Analyzing Serially Correlated Binary Data Using Partially Exchangeable Models

Xin Dang\textsuperscript{1,4}  
University of Mississippi  
Latonya C. Garner\textsuperscript{2}  
Mississippi Valley State University  
Hanxiang Peng\textsuperscript{*3,4}  
Indiana University-Purdue University at Indianapolis  

\textsuperscript{*} Corresponding author: Hanxiang Peng (hpeng@math.iupui.edu)

Run title: Analyzing Partially Exchangeable Data

\textsuperscript{1} Department of Mathematics, University of Mississippi, University, MS 38677, USA. Telephone: (662)915-7409. Fax: (662)915-2361. Email: xdang@olemiss.edu. URL: www.olemiss.edu/~xdang.

\textsuperscript{2} Department of Mathematics, Computer and Information Sciences, Mississippi Valley State University, 14000 HWY 82 West #7257, Itta Bena, MS 38941, USA. Email: lgarner@mvsu.edu

\textsuperscript{3} Department of Mathematical Sciences, Indiana University-Purdue University at Indianapolis, Indianapolis, IN 46202-3216, USA. Telephone: (317)274-8070. Fax: (317)274-3460. Email: hpeng@math.iupui.edu. URL: https://math.iupui.edu/~hpeng/.

\textsuperscript{4} This research is supported by the US National Science Foundation under Grant No. DMS-0940365.
Abstract

In this article, we propose a procedure for modeling serially correlated binary data based on the concept of partial exchangeability. This can be viewed as a model of the mixture of Markov chains. We introduce rectangular completely monotone (RCM) links to characterize partially exchangeable distributions. We present several methods to construct mixtures of Markov chains and show that the mixtures as a class is closed under convex linear combinations, products, and composites. We also introduce incomplete RCM links and demonstrate that the resulting distributions contain the Binomial distribution as a special case. We discuss model fitting, regression, parameter estimation, and model selection. Our procedure possesses the advantages of mathematical simplicity and computational feasibility. A small simulation is run to validate that the Binomial distribution is a special case of the partially exchangeable distribution. Finally, the proposed procedure is applied to analyze two real datasets from dairy and medical sciences, and the results are superior to some existing procedures.

Key words and phrases: complete monotonicity, correlated binary sequence, link, Markov chain mixture, partial exchangeability
1 Introduction

Binary responses commonly arise in many fields of science, including tumor recurrences in medical science, baseball hits in sports, occupational status in labor market surveys, presence of a pathogen in dairy science, and others, such as educational testing, molecular biology, animal breeding, and meteorology to name a few. Often each individual binary sequence exhibits serial dependence. One example that we will look into in this article is a bladder cancer study in medical science where each of a group of patients has multiple recurrences. Another example that we will analyze is from dairy science where each of a number of cows is examined for the presence of a pathogen infection in a given lactation cycle. When formulating models of such data, it is natural to allow serial dependence.

The literature describing the analysis of correlated binary data is extensive (see, e.g., George and Bowman (1995) [13]; Yu and Zelterman (2002) [30]; Stefanescu and Turnbull (2003) [25]; Xu and Prorok (2003) [28]; Kuk (2004) [16]; Liu and Agresti (2005) [17]; Garner (2007) [12]; Yu and Zelterman (2008) [29]; Fei Tan, et al. (2010) [27]). Different approaches measure different effects of covariates and account for dependence structure in different ways. In this article, we will exploit the perception of partial exchangeability to investigate multiple serially dependent binary responses. Introduced by de Finetti in 1938, partial exchangeability was studied by many authors, including Diaconis (1988) [5], Diaconis and Freedman (1980a, b, c) [6, 7, 8], Freedman (1962a, b) [9, 10], and Zaman (1984) [31]. A sequence of binary random variables is partially exchangeable (PE) if the joint distribution of any sub-sequence of the first finite terms is invariant under permutations that keep the initial state and the transitions from $i$ to $j$ for $i = 0, 1$ unchanged. A more stringent assumption is exchangeability which requires invariance under any permutation. Also introduced by de Finetti in the 1930’s and intensely studied over the past century,
exchangeability is meant to capture the notion of symmetry in a collection of random variables and often used as an alternative to independence. Partial exchangeability captures symmetry more generally and grasps dependence in a collection of random variables and can be utilized as a candidate for dependence.

Dang, et al. (2009) [3] proposed a unified approach for analyzing exchangeable binary data and applied it to developmental toxicity studies. We further develop their approach and propose a procedure to analyze partially exchangeable binary data. A model in which an individual binary sequence is assumed to be recurrent and partially exchangeable is, according to Diaconis and Freedman (1980a, c) [6, 8], a mixture of Markov chains; whereas the latter, as pointed out by Quintana and Newton (1998) [22], can be viewed as a random-effects model in which the unobservable latent transition matrices follow some unknown distribution. Markov chains and their mixtures are widely used to model serially correlated data for their structural simplicity and easy interpretation. A parametric mixture of Markov chains can be obtained via choosing the mixing distribution to be a parametric distribution. (See e.g. Neal (2000) [19] in which the Dirichlet distribution was chosen as the mixing distribution.) A mixing distribution can be viewed as a Bayesian prior. (See e.g. a Bayesian approach based on the method of Gibbs sampling by Chib (1996) [2].) Nevertheless, there are two main drawbacks for some existing Markov mixture models. One is that the likelihood function of the mixture is not available in a simple form. This makes it infeasible for the statistical inference that is based on the method of maximum likelihood estimation. Frühwirth-Schnatter (2001) [11] used the Markov Chain Monte Carlo method for estimation of mixture models. The other drawback is the computational complexity. It often requires a difficult numerical integration to be performed. Our proposed partially exchangeable model intends to avoid these drawbacks. For instance, the probability mass functions have mathematical simplicity
and computational ease, the likelihood functions can be computed without numerical integration, and parameter estimation can be conveniently obtained by maximum likelihood using conventional routines in packages, such as R or SAS.

By exploiting partial exchangeability, Quintana and Newton (1998) [22] investigated the assessment of the order of serial dependence for multiple binary responses. Relating to the Pólya sequence from a one-urn model, Quintana and Newton (1999) [23] constructed a two-urn model for the mixtures of Markov chains, providing interesting and useful interpretation. Their two-urn model (see their equation (2.4)) is a special case of our proposed partially exchangeable model. By choosing different RCM link functions, our proposed PE model also includes the Markov models of Muenz and Rubinstein (1985) [18], the random-effects model of Stiratelli, Laird and Ware (1984) [26], and the logistic regression model of Cessie and Houwelingen (1994) [1].

The article is organized as follows: In Section 2, we introduce parsimonious partially exchangeable models via RCM links. We provide constructive methods to obtain RCM links and demonstrate that the class of models is closed under convex linear combinations, products and composites. We also show that RCM link can be obtained from existing univariate completely monotone links. Incomplete RCM links are introduced. Section 3 is devoted to statistical inference and model fitting, regression analysis and model selection. We apply our procedure to two real data sets in Section 4. Section 5 includes the conclusion and further research topics. Technical details can be found in the Appendix.

2 Parsimonious PE Distributions

In this Section, we first briefly review the results about partially exchangeable (PE) distributions given by Peng, et al. (2010) [21]. Then we introduce parsimonious PE
distributions and rectangular completely monotone (RCM) links, and discuss methods about obtaining such distributions and links. Finally, we introduce incomplete RCM links.

2.1 The PE Distribution and CM Links

As elaborated by Diaconis and Freedman (1980a) [6], a sequence of binary random variables (r.v.’s) $B_0, B_1, \ldots$ is partially exchangeable if for any two binary sequences $\{b_i\}$ and $\{c_i\}$ which possess identical initial states and identical transition counts,

$$P(B_0 = b_0, B_1 = b_1, \ldots, B_n = b_n) = P(B_0 = c_0, B_1 = c_1, \ldots, B_n = c_n).$$

Recall that a sequence of r.v.’s $B_0, B_1, \ldots$ is recurrent if

$$P(B_n = B_0 \text{ for infinitely many } n) = 1.$$

Peng, et al. (2010) [21] derived the joint distribution of partially exchangeable binary r.v.’s. Here we re-state their result below.

