

Research Article

Estimation of Poisson-Dirichlet Parameters with Monotone Missing Data

Xueqin Zhou,¹ Jinlong Huang,² and Xianyi Wu¹

¹Department of Statistics and Actuarial Science, East China Normal University, Shanghai, China

²China Pacific Property Insurance Co., Ltd., Shanghai, China

Correspondence should be addressed to Xueqin Zhou; xueqinzhou@163.com

Received 22 March 2017; Accepted 29 August 2017; Published 12 October 2017

Academic Editor: Giuseppe Vairo

Copyright © 2017 Xueqin Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article considers the estimation of the unknown numerical parameters and the density of the base measure in a Poisson-Dirichlet process prior with grouped monotone missing data. The numerical parameters are estimated by the method of maximum likelihood estimates and the density function is estimated by kernel method. A set of simulations was conducted, which shows that the estimates perform well.

1. Introduction

As a young but fast-growing field of statistics, Bayesian nonparametrics (abbreviated as BNP below) focuses on Bayesian solutions of nonparametric and other infinite-dimensional statistical models. Compared with frequentist statistics and classical Bayesian statistics, BNP provides highly flexible and robust models for infinite-dimensional parameter spaces. The most extensively investigated priors in BNP include the famous Dirichlet process prior [1] and Polya tree process prior [2–5] which have played fundamental roles in the development of Bayesian nonparametrics.

Dirichlet processes are also referred to as one-parameter Poisson-Dirichlet processes by Kingman [6]. As the first and foremost generalization of Dirichlet processes, two-parameter Poisson-Dirichlet processes (abbreviated as Poisson-Dirichlet process below) were first discussed by Pitman and Yor [7] and from then have made huge success in Bayesian nonparametric modeling in language, images, ecology, biology, genomics, and so on. Remarkable examples include Goldwater et al. [8] who used Poisson-Dirichlet process as an adaptor to justify the appearance of type frequencies in formal analyses of natural language and improved the performance of an earlier model for unsupervised learning of morphology, Sudderth and Jordan [9] who modeled the object frequencies

and segment sizes by Poisson-Dirichlet processes and developed a statistical framework for the simultaneous, unsupervised segmentation and discovery of visual object categories from image databases, Favaro et al. [10] who used a Poisson-Dirichlet model to deal with the issue of prediction within species sampling problems, and Hoshino [11] who studied the microdata disclosure risk assessment with Pitman's sampling formula, clarified some of its theoretical implications, and compared various models based on the Akaike Information Criterion by applying them to real data sets. For more references related to the application of Poisson-Dirichlet process in the area of language learning, one can be referred to Johnson et al. [12], Wood and Teh [13], and Wallach et al. [14].

While the exact Bayesian methods take the assumption that prior distributions are completely specified, empirical Bayesian methods deal with the situations where prior distributions are at most partially specified and thus need to be estimated. Empirical Bayesian methods for parametric and semi-parametric models have been investigated in a huge volume of literature. However, the study of empirical Bayesian methods in the framework of Bayesian nonparametrics is quite limited. A recently published paper is that by Yang and Wu [15] who studied the problem of estimation of the priors with monotonically missing data when the prior is a Dirichlet process.

In this paper, we aim at estimating the unknown numerical parameters and the density of the base measure with independent and identically distributed (i.i.d.) groups of observations with a Poisson-Dirichlet process prior. Because the Dirichlet process prior is a special case of the Poisson-Dirichlet process prior, we in fact have extended the methodologies of Yang and Wu [15] to a bigger model. The estimation of the unknown parameters is carried out in two different methods, the maximum likelihood method and a naive method proposed by Carlton [16], of which performances are compared by a simulation study.

Because there are two numerical parameters in Poisson-Dirichlet process priors, the maximum likelihood estimates (MLE) for the unknown parameters are discussed under three different settings (see the next section for the definition of the parameters α and θ and the density function): (i) the discount parameter α is unknown but the concentration parameter θ is known; (ii) the concentration parameter θ is unknown but the discount parameter α is known; and (iii) both α and θ are unknown. Favaro et al. [10] gave the empirical Bayes estimates when both α and θ were unknown on complete data without missing. The comparisons between the estimates of Favaro et al. [10] and ours are presented with the same sample size in terms of bias, standard deviations (SD), and mean squared errors (MSEs).

The remainder of this paper is structured as follows. In Section 2, we review the basic model and the definition of the Poisson-Dirichlet processes priors. Data structure and model assumptions are also described in this section. In Section 3, the MLEs of the prior parameters are discussed in detail under three different aforementioned situations. A naive estimate for the discount parameter α is also discussed. Section 4 discusses the estimates of base distribution density by kernel method. Section 5 presents a small simulation study to show the performance of the estimates discussed in Section 3.

