Using informative Multinomial-Dirichlet prior and reversible jump estimation of nucleosome occupancy for genome-wide profiling

Khader Khadraoui∗, Rawane Samb†, M’Hamed Lajmi Lakhal Chaieb‡, Arnaud Droit§

Abstract

In this paper we approach the problem of nucleosome positioning/occupancy estimation for genome-wide profiling from a Bayesian perspective. The proposed strategy is based on a Multinomial-Dirichlet informative prior and a hierarchical t-mixture model. The number and the positions of nucleosomes are estimated using a reversible jump Markov chain Monte Carlo simulation technique. Our approach enables us to obtain specific nucleosome configurations for the complex positioning patterns from experimental data and to identify each nucleosome on a given region bound from the aligned reads. The convergence in total variation of our simulation scheme is proved. The Bayesian approach not only produces point estimators that outperform existing approaches, but also offers a wealth of information about the parameters of interest. An application to a two real data sets using MNase-Seq and sonicated ChIP-Seq is provided. Numerical experiment obtained show that our Bayesian method estimated binding sites more consistent, even for a high number of nucleosomes, than the alternative method PING. By using sonicated CHIP-Seq datasets, we confirm that our method compares favorably to PING method and is robust to the dimension of regions binding event.

Keywords: Bayesian t-mixture; Multinomial-Dirichlet prior; Nucleosomes positioning; Genome-wide profiling; MNase-Seq; ChIP-Seq; Reversible-jump MCMC; Convergence in total variation.

1 Introduction

Global gene expression patterns are established and maintained by the concerted actions of Transcription Factors (TFs) and the proteins that constitute chromatin. The key structural element of chromatin is the nucleosome, which consists of an octameric histone core wrapped by 146 bps of DNA and connected to its neighbor by approximately 10-80 pbs of linker DNA [Luger (2006)]. Nucleosomes are a physical barrier to transcription that block access of activators and TFs to their sites on DNA, thereby inhibiting the elongation of transcripts by engaged polymerases. The packaging of DNA into nucleosomes appears to affect all stages of transcription, thereby regulating gene expression. In particular, the precise position of nucleosomes around the transcriptional start sites (TSSs) has an important influence on the initiation of transcription. In the past decade, the main strategy for genome-wide mapping of chromatin modifications, histone marks and interactions between DNA and proteins, has been ChIP followed by micro-array analysis (ChIP-chip) [Mendenhall & Bernstein (2008)]. Recent improvements in the efficiency, quality, and cost of genome-wide sequencing prompted biologists to abandon micro-arrays in favor of next-generation sequencing, a method referred to as ChIP-Seq [Pepke et al. (2009)]. Other experimental protocols can be used to generate profiles for nucleosome positioning. For instance, in an MNase ChIP-Seq experiment, the sonication step is replaced by MNase digestion, an endo-exonuclease that digests the linkers between adjacent nucleosomes. MNase-seq experiments are used when we are interested in
positioning nucleosomes over the whole genome, and not for regions related to certain types of histones or histone modifications. For nucleosome positioning mapping, although the ChIP-Seq experimental protocols are generally similar to those for transcription factors, histones are typically more abundant than any TF, resulting in far more ChIP enrichment events in a single histone modification experiment compared to a TF experiment. In addition, nucleosomes tend to occur in linear arrays. For these reasons, nucleosome positioning typically requires more complex models than transcription factor profiling. Moreover, the development of methods to analyze histone modifications based on nucleosome positioning and their variants, as well as other epigenomic data to enhance our understanding of a given cellular process may not be optimal and require more elaborated approaches (Ganguli et al., 2012).

The literature on nucleosome positioning usually focuses on frequentist inference within parametric approaches; see for instance: Kuan et al. (2009), Chen et al. (2010), Xi et al. (2010), and Mitra & Gupta (2011). In these works, the authors address the detection of nucleosome positions thanks to the hidden Markov models with a known order by assumption. Recently, Polishko et al. (2012) proposed another method for the inference of nucleosome positions. Their strategy is based on a modified Gaussian mixture model where the number of clusters is unknown and estimated by an effective heuristic. In addition, and in the same spirit, Zhang et al. (2011) and Zhang et al. (2012) proposed to use the BIC criterion (Schwarz, 1978) for the estimation of the number of components in the mixture model. In particular, they estimate the parameters of the model before selecting the number of components using a penalized log likelihood. Schopflin et al. (2013) used a simulated annealing scheme, see for instance Bartoli & Del Moral (2001), to find an optimized placement of nucleosomes. The principal inconvenient of their optimization approach is that incapable to give a computational credible sets for optimized parameters. All of the cited approaches (hidden Markov models, template filtering, mixture models, peak calling on smoothed coverage data and sliding window statistics) need an elaborate specific parameterization (that cannot work in general case) or not enable to decide between overlapping peaks. Bayesian approaches on nucleosome positioning have seldom been considered though it seems of practical interest and very coherent with the raised problem. Hence, the estimation of nucleosomes positions is not well understood and remains a problem that deserves further investigation although some works appear in the literature to address this question.

The aim of our work is to derive a fully Bayesian hierarchical model for genome-wide profiling of nucleosome positions based on high-throughput short-read data (MNase or Sonicated short-read data). The coherence of the Bayesian paradigm with inference on nucleosome positioning/occupancy estimation for genome-wide profiling is argued. The contributions of this paper are fourfold. First, to propose a Bayesian inference that jointly models local concentrations of directional reads and satisfy our knowledge about genome regions thanks to the prior distribution. Second, to use a Multinomial-Dirichlet model in the construction of an informative prior distribution coupled to a t-mixture model. The choice of student mixture as that represents a flexible approach for fitting asymmetric distributions with heavy tails. Third, the number of nucleosomes will be considered as a random variable and refers to a prior distribution where, to our knowledge, this aspect has not been yet addressed and the set of unknown parameters are simultaneously estimated by posterior mean. The posterior mean and their 95% credible sets are computed using simulation from the posterior distribution. The problem of routine simulation from posterior distribution for all parameters is addressed using the Markov chain Monte Carlo simulation technique of reversible jump (Green, 1995). Fourth, to study the asymptotic behavior of our simulation scheme from the posterior distribution. Our study is based on tools for convergence analysis of Metropolis-Hastings kernel for general state space, e.g. (Tierney, 1998), i.e. properties of interest are reversibility, irreductibility and aperiodicity. Most technical details differ however, as we use a truncated prior distribution and it is necessarily to specify the normalizing constant.

The rest of the paper is organized as follows. Section 2 presents the nucleosomes determination problem and the Bayesian inference as a preliminary to the development of the Multinomial-Dirichlet informative prior. Section 3 describes the MCMC reversible jump scheme for simulations from the posterior and the convergence of the
trans-dimensional MCMC scheme considered. Section 4 illustrates the performances of the proposed method via simulation studies and an application to the MNase-Seq Kaplan data from budding yeast (Kaplan et al., 2009) and sonicated H3K4me1 Chip-Seq data in mouse puer cell line (Heinz et al., 2010). Finally, Section 5 draws some conclusive considerations and discusses extensions of the model considered and directions of future research. Detailed proofs are deferred to Appendix A.

2 Bayesian t-mixture model for nucleosomes positioning

The basic mixture model, described in Section 2.1, utilizes Student densities formulations for both the forward and reverse read position for genome candidate regions, along with prior specifications that may be used to encourage using of prior informative knowledge. The resulting model, which retain the form of a mixture density estimation model, permits (but does not force) shrinkage towards a specific parametric mixture model as well as allows one to incorporate specific prior knowledge. We describe our method for adaptive selection of the number and placement of nucleosomes in Section 3. We assume that the observed data structure \( y = (y_1, y_2, \ldots, y_n) \) be the aligned reads positions generated by high-throughput short-read sequencing of DNA obtained by MNase procedures or by Sonication protocols. Let \( d_i \in \{+1, -1\} \) denote the direction of the \( i \)-th short-read fragment: +1 for the forward read position, and -1 for the negative read position. In the sequel we assume that, as in Zhang et al. (2012), the read-data are first pre-processed by segmenting the genome into candidate regions, each of which has a minimum number of reads that aligned to forward and reverse strands.

2.1 Basic model structure

Under the indicated assumptions and conditionally upon the number of nucleosomes \( k \) in each region, we model all the aligned read positions as independent and identically distributed with some density thanks to the usual basic Student’s-t mixture model:

\[
y_i \sim \sum_{j=1}^{k} w_j t_4 \left( y_i, \mu_j - \frac{d_i \delta_j}{2}, \sigma_{d_i,j} \right),
\]

where \( t_4(\cdot, \mu, \sigma) \) is the density function of the Student’s t-distribution with four degree of freedom (to minimize computation), mean \( \mu \) and scale \( \sigma \). In the model (2.1), weight associated with each component is the same for both forward and reverse distributions \( (w_j \in [0, 1] \text{ such that } \sum_{j=1}^{k} w_j = 1) \). For the \( j \)-th nucleosome, \( \mu_j \) denotes the position of its center, while \( \delta_j \) represents the distance between the maxima of the forward and reverse read position densities. The parameters \( \sigma_{d_i,j}^2 \) evaluate the variability in DNA fragment end positions of the strands linked to the \( j \)-th nucleosome. In the sequel, we use the notations \( \mu_{fj} = \mu_j - \delta_j/2 \) and \( \sigma_{fj} = \sigma_{d_i,j} \) for forward read \( (d_i = +1) \). We use \( \mu_{rj} = \mu_j + \delta_j/2 \) and \( \sigma_{rj} = \sigma_{d_i,j} \) for reverse read \( (d_i = -1) \). We denote by \( \theta = (\theta_1, \ldots, \theta_k)^T \), where \( \theta_j = (\mu_j, \delta_j, \sigma_{fj}, \sigma_{rj})^T \), the parameters of our modeling. The set of parameters derived from the t-mixture model is illustrated in Figure 1. In a frequentist approach to parameter estimation, one would now maximize the marginal likelihood function that is obtained from (2.1). In practice, the EM algorithm or some related variant is typically used for this purpose, e.g. Zhang et al. (2011) and Zhang et al. (2012).

