The Distribution of Partially Exchangeable Random Variables

Hanxiang Peng[a,1,*], Xin Dang[b,1], Xueqin Wang[c,2]

[a] *Department of Mathematical Sciences, Indiana University-Purdue University at Indianapolis, Indianapolis, IN 46202-3216, USA*
[b] *Department of Mathematics, University of Mississippi, University, MS 38677, USA*
[c] *School of Mathematics & Computational Science and Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou City, Guangdong 510275, PRC*

Abstract

In this article, we derive the distribution of partially exchangeable binary random variables, generalizing the distribution of exchangeable binary random variables and hence the binomial distribution. The distribution can be viewed as a mixture of Markov chains. We introduce rectangular complete monotonicity and show that partial exchangebility can be characterized by rectangular complete monotonicity. The distribution aided with rectangular complete monotonicity can be used to analyze serially correlated data common in many areas of science.

Keywords:
Complete monotonicity, exchangeability, mixture of Markov chains, partial exchangeability, serial correlation.

1. Introduction

Binary responses commonly arise in many fields of science. The literature describing the analysis of correlated binary data is extensive. In this article, we propose to exploit the perception of *partial exchangeability* to investigate multiple serially dependent responses.

Introduced by de Finetti in 1938, partial exchangeability was studied by many authors including Diaconis (1988), Diaconis and Freedman (1980a, b, c), Freedman (1962a, b), and Zaman (1984). A sequence of binary random variables is partially exchangeable if the joint distribution of any sub-sequence of the first finite terms is invariant under permutations that keep the initial state and the numbers of 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

*Hanxiang Peng (hpeng@math.iupui.edu)

[1] This research was supported by the US National Science Foundation under Grant No. DMS-0940365.

[2] This research is supported by Doctoral Fund of Ministry of Education of China under Grant No. 20090171110017 and Tian Yuan Fund for Mathematics under Grant No. 10926200.

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.

A model in which an individual binary sequence is assumed to be recurrent and partially exchangeable is, according to Diaconis and Freedman (1980a, c), a mixture of Markov chains. Whereas the latter, as pointed out by Quintana and Newton (1998), can be viewed as a random-effects model in which the unobservable latent transition matrices follow some unknown distribution. Our proposed distribution can be viewed either as a Markov-chain-mixture model (both parametric and nonparametric), a random-effects model, or a Bayesian model. We represent the distribution of partially exchangeable binary random variables in moments of the mixing probability measure, an analog to the formula of the distribution of exchangeable binomial random variables by Kendall (1967), Chow and Teicher (1997), George and Bowman (1995), and Dang, Keeton and Peng (2009) among others. We show that partial exchangebility can be characterized by rectangular complete monotonicity. The distribution aided with rectangular complete monotonicity can be used to analyze serially correlated data. Peng, Garner and Dang (2009) used rectangular complete monotonic functions to give a class of parsimonious Markov chain mixtures, and applied the parsimonious mixtures to analyze two real datasets. Their method is related to the procedure used in Dang, *et al.* (2009) in which a unified approach is presented for analyzing exchangeable binary data with an application in developmental toxicity studies. By exploiting partial exchangeability, Quintana and Newton (1998) 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) constructed a two-urn model for the mixtures of Markov chains, providing interesting and useful interpretation.

The rest of the article is organized as follows. In Section 2, we derive the distribution of partial exchangeable binary random variables, discuss the sufficient statistics of the distribution, and introduce rectangular complete monotonicity to characterize partial exchangeability. Technical details can be found in the Appendix.

2. The Partially Exchangeable Model and Rectangular Complete Monotonicity

In this section, we first give the distribution of partially exchangeable binary r.v.'s, followed by rectangular complete monotonicity and characterization of partial exchangeability.