2. The Data and Model

The data are observed in I time periods and accordingly organized in groups $i = 1, 2, \dots, I$, with i representing the calendar time point the individuals in this group begin to be observed. Each group i contains n_i individuals. Denote by (i, k) the k th individual in group i which is represented by a random vector X_{ik} with X_{ikj} being the j th coordinate of individual (i, k) . The observations are presented by J -dimensional ($J \leq I$) vectors for which the coordinates are observed sequentially in time, so that the observations are subject to monotone missing. A clearer picture of the data structure is exhibited in Table 1.

Hence, the observed data up to time I are monotone missing; that is, the components of an individual (i, k) are ordered in such a way that if an observation of a variable X_{ikj} for individual (i, k) is missing, then so are the observations of all subsequent variables X_{ikl} , $l \geq j$ (if any) for the same individual. Clearly, Table 1 indicates that variables for individuals in the first $I - J + 1$ groups are completely observed

and only the first $I - i + 1$ components are observable for individuals in group $i = I - J + 2, \dots, I$.

Data structured as in Table 1 is frequently encountered in real world. A typical example is loss data for claims used for the purpose of loss reserving in the industry of non-life insurance (see, e.g., [15, 17, 18]). Assume that the evaluation time of loss reserving is accident year I ; each claim made at accident year i ($i \leq I$) will be paid at the end of the subsequent J years after the claim (the first is the one paid at the end of the year the claim is made) so that the payments of an individual (i, k) correspond to a J -vector $(X_{ik1}, \dots, X_{ikj}, \dots, X_{ikJ})$. Then observing at the end of the evaluation year I , the payments for a claim in accident year i have been paid only with j s such that $j \leq I - i + 1$; that is, the observable payments of claim (i, k) are just the subvector $\mathbf{X}_{ik}^o = (X_{ik1}, X_{ik2}, \dots, X_{ik, I-i+1})$, where the superscript “o” means “observable.” Other example data of this structure can also be found in Marini et al. [19] who discussed the maximum likelihood estimation in panel studies for the first time, Hao and Krishnamoorthy [20] who considered testing and estimation problems, and Raats et al. [21] who discussed the estimation and testing for multivariate regression model, among a large number of others, with monotone missing data.

Before presenting the probabilistic characteristics of the data under a BNP framework in Assumption 2, we recall the definition of Poisson-Dirichlet processes in terms of stick-breaking. Let (α, θ) be a pair of real numbers satisfying $\alpha \in [0, 1)$ and $\theta > -\alpha$, which is always respected and hence will not be mentioned everywhere.

Definition 1. (1) Let $(V_n)_{n \geq 1}$ be a sequence of mutually independent random variables with $V_n \sim \text{Beta}(1 - \alpha, \theta + n\alpha)$ and $\tilde{p}_1 = V_1$ and $\tilde{p}_k = V_k \prod_{i=1}^{k-1} (1 - V_i)$, $k \geq 2$, so that $\sum_{k=1}^{\infty} \tilde{p}_k = 1$. Then the ranked sequence $\mathbf{p} = (p_1, p_2, \dots)$ (i.e., $p_1 \geq p_2 \geq \dots$) of $\tilde{\mathbf{p}} = (\tilde{p}_1, \tilde{p}_2, \dots)$ is said to be Poisson-Dirichlet distributed, writing $\mathbf{p} \sim \text{PD}(\alpha, \theta)$, where α and θ are called discount and concentration parameter, respectively.

(2) Let $(Y_n)_{n \geq 1}$ be a sequence of random variables independent and identically distributed as a probability measure μ (called *base distribution* below), taking values in a measurable space $(\mathcal{X}, \mathcal{B})$, and $\mathbf{p} = (p_n)_{n \geq 1} \sim \text{PD}(\alpha, \theta)$. Then the random probability measure

$$Q = \sum_{k=1}^{\infty} p_k \delta_{Y_k}(\cdot) \quad (1)$$

on $(\mathcal{X}, \mathcal{B})$ is referred to as a Poisson-Dirichlet process (indexed by \mathcal{B}) with parameters α, θ and base distribution $\mu(\cdot)$ or, in symbol, writing $Q \sim \text{PD}(\mu, \alpha, \theta)$.

Clearly, for given \mathbf{p} and $(Y_n)_{n \geq 1}$, one has a realization of the Poisson-Dirichlet process Q that is a discrete distribution with mass p_n at Y_n , $n = 1, 2, \dots$ in the domain \mathcal{X} . Pitman and Yor [7] discussed in detail Poisson-Dirichlet processes in which a number of well-known properties of Dirichlet processes were generalized. Carlton [16] gave methods to estimate the parameters (α, θ) with completely observed data.

TABLE I: Data structure.

