Published in final edited form as:

J Stat Theory Pract. 2013 ; 7(2): 248–258. doi:10.1080/15598608.2013.772838.

Propriety conditions for the Bayesian autologistic model – Inference for histone modifications

Riten Mitra¹, Peter Müller², and Yuan Ji³

¹ ICES, University of Texas at Austin, USA

²Department of Mathematics, University of Texas at Austin, USA

³ Center for Clinical and Research Informatics, NorthShore University HealthSystem, USA

Abstract

Motivated by inference for a set of histone modifications we consider an improper prior for an autologistic model. We state sufficient conditions for posterior propriety under a constant prior on the coefficients of an autologistic model. We use known results for a multinomial logistic regression to prove posterior propriety under the autologistic model. The conditions are easily verified.

Keywords

Autologistic; Bayesian; Identifiability; Propriety

1 Introduction

We investigate the properties of the posterior distribution for an autologistic model using an improper uniform prior for the coefficients. Sufficient conditions for the propriety of the posterior distribution on the autologistic coefficients are obtained. While a proper prior specification always ensures a proper posterior in Bayesian analysis, investigators nevertheless often prefer to use improper priors to reflect ignorance, or simply for mathematical convenience. The motivating application for this study is inference for histone modifications (HMs), an important part of gene regulation. We use an autologistic model to represent the joint distribution of binary indicators for the presence or absence of HMs. The lack of extensive prior knowledge on the dependence of different HMs leads us to consider non-informative priors in this study.

In general, there are several important reasons to consider improper priors for the autologistic model. First, in the autologistic model it is difficult to derive standard noninformative priors such as Jeffreys' prior and reference priors (Bernardo et al., 1992). Second, simple noninformative priors are easy to implement. It is well known that improper priors in Bayesian statistics may lead to proper posterior distributions and useful inference procedures. However without careful examination of propriety conditions, simple

noninformative priors can be misused, sometimes unknowingly, and lead to meaningless inference and problems in posterior MCMC simulation (Hobert and Casella, 1996). From this perspective, the results proved for the autologistic model are not only of theoretical interest, but also relevant from a practical perspective.

There is a long history of research on suitable propriety conditions in the context of discrete response models. Zellner and Rossi (1984) examined propriety for independent binary responses. Ibrahim and Laud (1991) established a sufficient condition for the propriety of the posterior distribution under Jeffreys' prior in Generalized Linear Models (GLMs), while Natarajan and McCulloch (1995) provided necessary and sufficient conditions for the existence of the posterior distribution of the variance components in a class of mixed models for binomial responses. Chen and Shao (2001) established necessary and sufficient conditions for the propriety of the posterior distribution with a uniform improper prior on the regression coefficients for independent dichotomous response models with general links. However, there is a lack of literature on conditions for posterior propriety in autologistic models.

The article is organized as follows. We introduce the histone modifications data and related modeling issues in Section 2. In Section 3, we introduce the Bayesian autologistic model with a flat improper prior on the coefficients. In Section 4, we characterize sufficient conditions for posterior propriety in the autologistic model. We discuss in Section 5 some aspects of inference in autologistic models with and without a graphical model to characterize the dependence structure. In Section 6 we consider some computational issues related to the implementation of posterior inference. In Section 7 we report results for the earlier described histone modification data. Finally, we conclude with a short discussion in Section 8.

2 Histone Modifications

2.1 Data

Histones are strongly alkaline proteins that package the DNA and other proteins into densely structured units called *nucleosomes*. Post-translational modifications of histone proteins are believed to play an important role in gene regulations. These modifications are due to biochemical processes, such as methylation, acetylation, phosphorylation, and ubiquitination. A comprehensive list of histone modifications (HM) appears at <http://bioinfo.hrbmu.edu.cn/hhmd> (Zhang et al., 2010).

The interest in studying HMs arises from the well-known correlation of HMs with translational activity (Barski et al., 2005). While these correlations are well documented, correlations and dependence structure among HMs themselves are largely unknown. Dependence structure among HMs is hypothesized to regulate genes which is known as the histone code in epigenetics. We attempt to reveal the histone code by proposing an autologistic model for HM interactions. We use data for HMs from a ChIP-seq experiment for CD4 T lymphocytes, described in Wang et al. (2008). The data reports counts for $m = 39$ types of HMs including 18 acetylations, 20 methylations, and one special histone modification H2A.Z. A high count indicates an enrichment of the HM at the corresponding

genomic location. Counts are defined for windows of 3000 base pairs each. The data reports counts for $n = 50000$ such windows. We consider inference for a subset of $m = 11$ HMs out of the 39, assuming no prior knowledge on their dependency structure. The selected HMs are chosen for their known important role in gene regulation.

