

Markov Chain Monte Carlo

Metropolis-Hastings and related algorithms

I. Holmes

Department of Bioengineering
University of California, Berkeley

Spring semester

Outline

- 1 Conjugate prior distributions
 - Gamma distribution
 - Dirichlet distribution
 - Normal-gamma distribution
 - Summary
- 2 MCMC in theory
 - Metropolis-Hastings
 - Gibbs sampling
 - Jacobians
 - MCMC on continuous parameters
- 3 MCMC in bioinformatics
 - Alignments
 - Trees
 - Structures

Motivation

- $P(\theta|x)$ and $P(\theta)$ are **conjugate** if the posterior $P(\theta|x)$ is of the same family as the prior $P(\theta)$

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)} = \frac{P(x|\theta)P(\theta)}{\int P(x, \theta')d\theta'}$$

- The denominator is $P(x)$, the **Bayesian evidence**, which does not depend on θ .
- The dependence of $P(\theta|x)$ on θ is same as that of $P(x|\theta)P(\theta)$, but to get a normalized form for $P(\theta|x)$, we need to integrate out θ to find the evidence $P(x)$.
- The parameters of the conjugate prior distribution are called **hyperparameters**.

Motivation

$$\begin{aligned}L(\theta) &= P(\text{data } D | \text{params } \theta) &&= \text{likelihood} \\F(\theta, \alpha) &= P(\theta | \text{hyperparams } \alpha) &&= \text{prior} \\G(\theta) &= P(\theta | \alpha, D) &&= \text{posterior}\end{aligned}$$

L and F are conjugate if $G(\theta) = F(\theta, \alpha')$ where $\alpha' \equiv \alpha'(\alpha, D)$

Common conjugacies

Likelihood	Prior
Exponential	Gamma
Binomial	Beta
Multinomial	Dirichlet
Gaussian (fixed precision)	Gaussian
Gaussian (varying precision)	Gaussian + Gamma

Outline

- 1 Conjugate prior distributions
 - Gamma distribution
 - Dirichlet distribution
 - Normal-gamma distribution
 - Summary
- 2 MCMC in theory
 - Metropolis-Hastings
 - Gibbs sampling
 - Jacobians
 - MCMC on continuous parameters
- 3 MCMC in bioinformatics
 - Alignments
 - Trees
 - Structures

Gamma distribution

- Exponential, Poisson, Gamma distributions
 - Exponential distribution: pdf for time to first event, T , given that mean event rate is μ

$$P(T|\mu) = \mu \exp(-\mu T)$$

- Poisson: probability distribution of number of events, n , in time T , given that mean event rate is μ

$$P(n|\mu) = \frac{(\mu T)^n \exp(-\mu T)}{n!}$$

(NB exponential distribution can be obtained by setting $n = 0$ and differentiating w.r.t. T)

Gamma distribution

- Poisson

$$P(n|\mu) = \frac{(\mu T)^n \exp(-\mu T)}{n!}$$

- Gamma, conjugate to Poisson.

Shape parameter α , rate parameter β .

$$P(\mu|\alpha, \beta) = \frac{\mu^{\alpha-1} \beta^\alpha \exp(-\mu\beta)}{\Gamma(\alpha)}$$

where Γ is the **gamma function**

$$\Gamma(\alpha) = \int_0^\infty (\mu\beta)^{\alpha-1} \exp(-\mu\beta) d(\mu\beta)$$

Gamma function

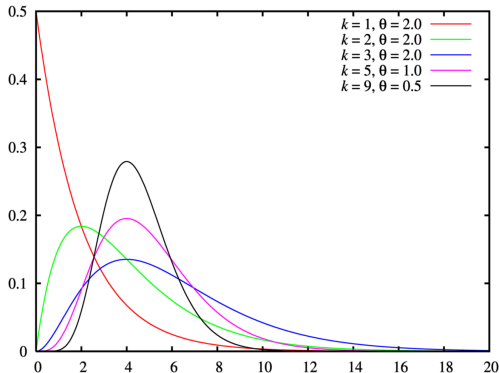
Changing variables in the gamma function integral

$$\Gamma(z) = \int_{u=0}^{\infty} u^{z-1} \exp(-u) du$$

Clearly $\Gamma(1) = 1$. Integrating by parts for positive integer z ,

$$\begin{aligned} \Gamma(z+1) &= \int_{u=0}^{\infty} u^z \exp(-u) du \\ &= [-u^z \exp(-u)]_{u=0}^{\infty} + z\Gamma(z) \\ &= z! \end{aligned}$$

Gamma distribution



$$k = \alpha, \theta = 1/\beta$$

θ is called the scale parameter.