![Figure 1: Diagram displaying one binding event in a DNA candidate region. Forward and reverse strand aligned reads are shown by thick and thin arrowheads, respectively.](image-url)
2.2 Multinomial-Dirichlet model and prior specifications

As indicated earlier, we intend to use a fully Bayesian approach where the parameters are assumed to be drawn from convenient prior distributions. Initially, let \( z_i \in \{1, \ldots, k\} \) be the hidden labels that indicate the observation of \( y_i \) arising from the \( j \)-th component and let us put \( z = (z_1, \ldots, z_n)^T \) and \( w = (w_1, \ldots, w_k)^T \). In a similar way as in [Richardson & Green (1997)], we consider that joint distribution of all variables mentioned above have the representation:

\[
\pi(k, w, z, \theta_k) = \pi(k)\pi(w|k)\pi(z|w, k)\pi(\theta_k|z, w, k),
\]

(2.2)

where \( \pi \) is a generic notation for some density with respect to the reference measurement. Furthermore, we consider that the conditional independence assumption \( \pi(\theta_k|z, w, k) = \pi(\theta_k|k) \) is imposed. Consequently, the joint prior distribution (2.2) simplifies and yields the Bayesian hierarchical model

\[
\pi(k, w, z, \theta_k) = \pi(k)\pi(w|k)\pi(z|w, k)\pi(\theta_k|k).
\]

(2.3)

Now, concerning the prior distribution of \( k \), a common choice is the Poisson distribution with fixed hyperparameter \( \lambda \). In this paper, we consider a truncated Poisson distribution: \( \pi(k) \propto \exp(-\lambda)\frac{\lambda^k}{k!}1_{\{1, \ldots, k_{\text{max}}\}} \) for a prespecified integer \( k_{\text{max}} \) and where \( 1_A \) denotes the indicator function of the set \( A \). The choice of \( \lambda \) is discussed when we come to our numerical experiments. The Poisson prior is strongly informative, and allows great control over the dimension of the resulting mixture, at the risk of overfitting. In contrast, less informative priors such as a Geometric, Negative Binomial or an uniform distribution \( U_{\{1, \ldots, k_{\text{max}}\}} \) can be used to penalize large number of components. For \( d \in \{f, r\} \), we put \( \mu_d = (\mu_{d,1}, \ldots, \mu_{d,k})^T \) and we consider the set \( \mathcal{E}_{d,k} = \{\mu_d \in \mathbb{R}^k, \mu_{d,1} \leq \cdots \leq \mu_{d,k}\} \). Prior distributions, with respect to Lebesgue measure, for the mixtures Student components are specified as follows:

\[
\begin{align*}
(\mu_d|k) &\sim N_k^{E_d}(\xi_d, \tau_d^{-1}V_d); \\
\sigma_{d,j} &\sim IG(\tau_1, \tau_2); \\
(\delta_j|\sigma_{f,j}^2, \sigma_{r,j}^2) &\sim N\left(\xi, 1/(\sigma_{f,j}^2 + \sigma_{r,j}^2)\right),
\end{align*}
\]

(2.4)

where \( N_k^{E_d}(\cdot, \cdot) \) is a multivariate normal distribution truncated to the set \( \mathcal{E}_{d,k}, \xi_d = (\xi_{d,1}, \ldots, \xi_{d,k}) \) is the prior expectation of \( \mu_d \) and \( V_d \) is a \( k \times k \) covariance matrix which will be defined in the following Proposition 1. For the bandwidth of components, \( \tau_1 \) and \( \tau_2 \) are a fixed hyperparameters of the inverse gamma prior. These prior distributions may be quite diffuse. The prior distributions of \( k \) and \( \mu_d|k \) corresponding to specific choices of the truncated spaces may be improper. Improper priors, while not uniformly accepted among Bayesian, are commonly used in statistics. In our setting, the benefits of using either the priors \( \pi(k) \) and \( \pi(\mu_d|k) \) stem mainly from the ability to incorporate related qualitative criteria in a natural way. We precise that, in (2.4), \( \xi \) can be chosen empirically according an exploratory analysis of data. The prior expectation \( \xi_d \) for \( \mu_d \) is taken to be rather flat over the corresponding observed range of the data. Note that the quantities \( \mu_d \) and \( \delta_j \) could depend on data. Although not purely Bayesian, data dependent priors are quite common in the literature. We complete the construction of our prior by considering a Bayesian approach for determining the mixing weights \( w \) and the missing data for components. This approach is based on a Multinomial-Dirichlet model. To simplify the modeling, we use \( \mu_d \) in order to calculate

\[
n_{d,k} = \sum_{i=1}^{n_d} 1\{a_{d,i} \leq y_{d,i} \leq a_{d,k+1}\} \quad \text{and} \quad n_{d,j} = \sum_{i=1}^{n_d} 1\{a_{d,i} \leq y_{d,i} \leq a_{d,j+1}\}, \quad j = 1, \ldots, k - 1,
\]

(2.5)

where \( y_{d,i} \) is the value of \( y_i \) according to the value of \( d \), \( n_d = \#\{y_i : d_i = d\} \) and \( (a_{d,1}, \ldots, a_{d,k+1}) \) is a random partition of \( y_d = \{y_{d,1}\}_{i=1}^{n_d} \) such that \( a_{d,1} = \min(y_d) \leq \mu_{d,1} < a_{d,2} < \cdots < \mu_{d,k} < a_{d,k+1} = \max(y_d) \). It is clear that \( \sum_{j=1}^{k} n_{d,j} = n_d \) and \( n_{d,j}, n_{d,k} \) represent the number of observations in the intervals \( [a_{d,j}, a_{d,j+1}) \), \( [a_{d,k}, a_{d,k+1}) \]. This idea can be seen as a classification of \( \{y_{d,i}\}_{i=1}^{n_d} \) where each class size is \( n_{d,1}, \ldots, n_{d,k} \). The above reformulation of the model helps in understanding how the data update the prior. This means that the
positions of nucleosomes are somewhere informative for the classification of data and the size of each class. We put \( \mathbf{n}_d = (n_{d,1}, \ldots, n_{d,k})^\top \) and \( w = (w_1, \ldots, w_k)^\top \) become the distribution of assigning an observation \( y_{d,i} \) to a some class such that \( \mathbb{P}(y_{d,i} \in [a_{d,j}, a_{d,j+1}]) = w_j, \ j = 1, \ldots, k \). Thus, for all \( j, w_j \geq 0 \) and \( \sum_{j=1}^k w_j = 1 \). We can now consider a multinomial prior on \( \mathbf{n}_d \ (\mathbf{n}_d \sim \text{Mult}(k, w)) \):

\[
\pi(\mathbf{n}_d|k, w) = \binom{n_d}{n_{d,1} \cdots n_{d,k}} \prod_{j=1}^k w_j^{n_{d,j}} = \frac{n_d!}{n_{d,1}! \cdots n_{d,k}!} \prod_{j=1}^k w_j^{n_{d,j}},
\]

and for \( w \) we consider a Dirichlet prior on the \((k-1)\)-dimensional positive simplex

\[
D_{k-1} = \left\{ (w_1, \ldots, w_{k-1}) \in \mathbb{R}^{k-1} : \forall j \in \{1, \ldots, k-1\}, \ w_j \geq 0 \text{ and } \sum_{j=1}^{k-1} w_j < 1 \right\},
\]

with parameter \( \alpha_d = (\alpha_{d,1}, \ldots, \alpha_{d,k})^\top \) where, \( \forall j = \{1, \ldots, k\}, \alpha_{d,j} > 0 \). The density of the prior of \( w|k \) with respect to the Lebesgue measure on \( \mathbb{R}^{k-1} \) is given by:

\[
\pi(w|k) = \frac{\Gamma(\sum_{j=1}^k \alpha_{d,j})}{\prod_{j=1}^k \Gamma(\alpha_{d,j})} \prod_{j=1}^{k-1} w_j^{\alpha_{d,j} - 1} \left( 1 - \sum_{j=1}^{k-1} w_j \right)^{\alpha_{d,k} - 1}.
\]

We denote by \( \mathcal{D}(\alpha_d, \ldots, \alpha_{d,k}) \) the above Dirichlet distribution. For the choice of \( \alpha_d \), we discuss its effect in the following. Let \( Y = (Y_1, \ldots, Y_k) \sim \mathcal{D}(\alpha_{d,1}, \ldots, \alpha_{d,k}) \). We know that, by putting \( \alpha_0 = \sum_{j=1}^k \alpha_{d,j}, \ E[Y_j] = \alpha_{d,j}/\alpha_0 \) and \( \forall j = \{1, \ldots, k\} = \alpha_{d,j}(\alpha_0 - \alpha_{d,j})/\alpha_0^2(\alpha_0 + 1) \). Then, in the case that \( \alpha_{d,j} \) are small, for all \( j = 1, \ldots, k \), it is easy to remark that the dispersion of \( Y_i \) is high. When \( \alpha_{d,j} \) are great and close, for all \( j = 1, \ldots, k \), we can remark that the variance of \( Y_i \) is small. To understand the effect of \( \alpha_d \) on nucleosomes positioning and weights mixing, we compare in simulations two choices for \( \alpha_{d,j} \): in one are small \( (\alpha_{d,j} = 1, j = 1, \ldots, k) \) and in the other are great \( (\alpha_{d,j} = n_{d,j} + 1, j = 1, \ldots, k) \). In the case of \( \alpha_{d,j} = n_{d,j} + 1 \), it is straightforward that

\[
\frac{\Gamma(\sum_{j=1}^k \alpha_{d,j})}{\prod_{j=1}^k \Gamma(\alpha_{d,j})} \prod_{j=1}^{k-1} w_j^{\alpha_{d,j} - 1} = \frac{\Gamma(n_d + k)}{n_d!} \left( n_{d,1} \cdots n_{d,k} \right) \prod_{j=1}^k w_j^{n_{d,j}}.
\]