Figure 1 summarizes inference on the dependence structure of these $m = 11$ HMs under a Gaussian graphical model (GGM) using the R package *deal* (<http://cran.r-project.org/web/packages/deal/deal.pdf>). The GGM is perhaps the most commonly used model for inference on high dimensional joint distributions and finds numerous applications in machine learning and statistics. See for example Heckerman and Geiger (1994, 1995). Let $\mathbf{c}_t$ denote the vector of 11 HM counts for the window t . For inference with the GGM, we first transform the original counts to $z_{ti} = \log(c_{ti} + \epsilon)$. The GGM assumes that the m -dimensional response vectors $\mathbf{z}_t$ arise from a multivariate normal sampling model. The GGM focuses on inference for the conditional independence structure, i.e., zeroes in the multivariate normal precision matrix. We refer to Heckerman and Geiger (1995) for a detailed description of the model. Figure 1 shows the reported conditional independence structure. The vertices of the graph correspond to the $m = 11$ HMs. The absence of a line between any two HMs k and l indicates conditional independence of the two HM counts, conditional on all other counts.

2.2 Sampling Model

Inference under the Gaussian graphical model (GGM) is a useful preliminary data analysis, but is limited by the assumption of multivariate normal sampling model.

We propose alternative inference based on a sampling model in the form of a mixture of Poisson and lognormal distributions for each HM count, with a Poisson distribution for the low counts, say $c_{ti} < \kappa_i$, and a mixture of two log normal distributions for moderate to high counts.

This sampling model is inspired by inspection of the marginal empirical distributions of the observed HM counts (not shown). We introduce latent indicators e_{ti} , defining e_{ti} as a latent binary variable that codes for presence ($c_{ti} > \kappa_i$) versus absence (low counts, $c_{ti} < \kappa_i$) of HM i at location t . Let $\text{Poi}(\lambda)$ denote a Poisson distribution with mean λ and let $\text{LN}(m, s^2)$ denote a log normal distribution with location and scale m and s . We assume the sampling model

$$p(c_{ti} | e_{ti}) \propto \begin{cases} \text{Poi}(\lambda_i) I(c_{ti} < \kappa_i) & e_{ti} = 0 \\ \pi_i \text{LN}(\mu_{1i}, \rho_{1i}^2) + (1 - \pi_i) \text{LN}(\mu_{2i}, \rho_{2i}^2) & e_{ti} = 1 \end{cases} \quad (2.1)$$

We interpret $e_{ti} = 0$ and the corresponding Poisson distribution for low counts as background when the HM is not present, i.e., under $e_{ti} = 0$.

In the rest of this discussion we focus on the prior $p(\mathbf{e}_i | G, \boldsymbol{\beta})$ for the indicators of presence of the HMs. The model is indexed by two parameters. The dependence structure G and the strength of the dependence $\boldsymbol{\beta}$. Dependence structure is usually represented as a graph $G = (V, E)$. The graph consists of a set of nodes V and a set of edges $E \subset \{(i_1, i_2), i_1, i_2 \in V\}$. Here $V = \{1, \dots, m\}$ corresponds to the m HMs. The absence of an edge (i_1, i_2) in E indicates that

HMs i_1 and i_2 are conditionally independent given all other HMs. We also use $i_1 \sim i_2$ as short notation for $(i_1, i_2) \in E$. The graph is a statement about the dependence structure of the indicators e_{it} . For the moment we will assume that G is known and focus on the model $p(\mathbf{e}_t | G, \boldsymbol{\beta})$.

Inference on G and an in-depth analysis of the HM data are reported in another paper (Mitra et al., 2011).

3 An Autologistic Model

Besag (1974) gives a convenient parametrization of all possible joint probability models for a multivariate binary vector $\mathbf{e} = (e_1, \dots, e_m) \in \{0, 1\}^m$ under positivity condition. The positivity condition simply states that the probability of the null vector is non zero. We assume that an underlying known graph $G = (V, E)$ specifies the conditional independence structure.

