Estimation for Restriction Sites Observed by Optical Mapping Using Reversible-Jump Markov Chain Monte Carlo

Jae K. Lee, Vlado Dančík, and Michael S. Waterman
Department of Mathematics, University of Southern California
1042 W. 36th Place, DRB 155, Los Angeles, CA 90089-1113

ABSTRACT

A fundamentally new molecular-biology technology in constructing restriction maps, *Optical Mapping*, has been developed by Schwartz et al. (1993). Using this method restriction maps are constructed by measuring the relevant fluorescence intensity and length measurements. However, it is difficult to directly estimate the restriction site locations of single DNA molecules based on these optical mapping data because of the precision of length measurements and the unknown number of true restriction sites in the data. We propose the use of a hierarchical Bayes model based on a mixture model with normals and random noise. In this model, we explicitly consider the missing observation structure of the data, such as the orientation of each molecule, the allocation of each cutting site to one of the normal distributions, and an indicator variable of whether an observed cut site is true or false. Because of the complexity of the model, the large number of missing data, and the unknown number of restriction sites, we use Reversible-Jump Markov chain Monte Carlo (MCMC) to estimate the number and the locations of the restriction sites. The study is highly computer-intensive and the development of an efficient algorithm is required.

1 Introduction

Restriction maps are one of the most fundamental data structures in molecular biology. However, the construction of a restriction map of a DNA molecule from fragment length data has proven difficult to automate. In addition to the time and expense required in running gel electrophoresis, the computational part of restriction mapping is not easy. Even ignoring length-dependent measurement errors, the double digestion problem is known to be NP-hard (Goldstein and Waterman, 1987) and there are often multiple solutions, only one of which is biologically relevant (Schmitt and Waterman, 1991). Recently an innovative new approach has been developed, *Optical Mapping*, that can produce ordered restriction maps using fluorescence microscopy (Schwartz et al., 1993). In this version restriction maps of individual molecules are constructed by measuring the relevant fluorescence intensity and length measurements.

First, restriction maps are constructed by imaging restriction endonuclease cutting events in single-stranded DNA molecules from yeast chromosomes with fluorescence microscopy. Cut sites appear as gaps that increase as the DNA fragments relaxed. Then for each molecule these gaps are rescaled within a unit interval (Figure 1). Thus, from this method data can be collected for the ordered restriction maps (fluorescence images) of thousand molecules in a few hours, up to 500 Kb in size currently with resolution about 200-250 basepairs (Schwartz et al., 1993). From these (orderd) cut sites we try to estimate the number and locations of restriction sites of a molecule. Notably, the measurement errors of these optical mapping data appear to be independent of length. Therefore, the advantage of this new technology is to eliminate the imprecision and expense in time and money of gel electrophoresis for determining the number and locations of restriction sites.

There are several complications in directly constructing a physical map from this kind of optical mapping data. For each molecule usually a few

RECOMB 98 New York NY USA
Copyright 1998 0-89791-976-9/98/ 3..$5.00

cut sites are observed from among all (unknown) restriction sites. Furthermore, there are false cut sites that appear with an unknown rate, and in some cases the orientation of each molecule from the fluorescence images is unknown. This has been studied with several different approaches, such as the pioneering Bayesian calculation of the model probability of the data by Anantharaman et al. (1997) and subsequent work by Dančík and Waterman (1997). However, the statistical problem of how to rigorously estimate the number of unknown restriction sites K has not been resolved. In addition to the difficulty of the unknown number of restriction sites, there is multimodality of the likelihood function due to unknown orientations of molecules. We here propose a full construction of a hierarchical Bayes model by explicitly defining the missing structure of the data. Because of the complexity and the presence of a large number of missing data, we use Markov chain Monte Carlo (MCMC) techniques to infer the parameters (e.g., Besag et al., 1995; Smith and Roberts, 1993; Thompson, 1995). First, at a fixed K we estimate separately the locations of the restriction sites, their relevant variances, and the constant rate of random noise. Then for an unknown K we use the *reversible-jump* MCMC technique suggested by Green (1995) and Richardson and Green (1997). Our model may appear a bit complex at the first glance. However, the hardest problem of applications of this technology is estimating the number of restriction sites. The complications of our model all arise from our desire to apply the powerful modern simulation methods such as MCMC to this central problem, since no analytic counterpart is tractable. Such methods invariably lead to making rigorous models for aspects of the problem that are often left implicit and unspecified such as the different likelihood spaces for differing numbers of sites where the numerical values of the likelihoods are not comparable.