Groups	1	2	...	$I - i + 1$...	$J - 1$	J
1	X_{111}	X_{112}	...	$X_{11(I-i+1)}$...	$X_{11(J-1)}$	X_{11J}
	X_{121}	X_{122}	...	$X_{12(I-i+1)}$...	$X_{12(J-1)}$	X_{12J}
	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
2	$X_{1n_1 1}$	$X_{1n_1 2}$...	$X_{1n_1(I-i+1)}$...	$X_{1n_1(J-1)}$	$X_{1n_1 J}$
	X_{211}	X_{212}	...	$X_{21(I-i+1)}$...	$X_{21(J-1)}$	X_{21J}
	X_{221}	X_{222}	...	$X_{22(I-i+1)}$...	$X_{22(J-1)}$	X_{22J}
	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
$I - J + 1$	$X_{2n_2 1}$	$X_{2n_2 2}$...	$X_{2n_2(I-i+1)}$...	$X_{2n_2(J-1)}$	$X_{2n_2 J}$
	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
	$X_{(I-J+1)11}$	$X_{(I-J+1)12}$...	$X_{(I-J+1)1(I-i+1)}$...	$X_{(I-J+1)1(J-1)}$	$X_{(I-J+1)1J}$
$I - J + 2$	$X_{(I-J+1)21}$	$X_{(I-J+1)22}$...	$X_{(I-J+1)2(I-i+1)}$...	$X_{(I-J+1)2(J-1)}$	$X_{(I-J+1)2J}$
	\vdots	\vdots	...	\vdots	...	\vdots	\vdots
	$X_{(I-J+1)n_{(I-J+1)}1}$	$X_{(I-J+1)n_{(I-J+1)}2}$...	$X_{(I-J+1)n_{(I-J+1)}(I-i+1)}$...	$X_{(I-J+1)n_{(I-J+1)}(J-1)}$	$X_{(I-J+1)n_{(I-J+1)}J}$
	$X_{(I-J+2)11}$	$X_{(I-J+2)12}$...	$X_{(I-J+2)1(I-i+1)}$...	$X_{(I-J+2)1(J-1)}$	
$I - J + 2$	$X_{(I-J+2)21}$	$X_{(I-J+2)22}$...	$X_{(I-J+2)2(I-i+1)}$...	$X_{(I-J+2)2(J-1)}$	
	\vdots	\vdots	...	\vdots	...	\vdots	
	$X_{(I-J+2)n_{(I-J+2)}1}$	$X_{(I-J+2)n_{(I-J+2)}2}$...	$X_{(I-J+2)n_{(I-J+2)}(I-i+1)}$...	$X_{(I-J+2)n_{(I-J+2)}(J-1)}$	
\vdots	\vdots	...	\vdots	...	\vdots		
i	X_{i11}	X_{i12}	...	$X_{i1(I-i+1)}$			
	X_{i21}	X_{i22}	...	$X_{i2(I-i+1)}$			
	\vdots	\vdots	...	\vdots			
	$X_{in_i 1}$	$X_{in_i 2}$...	$X_{in_i(I-i+1)}$			
\vdots	\vdots	...	\vdots				
$I - 1$	$X_{(I-1)11}$	$X_{(I-1)12}$					
	$X_{(I-1)21}$	$X_{(I-1)22}$					
	\vdots	\vdots					
I	$X_{(I-1)n_{I-1} 1}$	$X_{(I-1)n_{I-1} 2}$					
	X_{I11}						
	X_{I21}						
	\vdots						
	$X_{In_I 1}$						

Assumption 2.

- (a) For every $i = 1, 2, \dots, I$, the data $(\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{in_i})$ have a Poisson-Dirichlet structure: $\mathbf{P}_i \sim \text{PD}(\mu; \alpha, \theta)$ and, given \mathbf{P}_i , the data $\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{in_i} \stackrel{\text{i.i.d.}}{\sim} \mathbf{P}_i$.
- (b) The random elements $(\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{in_i}; \mathbf{P}_i)$ are mutually independent over $i = 1, \dots, I$.
- (c) The probability measure μ has a continuous density function, denoted also by $\mu(\mathbf{x})$, with respect to the J -dimensional Lebesgue measure on \mathbb{R}^J .

The intuition behind the assumption is as what follows. In many situations, the data are recorded chronologically so that the first subscript of X_{ikj} reflects the calendar time when the data vector begins to be recorded and the third

subscript represents the calendar time that component is recorded; hence the dependence between the data beginning at a same calendar time and independent over calendar times is reasonable. This structure is the typical case of loss reserving in general insurance with individual data (see, e.g., Huang et al. [17, 18]).

Given the above data structure and model assumption, the objective of this paper is to estimate the unknown parameters α and θ and the density $\mu(\mathbf{x})$ of the base measure.

3. Estimation of Parameters (α, θ)

The unknown parameters (α, θ) of the distribution of a Poisson-Dirichlet process can generally be estimated by means of maximum likelihood as shown in Section 3.2. To do it under the data structure of Table 1, we first introduce some

necessary preliminary results for Poisson-Dirichlet processes in Section 3.1

3.1. Preliminaries. In the following, write $\mathbf{X}_{ik}^j = (X_{ik1}, X_{ik2}, \dots, X_{ikj})$, $j = 1, 2, \dots, J$, for the first j components of \mathbf{X}_{ik} and \mathbf{P}_i^j and μ^j , respectively, for the marginal distributions of \mathbf{P}_i and μ on the first j -dimensional subspace \mathbb{R}^j of \mathbb{R}^J . The following lemma shows that if the observations $\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{in_i}$ have a Poisson-Dirichlet structure, then so are their subvectors.