Under positivity and given a set of coefficients $\boldsymbol{\beta}$,

$$p(\mathbf{e} | \boldsymbol{\beta}, G) = p(0 | \boldsymbol{\beta}) \times \exp \left\{ \sum_i \beta_i e_i + \sum_{i < j} \beta_{ij} e_i e_j + \sum_{i < j < k} \beta_{ijk} e_i e_j e_k + \dots + \beta_{1\dots m} e_1 \dots e_m \right\} \quad (3.1)$$

For any k -tuple of vertices, $\{i_1, \dots, i_k\}$ that do not form a *clique* in a graph $G = (V, E)$ the corresponding interaction coefficient $\beta_{i_1 \dots i_k}$ is zero. A clique is a set of vertices of which all pairs of vertices in the set are connected, i.e., $\{i_1, i_2\} \in E$ for all i_1, i_2 in the set. The conditional independence structure implied in G is thus equivalent to a restriction on $\boldsymbol{\beta}$.

Henceforth, we use $\boldsymbol{\beta}$ to denote the vector of all non-zero coefficients $\beta_{i_1 \dots i_k}$. For simplicity, we shall assume that all interaction coefficients of order higher than 2 are 0. In the following section, we shall prove sufficient conditions of posterior propriety for model (3.1) under a flat prior $p(\boldsymbol{\beta}) \propto c$.

In the inference for the HM data we use the autologistic model $p(\mathbf{e}_t | \boldsymbol{\beta}, G)$ as prior for the $\mathbf{e}_t$ indicators in the sampling model (2.1). Recall that $\mathbf{e}_t \in \{0, 1\}^m$ is a latent binary vector with $e_{it} = 1$ indicating presence of the i -th HM. We will use $\mathbf{Y}$ to indicate the full $(n \times m)$ data matrix of all indicators, with the t -th row equal to $\mathbf{e}_t$.

4 Sufficient Conditions for Propriety

The relative lack of substantive prior information about likely correlations of HMs leads us to consider a noninformative flat prior for $\boldsymbol{\beta}$. The use of this improper prior requires the investigation of posterior propriety. In this section we give an easily verifiable condition to assure posterior propriety. We shall use the following result from Maddala (1983).

Consider a generic multinomial logistic regression with categorical responses $y_i \in \{1, \dots, k\}$, covariate vectors $\mathbf{x}_{ij} \in \mathbb{R}^p$ and

$$p(y_t=j|\beta, x_t) \propto \exp(x_{tj}'\beta) \quad (4.1)$$

Here, the second index on the covariate vector x_{tj} allows the covariate vector to vary across the level j of the categorical response. The model can be motivated by the random utility model (McFadden, 1973). We define continuous latent random variables ϵ_{tj} with a standard type I extreme value distribution, i.e., $p(\epsilon_{tj} < x) = F(x) = \exp[-\exp(-x)]$ and let

$$U_{tj} = x_{tj}'\beta + \epsilon_{tj}$$

The multinomial logistic model is equivalent to $y_i = j$ if $U_{tj} > U_{tl}$ for all $l \neq j$.

In the context of the autologistic model, recall that $\{e_t, t = 1, \dots, n\}$ denote the n binary vectors with $e_t = (e_{ti}, i = 1, \dots, m)$. Let $p = m + m(m-1)/2$ denote the number of terms in (3.1), with the first m terms related to $e_{ts}, s = 1, \dots, m$, and the remaining $m(m-1)/2$ two-way interaction terms related to $e_{ts}e_{tr}, s < r$. We define a binary $(n \times p)$ matrix X to combine all p terms for all data. The first m columns are $X_{ti} = e_{ti}, i = 1, \dots, m$. The last $m(m-1)/2$ columns contain the interactions, with $X_{tj} = e_{ts}e_{tr}$, with $j = m+1, \dots, p$ indexing all possible pairs $s < r$. Let X_t denote the t -th row of X . Then the likelihood (3.1) of the autologistic model becomes

$$L(\beta) = \prod_t p(e_t|\beta) \propto \exp(X_t'\beta) \quad (4.2)$$

Comparing (4.2) and (4.1) we see that the autologistic model is a special case of a multinomial logistic model. The binary vector e_t codes a categorical outcome $y_t \in \{1, \dots, p\}$, and the covariate vector x_{tj} that is associated with the response for the t^{th} observation in (4.1) is equal to X_t . We will use $y_t \in \{0, \dots, 2^m\}$ to indicate the multinomial response e_t as an integer when it is technically more convenient to do so.

