Ann Inst Stat Math (2012) 64:633–655
DOI 10.1007/s10463-011-0325-x

Statistical modeling for discrete patterns in a sequence of exchangeable trials

Sigeo Aki

Received: 16 September 2009 / Revised: 9 August 2010 / Published online: 6 March 2011

Abstract This paper proposes a new method for constructing a sequence of infinitely exchangeable uniform random variables on the unit interval. For constructing the sequence, we utilize a Pólya urn partially. The resulting exchangeable sequence depends on the initial numbers of balls of the Pólya urn. We also derive the de Finetti measure for the exchangeable sequence. For an arbitrarily given one-dimensional distribution function, we generate sequences of exchangeable random variables with the one-dimensional marginal distribution by transforming the exchangeable uniform sequences with the inverse function of the distribution function. Among them we mainly investigate sequences of exchangeable discrete random variables. They differ from the well-known exchangeable sequence generated only by the Pólya urn scheme. Some examples are also given as applications of the results to exact distributions of some statistics based on sequences of exchangeable trials. Further, from the above exchangeable uniform sequence we construct partial or Markov exchangeable sequences. We also provide numerical examples of statistical inference based on the exchangeable and Markov exchangeable sequences.

Keywords Discrete pattern · de Finetti measure · Exchangeability · Waiting time · Generalized probability generating function · Pólya's urn · Beta distribution · Conditional expectation · Discrete distribution theory · Partial exchangeability

---

This research was partially supported by the Kansai University Grant-in-Aid for progress of research in graduate course and by Grant-in-Aid for Scientific Research (C) of the JSPS (Grant Number 22540159).

S. Aki (✉)
Department of Mathematics, Kansai University,
3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan
e-mail: aki@ipcku.kansai-u.ac.jp

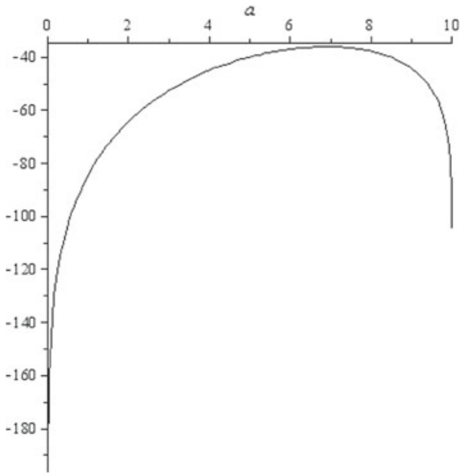
1 Introduction

Recently, some new approaches to discrete distribution theory have been successfully developed. Among them, the Markov chain embedding technique and the method of conditional probability generating functions enabled us to derive many new results on complicated enumeration problems based on various dependent trials. The Markov chain embedding technique was proposed by Fu and Koutras (1994) and developed by many researchers (see, e.g., Balakrishnan and Koutras (2002); Fu and Lou (2003)). The method of conditional probability generating functions is a natural technique of conditioning. In the literature of distribution theory, Ebneshahrashoob and Sobel (1990) used it sophisticatedly for deriving the exact distribution of the later waiting time between success and failure runs of specified length. The technique has been used by many researchers for solving complicated problems (see, e.g., Aki (2008); Aki and Hirano (2008)). Among various dependence models, exchangeable dependence is useful in some cases. For example, we can mention the reliability study of binary systems with n components if each component reliability p changes according to a random environment. Considering that the component reliability p is a random variable which values in the unit interval [0, 1] and assuming that the n components are conditionally independent and identically distributed given p, we obtain an exchangeable sequence (see Lau (1992)). Conversely, by de Finetti's theorem, if {0, 1}-valued random variables are infinitely exchangeable, there exists a random variable p which values in the unit interval such that the random variables are conditionally i.i.d. given p (see, e.g., de Finetti (1975); Durrett (2005); Billingsley (1995)). Therefore, the above reliability modeling is very general. We can find out other examples of exchangeable dependence modeling in George and Bowman (1995); Kolev et al. (2006); Kolev and Paiva (2008) and Eryilmaz and Demir (2007).

An infinite sequence of random variables $\{X_n\}$ is (infinitely) exchangeable if for all n the joint distribution of $(X_1, X_2, \dots, X_n)$ is invariant under permutations. By de Finetti's theorem, if $\{X_n\}$ is (infinitely) exchangeable, it is a mixture of i.i.d. sequences with a directing random measure, which is sometimes called the de Finetti measure (see, e.g., Aldous (1985)).

In the manuscript we study statistical inference of the directing random measure. Here, we suppose that our observations are sequences of random length from an exchangeable population. To understand the statistical problem, let us consider a simple example of Pólya's urn which contains a white and $(10 - a)$ red balls initially. Balls are randomly drawn one at a time and each ball is returned to the urn along with one additional ball of the same color. Based on the observations of waiting time (number of trials) until two consecutive white balls are drawn, we can estimate the value of the initial number of white balls a. For example, the waiting time is 7 if we observe the color of the balls drawn as (white, red, red, white, red, white, white). Here, when two consecutive white balls are drawn, the experiment is assumed to be repeated again from the initial situation of a white and $(10 - a)$ red balls. It is known that Pólya sequence is infinitely exchangeable and the directing random measure is described by the beta distribution $Beta(a, 10 - a)$. In this simple case, the directing random measure is parameterized by a. By estimating a, we can investigate the random measure which is directing the exchangeable distribution on $\{white, red\}^{\infty}$.

Fig. 1 The graph of the log-likelihood function based on the data set of the waiting time of the first occurrence of consecutive white balls



The next data set of waiting time for two consecutive white balls is simulated assuming that $a = 7$,

4, 2, 2, 8, 4, 2, 4, 2, 4, 2, 4, 9, 7, 2, 5, 4, 4, 2, 2, 4.

The maximum likelihood estimate of a based on the data can be calculated as $\hat{a} = 6.898$ and Fig. 1 shows the log-likelihood function of a.

The purpose of the manuscript is to propose some tractable parametric models for sequences of exchangeable trials like the above Pólya sequence. Generally, the characteristics (such as moments, probability generating function, etc.) of sample distributions of statistics based on the exchangeable binary random variables can be obtained by integrating the characteristics of the corresponding distributions of the statistics based on i.i.d. sequence using the de Finetti measure. It seems easy to assume a distribution with some parameters as the de Finetti measure. However, the above integration must be performed not numerically but analytically since we need the likelihood of the parameters for statistical inference. This has been an inevitable restriction for statistical inference and the distributions except for the beta distribution have been scarcely used for the purpose as far as the author knows. It is also well-known that we can construct sequentially an exchangeable binary sequence by using Pólya's urn, and the corresponding de Finetti measure is the beta distribution (see, e.g., Irwin (1954); Kemp and Kemp (1956); Freedman (1965)). Moreover, it should be noticed that the sample distributions can also be derived recursively without integration using the method of conditional probability generating functions by virtue of the sequential mechanism of Pólya's urn (see Inoue and Aki (2005)).

In Sect. 2 we propose a new method for constructing a sequence of infinitely exchangeable random variables whose one-dimensional marginal distribution is the uniform distribution on the unit interval [0, 1]. In Sect. 3 we treat the sequences of exchangeable random variables based on the method proposed in Sect. 2. It is shown that the de Finetti measures of our exchangeable discrete random variables often become appropriately scaled beta distribution. Even when the de Finetti measure becomes complicated in appearance, integration with the measure can be performed easily in many cases. In Sect. 4 we show that Markov exchangeable sequences can be constructed sequentially using our exchangeable uniform sequence proposed in Sect. 2. In Sect. 5 we provide illustrative examples for statistical inference in order to show the feasibility of our results. In the examples, we estimate simultaneously two parameters, one of which determines the de Finetti measure of the underlying exchangeable sequence.

2 Constructing exchangeable random variables using an urn

Let us consider an urn containing a white and b red balls. Assume that balls are randomly drawn one at a time and each ball is returned to the urn along with one additional ball of the same color before the next ball is drawn. Then, the sequence of random variables $\{X_n\}$ is defined as

$$X_n = \begin{cases} 0 & \text{if the } n\text{-th ball is white} \\ 1 & \text{if the } n\text{-th ball is red.} \end{cases}$$

Let $\{U_n\}$ be a sequence of independent uniformly distributed random variables on the unit interval $[0, 1]$. We also assume that $\{X_n\}$ and $\{U_n\}$ are independent. For $i = 1, 2, \ldots$, we set

$$Y_i = (1 - X_i)\frac{a}{a+b}U_i + X_i\left(\frac{a}{a+b} + \frac{b}{a+b}U_i\right).$$

Then, $\{Y_n\}$ is exchangeable since $\{X_n\}$ is exchangeable.