Lemma 3. *If the observations $\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{in_i} \mid \mathbf{P}_i \stackrel{\text{i.i.d.}}{\sim} \mathbf{P}_i$ and $\mathbf{P}_i \sim PD(\mu; \alpha, \theta)$, then, for every $j = 1, 2, \dots, J$, given \mathbf{P}_i^j , the partial observations $\mathbf{X}_{i1}^j, \mathbf{X}_{i2}^j, \dots, \mathbf{X}_{in_i}^j \stackrel{\text{i.i.d.}}{\sim} \mathbf{P}_i^j$ and $\mathbf{P}_i^j \sim PD(\mu^j; \alpha, \theta)$.*

Proof. The assertion that $\mathbf{P}_i^j \sim PD(\mu^j; \alpha, \theta)$ follows immediately from the definition of $PD(\mu; \alpha, \theta)$. For any \mathbb{B}^j -measurable set B_k^j , $k = 1, 2, \dots, n_i$, we have

$$\begin{aligned} \Pr(\mathbf{X}_{i1}^j \in B_1^j, \dots, \mathbf{X}_{in_i}^j \in B_{n_i}^j \mid \mathbf{P}_i^j) \\ = E \left[\Pr(\mathbf{X}_{i1}^j \in B_1^j, \dots, \mathbf{X}_{in_i}^j \in B_{n_i}^j \mid \mathbf{P}_i) \mid \mathbf{P}_i^j \right] \\ = E \left[\prod_{k=1}^{n_i} P_i(B_k^j \times \mathbb{R}^{J-j}) \mid \mathbf{P}_i^j \right] = \prod_{k=1}^{n_i} \mathbf{P}_i^j(B_k^j). \end{aligned} \quad (2)$$

It indicates the desired result $\mathbf{X}_{i1}^j, \mathbf{X}_{i2}^j, \dots, \mathbf{X}_{in_i}^j \mid \mathbf{P}_i^j \stackrel{\text{i.i.d.}}{\sim} \mathbf{P}_i^j$. \square

The next lemma shows that, with a Poisson-Dirichlet process structure, observations of any two individuals are identical with probability of 1 if and only if so are their corresponding components. This provides an easy and convenient way to judge whether any two observations are the same only by their any (e.g., the first) component.

Lemma 4. *For any $s, t \in \{1, 2, \dots, n_i\}$, the equality $X_{is1} = X_{it1}$ implies that $\mathbf{X}_{is}^j = \mathbf{X}_{it}^j$ for all $j = 1, 2, \dots, J$ almost surely.*

Proof. We verify the case $j = 2$ and the proof applies for any $j > 2$. By the definition of Poisson-Dirichlet process, we take $\mathbf{P}_i = \sum_{n=1}^{\infty} p_n \delta_{Y_n}$ with an i.i.d. sequence $\{Y_i\}_{i \geq 1}$ drawn from μ . Because μ is continuous, then we have $\Pr(Y_i = Y_j \text{ for some } i, j \in \{1, 2, \dots\}) = 0$. Therefore,

$$\begin{aligned} \Pr(\mathbf{X}_1 = \mathbf{X}_2) &= E \left[\Pr(\mathbf{X}_1 = \mathbf{X}_2 \mid \mathbf{P}_i) \right] \\ &= E \left[\sum_{n=1}^{\infty} \Pr(\mathbf{X}_1 = \mathbf{X}_2 = Y_n \mid \mathbf{P}_i) \right] \\ &= E \left[\sum_{n=1}^{\infty} p_n^2 \right]. \end{aligned} \quad (3)$$

Corollary 3 of Pitman and Yor [7] implies that

$$\Pr(\mathbf{X}_1 = \mathbf{X}_2) = E \left[\frac{W^2}{W} \right] = E[W] = \frac{1-\alpha}{1+\theta}, \quad (4)$$

where $W \sim \text{Beta}(1-\alpha, \theta+\alpha)$. On the other hand, because Lemma 3 implies that X_{11} and X_{21} are two observations from a PD structure with the parameters α and θ and $\mu^1(t_1) = \int \mu(t_1, t_2, \dots, t_j) dt_2 \cdots dt_j$, similar discussion as above implies that $\Pr(X_{11} = X_{21}) = (1-\alpha)/(1+\theta)$. Therefore,

$$\Pr(\mathbf{X}_1 = \mathbf{X}_2 \mid X_{11} = X_{21}) = \frac{\Pr(\mathbf{X}_1 = \mathbf{X}_2)}{\Pr(X_{11} = X_{21})} = 1. \quad (5)$$