Gamma distribution

- Properties of gamma distribution:
mean of μ is α/β and variance is α/β^2 .
- Conjugacy:

$$P(n) = \frac{\Gamma(\alpha')}{\Gamma(n+1)\Gamma(\alpha)} \frac{T^n \beta^\alpha}{(\beta')^{(\alpha')}} \\ P(\mu|n) = \frac{\mu^{\alpha'-1} (\beta')^{\alpha'} \exp(-\mu\beta')}{\Gamma(\alpha')}$$

i.e. posterior for μ is a gamma distribution with shape $\alpha' = \alpha + n$ and rate $\beta' = \beta + t$.

- α and β are like a *pseudocount* and *pseudotime*.

Gamma distribution

- Lengthier derivation:

$$\begin{aligned}
 P(n) &= \int_{\mu=0}^{\infty} P(n|\mu)P(\mu)d\mu \\
 &= \int_{\mu=0}^{\infty} \frac{(\mu T)^n \exp(-\mu T)}{\Gamma(n+1)} \frac{\mu^{\alpha-1} \beta^{\alpha} \exp(-\mu\beta)}{\Gamma(\alpha)} d\mu \\
 &= \frac{T^n \beta^{\alpha}}{\Gamma(n+1)\Gamma(\alpha)} \int_{\mu=0}^{\infty} \mu^{n+\alpha-1} \exp(-\mu(T+\beta)) d\mu \\
 &= \frac{T^n \beta^{\alpha}}{\Gamma(n+1)\Gamma(\alpha)} \frac{1}{(T+\beta)^{n+\alpha}} \int_{u=0}^{\infty} u^{n+\alpha-1} \exp(-u) du \\
 &= \frac{\Gamma(\alpha')}{\Gamma(n+1)\Gamma(\alpha)} \frac{T^n \beta^{\alpha}}{(\beta')^{(\alpha')}}
 \end{aligned}$$

Outline

- 1 **Conjugate prior distributions**
 - Gamma distribution
 - **Dirichlet distribution**
 - Normal-gamma distribution
 - Summary
- 2 **MCMC in theory**
 - Metropolis-Hastings
 - Gibbs sampling
 - Jacobians
 - MCMC on continuous parameters
- 3 **MCMC in bioinformatics**
 - Alignments
 - Trees
 - Structures

Dirichlet distribution

- Multinomial, Dirichlet distributions
 - Multinomial: K possible outcomes, outcome probabilities \mathbf{p} , outcome frequencies \mathbf{n} in $N = \sum_k n_k$ trials

$$P(\mathbf{n}|\mathbf{p}) = \frac{N!}{\prod_k n_k!} \prod_k p_k^{n_k}$$

- Conjugate prior: let \mathbf{a} be a vector of **pseudocounts**, $a_1 \dots a_K$. The **Dirichlet distribution** for \mathbf{p} is

$$P(\mathbf{p}|\mathbf{a}) = \frac{\prod_i p_i^{a_i-1}}{\mathcal{B}(\mathbf{a})} \delta \left(\sum_i p_i - 1 \right)$$

where $\mathcal{B}()$ is the **type one Dirichlet integral** or the **multinomial beta function**

$$\mathcal{B}(\mathbf{a}) = \int \left(\prod_i p_i^{a_i-1} \right) \delta \left(\sum_i p_i - 1 \right) d\mathbf{p} = \frac{\prod_k \Gamma(a_k)}{\Gamma(\sum_k a_k)}$$

Dirichlet distribution

- Properties: mean value of p_k is $a_k / \sum_j a_j$. Modal value is $p_k = (a_k - 1) / (\sum_j a_j - 1)$.
- Note the relationship between the multinomial coefficient and \mathcal{B}

$$\frac{N!}{\prod_k n_k!} = \frac{1}{\mathcal{B}(\mathbf{n} + \mathbf{1})} \times \frac{\Gamma(N + 1)}{\Gamma(N + K)}$$

- Conjugacy:

$$P(\mathbf{n}|\mathbf{a}) = \frac{\mathcal{B}(\mathbf{a}')}{\mathcal{B}(\mathbf{a})} \frac{N!}{\prod_k n_k!}$$

$$P(\mathbf{p}|\mathbf{n}, \mathbf{a}) = \frac{\prod_i p_i^{a'_i - 1}}{\mathcal{B}(\mathbf{a}')} \delta \left(\sum_i p_i - 1 \right)$$

where $\mathbf{a}' = \mathbf{a} + \mathbf{n}$ (hence the name “pseudocount”).