We now state the sufficient conditions for propriety. Let $C(i) = \{j | j \sim i\}$ be the set of nodes with which node i shares an edge in the graph $G = (V, E)$. We define a matrix X^* of all possible covariate vectors for a realization e of the autologistic model. Under the positivity constraint the number of possible covariate vectors is 2^m . Thus X^* is a $(2^m \times p)$ matrix. Note that X^* is different from the $(n \times p)$ matrix X , which is the matrix of covariate vectors for the actually observed instances $e_t, t = 1, \dots, n$. By construction of X^* , the first m columns span the set of all possible binary vectors e that have non-zero probability. Under the positivity constraint, these are all 2^m possible size m binary vectors. Also by construction, the matrix X^* satisfies $X_{tj}^* = X_{tr}^* X_{ts}^*$ and corresponding columns r and s .

Theorem 1. Assume that the following conditions hold

1. For any i there exists a row vector $\mu \in X$ such that $\mu_i = 1$ and $\mu_k = 0 \forall k \in C(i)$
2. For any i there exists a row vector $\nu \in X$ such that $\nu_i = 0$ and $\nu_k = 0 \forall k \in C(i)$

3. For a pair i and j , ($i \sim j$), there exists a row vector $\mathbf{z} \in \mathbf{X}$ such that $z_i = z_j = 1$ and either

$$(a) z_l = 0 \quad \forall l \in C(i), l \neq j, \text{ or}$$

$$(b) z_l = 0 \quad \forall l \in C(j), l \neq i;$$

and a vector ω such that either

$$(c) \omega_i = 0, \omega_j = 1 \text{ and } \omega_l = 0 \quad \forall l \in C(i); \text{ or}$$

$$(d) \omega_j = 0, \omega_i = 1 \text{ and } \omega_l = 0 \quad \forall l \in C(j).$$

Under the above conditions, the posterior distribution $p(\beta | Y)$ is proper based on the autologistic model (3.1) and an improper prior $p(\beta) \propto 1$.

We shall generically denote the t^{th} row vector of $\mathbf{X}$ as $\mathbf{X}_t$ and the j^{th} row vector of $\mathbf{X}^*$ as $\mathbf{X}_j^*$. We precede the main proof with two lemmas.

Lemma 2. Under the conditions of Theorem 1, any $\beta \in \mathbb{R}^p$ can be written as

$$\beta = \sum_{t, l | X_t \neq X_l^*} a_t (\mathbf{X}_t - \mathbf{X}_l^*)$$

where $a_t \geq 0$ and $\sum a_t > 0$.

Proof of Lemma 2. We define as set of basis vectors the p -tuples $f_h \in \{0, 1\}^p$ with 1 in the h^{th} position and 0 otherwise. It is sufficient to prove that any f_h can be written as

$$f_h = \begin{cases} \mathbf{X}_t - \mathbf{X}_l^*, & \mathbf{X}_l^* \neq \mathbf{X}_t \\ -(\mathbf{X}_j - \mathbf{X}_k^*), & \mathbf{X}_k^* \neq \mathbf{X}_j. \end{cases} \quad (4.3)$$

for some $1 \leq t, j \leq n$ and $1 \leq l, k \leq 2^m$.

First, consider any $h \leq m$. From Conditions 1 and 2, we can choose a $\mu \in \mathbf{X}$ and $\nu \in \mathbf{X}$ such that

$$\mu_h = 1 \quad \text{and} \quad \mu_k = 0 \quad \forall k \in C(h)$$

and

$$\nu_h = 0 \text{ and } \nu_k = 0 \quad \forall k \in C(h)$$

To prove (4.3) first, let $\mathbf{z} = T(\mu, h)$. Clearly $f_h = (\mathbf{X}_t - \mathbf{z})$. By Lemma 3 and the construction of μ , $\mathbf{z}$ is a row vector in $\mathbf{X}^*$ for some row index l . Thus $\mathbf{z} = \mathbf{X}_l^* - \mathbf{X}_t$ and $f_h = (\mathbf{X}_t - \mathbf{X}_l^*)$. A similar argument using Condition 2 proves that $f_h = -(\mathbf{X}_j - \mathbf{X}_k^*)$ for some j and $\mathbf{X}_k^* - \mathbf{X}_j$.