The proof is thus completed. \square

3.2. Maximum Likelihood Estimation of the Unknown Parameters (α, θ) . For $i = 1, 2, \dots, I$, let K_i be the number of distinct observations among the n_i individuals of the i th group. Each distinct observation represents a species or cluster in biological perspective [7]. Let A_j^i be the number of those clusters, each of which appears j times in group i . For example, if $A_j^i = 3$ for some $j \in \{1, 2, \dots, n_i\}$, then there are total 3 different species, each of which appears just j times in group i . Clearly,

- (1) there might exist many $j \in \{1, 2, \dots, n_i\}$ with $A_j^i = 0$,
- (2) generally,

$$A_j^i \geq 0 \quad \text{for } j = 1, 2, \dots, I,$$

$$K_i = \sum_{j=1}^{n_i} A_j^i, \quad (6)$$

$$\sum_{j=1}^{n_i} j A_j^i = n_i.$$

Write

$$\mathbf{A}^i = (A_1^i, A_2^i, \dots, A_{n_i}^i), \quad i = 1, 2, \dots, I. \quad (7)$$

Similar to Carlton [16] who used \mathbf{A}^i to estimate the unknown parameters by maximum likelihood method when individuals are completely observed, we here discuss the MLE of (α, θ) in terms of the random variables \mathbf{A}^i , $i = 1, 2, \dots, I$. Under Assumption 2, the log-likelihood function for (α, θ) given the observations $\mathbf{A}^i = \mathbf{a}^i$ in all the I groups is

$$\begin{aligned} \ell(\alpha, \theta) &= \sum_{i=1}^I \log \Pr(\mathbf{A}^i = \mathbf{a}^i) \\ &= \sum_{i=1}^I \log N(\mathbf{a}^i) - \sum_{i=1}^I \sum_{l=1}^{n_i-1} \log(\theta + l) \\ &\quad + \sum_{i=1}^I \sum_{l=1}^{K_i-1} \log(\theta + l\alpha) \\ &\quad + \sum_{i=1}^I \sum_{j=2}^{n_i} a_j^i \sum_{l=1}^{j-1} \log(l - \alpha), \end{aligned} \quad (8)$$

where $N(\mathbf{a}^i) = n_i! / \prod_{j=1}^{n_i} (j!)^{a_j^i}$ is irrelevant to the parameter (α, θ) .

The estimation is analyzed in three different situations as what has been done in Carlton [16]:

- (i) α is unknown and θ is known.
- (ii) α is known and θ is unknown.
- (iii) Both α and θ are unknown.

Remark 5. In the following, we assume that $K_i \neq 1$ and $K_i \neq n_i$ for some $i = 1, 2, \dots, I$. Note that, in group i , $K_i = 1$ means that all observations are the same and $K_i = n_i$ means that all the observations are different from one another. Thus, if $K_i = 1$ or $K_i = n_i$ for all $i = 1, 2, \dots, I$, the MLE will be attained at the boundary of the parameter spaces. However, the joint events $\{K_i = n_i, i = 1, 2, \dots, I\}$ and $\{K_i = 1, i = 1, 2, \dots, I\}$ will become increasingly unlikely as $n = \sum_{i=1}^I n_i$ tends to infinity; see Chapter 3 of Carlton [16].

3.2.1. *Estimate α with Known θ .* In this case, note that

$$\ell_\alpha(\alpha, \theta) := \frac{\partial \ell(\alpha, \theta)}{\partial \alpha} = \sum_{i=1}^I \sum_{l=1}^{K_i-1} \frac{l}{\theta + l\alpha} - \sum_{i=1}^I \sum_{j=2}^{n_i} a_j^i \sum_{l=1}^{j-1} \frac{1}{l-\alpha} \quad (9)$$

is a decreasing function of α and $\ell_\alpha(\alpha, \theta) \rightarrow -\infty$ as $\alpha \rightarrow 1$. The equality $\ell_\alpha(\alpha, \theta) = 0$ has a unique solution if $\ell_\alpha(0, \theta) \geq 0$, as stated in the theorem below.

Theorem 6. *The MLE $\hat{\alpha}(\theta)$ of α is unique and*

$$\hat{\alpha}(\theta) > 0 \quad \text{if } \theta < \frac{\sum_{i=1}^I K_i (K_i - 1)}{2 \sum_{i=1}^I \sum_{j=2}^{n_i} a_j^i (\ln(j-1) + C)}, \quad (10)$$

$$\hat{\alpha}(\theta) = 0 \quad \text{otherwise,}$$

where C is Euler's constant.