Dirichlet distribution

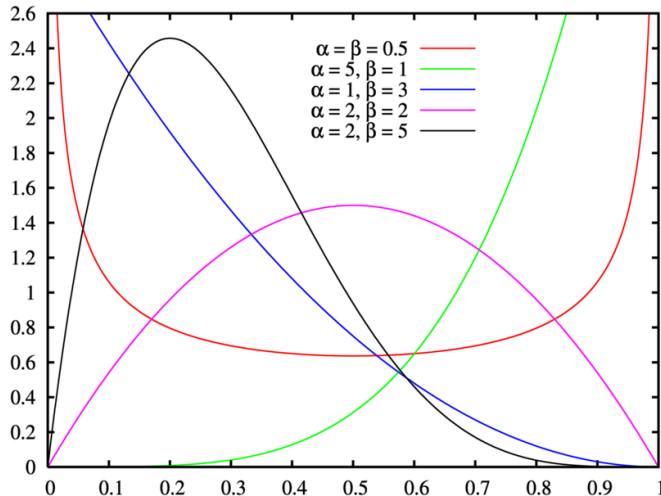
- Special case of {Multinomial, Dirichlet} when $K = 2$ is {Binomial, Beta}, and \mathcal{B} is the **beta function**.
- The following snippets from Mathworld point to a more rigorous solution of the type one Dirichlet integral
 - The beta integral may be obtained by writing $m!n!$ as a product of gamma functions with integrands u, v , transforming to $(x, y) = (\sqrt{u}, \sqrt{v})$ and then to polar co-ordinates $(x, y) = r(\cos \theta, \sin \theta)$. This yields

$$\mathcal{B}(m+1, n+1) = 2 \int_0^{\pi/2} \cos^{2m+1} \theta \sin^{2n+1} \theta d\theta = \frac{m!n!}{(m+n+1)!}$$

See mathworld.wolfram.com/BetaFunction.html for details.

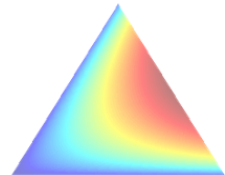
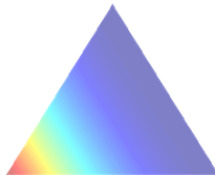
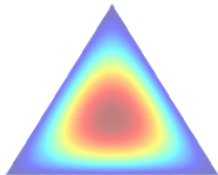
- More info on Dirichlet type one integral at mathworld.wolfram.com/DirichletIntegrals.html

Beta distribution



Dirichlet distribution

- For $K = 3$, the probabilities (p_x, p_y, p_z) satisfy $p_x + p_y + p_z = 1$
- Like 3D points lying on the 2D triangle with corners at $(0, 0, 1)$, $(0, 1, 0)$ and $(1, 0, 0)$



Outline

- 1 **Conjugate prior distributions**
 - Gamma distribution
 - Dirichlet distribution
 - **Normal-gamma distribution**
 - Summary
- 2 **MCMC in theory**
 - Metropolis-Hastings
 - Gibbs sampling
 - Jacobians
 - MCMC on continuous parameters
- 3 **MCMC in bioinformatics**
 - Alignments
 - Trees
 - Structures

Normal-gamma distribution

- Normal (Gaussian) and Normal-gamma distributions
 - Mean μ , precision τ (precision is reciprocal of variance).
Data $\{x_i\}$ with moments $m_k = \sum_i x_i^k$

$$\begin{aligned} P(\mathbf{x}|\mu, \tau) &= \prod_i \sqrt{\frac{\tau}{2\pi}} \exp\left(-\frac{\tau}{2}(x_i - \mu)^2\right) \\ &= \left(\frac{\tau}{2\pi}\right)^{m_0/2} \exp\left(-\frac{\tau}{2}(m_2 - 2\mu m_1 + \mu^2 m_0)\right) \end{aligned}$$

Normal-gamma distribution

- Normal

$$P(\mathbf{x}|\mu, \tau) = \left(\frac{\tau}{2\pi}\right)^{m_0/2} \exp\left(-\frac{\tau}{2}(m_2 - 2\mu m_1 + \mu^2 m_0)\right)$$