Now consider any $h > m$. We shall show that there exist indices p, q, k, l, m, n, a, b such that f_h can be written as

$$f_h = \begin{cases} (X_p - X^*_l) + (X_q - X^*_k) & \text{and } X^*_l \neq X_p, X^*_k \neq X_q. \\ -(X_m - X^*_a) - (X_n - X^*_b) & \text{and } X^*_a \neq X_m, X^*_b \neq X_n \end{cases} \quad (4.4)$$

Now let i and j be the indices whose interaction terms are indexed by h . From Condition 3 there exists a row vector $\mathbf{z} \in \mathbf{X}$ such that $z_i = 1, z_j = 1$ and $z_d = 0 \forall d \in C(i), d \neq j$, without any loss of generality. Let $\mathbf{w} = T(\mathbf{z}, i)$. By Lemma 3, $\mathbf{w}$ is a row vector in $\mathbf{X}^*$. Clearly

$$f_h + f_i = z - w$$

From the proof of $h < m$ (which uses Condition 1) there exists $\mathbf{v} \in \mathbf{X}$ and a row $\mathbf{X}^*_l$ of $\mathbf{X}^*$ with $\mathbf{X}^*_l \neq \mathbf{v}$ such that $-f_i = (\mathbf{v} - \mathbf{X}^*_l)$. Then

$$f_h = f_h + f_i - f_i = (z - w) + (v - X^*_l)$$

Using the remainder of Condition 3, and Condition 2, we get the second representation.

Lemma 3. Let $T(\mathbf{x}, h)$ be a function $(\{0, 1\}^p \times \mathbb{N}) \rightarrow \{0, 1\}^p, 1 \leq h \leq m$ such that $T(\mathbf{x}, h)_j = x_j \forall j \neq h$ and $T(\mathbf{x}, h)_h = 1 - x_h \forall j = h$. i.e., the transformation T only inverts the h^{th} element of $\mathbf{x}$ while preserving the rest. If $x_g = 0 \forall g \in C(h)$, and $\mathbf{x}$ is a row vector $\in \mathbf{X}^*$, then $T(\mathbf{x}, h) \in \mathbf{X}^*$.

The proof of Lemma 3 is straightforward.

Proof of Theorem 1. Recall that $y_t \in \{0, \dots, 2^m\}$ is $\mathbf{e}_t$ coded as an integer. Exploiting the equivalence with the multinomial logistic model we can use the result of Maddala (1983) to state that for any β

$$L(\beta) = \prod_t p(\mathbf{e}_t | \beta) = \prod_t p \left[\epsilon_{tl} - \epsilon_{tyt} \leq (X_t - X^*_l)' \beta \forall l \neq y_t \right] \leq \prod_t p \left\{ \min_{l | X^*_l \neq X_t} (\epsilon_{tl} - \epsilon_{tyt}) \leq \min_{l \neq y_t} (X_t - X^*_l)' \beta \right\} = \prod_t F \left[\min_{l | X^*_l \neq X_t} (X_t - X^*_l)' \beta \right]$$

where $F(\cdot)$ is the distribution function of $\min_{l=1}^L (E_{tl} - E_{t1})$. Let f denote the corresponding density function. Since $F \in [0, 1]$ and monotone increasing,

$$L(\beta) \leq \min_t F \left(\min_{l | X^*_l \neq X_t} (X_t - X^*_l)' \beta \right) \leq F \left(\min_{l,t} |X^*_l \neq X_t, 1 \leq t \leq n \{X_t - X^*_l'\beta\} \right)$$

Under the conditions of the theorem, by Lemma 2 any $\beta \in \mathbb{R}^p$ can be written as

$$\sum_{t,l|X_l^* \neq X_t} a_t (X_t - X_l^*) \quad a_t \geq 0 \quad \text{where } a_t \text{ are elements of a non-zero vector } \mathbf{a}.$$