Proposition 1 *Each Y_n follows the uniform distribution on the unit interval* $[0, 1]$.

Proof Since $\{Y_n\}$ is exchangeable, it suffices to show that Y_1 follows the uniform distribution. We see that for $0 \le y \le \frac{a}{a+b}$,

$$\begin{aligned} P(Y_1 \le y) &= P(X_1 = 0)P\left(\frac{a}{a+b}U_1 \le y\right) \\ &\quad + P(X_1 = 1)P\left(\frac{a}{a+b} + \frac{b}{a+b}U_1 \le y\right) \\ &= \frac{a}{a+b} \cdot \frac{a+b}{a}y + \frac{b}{a+b} \cdot 0 = y \end{aligned}$$

and for $\frac{a}{a+b} < y \le 1$,

$$
\begin{aligned}
P(Y_1 \leq y) &= P(X_1 = 0)P\left(\frac{a}{a+b}U_1 \leq y\right) \\
&\quad + P(X_1 = 1)P\left(\frac{a}{a+b} + \frac{b}{a+b}U_1 \leq y\right) \\
&= \frac{a}{a+b} \cdot 1 + \frac{b}{a+b} \cdot \frac{a+b}{b}\left(y - \frac{a}{a+b}\right) = y.
\end{aligned}
$$

This completes the proof. □

For $0 < z < 1$, we define

$$
h(z; x) = \frac{a+b}{a} x z I_{\left[0, \frac{a}{a+b}\right]}(x) + I_{\left(\frac{a}{a+b}, 1\right]}(x)\left\{z + \frac{a+b}{b}\left(x - \frac{a}{a+b}\right)(1-z)\right\}.
$$

Let A, B and C be the points with the coordinates $(0, 0)$, $(\frac{a}{a+b}, z)$, and $(1, 1)$, respectively. Then $h(z; x)$ is the distribution function whose graph is piecewise linear with the segments AB and BC. We have the next theorem.

Theorem 1 *There exists a random variable $Z = Z(\omega)$ such that Z follows the beta distribution $Beta(a, b)$ and the empirical distribution function $F_n(t)$ of $Y_1, \ldots, Y_n$ converges to the random distribution function $h(Z(\omega); t)$ uniformly in t with probability one as n tends to infinity. Further, given $h(Z(\omega); t)$, $Y_1, Y_2, \ldots$ are conditionally independent random variables whose distribution functions are $h(Z(\omega); t)$.*

Proof From de Finetti's theorem (see, e.g., Johnson and Kotz (1977); Kingman (1978); Bertoin (2006); Pitman (2006)), we see that $F_n(t)$ converges to a random distribution function $\varphi(t)$ with probability one uniformly in t as $n \to \infty$. Further, given $\varphi(t)$, $Y_1, Y_2, \ldots$ are conditionally independent random variables whose distribution functions are $\varphi(t)$. But, from the definition of Y_i, we can write

$$
F_n\left(\frac{a}{a+b}\right) = \frac{\#\{i : X_i = 0, i = 1, 2, \ldots, n\}}{n}.
$$

If we set $Z = \varphi(\frac{a}{a+b})$, Z follows the beta distribution $Beta(a, b)$ (see, e.g., (Durrett, 2005, p. 238)). Noting that $\{X_n\}$ and $\{U_n\}$ are independent, we observe that given $Z = \varphi(\frac{a}{a+b})$, conditional distributions of $\varphi(t)$ on the intervals $[0, \frac{a}{a+b}]$ and $[\frac{a}{a+b}, 1]$ are uniform. Therefore, $\varphi(t)$ is piecewise linear and hence $\varphi(t) = h(Z; t)$ holds. □

Let us obtain the joint distribution of $(Y_1, Y_2, \ldots, Y_m)$. To calculate $P(Y_1 \leq y_1, Y_2 \leq y_2, \ldots, Y_m \leq y_m)$, without loss of generality, we assume $y_1, \ldots, y_k \in [0, \frac{a}{a+b}]$ and $y_{k+1}, \ldots, y_m \in (\frac{a}{a+b}, 1]$, since $(Y_1, Y_2, \ldots, Y_m)$ are exchangeable.

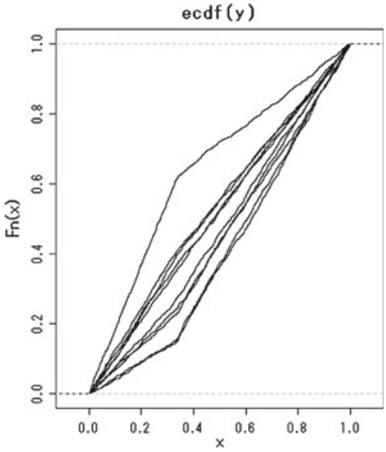


Fig. 2 Sample paths of empirical distribution functions of $\{Y_n\}$ of size 2,000 with $a = 5$ and $b = 10$

Proposition 2 *For* $y_1, \dots, y_k \in \left[0, \frac{a}{a+b}\right]$ *and* $y_{k+1}, \dots, y_m \in (\frac{a}{a+b}, 1]$,

$$P(Y_1 \le y_1, \dots, Y_m \le y_m) = \left(\prod_{j=1}^{k} \frac{a+b}{a} y_j\right) \sum_{\ell=0}^{m-k} \sum_{\substack{L \subset \{k+1,\dots,m\} \\ |L| = \ell}} \prod_{j \in L} \left(\frac{a+b}{b}\left(y_j - \frac{a}{a+b}\right)\right) \times \frac{(a)_{\uparrow(m-\ell)}(b)_{\uparrow\ell}}{(a+b)_{\uparrow m}}$$

holds, where $(c)_{\uparrow d} = c(c+1)\dots(c+d-1)$.

We illustrate in Fig. 2 ten sample paths of empirical distribution functions of $Y's$ of size 2,000 starting from the urn with $a = 5$ and $b = 10$.

Proof From Theorem 1, we can calculate the joint distribution as follows.

$$\begin{aligned} P(Y_1 \le y_1, \dots, Y_m \le y_m) &= \int_0^1 \left(\prod_{j=1}^{k} \frac{a+b}{a} y_j z\right) \prod_{j=k+1}^{m} \left(z + \frac{a+b}{b}\left(y_j - \frac{a}{a+b}\right)(1-z)\right) \\ &\quad \times \frac{1}{B(a,b)} z^{a-1}(1-z)^{b-1} dz \\ &= \left(\prod_{j=1}^{k} \frac{a+b}{a} y_j\right) \int_0^1 \prod_{j=k+1}^{m} \left(z + \frac{a+b}{b}\left(y_j - \frac{a}{a+b}\right)(1-z)\right) \\ &\quad \times \frac{1}{B(a,b)} z^{a+k-1}(1-z)^{b-1} dz \end{aligned}$$

$$= \left(\prod_{j=1}^{k} \frac{a+b}{a} y_j \right) \sum_{\ell=0}^{m-k} \sum_{\substack{L \subset \{k+1,\ldots,m\} \\ |L|=\ell}} \prod_{j \in L} \left(\frac{a+b}{b} \left(y_j - \frac{a}{a+b} \right) \right)$$

$$\times \int_0^1 \frac{1}{B(a,b)} z^{a+k+m-k-\ell-1} (1-z)^{b+\ell-1} dz$$

$$= \left(\prod_{j=1}^{k} \frac{a+b}{a} y_j \right) \sum_{\ell=0}^{m-k} \sum_{\substack{L \subset \{k+1,\ldots,m\} \\ |L|=\ell}}$$

$$\times \prod_{j \in L} \left(\frac{a+b}{b} \left(y_j - \frac{a}{a+b} \right) \right) \frac{(a)_{\uparrow(m-\ell)} (b)_{\uparrow \ell}}{(a+b)_{\uparrow m}}.$$

This completes the proof. □

For a one-dimensional distribution function $G(x)$, we define

$$G^{-1}(t) = \inf\{x : G(x) \geq t\} \quad \text{for } 0 < t < 1$$

and we set $W_n = G^{-1}(Y_n)$. Since $\{W_n \leq x\} = \{Y_n \leq G(x)\}$ holds for all x, we obtain the next two results from Theorem 1 and Proposition 2.