Considering the Multinomial-Dirichlet model described above and given by:

\[
\begin{cases}
( \mathbf{n}_d|k, w) \sim \text{Mult}(k, w), \\
( w|k) \sim \mathcal{D}(\alpha_{d,1}, \ldots, \alpha_{d,k}),
\end{cases}
\tag{2.6}
\]

we obtain immediately the conditional distribution of \( (w|k, \mathbf{n}_d) \) thanks to the conjugacy of the model \( (2.6) \). This conditional distribution is a Dirichlet with an update parameters \( \alpha_d' = \alpha_d + \mathbf{n}_d \) and with density proportional to

\[
\pi(w|k, \mathbf{n}_d, \cdots) \propto \pi(\mathbf{n}_d|k, w) \pi(w|k) \propto \prod_{j=1}^{k-1} w_j^{n_{d,j} + \alpha_{d,j} - 1} \left( 1 - \sum_{j=1}^{k-1} w_j \right)^{\alpha_{d,k} + 1},
\tag{2.7}
\]

where here and later we use "\( \cdot \)" to denote conditioning on all other variables. We denote by \( \mathcal{D}(\alpha_{d,1} + n_{d,1}, \ldots, \alpha_{d,k} + n_{d,k}) \) the Dirichlet conditional distribution \( (2.7) \). It is clear that we can make a sampling from this full conditional and updated \( w \) thanks to a Gibbs move. Thus, by defining \( n_{d,j} = \# \{ i : d_{i,j} = j \} \), we relate \( z_d = (z_{d,1}, \ldots, z_{d,n_d})^\top \) to \( \mathbf{n}_d \) and there is no need to infer the variable \( z_d \) because the useful information about positions and size of data corresponding to nucleosomes is contained in the variable \( \mathbf{n}_d \).
2.3 Consistency

We briefly discuss posterior consistency of our model, noting that they have been extensively examined by Van
deer Vaart and co-authors [see in particular, Ghosal & deer Vaart (2007) and Kruijer et al. (2010)]. Despite that
much progress has been made regarding the computational problems in nonparametric Bayesian inference [see for
instance the review by Marin et al. (2005)], results on convergence rates were found only recently. Kruijer et al.
(2010) consider location-scale mixtures to considerably generalize results on posterior convergence using this type
of mixtures by finding conditions under which a probability density of any Hölder-smoothness can be efficiently
approximated by a location-scale mixture.

Kruijer et al. (2010) showed that for the location-scale mixture models, the conditions of the general consistency
theory are satisfied for a family of priors that is usually used in finite mixture models. In particular, the bandwidth
prior can be any inverse-gamma distribution, whose support neither has to be bounded away from zero, nor to
depend on the sample size. The prior on \( k \) can be any distribution satisfying

\[
B_0 \exp(-b_0 k (\log k)^{r_0}) \leq \pi(k) \leq B_1 \exp(-b_1 k (\log k)^{r_0}),
\]

for some constants \( 0 < B_0 \leq B_1, 0 < b_1 \leq b_0 \) and \( r_0 \geq 0 \). It is clear that the prior on \( k \) can be any Poisson
distribution \((r_0 = 1)\) or geometric distribution \((r_0 = 0)\). Given \( k \), the locations \( \mu_d \) can be drawn from a prior
density that may not have polynomial tails, which would increase to much the entropy of the mixture model,
or super-exponential tails, which would decrease the approximative properties of the mixture model. Given \( k \),
the prior of the weight \( w \) is independent of \( \mu_d \) and can be, for some nonnegative constant \( b \), any distribution satisfying,

\[
\pi(w \in D_{k-1}(w^0, \epsilon)|k) \gtrsim \exp \left\{ d_1 k (\log k)^b \log \frac{1}{\epsilon} \right\},
\]

where \( D_{k-1}(w^0, \epsilon) = \{ \omega \in D_{k-1}; \sum_{i=1}^{k} |\omega_i - w^0_i| \leq \epsilon \} \), there is a constant \( d_1 \) such that (2.9) is obtained for \( \epsilon < \frac{1}{k} \)
and \( w^0 \in D_{k-1} \) is the true weight. The notation \( \gtrsim \) denotes inequality up to a multiplicative constant. Conditions
on prior of \( \mu_d \) and the prior \((2.9)\) of the weight \( w \) in fact require that there is a minimal amount of prior mass
around the true location \( \mu_d^0 \) and the true weight \( w^0 \). However, the prior distributions are well specified to satisfy
all these conditions, and thus posterior consistency holds. Therefore, it follows that the posterior of our proposed
Student’s-t mixture model is consistent.

3 Reversible jump scheme for nucleosomes determination

We describe in this section our reversible jump Metropolis-Hastings within Gibbs scheme for nucleosome position-
ing. From the mixture of Student’s-t observations given in (2.1), the likelihood of the data \( y_d = (y_{d,1}, \ldots, y_{d,n_d}) \)
is

\[
L(y_d|\theta_{d,k}, k, \cdots) = \prod_{i=1}^{n_d} \left[ \sum_{j=1}^{k} w_j f_d(y_{d,i}, \mu_{d,j}, \sigma_{d,j}) \right],
\]

and from (3.1) we obtain the joint posterior density (up to a constant) by incorporating the prior information
(2.3). This posterior density of all parameters is specifically proportional to

\[
\pi(\theta_{d,k}, k, \cdots | y_d) \propto L(y_d|\theta_{d,k}, k, \cdots) \pi(k) \pi(\theta_{d,k}|k) \pi(w|k) \pi(n_d | k, w).
\]

Sampling from the posterior distribution \( \pi(\theta_{d,k}, k, \cdots | y_d) \) can be done by several ways. Note that the full
conditional posterior distribution is truncated to \( \mathcal{E}_{d,k} \). The main difficulty in using a Gibbs sampler is to simulate
from the truncated multivariate normal distribution of \( \mu_d \). This can be avoided by sampling sequentially from
the full conditional distribution of each coordinate $\mu_{d,j}$, which requires to sample from a truncated univariate normal distribution. This can also involve numerical difficulties when the truncation is located in the tails of the distribution. Numerical problems and calculation of full conditional posterior distributions for all parameters can be avoided by using the following reversible jump Metropolis-Hastings within Gibbs algorithm. Because the prior distribution $\pi(\theta_d, n_d, k, w)$ is expressed as a trans-dimensional prior, implementation of the MCMC algorithm require exact knowledge of $\pi(\mu_d|k)$ (i.e. the normalizing constant of the truncated Gaussian density $N_{\xi_d,\tau_d^{-1}}(\mu_d)$).

**Proposition 1** Assume that $\mu_{d,j} \sim N(\xi_{d,j}, \tau_d^{-1})$ for all $j = 1, \ldots, k$. Then, the exact prior density restricted to \{\mu_{d,1} \leq \cdots \leq \mu_{d,k}\} of $\mu_d|k$ is given by

$$
\pi(\mu_d|k) = \left\{ \frac{\sqrt{\pi}}{2\tau_d} \right\}^{-k} \exp \left\{ -\tau_d \frac{(\mu_d - \xi_d)^T V^{-1}_d (\mu_d - \xi_d)}{2} \right\} 1_{\{\mu_{d,1} \leq \cdots \leq \mu_{d,k}\}},
$$

(3.3)

where the matrix $V^{-1}_d = (\omega_{i,j})_{1 \leq i,j \leq k}$ is a $k \times k$ matrix such that

$$
\omega_{i,j} = \begin{cases} 
2, & \text{if } i = j = 1, \ldots, k - 1; \\
1, & \text{if } i = j = k; \\
-1, & \text{if } j = i \pm 1, j = 1, \ldots, k; \\
0, & \text{otherwise.}
\end{cases}
$$

For a proof, see Appendix A.1.

The following show in details the Markov chain Monte Carlo scheme used to sample from the joint posterior distribution (3.2) consists of updating unknown parameters via reversible jump Metropolis-Hastings within Gibbs sampler. After initialization, posterior samples of all parameters can be drawn by the MCMC steps, repeated as long as needed to ensure convergence of the chain and a sufficient number of samples from the posterior.