2 The Model

Suppose that for molecule $i = 1, \dots, M$ we observe n_i cutting sites $x_{i,1}, \dots, x_{i,n_i}$, that are rescaled to lie within a unit interval. Notably, the orientation of these molecules are unknown; so the true location of each cut site is either $x_{i,j}$ or $1 - x_{i,j}$ depending on the orientation. For our statistical construction we consider a mixture model with normals and a random noise. Let an integer K be the number of the restriction sites. Cut sites corresponding to the k-th restriction site are observed as a normal distribution with mean θ_k and variance σ_k^2, $k = 1, \dots, K$. For $i = 1, \dots, M$ let the indicator variable w_i for the orientation of the molecules be distributed as

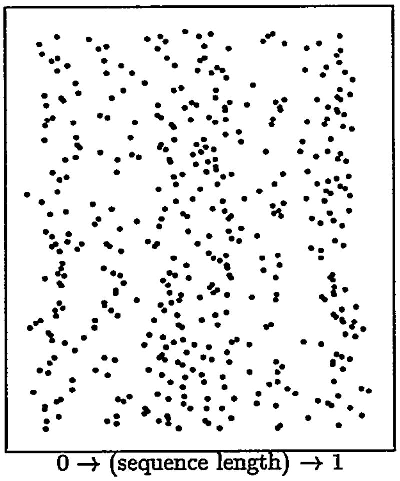

Figure 1: λ DNA, *BamH* I Enzyme Data. Each row represents a molecule cut by the enzyme.

$$w_i \sim \text{Bernoulli}(\frac{1}{2}),$$

with zero being for the current values $x_{i,j}$ and one for the reverse $(1-x_{i,j})$, all M independent of each other. Let $v_{i,j}$ be the index variable for false cuts

$$v_{i,j} \sim \text{Bernoulli}(1-p),$$

where p is the constant fraction of random noise, and $v_{i,j}$s are all independent of each other for $i = 1, \dots, M$ and $j = 1, \dots, n_i$. Then, given w_i and $v_{i,j}$, let the allocation variable $u_{i,j}$ have probabilities

$$\mathbb{P}r\{u_{i,j} = k\} = \delta_k, \ (k = 1, \dots, K, \delta_1 + \cdots + \delta_K = 1),$$

all independent of each other. Finally, conditional on w_i, $v_{i,j}$, $u_{i,j}$, $\sigma^2_{u_{i,j}}$, and $\theta_{u_{i,j}}$, let

$$y_{i,j} \sim \begin{cases} \text{Uniform}(0,1), & v_{i,j} = 0 \\ \text{Normal}(\theta_{u_{i,j}}, \sigma^2_{u_{i,j}}), & v_{i,j} = 1 \end{cases}$$

all independent of each other, where $y_{i,j}$, given w_i, are the (true) cut site after knowing the orientation

$$y_{i,j} = \begin{cases} x_{i,j}, & w_i = 0 \\ 1 - x_{i,j}, & w_i = 1. \end{cases}$$

We note that in our modeling only the $x_{i,j}$'s are observed and all $w_i, v_{i,j}, u_{i,j}$, $j = 1, \dots, n_i$, $i = 1, \dots, M$ are missing. So, the parameters of interest are