Theorem 2 *$\{W_n\}$ is a sequence of exchangeable random variables and the distribution function of W_n is $G(x)$ for each n. Further, there exists a random variable $Z = Z(\omega)$ with the beta distribution $Beta(a,b)$ such that the empirical distribution function $G_n(x)$ of $W_1, \ldots, W_n$ converges uniformly in x to the random distribution function $h(Z(\omega); G(x))$ with probability one as $n \to \infty$, and given $h(Z(\omega); G(x))$, $\{W_n\}$ is a sequence of conditionally independent random variables with distribution function $h(Z(\omega); G(x))$.*

Proposition 3 *For $G(w_1), \ldots, G(w_k) \in [0, \frac{a}{a+b}]$, and $G(w_{k+1}), \ldots, G(w_m) \in (\frac{a}{a+b}, 1]$, the joint distribution of $W_1, \ldots, W_m$ can be written as*

$$P(W_1 \leq w_1, \ldots, W_m \leq w_m)$$
$$= \left(\prod_{j=1}^{k} \frac{a+b}{a} G(w_j) \right) \sum_{\ell=0}^{m-k} \sum_{\substack{L \subset \{k+1,\ldots,m\} \\ |L|=\ell}}$$
$$\times \prod_{j \in L} \left(\frac{a+b}{b} \left(G(w_j) - \frac{a}{a+b} \right) \right) \frac{(a)_{\uparrow(m-\ell)} (b)_{\uparrow \ell}}{(a+b)_{\uparrow m}}.$$

We can calculate the moments of $\{W_n\}$. Let us obtain the second moment.

Proposition 4 *Since Y_n follows the uniform distribution on the unit interval, the distribution function of $W_n (= G^{-1}(Y_n))$ is G. Then, the moments of $W_n (= G^{-1}(Y_n))$*

is the moments of G if they exist. Setting $G(a,b) = G^{-1}\left(\frac{a}{a+b}\right)$*,* $E(W_1W_2)$ *can be written as follows.*

$$\begin{aligned}
&E(W_1W_2)\\
&= \frac{a(a+1)}{(a+b)(a+b+1)}\left(\frac{a+b}{a}\int_{-\infty}^{G(a,b)} x dG(x)\right)^2\\
&\quad + \frac{2ab}{(a+b)(a+b+1)}\left(\frac{a+b}{a}\int_{-\infty}^{G(a,b)} x dG(x)\right)\left(\frac{a+b}{b}\int_{G(a,b)}^{\infty} x dG(x)\right)\\
&\quad + \frac{b(b+1)}{(a+b)(a+b+1)}\left(\frac{a+b}{b}\int_{G(a,b)}^{\infty} x dG(x)\right)^2.
\end{aligned}$$

In particular, by setting $G(t) = t \quad (0 < t < 1)$*, we obtain the second moment of* $\{Y_n\}$ *as*

$$E(Y_1Y_2) = \frac{(a+b)^3 + a^2 + 3ab + b^2}{4(a+b)^2(a+b+1)}.$$

Further, as Y_n *follows the uniform distribution on the unit interval,* $E(Y_n) = \frac{1}{2}$ *and* $Var(Y_n) = \frac{1}{12}$ *hold. Thus, we have*

$$Cov(Y_1, Y_2) = \frac{ab}{4(a+b)^2(a+b+1)},$$

and

$$Cor(Y_1, Y_2) = \frac{3ab}{(a+b)^2(a+b+1)}.$$

Proof Since $\{X_n\}$ and $\{U_n\}$ are independent, it holds that

$$\begin{aligned}
&E(W_1W_2)\\
&= P(X_1 = 0, X_2 = 0)E\left[G^{-1}\left(\frac{a}{a+b}U_1\right)G^{-1}\left(\frac{a}{a+b}U_2\right)\right]\\
&\quad + P(X_1 = 0, X_2 = 1)E\left[G^{-1}\left(\frac{a}{a+b}U_1\right)G^{-1}\left(\frac{a}{a+b} + \frac{b}{a+b}U_2\right)\right]\\
&\quad + P(X_1 = 1, X_2 = 0)E\left[G^{-1}\left(\frac{a}{a+b} + \frac{b}{a+b}U_1\right)G^{-1}\left(\frac{a}{a+b}U_2\right)\right]\\
&\quad + P(X_1 = 1, X_2 = 1)\\
&\quad \times E\left[G^{-1}\left(\frac{a}{a+b} + \frac{b}{a+b}U_1\right)G^{-1}\left(\frac{a}{a+b} + \frac{b}{a+b}U_2\right)\right].
\end{aligned}$$

Noting that U_1 and U_2 are independent, we have the desired results. □

Let us derive the distribution of the order statistics $W_{(1)} \leq W_{(2)} \leq \cdots \leq W_{(m)}$ of $W_1, W_2, \ldots, W_m$.

Proposition 5 *For* $1 \leq j \leq m$, *if* $G(w) \leq \frac{a}{a+b}$, *then*

$$
\begin{aligned}
&P(W_{(j)} \leq w) \\
&= \sum_{r=j}^{m} \sum_{k=0}^{m-r} (-1)^k \binom{m}{r} \binom{m-r}{k} \left(\frac{a+b}{a} G(w) \right)^{r+k} \frac{B(a+r+k, b)}{B(a, b)},
\end{aligned}
$$

and if $G(w) > \frac{a}{a+b}$, *then*

$$
\begin{aligned}
&P(W_{(j)} \leq w) \\
&= \sum_{r=j}^{m} \sum_{k=0}^{r} \binom{m}{r} \binom{r}{k} (1 - A(w))^{k+m-r} A(w)^{r-k} \frac{B(a+k, b+m-r)}{B(a, b)},
\end{aligned}
$$

where $A(w) = \frac{a+b}{b} \left(G(w) - \frac{a}{a+b} \right)$.

Proof Since $W_1, \ldots, W_m$ are exchangeable,

$$
\begin{aligned}
&P(W_{(j)} \leq w) \\
&= \sum_{r=j}^{m} \binom{m}{r} P(W_1 \leq w, \ldots, W_r \leq w, W_{r+1} > w, \ldots, W_m > w).
\end{aligned}
$$

If $G(w) \leq \frac{a}{a+b}$, we observe

$$
\begin{aligned}
&P(W_1 \leq w, \ldots, W_r \leq w, W_{r+1} > w, \ldots, W_m > w) \\
&= P(Y_1 \leq G(w), \ldots, Y_r \leq G(w), Y_{r+1} > G(w), \ldots, Y_m > G(w)) \\
&= \int_0^1 \left(\prod_{j=1}^{r} \frac{a+b}{a} G(w) z \right) \left(\prod_{j=r+1}^{m} (1 - \frac{a+b}{a} G(w) z) \right) \\
&\quad \times \frac{1}{B(a, b)} z^{a-1} (1-z)^{b-1} dz \\
&= \sum_{k=0}^{m-r} (-1)^k \binom{m-r}{k} \left(\frac{a+b}{a} G(w) \right)^{r+k} \frac{B(a+r+k, b)}{B(a, b)}.
\end{aligned}
$$

If $G(w) > \frac{a}{a+b}$, we have

$$
\begin{aligned}
&P(W_1 \leq w, \dots, W_r \leq w, W_{r+1} > w, \dots, W_m > w) \\
&\quad = P(Y_1 \leq G(w), \dots, Y_r \leq G(w), Y_{r+1} > G(w), \dots, Y_m > G(w)) \\
&\quad = \int_0^1 (A(w) + (1 - A(w))z)^r (1 - A(w) - (1 - A(w))z)^{m-r} \\
&\qquad \times \frac{1}{B(a,b)} z^{a-1}(1-z)^{b-1} dz \\
&\quad = \sum_{k=0}^{r} \binom{r}{k} (1 - A(w))^{k+m-r} A(w)^{r-k} \frac{B(a+k, b+m-r)}{B(a,b)}.
\end{aligned}
$$

This completes the proof. □

Corollary 1 *Let $M = \max_{1 \leq j \leq m} W_j$. Then, the distribution function of the maximum M can be written by using a random variable ξ, which follows the beta distribution $Beta(a, b)$, as*

$$
P(M \leq w) = \begin{cases} E\left[\left(\frac{a+b}{a} G(w)\xi\right)^m\right] & \text{if } G(w) \leq \frac{a}{a+b} \\ E\left[\left\{\xi + \frac{a+b}{b}\left(G(w) - \frac{a}{a+b}\right)(1-\xi)\right\}^m\right] & \text{if } G(w) > \frac{a}{a+b} \end{cases}.
$$

Proof From Proposition 5, we see that

$$
\begin{aligned}
&P(M \leq w) \\
&\quad = \begin{cases} \left(\frac{a+b}{a} G(w)\right)^m \frac{(a)_{\uparrow m}}{(a+b)_{\uparrow m}} & \text{if } G(w) \leq \frac{a}{a+b} \\ \sum_{\ell=0}^{m} \frac{(a)_{\uparrow(m-\ell)}(b)_{\uparrow \ell}}{(a+b)_{\uparrow m}} \left\{\frac{a+b}{b}\left(G(w) - \frac{a}{a+b}\right)\right\}^\ell & \text{if } G(w) > \frac{a}{a+b} \end{cases}.
\end{aligned}
$$

Then, noting that

$$
E[\xi^\ell (1-\xi)^m] = \frac{(a)_{\uparrow \ell}(b)_{\uparrow m}}{(a+b)_{\uparrow(m+\ell)}}
$$

for every positive integers ℓ and m, we have the desired result. □

Figure 3 shows the distribution function of the maximum of five exchangeable variables whose one-dimensional marginal distributions are exponential of mean 1 with $a = 1$ and $b = 1$. We observe that the shape of the c.d.f. changes at $w = -\log(\frac{b}{a+b})$ though it is continuous for all w.