As Diaconis and Freedman (1980a) elaborated, a sequence of binary r.v.'s $B_0, B_1, ...$ are *partially exchangeable* if for any two binary sequences $\{b_i : i = 0, 1, ..., n\}$ and $\{c_i : i = 0, 1, ..., n\}$ which have the same initial state and the same transition counts,

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

Recall that a sequence of binary r.v.'s $B_0, B_1, ...$ is recurrent if

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

Diaconis and Freedman (1980a) show that if $B_0, B_1, ..., B_n$ are recurrent and partially exchangeable, then the joint probability of them is a mixture of Markov chains. Specifically, for the binary $\{0, 1\}$ state space there exists a unique probability measure Q_{b_0} on $[0, 1]^2$ such that

$$P(B_0 = b_0, B_1 = b_1, ..., B_n = b_n) = \int_0^1 \int_0^1 u^{t_{00}}(1-u)^{t_{01}} v^{t_{11}}(1-v)^{t_{10}}\, dQ_{b_0}(u, v), \quad (1)$$

where t_{ij} is the number of transitions from i to j for i, $j = 0$, 1. From this and the binomial formula, we have the following result.

Theorem 1. *Suppose that $B_0, B_1, B_2, ...$ are recurrent and partially exchangeable. Then*

$$\mathbb{P}(B_0 = b_0, B_1 = b_1, ..., B_n = b_n) = \sum_{i=0}^{t_{01}} \sum_{j=0}^{t_{10}} (-1)^{i+j} \binom{t_{01}}{i} \binom{t_{10}}{j} \lambda^{(b_0)}_{t_{00}+i, t_{11}+j}, \quad (2)$$

where $b_0, b_1, ..., b_n$ are $\{0, 1\}$-valued, and $\lambda^{(b_0)}_{i,j}$ is the joint probability that the transitions $0 \to 0$ and $1 \to 1$ in $B_0, B_1, ..., B_n$ simultaneously occur i and j times with initial state b_0, or symbolically,

$$\lambda^{(b_0)}_{i,j} = \mathbb{P}\left(B_0 = b_0, (0 \to 0)^i, (1 \to 1)^j\right), \; b_0 = 0, 1, \; 0 \le i \le t_{00} + t_{01}, \; 0 \le j \le t_{11} + t_{10}.$$

We refer (2) the *partially exchangeable model* and denote it by $\mathbf{PE}(\boldsymbol{\lambda})$, where $\boldsymbol{\lambda} = \{\lambda^{(b_0)}_{i,j} : b_0 = 0, 1, \; i = 0, ..., t_{00} + t_{11}, \; j = 0, ..., t_{11} + t_{10}\}$. Representation (2) is an analog to the formula of the distribution of exchangeable binomial random variables by Kendall (1967), Chow and Teicher (1997), George and Bowman (1995), and Dang, *et al.* (2009) among others. It is interesting to observe that a (parametric or nonparametric) mixture of Markov chains is a parametric model with a sequence of parameters $\boldsymbol{\lambda}$ in which the number of parameters increases with the length of the binary sequence. The following remark describes the relationships among the distributions of partially exchangeable, exchangeable and independent and identically distributed r.v.'s.

Remark 1. If $B_0, ..., B_n$ are exchangeable, then

$$\lambda^{(b_0)}_{j,k} = \sum_{i=0}^{j} (-1)^i \binom{j}{i} \pi^{(b_0)}_{k+i}, \quad b_0 \in \{0, 1\}, \; 0 \le j \le t_{00} + t_{01}, \; 0 \le k \le t_{11} + t_{10}. \quad (3)$$

where $\pi^{(b_0)}_i = P(B_0 = b_0, B_1 = 1, ..., B_i = 1)$, $i \ge 0$. Hence

$$\mathbb{P}(B_1 = b_1, ..., B_n = b_n) = \sum_{i=0}^{t_{00}+t_{10}} (-1)^i \binom{t_{00} + t_{10}}{i} \pi_{t_{01}+t_{11}+i}. \quad (4)$$

where $\pi_j = P(B_0 = 1, B_1 = 1, ..., B_{j-1} = 1)$, $j \geq 1$ with $\pi_0 = 1$. This is the probability under exchangeability, see Dang, *et al.* (2009). Accordingly, the Binomial distribution is a special case of the partial exchangeable model. Further, under exchangeability,

$$\pi_j^{(0)} = \pi_j - \pi_{j+1}, \quad \pi_j^{(1)} = \pi_{j+1},$$

where $\pi_j = P(B_0 = 1, B_1 = 1, ..., B_{j-1} = 1)$, $j \geq 1$ with $\pi_0 = 1$.

Remark 2. By Diaconis and Freedman (1980a), the binary state space in Theorem 1 can be generalized to a multi-state space.