**Theorem 1** Suppose $B_0, B_1, B_2, \ldots$ is a sequence of binary random variables which are recurrent and partially exchangeable. Then

$$P(B_0 = b_0, B_1 = b_1, \ldots, B_n = b_n) = \sum_{i=0}^{t_{01}} \sum_{j=0}^{t_{10}} (-1)^{i+j} \binom{t_{01}}{i} \binom{t_{10}}{j} \lambda_{i,j}^{(b_0)},$$

where $b_0, b_1, \ldots, b_n$ take values in $\{0, 1\}$, $t_{ij}$ is the number of transitions in $B_0, B_1, \ldots, B_n$ from $i$ to $j$, and $\lambda_{i,j}^{(b_0)}$ is the joint probability that the transitions $0 \rightarrow 0$ and $1 \rightarrow 1$ in $B_0, B_1, \ldots, B_n$ simultaneously occur $i$ and $j$ times with initial state $b_0$, or symbolically,

$$\lambda_{i,j}^{(b_0)} = P(B_0 = b_0, (0 \rightarrow 0)^i, (1 \rightarrow 1)^j), \quad 0 \leq i \leq t_{00} + t_{01}, \quad 0 \leq j \leq t_{11} + t_{10}.$$

Let us write $\mathbf{\lambda} = \{\lambda_{i,j}^{(b_0)} : b_0 = 0, 1, \ i = 0, \ldots, t_{00} + t_{01}, \ j = 0, \ldots, t_{11} + t_{10}\}$, and denote $\text{PE}(\mathbf{\lambda})$ the above PE distribution. Clearly $\text{PE}(\mathbf{\lambda})$ is a parametric distribution with a sequence of parameters $\lambda$ in which the number of parameters grows with the
length of the binary sequence. These $\lambda_{ij}^{(b_0)}$’s are referred to as *marginal probabilities.* Notice that the overall transition probabilities $P_{00}$ and $P_{11}$ from zero to zero and from one to one respectively are useful quantities. They can be expressed by

$$ P_{00} = \lambda_{1,0}^{(0)} + \lambda_{1,0}^{(1)}, \quad P_{11} = \lambda_{0,1}^{(0)} + \lambda_{0,1}^{(1)}. \quad (2) $$

The observed values of $P_{00}, P_{11}$ can be computed from the transition counts of the data by

$$ p_{00} = t_{00} / (t_{00} + t_{01}), \quad p_{11} = t_{11} / (t_{11} + t_{10}). $$

These quantities can be used for model comparison; a better model gives closer estimated overall transition probabilities to these observed values.

It is interesting to observe that the sequence of marginal probabilities $\lambda = \{\lambda_{j,k}^{(b_0)}\}$ is comprised of two sequences $\{\lambda_{j,k}^{(0)}\}$ and $\{\lambda_{j,k}^{(1)}\}$ that are *rectangular completely monotone (RCM)* satisfying $\lambda_{0,0}^{(0)} + \lambda_{0,0}^{(1)} = 1$. That is, for $r_1, r_2 \geq 0$ and $\lambda_{0,0}^{(0)} + \lambda_{0,0}^{(1)} = 1$,

$$ (-1)^{r_1 + r_2} \Delta_1^{r_1} \Delta_2^{r_2} \lambda_{j,k}^{(b_0)} \geq 0, \quad 0 \leq j + r_1 \leq t_{00} + t_{01}, \quad 0 \leq k + r_2 \leq t_{11} + t_{10}, \quad (3) $$

where $\Delta_1, \Delta_2$ are the (univariate) marginal difference operators, i.e., $\Delta_1 a_{j,k} = a_{j+1,k} - a_{j,k}$ and $\Delta_2 a_{j,k} = a_{j,k+1} - a_{j,k}$ with $\Delta_1^2 = \Delta_1 (\Delta_1)$ and $\Delta_0^2 = \Delta_2 = I$ the identity operator. A useful formula for verifying RCM is

$$ \Delta_1^r \Delta_2^s a_{u,v} = \sum_{j=0}^{r} \sum_{k=0}^{s} \binom{r}{j} \binom{s}{k} (-1)^{r+s-j-k} a_{u+j,v+k} \quad r, s, u, v = 0, 1, \ldots, \quad (4) $$

where $\{a_{j,k} : j, k = 0, 1, \ldots\}$ is a two-dimensional sequence of reals.

One can easily check that the matrix $t = \{t_{ij} : i, j = 0, 1\}$ of transition counts and the initial state $b_0$ are sufficient statistics for the PE distribution. Let $p(t, b_0; \lambda)$ be the probability that $B_0, B_1, \ldots, B_n$ have a common transition matrix $t$ with initial state $b_0$. Then by Theorem 1 we derive

$$ p(t, b_0; \lambda) = \binom{n_0 - 1}{t_{00}} \binom{n_1 - 1}{t_{11}} \sum_{i=0}^{t_{01}} \sum_{j=0}^{t_{10}} (-1)^{i+j} \binom{t_{01}}{i} \binom{t_{10}}{j} \lambda_{t_{00}+i, t_{11}+j}^{(b_0)}, \quad (5) $$
where \( n_0 \) and \( n_1 \) are the number of zero’s and one’s respectively. Here we define \((-1)^i = 0\) for \( i \geq 0\). For more discussions about these results, please see [21].

Since \( \lambda_{0,0}^{(0)} + \lambda_{0,0}^{(1)} = 1 \), it follows that \( \text{PE}(\mathbf{\lambda}) \) has \( N = 2(t_{00} + t_{01} + 1)(t_{11} + t_{10} + 1) - 1 \) independent parameters and is the saturated model with parameter space \( \Lambda = \{ \mathbf{\lambda} \in \mathbb{R}^N : \mathbf{\lambda} \text{ satisfies (3)} \} \). Due to its high number of parameters and the complicated structure of the parameter space, the saturated model is seldom useful in fitting real data. To tackle this difficulty, we seek dimension reduction. Specifically, we map \( \Lambda \) onto a lower \( d \)-dimensional space \( \Theta \subset \mathbb{R}^d \), where \( d << N \), so that we obtain a parsimonious model. To be more specific, we are in pursuit of a map \( h \) from \( \Theta \) into \( \Lambda \) so that \( \mathbf{\lambda} = h(\mathbf{\theta}) \in \Lambda \) for every \( \mathbf{\theta} \in \Theta \). This can be expressed in the usual form,

\[
\lambda_{j,k}^{(b_0)} = h_{b_0,j,k}(\mathbf{\theta}), \quad b_0 = 0, 1, \ 0 \leq j \leq t_{00} + t_{01}, \ 0 \leq k \leq t_{11} + t_{10}, \ \mathbf{\theta} \in \Theta, \quad (6)
\]

where \( h = \{ h_{b_0,j,k} : b_0 = 0, 1, \ 0 \leq j \leq t_{00} + t_{01}, \ 0 \leq k \leq t_{11} + t_{10} \} \). Substitution of this into the density (5) yields a parsimonious model with the density given by

\[
f(t, b_0; \mathbf{\theta}) = \binom{n_0 - 1}{t_{00}} \binom{n_1 - 1}{t_{11}} \sum_{i=0}^{t_{01}} \sum_{j=0}^{t_{10}} (-1)^{i+j} \binom{t_{01}}{i} \binom{t_{10}}{j} h_{b_0,t_{00}+i,t_{11}+j}(\mathbf{\theta}), \quad (7)
\]

where \( h_{0,0,0}(\mathbf{\theta}) + h_{1,0,0}(\mathbf{\theta}) = 1 \) for every \( \mathbf{\theta} \in \Theta \). In order for the above expression to be a legitimate probability mass function, the two sequences \( \{ h_{b_0,j,k}(\mathbf{\theta}) \} \) for \( b_0 = 0, 1 \) must be RCM respectively, i.e., they must satisfy (3) with \( h_{0,0,0}(\mathbf{\theta}) + h_{1,0,0}(\mathbf{\theta}) = 1 \) for every \( \mathbf{\theta} \in \Theta \). It is not difficult to show that (7) indeed constitutes a probability mass function for every \( h \) which satisfies (3). Such an \( h \) is referred to as a rectangular completely monotone (inverse) link.