Remark 1 Let $\{\xi_n\}$ be a sequence of $\{0, 1, 2\}$-valued random variables which are generated by the Pólya sampling from an urn with a, b, and c balls respectively with color code 0, 1, and 2, initially. Let $\{U_n\}$ be a sequence of independent uniformly distributed

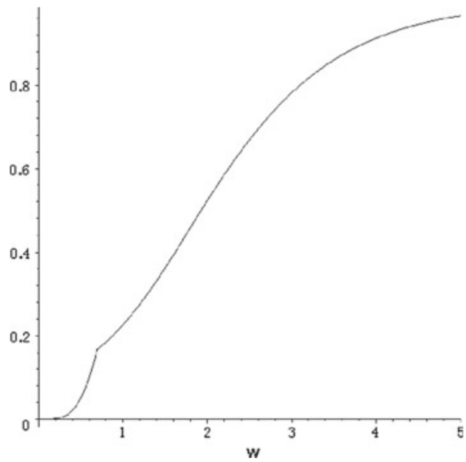


Fig. 3 The distribution function of the maximum of five exchangeable random variables whose one-dimensional marginal distributions are the exponential of mean 1 with $a = 1$ and $b = 1$

random variables on the unit interval. Further, $\{\xi_n\}$ and $\{U_n\}$ are assumed to be independent. Then, similarly as Proposition 1, we can define a sequence of exchangeable uniform random variables $\{\eta_n\}$ by

$$\eta_n = \begin{cases} \frac{aU_n}{a+b+c} & \text{if } \xi_0 = 0 \\ \frac{a}{a+b+c} + \frac{bU_n}{a+b+c} & \text{if } \xi_n = 1 \\ \frac{a+b}{a+b+c} + \frac{cU_n}{a+b+c} & \text{if } \xi_n = 2 \end{cases}$$

It is not difficult to see asymptotic behavior of the empirical distribution function of $\eta_1, \ldots, \eta_n$ by extending Theorem 1.

3 Discrete models for exchangeable trials

It is well-known that a two-color Pólya urn generates a sequence of exchangeable binary trials sequentially and the corresponding de Finetti measure is a beta distribution. It is also known that $(k+1)$-color Pólya urn generates a sequence of exchangeable $\{0, 1, 2, \ldots, k\}$-valued random variables sequentially and the corresponding de Finetti measure is a Dirichlet distribution (e.g. Hill et al. (1987); Mauldin et al. (1992)). In this section we construct a sequence of exchangeable $\{0, 1\}$-valued random variables sequentially using $\{Y_n\}$ defined in the previous section. The de Finetti measure for the new sequence of exchangeable binary variables is not a beta distribution but a properly scaled beta distribution.

First, we transform $\{Y_n\}$ to $\{Z_n\} = \{G_p^{-1}(Y_n)\}$, where $G_p(x)$ is the cumulative distribution function on $\{0, 1\}$,

$$G_p(t) = \begin{cases} 0 & \text{if } t < 0 \\ p & \text{if } 0 \le t < 1 \\ 1 & \text{if } 1 \le t. \end{cases}$$

The de Finetti measure of the sequence $\{Z_n\}$ is the distribution of a random variable Θ, which values on the unit interval [0, 1], and $\frac{S_n}{n}$ converges to Θ with probability one, where $S_n = Z_1 + Z_2 + \cdots + Z_n$. Further, given the random variable Θ, the sequence $\{Z_n\}$ is conditionally i.i.d. and each Z_n follows the distribution of Θ.

Theorem 3 *The density of the de Finetti measure of* $\{Z_n\}$ *is given as follows.*

(1) *When* $0 < p \le \frac{a}{a+b}$, *the de Finetti measure of* $\{Z_n\}$ *has a support on the interval* $\left(1 - \frac{a+b}{a}p, 1\right)$ *and the density w.r.t. Lebesgue measure is*

$$f_1(x) = \frac{1}{\alpha B(a,b)} \left(\frac{1}{\alpha}(1-x)\right)^{a-1} \left(1 - \frac{1}{\alpha}(1-x)\right)^{b-1} \quad (1-\alpha < x < 1),$$

where, $\alpha = \frac{a+b}{a}p$.

(2) *When* $\frac{a}{a+b} < p < 1$, *if we set* $\beta = 1 - \frac{a+b}{b}(p - \frac{a}{a+b})$, *the de Finetti measure of* $\{Z_n\}$ *has a support on the interval* $(0, \beta)$ *and the density w.r.t. Lebesgue measure is*

$$f_2(x) = \frac{1}{\beta B(b,a)} \left(\frac{1}{\beta}x\right)^{b-1} \left(1 - \frac{1}{\beta}x\right)^{a-1} \qquad (0 < x < \beta).$$

Proof Denoting the de Finetti measure by μ we can write $P(Z_1 = 0, \ldots, Z_m = 0) = \int_0^1 (1-\theta)^m d\mu(\theta)$, since $P(Z_1 = 1, \ldots, Z_m = 1) = \int_0^1 \theta^m d\mu(\theta)$. Then, letting Θ be a random variable with the distribution μ, we can write $P(Z_1 = 0, \ldots, Z_m = 0) = E[(1-\Theta)^m]$. Further, by using Proposition 2, we can also write $P(Z_1 = 0, \ldots, Z_m = 0)$ as

$$\begin{aligned} &P(Z_1 = 0, \ldots, Z_m = 0) \\ &\quad = P(Y_1 \le p, \ldots, Y_m \le p) \\ &\quad = \begin{cases} \frac{(a)_{\uparrow m}}{(a+b)_{\uparrow m}} \cdot \left(\frac{a+b}{a}p\right)^m & \text{if } 0 \le p \le \frac{a}{a+b} \\ \sum_{\ell=0}^m \binom{m}{\ell} \frac{(a)_{\uparrow(m-\ell)}(b)_{\uparrow \ell}}{(a+b)_{\uparrow m}} \left\{\frac{a+b}{b}\left(p - \frac{a}{a+b}\right)\right\}^\ell & \text{if } \frac{a}{a+b} < p \le 1 \end{cases}. \end{aligned}$$

Here, we note that denoting by ξ a random variable with the beta distribution $Beta(a, b)$,

$$E[\xi^\ell (1-\xi)^m] = \frac{(a)_{\uparrow \ell}(b)_{\uparrow m}}{(a+b)_{\uparrow(m+\ell)}}$$

holds for every pair of positive integers ℓ and m. Then the r.h.s. of the above formula can be written as $E[(\frac{a+b}{a}p\xi)^m]$ when $0 \le p \le \frac{a}{a+b}$, and as $E[\{\xi + \frac{a+b}{b}(p - \frac{a}{a+b})(1-\xi)\}^m]$ when $\frac{a}{a+b} < p \le 1$. Since the result holds for every integer m, by the moment problem, the distribution is determined uniquely. Consequently, we see that the distribution of $1-\Theta$ is the same as that of $\frac{a+b}{a}p\xi$ when $0 \le p \le \frac{a}{a+b}$, and $\xi + \frac{a+b}{b}(p - \frac{a}{a+b})(1-\xi)$ when $\frac{a}{a+b} < p \le 1$. This completes the proof. □

Remark 2 In Theorem 3, if $p = \frac{a}{a+b}$ holds, the de Finetti measure becomes a beta distribution, since

$$G_p^{-1}(t) = \begin{cases} 0 & \text{if } 0 < t \leq p \\ 1 & \text{if } p < t < 1 \end{cases}$$

holds and hence we see that $Z_n = G_p^{-1}(Y_n) = X_n$. Thus, in the case, to observe the sequence $\{Z_n\}$ is equivalent to observe the Pólya urn. Consequently, the sequence $\{Z_n\}$ is a proper extension of the Pólya urn scheme.