- Conjugate prior

$$P(\mu, \tau) = P(\tau)P(\mu|\tau)$$

$$P(\tau) = \frac{e^{-\beta\tau} \tau^{\alpha-1} \beta^\alpha}{\Gamma(\alpha)}$$

$$P(\mu|\tau) = \sqrt{\frac{\lambda\tau}{2\pi}} \exp\left(-\frac{\lambda\tau}{2}(\mu - \epsilon)^2\right)$$

- $P(\tau)$ is gamma (shape α , rate β)
- $P(\mu|\tau)$ is Normal (mean ϵ , precision $\lambda\tau$)
- Full set of hyperparameters is $\{\alpha, \beta, \epsilon, \lambda\}$

Normal-gamma distribution

- Conjugacy:

$$P(\mathbf{x}) = (2\pi)^{-m_0/2} \frac{\Gamma(\alpha')}{\Gamma(\alpha)} \frac{\beta^\alpha}{\beta'^{\alpha'}} \sqrt{\frac{\lambda}{\lambda'}}$$

$$P(\mu, \tau | \mathbf{x}) = \frac{e^{-\beta' \tau} \tau^{\alpha'-1} \beta'^{\alpha'}}{\Gamma(\alpha')} \times \sqrt{\frac{\lambda' \tau}{2\pi}} \exp\left(-\frac{\lambda' \tau}{2} (\mu - \epsilon')^2\right)$$

where $\epsilon' = \frac{\lambda\epsilon + m_1}{\lambda + m_0}$, $\lambda' = \lambda + m_0$, $\alpha' = \alpha + \frac{m_0}{2}$ and $\beta' = \beta + \frac{1}{2} \left(\epsilon^2 + m_2 - \frac{(\lambda\epsilon + m_1)^2}{\lambda + m_0} \right)$.

- NB if we regard τ as fixed, then Gaussian is auto-conjugate.

χ^2 distribution

$$P(\mu, \tau | \mathbf{x}) = \frac{e^{-\beta' \tau} \tau^{\alpha' - 1} \beta'^{\alpha'}}{\Gamma(\alpha')} \times \sqrt{\frac{\lambda' \tau}{2\pi}} \exp\left(-\frac{\lambda' \tau}{2} (\mu - \epsilon')^2\right)$$

- Rescale the posterior distribution for τ : divide by $(m_2/m_0 - m_1^2/m_0)^{-1}$ (the posterior mean estimate for τ)
- Resulting gamma distribution with shape $m_0/2$ and scale $1/2$ is the “ χ^2 distribution”.
- Canonical definition of the quantity χ^2 is in the special case where $\mu = 0$ and $\tau = 1$, in which case $\chi^2 = m_2$

Outline

- 1 **Conjugate prior distributions**
 - Gamma distribution
 - Dirichlet distribution
 - Normal-gamma distribution
 - **Summary**
- 2 **MCMC in theory**
 - Metropolis-Hastings
 - Gibbs sampling
 - Jacobians
 - MCMC on continuous parameters
- 3 **MCMC in bioinformatics**
 - Alignments
 - Trees
 - Structures

Common conjugacies

- Exponential/Poisson \leftrightarrow Gamma
- Binomial \leftrightarrow Beta
- Multinomial \leftrightarrow Dirichlet
- Normal \leftrightarrow Normal-gamma

Outline

- 1 Conjugate prior distributions
 - Gamma distribution
 - Dirichlet distribution
 - Normal-gamma distribution
 - Summary
- 2 MCMC in theory
 - Metropolis-Hastings
 - Gibbs sampling
 - Jacobians
 - MCMC on continuous parameters
- 3 MCMC in bioinformatics
 - Alignments
 - Trees
 - Structures

Motivation

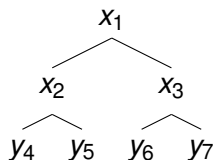
- Suppose we have a model e.g.
 - Missing data X , observed data Y , parameters θ
 - Prior $P(\theta)$, likelihood $P(X, Y|\theta)$
- Want to sample from some related probability distribution, e.g.
 - Posterior for missing data: $P(X|Y, \theta)$
 - Posterior for parameters: $P(\theta|Y)$

Phylogenetic MCMC

Tree-related applications of MCMC

- Sampling from posterior of models, parameters
 - Tree topologies, branch lengths
 - Structures, alignments
 - Phylogeography
- What about ancestral sequence itself?
 - Felsenstein's algorithm works great when state space is small
 - Nucleotides, amino acids
 - What about when state space is large?