$\gamma = (\theta_1, \ldots, \theta_K, \delta_1, \ldots, \delta_K, \sigma_1^2, \ldots, \sigma_K^2, p)$ and the missing data are $m = (u_{i,j}, v_{i,j}, w_i)$ for $i = 1, \ldots, M$ and $j = 1, \ldots, n_i$. Thus, the joint probability of the observed (x) and missing data (m) is $\mathbb{P}r(x, m \mid \gamma)$

$$\begin{aligned}
&= \mathbb{P}r(x \mid m, \gamma)\mathbb{P}r(u)\mathbb{P}r(v)\mathbb{P}r(w) \\
&= \prod_{i=1}^{M}\prod_{j=1}^{n_i}\left\{[\frac{1}{\sqrt{2\pi}\sigma_{u_{i,j}}}\exp\{-\frac{(x_{i,j}-\theta_{u_{i,j}})^2}{2\sigma^2_{u_{i,j}}}\}]^{v_{i,j}}\right. \\
&\quad \left.\delta_{u_{i,j}}^{v_{i,j}}(1-p)^{v_{i,j}}p^{1-v_{i,j}}\right\}^{1-w_i} \times \\
&\quad \left\{[\frac{1}{\sqrt{2\pi}\sigma_{u^*_{i,j}}}\exp\{-\frac{(1-x_{i,j}-\theta_{u^*_{i,j}})^2}{2\sigma^2_{u^*_{i,j}}}\}]^{v^*_{i,j}}\right. \\
&\quad \left.\delta_{u^*_{i,j}}^{v^*_{i,j}}(1-p)^{v^*_{i,j}}p^{1-v^*_{i,j}}\right\}^{w_i}
\end{aligned}$$

where $v_{i,j}$ and $v^*_{i,j}$ are different depending on the orientation; so are $u_{i,j}$ and $u^*_{i,j}$.

Under the Bayesian paradigm we consider independent prior distributions on the parameters as follows: the order statistics of *iid* K Uniform(0,1) variables for $\theta = (\theta_1, \ldots, \theta_K)$, Beta$(\alpha, \beta)$ for p, $\Gamma(\xi, \nu)$ (Gamma) for σ_k^2, and Dirichlet$(\eta, \ldots, \eta)$ prior for $(\delta_1, \ldots, \delta_K)$. (For simplicity we use the same notation of the order statistics as those of the uniform variables θ_k, and unless explicitly mentioned, we assume the former from now on.) The estimated mixture is sensitive to the choice of the parameters in $\Gamma(\xi, \nu)$, so that we need to consider another hyper prior distribution for σ_k^2 (Richardson and Green, 1997). Fixing $\xi = 2$, we choose a $\Gamma(g, h)$ distribution for ν with $g = 0.2$ and $h = 40$ for lb.dat (119 molecules, 401 cut sites). Under several assumptions about our modeling on optical mapping, such as false cut rate (less than 15%), resolution of optical mapping ($\sigma_k < .05$), and proportion of cut sites allocated to each restriction site (max p_i ¡ 1.5 min p_i), some hyper prior parameter values are to be changed based on the size of each data set. For instance, if the number of cut sites is N, we use $\eta = N/10, \alpha = N/4, \beta = 2.25N$, and $g = N/10$.

Then the joint posterior distribution $\pi(\gamma, m \mid x)$ of the missing data and the parameters is proportional to

$$\begin{aligned}
&\mathbb{P}r(x, m \mid \gamma) \times \; K!\, p^{\alpha-1}(1-p)^{\beta-1} \times \\
&\prod_{k=1}^{K}\left(\delta_k^{\eta-1}\;\frac{\nu^{\xi}(\sigma_k^{-2})^{\xi-1}e^{-\sigma_k^{-2}\nu}}{\Gamma(\xi)}\right) \times \qquad (1) \\
&\frac{\Gamma(K\eta)}{\Gamma^K(\eta)}\;\frac{\nu^{g-1}h^g e^{-h\nu}}{\Gamma(g)}
\end{aligned}$$

3 Markov Chain Monte Carlo

3.1 Basic Updating

Since a direct inference on the model is rather difficult, we utilize Markov chain Monte Carlo (MCMC), a recent statistical simulation technique (Smith and Roberts, 1993; Thompson, 1994). This technique enables us to sample both the parameters and missing data directly from a complex (large-dimensional) likelihood function known up to constant. To do this we need to derive each full conditional distributions of the missing data and the parameters. First, given the orientation w_i, we jointly update $u_{i,j}$ and $v_{i,j}$. Note that the conditional distribution of $v_{i,j}$, given $u_{i,j}$, is completely deterministic. Given w_i and the parameters, the joint conditional distribution of $u_{i,j}$ and $v_{i,j}$ has probabilities:

$$\begin{aligned}
r_0 &= \mathbb{P}r[u_{i,j} = 0, v_{i,j} = 0 \mid \text{rest}] = c\, p, \\
r_k &= \mathbb{P}r[u_{i,j} = k, v_{i,j} = 1 \mid \text{rest}] \qquad (2) \\
&= c\,\frac{(1-p)\delta_k}{\sqrt{2\pi}\sigma_k}\exp\{-\frac{(y_{i,j}-\theta_k)^2}{2\sigma_k^2}\},
\end{aligned}$$

where $k = 1, \ldots, K$ and $r_0 + r_1 + \cdots + r_K = 1$ with some constant c. Since some false cuts may become true cuts as the orientation is flipped, and vice versa, we actually sample both $u^+_{i,j}$ (for $w_i = 0$) and $u^-_{i,j}$ (for $w_i = 1$); the values for $v^+_{i,j}$ and $v^-_{i,j}$ then follow from $u^+_{i,j}$ and $u^-_{i,j}$.

The conditional distribution of w_i, given the others, is proportional to

$$\prod_j(\frac{(1-p)\delta_{u_{i,j}}}{p\sqrt{2\pi}\sigma_{u_{i,j}}})^{v_{i,j}}\; e^{-\frac{1}{2\sigma^2_{u_{i,j}}}v_{i,j}(y_{i,j}-\theta_{u_{i,j}})^2}$$

Thus, w_i is updated from a Bernoulli trial with probability

$$\begin{aligned}
q_0 &= \mathbb{P}r[w_i = 0 \mid \text{rest}] \\
&= c\prod_j(\frac{(1-p)\delta_{u^+_{i,j}}}{p\sqrt{2\pi}\sigma_{u^+_{i,j}}})^{v^+_{i,j}} \\
&\quad \exp\{-\frac{1}{2\sigma^2_{u^+_{i,j}}}v^+_{i,j}(x_{i,j}-\theta_{u^+_{i,j}})^2\} \quad \text{and} \\
q_1 &= \mathbb{P}r[w_i = 1 \mid \text{rest}] \\
&= c\prod_j(\frac{(1-p)\delta_{u^-_{i,j}}}{p\sqrt{2\pi}\sigma_{u^-_{i,j}}})^{v^-_{i,j}} \\
&\quad \exp\{-\frac{1}{2\sigma^2_{u^-_{i,j}}}v^-_{i,j}(1-x_{i,j}-\theta_{u^-_{i,j}})^2\},
\end{aligned}$$

where $q_0 + q_1 = 1$ with some constant c.

After deciding the orientation we set $u_{i,j}$ (and $v_{i,j}$) to be either $u^{+}_{i,j}$ $(v^{+}_{i,j})$ or $u^{-}_{i,j}$ $(v^{-}_{i,j})$ depending on w_i. Also note that even though we describe the updating of u, v, and w separately, these are jointly sampled in our actual updating.

The conditional distribution of θ_k, $k = 1, \ldots, K$, is derived as a double-truncated normal, the exact form being found by *completing the square.* Let

$$\bar{x}_k = \frac{1}{M_k} \sum_{i=1}^{L} \sum_{j=1}^{n_i} y_{i,j} \ \mathrm{I}(u_{i,j} = k)$$

the mean of the points in the k-th component, where $M_k = \sum_i \sum_j \mathrm{I}(u_{i,j} = k)$ with $\mathrm{I}(u_{i,j} = k)$ being the index variable whether or not the (i,j)-th observation is classified into the k-th subpopulation. Then the conditional distribution of θ_k is proportional to

$$\exp\left\{ -\frac{(\theta_k - \bar{x}_k)^2}{2\sigma_k^2 / M_k} \right\}$$

if $M_k > 0$. From this we see that the θ_k's are conditionally independent given the missing data, and their distribution is a double-truncated normal subject to $\theta_{k-1} \leq \theta_k \leq \theta_{k+1}$, $k = 1, \ldots, K$ $(\theta_0 = 0, \ \ \theta_K = 1$ for notation), with mean $\bar{x}_k$ and variance σ_k^2 / M_k.