The marginal probabilities $\boldsymbol{\lambda} = \{\lambda_{j,k}^{(b_0)}\}$ are moments of the mixing probability measure. Specifically,

$$\lambda_{ij}^{(b_0)} = \int_0^1 \int_0^1 u^i v^j \, dQ_{b_0}(u, v), \quad b_0 = 0,\ 1,\ i,\ j = 0, 1, ...$$

The moment problem is a folklore in functional analysis and one-dimensional Hausdorff moment problem is related to *complete monotonicity (CM)*. Here we extend it to a two-dimensional sequence, which is related to the two dimensional Hausdorff moment problem. A finite bivariate sequence $\{a_{j,k} : 0 \leq j \leq J,\ 0 \leq k \leq K\}$ is said to be *(finite) rectangular completely monotone (RCM)* if

$$(-1)^{r_1+r_2} \Delta_1^{r_1} \Delta_2^{r_2} a_{j,k} \geq 0, \quad 0 \leq j + r_1 \leq J,\ 0 \leq k + r_2 \leq K, \tag{5}$$

where Δ_1, Δ_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 $\Delta_1^0 = \Delta_2^0 = I$ being the identity operator. It is not difficult to verify the useful expansion

$$\Delta_1^r \Delta_2^s a_{u,v} = \sum_{j=0}^{r} \sum_{k=0}^{s} \binom{r}{j} \binom{s}{k} (-1)^{r+s-j-k} a_{u+j,v+k} \quad r, s, u, v = 0, 1, ... \tag{6}$$

for a two-dimensional sequence $\{a_{j,k} : j, k = 0, 1, ...\}$. The above expansion can be used to verify RCM. Applying (6) to (5) for $r = s = 1$ gives $a_{u+1,v+1} - a_{u+1,v} - a_{u,v+1} + a_{u,v} \geq 0$, which motivates us to call rectangular.

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

$$(-1)^{r_1+r_2} \Delta_1^{r_1} \Delta_2^{r_2} \lambda_{j,k}^{(b_0)} \geq 0,\ 0 \leq j + r_1 \leq t_{00} + t_{01},\ 0 \leq k + r_2 \leq t_{11} + t_{10}. \tag{7}$$

An infinite bivariate sequence $\{\bar{\lambda}(i, j) : i, j = 0, 1, ...\}$ with $\bar{\lambda}(0, 0) = 1$ is called *infinite rectangular completely monotone (IRCM)* if $\bar{\lambda}(i, j)$ satisfies (5) for all nonnegative integers i, j and r_1, r_2. IRCM characterizes partial exchangeability as stated in the following theorem. The simple proof is based on the well-known result in the two-dimensional Hausdorff moment problem by Haviland (1935) and is relegated in the Appendix. This characterization is useful in obtaining parsimonious mixtures of Markov chains, see Peng, Dang and Garner (2009).

Theorem 2. *Two infinite sequences* $\{\bar{\lambda}^{(b_0)}(i,j) : i,j = 0,1,...\}$ *for* $b_0 = 0,1$ *satisfying* $\bar{\lambda}^{(0)}(0,0) + \bar{\lambda}^{(1)}(0,0) = 1$ *induce a probability distribution (2) if and only if they are IRCM.*

Clearly the matrix $\mathbf{t} = \{t_{ij} : i,j = 0,1\}$ of transition counts and the initial state b_0 are sufficient statistics of the PE distribution. The number of the strings $B_0, B_1, ..., B_n$ that have a common transition matrix $\mathbf{t}$ is $\binom{n_0-1}{t_{00}}\binom{n_1-1}{t_{11}}$, where n_0, n_1 are the number of zero's and one's respectively. See Quintana and Newton (1998) for details. Let $p(\mathbf{t}, b_0; \boldsymbol{\lambda})$ be the probability that $B_0, B_1, ..., B_n$ have a common transition matrix $\mathbf{t}$ with initial state b_0. Then by (2),

$$p(\mathbf{t}, b_0; \boldsymbol{\lambda}) = \binom{n_0-1}{t_{00}}\binom{n_1-1}{t_{11}} \sum_{i=0}^{t_{01}} \sum_{j=0}^{t_{10}} (-1)^{i+j} \binom{t_{01}}{i}\binom{t_{10}}{j} \lambda^{(b_0)}_{t_{00}+i, t_{11}+j}. \tag{8}$$