One of the merits of the new model $\{Z_n\}$ is that the sequence can be generated sequentially like the Pólya urn scheme. In the following examples, we show that the method of conditional probability generating functions can be applied to derive sample distributions based on the sequence $\{Z_n\}$ by virtue of the sequential construction of $\{Z_n\}$.

Example 1 We show two methods for deriving the exact sample distribution of $S_n = \sum_{i=1}^{n} Z_i$ when $0 < p \leq \frac{a}{a+b}$. First, starting from an urn with w white balls and r red balls, we repeat the trials m times in the Pólya urn scheme. After the m-th trial, we denote by $\varphi_1(m, w, r)$ the conditional probability generating function of the number of 1's. Then, the next recurrence relations hold by virtue of the sequential construction of Z_n.

$$\begin{cases} \varphi_1(0, w, r) = 1 \\ \varphi_1(m, w, r) = \frac{w}{w+r}((1-\alpha)t + \alpha)\varphi_1(m-1, w+1, r), \\ \qquad\qquad\quad + \frac{r}{w+r} t\varphi_1(m-1, w, r+1) \end{cases}$$

where $\alpha = \frac{a+b}{a}p$. To be precise, we observe from the definition of $\varphi_1(m, w, r)$,

$$\begin{aligned} &\varphi_1(m, w, r) \\ &\quad = P(X_1 = 0)(P(Z_1 = 1|X_1 = 0)t + P(Z_1 = 0|X_1 = 0)) \\ &\qquad \times \varphi_1(m-1, w+1, r) \\ &\qquad + P(X_1 = 1)(P(Z_1 = 1|X_1 = 1)t + P(Z_1 = 0|X_1 = 1)) \\ &\qquad \times \varphi_1(m-1, w, r+1). \end{aligned}$$

Noting that

$$\begin{aligned} P(Z_1 = 1|X_1 = 0) &= P\left(\frac{a}{a+b}U_1 > p\right) = 1 - \alpha, \\ P(Z_1 = 0|X_1 = 0) &= P\left(\frac{a}{a+b}U_1 \leq p\right) = \alpha, \\ P(Z_1 = 1|X_1 = 1) &= P\left(\frac{a}{a+b} + \frac{b}{a+b}U_1 > p\right) = 1 \quad \text{and} \\ P(Z_1 = 0|X_1 = 1) &= P\left(\frac{a}{a+b} + \frac{b}{a+b}U_1 \leq p\right) = 0, \end{aligned}$$

we have the above recurrence relations. By solving them recursively, we can obtain the probability generating function $\varphi_1(n, a, b)$ for arbitrarily given n.

Next, by using de Finetti's theorem, we obtain $\varphi_1(n, a, b)$ as

$$\varphi_1(n, a, b) = \int_{1-\alpha}^{1} (xt + (1 - x))^n f_1(x)dx,$$

where $f_1(x)$ is the density given in (1) of Theorem 3. In the integral, changing variables $\frac{(1-x)}{\alpha} = u$, the integral can be written as

$$\int_0^1 ((1 - \alpha u)t + \alpha u)^n \frac{1}{B(a, b)} u^{a-1}(1 - u)^{b-1} du.$$

If we expand $((1-\alpha u)t+\alpha u)^n$ in the above integral, the integration can be performed and we see that the result is the same as the result obtained by solving the recurrence relations above.

Further, using the next formula for the hypergeometric function

$$F(\kappa, \lambda, \mu; z) = \frac{\Gamma(\mu)}{\Gamma(\lambda)\Gamma(\mu - \lambda)} \int_0^1 t^{\lambda-1}(1 - t)^{\mu-\lambda-1}(1 - tz)^{-\kappa} dt \tag{1}$$
$$(\Re(\mu) > \Re(\lambda) > 0), \text{ where } \Re(z) \text{ denotes the real part of } z,$$

we can also write $\varphi_1(n, a, b)$ as

$$\varphi_1(n, a, b) = t^n F\left(-n, a, a + b; \alpha\left(1 - \frac{1}{t}\right)\right).$$

Example 2 Like the previous example, when $\frac{a}{a+b} < p < 1$, we derive the exact distribution of $S_n = \sum_{i=1}^n Z_i$ by the two methods. First, we shall explain how to obtain the exact distribution by using recurrence relations of conditional probability generating functions. Starting from an urn with w white balls and r red balls, we repeat the trials m times in the Pólya urn scheme. After the m-th trial, we denote by $\varphi_2(m, w, r)$ the conditional probability generating function of the number of 1's. Then, similarly as Example 1, we can see that $\varphi_2(m, w, r)$ satisfies the next recurrence relations:

$$\begin{cases} \varphi_2(0, w, r) = & 1 \\ \varphi_2(m, w, r) = & \frac{w}{w+r}\varphi_2(n - 1, w + 1, r) \\ & + \frac{r}{w+r}((\beta t + (1 - \beta))\varphi_2(m - 1, w, r + 1), \end{cases}$$

where $\beta = 1 - \frac{a+b}{b}\left(p - \frac{a}{a+b}\right)$. By solving them recursively, we can obtain the probability generating function $\varphi_2(n, a, b)$.

Next, by using de Finetti's theorem, we can write $\varphi_2(n, a, b)$ directly as

$$\varphi_2(n, a, b) = \int_0^{\beta} (xt + (1 - x))^n f_2(x)dx,$$

where $f_2(x)$ is the density given in (2) of Theorem 3. Changing variables $\frac{x}{\beta} = u$, we obtain

$$\int_0^1 (\beta u t + (1 - \beta u))^n \frac{1}{B(b, a)} u^{b-1} (1 - u)^{a-1} du.$$

By expanding $(\beta u t + (1 - \beta u))^n$, we can perform the above integration for arbitrarily given n, and we can write $\varphi_2(n, a, b)$ as a polynomial in t. Furthermore, by using the formula of the hypergeometric function (1), we can also represent $\varphi_2(n, a, b)$ as

$$\varphi_2(n, a, b) = F(-n, b, a + b; \beta(1 - t)).$$

Although the next example is essentially equivalent to the previous examples, we state the next example in order to compare our results with those previously investigated in the literature of discrete distribution theory.

Example 3 By using the sequence of exchangeable uniform random variables $\{Y_n\}$, we define the exchangeable $\{0, 1\}$-valued random variables $\{V_n\}$ as

$$V_n = \begin{cases} 1 & \text{if } Y_n \le p \\ 0 & \text{if } Y_n > p \end{cases}$$

For simplicity, we assume that $0 < p \le \frac{a}{a+b}$. Though $\{V_n\}$ is essentially equivalent to $\{Z_n\}$, we give the de Finetti measure, since it slightly differs from that of $\{Z_n\}$. First, we note that

$$\begin{aligned} &P(V_1 = 1, V_2 = 1, \dots, V_m = 1) \\ &\quad = P(X_1 = 0, X_2 = 0, \dots, X_m = 0) P\left(\frac{aU_1}{a+b} \le p, \dots, \frac{aU_m}{a+b} \le p\right) \\ &\quad = \frac{(a)_{\uparrow m}}{(a+b)_{\uparrow m}} \left(\frac{a+b}{a} p\right)^m . \end{aligned}$$

Denoting by μ the de Finetti measure of $\{V_n\}$, let Θ be a random variable which follows μ and let ξ be a random variable with beta distribution $Beta(a, b)$. For every $m = 1, 2 \dots$, we have

$$\begin{aligned} P(V_1 = 1, V_2 = 1, \dots, V_m = 1) &= \int_0^1 P(V_1 = 1, \dots, V_m = 1|\theta) d\mu(\theta) \\ &= \int_0^1 \theta^m d\mu(\theta) \\ &= \frac{(a)_{\uparrow m}}{(a+b)_{\uparrow m}} \left(\frac{a+b}{a} p\right)^m \\ &= E\left[\left(\frac{a+b}{a} p\xi\right)^m\right]. \end{aligned}$$

Therefore, the distributions of Θ and $\frac{a+b}{a}p\xi$ are the same and hence the density function of the de Finetti measure μ can be written as

$$f_3(u) = \frac{1}{\alpha}\frac{1}{B(a,b)}\left(\frac{u}{\alpha}\right)^{a-1}\left(1-\frac{u}{\alpha}\right)^{b-1} \quad (0 < u < \alpha \le 1), \tag{2}$$

where $\alpha = \frac{a+b}{a}p$.