### 3.1 Sampling from the posterior distribution

Specifically we describe in the sequel the structure of the algorithm where we briefly discuss initialization of the algorithm and the reversible jump Metropolis-Hastings update steps. First, let us express roughly our opinion of the unknown mixture of densities by drawing a histogram of data. Unfortunately, we find that good starting values hasten the convergence of the chain and reduce the risk of numerical problems of course even though the chain can in principle be initialized at any value. Once initial values have been obtained, the algorithm proceeds through successive proposal and acceptance/rejection steps in which each set of parameters is updated in turn. Throughout this steps and in order to enable adaptive component selection we not only allow components to move, but also permit changes in dimension, such as adding a component (birth submove) or deleting a component (death submove). For more details about reversible-jump MCMC subject to a dimension-matching we refer the reader to [Green (1995)](#). Typically, at each iteration we choose randomly whether to execute a birth, death, or adjustment submove. Let

$$
\Theta_d = \bigcup_{k \geq 1} \left\{ k \right\} \times \Theta_{d,k}
$$

where $\Theta_{d,k}$ is a subspace of the Euclidean space $\mathcal{E}_{d,k} \times [0, \infty)^k \times \mathbb{R}^k$. We denote by $\theta_{d,k}$ a generic element of $\Theta_{d,k}$ such that $\theta_{d,k} = (\mu_d, \sigma_d, \delta)$. Suppose that for each $k \in \mathcal{K} \subseteq \mathbb{N}$, where $|\mathcal{K}| > 1$, the formulation of the MCMC simulation from the posterior leads to a involved trans-dimensional algorithm on $\Theta_{d,k} \times [0, 1]^k \times \{-1, +1\}^n \times \{1, \ldots, n\}^k \times \mathcal{K}$. Trans-dimensional chains can proceed in a variety of ways. On the one hand, if $k \geq 2$ we select a submove from
scheme: unoccupied candidate center of the new component is chosen to be added to the current set of centers \( \mu^k \). In the birth submove (B), \( q(n, \tilde{\theta}) \) is then accepted or rejected according to

\[
\rho_b = \min \left\{ 1, \frac{L(y_n | \tilde{\theta}_d, \tilde{n}_d, \tilde{k}, w)}{L(y_n | \tilde{\theta}_d, n_d, k, w)} \frac{\pi(\tilde{n}_d k) \pi(\tilde{n}_d k) \pi(n_d k) \pi(n_d k) q_l(\tilde{\mu}_{d,L})}{q_l(\mu_{d,L})} \right\},
\]

where \( q_l(\tilde{\mu}_{d,L}) = \frac{g(\tilde{n}_d \mid n_d)}{g(n_d \mid n_d)} \) is a ratio of Multinational densities kernels and \( q_l(\tilde{\mu}_{d,L}) \) is the uniform density of \( \tilde{\mu}_{d,L} \).

3.1.1 Updating by (B)_k submove

In the birth submove (B)_k, the algorithm attempts to add a component for a new nucleosome where a random unoccupied candidate center of the new component is chosen to be added to the current set of centers \( \mu_d \). To do this, we compute the candidate component parameters as follows: With probability \( d_k \), we can update \( (\theta_d, k, w, n_d) \) through the proposal \( (\tilde{\theta}_d, \tilde{k}, \tilde{w}, \tilde{n}_d) \), where \( \tilde{\mu}_d \in \mathcal{E}_{d,k+1} \), generated by a dimension-matching scheme:

- Draw \( \ell \sim \mathcal{U}(1, \ldots, k) \) and \( \delta_{d,\ell} \sim \pi(\delta_{d,\ell}) \) and \( \delta_{d,\ell} \sim \pi(\delta_{d,\ell}) \)
- Draw \( \tilde{\mu}_{d,\ell} \sim \mathcal{U}(\mathcal{E}_{d,k+1}(\tilde{\mu}_d, \ell) \cap |\mu_d, \ell \pm \epsilon|) \) where
  \[ \mathcal{E}_{d,k+1}(\tilde{\mu}_d, \ell) = \{ \tilde{\mu}_{d,\ell} : (\mu_{d,1}, \ldots, \mu_{d,\ell}, \ldots, \mu_{d,k+1})^\top \in \mathcal{E}_{d,k+1} \} \]
  and \( \epsilon \) is a constant that controls the variation of the proposal.
- Draw \( \tilde{n}_d \sim \mathcal{P}(\tilde{\mu}_d) \) where \( \mathcal{P}(\cdot) \) is a Multinational proposal using sizes of class defined in (2.5).
- Update \( (w | \tilde{k}, \tilde{n}_d, \cdots) \) from the full conditional posterior given by (2.7).
- This proposal is then accepted or rejected according to
  \[ \rho_d = \min \left\{ 1, \frac{L(y_n | \tilde{\theta}_d, \tilde{n}_d, \tilde{k}, w)}{L(y_n | \tilde{\theta}_d, n_d, k, w)} \frac{\pi(\tilde{n}_d k) \pi(\tilde{n}_d k) \pi(n_d k) \pi(n_d k) q_l(\tilde{\mu}_{d,L})}{q_l(\mu_{d,L})} \right\}, \]

3.1.2 Updating by (D)_k submove

If the death submove (D)_k is selected the algorithm randomly nominates a component for deletion uniformly from the existing components \( \{1, \ldots, k\} \). The component parameters are correspondingly adjusted by the inverse of the transformation in (B)_k submove. Hence, with probability \( d_k \), we can update \( (\theta_d, k, w, n_d) \) through the proposal \( (\tilde{\theta}_d, k, \tilde{w}, \tilde{n}_d) \) where \( \tilde{\mu}_d \in \mathcal{E}_{d,k-1} \). The proposal scheme is as follow:

- Draw \( \ell \sim \mathcal{U}(1, \ldots, k) \), remove \( \mu_{d,\ell} \), \( \sigma_{d,\ell} \) and \( \delta_{d,\ell} \) from \( \mu_d, \sigma_d^2 \) and \( \delta \) then relabel \( \tilde{\theta}_d, k, (\tilde{\mu}_{d,k-1})^\top \).
- Draw \( \tilde{n}_d \sim \mathcal{P}(\tilde{\mu}_d) \) similarly in (B)_k.
- Update \( (w | \tilde{k}, \tilde{n}_d, \cdots) \) from the full conditional posterior given by (2.7).
- This proposal is then accepted or rejected according to \( \rho_d = 1/\rho_b \) with \( k \) replaced by \( k - 1 \).

Because the birth and death submoves are symmetrically defined, the likelihood ratio, prior ratio, transition ratio are the inverses of those in (3.4). Hence, using these directly would not violate the reversibility and dimension-matching constraint between the birth and death submoves.
3.1.3 Updating by adjustment (MH) submove

In the adjustment submove (MH) the number of components remains constant where a single component to be moved is chosen uniformly from the set \( \{1, \ldots, k\} \) and changed to a random new candidate component located between its neighboring components. Hence, its probability \( mh_k \), we can update \((\theta_{d,k}, k, w, n_d)\) through the proposal \((\tilde{\theta}_{d,k}, \tilde{k}, \tilde{w}, \tilde{n}_d)\) where \(\tilde{\mu}_d \in \mathcal{E}_{d,k}\). This move is given by the usual Metropolis-Hastings move, see for instance [Tierney 1994]. The proposal scheme is as follow:

- Draw \( \ell \sim \mathcal{U}_{\{1, \ldots, k\}} \) and \( \tilde{\sigma}_{d,\ell} \sim \pi(\sigma_{d,\ell}) \) and \( \delta_\ell \sim \pi(\delta_\ell) \).
- Draw \( \tilde{\mu}_{d,\ell} \) similarly in (B) and put \( \tilde{\mu}_d = (\mu_{d,1}, \ldots, \tilde{\mu}_{d,\ell}, \ldots, \mu_{d,k})^\top \).
- Draw \( \tilde{n}_d \sim \mathcal{P}(\tilde{\mu}_d) \) similarly in (B).
- Update \((w|\tilde{k}, \tilde{n}_d, \ldots)\) from the full conditional posterior given by (2.7).
- The acceptance probability turns out to be

\[
\rho_{mh} = \min \left\{ 1, \frac{L(y_d|\tilde{\theta}_{d,k}, \tilde{n}_d, w, \tilde{k})\pi(\tilde{\mu}_d|\tilde{k})\pi(\tilde{n}_d|\tilde{k}, w)q(n_d, \tilde{n}_d)}{L(y_d|\theta_{d,k}, n_d, w, k)\pi(\mu_d|k)\pi(n_d|k, w)} \right\}.
\]

Since the proposition distribution on the components positions is continuous uniform using constant \( \epsilon \), the instrumental densities of transition for positions \( \mu_d \) and the candidate \( \tilde{\mu}_d \) are identical. Since no dimension change is required, the new component parameters are accepted with a Metropolis-Hastings acceptance/rejection probability.