From Schechter (2002), if for all $\beta \in \mathbb{R}^p$ there exists a non-zero vector $\mathbf{a}$ such that $a_t \geq 0$ and

$$\beta = \sum_t a_t z_t \text{ for a set of } z_t \in \mathbb{R}^p, t = 1 \dots n, \text{ then there must exist a positive } K \text{ satisfying } \min_t z_t' \beta < -K \|\beta\|. \text{ Let } \min_{t,l|X_l^* \neq X_t, 1 \leq t \leq n} (X_t - X_l^*)' \beta = d(X^*, y, \beta). \text{ We get } d(X^*, y, \beta) < -K \|\beta\|. \text{ Since } F \text{ is monotone, we have thus } L(\beta) < F(-K \|\beta\|) \text{ and}$$

$$\int L(\beta) d\beta < \int_{-\infty}^{\infty} F(-K \|\beta\|) \leq K \int_0^{\infty} \theta^{m-1} F(\theta) d\theta = K \int_0^{\infty} \theta^{m-1} \int_0^{\theta} f(u) du d\theta$$

Using Fubini's theorem, the integral reduces to

$$K \int_{-\infty}^0 \left[\int_0^{-u} \theta^{m-1} d\theta \right] f(u) du \leq K \int u^m f(u) d(u)$$

Since the density f admits finite moments, we have $\int L(\beta) d\beta < \infty$.

All the conditions of Theorem 1 suggest a basic feature. The main effects and the interaction parameters of the autologistic model appear at least once in the likelihood. Without this, trivially a flat prior for a parameter (not appearing in the likelihood equation) would yield an improper posterior. The first two conditions further imply that a main effect must appear at least once in the data, not entangled with other effects with which it has interaction. Intuitively, this is related to the estimability and identifiability of these main effects. Similarly the third condition implies that each two-way interaction could be extracted from the data and is not confounded in the likelihood with other interaction parameters. Direct relationships between identifiability, estimability and propriety exist in the context of GLMs. For autologistic models, such relationships are not so easy to arrive at, because the likelihood is complicated by the presence of normalizing constants. However, the sufficient conditions suggest a data structure that intuitively accommodates some estimability features.

5 Conditioning on a Graph

The statement of the autologistic model in Section 3 is conditional on a known graph G , i.e., we focus on the situation where the dependence structure is known *a priori*. Here the goal is to estimate β conditional on the graph, and infer the strength and direction of associations between pairs of nodes. In general, however, the graph itself is an unknown quantity and estimated along with β .

However, when the number of nodes is small, formal inference on G can be avoided by assuming a fully connected graph as the underlying structure and estimating the posterior distribution of β under that assumption. In this framework, no element of β has a non-zero prior mass at 0. Inference is obtained by examining the marginal posterior distribution of each β_{ij} . If the central 95% posterior credible interval excludes 0, we conclude that the

association is significant. Finally we summarize the inference through a graph where an edge is drawn between a pair of nodes satisfying the above criterion. This graph simply represents a summary of the posterior inference on β . This summary is not formal posterior inference on the graph under a graphical model which includes the graphs as unknown parameters, but serves a similar purpose.

For higher dimensional distributions, this approach to inference, though feasible, involves prohibitive computational effort. A graphical model significantly reduces the computation by limiting the number of all non-zero β_{ij} . This benefit in computational cost is counterbalanced by some difficulties in determining the prior distribution of non zero β_{ij} , i.e., a prior model $p(G)$. An improper prior for the coefficients β in a graphical model G would almost surely lead to an improper posterior when G is random itself. Essentially, the problem is that the marginal prior for a graph G can be written as the marginal integral of the $p(\beta|G)$, marginalized with respect to β . If $p(\beta|G)$ is improper, then the integral becomes meaningless.

6 Posterior Inference

We use MCMC simulation to carry out posterior inference. Most transition probabilities are straightforward. Only updating β is non-trivial because its complete conditional posterior distribution required evaluation of the normalization constant that appears in (3.1) given by

$$c(\beta) = 1/p(0|\beta) = \sum_e \exp \left\{ \sum_i \beta_i e_i + \sum_{i < j} \beta_{ij} e_i e_j \right\} \quad (6.1)$$

Since G is assumed to be given, we drop the conditioning on G for notational ease. Let