This formula seems to suggest that the possible values of t_{00} and t_{11} with which $\{p(\mathbf{t}, b_0; \boldsymbol{\lambda}) : b_0 = 0,1, \forall \mathbf{t}\}$ constitutes a probability distribution are $\{0, ..., n_0 - 1\}$ and $\{0, ..., n_1 - 1\}$ respectively. A further glance reveals that the constraints on t_{01}, t_{10} are not taken into account, that is, t_{01}, t_{10} can take at most half the value of the length $n+1$ of the string because the number of transitions from zero to one or from one to zero is determined by $n_0 \wedge n_1 = \min(n_0, n_1)$. Indeed, the values of t_{01} and t_{10} are equal if the initial state b_0 and terminal state b_n are identical $b_0 = b_n$, otherwise they differ by one, i.e., $t_{01} - 1 = t_{10}$ if $b_0 = 0, b_n = 1$ and $t_{10} - 1 = t_{01}$ if $b_0 = 1$ and $b_n = 0$. Thus given the initial and terminal states b_0, b_n, either of t_{01} and t_{10} determines the other. Note that the length of the string $B_0, ..., B_n$ can be calculated by $n+1 = t_{00}+t_{01}+t_{10}+t_{11}+1$. Also $n_0 - \mathbf{1}[b_0 = 0] = t_{00}+t_{10}$, $n_1 - \mathbf{1}[b_0 = 1] = t_{11}+t_{01}$, and $n_0 + n_1 = n+1$. Thus the pentuple $(\mathbf{t}, b_0)$ can be equivalently expressed as the pentuple $(s, n_0, b_0, b_n, n+1)$ where s is either t_{01} or t_{10}. In fact, the latter is more convenient in describing the distribution. For this reason, we shall write $p(\mathbf{t}, b_0; \boldsymbol{\lambda})$ as $p(s, n_0, b_0, b_n, n+1; \boldsymbol{\lambda})$ or as usual omit $n+1$ and write as $p(s, n_0, b_0, b_n; \boldsymbol{\lambda})$. We now derive the formula for $p(s, n_0, b_0, b_n; \boldsymbol{\lambda})$. To simplify our description, let us consider the string starting with state $b_0 = 0$ and denote $s = t_{01}$. Then $n_0 - 1 = t_{00} + t_{10}$, $n_1 = t_{11} + t_{01}$, and $n_1 = n + 1 - n_0$. There are now two cases.
Case 1. The terminal state is 0. Then $t_{01} = t_{10} = s$, $t_{00} = n_0 - 1 - s$, $t_{11} = n_1 - s$. The range of s is the integers in the interval $[0,\ (n_0 - 1) \wedge n_1]$.
Case 2. The terminal state is 1. Then $t_{01} = s, t_{10} = s - 1$, $t_{00} = n_0 - s$, $t_{11} = n_1 - s$. The range of s is the integers in the interval $[1,\ n_0 \wedge n_1]$.

Accordingly, we have

$$\begin{aligned} p(s, n_0, 0, b_n; \boldsymbol{\lambda}) = & \binom{n_0-1}{s-\mathbf{1}_1}\binom{n_1-1}{s-1} \times \\ & \sum_{i=0}^{s} \sum_{j=0}^{s-\mathbf{1}_1} (-1)^{i+j} \binom{s}{i}\binom{s-\mathbf{1}_1}{j} \lambda^{(0)}_{n_0-\mathbf{1}_0-s+i,\, n_1-s+j}, \\ & b_n = 0,1,\ n_0 = 1, ..., n+1, s = \mathbf{1}_1, ..., (n_0 - \mathbf{1}_0) \wedge n_1. \end{aligned} \tag{9}$$

where $\mathbf{1}_1 = \mathbf{1}[b_n = 1]$ and $\mathbf{1}_0 = 1 - \mathbf{1}_1$. Here we set $\binom{m}{i} = 0$ if either $i > m$, or $m < 0$ or $i < 0$ except $\binom{m}{i} = 1$ when $m = i \in \{0, -1\}$. Analogously one has the formula for the string starting with one. The two formulas can be combined in one as follows:

$$p(s, n_0, b_0, b_n; \boldsymbol{\lambda}) = \binom{n_0 - 1}{s - 1 + \mathbf{1}_{00}} \binom{n_1 - 1}{s - 1 + \mathbf{1}_{11}} \times$$
$$\sum_{i=0}^{s-\mathbf{1}_{10}} \sum_{j=0}^{s-\mathbf{1}_{01}} (-1)^{i+j} \binom{s - \mathbf{1}_{10}}{i} \binom{s - \mathbf{1}_{01}}{j} \lambda^{(b_0)}_{n_{00}-s+i,\, n_{11}-s+j},$$
$$b_0, b_n = 0, 1,\ n_0 = \mathbf{i}_0, ..., n + 1 - \mathbf{i}_1,\ s = \mathbf{1}_{01} + \mathbf{1}_{10}, ..., n_{00} \wedge n_{11}, \quad (10)$$

where $\mathbf{1}_{ij} = \mathbf{1}[b_0 = i, b_n = j], i, j = 0, 1$, $n_{00} = n_0 - \mathbf{1}_{00}$, $n_{11} = n_1 - \mathbf{1}_{11}$, $\mathbf{i}_0 = \mathbf{1}[b_0 = 0]$ and $\mathbf{i}_1 = 1 - \mathbf{i}_0$. Since the sum of all the probabilities $P(B_0 = b_0, B_1 = b_1, ..., B_n = b_n)$ for $b_0, ..., b_n = 0, 1$ is one, it follows

$$\sum_{b_0=0}^{1} \sum_{b_n=0}^{1} \sum_{n_0=\mathbf{i}_0}^{n+1-\mathbf{i}_1} \sum_{s=\mathbf{1}_{01}+\mathbf{1}_{10}}^{n_{00}\wedge n_{11}} p(s, n_0, b_0, b_n; \boldsymbol{\lambda}) = 1. \quad (11)$$

This shows the following result.

Theorem 3. $\{p(s, n_0, b_0, b_n; \boldsymbol{\lambda}) : b_0, b_n = 0, 1,\ n_0 = \mathbf{i}_0, ..., n + 1 - \mathbf{i}_1,\ s = \mathbf{1}_{01} + \mathbf{1}_{10}, ..., n_{00} \wedge n_{11}\}$ *is a probability distribution.*

The justification that $p(s, n_0, b_0, b_n; \boldsymbol{\lambda})$ constitutes a probability distribution provides the theoretical basis for the statistical inference with it, such as the use of the log likelihood as a criterion of model selection, see the application of the distribution in real data by Peng, *et al.* (2009).

From the above discussion, it is easily checked these relations:

$$s = t_{01}\mathbf{i}_0 + t_{10}\mathbf{i}_1, \quad n_0 = t_{00} + t_{10} - \mathbf{i}_0, \quad b_n = \mathbf{1}[t_{01} = t_{10}]b_0 + \mathbf{1}[t_{01} = t_{10} + 1].$$

Let $S = B_0 + ... + B_n$. Then an immediate consequence of the above theorem is

$$\mathbb{P}(S = n + 1 - n_0) = \sum_{b_0=0}^{1} \sum_{b_n=0}^{1} \sum_{s=\mathbf{1}_{01}+\mathbf{1}_{10}}^{n_{00}\wedge n_{11}} p(s, n_0, b_0, b_n; \boldsymbol{\lambda}), \quad n_0 = 0, 1, ..., n + 1. \quad (12)$$

This generalizes the binomial distribution under exchangeability and hence under independence (binomial). The proof of (11) is a probabilistic one. An algebraic proof is interesting and gives additional insight. Using the integral representation (1), we have surprisingly found that we can prove the following stronger result:

$$q_{n+1}(u, v|b_0) \equiv 1, \quad b_0 = 0, 1,\ 0 \le u, v \le 1, \quad (13)$$

where $q_{n+1}(u, v|b_0)$ equals

$$\sum_{b_n=0}^{1} \sum_{n_0=\mathbf{i}_0}^{n+1} \sum_{s=\mathbf{1}_{01}+\mathbf{1}_{10}}^{n_{00}\wedge n_{11}} \binom{n_0 - 1}{s - 1 + \mathbf{1}_{00}} \binom{n_1 - 1}{s - 1 + \mathbf{1}_{11}} u^{n_{00}-s}(1 - u)^{s-\mathbf{1}_{10}} v^{n_{11}-s}(1 - v)^{s-\mathbf{1}_{01}}.$$