Note that every RCM link \( h \) with its natural domain as \( \Theta \) corresponds to a parsimonious model of the saturated model \( \text{PE}(\mathbf{\lambda}) \). Let us denote this submodel by \( \text{PE}(h) \). If \( B_0 \) is degenerate at 1 (i.e. \( \mathbb{P}(B_0 = 1) = 1 \)), then \( \lambda_{0,0}^{(1)} = 1 \) and \( \lambda_{i,j}^{(0)} = 0 \) for all \( i, j \). In applications, this corresponds to the case that all observations start with the same initial state 1. An example of this is the Bladder Cancer study in which
each patient had the tumor when entering the study. Even though degeneracy is a
special case, it is indeed quite general as we explain below. Toward this, conditional
on the initial state $b_0$, we have two independent RCM sequences $\{\kappa_{j,k}^{(b_0)} = \lambda_{j,k}^{(b_0)} / \lambda_{0,0}^{(b_0)}\} : b_0 = 0, 1$, both satisfying $\kappa_{0,0}^{(b_0)} = 1$. For such two RCM sequences, two independent
links $\{g_{j,k}(\vartheta|b_0)\} : \vartheta \in \Theta'$ for $b_0 = 0, 1$ are required to obtain dimension reduction,
either of which must satisfy (3) for $b_0 = 0, 1$ respectively and $\vartheta \in \Theta'$. As a result, in
applications we shall seek RCM sequences $g = \{g_{j,k}(\vartheta|b_0)\}$ satisfying $g_{0,0}(\vartheta|b_0) = 1$
for $b_0 = 0, 1$. This is possible in view of Remark 1. With such RCM $\{g_{j,k}(\vartheta|b_0)\}$, one
immediately obtains unconditional RCM $h = \{h_{b_0,j,k}(\theta)\}$ by setting

$$h_{0,j,k}(\theta) = \alpha g_{j,k}(\vartheta|b_0 = 0), \quad h_{1,j,k}(\theta) = (1 - \alpha) g_{j,k}(\vartheta|b_0 = 1), \quad \theta = (\alpha, \vartheta) \in \Theta,$$

where $\lambda_{0,0}^{(b)} = \alpha$ for some parameter $\alpha \in [0, 1]$. One can easily show that the maximum
likelihood estimator of $\alpha$ is $\hat{\alpha} = n_0/(n + 1)$, the proportion of the number of zero’s
in the binary sequence. (See details in Section 3.) If the transition probabilities do
not depend upon the initial state, i.e. $g_{i,j}(\vartheta|b_0 = 0) = g_{i,j}(\vartheta|b_0 = 1)$, then one only
has to get one RCM link. In application, one may first perform a likelihood ratio
test about whether the transition probabilities are independent of the initial states
so that model complexity can be reduced. We perform this test for the Cow data in
our application and conclude the independence.

2.2 RCM Links

Here we give several useful RCM links. Constructive methods and justifications of
these links are given in the Appendix.

We begin with perhaps the easiest link first, and its RCM can be proved by (4).
Note here $s, t \geq 0$.

**Independence Link**: $h(s, t; \theta) = \theta^s (1 - \theta)^t, \theta \in (0, 1)$,
Bivariate Gamma Link (Biga): Let $\theta = (\theta_1, \theta_2, \theta_3, \rho)$,

$$h(s, t; \theta) = \frac{(1 + s)^{\theta_3 - \theta_1}(1 + t)^{\theta_3 - \theta_2}}{(1 + s + t + \rho st)^{\theta_3}}, \quad \theta \in [0, \infty)^3 \times [0, 1].$$  \hspace{1cm} (9)

When $\rho = 1$, this reduces to the product of two Gamma links given in Table 1 with power $\nu = 1$.

Kibble’s Bivariate Gamma Link (Kbiga): Let $\theta = (\theta_1, \theta_2, \rho, v)$,

$$h(s, t; \theta) = \frac{(\theta_1 \theta_2)^v}{((\theta_1 + s)(\theta_2 + t) - \rho st)^v}, \quad \theta \in (0, \infty)^2 \times (0, 1) \times (0, \infty).$$  \hspace{1cm} (10)

When $\rho = 0$ and $v = 1$, this reduces to the product of two MM links given in Table 1 with power $\nu = 1$.

Bivariate Beta Link: Let $\theta = (a, b, c)$ and denote by $3F_2$ the hypergeometric function,

$$h(s, t; \theta) = 3F_2(a + s, b + t, d; d + s, d + t; 1), \quad \theta = (a, b, c) \in (0, \infty)^3.$$  \hspace{1cm} (11)

where $d = a + b + c$. This can easily be implemented with S-plus or R softwares.

The following is a useful result which is easy to verify.

Remark 1 For a RCM $\bar{h}$, let $h(s, t) = \bar{h}(s_0 + s, t_0 + t)/\bar{h}(s_0, t_0)$. Then $\{h(s, t) : s, t \geq 0\}$ is RCM with $h(0, 0) = 1$.

We can create new RCM links from existing links based on the following theorem with the proof delayed to the Appendix.

Theorem 2 (LP) If $\{g_{i,j}\}$ and $\{h_{i,j}\}$ are RCM, so are the convex linear combinations $\{\delta g_{i,j} + (1 - \delta)h_{i,j}\}$ for $\delta \in [0, 1]$ and the product $\{g_{i,j} \ast h_{i,j}\}$.

(C1) If $\{h_{i,j}\}$ is RCM and $\{\psi_1(s) : s \geq 0\}$ and $\{\psi_2(t) : t \geq 0\}$ are positive functions with completely monotone derivatives, then the composite $\{h_{\psi_1(i),\psi_2(j)}\}$ is RCM.

(C2) Suppose $\varphi(u, v) = \sum_{j,k} \theta_{j,k} u^j v^k$ is a non-negative polynomial in $u, v$, i.e., all the coefficients are non-negative $\theta_{j,k} \geq 0$. If $\{\phi_1(s, t) : s, t \geq 0\}$ and $\{\phi_2(s, t) : s, t \geq 0\}$ are RCM, then the composite $\{\varphi(\phi_1(i, j), \phi_2(i, j))\}$ is RCM.
Table 1: **UCM Links**, $\nu \in [0, 1]$ and $N = \{0, 1, \ldots\}$.

<table>
<thead>
<tr>
<th>Name</th>
<th>Link</th>
<th>Parameter Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power-Power(Kuk’s)</td>
<td>$\theta^\nu$</td>
<td>$\theta \in (0, 1)$</td>
</tr>
<tr>
<td>MM-Power</td>
<td>$\theta/(\theta + t^\nu)$</td>
<td>$\theta \in (0, \infty)$</td>
</tr>
<tr>
<td>Beta-Power</td>
<td>$B(\theta_1 + t^\nu, \theta_2)/B(\theta_1, \theta_2)$</td>
<td>$\theta \in (0, \infty)^2$</td>
</tr>
<tr>
<td>Gamma-Power</td>
<td>$(1 + \theta_1 t^\nu)^{-\theta_2}$</td>
<td>$\theta \in (0, \infty)^2$</td>
</tr>
<tr>
<td>Poisson-Power</td>
<td>$\exp(\theta(e^{-t^\nu} - 1))$</td>
<td>$\theta \in (0, \infty)$</td>
</tr>
<tr>
<td>Binomial-Power</td>
<td>$(pe^{-t^\nu} + 1 - p)^N$</td>
<td>$(p, N) \in [0,1] \times N$</td>
</tr>
<tr>
<td>Normal-Power</td>
<td>$2 \exp(\sigma^2 t^{2\nu}/2)(1 - \Phi(\sigma t^\nu))$</td>
<td>$\sigma^2 \in (0, \infty)$</td>
</tr>
</tbody>
</table>

**Remark 2**  
1. The logarithm function $\psi_1(t; \theta) = \theta \log(1 + t)$ with $\theta > 0$, the power function $\psi_2(t; \nu) = t^\nu$ with $0 \leq \nu \leq 1$, and $\psi_3(t) = 1 - \exp(-t)$ are *positive with RCM derivatives*.  
2. The exponential function $\varphi_2(t; \theta) = \theta^t$ with $\theta > 1$ are *absolutely monotone*.

Dang, *et al.* (2009) [3] provided numerous useful univariate completely monotone (UCM) links from which we can obtain RCM links based on the following theorem.

**Theorem 3** *If* $\{f_i\}$ *and* $\{g_j\}$ *are UCM, then* $\{h_{i,j} = f_{i+j}\}$ *and* $\{h_{i,j} = f_i * g_j\}$ *are RCM.*

Table 1 lists some of UCM links given in Dang, *et al.* (2009) [3]. Our applications below will focus on three one-parameter UCM links, i.e., the Gamma link $Ga_t(\theta) = 1/(1 + t)^\theta$, the MM link $MM_t(\theta) = \theta/(t + \theta)$, and the Independence link $Bin_t(\theta) = \theta^t$. If $\{f_i\}$ and $\{g_j\}$ are chosen to be the Beta-power link given in Table 1 with power $\nu = 1$, then the resulting parsimonious model $\{f_i * g_j\}$ is the two-urn model given by Quintana and Newton (1996) [23].