Proof. Note that

$$\begin{aligned} \ell_\alpha(0, \theta) &= \frac{1}{\theta} \sum_{i=1}^I \sum_{l=1}^{K_i-1} l - \sum_{i=1}^I \sum_{j=2}^{n_i} a_j^i \sum_{l=1}^{j-1} \frac{1}{l} \\ &= \frac{1}{2\theta} \sum_{i=1}^I K_i (K_i - 1) \\ &\quad - \sum_{i=1}^I \sum_{j=2}^{n_i} a_j^i (\ln(j-1) + C). \end{aligned} \quad (11)$$

The proof is thus completed. \square

3.2.2. *Estimate θ with Known α .* If $\alpha = 0$, then the Poisson-Dirichlet process reduces to a Dirichlet process, for which the estimation problem of θ can be referred to Yang and Wu [15]. We thus consider only the case of $\alpha \in (0, 1)$. By the log-likelihood function $\ell(\alpha, \theta)$ given in (8), we have

$$\ell_\theta(\alpha, \theta) := \frac{\partial \ell(\alpha, \theta)}{\partial \theta} = -\sum_{i=1}^I \sum_{l=1}^{n_i-1} \frac{1}{\theta + l} + \sum_{i=1}^I \sum_{l=1}^{K_i-1} \frac{1}{\theta + l\alpha}. \quad (12)$$

A condition under which the MLE is unique is presented in the following theorem.

Theorem 7. *Set $k_{\max} = \max\{K_1, K_2, \dots, K_I\}$ and $k_{\min} = \min\{K_1, K_2, \dots, K_I\}$. Then the solution to the equation $\ell_\theta(\alpha, \theta) = 0$ is unique if $k_{\max} - 2k_{\min} \leq 0$.*

Proof. For given $\alpha \in (0, 1)$, notice that the variable for $\ell_\theta(\alpha, \theta)$ is only θ ; we write

$$\begin{aligned} g(\theta) &= \ell_\theta(\alpha, \theta) \\ &= \sum_{i=1}^I \left[\sum_{l=1}^{K_i-1} \left(\frac{1}{\theta + l\alpha} - \frac{1}{\theta + l} \right) - \sum_{l=K_i}^{n_i-1} \frac{1}{\theta + l} \right]. \end{aligned} \quad (13)$$

Further recall that $K_i > 1$ for $i = 1, 2, \dots, I$; it follows that

$$\begin{aligned} g(\theta) &= \frac{I}{\theta + \alpha} \\ &\quad + \sum_{i=1}^I \left[\sum_{l=2}^{K_i-1} \frac{1}{\theta + l\alpha} - \sum_{l=1}^{K_i-1} \frac{1}{\theta + l} - \sum_{l=K_i}^{n_i-1} \frac{1}{\theta + l} \right], \end{aligned} \quad (14)$$

and consequently $g(\theta) \rightarrow \infty$ as $\theta \rightarrow -\alpha$ for the second term is limited in the upper equation. With some algebraic transformation to (13), we obtain

$$g(\theta) = \sum_{i=1}^I \left[(1 - \alpha) \sum_{l=1}^{K_i-1} \frac{l}{(\theta + l\alpha)(\theta + l)} - \sum_{l=K_i}^{n_i-1} \frac{1}{\theta + l} \right], \quad (15)$$

which implies that, for sufficiently large θ ,

$$g(\theta) \leq \sum_{i=1}^I \left[\frac{(K_i)^2 (1 - \alpha) - (n_i - K_i - 1)\theta/2}{\theta^2} \right] < 0. \quad (16)$$

Then, by the continuity of $g(\theta)$, the Intermediate Value Theorem implies that $c \in (-\alpha, \infty)$ such that $g(c) = 0$.

For the uniqueness, it suffices to show that $g(\theta) < 0$ whenever $g'(\theta) \geq 0$. By (13), we have

$$\begin{aligned} g'(\theta) &= \sum_{i=1}^I \left[\sum_{l=1}^{n_i-1} \frac{1}{(\theta + l)^2} - \sum_{l=1}^{K_i-1} \frac{1}{(\theta + l\alpha)^2} \right] \\ &= \sum_{i=1}^I \left[(\alpha - 1) \sum_{l=1}^{K_i-1} \frac{l(l\alpha + l + 2\theta)}{(\theta + l)^2 (\theta + l\alpha)^2} + \sum_{l=K_i}^{n_i-1} \frac{1}{(\theta + l)^2} \right], \end{aligned} \quad (17)$$

which implies that

$$g'(\theta) < \sum_{i=1}^I \left[\sum_{l=1}^{K_i-1} \frac{l(\alpha - 1)}{(\theta + l)^2 (\theta + l\alpha)} + \sum_{l=K_i}^{n_i-1} \frac{1}{(\theta + l)^2} \right]. \quad (18)$$

By the definition of k_{\max} and k_{\min} ,

$$\begin{aligned} g'(\theta) &< \sum_{i=1}^I \left[\frac{1}{\theta + K_i - 1} \sum_{l=1}^{K_i-1} \frac{2l(\alpha - 1)}{(\theta + l)(\theta + l\alpha)} \right. \\ &\quad \left. + \frac{1}{\theta + K_i} \sum_{l=K_i}^{n_i-1} \frac{1}{(\theta + l)} \right] < \frac{2}{\theta + k_{\max} - 1} \end{aligned} \quad (19)$$

$$\cdot \sum_{i=1}^I \sum_{l=1}^{K_i-1} \frac{l(\alpha - 1)}{(\theta + l)(\theta + l\alpha)} + \frac{1}{\theta + k_{\min}} \sum_{i=1}^I \sum_{l=K_i}^{n_i-1} \frac{1}{(\theta + l)}.$$