Due to conjugacy, the full conditional distribution of weight δ remains in the form of Dirichlet where

$$(\delta_1, \ldots, \delta_{K-1}) \sim \text{Dirichlet}(\eta + M_1, \ldots, \eta + M_K).$$

The update of p is simply derived as a Beta distribution with parameters $\acute{\alpha} = N - \sum_i \sum_j v_{i,j} + \alpha$ and $\acute{\beta} = \sum_i \sum_j v_{i,j} + \beta$.

The conditional distribution of σ_k^{-2} is proportional to

$$(\sigma^{-2})^{\frac{M_k}{2} + \xi - 1} \ \ e^{-\sigma_k^{-2} (\sum_{u_{i,j}=k} (y_{i,j} - \theta_k)^2 / 2 + \nu)},$$

that is a gamma distribution with parameters $(\frac{M_k}{2} + \xi)$ and $(\sum_{u_{i,j}=k} (y_{i,j} - \theta_k)^2 / 2 + \nu)$.
Finally, the hyperparameter ν is updated by $\Gamma(\xi + g, \sum_k \sigma_k^{-2} + h)$.

3.2 Reversible Jump for Number of Restriction Sites

Since the number of restriction sites is unknown, we need to devise an updating scheme that allows us to jump between two models with different numbers of restriction sites. This cannot be achieved by the standard MCMC approach in the previous section because we need to jump between two state spaces with different dimensionality, for which the existence of common dominating measure is not generally ensured. Green (1995) and Richardson and Green (1997) proposed the *reversible-jump* chain, a way to circumvent these difficulties by introducing some auxiliary independent variables for balancing the dimensionality. Suppose that a move based on transition q is proposed from $z = (\gamma, m)$ to a point $z' = (\gamma', m')$ for both parameters and missing data, with z' in a higher-dimensional space. The dimension-matching between them can be accomplished by drawing independent random vector t having the same degree of freedom as the difference of the dimensionality between the two state spaces of z and z'. Then, we effectively set an invertible deterministic relationship $z' = z'(z, t)$. Note that the reverse of the move (from z' to (z, t)) can be implemented by using the inverse transformation of the relationship. The acceptance probability of the move from z to z' is the Metropolis-Hastings ratio r (MH; Hastings, 1971) as

$$\min\{1, \ \frac{\pi(z'|x)}{\pi(z|x)} \frac{q(z', z)}{q(z, z')\, p(t)} \left| J_{z'(z,t)} \right| \}, \qquad (3)$$

where $q(\cdot,\cdot)$ is the probability of the transition q, $p(t)$ is the probability density function of t, and $J_{z'(z,t)}$ is the Jacobian of the transformation from (z, t) to z'. To make an efficient jump between the two different spaces, we need to devise a "good" transition q and invertible relationship $z'(z, t)$. Since the move between two spaces with a large difference of dimensionality is hard to achieve, we only consider the move adding or removing one component of the mixture (corresponding to one restriction site) at a time. We here propose a *split-combine* transition.

To increase K we split one of existing components into two, and to decrease K we combine two adjacent components into one. A transition can be constructed by the following two steps. First, decide to add one component with probability a_K, or remove one with probability $1 - a_K$. In our study we use $a_K = .5$ if $1 < K < K\text{max}$, $a_K = 0$ if $K = K\text{max}$, and $a_K = 1$ if $K = 1$, with a sufficiently large, pre-determined $K\text{max}$. Next, choose which site is to be split or which pair of adjacent sites is to be combined. Since the component having a larger number of cut sites allocated is more likely to be split into two (similarly, two adjacent components having fewer cut sites allocated are more appropriate candidates to be combined), we give different weights as $p'_k = \frac{m_k^2}{m_1^2 + \cdots + m_K^2}, k = 1, \ldots, K$ for *split* and $p_k = \frac{1/(m_k^2 + m_{k+1}^2)}{1/(m_1^2 + m_2^2) + \cdots + 1/(m_{K-1}^2 + m_K^2)}, k =$

$1, \dots, K-1$ for *combine*. These addition or deletion schemes will be effective under the assumption that there are about the same number of cut sites observed at each true restriction site. (However, it has been observed from experimental data that if two restriction sites are very close, the sum of the numbers of cut sites observed in the two sites is slightly less than twice of the average number of cut sites of the other restriction sites.)