Ancestral reconstruction MCMC



$$P(X|Y) = \frac{P(X,Y)}{P(Y)} = \frac{P(X,Y)}{\sum_{X'} P(X',Y)}$$

- Let A = alphabet size
- Felsenstein's pruning algorithm for $P(Y)$
 - Time $\mathcal{O}(NA^2)$
 - Memory $\mathcal{O}(NA)$
- Exponentiating an $A \times A$ matrix N times
 - Time $\mathcal{O}(NA^3)$
- OK: bases ($A = 4$), amino acids ($A = 20$), codons ($A = 64$)
 - Painful: Gene Ontology ($A = 33,587$), GO-Slim ($A = 127$)

MCMC: general idea

- Want to sample from some pdf $P(X) = \frac{f(X)}{Z}$ where $X \in \mathcal{X}$
- Problem: set \mathcal{X} is large; computing $Z = \sum_{X \in \mathcal{X}} f(X)$ is hard
- Construct a stochastic process $X_1, X_2, X_3, X_4, \dots$ such that the equilibrium distribution is $P(X)$
- If the process is ergodic, then $X_t \sim P(X)$ for “large” t
 - How large must t be? The **burn-in time** or **mixing time**

MCMC: construction

- Want a Markov process X_1, X_2, \dots w/equilibrium $P(X)$
- Let $Q(i, j) = P(X_{t+1} = j | X_t = i)$ be the **transition matrix**
- One way to make $P(X)$ the equilibrium distribution is to force Q to satisfy **detailed balance**

$$P(i)Q(i, j) = P(j)Q(j, i) \quad \forall i, j \in \mathcal{X}$$

- If $P(X) = \frac{f(X)}{Z}$ we can drop the Z

$$f(i)Q(i, j) = f(j)Q(j, i)$$

- Note this process is **reversible** (MCMC doesn't have to be, but usually is)

Metropolis-Hastings algorithm: the proposal

- One way to construct a Q that satisfies detailed balance

$$f(i)Q(i, j) = f(j)Q(j, i)$$

- Start with a **proposal distribution**

$$U(i, j) = P(X_{t+1} = j | X_t = i)$$

- Metropolis/Teller (1953): **symmetric** proposal,
 $U(i, j) = U(j, i)$
- Hastings (1970) generalized to **reversible** proposal

$$g(i)U(i, j) = g(j)U(j, i)$$

- Note $g \neq f$: proposal does **not** have to converge on f
- Now we adapt this U to construct Q

Metropolis-Hastings algorithm: accept/reject

- We have a proposal density, U , satisfying (for some g)

$$g(i)U(i, j) = g(j)U(j, i)$$

- Process:

- Given X_n , **propose a move** by sampling X' from

$$P(X' = j | X_n = i) = U(i, j)$$

- Calculate the **Hastings ratio**

$$h(X, X') = \frac{f(X')}{f(X)} \frac{g(X)}{g(X')} = \frac{f(X')}{f(X)} \frac{U(X', X)}{U(X, X')}$$

- If $h(X, X') \geq 1$, then **accept the move**: set $X_{n+1} \leftarrow X'$
- If $h(X, X') < 1$, then
 - Sample an r.v. α from $U(0, 1)$ (uniform distribution)
 - If $\alpha < h(X, X')$, then accept the move
 - If $\alpha \geq h(X, X')$, then **reject the move**: set $X_{n+1} \leftarrow X_n$

Metropolis-Hastings: proof of detailed balance

$$h(i, j) = \frac{f(j)}{f(i)} \frac{U(j, i)}{U(i, j)}$$

$$Q(i, j) = U(i, j) \times \min(1, h(i, j))$$

- Suppose (with no loss of generality) that $h(i, j) \leq 1$

$$\begin{aligned} Q(i, j) &= U(i, j) h(i, j) \\ &= U(i, j) \frac{f(j)}{f(i)} \frac{U(j, i)}{U(i, j)} \\ &= \frac{f(j)}{f(i)} U(j, i) \end{aligned}$$

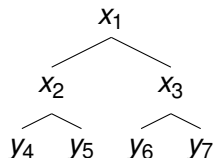
- Since $h(j, i) \geq 1$, we must have $Q(j, i) = U(j, i)$, and so

$$f(i) Q(i, j) = f(j) Q(j, i)$$

Outline

- 1 Conjugate prior distributions
 - Gamma distribution
 - Dirichlet distribution
 - Normal-gamma distribution
 - Summary
- 2 MCMC in theory
 - Metropolis-Hastings
 - **Gibbs sampling**
 - Jacobians
 - MCMC on continuous parameters
- 3 MCMC in bioinformatics
 - Alignments
 - Trees
 - Structures