3.2 Convergence of the trans-dimensional MCMC scheme

We study in the sequel the convergence properties of the reversible jump Metropolis-Hastings within Gibbs scheme described in section 3.1. Thus, in investigating convergence, we check that our Metropolis-Hastings transition kernel which makes transitions according to a mechanism as in submove (MH) satisfies the reversibility condition (also called detailed balance) given in [Tierney 1998]. For reason of simplicity, we omit the discussion for (B) and (D) submoves, because it is similar to that for (MH) submove. The distribution \( \pi^*(dw|y_d) \) denotes the marginal posterior of \( w \) and, by putting \( \nu_d = (\sigma_{d,j}, \delta_j, n_d, k) \), \( K(\theta, d(\tilde{\mu}_d, \tilde{\nu}_d)) \) denotes the transition kernel of replacing \( \mu_d \) by \( \tilde{\mu}_d \), \( \sigma_{d,j} \) by \( \tilde{\sigma}_{d,j} \), \( \delta_j \) by \( \tilde{\delta}_j \), \( n_d \) by \( \tilde{n}_d \) and \( k \) by \( \tilde{k} \). Using the kernel \( K \) and the posterior conditional distribution \( \pi^*(dw|y_d) \), we can write

\[
K^{MH}(\theta, d(\tilde{\mu}_d, \tilde{\nu}_d), w)) = \pi^*(dw|y_d)K(\theta, d(\tilde{\mu}_d, \tilde{\nu}_d)).
\]

To show that \( \pi^* \) is an invariant distribution for \( K^{MH} \), it suffices to prove that \( \pi^*(dw|y_d) \) is an invariant distribution for \( K \) [Chib & Greenberg 1995]. We note by \( \pi^*(d(\mu_d, \nu_d)|y_d) \) the posterior conditional distribution of \((\mu_d, \delta_j, \sigma_{d,j}, n_d, k)\) conditional to \( y_d \) and \( w \). By definition of \( K \), we have

\[
K(\theta, d(\tilde{\mu}_d, \tilde{\nu}_d)) = k^{-1} \sum_{j=1}^k K_j(\theta, d(\tilde{\mu}_d, \tilde{\nu}_d)),
\]

where \( K_j(\theta, d(\tilde{\mu}_d, \tilde{\nu}_d)) \) is the transition kernel associated to replacing \( \mu_{d,j} \) by \( \tilde{\mu}_{d,j} \), \( \sigma_{d,j} \) by \( \tilde{\sigma}_{d,j} \), \( \delta_j \) by \( \tilde{\delta}_j \), \( n_{j-1} \) by \( \tilde{n}_{j-1} \), \( n_j \) by \( \tilde{n}_j \) and \( k \) by \( \tilde{k} \). Let \( C \) be a compact of \( \mathbb{R}^k \), we define the set \( E \) by

\[
E := \left\{ \tilde{\mu}_d = (\tilde{\mu}_{d,1}, \ldots, \tilde{\mu}_{d,k})' \in C : \tilde{\mu}_d \in \mathcal{E}_{d,k} \right\}.
\]
For all $\bar{\mu}_d \in \mathcal{E}$ and $j \in \{1, \ldots, k\}$, we define

\[ E_{\bar{\mu}_d,j} := \{(\mu_{d,j} \in \mathbb{R} : (\bar{\mu}_{d,1}, \ldots, \bar{\mu}_{d,j}, \ldots, \bar{\mu}_{d,k}) \in \mathcal{E}_{d,k}\}. \tag{3.6} \]

Thus, according to a mechanism as in move (MH)$_k$ where $k = k$, we can write

\[ K_j(\theta, d(\bar{\mu}_d, \bar{\nu}_d)) = \rho_{mh} Q_j(\mu_d, d\bar{\mu}_d) q(\nu_d, d\bar{\nu}_d) + r(\theta) \delta_{\mu_d}(d\bar{\mu}_d) \delta_{\nu_d}(d\bar{\nu}_d), \]

where $\delta_x(A)$ denotes the Dirac mass and

\[ r(\theta) = 1 - \sum_{\bar{n}_d \in \{1, \ldots, n\}^k} \int \int \rho_{mh} Q_j(\mu_d, d\bar{\mu}_d) q(\nu_d, d\bar{\nu}_d), \]

\[ Q_j(\mu_d, d\bar{\mu}_d) = C_{\mu_d,j} 1_{E_{\mu_d,j}} (d\bar{\mu}_d) d\bar{\nu}_d, \]

\[ C_{\mu_d,j} = \left( \int_{E_{\mu_d,j}} (d\bar{\mu}_d) \right)^{-1}, \]

\[ q(\nu_d, d\bar{\nu}_d) = d\delta, d\tilde{\sigma}_{d,j} d\bar{\nu}_{d,j-1} d\bar{\nu}_{d,j} \prod_{l \neq j} d\bar{\bar{\nu}}_{d,l}. \]

Clearly, $Q_j(\mu_d, d\bar{\mu}_d)$ change one element in $\mu_d$ of index $j$ uniformly on $E_{\mu_d,j}$ and $q(\nu_d, d\bar{\nu}_d)$ change the elements $(\delta_j, \sigma_{d,j}, n_{j-1}, n_j)$ in $\mathbb{R} \times [0, \infty) \times \{1, \ldots, n\}^2$. The posterior conditional distribution $\pi^*(d(\mu_d, \nu_d)|y_d)$ is given by

\[ \pi^*(d(\mu_d, \nu_d)|y_d) = C 1_{E}(\mu_d) p^*(\mu_d, \nu_d|y_d) d\mu_d d\nu_d, \tag{3.7} \]

where $d\mu_d = \prod_{j=1}^k d\mu_{d,j}$ is the Lebesgue measure on $\mathbb{R}^k$, $d\mathcal{M}_d$ is the counting measure on $\{1, \ldots, n\}^k$ and

\[ p^*(\mu_d, \nu_d|y_d) = L(y_d|\theta_d,k, n_d, w, k) \pi(\mu_d) \pi(\nu_d) d\mu_d d\nu_d, \]

\[ C = \left( \sum_{n_d \in \{1, \ldots, n\}^k} \int \int 1_{E}(\mu_d) p^*(\mu_d, \nu_d|y_d) d\mu_d d\nu_d \right)^{-1}. \]

We denote by $\beta = (\mu_d, \nu_d)$ and $\mathcal{M}_j$ the measure on $\mathbb{R}^{k+1} \times [0, \infty) \times \{1, \ldots, n\}^k \times \mathbb{R}^{k+1} \times [0, \infty) \times \{1, \ldots, n\}^k$ given by

\[ \mathcal{M}_j(d\beta, d\check{\beta}) = d\delta, d\check{\delta}, d\tilde{\sigma}_{d,j} d\tilde{\sigma}_{d,j} d\tilde{\mu}_d d\tilde{\nu}_{d,j-1} d\tilde{\nu}_{d,j} \prod_{l \neq j} d\tilde{\nu}_{d,l} d\tilde{\nu}_{d,l} \prod_{l \neq j} d\tilde{\bar{\nu}}_{d,l}. \]

The measure $\mathcal{M}_j$ will be used to verify that the transition kernel satisfies the detailed balance condition. To do this, let’s consider the following Lemma.

**Lemma 1** For all measurable function $\varphi \geq 0$, we have

\[ \int \varphi(\beta, \check{\beta}) \mathcal{M}_j(d\beta, d\check{\beta}) = \int \varphi(\check{\beta}, \beta) \mathcal{M}_j(d\beta, d\check{\beta}). \tag{3.8} \]

For a proof, see Appendix A.2.

Now, in order to check the reversibility condition, it is reasonable to take $\varphi$ of the form:
\[ \varphi(\tilde{\beta}, \beta) = \left( \frac{p^*(\tilde{\mu}_d, \tilde{\nu}_d | \mathbf{y}_d)}{p^*(\mu_d, \nu_d | \mathbf{y}_d)} q(n_d, \tilde{n}_d) \lor 1 \right) p^*(\mu_d, \nu_d | \mathbf{y}_d) C_{\mu_d, j} 1_{E_{\mu_d, j}} (\mu_d,j) C_{1E} (\mu_d). \] (3.9)

Thus, it is straightforward that

\[ \rho_{mh} Q_j (d\tilde{\mu}_d, \mu_d) q(d\tilde{\nu}_d, \nu_d) \pi^* (d(\mu_d, \nu_d) | \mathbf{y}_d) = \varphi(\tilde{\beta}, \beta) m_j (d\tilde{\beta}, d\beta). \]

Let us now verify that \( \varphi(\tilde{\beta}, \beta) = \varphi(\beta, \tilde{\beta}) \) of course according to \( \varphi \) given by (3.9). As for all \( l \neq j \) we have \( \tilde{\mu}_{d,l} = \mu_{d,l}, m_j \)-almost surely, we have that, for \( m_j \)-almost all \( (\beta, \tilde{\beta}) \):

\[ \varphi(\tilde{\beta}, \beta) = \left( \frac{p^*(\tilde{\mu}_d, \tilde{\nu}_d | \mathbf{y}_d)}{p^*(\mu_d, \nu_d | \mathbf{y}_d)} q(n_d, \tilde{n}_d) \lor 1 \right) p^*(\mu_d, \nu_d | \mathbf{y}_d) C_{\mu_d, j} 1_{E_{\mu_d, j}} (\mu_d,j) C_{1E} (\mu_d) = \varphi(\beta, \tilde{\beta}). \]

Now, we are in position to give our theorem for the MCMC scheme convergence. Before this, let check that the Markov chain defined is irreductible and aperiodic. Irreductibility is easily established, since the chain can move from any value of \( k \) to any other value in steps of one at a time and the parameters are updated by drawing from continuous distributions whose supports are the natural parameter spaces. Furthermore, aperiodicity is clear, since given any arbitrarily small neighborhood of a current state \( (\mu_d, \nu_d) \) there is positive probability that after one sweep through moves \( (B)_k \), \( (D)_k \) and \( (MH)_k \) the chain lies in that neighborhood.

**Theorem 1** For all measurable sets \( A \) and \( B \), the identity

\[ \int_B \int_A \rho_{mh} Q_j (\mu_d, d\tilde{\mu}_d) q(\nu_d, d\tilde{\nu}_d) \pi^* (d(\mu_d, \nu_d) | \mathbf{y}_d) = \int_A \int_B \rho_{mh} Q_j (\mu_d, d\tilde{\mu}_d) q(\nu_d, d\tilde{\nu}_d) \pi^* (d(\mu_d, \nu_d) | \mathbf{y}_d) \] (3.10)

holds and thus \( \pi^* (d(\mu_d, \nu_d) | \mathbf{y}_d) \) is invariant for \( K \). Furthermore, for \( \pi^* \)-almost all \( \beta \),

\[ ||K^r(\theta, d(\tilde{\mu}_d, \tilde{\nu}_d), \cdot) - \pi^*||_{TV} \rightarrow 0 \quad \text{as} \quad r \rightarrow \infty. \] (3.11)

For a proof, see Appendix A.3.