The proof of (13) is given in the Appendix. The identity (13) certainly implies (11). In addition, it is noteworthy to mention that (13) is an identity which generalizes the binomial identity, i.e., when $u+v=1$, it simplifies to the binomial identity

$$\sum_{n_0=0}^{n+1} \binom{n+1}{n_0} u^{n_0} v^{n+1-n_0} = (u+v)^{n+1} = 1.$$

Appendix: Technical Details

Proof of Theorem 2. We only have to show the sufficiency. It follows from the case 6 (page 168) of the theorem in Haviland (1935) that an infinite sequence $\{\bar{\lambda}(i,j)\}$ is IRCM if and only if it can be represented as

$$\bar{\lambda}(i,j) = \int_0^\infty \int_0^\infty e^{-ix-jy}\, d\bar{P}(x,y), \quad i,j = 0,1,...$$

for a probability $\bar{P}$ on $[0,\infty)^2$. An application of the result to $\{\bar{\lambda}(b_0,i,j) : i,j = 0,1,...\}$ for $b_0 = 0,1$ and substitution of the representations in (4) yields the desired result. □

Proof of Theorem 3. We shall prove (13) for $b_0 = 0$ and the case for $b_0 = 1$ holds by symmetry. Let $F_n(u,v) = q_{n+1}(u,v|b_0 = 0)$. Then (13) immediately follows from the iterative formula:

$$F_n(u,v) = uF_{n-1}(u,v) + (1-u)F_{n-1}(v,u), \quad 0 \le u,v \le 1,\ n = 1,2,... \tag{14}$$

We shall prove this by mathematical induction. For $n = 2k$, we have

$$\begin{aligned}
F_{2k}(u,v) &= \sum_{m=1}^{k}\sum_{s=0}^{m-1} \binom{m-1}{s}\binom{2k-m}{s-1} u^{m-1-s}(1-u)^s v^{2k+1-m-s}(1-v)^s \\
&+ \sum_{m=k+1}^{2k+1}\sum_{s=0}^{2k+1-m} \binom{m-1}{s}\binom{2k-m}{s-1} u^{m-1-s}(1-u)^s v^{2k+1-m-s}(1-v)^s \\
&+ \sum_{m=1}^{k}\sum_{s=0}^{m} \binom{m-1}{s-1}\binom{2k-m}{s-1} u^{m-s}(1-u)^s v^{2k+1-m-s}(1-v)^{s-1} \\
&+ \sum_{m=k+1}^{2k}\sum_{s=0}^{2k+1-m} \binom{m-1}{s-1}\binom{2k-m}{s-1} u^{m-s}(1-u)^s v^{2k+1-m-s}(1-v)^{s-1}.
\end{aligned}$$

Keep in mind that we make use of the convention $\binom{m}{i} = 0$ if either $i > m$, or $m < 0$

or $i < 0$ except $\binom{m}{i} = 1$ when $m = i \in \{0, -1\}$. For $n = 2k+1$, we have

$$\begin{aligned}
F_{2k+1}(u,v) &= \sum_{m=1}^{k+1}\sum_{s=0}^{m-1}\binom{m-1}{s}\binom{2k+1-m}{s-1}u^{m-1-s}(1-u)^s v^{2k+2-m-s}(1-v)^s \\
&+ \sum_{m=k+2}^{2k+2}\sum_{s=0}^{2k+2-m}\binom{m-1}{s}\binom{2k+1-m}{s-1}u^{m-1-s}(1-u)^s v^{2k+2-m-s}(1-v)^s \\
&+ \sum_{m=1}^{k+1}\sum_{s=0}^{m}\binom{m-1}{s-1}\binom{2k+1-m}{s-1}u^{m-s}(1-u)^s v^{2k+2-m-s}(1-v)^{s-1} \\
&+ \sum_{m=k+2}^{2k+1}\sum_{s=0}^{2k+2-m}\binom{m-1}{s-1}\binom{2k+1-m}{s-1}u^{m-s}(1-u)^s v^{2k+2-m-s}(1-v)^{s-1}.
\end{aligned}$$