Proposition 6 *Let τ be the number of trials until the first occurrence of "1"-run of length k in the sequence $\{V_n\}$ in Example 3. Then the probability generating function of τ can be written as*

$$\begin{aligned}E[t^{\tau}] = &\sum_{x=k}^{\infty}\sum_{x_1+2x_2+\cdots+kx_k=x-k}\binom{x_1+\ldots x_k}{x_1,\ldots,x_k}\alpha^{x-(x_1+\cdots+x_k)}\\ &\times\frac{\Gamma(a+b)}{\Gamma(a)}\frac{\Gamma(x-(x_1+\cdots+x_k)+a)}{\Gamma(x-(x_1+\cdots+x_k)+a+b)}\\ &\times F(-(x_1+\cdots+x_k), x-(x_1+\cdots+x_k)+a,\\ &x-(x_1+\cdots+x_k)+a+b;\alpha),\end{aligned}$$

where $\alpha = \frac{a+b}{a}p$.

Proof It is well known that the probability generating function of the waiting time for the first 1-run of length k. By using de Finetti's theorem and the de Finetti measure (2), we obtain

$$\begin{aligned}E[t^{\tau}] = &\sum_{x=k}^{\infty}\sum_{x_1+2x_2+\cdots+kx_k=x-k}\binom{x_1+\cdots x_k}{x_1,\ldots,x_k}\\ &\times\int_0^{\alpha}u^x\left(\frac{1-u}{u}\right)^{x_1+\cdots+x_k}t^x\frac{1}{\alpha}\frac{1}{B(a,b)}\left(\frac{u}{\alpha}\right)^{a-1}\left(1-\frac{u}{\alpha}\right)^{b-1}du.\end{aligned}$$

Changing variables $u = \alpha v$, we can write

$$\begin{aligned}&\int_0^{\alpha}u^x\left(\frac{1-u}{u}\right)^{x_1+\cdots+x_k}t^x\frac{1}{\alpha}\frac{1}{B(a,b)}\left(\frac{u}{\alpha}\right)^{a-1}\left(1-\frac{u}{\alpha}\right)^{b-1}du\\ &\quad= t^x\int_0^1(\alpha v)^{x-(x_1+\cdots+x_k)}(1-\alpha v)^{x_1+\cdots+x_k}\frac{1}{B(a,b)}v^{a-1}(1-v)^{b-1}dv.\end{aligned}$$

Then, the desired result can easily be shown by the formula of the hypergeometric function (1). □

Corollary 2 (Panaretos and Xekalaki (1986)) *When we repeat sampling from Pólya's urn with a white and b red balls, the probability generating function of the waiting* *time for the first occurrence of a run of white balls of length k can be written as*

$$E[t^{\tau}] = \sum_{x=k}^{\infty} \sum_{x_1+2x_2+\cdots+kx_k=x-k} \binom{x_1+\cdots x_k}{x_1,\ldots,x_k} \alpha^{x-(x_1+\cdots+x_k)} \times \frac{B(a+x-x_1-\cdots-x_k, b+x_1+\cdots+x_k)}{B(a,b)} t^x.$$

Proof It suffices to set $\alpha = 1$ in Proposition 6.

When $\alpha = 1$, $p = \frac{a}{a+b}$ holds. Then, from the definition of $\{V_n\}$, it is reduced to the case that we observe the first occurrence of 1-run of length k in the sequence of Pólya sampling. □

Example 4 (Continuation of Example 3) We shall derive the probability generating function of τ in Example 3 by solving recurrence relations. Let $\psi(w, r, N; t)$ be the conditional probability generating function of number of trials until the first occurrence of 1-run (white balls) which will be possibly observed in N trials starting from the urn with w white and r red balls. Of course, $\psi(w, r, N; t)$ may be a probability generating function of a defective distribution, since the first 1-run may not occur until the N-th trial. By definition, it is clear that $\psi(w, r, 0; t) = 0$ and for $N > 0$, we see that the next recurrence relation holds:

$$\begin{aligned}
&\psi(w, r, N; t)\\
&\quad = P(V_1 = 0, X_1 = 0)t\psi(w+1, r, N-1; t)\\
&\qquad + P(V_1 = 0, X_1 = 1)t\psi(w, r+1, N-1; t)\\
&\qquad + \sum_{\ell=1}^{k-1} \sum_{\boldsymbol{x}\in\{0,1\}^{\ell+1}} P(V_1 = 1, \ldots, V_\ell = 1, V_{\ell+1} = 0, (X_1, \ldots, X_{\ell+1}) = \boldsymbol{x})\\
&\qquad \times t^{\ell+1}\psi(w+s, r+\ell+1-s, N-\ell-1; t)\\
&\qquad + \sum_{\boldsymbol{x}\in\{0,1\}^k} P(V_1 = 1, \ldots, V_k = 1, (X_1, \ldots, X_k) = \boldsymbol{x})t^k,
\end{aligned} \tag{3}$$

where $s = \sum_{i=1}^{\ell+1} x_i$. This can be seen by noting that in case of occurrence of the first '0' before occurrence of the first 1-run of length k we have to start counting from scratch, where balls in the urn will have been changed from the initial state. Here, we can see that

$$\begin{aligned}
&P(V_1 = 1, \ldots, V_\ell = 1, V_{\ell+1} = 0, (X_1, \ldots, X_{\ell+1}) = \boldsymbol{x})\\
&\quad = P((X_1, \ldots, X_{\ell+1}) = \boldsymbol{x})\\
&\qquad \times P(V_1 = 1, \ldots, V_\ell = 1, V_{\ell+1} = 0 | (X_1, \ldots, X_{\ell+1}) = \boldsymbol{x})\\
&\quad = \frac{(w)_{\uparrow s}(r)_{\uparrow(\ell+1-s)}}{(w+r)_{\uparrow(\ell+1)}} P(V_1 = 1, \ldots, V_\ell = 1, V_{\ell+1} = 0 | (X_1, \ldots, X_{\ell+1}) = \boldsymbol{x}).
\end{aligned}$$

When $X_i = 0$, it holds that

$$V_i = 0 \Longleftrightarrow \frac{a}{a+b} U_i > p$$

and

$$V_i = 1 \Longleftrightarrow \frac{a}{a+b} U_i \le p.$$

When $X_i = 1$, we see that

$$V_i = 0 \Longleftrightarrow \frac{a}{a+b} + \frac{b}{a+b} U_i > p \quad \text{(necessarily holds)}$$

and

$$V_i = 1 \quad \Longleftrightarrow \quad \frac{a}{a+b} + \frac{b}{a+b} U_i \le p \quad \text{(never holds).}$$

Thus, if at least one of $\{x_1, x_2, \ldots, x_\ell\}$ is 1, then the corresponding conditional probability vanishes. Therefore, (3) can be written as

$$\begin{aligned}
&\psi(w, r, N; t) \\
&\quad = \frac{w}{w+r}(1-\alpha) t \psi(w+1, r, N-1; t) \\
&\quad\quad + \sum_{\ell=1}^{k-1} \left\{ \frac{(w)_{\uparrow(\ell+1)}}{(w+r)_{\uparrow(\ell+1)}} \alpha^\ell (1-\alpha) t^{\ell+1} \psi(w+\ell+1, r, N-\ell-1; t) \right. \\
&\quad\quad\quad\quad \left. + \frac{(w)_{\uparrow\ell} r}{(w+r)_{\uparrow(\ell+1)}} \alpha^\ell t^{\ell+1} \psi(w+\ell, r+1, N-\ell-1; t) \right\} \\
&\quad\quad + \frac{(w)_{\uparrow k}}{(w+r)_{\uparrow k}} \alpha^k t^k.
\end{aligned} \tag{4}$$

By using the above relation (4), we can derive recursively $E[t^\tau I(\tau \le N)] = \psi(a, b, N; t)$.