### 2.3 Incomplete RCM Links

In the Appendix, we have introduced incomplete RCM links. Parsimonious distributions resulting from these links contain extra parameters and offer additional flexibility; in particular, they contain the Binomial distribution as a special case in their
parameter domains, while existing distributions in the literature such as the Beta-binomial distribution do not possess this property in its parameter domain. More details are given in Theorem 5 in the Appendix and the discussions therein, and in Dang, et al. (2009) [3] where parsimonious distributions resulting from incomplete CM links also take the Binomial as its special case.

If \( \{ f_i \} \) and \( \{ g_j \} \) are chosen to be the univariate incomplete A-link given in Dang, et al. (2009) [3], then

\[
h_{s,t}(\theta_1, \theta_2, \varphi_1, \varphi_2) = \frac{\theta_1^{s} - \theta_2^{s}}{s \log(\theta_1/\theta_2)} \frac{\varphi_1^{t} - \varphi_2^{t}}{t \log(\varphi_1/\varphi_2)}
\]

is an incomplete RCM link. The parameter relationships \( \theta_1 = \theta_2 \) and \( \varphi_1 = \varphi_2 \) recover the Binomial distribution. We refer it to as the Incomplete RCM A-Link. It is used in the simulation study below to validate the inclusion of the Binomial distribution as a special case of the PE distribution.

3 Estimation and Inference

This section discusses model fitting, estimation, regression, and model selection.

Model Fitting Consider the parsimonious model PE(h) resulting from RCM link (8). Let \( (T_i, b_i) : i = 1, ..., n \) be independent observations of transition matrix and initial state \( (T, b_0) \). Based on the observations, we estimate the parameter \( \theta \) by the maximum likelihood estimator (MLE) \( \hat{\theta}_n \). In view of the discussion in Subsection 2.1, the log-likelihood of PE(h) can be calculated by

\[
L(\alpha, \theta) = n_0 \log \alpha + n_1 \log(1 - \alpha) + \sum_{i=1}^{n} \log f(T_i, \theta),
\]

where \( f(t, \theta) \) is defined analogously to (7) with \( h_{b_0,j,k} \) being replaced with \( g_{j,k}(\cdot | b_0) \) given in (8). Solving \( \partial L/\partial \alpha(\alpha, \theta) = 0 \) gives the MLE \( \hat{\alpha} = n_0/(n + 1) \) of \( \alpha \), (see (8) and the remarks therein). The MLE \( \hat{\theta}_n \) can be found as the solution to the score equation, \( S_n(\alpha, \theta) = \partial L/\partial (\alpha, \theta) = 0 \). Numerical solutions can be obtained by using
the Newton Raphson iteration or other optimization algorithms such as \textit{nlminb} in R. Under the usual regularity assumptions, the MLE $\hat{\theta}_n$ is asymptotically normal, so that the standard errors of the MLE can be computed from the usual Hessian matrix. These details can be found in a standard textbook describing MLE’s.

\textbf{Regression} Suppose to each observation $(T, b)$ from $\text{PE}(h)$, there is associated with a covariate $X$ such as dose, weight, etc. Let $(T_i, b_i, X_i), i = 1, ..., n$ be independent observations of $(T, b, X)$. Let us write $\theta = (\theta, \vartheta)$ where $\theta \in \mathbb{R}$ is a parameter of interest, while $\vartheta$ is treated as a nuisance parameter. Allowing $\theta$ to relate to the linear systematic part $\eta = \beta^\top X$ leads us to taking $\theta = \eta$, where $\beta$ is a regression parameter. A common but equivalent expression is

$$\lambda_{j,k}^{(b_0)} = h_{b_0,j,k}(\beta^\top X; \vartheta), \quad b_0 = 0, 1, \quad j, k = 0, 1, 2, ... \quad (14)$$

These equations suggest two extensions of the proposed partially exchangeable model from the Binomial model in generalized linear models, namely, the PE distribution extends the Binomial distribution and the RCM links extend the usual links.

Again as above, the parameters $\alpha, \beta, \vartheta$ can be estimated by maximum likelihood. Often $\theta$ is subject to a box constraint such as $\theta$ in $[0, 1]$ or $[0, \infty)$. For regression $\theta = \alpha + \beta X$, the box constraint forces the parameter $\alpha, \beta$ to be linearly constrained. One may use optimization algorithms such as \textit{constrOptim} in R or \textit{nlpnra} in SAS to find the numerical values of the MLE’s. Under the usual regularity conditions, these estimators will converge in distribution to a multinormal normal distribution. The significance of coefficient $\beta$ can be tested by hypotheses $H_0: \beta = 0$ vs $H_a: \beta \neq 0$ using either the Wald test or the likelihood ratio test (LRT).

\textbf{Remark 3} As Diaconis and Freedman (1980a) [6] pointed out, De Finetti’s theorem for Markov chains does not hold without the recurrence of the partially exchangeable sequence. For an finite exchangeable sequence, Diaconis and Freedman (1980c) [8]
showed that de Finetti’s theorem does not hold either, but it approximately holds if the finite sequence is the beginning of a long exchangeable sequence, see their Theorem 3. We conjecture that a similar approximation may hold for a finite partially exchangeable sequence. Thus, we view a finite sequence of observations as the beginning of a long partially exchangeable sequence, so that we can apply our proposed procedure.

**Criteria for Model Comparison** We use the well-known Akaike and Bayesian information criteria (AIC and BIC) for model comparison. Recall

$$AIC = -2 \log L + 2 \times npr, \quad BIC = -2 \log L + npr \times \log(nobs),$$

where $npr$ and $nobs$ are the number of parameters and number of observations respectively. Both criteria reward goodness of fit and penalize the increasing number of parameters. The penalty discourages over-fitting of the model. Although the BIC has a heavier penalty than AIC, the preferred model is the one with the lowest AIC or BIC value. We also look at the overall transition probabilities $P_{00}$ and $P_{11}$. Models which have estimated probabilities closer to observed ones are preferred, see (2) and the discussions therein.

**Model Selection** As discussed previously and in the Appendix, we can obtain RCM links from Laplace transforms, moment generating functions, and characteristic functions; we can acquire new RCM links from existing RCM via convex linear combinations, products, and two types of composite. In particular, we can construct RCM links from UCM links. Thus, we have a rich family of parsimonious PE distributions, giving great flexibility in model fitting, regression analysis, etc. We now give a forward model selection procedure for finding a possible optimal model from such a rich family. Borrowed from the forward stepwise variable selection commonly used in multiple regression, Dang, *et al.* (2009) [3] gave a forward model selection
procedure. Notice that increasing the complexity of a model means increasing the number of parameters in the model, rather than increasing the number of covariates. Here we adopt a similar idea.

Start from simple RCM models with the fewest number of parameters, and select the best two or three among them according to certain criteria. Then build up larger models from the chosen ones by including more parameters via linear combinations, power composites, and nonnegative polynomial composites of chosen models. At each step, perform a test of significance of the larger model using the likelihood ratio test. The procedure continues until the significance of all additional parameters is rejected. Just as the forward variable selection in multiple regression, this procedure may identify a possible best model with considerably less computing time than the procedure that tries all possible models. We used this procedure in our applications to two real datasets and obtained satisfactory results.

4 A Simulation and Applications

In this section, we first run a small simulation to validate the inclusion of the Binomial distribution as a special case of the partially exchangeable distribution. We then apply the proposed procedure to analyze two real data sets.

A Small Simulation We now run a simulation to validate the partially exchangeable distribution includes the Binomial distribution as a special case. We generate 1000*15 random variables from the Bernoulli distribution with probability of success $p = 0.1$ to form 1000 binary sequences, each of which is 15 bits long. Here we consider the case of $p$ close to zero because estimation in this situation may become difficult. For each sequence, we record the initial state and count the numbers of transitions. We fit the simulated data with the Incomplete A-RCM model specified by link (12), and computed the MLE’s of the parameters. The results are as follows: $\hat{\theta}_1 = 0.891$, ...
\[ \hat{\theta}_2 = 0.905, \hat{\vartheta}_1 = 0.108, \text{ and } \hat{\vartheta}_2 = 0.108. \] These values indicate that the parameters of the underlying model satisfy \( \theta_1 = \theta_2, \vartheta_1 = \vartheta_2 \text{ and } \theta_1 + \vartheta_1 = 1. \) By Theorem 5, the data are from the Binomial model.