$K(e_t; \beta) = \exp \left\{ \sum_i \beta_i e_{ti} + \sum_{i < j} \beta_{ij} e_{ti} e_{tj} \right\}$ denote the exponential factor in (3.1), i.e., the un-normalized autologistic conditional probability for e_t . The un-normalized joint

probability model for e is thus given as $K(e; \beta) = \prod_t K(e_t; \beta)$. The constant $c(\beta)$ is computationally intractable for the massively repeated evaluation that is needed in MCMC simulation. The problem of evaluating the normalization constant in the autologistic model is known to be a challenging computational problem. See, for example, Welsh (1990) for a detailed discussion. We implemented an importance sampling estimate to tackle this problem. The main idea is that at each iteration step we sample k binary vectors $v_i \sim p(e = v_i | \beta)$ from the autologistic prior model. We then propose a new $\tilde{\beta}$ by generating from

$\tilde{\beta}_i \sim q(\tilde{\beta}_i; \beta_i) = N(\beta_i, c)$. The sample average $\tilde{R} = 1/k \sum_{i=1}^k K(v_i; \tilde{\beta}) / K(v_i; \beta)$ can be shown

to be a Monte Carlo estimate for the ratio $R = c(\tilde{\beta}) / c(\beta)$. We use $\tilde{R}$ to evaluate the Metropolis-Hastings acceptance probability for the proposal $\tilde{\beta}$.

The transition kernels for the sampling model parameters involve simple MH steps and are easy to sample from.

7 Results

We used the autologistic setup to model the dependence between the unobserved binary signals of HM presence. We assumed a sampling model consisting of a mixture of Poisson and log-normals to model the observed count data conditioned on the latent signal. For purposes of this analysis, we randomly sampled 400 genomic locations from the preprocessed dataset. The Markov chain included an initial burn-in of 4500 iterations. We used posterior samples from the subsequent 5000 iterations to evaluate the desired posterior summaries. The posterior estimates of the parameters β that index the auto-logistic model were obtained by averaging over the 5000 posterior samples, after discarding the initial 4500 iterations.

We summarized inference in a graph G by drawing a connection between pair of vertices i and j whenever the credible interval of β_{ij} excluded 0. We used 95% central credible intervals.

The estimated graph is shown in Figure 2a. The colors (and line types) of the edges were determined by $p(\beta_{ij} > 0 \mid \beta_{ij} = 0, \mathbf{Y})$. If the reported probability is greater than 1/2, then we say that the variables connected by the edge have a positive relationship. Otherwise we say that they have an inhibiting or negative relationship. We denote the positive and negative edges by solid lines and dotted lines, respectively. Existing literature lends support to the findings in Figure 2a. Yu et al. (2008) mention the central roles of the HMs, H3K4me3 and H3K27me3. There is mounting experimental evidence for significant interactions between these HMs. It is believed that they reinforce each other and jointly act on polycomb complexes. In embryonic stem cells, genes involved in cell differentiation are bivalently modified with both H3K4me3 and H3K27me3 (Bernstein, 2006). The solid edge in Figure 2a between the two HMs in the estimated graph supports the hypothesis of strong positive correlation between the two. There are three prominent triangles of HMs with positive association among themselves. The first group consists of H3K9me3, H3K27me3 and H3K36me3, the second group is formed by H3K9ac, H3K27ac and H3K4me3, and the third group comprises H4K20me1, H3K79me3 and H3BK5me1. The relationships between these groups of HMs have not been explored in depth. On the other hand, there is evidence to suggest that the HMs H3K27me3 and H3K27ac act antagonistically in transcription of polycomb target genes (Tie et al., 2009). In the figure, the dotted edge between these HMs implies a negative relationship between these variables.

Finally, we compared inference under the improper prior with inference under independent $N(0, 1)$ priors for the autologistic coefficients. We use the standard normal prior as a proxy for a technical convenience prior, as it is often used. Figure 2 compares the posterior networks under the two priors. Inference under the informative $N(0, 1)$ prior misses several edges that are reported under the improper prior. In fact, 28% of the edges are missing in (b), compared to (a). This is strong evidence for substantial differences in the posterior distributions under the improper and the informative model. When the informative prior was based on expert opinion, this discrepancy would be a desirable feature. However, the same differences raise concerns when the informative prior is only a technically convenient choice.