The above result (3.11) ensures convergence of the Markov chain to the distribution \( \pi^* \), but it does not provide a convergence rate. It is known that, in general, a relevant feature for the convergence rate is the correlation among the components of the Markov chain implied in the stationary distribution \( \pi^* \). Convergence is in general slower if the correlation is strong. Roughly speaking, to provide a convergence rate, we should show a quantitative bound for \( ||K^r(\theta, d(\tilde{\mu}_d, \tilde{\nu}_d), \cdot) - \pi^*||_{TV} \) by showing that the transition kernel \( K \) satisfies the Doeblin condition.

### 4 Numerical computations

In this section we implemented the methodology described in Section 3.1 in order to establish the performance and flexibility of the method. We conducted simulation studies under a variety of settings that intend to investigate the capacity of the method to correctly identify the positions of the underlying nucleosomes, for different numbers of components and component sizes. In particular, we desired to show that the estimated components parameters

\[ 11 \]
and specifically position and variance parameters can be accurately estimated. We consider two scenarios within which to test the method, differing in the nature of data of the "true" mixture densities. The first, referred to as the simulated scenario, is characterized by a two simulated examples of mixture of Student’s-t distributions. In the real scenario, our method is evaluated on real data: MNase-Seq Kaplan data for budding yeast and sonicated ChIP-Seq Heinz data for mouse Puer cell line.

4.1 Simulated scenario

We propose two examples of mixture of Student’s-t distributions, as described in (2.1), with \( k = 4 \) and \( k = 10 \), respectively. In each example, we consider \( \mu_j = 150j, \sigma_j^2 = 400, \delta_j = 50 \) and \( w_j = 1/k, j = 1, \ldots, k \). For the results built from the RJMCMC strategy, we used two approaches. The first approach, called RJMCMC1, consisted to apply the RJMCMC algorithm to the forward strands and the reverse strands, separately. The second approach, named RJMCMC2, applied the RJMCMC algorithm simultaneously to the positive and negative strands. The results are based on 50 samples of the RJMCMC algorithms corresponding to \( T = 100,000 \) sweeps, following a burn-in period of 10,000 sweeps. The size of data is \( n = 1000 \) and the starting state being chosen by setting \( k = 3 \) and sampling the other parameters from their prior distributions. For the parameters \( \alpha \) and \( \lambda \), we choose \( \alpha_{d,j} = n_{d,j} + 1, d \in \{f, r\} \) and \( \alpha_j = 1 \) for RJMCMC2, and \( \lambda = 1 \) for both RJMCMC methods as this choice give good numerical results.

We present in Figure 2 some indications about the Multinomial-Dirichlet classification and the exploratory research of the RJMCMC algorithms. We notice that the RJMCMC1 procedure estimates well the number of components after only 5000 iterations, whereas the RJMCMC2 method finds the true value of \( k \) only after 10,000 iterations. In Figure 3 we compare the curves of the true densities and the estimated densities for the strand read positions when \( k = 4 \). We see that the RJMCMC1 method provides better results because that, in particular, it estimate very well the unknown variances as the posterior mean of variances under this method are closer to the true variances. In addition, for each of the three approaches, we report in Table 1 results of estimations for unknown parameters computed by the empirical means using only the iterations of the Markov chain where the number of components is \( k = 4 \). The results of the three approaches are relatively different for the estimated parameters, especially for \( \delta, \sigma_f, \sigma_r \). Indeed, the estimation of \( \delta \) based on PING is far from the true value of this parameter because the PING strategy is not adapted to estimate \( \delta \) in general case. Estimated 95% credible intervals for posterior mean values are given in Table 2. We notice that credible intervals obtained with RJMCMC1 are generally narrower than the ones built from RJMCMC2. Figures 4 and 5 display posterior distributions of \( \mu, \delta, \sigma_f \) and \( \sigma_r \) for the true model with \( k = 4 \). This gives indications on the estimated values and on the distributions of these parameters. Figure 6 shows the number of components \( k \) visited in the MCMC iterations for the true model with 4 and 10 components. Under this model, the RJMCMC1 method estimates well the true value of \( k = 10 \), contrarily to PING (because PING use the BIC to decide between \( k \) and the BIC is a penalized likelihood by a factor of dimension). This suggests that the RJMCMC1 strategy is more powerful than PING for the estimation of \( k \) when the dimension of the model is substantially high. Figure 7 displays the curves of the true densities and the ones of the estimated densities for the reads positions when \( k = 10 \). We notice that the RJMCMC1 method provide good results for the estimation of the parameters, since the curves of the true densities are very close to those of the estimated densities.

4.2 Real scenario

In this section, two real published data sets are used throughout the paper, as a basis for our comparisons between our method and the PING. The two data sets are: Kaplan MNase data from *Saccharomyces cerevisiae* (Kaplan et al., 2009), and Sonicated CHIP-Seq Heinz data from Puer cell line (Heinz et al., 2010). The results based on PING strategy are computed using the PING R-Package. For each of the experimental datasets, we splitted the genome in candidate regions that contain a minimum of reverse and forward strands, and compared the performance of PING and RJMCMC methods.
Figure 2: Some indications about the Multinomial-Dirichlet classification and the exploratory research of the RJMCMC algorithms for $T = 5000$ (first row), $T = 10000$ (second row) and $T = 50000$ (third row).
Figure 3: Curves of the true strand densities (in solid lines) and the estimated strand densities (in dashed lines) for the true mixture model with $k = 4$, $\mu_j = 150j$, $\sigma^2_f = \sigma^2_r = 400$, $\delta_j = 50$ and $w_j = 1/k$, $j = 1, \ldots, k$. The curves of the forward strands are represented in blue while the ones of the reverse strands are represented in red. The results based on 50 samples of 100000 iterations from the RJMCMC algorithms.

Figure 4: Posterior density of $\mu$ and $\delta$ for the true mixture model with $k = 4$, $\mu_j = 150j$, $\sigma^2_f = \sigma^2_r = 400$, $\delta_j = 50$ and $w_j = 1/k$, $j = 1, \ldots, k$. The results are derived 50 samples of 100000 runs from the RJMCMC1 algorithm.
Figure 5: Posterior density of \( \sigma_f \) and \( \sigma_r \) for the true mixture model with \( k = 4, \mu_j = 150j, \sigma^2_f = \sigma^2_r = 400, \delta_j = 50 \) and \( w_j = 1/k, j = 1, \ldots, k \). The results are obtained from 50 samples of 100000 runs from the RJMCMC1 algorithm.

Figure 6: Number of components \( k \) visited in the RJMCMC1 iterations for \( k = 4 \) (first row) and \( k = 10 \) (second row).
Figure 7: Curves of the true strand densities (in solid lines) and the estimated strand densities (in dashed lines) for the true mixture model with \( k = 10, \mu_j = 150j, \sigma_{fj}^2 = \sigma_{rj}^2 = 400, \delta_j = 50 \) and \( w_j = 1/k, j = 1, \ldots, k \). The curves of the forward strands are represented in blue while the ones of the reverse strands are represented in red. The results based on 50 samples of 100000 runs from the RJMCMC1 algorithm.