**Application in Dairy Science** A pathogen is any disease producing microorganism. The data, kindly provided by F. A. Quintana, consists of series of measurements of presence(1)/absence(0) about pathogen infection on cows. Pathogen infection causes mastitis, which is inflammation of the mammary gland or breast. The presence or absence of a pathogen can be determined from examining glandular organs for mastitis. Measurements contained in the data set were collected in each of the four quarters of each cow’s udder (the glandular organ in which milk is secreted and stored) at 11 time periods after lactation (period of the secretion of milk). Many of these cows were measured on several consecutive lactation periods, which took about 1 year each. One period of each cow was randomly selected, and restriction was made to those cases where at least five observations were available. Quintana and Newton (1998) [22] analyzed the data of 278 cows from the 374 originally available cows. They tested the order of dependence and concluded that the binary sequence of each cow is the realization of a first-order Markov chain, which exhibits serial dependence on the most recent binary variable. The observed overall transition probabilities are \( p_{00} = 0.896 \) and \( p_{11} = 0.967. \) On one hand, if a cow had a pathogen infection

<p>| Table 2: Estimation in the Cow Data. The optimal models are in bold. |
|---------------------|---------------------|---------------------|</p>
<table>
<thead>
<tr>
<th>Models</th>
<th>Parameter Estimates(s.d)</th>
<th>-loglik</th>
<th>( P_{00} ) (s.d)</th>
<th>( P_{11} ) (s.d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td></td>
<td>503.8</td>
<td>.896</td>
<td>.967</td>
</tr>
<tr>
<td>Bin_{1}(\theta_1) * Ga_{D}(\theta_2)</td>
<td>0.896(0.01) 0.108(0.01)</td>
<td></td>
<td>.896(.010) 0.928(.009)</td>
<td></td>
</tr>
<tr>
<td>MM_{1}(\theta_1) * MM_{1}(\theta_2)</td>
<td>6.480(1.03) 25.95(3.44)</td>
<td></td>
<td>6.866(.018) .963(.005)</td>
<td></td>
</tr>
<tr>
<td>Bin_{1}(\theta_1) * MM_{1}(\theta_2)</td>
<td>0.896(0.01) 25.95(3.38)</td>
<td></td>
<td>.896(.010) <strong>.963</strong>( .005)</td>
<td></td>
</tr>
<tr>
<td>Bin_{1}(\theta_1) * Ga_{D}(\theta_2, \theta_3)</td>
<td>0.896(0.01) 0.191(0.06) 0.315(0.16)</td>
<td>501.2</td>
<td>.896(.010) .949(.010)</td>
<td></td>
</tr>
<tr>
<td>Kbg_{s,t}(\theta_1, \theta_2, \nu, \rho = 0)</td>
<td>5.900(1.02) 23.83(2.50) 0.928(0.19)</td>
<td></td>
<td>.864(.032) .962(.008)</td>
<td></td>
</tr>
<tr>
<td>Kbg_{s,t}(\theta_1, \theta_2, \nu, \rho = 1)</td>
<td>6.480(1.04) 25.95(3.48) 0.000(0.00)</td>
<td></td>
<td>.866(.019) .963(.006)</td>
<td></td>
</tr>
<tr>
<td>Bin_{s}(\theta_1, \nu) * MM_{1}(\theta_2)</td>
<td>0.876(0.03) 0.901(0.12) 25.94(3.90)</td>
<td></td>
<td>.876(.031) .963(.005)</td>
<td></td>
</tr>
</tbody>
</table>
in the previous period, the chance to get infection at the next period was as high as 96.7%. On the other hand, a cow without a pathogen in the previous period only had 10.4% probability to get infection in the next period. The mixture of Markov chains depends on the initial value, so we split the data set into two parts, Cow0 and Cow1. The former includes the sequences with initial state 0, and the latter contains the sequences with initial state 1. We find that there is no significant difference between the transition probabilities of Cow0 and Cow1. Specifically, in Cow0, \( p_{00} = 0.897 \) and \( p_{11} = 0.960 \); while in Cow1, \( p_{00} = 0.890 \) and \( p_{11} = 0.970 \).

We fit Cow0, Cow1 and the combined Cow data separately and obtained the log-likelihoods -151.08, -351.22, and -502.3 respectively. Note that the Cow data had a slight smaller log-likelihood with two parameters fewer. The likelihood ratio test on the hypothesis that the two models fitted to Cow0 and Cow1 are different had a p-value 0.22. Thus, we assume all individual binary sequences share a mixture of Markov chains independent of the initial state, i.e. in (8),

\[
 g_{i,j}(\cdot | b_0 = 0) = g_{i,j}(\cdot | b_0 = 1) := g_{i,j}(\cdot).
\]

Although we can obtain a RCM link directly from a UCM link via \( g(i, j) = f(i + j) \) where \( f \) is a UCM, the resulting RCM \( g(i, j) \) is symmetric in \( i \) and \( j \), leading to the same probability for \( P_{00} \) and \( P_{11} \). Such links are inappropriate in this data. Thus, we shall consider three types of RCM links. The first type is the product of two different UCM links. Dang, et al. (2009) provided a rich family of UCM links. We shall illustrate this type of links with the three one-parameter UCM links in Table 1: the Binomial, Gamma, and MM link. The second type is the RCM links resulting from the Laplace transforms or moments of a bivariate distribution. Here we shall consider the gamma links (9) and (10). The third type is the RCM links resulting from the linear combinations, products, and composites of existing RCM links.

The parameters are estimated by maximum likelihood using the R routine \textit{nlminb}. 

15
We then estimate the marginal probability \( \lambda^{(bn)}_{i,j} \) via formulas (6) and (8), noting that the MLE of \( \alpha \) is \( \hat{\alpha}_n = n_0/(n+1) = 0.457 \). In particular, we are interested in estimating the transition probabilities \( P_{00} \) and \( P_{11} \) which can be estimated via (2) by \( \hat{\lambda}_{1,0} \) and \( \hat{\lambda}_{0,1} \). The standard deviations of these estimates can be found by the asymptotic normality of the estimates and the delta method. Reported in Table 2 are these estimates, the negative log-likelihood, and the estimated transition probabilities.

Here we use the proposed forward model selection procedure to choose the optimal models. We start with simple two-parameter RCM links constructed via nine pairwise products of the above mentioned three one-parameter UCM links, i.e., the Ind-Bin, Gamma-Bin and MM-Bin link. We use the criterion of highest likelihood to select the optimal models. Note that this is equivalent to the criteria of AIC and BIC because of equal number of parameters. We have selected three optimal models, which are listed in the first three rows in Table 2. We continue to consider larger models which contain the selected ones as sub-models. Note that \( \text{Bin}_s(\theta_1) \ast \text{Ga}_t(\theta_2, \theta_3) \) reduces to \( \text{Bin}_s(\theta_1) \ast \text{Ga}_t(\theta_2) \) when \( \theta_3 = 1 \). The additional parameter \( \theta_3 \) increases the likelihood from -503.8 to -501.2, whereas the asymptotic likelihood ratio test rejects \( \theta_3 = 1 \) at a significant level 0.02. Besides, \( \theta_3 \) also improves the estimated transition probability of \( P_{11} \) which is closer to observed one. Kibble’s bivariate gamma model (10) includes \( \text{MM}_s(\theta_1) \ast \text{MM}_t(\theta_2) \) as a sub-model. However, there is strong evidence from the data to accept \( v = 1 \) and \( \rho = 0 \) with p-values 0.841 and 1 respectively. So we are clearly in favor of \( \text{MM}_s(\theta_1) \ast \text{MM}_t(\theta_2) \) rather than either of Kbiga models due to model simplicity. The composite of \( \text{Bin}_s(\theta_1) \ast \text{MM}_t(\theta_2) \) with a power yields the RCM link:

\[
h(s, t; \theta) = \text{Bin}_s(\theta_1, \nu) \ast \text{MM}_t(\theta_2) = \theta_1^{\nu} \frac{\theta_2}{t + \theta_2}, \quad \theta = (\theta_1, \theta_2, \nu) \in (0, \infty)^2 \times [0, 1].
\]

The extra power parameter \( \nu \) does not yield significant gain in likelihood. Summarizing our findings, we have chosen the optimal models \( \text{Bin}_s(\theta_1) \ast \text{Ga}_t(\theta_2, \theta_3) \) and \( \text{Bin}_s(\theta_1) \ast \text{MM}_t(\theta_2) \). Perhaps the latter has a slight edge because of fewer parameters.
and more accuracy in the estimates of the transition probabilities. For comparison, we have fitted the Cow data with the two-urn model of Quintana and Newton (1999). The MLE’s of their model are 15.16, 2.076, 3.544, and 0.190 respectively. The negative likelihood is 575.5 and the estimated overall transition probabilities are \( \hat{P}_{00} = 0.879 \) and \( \hat{P}_{11} = 0.949 \). It can be seen that the two-urn model was inferior to our optimal models in this dataset.