8 Discussion

We have given an easily verifiable condition for posterior propriety under an improper prior for the coefficients of an autologistic model. We believe that the autologistic model could be useful for statistical inference for many high throughput genomic data. For many genomic data, including gene expression data, protein activation data, histone modification and many others, the underlying biologic signal is best thought of as a binary indicator for presence versus absence, or mathematically almost equivalently, as a trinary indicator of over-, normal- and under-expression. In such problems it is natural to consider an autologistic model for the joint probability model across many genes (proteins, histone modifications etc.). There are several reasons why such models are not used more extensively in the current literature. In this discussion we addressed one reason, the need for improper priors.

References

- Bernardo J, Berger J, Dawid A, Smith A. Bernardo, et al. On the development of reference priors. *Bayesian Statistics*. 1992; 4:35–60.
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006; 125:315–326. [PubMed: 16630819]
- Besag J. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of Royal Statistical Society Series B*. 1974; 135:192–236.
- Chen M, Shao Q. Propriety of posterior distribution for dichotomous quantal response models. *Proceedings-American Mathematical Society*. 2001; 129(1):293–302.
- Heckerman D, Geiger D. Learning Gaussian networks. In *Proceedings of Tenth Conference on Uncertainty in Artificial Intelligence*, Seattle, WA. 1994:274–284.
- Heckerman D, Geiger D. Learning Bayesian networks: A unification for discrete and Gaussian domains. In *Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence*, Montreal, Quebec. 1995:293–301.
- Hobert J, Casella G. The effect of improper priors on gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*. 1996:1461–1473.
- Ibrahim J, Laud P. On bayesian analysis of generalized linear models using jeffreys's prior. *Journal of the American Statistical Association*. 1991:981–986.
- Maddala, G. Limited-dependent and qualitative variables in econometrics. Vol. 3. Cambridge Univ Pr; 1983.
- Mitra R, Müller P, Liang S, Yue L, Ji Y. A bayesian graphical model for chip-seq data on histone modifications. *Journal of American Statistical Association, Applications and Case Studies*. 2011 To appear.
- Natarajan R, McCulloch C. A note on the existence of the posterior distribution for a class of mixed models for binomial responses. *Biometrika*. 1995; 82(3):639–643.
- Schechter, M. Principles of functional analysis. Vol. 36. Amer Mathematical Society; 2002.
- Tie F, Banerjee R, Stratton CA, Prasad-Sinha J, Stepanik V, Zlobin A, Diaz MO, Scacheri PC, Harte PJ. CBP-mediated acetylation of histone H3 lysine 27 antagonizes *Drosophila* Polycomb silencing. *Development*. 2009; 136:3131–3141. [PubMed: 19700617]
- Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet*. 2008; 40:897–903. [PubMed: 18552846]
- Welsh, DJA. In *Disorder in Physical Systems*. Clarendon Press; 1990. The computational complexity of some classical problems from statistical physics; p. 307-321.

- Yu H, Zhu S, Zhou B, Xue H, Han JD. Inferring causal relationships among different histone modifications and gene expression. *Genome Res.* 2008; 18:1314–1324. [PubMed: 18562678]
- Zellner A, Rossi P. Bayesian analysis of dichotomous quantal response models. *Journal of Econometrics.* 1984; 25(3):365–393.
- Zhang Y, Lv J, Liu H, Zhu J, Su J, Wu Q, Qi Y, Wang F, Li X. HHMD: the human histone modification database. *Nucleic Acids Res.* 2010; 38:D149–154. [PubMed: 19892823]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Conditional independence structure for the $m = 11$ HMs. The graph shows inference under a Gaussian graphical model as implemented in the R package *deal*.

Figure 2.

The estimated posterior graph. Panel (a) shows inference under the improper prior $p(\beta) \propto c$. Edges are included in the graph when a central 95% credible interval excludes 0. The dotted lines represent the HM pairs with negative relationships, while the solid lines depict the positive ones. The thickness of the edges indicate the strength of the relationship, formalized as $E(\beta_{ij} | Y)$. The numbers in the nodes refer to H3K27AC(1), H3K36AC(2), H3K4AC(3), H3K9AC(4), H3BK5ME1(5), H3K27ME3(6), H3K36ME3(7), H3K4ME3(8), H3K79ME3(9), H3K9ME3(10), and H4K20ME1(11). For comparison, panel (b) shows the same inference under a proper prior $p(\beta_{ij}) = N(0, 1)$. Some edges are dropped under the $N(0, 1)$ prior which shrinks inference on β_{ij} to 0.