Table 1: Posterior mean of parameters for the true model with parameter values \( k = 4, \mu_j = 150j, \sigma_{fj}^2 = \sigma_{rj}^2 = 400, \delta_j = 50 \) and \( w_j = 1/k, j = 1, \ldots, k \).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>PING</th>
<th>RJMCMC1</th>
<th>RJMCMC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k )</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>( \hat{\mu} )</td>
<td>\begin{align*} \hat{\mu}_1 &amp;= 150.0233 \ \hat{\mu}_2 &amp;= 299.1054 \ \hat{\mu}_3 &amp;= 448.4577 \ \hat{\mu}_4 &amp;= 600.2255 \end{align*}</td>
<td>\begin{align*} \hat{\mu}_1 &amp;= 150.7673 \ \hat{\mu}_2 &amp;= 300.0314 \ \hat{\mu}_3 &amp;= 449.6801 \ \hat{\mu}_4 &amp;= 600.5186 \end{align*}</td>
<td>\begin{align*} \hat{\mu}_1 &amp;= 150.6413 \ \hat{\mu}_2 &amp;= 299.5103 \ \hat{\mu}_3 &amp;= 449.3746 \ \hat{\mu}_4 &amp;= 600.3231 \end{align*}</td>
</tr>
<tr>
<td>( \hat{\sigma}_f )</td>
<td>\begin{align*} \hat{\sigma}<em>{f1} &amp;= 24.29940 \ \hat{\sigma}</em>{f2} &amp;= 23.81048 \ \hat{\sigma}<em>{f3} &amp;= 23.74606 \ \hat{\sigma}</em>{f4} &amp;= 23.83906 \end{align*}</td>
<td>\begin{align*} \hat{\sigma}<em>{f1} &amp;= 19.78066 \ \hat{\sigma}</em>{f2} &amp;= 19.54551 \ \hat{\sigma}<em>{f3} &amp;= 19.69373 \ \hat{\sigma}</em>{f4} &amp;= 19.49507 \end{align*}</td>
<td>\begin{align*} \hat{\sigma}<em>{f1} &amp;= 22.25564 \ \hat{\sigma}</em>{f2} &amp;= 25.99794 \ \hat{\sigma}<em>{f3} &amp;= 18.69803 \ \hat{\sigma}</em>{f4} &amp;= 19.48980 \end{align*}</td>
</tr>
<tr>
<td>( \hat{\sigma}_r )</td>
<td>\begin{align*} \hat{\sigma}<em>{r1} &amp;= 26.40856 \ \hat{\sigma}</em>{r2} &amp;= 27.32508 \ \hat{\sigma}<em>{r3} &amp;= 26.98290 \ \hat{\sigma}</em>{r4} &amp;= 26.02905 \end{align*}</td>
<td>\begin{align*} \hat{\sigma}<em>{r1} &amp;= 19.50021 \ \hat{\sigma}</em>{r2} &amp;= 21.78654 \ \hat{\sigma}<em>{r3} &amp;= 21.39388 \ \hat{\sigma}</em>{r4} &amp;= 19.17606 \end{align*}</td>
<td>\begin{align*} \hat{\sigma}<em>{r1} &amp;= 16.36062 \ \hat{\sigma}</em>{r2} &amp;= 34.48250 \ \hat{\sigma}<em>{r3} &amp;= 28.26934 \ \hat{\sigma}</em>{r4} &amp;= 31.56446 \end{align*}</td>
</tr>
<tr>
<td>( \hat{\delta} )</td>
<td>\begin{align*} \hat{\delta}_1 &amp;= 149.3080 \ \hat{\delta}_2 &amp;= 148.9967 \ \hat{\delta}_3 &amp;= 148.2038 \ \hat{\delta}_4 &amp;= 146.6910 \end{align*}</td>
<td>\begin{align*} \hat{\delta}_1 &amp;= 51.55482 \ \hat{\delta}_2 &amp;= 51.06095 \ \hat{\delta}_3 &amp;= 52.22995 \ \hat{\delta}_4 &amp;= 52.05643 \end{align*}</td>
<td>\begin{align*} \hat{\delta}_1 &amp;= 53.11760 \ \hat{\delta}_2 &amp;= 52.60536 \ \hat{\delta}_3 &amp;= 46.89090 \ \hat{\delta}_4 &amp;= 40.55607 \end{align*}</td>
</tr>
</tbody>
</table>
Table 2: The 95% credible intervals of parameters of the true mixture model with $k = 4$, $\mu_j = 150j$, $\sigma_f^2 = \sigma_r^2 = 400$, $\delta_j = 50$ and $w_j = 1/k$, $j = 1, \ldots, k$. The results are based on 50 samples of 100000 runs from the RJMCMC algorithms.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>RJMCMC1</th>
<th>RJMCMC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\mu}$</td>
<td>$\mu_1$</td>
<td>$[149.5646, 152.5463]$</td>
</tr>
<tr>
<td></td>
<td>$\mu_2$</td>
<td>$[297.1576, 304.4735]$</td>
</tr>
<tr>
<td></td>
<td>$\mu_3$</td>
<td>$[446.9918, 454.2187]$</td>
</tr>
<tr>
<td></td>
<td>$\mu_4$</td>
<td>$[597.7653, 604.9235]$</td>
</tr>
<tr>
<td>$\hat{\sigma_f}$</td>
<td>$\sigma_{f1}$</td>
<td>$[17.96529, 20.94329]$</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{f2}$</td>
<td>$[16.88599, 22.32086]$</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{f3}$</td>
<td>$[17.65742, 22.06642]$</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{f4}$</td>
<td>$[18.05622, 21.70781]$</td>
</tr>
<tr>
<td>$\hat{\sigma_r}$</td>
<td>$\sigma_{r1}$</td>
<td>$[17.13639, 21.96415]$</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{r2}$</td>
<td>$[19.96422, 25.74073]$</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{r3}$</td>
<td>$[20.59076, 22.44596]$</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{r4}$</td>
<td>$[17.47192, 20.30170]$</td>
</tr>
<tr>
<td>$\hat{\delta}$</td>
<td>$\delta_1$</td>
<td>$[39.65471, 64.40973]$</td>
</tr>
<tr>
<td></td>
<td>$\delta_2$</td>
<td>$[41.42344, 63.35966]$</td>
</tr>
<tr>
<td></td>
<td>$\delta_3$</td>
<td>$[43.83419, 61.89987]$</td>
</tr>
<tr>
<td></td>
<td>$\delta_4$</td>
<td>$[40.12432, 64.28964]$</td>
</tr>
</tbody>
</table>

4.2.1 Sonicated CHIP-Seq data

We employed the sonicated CHIP-Seq data for chromosome 1 from mouse PUER cell line. After having fragmented the data in candidate regions, we selected three regions (named A, B, C) on which we present a comparative study between the RJMCMC methods and the PING strategy. The candidates regions A, B and C were selected with respect to the number of reads. Region A contains 8 forward strands and 11 reverse strands, region B counts 22 forward reads and 18 reverse reads, and region C includes 29 forward strands and 18 reverse strands. Our aim is to show that the RJMCMC methods provide good results as well for regions with enough reads as for regions with a small number of reads. The three selected candidate regions are visualized in Figure 8 where we give indications on the number of nucleosomes that should be present in each of the three candidate regions. We notice that the number of nucleosomes in each of the region A, B and C should be equal to 2. We present in Figures 9, 10 and 11 the results of the numerical computations by giving the estimated density curves of data in regions A, B and C, respectively. In each of these figures, the curves of the RJMCMC methods are in adequateness with Figure 8, contrarily to the curves based on the PING strategy (for Regions B and C). In this data set for the sonicated CHIP-Seq data from mouse PUER cell line, it is immediately apparent that the RJMCMC procedures are more performant than the PING method.

4.2.2 MNase-Seq data

The data set based on the chromosome 1 of Kaplan MNase-Seq data from budding yeast has been analyzed with our methodology and the PING. The comparison study between the three approaches are performed on two candidate regions, named A and B. The first region contains 26 forward reads and 35 reverse strands. The second region includes 153 forward reads and 209 reverse reads. The reads positions are visualized in Figure 12. Moreover, the Figures 13 and 14 display the estimations of the density curves of data in regions A and B, respectively. The estimated density curves identify 2 nucleosomes in region A and 3 nucleosomes in region B, which is in phase with Figure 12. We report in Tables 3 and 4 the computed posterior mean and the computed 95% credible set for parameters estimated from Kaplan MNase-Seq data belonging to region B. It is seen that for each estimated parameter, results are again relatively different between PING and both the RJMCMC methods.
5 Discussion

We have provided a Bayesian framework for analysis of nucleosomes occupancy. On the one hand, the prior knowledge about the genetic structure of nucleosomes is fully respected. On the other hand, for identifying the number and position of nucleosomes, the Multinomial-Dirichlet model and priors distributions incorporates both data and estimation uncertainty. Actual computations were performed by using a trans-dimensional MCMC scheme. Software written in R may be obtained from the authors.

Several interlinked aspects of the methodology proposed are demonstrated in illustrating its performance. We have displayed examples of the results that we obtain from simulated and real data sets. The numerical studies for the considered data sets have showed that our methodology is performant as well for MNase-Seq Kaplan data as for sonicated Chip-Seq Heinz data. Furthermore, we notice that the RJMCMC1 strategy is more robust than PING for sonicated CHIP-Seq data from mouse PUER cell line, and estimates more precisely the number, the position and specially the dispersion (variance) of nucleosome distributions.

Finally, in a conceptually viewpoint, it is not difficult to see how to generalize the approach to non-parametric inference. Practically, however, there are some serious obstacles to be overcome. In non-parametric model, the density of the mixture may be an arbitrary density of probability measure satisfying a certain tail constraint.

Figure 8: Sonicated CHIP-Seq data of three candidate regions for chromosome 1 from mouse PUER cell. The blue color corresponds to the forward strands and the red color illustrates the reverse strands.

Estimations based on data from region A was investigated and led to the same concluding remarks (results not reported here).
Figure 9: Estimation of the density curves of the forward read positions (in blue) and the reverse read positions (in red) of the candidate Region A for the sonicated Heinz data from mouse PUER cell line.

Figure 10: Estimation of the density curves of the forward read positions (in blue) and the reverse read positions (in red) of the candidate Region B for the sonicated CHIP-Seq data from mouse PUER cell line.
Figure 11: Estimation of the density curves of the forward read positions (in blue) and the reverse read positions (in red) of the candidate Region C for the sonicated Heinz data from mouse PUER cell line.

Figure 12: MNase-Seq data of two candidate regions for chromosome 1 from *S. cerevisiae* budding yeast. The blue color corresponds to the forward strands and the red color illustrates the reverse strands.
Figure 13: Estimation of the density curves of the forward read positions (in blue) and the reverse read positions (in red) of the candidate Region A for the MNase-Seq Kaplan data from budding yeast.