**Application in Medical Science** The bladder cancer data (kindly provided by F. A. Quintana) was analyzed by Quintana and Müller (2004) [24] and Davis and Wei (1988) [4]. Here we will apply the proposed partially exchangeable model to re-analyze it. This bladder cancer study was conducted by the Veterans Administration Cooperative Urological Research Group (VACURG). State I bladder tumors can usually be completely removed by transurethral resection, that is performed through a tube (urethra) which conveys urine from the bladder to the outside. Unfortunately, many patients have multiple recurrences. The subsequent tumors sometimes show a higher degree of malignancy and may even progress to invasive carcinoma. In order to determine if recurrences of Stage I bladder cancer can be prevented, a randomized clinical trial was conducted. All patients had superficial bladder tumors when they entered the trial. These tumors were removed transurethrally and patients were assigned to one of the three groups: placebo, thiotepa, and pyridoxine (Vitamin B6). Since the size of study was relatively small, patients with thiotepa and pyridoxine were combined in the treatment group, while patients with placebo were in the control group. At subsequent follow-up visits, any recurrent tumors were removed and the treatment was continued.

Although the patients were scheduled to be re-examined every 3 months for tumor recurrences during the three-year period, there were many missing observations. The bladder cancer data consists 73 measurements out of 82 patients with the maximum
### Table 3: Estimation and Model Comparison in the Bladder Cancer Study.
The optimal model is in bold.

<table>
<thead>
<tr>
<th>Models</th>
<th>Parameter Estimates</th>
<th>-loglik</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Ga_1(\alpha_1, \beta_1) \ast Ga_1(\alpha_2, \beta_2))</td>
<td></td>
<td>.0344</td>
<td>-0.197</td>
<td>1.683</td>
</tr>
<tr>
<td>St.Dev</td>
<td></td>
<td>0.067</td>
<td>0.080</td>
<td>0.264</td>
</tr>
<tr>
<td>(Ga_1(\alpha_1, \beta_1) \ast MM_1(\alpha_2, \beta_2))</td>
<td></td>
<td>.0344</td>
<td>-0.197</td>
<td>0.568</td>
</tr>
<tr>
<td>St.Dev</td>
<td></td>
<td>0.067</td>
<td>0.080</td>
<td>0.156</td>
</tr>
<tr>
<td>(Bin_1(\alpha_1, \beta_1) \ast MM_1(\alpha_2, \beta_2))</td>
<td></td>
<td>.0829</td>
<td>0.113</td>
<td>0.565</td>
</tr>
<tr>
<td>St.Dev</td>
<td></td>
<td>0.014</td>
<td>0.017</td>
<td>0.152</td>
</tr>
<tr>
<td>(MM_1(\alpha_1, \beta_1) \ast MM_1(\alpha_2, \beta_2))</td>
<td></td>
<td>2.968</td>
<td>10.280</td>
<td>0.568</td>
</tr>
<tr>
<td>St.Dev</td>
<td></td>
<td>0.570</td>
<td>3.390</td>
<td>0.157</td>
</tr>
<tr>
<td>(Bg_s,t(\alpha_1, \beta_1, \alpha_2, \beta_2, \theta_3, \rho))</td>
<td></td>
<td>.345</td>
<td>-0.198</td>
<td>1.681</td>
</tr>
<tr>
<td>St.Dev</td>
<td></td>
<td>0.067</td>
<td>0.080</td>
<td>0.265</td>
</tr>
<tr>
<td>(Ga_1(\alpha_1, \beta_1, \theta) \ast MM_1(\alpha_2, \beta_2))</td>
<td></td>
<td>.338</td>
<td>-0.193</td>
<td>1.044</td>
</tr>
<tr>
<td>St.Dev</td>
<td></td>
<td>0.089</td>
<td>0.088</td>
<td>0.389</td>
</tr>
<tr>
<td>(Kbg_s,t(\alpha_1, \beta_1, \alpha_2, \beta_2, \rho, v = 1))</td>
<td></td>
<td>2.988</td>
<td>10.260</td>
<td>0.574</td>
</tr>
<tr>
<td>St.Dev</td>
<td></td>
<td>0.570</td>
<td>3.400</td>
<td>0.157</td>
</tr>
<tr>
<td>(Kbg_s,t(\alpha_1, \beta_1, \alpha_2, \beta_2, v = 0))</td>
<td></td>
<td>1.375</td>
<td>4.872</td>
<td>0.222</td>
</tr>
<tr>
<td>St.Dev</td>
<td></td>
<td>2.810</td>
<td>6.230</td>
<td>0.455</td>
</tr>
</tbody>
</table>

sequence length being 12 observations. A patient in the control group has the group indicator \(x = 0\) and in the treatment group \(x = 1\). There are 41 patients in the control group and 32 patients in the treatment group. In the control group, the observed transition probabilities are \(p_{00} = 0.829\) and \(p_{11} = 0.374\), while in the treatment group, \(p_{00} = 0.942\) and \(p_{11} = 0.333\). It seems that the treatment is effective in terms of higher \(p_{00}\) and lower \(p_{11}\). In the treatment group, a patient with a bladder cancer tumor in the previous period has a lower probability of recurrence in the next period than a patient in the control group, and the chance to stay without tumor is also more than 10% higher.

Here we are interested in the relationship between the cancer recurrence and the covariate variable \(x = 0\) and 1 (control and treatment). We model \(\theta_1 = \alpha_1 + \beta_1 x\) and \(\theta_2 = \alpha_2 + \beta_2 x\), while other parameters are treated as nuisance parameters. Note that for \(i = 1, 2\), \((\alpha_i, \beta_i)\) are linearly constrained. For example, the model from the product of two Gamma-Bin links requires \(\theta_i > 0\), which results in linear constraint
Table 4: Estimated & Observed Transition Probabilities in the Bladder Cancer Study. The optimal models are in bold.

<table>
<thead>
<tr>
<th></th>
<th>Control P00 (s.d.)</th>
<th>Control P11 (s.d.)</th>
<th>Treatment P00 (s.d.)</th>
<th>Treatment P11 (s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed</td>
<td>0.829</td>
<td>0.374</td>
<td>0.942</td>
<td>0.333</td>
</tr>
<tr>
<td>Ga_s(α1, β1) * Ga_t(α2, β2)</td>
<td>0.788 (.036)</td>
<td>0.311 (.057)</td>
<td>0.903 (.027)</td>
<td>0.264 (.087)</td>
</tr>
<tr>
<td>Ga_s(α1, β1) * MM_t(α2, β2)</td>
<td>0.788 (.036)</td>
<td>0.362 (.064)</td>
<td>0.903 (.027)</td>
<td>0.307 (.098)</td>
</tr>
<tr>
<td>Bin_s(α1, β1) * MM_t(α2, β2)</td>
<td><strong>0.829 (.005)</strong></td>
<td><strong>0.361 (.062)</strong></td>
<td><strong>0.943 (.004)</strong></td>
<td><strong>0.320 (.098)</strong></td>
</tr>
<tr>
<td>MM_s(α1, β1) * MM_t(α2, β2)</td>
<td>0.748 (.036)</td>
<td>0.362 (.064)</td>
<td>0.930 (.016)</td>
<td>0.308 (.094)</td>
</tr>
<tr>
<td>Bg_s,t(α1, β1, α2, β2, θ3, ρ)</td>
<td>0.787 (.036)</td>
<td>0.312 (.058)</td>
<td>0.903 (.027)</td>
<td>0.264 (.087)</td>
</tr>
<tr>
<td>Ga_s(α1, β1, θ) * MM_t(α2, β2)</td>
<td>0.785 (.041)</td>
<td>0.362 (.064)</td>
<td>0.901 (.030)</td>
<td>0.307 (.098)</td>
</tr>
<tr>
<td>Kbg_s,t(α1, β1, α2, β2, ρ, v = 1)</td>
<td>0.749 (.036)</td>
<td>0.365 (.064)</td>
<td>0.930 (.016)</td>
<td>0.307 (.094)</td>
</tr>
<tr>
<td>Kbg_s,t(α1, β1, α2, β2, v, ρ = 0)</td>
<td>0.753 (.195)</td>
<td>0.412 (.145)</td>
<td>0.926 (.109)</td>
<td>0.351 (.162)</td>
</tr>
</tbody>
</table>

α_i + β_i > 0 for i = 1, 2. We employ the sub-routine constrOptim to find the numerical values of the MLE’s.