By the identity $\binom{n}{s} = \binom{n-1}{s} + \binom{n-1}{s-1}$, we have $F_{2k+1}(u,v) = D_1 + D_2$, where

$$\begin{aligned}
D_1 &= \sum_{m=1}^{k+1}\sum_{s=0}^{m-1}\binom{m-2}{s}\binom{2k+1-m}{s-1}u^{m-1-s}(1-u)^s v^{2k+2-m-s}(1-v)^s \\
&+ \sum_{m=k+2}^{2k+2}\sum_{s=0}^{2k+2-m}\binom{m-2}{s}\binom{2k+1-m}{s-1}u^{m-1-s}(1-u)^s v^{2k+2-m-s}(1-v)^s \\
&+ \sum_{m=1}^{k+1}\sum_{s=0}^{m}\binom{m-2}{s-1}\binom{2k+1-m}{s-1}u^{m-s}(1-u)^s v^{2k+2-m-s}(1-v)^{s-1} \\
&+ \sum_{m=k+2}^{2k+1}\sum_{s=0}^{2k+2-m}\binom{m-2}{s-1}\binom{2k+1-m}{s-1}u^{m-s}(1-u)^s v^{2k+2-m-s}(1-v)^{s-1}
\end{aligned}$$

$$\begin{aligned}
D_2 &= \sum_{m=1}^{k+1}\sum_{s=0}^{m-1}\binom{m-2}{s-1}\binom{2k+1-m}{s-1}u^{m-1-s}(1-u)^s v^{2k+2-m-s}(1-v)^s \\
&+ \sum_{m=k+2}^{2k+1}\sum_{s=0}^{2k+2-m}\binom{m-2}{s-1}\binom{2k+1-m}{s-1}u^{m-1-s}(1-u)^s v^{2k+2-m-s}(1-v)^s \\
&+ \sum_{m=1}^{k+1}\sum_{s=0}^{m}\binom{m-2}{s-2}\binom{2k+1-m}{s-1}u^{m-s}(1-u)^s v^{2k+2-m-s}(1-v)^{s-1} \\
&+ \sum_{m=k+2}^{2k+1}\sum_{s=0}^{2k+2-m}\binom{m-2}{s-2}\binom{2k+1-m}{s-1}u^{m-s}(1-u)^s v^{2k+2-m-s}(1-v)^{s-1}.
\end{aligned}$$

By the change $m-1 \to m$ of the variable, we have

$$\begin{aligned}
D_1 &= \sum_{m=1}^{k}\sum_{s=0}^{m}\binom{m-1}{s}\binom{2k-m}{s-1}u^{m-s}(1-u)^s v^{2k+1-m-s}(1-v)^s \\
&+ \sum_{m=k+1}^{2k+1}\sum_{s=0}^{2k+1-m}\binom{m-1}{s}\binom{2k-m}{s-1}u^{m-s}(1-u)^s v^{2k+1-m-s}(1-v)^s \\
&+ \sum_{m=1}^{k}\sum_{s=0}^{m+1}\binom{m-1}{s-1}\binom{2k-m}{s-1}u^{m+1-s}(1-u)^s v^{2k+1-m-s}(1-v)^{s-1} \\
&+ \sum_{m=k+1}^{2k}\sum_{s=0}^{2k+1-m}\binom{m-1}{s-1}\binom{2k-m}{s-1}u^{m+1-s}(1-u)^s v^{2k+1-m-s}(1-v)^{s-1} = uF_{2k}(u,v).
\end{aligned}$$

$$\begin{aligned}
D_2 &= \sum_{m=0}^{k}\sum_{s=0}^{m}\binom{m-1}{s-1}\binom{2k-m}{s-1}u^{m-s}(1-u)^s v^{2k+1-m-s}(1-v)^s \\
&+ \sum_{m=k+1}^{2k}\sum_{s=0}^{2k-m+1}\binom{m-1}{s-1}\binom{2k-m}{s-1}u^{m-s}(1-u)^s v^{2k+1-m-s}(1-v)^s \\
&+ \sum_{m=1}^{k}\sum_{s=0}^{m}\binom{m-1}{s-1}\binom{2k-m}{s}u^{m-s}(1-u)^{s+1} v^{2k-m-s}(1-v)^s \\
&+ \sum_{m=k+1}^{2k}\sum_{s=0}^{2k-m}\binom{m-1}{s-1}\binom{2k-m}{s}u^{m-s}(1-u)^{s+1} v^{2k-m-s}(1-v)^s,
\end{aligned}$$