By setting $\alpha = 1$ ($\Leftrightarrow p = \frac{a}{a+b}$) in the recurrence relation (4), we obtain

$$\begin{aligned}
\psi(w, r, N; t) = &\frac{r}{w+r} t \psi(w, r+1, N-1; t) \\
&+ \sum_{\ell=1}^{k-1} \frac{(w)_{\uparrow\ell}\, r}{(w+r)_{\uparrow(\ell+1)}} t^{\ell+1} \psi(w+\ell, r+1, N-\ell-1; t) \\
&+ \frac{(w)_{\uparrow k}}{(w+r)_{\uparrow k}} t^k.
\end{aligned}$$

Further, if we set $b = 0$ in the definition of Y_n, then $X_n = 0$ and $Y_n = U_n$ hold with probability one. In the case, the sequence $\{V_n\}$ becomes independent identically distributed sequence with $P(V_n = 1) = p$ and $\alpha = p$. Then, (4) can be written as

$$\psi(a, 0, N; t) = (1-p)t\psi(a+1, 0, N-1; t) + \sum_{\ell=1}^{k-1} p^{\ell}(1-p)t^{\ell+1}\psi(a+\ell+1, 0, N-\ell-1; t) + p^k t^k.$$

Since $\lim_{N\to\infty} \psi(a, 0, N; t)$ does not depend on a, denoting the limit by $\phi(t)$, we obtain

$$\phi(t) = \sum_{\ell=0}^{k-1} p^{\ell}(1-p)t^{\ell+1}\phi(t) + p^k t^k.$$

This can be solved easily and we have the probability generating function of the geometric distribution of order k (see, e.g., Balakrishnan and Koutras (2002); Johnson et al. (2005)).

4 Sequential construction of Markov exchangeable sequences

de Finetti's theorem has been developed to mixtures of Markov chains and some kinds of conditions of partial exchangeability have been investigated (see Diaconis and Freedman (1980); Zabell (1995); Fortini et al. (2002); Quintana and Newton (1998)). Here, as a simple parametric model for Markov exchangeable sequence we sequentially construct mixtures of Markov chains using $\{Y_n\}$ defined in Sect. 2.

Let $0 < p_0, p_1 < 1$ be real numbers. We define

$$W_0 = 0 \ \ a.s.$$

and for $n = 1, 2, \ldots,$

$$W_n = \begin{cases} I(Y_n \le p_0) & \text{if } \ W_{n-1} = 0 \\ I(Y_n \le p_1) & \text{if } \ W_{n-1} = 1. \end{cases}$$

Then, we see that $P(W_1 = 1|\varphi) = P(Y_1 \le p_0|\varphi) = \varphi(p_0)$, and

$$\begin{aligned} P(W_1 = 1, W_2 = 1|\varphi) &= P(Y_1 \le p_0, Y_2 \le p_1|\varphi) \\ &= P(Y_1 \le p_0|\varphi)P(Y_2 \le p_1|\varphi) = \varphi(p_0)\varphi(p_1). \end{aligned}$$

For a $\{0,1\}$-sequence $\{i_j\}$, we set

$$Q(i_{j-1}, i_j) = \begin{cases} \varphi(p_{i_{j-1}}) & \text{if } \ i_j = 1 \\ 1 - \varphi(p_{i_{j-1}}) & \text{if } \ i_j = 0. \end{cases}$$

Then, similarly we obtain that

$$\begin{aligned}
&P(W_n = i_n | W_1 = i_1, W_2 = i_2, \dots, W_{n-1} = i_{n-1}, \varphi) \\
&= \frac{P(W_1 = i_1, \dots, W_{n-1} = i_{n-1}, W_n = i_n | \varphi)}{P(W_1 = i_1, \dots, W_{n-1} = i_{n-1} | \varphi)} \\
&= \frac{Q(i_0, i_1) \cdots Q(i_{n-2}, i_{n-1}) Q(i_{n-1}, i_n)}{Q(i_0, i_1) \cdots Q(i_{n-2}, i_{n-1})} \\
&= Q(i_n - 1, i_n) = P(W_n = i_n | W_{n-1} = i_{n-1}, \varphi).
\end{aligned}$$

Therefore, given φ, $\{W_n\}$ are conditionally Markov chain with $P(W_n = 1 | W_{n-1} = 1, \varphi) = \varphi(p_1)$ and $P(W_n = 1 | W_{n-1} = 0, \varphi) = \varphi(p_0)$. In particular, when $p_0, p_1 \le \dfrac{a}{a+b}$, we can write with a random variable ξ, which follows the beta distribution $Beta(a, b)$, $\varphi(p_1) = \frac{a+b}{a} p_1 \xi$ and $\varphi(p_0) = \frac{a+b}{a} p_0 \xi$.

Example 5 Let τ be the waiting time for the first 1-run of length k in the Markov exchangeable sequence $W_1, W_2, \dots$. Let ν be the waiting time for the first 0 in $W_1, W_2, \dots$. Noting that the sequence is homogeneous, we obtain

$$\begin{aligned}
E[t^\tau | \varphi] &= E\left[t^\tau I(\nu = 1) + \sum_{j=2}^{k} t^\tau I(s = j) + t^\tau I(\nu > k) \,\middle|\, \varphi \right] \\
&= P(\nu = 1) E[t^\tau | \varphi, W_1 = 0] \\
&\quad + \sum_{j=2}^{k} P(\nu = j) E[t^\tau | \varphi, W_1 = 1, \dots, W_{j-1} = 1, W_j = 0] \\
&\quad + \ t^k P(W_1, \dots, W_k = 1 | \varphi) \\
&= (1 - \varphi(p_0)) t E[t^\tau | \varphi] \\
&\quad + \sum_{j=2}^{k} \varphi(p_0) \varphi(p_1)^{j-2} (1 - \varphi(p_1)) t^j E[t^\tau | \varphi] + \varphi(p_0) \varphi(p_1)^{k-1} t^k.
\end{aligned}$$

Therefore, we have

$$E[t^\tau | \varphi] = \frac{\varphi(p_0) \varphi(p_1)^{k-1} t^k}{1 - (1 - \varphi(p_0)) t - \sum_{j=2}^{k} \varphi(p_0) \varphi(p_1)^{j-2} (1 - \varphi(p_1)) t^j}. \tag{5}$$

5 Statistical inference

In this section we shall show the feasibility of parametric estimation of our exchangeable models.

Example 6 In the exchangeable sequence proposed in Example 3 of Sect. 3, we observe waiting time for the first 1-run of length 2. Based on the data of the waiting time, we estimate the parameters a and p, where we posed the additional restriction that $a + b = 10$ in order to reduce the number of parameters. The next data set is a simulated sample assuming that $a = 8$, $b = 2$ and $p = 0.7$,

2, 2, 2, 2, 2, 3, 4, 2, 9, 6, 2, 2, 2, 3, 2, 2, 2, 5, 2, 2, 2, 3, 2, 2, 2, 4, 3, 2, 3, 2, 2, 2, 8, 2, 2, 2, 2, 12, 2, 3, 2, 2, 5, 2, 2, 2, 6, 8, 2, 12, 7, 5, 3, 4, 3, 2, 4, 2, 2, 4, 2, 6, 3, 2, 2, 2, 2, 4, 2, 2, 2, 2, 4, 2, 4, 2, 3, 2, 4, 2, 5, 4, 5, 4, 3, 3, 6, 4, 6, 2, 2, 2, 7, 4, 2, 12, 2, 2, 2, 3, 5, 4, 6, 7, 3, 3, 2, 3, 12, 6, 3, 4, 5, 2, 12, 2, 4, 2, 3, 6, 12, 2, 2, 2, 2, 2, 3, 4, 2, 5, 2, 2, 2, 2, 6, 2, 2, 2, 2, 9, 3, 2, 6, 2, 3, 2, 4, 5, 2, 3, 4, 2, 4, 2, 2, 4, 3, 2, 4, 2, 2, 11, 3, 2, 2, 2, 2, 2, 4, 11, 4, 2, 4, 3, 2, 4, 2, 4, 4, 5, 2, 2, 2, 2, 12, 7, 7, 6, 9, 2, 3, 6, 3, 2, 8, 2, 3, 2, 5, 2.

In generating the data, we censored the waiting time at 11 for simplicity. In the above data set, 12 means that the waiting time is over 11. As we explained in Example 4, the exact probability of the censored waiting time can be obtained as the function of the parameters. Therefore, we could calculate the exact log-likelihood function from the data set. By maximizing the function with respect to a and p, we obtained the maximum likelihood estimate $\hat{a} = 7.502$, $\hat{p} = 0.706$. The graph of the log-likelihood function is in Fig. 4.