Reported in Table 3 are the estimates of the parameters, the values of negative log likelihood, AIC and BIC. As in the Cow data, we build up our optimal models by the aforementioned forward model selection. It turns out that the four product links, listed in the first four rows of Table 3, of the UCM links outperform the other links according to the criteria of the log-likelihood, AIC or BIC. Note that the last four rows in Table 3 are four larger models. The bivariate Gamma link (9) reduces to Ga*Ga when ρ = 1. The LRT accepts ρ = 1 with p-value 1 so Ga*Ga is preferred by both AIC and BIC. The power-gamma link,
\[
h_{s,t}(θ) = \frac{1}{(1 + θ_3 s)^{α_1 + β_1 x} t + α_2 + β_2 x},
\]
includes Ga*MM as a sub-model. However, the extra parameter θ_3 is not significant.

Note that MM*MM is a sub-model of the Kibble’s bivariate gamma. The likelihood ratio test fails to reject ρ = 0 with the p-value approximately one and rejects ν = 1 with the p-value 0.0007.

Summing up the above discussion, we obtain the final four optimal models: Ga*MM,
Ga*MM, Bin*MM, and Kbiga(\(\rho = 0\)). According to AIC and BIC, Ga(s) * Ga(t) results in the best model. But since the numerical value of an MLE is highly sensitive, especially for data with small or moderate sample size, the model with the highest maximum likelihood may not yield the best fitting. In fact, if we are interested in estimating the transition probabilities, then it would make more sense to select a model by comparing the observed and estimated transition probabilities. Table 4 reports the observed and estimated transition probabilities for each model together with the standard deviations (s.d.). Clearly, all models capture the trend of effectiveness of the treatment. That is, the treatment group has lower cancer recurrence probability \(P_{11}\) and higher probability to keep the status with no tumor than those in the control group. The \(P_{00}\) in the control and treatment groups are underestimated by all models except for the model Bin(s) * MM(t). Model Kbiga(s, t)(\(\rho = 0\)) overestimates \(P_{11}\) in both groups and has the largest standard deviations of the estimated probabilities among all models. Model Bin(s) * MM(t) gives the balance between the estimated transition probabilities and the accuracy of the estimates in both groups and has demonstrated superior behavior to other models.

Let us look at the model Bin(s) * MM(t) a bit further. The Wald test strongly rejects \(\beta_1 = 0\) and accepts \(\beta_2 = 0\), indicating that the treatment is very useful in prevention of Stage I bladder cancer but not quite effective in treating recurrences. The treatment improves \(P_{00}\) from the value of 0.829 in the control group to 0.943 in the treatment grouped and reduces \(P_{11}\) from the value of 0.361 in the control group to 0.320 in the treatment group, even though the decrease is not statistically significant. Similar conclusions can be made for the other models.
5 Discussion and Further Research

In this article, we have proposed to use partial exchangeability to model serially correlated data. A generalization of exchangeability, PE is adequate in expressing serially correlated sequences from the perspective of its generality that a recurrent PE sequence is a mixture of Markov chains. Our main approach is the generalization of the methodology in generalized linear models.

We have presented constructive methods to find RCM links and provided numerous RCM links. We have furnished rules for constructing RCM links from existing CM links; in particular, from UCM links via linear combinations, products and composites. A forward model selection method is provided as a guideline for practitioners to build a possible optimal model. The proposed procedure is demonstrated with real datasets. The R codes that were used to analyze the data can be found at the web page http://www.olemiss.edu/~xdang. We have suggested several simple one-parameter UCM links, the Binomial, Gamma, and MM links, to be used for constructing RCM links. We have used one incomplete RCM link in a small simulation.

The present work opens up several topics that may deserve further investigation. One topic is multi-state spaces. Our focus in this article is on sequences of partially exchangeable \{0,1\}-valued r.v.’s. The approach can be generalized to sequences of partially exchangeable multi-valued r.v.’s. An example of this the Poisson r.v.’s which take nonnegative values. Another topic is finite partially exchangeable sequences. In view of Remark 3, it would make sense to have an analog of Theorem 3 of Diaconis, P. and Freedman, D. (1980c) [8] for partially exchangeable r.v.’s.

Acknowledgments We are grateful to Fernando A. Quintana for kindly providing and discussing the two data sets.
Appendix: Constructive Methods and Proofs

In this section, we first discuss constructive methods. Incomplete RCM links are introduced. In the end, we give the proofs to the theorems.

A useful method for creating RCM links is to take the advantage of the Laplace transforms as stated below with the proof given at the end.

**Theorem 4** Let $H$ be a probability measure on $[0, \infty)^2$. Then the Laplace transform of $H$,$$ar{h}(s, t) = \int_0^\infty \int_0^\infty \exp(-sx - ty) \, dH(x, y), \quad s, t \in [0, \infty),$$is RCM.

The Bivariate Gamma Link, Kibble’s Bivariate Gamma Link, and Bivariate Beta Link are the Laplace transforms of bivariate gamma distribution (Krishnamoorthy and Parthasarathy, 1951) [15], Kibble’s bivariate gamma distribution (Kibble, 1941) [14], and bivariate Beta distribution (Olkin and Liu, 2003) [20] respectively. Hence they are RCM.

Hereafter we shall assume that every RCM sequence $\{h_{i,j} : i, j = 0, 1, \ldots\}$ is obtained from $\{\bar{h}(s, t) : s, t \geq 0\}$ which allows a Laplace transform (16). Thus we may obtain RCM links from the tabulated Laplace transforms and, in particular, from the moment generating functions. Specifically, suppose that the moment generating function $M_H(s, t)$ of distribution $H$ exists for all $s \leq s_0, t \leq t_0$ for some $s_0, t_0 \geq 0$, then one has a RCM link given by$$\bar{h}(s, t) = M_H(-s, -t), \quad s, t \in [0, \infty). \quad (17)$$

Using the relation between the Laplace transform and the characteristic function, we can also obtain RCM links. Suppose that the characteristic function of $H$ is $\varphi_H$. Then it is easy to prove that$$\bar{h}(s, t) = \varphi_H(is, it), \quad s, t \in [0, \infty), \quad (18)$$
is a RCM link, where $i = \sqrt{-1}$ is the imaginary unit. A substitution $x = -\log u, y = -\log v$ in (16) yields a useful representation

$$\bar{h}(s, t) = \int_0^1 \int_0^1 u^s v^t dG(u, v), \quad s, t \in [0, \infty),$$

where $G = H \circ \log$ is the induced probability measure. Thus we may obtain RCM links from the moments of distribution $G$. Suppose that $G$ has $(s, t)$-th moment $E_G(U^s V^t)$ for $(U, V) \sim G$ for $s, t = 0, 1, 2, \ldots$. Then we have a RCM link given by

$$\bar{h}(s, t) = E_G(U^s V^t), \quad s, t = 0, 1, 2, \ldots,$$

where $E_G$ denotes the expectation under $G$.

Note that if the mixing distribution $G$ is parametric $G = G_\theta$ with $\theta \in \Theta$ then the resulting RCM link $\bar{h}(t; \theta)$ is parametric.