Gibbs sampling



- The X in $P(X)$ is N -dimensional, $X = (x_1, x_2 \dots x_N)$
- **Gibbs Sampling** proposal distribution is as follows
 - Pick a dimension m , where $1 \leq m \leq N$
 - Conceptually, fix all of the $\{x_n\}$ except for x_m ; sample x_m from its marginal distribution conditional on all the others
 - Proposal distribution is

$$\begin{aligned}
 U(X, X') &= U(\{x_n\}, \{x'_n\}) \\
 &= P(x'_m | \{x_n : n \neq m\}) \times \prod_{n \neq m} \delta(x'_n = x_m)
 \end{aligned}$$

- All proposals accepted: Hastings ratio always 1 (prove it!)

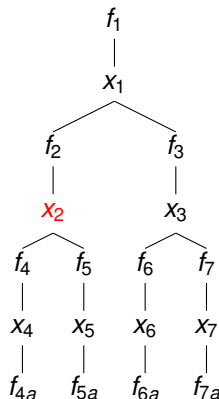
Gibbs sampling: the marginal distribution

Let $X_{\text{other}} = \{x_n : n \neq m\}$

$$\begin{aligned} P(x'_m | X_{\text{other}}) &= \frac{P(x'_m, X_{\text{other}})}{P(X_{\text{other}})} \\ &= \frac{P(x'_m, X_{\text{other}})}{\sum_k P(x'_m = k, X_{\text{other}})} \end{aligned}$$

If $P(X)$ is a product of functions of sparse subsets of X (c.f. factor graphs), then many of these functions will cancel in numerator and denominator

Gibbs sampling: factor graph example



$$P(x'_2 | X_{\text{other}}) = \frac{f_2(x_1, x'_2) f_4(x'_2, x_4) f_5(x'_2, x_5)}{\sum_k f_2(x_1, k) f_4(k, x_4) f_5(k, x_5)}$$

Outline

- 1 Conjugate prior distributions
 - Gamma distribution
 - Dirichlet distribution
 - Normal-gamma distribution
 - Summary
- 2 MCMC in theory
 - Metropolis-Hastings
 - Gibbs sampling
 - **Jacobians**
 - MCMC on continuous parameters
- 3 MCMC in bioinformatics
 - Alignments
 - Trees
 - Structures

The Jacobian matrix and its determinant

- Function $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$
- Let $(y_1 \dots y_m) = F(x_1 \dots x_n)$
- The **Jacobian matrix** of F is

$$J_F(x_1, \dots, x_n) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \dots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

- For square matrices ($m = n$), the magnitude of the **Jacobian determinant**, $|\det(J_F)|$, gives the factor by which the function F expands or shrinks volumes in \mathbf{x}

Multidimensional change of variables

- If $p(x)$ is pdf of x , and $y = F(x)$, what is pdf $q(y)$ of y ?
 - Now with multidimensional $x = (x_1 \dots x_n)$ and $y = (y_1 \dots y_n)$
- Derivation:

$$\begin{aligned}
 dy_1 dy_2 \dots dy_n &= |\det(J_F)| dx_1 dx_2 \dots dx_n \\
 q(y) dy_1 dy_2 \dots dy_n &= p(x) dx_1 dx_2 \dots dx_n \\
 q(y) &= p(x) / |\det(J_F)| \\
 &= p(F^{-1}(y)) |\det(J_{F^{-1}})|
 \end{aligned}$$

where the last line uses the **inverse function theorem**,

$$J_{F^{-1}} = (J_F)^{-1} \quad \dots \text{ c.f. } \textbf{chain rule: } J_{F \circ G}(x) = J_F(G(x)) J_G(x)$$

Outline

- 1 Conjugate prior distributions
 - Gamma distribution
 - Dirichlet distribution
 - Normal-gamma distribution
 - Summary
- 2 MCMC in theory
 - Metropolis-Hastings
 - Gibbs sampling
 - Jacobians
 - MCMC on continuous parameters
- 3 MCMC in bioinformatics
 - Alignments
 - Trees
 - Structures

MCMC on continuous variables

- Let x be a continuous state
- Suppose the proposal generates x' by sampling some parameter u from a p.d.f. $g(u)$, so $x' = x'(x, u)$
 - Let $u' = u'(x, u)$ be the **inverse** parameter that takes us back from $x' \rightarrow x$
 - For convenience, define the function $F(x, u) = (x', u')$
- The Hastings ratio then depends on the Jacobian J_F

$$h(x, x') = \frac{P(x')}{P(x)} \frac{g(u')}{g(u)} |\det(J_F)|$$

- Interpretation: we are moving through (x, u) -space rather than just x -space
- Green (1995, 2003)

There's much more

- Summarizing results of an MCMC run
- Diagnosing MCMC convergence
 - Replicates
 - Intra- and inter-chain variance

Not covered in depth here.