For *combine*, after choosing two adjacent components to combine according to probability p_k, we match the 0th (weights), 1st, and 2nd moments of the new component to those of the two current ones chosen to combine. We merge all cut sites of the two selected components into the new one and calculate the parameters for the new combined component as

$$\begin{aligned} \delta_{k^\bullet} &= \delta_k + \delta_{k+1} \\ \delta_{k^\bullet}\theta_{k^\bullet} &= \delta_k\theta_k + \delta_{k+1}\theta_{k+1} \\ \delta_{k^\bullet}(\theta^2_{k^\bullet} + \sigma^2_{k^\bullet}) &= \delta_k(\theta^2_k + \sigma^2_k) + \\ &\quad \delta_{k+1}(\theta^2_{k+1} + \sigma^2_{k+1}) \end{aligned} \tag{4}$$

For *split*, in addition to choosing a component to split with probability p'_k, we need to introduce three independent random variables t_1, t_2, and t_3 to match the dimensions of the two spaces; these are all generated from Beta(2,2) in our study. Then, using these generated values, we construct an invertible function $z' = z'(\mu_{k^\bullet}, \sigma_{k^\bullet}, \delta_{k*}, t_1, t_2, t_3)$ satisfying the relationship in (4). We, for example, set:

$$\begin{aligned} &\delta_k = t_1\delta_{k^\bullet} \\ &\delta_{k+1} = (1-t_1)\delta_{k^\bullet} \\ &\theta_k = \theta_{k*} - \sigma_{k^\bullet} t_2\sqrt{t_1/(1-t_1)} \\ &\theta_{k+1} = \theta_{k^\bullet} + \sigma_{k^\bullet} t_2\sqrt{(1-t_1)/t_1} \\ &\sigma^2_k = \frac{t_3(1-t^2_2)}{t_1}\sigma^2_{k^\bullet} \\ &\sigma^2_{k+1} = \frac{(1-t_3)(1-t^2_2)}{1-t_1}\sigma^2_{k^\bullet}. \end{aligned} \tag{5}$$

We now need to reallocate the cut sites belonging to the k^*th component into the new k-th and k+1-th components analogous to (2). That is, we assign each cut site into one of the two components with probabilities r_k and r_{k+1} as in (2) subject to $r_k + r_{k+1} = 1$. Then the MH ratio in (3) for the *split* move reduces to

$$\begin{aligned} r &= \min\{1, \frac{\pi(z'|x)}{\pi(z|x)} \frac{(1-a_{K+1})\, p_k}{a_K\, p'_k\, p(t)\, P_{alloc}} \times \\ &\qquad \left|J_{z'(\delta_k,\mu_k,\sigma^2_k,t)}\right|\}, \end{aligned} \tag{6}$$

where $\frac{\pi(z'|x)}{\pi(z|x)}$ is the posterior probability ratio of (2) for the new point z' against the old point z, $p(t) = p(t_1, t_2, t_3)$ is the product of three independent density functions of Beta(2,2), P_{alloc} is the probability that this particular allocation is made, and $\left|J_{z'(\delta_k,\mu_k,\sigma^2_k,t)}\right| = \frac{\delta_k\sigma^3_k(1-t^2_2)}{(t_1(1-t_1))^{3/2}}$. The MH ratio for the *combine* move can be calculated as the reciprocal of (6) with some obvious substitutions. In this case $\delta_{k^\bullet}, \mu_{k^\bullet}, \sigma^2_{k^\bullet}, t_1, t_2$ and t_3 should be back-calculated from (4) and (5).