The inequality is obtained when $g'(\theta) \geq 0$:

$$\sum_{i=1}^I \sum_{l=1}^{K_i-1} \frac{l(\alpha-1)}{(\theta+l)(\theta+l\alpha)} < \frac{\theta+k_{\max}-1}{2(\theta+k_{\min})} \sum_{i=1}^I \sum_{l=K_i}^{n_i-1} \frac{1}{(\theta+l)}. \quad (20)$$

With (15) and (20), we have

$$g(\theta) < \frac{(\alpha-1)+k_{\max}-2k_{\min}}{2(\theta+k_{\min})} \sum_{i=1}^I \sum_{l=K_i}^{n_i-1} \frac{1}{(\theta+l)}. \quad (21)$$

When $k_{\max} - 2k_{\min} \leq 0$, $g(\theta) < 0$. Particularly, there is a unique solution when $I = 1$ for $k_{\max} = k_{\min} = k_1$. \square

The solution to $\ell_\theta(\alpha, \theta) = 0$ is the MLE of θ and we denote it by $\hat{\theta}_{\text{mle}}$. When $I = 1$, by Pitman and Yor [7], the PD(α, θ) distributions are mutually absolutely continuous when $\alpha \in (0, 1)$ is fixed and θ is varying; hence $\hat{\theta}_{\text{mle}}$ is not a consistent estimator. Some simulation results of $\hat{\theta}_{\text{mle}}$ are reported in Section 5 when $I > 1$; theoretically I can be ∞ .

3.2.3. Estimate Both α and θ . As discussed above, the MLEs for α and θ when both are unknown can be obtained by solving the equations $\ell_\alpha(\alpha, \theta) = \ell_\theta(\alpha, \theta) = 0$. To guarantee the existence of the MLEs for α and θ , we need $(\hat{\alpha}, \hat{\theta})$ of $\ell_\alpha(\alpha, \theta) = \ell_\theta(\alpha, \theta) = 0$ to be such that

$$\det H(\alpha, \theta) = \ell_{\alpha\alpha} \ell_{\theta\theta} - \ell_{\alpha\theta}^2 > 0, \quad (22)$$

where $H(\alpha, \theta) := \begin{pmatrix} \ell_{\alpha\alpha} & \ell_{\alpha\theta} \\ \ell_{\alpha\theta} & \ell_{\theta\theta} \end{pmatrix}$ is the Hessian matrix; that is,

$$\begin{aligned} \ell_{\alpha\alpha} &:= \frac{\partial \ell_\alpha}{\partial \alpha} = -\sum_{i=1}^I \sum_{l=1}^{K_i-1} \frac{l^2}{(\theta+l\alpha)^2} - \sum_{i=1}^I \sum_{j=2}^{n_i} a_j^i \sum_{l=1}^{j-1} \frac{1}{(l-\alpha)^2}, \\ \ell_{\alpha\theta} &:= \frac{\partial \ell_\alpha}{\partial \theta} = -\sum_{i=1}^I \sum_{l=1}^{K_i-1} \frac{l}{(\theta+l\alpha)^2} = \frac{\partial \ell_\theta}{\partial \alpha} =: \ell_{\theta\alpha}, \\ \ell_{\theta\theta} &:= \frac{\partial \ell_\theta}{\partial \alpha} = \sum_{i=1}^I \sum_{l=1}^{n_i-1} \frac{1}{(\theta+l)^2} - \sum_{i=1}^I \sum_{l=1}^{K_i-1} \frac{1}{(\theta+l\alpha)^2}. \end{aligned} \quad (23)$$

3.2.4. Remarks. While we have so far proved the existence of MLEs under the three different cases, analytical expressions of the estimates under none of the cases are given and, thus, the MLEs can only be obtained with numerical methods, for example, the *Newton-Raphson iteration method* or *dichotomy method*.

In addition, by Carlton [16], a naive estimate of α in terms of merely the observations in group i is $\hat{\alpha}_i = \log K_i / \log n_i$, which is consistent as n_i tends to ∞ . Now with the I groups of data, we can estimate α by

$$\begin{aligned} \tilde{\alpha} &= \frac{1}{I} \sum_{i=1}^I \hat{\alpha}_i = \frac{1}{I} \sum_{i=1}^I \frac{\log K_i}{\log n_i} \\ \text{or } \tilde{\alpha}_w &= \frac{\sum_{i=1}^I \log K_i}{\sum_{i=1}^I \log n_i} = \sum_{i=1}^I \frac{\log K_i}{\log n_i} \cdot \frac{\log n_i}{\sum_{i=1}^I \log n_i} \\ &= \sum_{i=1}^I \omega_i \cdot \frac{\log K_i}{\log n_i}, \end{aligned} \quad (24)$$

where $\omega_i = \log n_i / \sum_{i=1}^I \log n_i$. It is then straightforward from Lemma 5.2 in Carlton [16] that both $\tilde{\alpha}$ and $\tilde{\alpha}_w$ are strongly consistent with the true value of α as $\max_{1 \leq i \leq I} n_i \rightarrow \infty$.