**Incomplete RCM Links** Additional parameters in a distribution give more flexibility. Analogous to the tolerance method of Dang, et al. (2009) [3], a RCM link containing additional parameters can be obtained from putting additional parameters in the limits of the integration in (19), namely,

$$\bar{h}(s, t; \theta) = \int_{\theta_1}^{\theta_2} \int_{\theta_1}^{\theta_2} u^s v^t dG_\theta(u, v), \quad \theta = (\theta_1, \theta_2, \vartheta_1, \vartheta_2, \vartheta) \in [0, 1]^4 \times \Theta. \quad (21)$$

By Remark 1, the following is a RCM link with $h_{0,0}(\theta) = 1$,

$$h_{s,t}(\theta) = \bar{h}(s + s_0, t + t_0; \theta)/\bar{h}(s_0, t_0; \theta), \quad \theta \in [0, 1]^4 \times \Theta \quad (22)$$

for $s_0, t_0 \geq 0$. It should be noted that the above RCM link is not defined when $\theta_1 = \theta_2$ or $\vartheta_1 = \vartheta_2$. Nevertheless it is readily verified that

$$\lim_{\theta_2 \to \theta_1} h_{s,t}(\theta_1, \theta_2, \vartheta_1, \vartheta_2, \vartheta) = \frac{\theta_1^s \bar{h}_2(t + t_0; \theta_1, \vartheta_1, \vartheta_2, \vartheta)}{\bar{h}_2(t_0; \theta_1, \vartheta_1, \vartheta_2, \vartheta)}, \quad \vartheta_1 \neq \vartheta_2, \quad (23)$$

where $\bar{h}_2(t; \theta_1, \vartheta_1, \vartheta_2, \vartheta) = \int_{\vartheta_1}^{\vartheta_2} v^t G_\theta(\theta_1, dv)$. Let us assume without loss of generality that $\bar{h}_2(s_0, t_0; \theta) = 1$ (e.g. $s_0 = t_0 = 0$). Interestingly, substitution of (23) in (7) yields
a representation
\[ \theta_{10}^t (1 - \theta_1)^{t_01} \int_0^1 v^{t_{11}} (1 - v)^{t_{10}} G_{\vartheta} (\theta_1, dv). \] (24)
This is a mixture of Markov chains corresponding to the stochastic transition matrices
\[ \left( \begin{array}{cc} \theta_1 & 1 - \theta_1 \\ 1 - v & v \end{array} \right), \quad 0 \leq v \leq 1, \]
where \( \theta_1 \) is the probability of transition from zero to zero and \( v \in [0, 1] \) is the probability of transition from one to one, given the distribution \( G_{\vartheta} \). Observe that the mixture in (24) is the product of Bernoulli distribution with probability \( \theta_1 \) of success and a Binomial mixture. In this case, the zeros in the binary sequence \( B_1, ..., B_n \) are independent and identically distribution with probability \( \theta_1 \) from zero to zero and with probability \( 1 - \theta_1 \) from zero to one, while the one’s in the sequence are exchangeable by the de Finetti theorem. It can be analogously interpreted for \( \vartheta_2 \to \vartheta_1 \).

Further,
\[ \lim_{\theta_2 \to \theta_1} \lim_{\vartheta_2 \to \vartheta_1} h_{s,t}(\theta_1, \theta_2, \vartheta_1, \vartheta_2, \vartheta) = \theta_1^s \vartheta_1^t. \] (25)
Substitution of the above in (7) yields a representation
\[ \theta_{10}^t (1 - \theta_1)^{t_01} \vartheta_{11}^t (1 - \vartheta_1)^{t_{10}}. \] (26)
This is a mixture corresponding to the stochastic transition matrix
\[ \left( \begin{array}{cc} \theta_1 & 1 - \theta_1 \\ 1 - \vartheta_1 & \vartheta_1 \end{array} \right). \]
Now the mixture in (26) is a product of two Bernoulli distributions, the zero’s and the one’s in the binary sequence are i.i.d. sequences with probabilities \( \theta_1 \) and \( \vartheta_1 \) of success respectively. Furthermore, when \( \theta_1 + \vartheta_1 = 1 \), it recovers the Bernoulli distribution, \( \theta_{10}^t (1 - \theta_1)^{t_01} (1 - \vartheta_1)^{t_{10}} \). We can now extend the definition of \( h_{s,t}(\theta_1, \theta_2, \vartheta_1, \vartheta_2, \vartheta) \) to admit the equality by defining it to be the limit, i.e.,
\[ h_{s,t}(\theta_1, \theta_2, \vartheta_1, \vartheta_2, \vartheta) = \begin{cases} \theta_1^s \vartheta_1^t, & \theta_1 = \theta_2, \vartheta_1 \neq \vartheta_2, \\ \theta_1^s \vartheta_1^t, & \theta_1 \neq \theta_2, \vartheta_1 = \vartheta_2, \\ \vartheta_1^s \theta_1^t, & \theta_1 = \theta_2, \vartheta_1 = \vartheta_2. \end{cases} \] (27)
Thus $h_{s,t}(\theta)$ is well defined for all $\theta \in [0,1]^4 \times \Theta$. Hereafter we shall adopt this definition and refer it to as an incomplete RCM link. The Bernoulli model is recovered when $\theta_1 = \theta_2$, $\varphi_1 = \varphi_2$ and $\theta_1 + \varphi_1 = 1$. Summing up our findings, we have the following result.

**Theorem 5** Suppose that $\{\bar{h}(s,t; \theta) : s,t \geq 0\}$ is given in (16). If (27) holds with $h_{s,t}$ given in (22), then $h(\theta) = \{h_{i,j}(\theta)\}$ well defines a mixture model $\text{PE}(h)$. The mixture model recovers the Bernoulli model when $\theta_1 = \theta_2$, $\varphi_1 = \varphi_2$ and $\theta_1 + \varphi_1 = 1$.

**Proof of Theorem 2.** We first prove (LP). The RCM of the convex linear combination follows easily from the definition. Since $\bar{g}$ and $\bar{h}$ are RCM, it follows by Theorem 4 that

$$\bar{g}(s,t) = \int_0^\infty \int_0^\infty \exp(-sx - ty) dG(x,y), \quad \bar{h}(s,t) = \int_0^\infty \int_0^\infty \exp(-su - tv) dH(u,v)$$

for some probability measures $G$ and $H$ on $[0,\infty)^2$. Hence by (4),

$$(-1)^{r_1+r_2} \Delta_1^{r_1} \Delta_2^{r_2} \bar{g}(s,t)\bar{h}(s,t)$$

$$= (-1)^{r_1+r_2} \Delta_1^{r_1} \Delta_2^{r_2} \int_0^\infty e^{-(x+u)} e^{-(y+v)t} dH_1(x,y) dH_2(u,v)$$

$$= \int_0^\infty \sum_{k=0}^{r_1} \sum_{l=0}^{r_2} \binom{r_1}{k} \binom{r_2}{l} (-1)^{k+l} e^{-(x+u)(k+s)} e^{-(y+v)(l+t)} dH_1(x,y) dH_2(u,v)$$

$$= \int_0^\infty e^{-(x+u)(1-s) - (y+v)t} (1 - e^{-(x-u)\psi_1(t)}) (1 - e^{-(y-v)\psi_2(t)}) dH_1(x,y) dH_2(u,v) \geq 0.$$ 

This shows the RCM of the product. We now show (C1). Since $\psi_1(s)$ and $\psi_2(t)$ are positive with RCM derivatives, it follows from Theorem (C.1) of Dang, et al. (2009) that $p^{\psi_1(s)}$ and $p^{\psi_2(t)}$ are CM for $0 < p < 1$. Thus (C1) follows immediately from

$$(-1)^{r_1+r_2} \Delta_1^{r_1} \Delta_2^{r_2} \bar{h}(\psi_1(s),\psi_2(t)) = (-1)^{r_1+r_2} \Delta_1^{r_1} \Delta_2^{r_2} \int_0^\infty \int_0^\infty e^{-\psi_1(s)x - \psi_2(t)y} dH(x,y)$$

$$= \int_0^\infty \int_0^\infty (-1)^{r_1} \Delta_1^{r_1} (e^{-x}\psi_1(s)) (-1)^{r_2} \Delta_2^{r_2} (e^{-y}\psi_2(t)) dH(x,y) \geq 0, \quad r_1, r_2 = 0, 1, ...$$

For (C2), $h(\phi_1(s,t), \phi_2(s,t)) = \sum_{j,k} \theta_{j,k} \phi_1^j(s,t) \phi_2^k(s,t)$. By (LP), $\phi_1^j(s,t) \phi_2^k(s,t)$ is RCM for all $j, k$. Thus again by (LP), the desired result follows in view of $\theta_{j,k} \geq 0$ for all $j, k$. □
Proof of Theorem 4  
By (4),
\[
(-1)^{r_1 + r_2} \Delta_{r_1} \Delta_{r_2} \bar{h}(s,t) = \sum_{j=0}^{r_1} \sum_{k=0}^{r_2} \binom{r_1}{j} (1-r_1)^j \binom{r_2}{k} (1-r_2)^k \bar{h}_{s+j,t+k}
\]
\[
= \int_0^\infty \int_0^\infty e^{-sx-ty} (1-e^{-x})^{r_1} (1-e^{-y})^{r_2} dH(x,y) \geq 0, \quad s, t \geq 0.
\]
It follows from (5) of [21] that \( \bar{h} \) is RCM. This completes the proof. □

References