Outline

- 1 Conjugate prior distributions
 - Gamma distribution
 - Dirichlet distribution
 - Normal-gamma distribution
 - Summary
- 2 MCMC in theory
 - Metropolis-Hastings
 - Gibbs sampling
 - Jacobians
 - MCMC on continuous parameters
- 3 MCMC in bioinformatics
 - **Alignments**
 - Trees
 - Structures

Sampling ungapped sequence alignments

- Example: ungapped sequence alignments (Lawrence *et al*)
 - Suppose you have K sequences, $\{S^{(1)} \dots S^{(K)}\}$, and that $S_i^{(k)}$ is the i 'th residue of the k 'th sequence.
 - Let x_k be the indentation of the k 'th sequence. The motif (of length M) runs from x_k to $x_k + M - 1$.

Sampling ungapped alignments

- Let \mathcal{U} denote the null hypothesis that the sequences are unrelated. Then

$$\frac{P(\mathbf{x})}{P(\mathcal{U})} = \prod_{m=0}^{M-1} \frac{f(\mathbf{n}(m))}{\prod_{\omega} q(\omega)^{n_{\omega}(m)}}$$

where $q(\omega)$ is a background distribution over alphabet symbols ω , and $n_{\omega}(m)$ is the number of times symbol ω appears in column m of the motif

$$n_{\omega}(m) = \sum_{k=1}^K \delta(S_{x_k+m} = \omega)$$

Sampling ungapped alignments

- The $f(\mathbf{n})$ function should reward conserved columns.
Lawrence *et al* use an entropy-like measure

$$f(\mathbf{n}) = \prod_{\omega} p_{\omega}^{n_{\omega}} = \exp(-KS[p])$$

where $p_k = n_k/K$. Note that this is the distribution \mathbf{p} that maximizes $\prod_{\omega} p_{\omega}^{n_{\omega}}$ for a given \mathbf{n} . Due to this implicit maximization, the above $f(\mathbf{n})$ is not strictly a probability distribution for \mathbf{n} .

- Typically a vector of pseudocounts \mathbf{a} is added to \mathbf{n} . This suggests the Dirichlet evidence as an alternative form for f that does not involve any implicit maximization:

$$f(\mathbf{n}) = \frac{\mathcal{B}(\mathbf{n} + \mathbf{a})}{\mathcal{B}(\mathbf{a})\mathcal{B}(\mathbf{n})}$$

Sampling ungapped alignments

- Another way to formulate this problem is as a chain over $(\mathbf{x}, \{\mathbf{p}(m)\})$ with Gibbs-sampling steps that alternate between resampling one of the x_k 's and resampling all the \mathbf{p} 's.
- For good mixing it's also useful to allow moves that slide the entire motif window, i.e. $\mathbf{x} \leftarrow \mathbf{x} \pm \mathbf{1}$

Outline

- 1 Conjugate prior distributions
 - Gamma distribution
 - Dirichlet distribution
 - Normal-gamma distribution
 - Summary
- 2 MCMC in theory
 - Metropolis-Hastings
 - Gibbs sampling
 - Jacobians
 - MCMC on continuous parameters
- 3 MCMC in bioinformatics
 - Alignments
 - **Trees**
 - Structures

Sampling trees

- Priors: coalescent
- Moves:
 - Topology: prune-and-graft operations, swapping nodes & branches
 - Branch lengths: rescaling, sliding
- Equivalent ML moves:
 - quartet-puzzling (Strimmer & von Haeseler)
 - pplacer's "pendant branch length" (Matsen)

Drummond et al (2002)

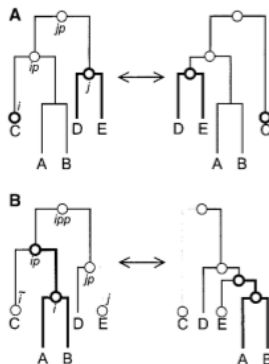


FIGURE 1.—Diagrams of two proposal mechanisms used to modify tree topology during an MCMC analysis. (A) This move is called the “narrow exchange” and is similar to a nearest neighbor interchange. This move picks two subtrees at random under the constraint that they have an aunt-niece relationship; *i.e.*, the parent of one is the grandparent of the other, but neither is parent of the other. Once picked these two subtrees are swapped so long as doing so does not require any modifications in node heights to maintain parent-child order constraints. (B) This move is similar to one proposed by WILSON and BALDING (1998) and involves removing a subtree and reattaching it on a new parent branch.