4 Results

For our example we have used a data set from Fig.1, which has five (true) restriction sites. In this case we have implemented our MCMC algorithm with two different starting points of the number of restriction sites—two and eight. The case started with two restriction sites leads us quickly to the true number of restriction sites (five), and the mean estimates of our MCMC sample precisely captured their true locations. Unfortunately, starting with eight restriction sites, we ended at six restriction sites, creating a false restriction site between the true first and second ones; however, the other five restriction sites were correctly estimated (Table 1). We believe the reason why we have more restriction sites than the true map (when started with eight) is that since the orientations of molecules are initially unknown, we created a false restriction site where the gap between two adjacent (true) restriction sites is large. This may have occurred by clustering single-cut sites and noise in the early stage of our MCMC run, and the algorithm may not be able to remove this in later iterations. Therefore, to overcome such multimodality, we believe it is better to start with a small number of restriction sites. In Table 1 we can see that our MCMC run has stayed at the true number of restriction sites (five) for most of the time (frequency 99.71%). Their true locations were also accurately captured by their MCMC mean estimates from the MCMC sample with five restriction sites.

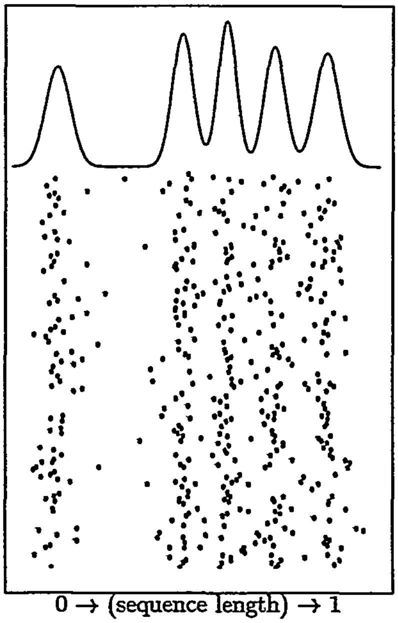

Figure 2: MCMC estimates of restriction sites for λ DNA Data.

```
    noise  r1   r2   r3   r4   r5
  2 (freq  0.12%)
w   : .084 .228 .687
mean:      .126 .604
  3 (freq  0.15%)
w   : .082 .194 .530 .192
mean:      .133 .497 .869
  4 (freq  0.02%)
w   : .087 .185 .372 .160 .194
mean:      .133 .498 .715 .862
  5 (freq 99.71%)
w   : .100 .168 .188 .185 .169 .187
mean:      .123 .462 .584 .713 .856
```

Table 1. Mean estimates and frequencies of the MCMC sample for λ DNA Data. The MCMC run started from two restriction sites.

Acknowledgments

We thank Dr. David Schwartz for introducing us into the problem and for generously providing the bacteriophage *lambda* data set in Section 4. This work was supported by NIH grant GM36230 and NFS grant BIR 95-04393.

References

Anantharaman, T. S., Mishra, B., and Schwartz D. C. (1997). "Genomics via Optical Mapping II: Ordered Restriction Maps." *Journal of Computational Biology* 4, 91-118.

Besag, J., Green, P. J., Higdon, D., and Mengersen, K. (1995), "Bayesian computation and stochastic systems with comments," *Statistical Science*, 10, 3-66.

Dančík, V. and Waterman, M. S. (1997). "Simple maximum likelihood methods for the optical mapping problem," *to appear in the Proceedings of the Workshop on Genome Informatics (GIW '97).*

Goldstein, L. and Waterman, M. S. (1987). "Mapping DNA by stochastic relaxation," *Adv. Appl. Math.*, 8, 194-207.

Green, P. J. (1995). "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711-32.

Richardson, S. and Green, P. (1997). "On Bayesian analysis of mixtures with an unknown number of components" (with discussion), *Journal of the Royal Statistical Society: Series B*, 59, 4, 731-792.

Schmitt, W. R. and Waterman, M. S. (1991). "Multiple solutions of DNA restriction mapping problems," *Adv. Appl. Math.*, 12, 412-427.

Schwartz, D. C., Li, X., Hernandez, L. I., Ramnarain, S. P., Huff, E. J., and Wang, Y.-K. (1993). "Ordered restriction maps of *Saccharomyces cerevisiae* chromosomes constructed by optical mapping," *Science*, Vol. 262, 110-114.

Smith, A. M. F. and Roberts, G. O. (1993), "Bayesian computation via Gibbs sampler and related Markov chain Monte Carlo methods," *J. Roy. Statist. Soc. B*, 55, 3-23.

Thompson, E. A. (1994), "Monte Carlo likelihood in genetic mapping," *Statistical Science*, 9-4, 355-366.