Example 7 In the Markov exchangeable sequence introduced in Sect. 4, we observe the waiting time for the first 1-run of length 2. In the model, there are four parameters a, b, p_0, p_1. In order to reduce the number of parameters, we pose the additional restriction $a+b = 10$, $p_0 < \frac{a}{a+b}$ and $p_1 = \dfrac{p_0 + \frac{a}{a+b}}{2}$. Further, we censor the waiting time at 11 for simplicity. Then, the parameters to be estimated are a and p_0. We can calculate the likelihood function as follows. From (5), given the random distribution function φ, the conditional probability generating function of the waiting time for the first 1-run of length 2 is written as

$$\phi(t) = \frac{\varphi(p_0)\varphi(p_1)t^2}{1 - (1 - \varphi(p_0))t - \varphi(p_0)(1 - \varphi(p_1))t^2}.$$

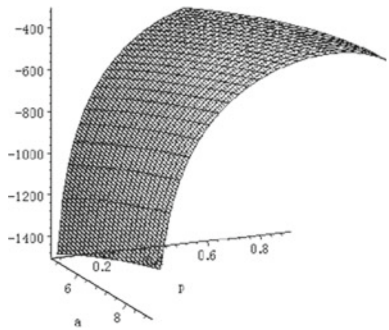
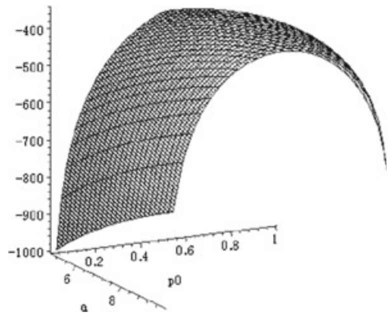


Fig. 4 The graph of the log-likelihood function based on the data set of the waiting time of the first 1-run of length 2

Fig. 5 The graph of the log-likelihood function of a and p_0 based on the data set of the waiting time of the first 1-run of length 2



Expanding it up to the 11th term w.r.t. t, we substitute $\varphi(p_1) = \frac{a+b}{a} p_1 \xi$, $\varphi(p_0) = \frac{a+b}{a} p_0 \xi$. Further, multiplying it by $\frac{1}{B(a,b)} \xi^{a-1}(1-\xi)^{b-1}$, and integrating the formula w.r.t. ξ from 0 to 1, we obtain the probability generating function of the consored waiting time. The integration need not be performed, since it is equivalent to replace all the ξ^n's by

$$\frac{a(a+1)\cdots(a+n-1)}{(a+b)(a+b+1)\cdots(a+b+n-1)}.$$

By taking out the coefficient of t^k from the probability generating function, we get the probability that the waiting time becomes k. The next data set is simulated assuming that $a = 8$, $b = 2$, $p_0 = 0.6$ and $p_1 = 0.7$,

2, 2, 2, 2, 2, 3, 4, 2, 9, 6, 2, 2, 2, 4, 2, 2, 2, 5, 2, 2, 5, 3, 2, 2, 2, 4, 3, 2, 6, 2, 2, 2, 8, 2, 10, 2, 2, 12, 2, 3, 2, 2, 5, 2, 3, 2, 6, 8, 3, 12, 7, 5, 3, 5, 3, 2, 4, 2, 2, 4, 2, 6, 3, 2, 2, 2, 2, 4, 3, 2, 2, 2, 4, 2, 4, 2, 4, 2, 4, 2, 5, 4, 5, 4, 3, 3, 7, 4, 6, 2, 3, 6, 7, 4, 3, 12, 2, 2, 2, 3, 5, 10, 6, 7, 3, 3, 2, 3, 12, 6, 3, 4, 5, 3, 12, 2, 4, 3, 3, 6, 12, 2, 2, 2, 2, 3, 3, 8, 2, 5, 2, 2, 2, 2, 6, 2, 2, 2, 2, 9, 3, 2, 6, 2, 3, 2, 4, 5, 2, 3, 4, 2, 4, 2, 3, 4, 3, 2, 4, 2, 2, 11, 3, 2, 2, 2, 2, 2, 5, 11, 4, 2, 4, 3, 2, 4, 2, 5, 4, 5, 2, 2, 2, 2, 12, 7, 7, 6, 9, 2, 3, 7, 3, 2, 8, 2, 4, 2, 5, 3.

By maximizing the log-likelihood function with respect to a and p_0, we obtained the maximum likelihood estimate $\hat{a} = 8.075$, $\hat{p} = 0.611$. The graph of the log-likelihood function is in Fig. 5.

References

Aki, S. (2008). Joint distributions of numbers of occurrences of a discrete pattern and weak convergence of an empirical process for the pattern. *Journal of Multivariate Analysis, 99*, 1460–1473.

Aki, S., Hirano, K. (2008). Waiting time distributions for a run with additional constraints. *Journal of Statististical Planning and Inference, 138*, 3492–3501.

Aldous, D.J. (1985). Exchangeability and related topics, Lectures from the Summer School on Probability Theory held in Saint-Flour, Lecture Notes in Mathematics, 1117 (pp. 1–198). Berlin: Springer.

Balakrishnan, N., Koutras, M.V. (2002). *Runs and scans with applications*. New York: Wiley.

Bertoin, J. (2006). *Random fragmentation and coagulation processes*. Cambridge: Cambridge University Press.

Billingsley, P. (1995). *Probability and measure* (3rd ed). New York: Wiley.

de Finetti, B. (1975). *Theory of probability* (Vol. 2). New York: Wiley.

Diaconis, P., Freedman, D. (1980). de Finetti's theorem for Markov chain. *Annals of Probability, 8*, 115–130.

Durrett, R. (2005). *Probability: theory and examples* (3rd ed). Belmont: Brooks/Cole-Thomson Learning.

Ebneshahrashoob, M., Sobel, M. (1990). Sooner and later waiting time problems for Bernoulli trials: frequency and run quotas. *Statistics & Probability Letters, 9*, 5–11.

Eryilmaz, S., Demir, S. (2007). Success runs in a sequence of exchangeable binary trials. *Journal of Statistical Planning and Inference, 137*, 2954–2963.

Fortini, S., Ladelli, L., Petris, G. (2002). On mixtures of distributions of Markov chains. *Stochastic Processes and their Applications, 100*, 147–165.

Freedman, D. (1965). Bernard Friedman's urn. *Annals of Mathematical Statistics, 36*(3), 956–970.

Fu, J. C. , Koutras, M. V. (1994). Distribution theory of run: a Markov chain approach. *Journal of American Statistical Association 89*, 1050–1058.

Fu, J. C., Lou, W. Y. W. (2003). *Distribution theory of runs and patterns and its applications*. Singapore: World Scientific.

George, E. O., Bowman. D. (1995). A full likelihood procedure for analysing exchangeable binary data. *Biometrics, 51*, 512–523.

Hill, B. M., Lane, D., Sudderth, W. (1987). Exchengeable urn processes. *Annals of Probability, 15*, 1586–1592.

Inoue, K., Aki, S. (2005). A generalized Pólya urn model and related multivariate distributions. *Annals of the Institute of Statistical Mathematics, 57*, 49–59.

Irwin, J. O. (1954). A distribution arising in the study of infectious diseases. *Biometrika, 41*, 266–268.

Johnson, N. L., Kotz, S. (1977). *Urn models and their applications*. New York: Wiley.

Johnson, N. L., Kemp, A. W., Kotz, S. (2005). *Univariate discrete distributions* (3rd ed). New York: Wiley.

Kemp, C. D., Kemp, A. W. (1956). Generalized hypergeometric distributions. *Journal of the Royal Statistical Society, Series B, 18*, 202–211.

Kingman, J. F. C. (1978). Uses of exchangeability. *Annals of Probability, 6*, 183–197.

Kolev, N., Paiva, D. (2008). Random sums of exchangeable variables and acturial applications. *Insurance: Mathematics and Economics, 42*, 147–153.

Kolev, N., Kolkovska, E. T., López-Mimbela, J. A. (2006). Joint probability generating function for a vector of arbitrary indicator variables. *Journal of Computational and Applied Mathematics, 186*, 89–98.

Lau, T. S. (1992). The reliability of exchangeable binary systems. *Statistics & Probability Letters, 13*, 153–158.

Mauldin, R., Sudderth, W. D., Williams, S. C. (1992). Pólya trees and random distributions. *Annals of Statistics, 20*, 1203–1221.

Panaretos, J., Xekalaki, E. (1986). On some distributions arising from certain generalized sampling schemes. *Communications in Statistics-Theory and Methods, 15*, 873–891.

Pitman, J. (2006). Combinatorial stochastic processes, Ecole d'Eté de Probabilités de Saint-Flour XXXII-2002, Lecture Notes in Mathematics 1875. Berlin: Springer.

Quintana, F. A., Newton, M. A. (1998). Assessing the order of dependence for partially exchangeable binary data. *Journal of American Statistical Association, 93*, 194–202.

Zabell, S. L. (1995). Characterizing Markov exchangeable sequences. *Journal of Theoretical Probability, 8*, 175–178.