Larget et al (2005)

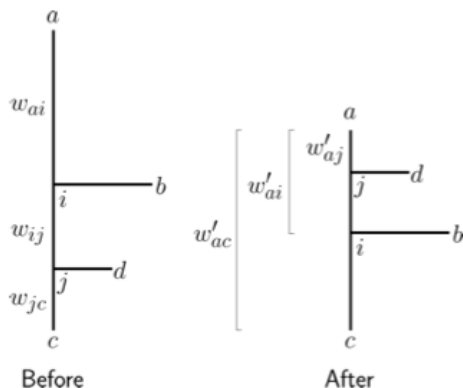


FIGURE 1. Illustration of the notation used to describe the LOCAL move. The panel labeled "Before" shows a tree before a LOCAL move. The panel labeled "After" shows one of the possible trees that can be produced by a LOCAL move if node j is moved. In this example, the topology of the tree has changed.

Coalescent prior

- Tree g has n leaves, age of i 'th node is t_i , number of lineages between $i - 1$ and i is k_i
- Mean time to coalescence is θ (=pop.size \times generation time)
- Probability density for g is

$$p(g|\theta) = \frac{1}{\theta^{n-1}} \prod_{i=2}^{2n-1} \exp\left(-\frac{k_i(k_i-1)}{2\theta}(t_i - t_{i-1})\right)$$

- See e.g. Drummond et al (2002)

Skyline Plots

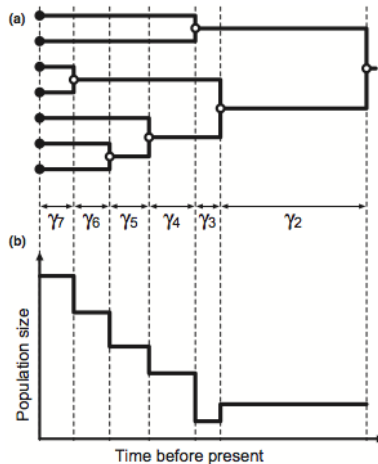
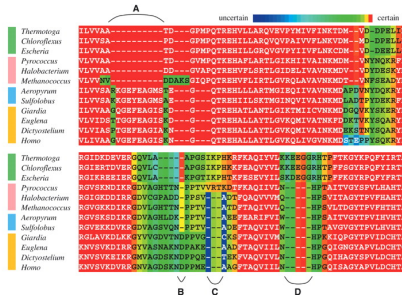
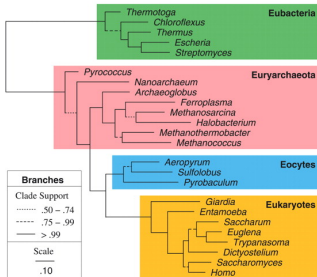


Figure: Ho and Shapiro (2011)

Extensions

- Co-sampling alignment and tree (and secondary/3D structure, ...)
- Models and approximations that sum out alignment, ancestral sequence, tree

Bali-Phy



Outline

- 1 Conjugate prior distributions
 - Gamma distribution
 - Dirichlet distribution
 - Normal-gamma distribution
 - Summary
- 2 MCMC in theory
 - Metropolis-Hastings
 - Gibbs sampling
 - Jacobians
 - MCMC on continuous parameters
- 3 MCMC in bioinformatics
 - Alignments
 - Trees
 - Structures

Molecular dynamics

- Folding
- Docking
- Transport
- Kinetics
- etc.

Protein, RNA, phospholipids, complexes, cells, etc etc.

RNA folding

- KinFold (Vienna): secondary structure
- KineFold (Isambert *et al*): pseudoknots, entropy
- SimulFold (Meyer *et al*): alignment, phylogeny, structure
- Multi-resolution Markov State Models (MSMs) (Huang *et al*): molecular dynamics + mesoscale Markov chain

MSMs

