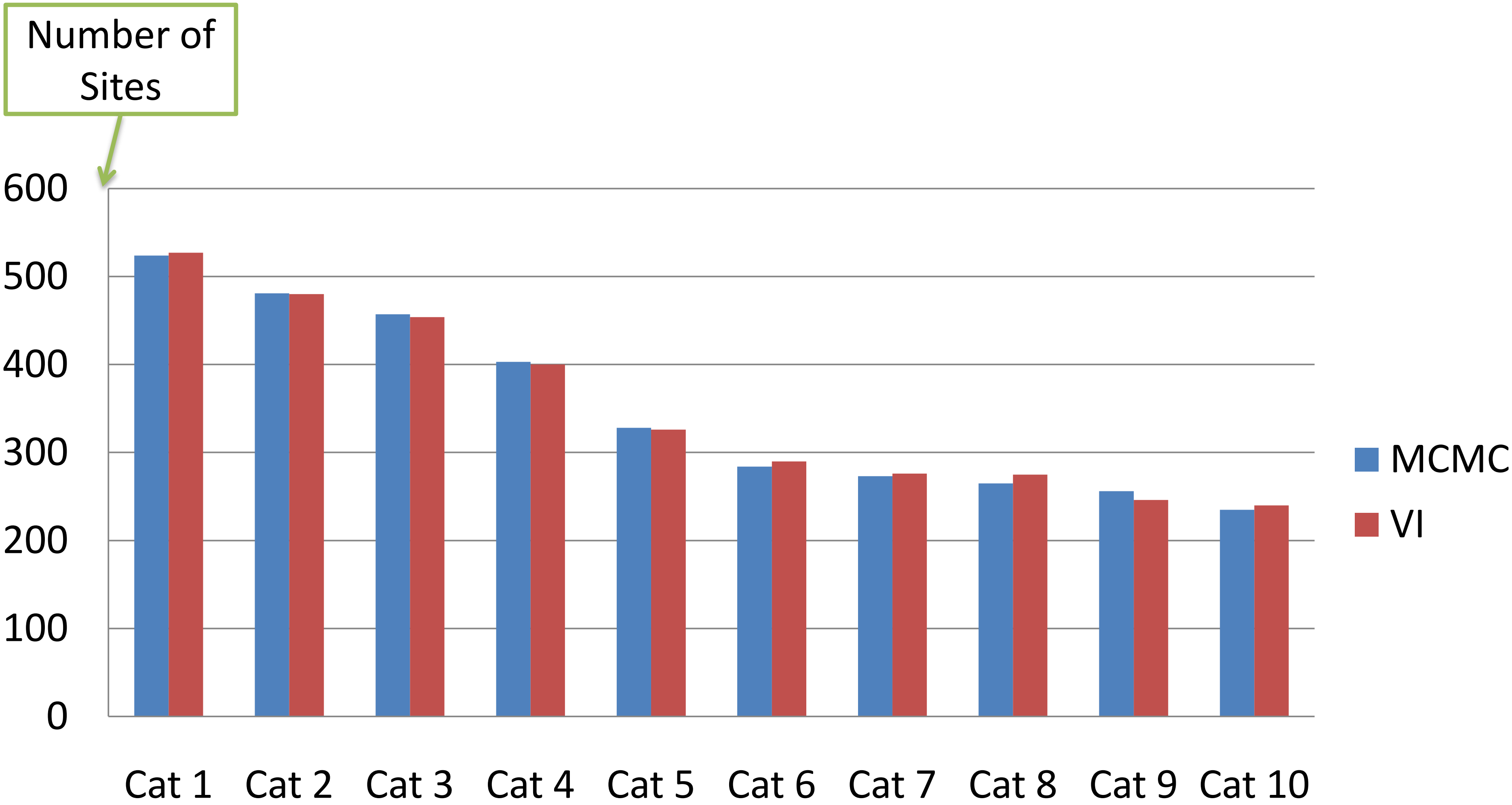


VI



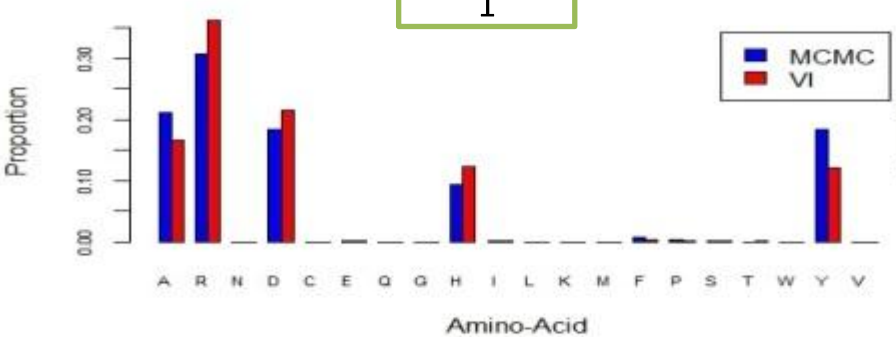
A mitochondrial data set (13 taxa and 6,622 amino acid positions (Rodr´ıguez-Ezpeleta *et al.* 2006))

Size of profile categories

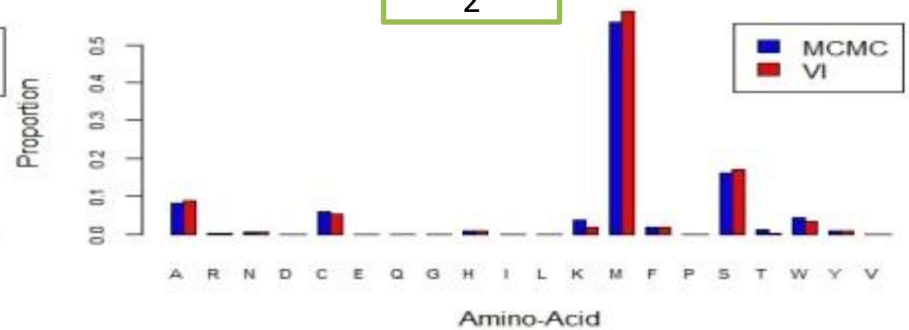


Look into the profile of each category

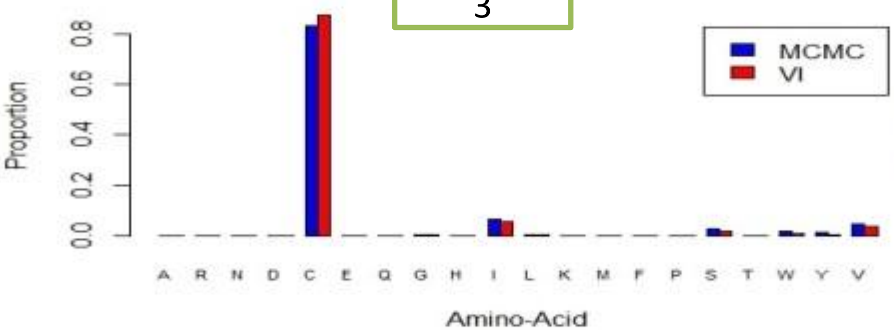
Category 1



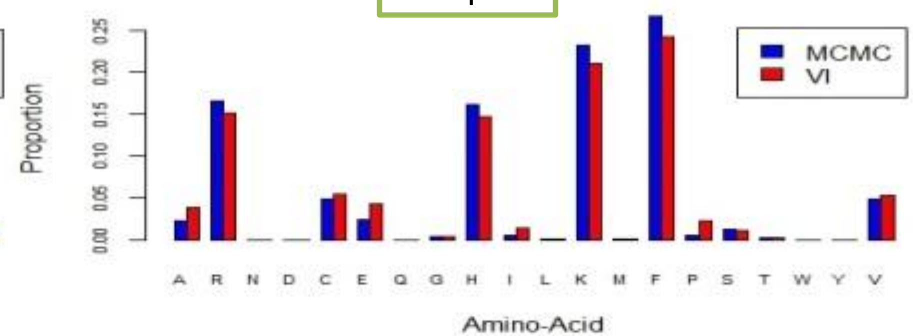
Category 2



Category 3



Category 4



A mitochondrial data set (13 taxa and 6,622 amino acid positions (Rodr´ıguez-Ezpeleta *et al.* 2006))

Summary

- The pattern of molecular evolution varies among gene sites and genes in a genome.
- By taking into account the complex heterogeneity of evolutionary processes among sites in a genome, Bayesian infinite mixture models of genomic evolution enable robust phylogenetic inference.
- With large modern data sets, however, the computational burden of Markov chain Monte Carlo sampling techniques becomes prohibitive.
- Here, we have developed a variational Bayesian procedure to speed up the widely used PhyloBayes MPI program, which deals with the heterogeneity of amino acid propensity.