Again by the change $l = 2k+1-m \to m$ of the variable, we have

$$\begin{aligned}
D_2 &= \sum_{l=k+1}^{2k+1}\sum_{s=0}^{2k+1-l}\binom{2k-l}{s-1}\binom{l-1}{s-1}v^{l-s}(1-v)^s u^{2k+1-l-s}(1-u)^s \\
&+ \sum_{l=1}^{k}\sum_{s=0}^{l}\binom{2k-l}{s-1}\binom{l-1}{s-1}v^{l-s}(1-v)^s u^{2k+1-l-s}(1-u)^s \\
&+ \sum_{l=k+1}^{2k+1}\sum_{s=0}^{2k+1-l}\binom{2k-l}{s-1}\binom{l-1}{s}v^{l-1-s}(1-v)^s u^{2k+1-l-s}(1-u)^{s+1} \\
&+ \sum_{l=1}^{k}\sum_{s=0}^{l-1}\binom{2k-l}{s-1}\binom{l-1}{s}v^{l-1-s}(1-v)^s u^{2k+1-l-s}(1-u)^{s+1} = (1-u)F_{2k}(v,u).
\end{aligned}$$

Thus the iterative formula (14) holds for odd n. Analogously, it holds for even n. $\square$

References

[1] Chow, Y. S. and Teicher, H. (1997). *Probability Theory: Independence, Interchangeability, Martingales*, 3rd Edition, Springer.

[2] Dang, X., Keeton, S. L. and Peng, H. (2009). "A Unified Approach for Analyzing Exchangeable Binary Data with Applications to Clinical and Developmental Toxicity Studies." *Statist. Med.* **28**: 2580-2604. DOI: 10.1002/sim.3638.

[3] Diaconis, P. (1988). "Recent Progress on de Finetti's Notion of Exchangeability" in *Bayesian Statistics* **3**,111-125, eds. J. M. Bernardo, M. H. De Groot, D. V. Lindley and A. F. M. Smith, Oxford University Press, Oxford, UK.

[4] Diaconis, P. and Freedman, D. (1980a). "de Finetti's Theorem for Markov Chains." *Ann. Probabil.* **8**, 115-130.

[5] Diaconis, P. and Freedman, D. (1980b). "de Finetti's Generalizations of Exchangeability. " in *Studies in Inductive Logic ond Probability* **11**, ed. R. C. Jeffrey, Berkeley, University of California Press.

[6] Diaconis, P. and Freedman, D. (1980c). "Finite Exchangeable Sequences". *Ann. Probabil.* **8**, 745-764.

[7] Freedman, D. (1962a). "Mixtures of Markov Processes." *Ann. Math. Statist.* **33**, 114-118.

[8] Freedman, D. (1962b). "Invariants Under Mixing Which Generalize de Finetti's Theorem." *Ann. Math. Statist.* **33**, 916-923.

[9] George, E. O. and Bowman, D. (1995). "A Full Likelihood Procedure for Analysing Exchangeable Binary Data." *Biometrics* **51**, 512-523.

[10] Haviland, E. K. (1935). "On the Momentum Problem for Distribution Functions in More than One Dimension." *Amer. J. Math.* **57**, 562-568.

[11] Kendall, D. G. (1967). "On Finite and Infinite Sequences of Exchangeable Events." *Studia Scientiarum Mathematicarum Hungarica* **2**, 319-327.

[12] Peng, H., Dang, X. and Garner, L. (2009). "Analyzing Partially Exchangeable Data", *Manuscript*, Available at URL: *http://www.math.iupui.edu/∼hpeng.*

[13] Quintana, F. A. and Newton, M. A. (1998). "Assessing the Order of Dependence for Partially Exchangeable Binary Data." *J. Amer. Statist. Assoc.* **93**, 194-202.

[14] Zaman, A. (1984). "Urn Models for Markov Exchangeability." *Ann. Probabil.* **11**, 223-229.