Figure 14: Estimation of the density curves of the forward read positions (in blue) and the reverse read positions (in red) of the candidate Region B for the MNase-Seq Kaplan data from budding yeast.
Table 3: Estimation of the parameters of the model based on the data of Region A from MNase-Seq Kaplan data.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>PING</th>
<th>RJMCMC1</th>
<th>RJMCMC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>$\hat{\mu}_1$, $\hat{\mu}_2$</td>
<td>73524.58, 73694.01</td>
<td>73537.41, 73705.40</td>
</tr>
<tr>
<td>$\hat{\sigma}_f$</td>
<td>$\hat{\sigma}<em>{f_1}$, $\hat{\sigma}</em>{f_2}$</td>
<td>29.15067, 31.83214</td>
<td>5.935524, 36.74604</td>
</tr>
<tr>
<td>$\hat{\sigma}_r$</td>
<td>$\hat{\sigma}<em>{r_1}$, $\hat{\sigma}</em>{r_2}$</td>
<td>28.40015, 31.12998</td>
<td>14.90607, 25.47253</td>
</tr>
<tr>
<td>$\hat{\delta}$</td>
<td>$\hat{\delta}_1$, $\hat{\delta}_2$</td>
<td>145.4191, 153.7028</td>
<td>134.6065, 134.9570</td>
</tr>
</tbody>
</table>

Table 4: Estimation of the parameters of the model based on the data of Region B from MNase-Seq Kaplan data.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>PING</th>
<th>RJMCMC1</th>
<th>RJMCMC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>$\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\mu}_3$</td>
<td>154188.4, 154375.6, 154539.4</td>
<td>154048.6, 154240.7, 154404.9</td>
</tr>
<tr>
<td>$\hat{\sigma}_f$</td>
<td>$\hat{\sigma}<em>{f_1}$, $\hat{\sigma}</em>{f_2}$, $\hat{\sigma}_{f_3}$</td>
<td>35.00355, 29.58954, 25.70207</td>
<td>38.07391, 26.81227, 16.27982</td>
</tr>
<tr>
<td>$\hat{\sigma}_r$</td>
<td>$\hat{\sigma}<em>{r_1}$, $\hat{\sigma}</em>{r_2}$, $\hat{\sigma}_{r_3}$</td>
<td>36.97024, 27.35775, 25.39520</td>
<td>55.46999, 28.53077, 10.31662</td>
</tr>
<tr>
<td>$\hat{\delta}$</td>
<td>$\hat{\delta}_1$, $\hat{\delta}_2$, $\hat{\delta}_3$</td>
<td>154.2777, 148.1532, 146.5213</td>
<td>125.5518, 131.0992, 128.4602</td>
</tr>
</tbody>
</table>
First, a prior needs to be specified via splines or kernels or some other set of basis functions. Second, efficient MCMC scheme should be performed for numerical computations.

Acknowledgments

Arnaud Droit holds a Réseau de médecine génétique appliquée (RMGA) salary award. Rawane Samb holds a Ministère du Développement Économique, Innovation et Exportation (MDEIE) award.

Appendix A

A.1. Proof of proposition 1

For reasons of simplicity and in order to derive an exact expression for \( \pi(\mu|k) \), we put \( \beta = (\beta_1, \ldots, \beta_k)' \) a random vector with Gaussian distribution truncated to positives components \( \beta \sim N_k^+(0, \tau^{-1}I_k) \) where \( I_k \) is the \( k \times k \) identity matrix. Thus, the density of \( \beta \) is given by

\[
\pi(\beta|k) = C \exp\left(-\frac{\beta' \beta}{2}\right) \prod_{j=1}^{k} \mathbf{1}_{\beta_j > 0},
\]

where \( C \) is a constant given by

\[
C^{-1} = \int_{\beta_1 \geq 0} \exp\left(-\frac{\beta_1^2}{2}\right) d\beta_1 \cdots \int_{\beta_k \geq 0} \exp\left(-\frac{\beta_k^2}{2}\right) d\beta_k = \left(\sqrt{\frac{\pi}{2\tau}}\right)^k.
\]

Now, using the random vector \( \beta \), we can easily construct a vector \( \nu \) with Gaussian distribution restricted to the set \( \{\nu_1 \leq \cdots \leq \nu_k\} \) as follows: \( \nu = T_k^{-1} \beta \), where \( T_k^{-1} = (t'_{i,j})_{1 \leq i,j \leq k} \) is a \( k \times k \) matrix such that

\[
t'_{i,j} = \begin{cases}
1, & \text{if } i \geq j, \\
0, & \text{else}.
\end{cases}
\]

By inverting the matrix \( T_k^{-1} \), we can also express \( \beta \) in terms of \( \nu \) in the following way: \( \beta = T_k \nu \) where \( T_k \) is the matrix inverse of \( T_k^{-1} \) given by

\[
t_{i,j} = \begin{cases}
1, & \text{if } i = j, \\
-1, & \text{if } j = i - 1, i = 2, \ldots, k, \\
0, & \text{else}.
\end{cases}
\]

Therefore, the probability density of \( \nu \) can be deduced from that of \( \beta \) as follows:

\[
\pi(\beta|k)d\beta = \pi(T_k \nu |k) \left| \frac{d\beta}{d\nu} \right| d\nu
\]

\[
= C \mathbf{1}_{\{T_k \nu \geq 0\}} \exp\left(-\frac{(T_k \nu)'(T_k \nu)}{2}\right) \left| \frac{d\beta}{d\nu} \right| d\nu
\]

\[
= \left(\sqrt{\frac{\pi}{2\tau}}\right)^{-k} \exp\left(-\frac{\nu' (T_k'^{-1} T_k) \nu}{2}\right) \mathbf{1}_{\{\nu_1 \leq \cdots \leq \nu_k\}} d\nu,
\]

where the Jacobian is 1. It suffice to put \( \mu_d = \nu + \xi_d \) to complete the proof.

A.2. Proof of lemma 1
For all measurable sets $A_1, \ldots, A_{k+1}$ and $B_1, \ldots, B_{k+1}$, for all measurable sets $E$ and $F$ in $[0, \infty)$, for all countable sets $G_1, \ldots, G_k$ and $H_1, \ldots, H_k$ in $\{1, \ldots, n\}$, it is easy to verify that

$$m_j\left(\sum_{i=1}^{k+1} E_i \times G_i \times H_i \times \cdots \times \mu_{n+1} \times n_{n+1} \times \cdots \times \mu_n \times n_n \right) = \left( \int_{A_j} \mu_{d,j} \int_{B_j} \mu_{d,j} \right) \prod_{i \neq j} \int_{A_i \cap B_i} \mu_{d,j}$$

$$= \left( \int_{A_{k+1}} d\delta_j \int_{B_{k+1}} f\delta_j \right) \left( \int_{E} d\sigma_{d,j} \int_{F} d\sigma_{d,j} \right) \left( \sum_{j=1}^{k+1} \mu_{d,j} \sum_{j-1}^{k+1} \int_{H_{j-1}} \sum_{H_j} \int_{H_{j+1}} \cdots \sum_{H_{k+1}} \mu_{d,j} \right) \prod_{i \neq j-1}^{k+1} \sum_{H_i} \mu_{d,j}$$

We can check that (3.8) is true for the function $\varphi$ given by

$$\varphi(\beta, \tilde{\beta}) = 1_{A_1 \times \cdots \times A_{k+1} \times E \times G_1 \times \cdots \times G_k \times B_1 \times \cdots \times B_{k+1} \times F \times H_1 \times \cdots \times H_k} (\beta, \tilde{\beta}).$$

As the sets of the form $\{A_1 \times \cdots \times A_{k+1} \times E \times G_1 \times \cdots \times G_k \times B_1 \times \cdots \times B_{k+1} \times F \times H_1 \times \cdots \times H_k\}$ is a $\pi$-system that generates the Borel $\sigma$-field of $\mathbb{R}^{k+1} \times [0, \infty) \times \{1, \ldots, n\}^k \times \mathbb{R}^{k+1} \times [0, \infty) \times \{1, \ldots, n\}^k$, we deduce that (3.8) is true for a function $\varphi$ of the form $\varphi(\beta, \tilde{\beta}) = 1_{A_1 \times B_1} (\beta, \tilde{\beta})$ and for any measurable set $\mathfrak{A}$ and countable set $\mathfrak{B}$. Thus, by the theorem of Beppo-Levy, we can conclude that (3.8) is true for all function $\varphi \geq 0$ which complete the proof.

### A.3. Proof of theorem 1

For all measurable sets $A$ and $B$, we can write

$$\int_{A} \int_{B} \varphi(\beta, \tilde{\beta}) m_j(d\beta, d\tilde{\beta}) = \int \int 1_A(\beta) 1_B(\tilde{\beta}) \varphi(\beta, \tilde{\beta}) m_j(d\beta, d\tilde{\beta})$$

$$= \int 1_B(\tilde{\beta}) \int 1_A(\beta) \varphi(\beta, \tilde{\beta}) m_j(d\beta, d\tilde{\beta})$$

$$= \int_{B} \int_{A} \varphi(\beta, \tilde{\beta}) m_j(d\beta, d\tilde{\beta}),$$

and it is straightforward that the identity (3.10) holds. Then, by using the identity (3.10) and considering the kernel $K_j(\theta, d(\tilde{\mu_d}, \tilde{\nu_d}))$, we can write the following calculation

$$\sum_{n=1}^{\infty} \int_{\mathbb{R}^{k+1}} K_j(\theta, A) \pi^*(d(\mu_d, \nu_d)) |y_d|$$

$$= \sum_{n=1}^{\infty} \int_{\mathbb{R}^{k+1}} K_j(\theta, d(\tilde{\mu}_d, \tilde{\nu}_d)) \pi^*(d(\mu_d, \nu_d)) |y_d|$$

$$= \sum_{n=1}^{\infty} \int_{\mathbb{R}^{k+1}} Q_j(\mu_d, d(\tilde{\mu}_d)) \pi^*(d(\mu_d, \nu_d)) |y_d|$$

$$+ \int_{A} r(\theta) \pi^*(d(\mu_d, \nu_d)) |y_d|$$

Thus, $\pi^*(d(\mu_d, \nu_d)) |y_d|$ is invariant for $K_j(\theta, d(\tilde{\mu}_d, \tilde{\nu}_d))$ and consequently we have $\pi^*(d(\mu_d, \nu_d)) |y_d|$ is also invariant for $K$. With detailed balance satisfied and as $K$ is $\pi^*$-irreducible, aperiodic, we conclude that, for $\pi^*$-almost all $\beta$, the convergence in total variation (3.11) holds.
References