4. The Estimation of Base Distribution

In this section, we estimate the base distribution μ of PD($\mu; \alpha, \theta$) under the monotone missing data structure. In Section 4.1, we pick up a set of i.i.d. observations from μ and then estimate the density of μ by kernel estimation method. With the i.i.d. observations, we can estimate the base density μ by multivariate kernel density method that is similar to the one used by Yang and Wu [15] who studied the multivariate density estimation in a Dirichlet process prior under the same missing mechanism and showed that their estimate is superior to Titterton and Mill [22] under asymptotic mean integrated squared errors (AMISE) criterion (note that the classical versions of [22] for complete data can be found in, e.g., [23, 24]).

4.1. Deduce i.i.d. Observations from μ . Denote by $\mathbf{Z}_{i1}, \mathbf{Z}_{i2}, \dots, \mathbf{Z}_{i, K_i}$ the distinct observations of $\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{i, n_i}$. The following theorem shows that, given $K_i, \mathbf{Z}_{i1}, \mathbf{Z}_{i2}, \dots, \mathbf{Z}_{i, K_i}$ are independent and identically $\mu(t_1, t_2, \dots, t_j)$ -distributed—note that, by Lemmas 3 and 4, under the structure of this model, for every group i , though the data are partly missing, the value of K_i can be easily worked out.

Theorem 8. Suppose that $\mu(t_1, t_2, \dots, t_j)$ is continuous; then, given K_i , the distinct observations $\mathbf{Z}_{i1}, \mathbf{Z}_{i2}, \dots, \mathbf{Z}_{i, K_i}$ are i.i.d. with distribution $\mu(t_1, t_2, \dots, t_j)$.

Proof. The proof can be found in Theorem 2.5 in Korwar and Hollander [25]. \square

By Theorem 8, there are K_i i.i.d. observations in group i and hence the number of i.i.d. observations in the whole I groups is $K = \sum_{i=1}^I K_i$. Note that the observed part of the K i.i.d. observations, namely,

$$\{(\mathbf{Z}_{ik1}, \mathbf{Z}_{ik2}, \dots, \mathbf{Z}_{ik, (I-i) \wedge j}) : k = 1, 2, \dots, K_i\}, \quad (25)$$

is still with a monotone missing structure. For every $j = 1, 2, \dots, J$, write $\mathbf{Z}_{ik}^j = (Z_{ik1}, Z_{ik2}, \dots, Z_{ikj})$ for the subvector of the first j components of \mathbf{Z}_{ik} from individual (i, k) , $k = 1, 2, \dots, K_i$, and

$$S_j = \{\mathbf{Z}_{im}^j : 1 \leq m \leq K_i, 1 \leq i \leq I - j + 1\} \quad (26)$$

for the maximal set of the subvectors \mathbf{Z}_{ik}^j , $k = 1, 2, \dots, K_i$, $i = 1, 2, \dots, I$, which are completely observed. Obviously, S_j is made of all the observations with no missing components and S_1 is the set of the first components of all \mathbf{Z}_{ik} in the I groups. Similarly, define

$$\tilde{S}_j = \{\mathbf{Z}_{im}^j : 1 \leq m \leq K_i, 1 \leq i \leq I - j\}. \quad (27)$$

Obviously, $\tilde{S}_j \subset S_j$, for all j .

4.2. *Kernel Density Estimation.* There are two methods to estimate the joint density μ of the random vector \mathbf{Z}_{ik} , proposed by Yang and Wu [15] and Titterington and Mill [22], respectively. They are described below.

(1) *Yang and Wu's Method.* Let $\{\omega_j, j = 1, 2, \dots, J\}$ be a sequence of *consistent* kernel functions—a terminology used in Yang and Wu [15] to indicate that ω_j are density

functions on space \mathbb{R}^j satisfying $\omega_{j-1}(t_1, t_2, \dots, t_{j-1}) = \int \omega_j(t_1, t_2, \dots, t_{j-1}, t_j) dt_j$ for $j = 2, \dots, J$. Then, under the monotone missing data structure, μ_j can be estimated in two different ways: (i) using the consistent property of kernel ω_{j+1} and the data S_{j+1} and (ii) using the ordinary kernel density estimation with data $S_j \setminus \tilde{S}_j$. Formally, the two estimates (denoted by $\hat{\mu}_{j1}$ and $\hat{\mu}_{j2}$, resp.) are given by

$$\begin{aligned} \hat{\mu}_{j1}(t_1, t_2, \dots, t_j) &= \int \frac{1}{h_{j+1}^{j+1} \sum_{i=1}^{I-j} K_i} \sum_{i=1}^{I-j} \sum_{k=1}^{K_i} \omega_{j+1} \left(\frac{t_1 - Z_{ik1}}{h_{j+1}}, \dots, \frac{t_{j+1} - Z_{ik(j+1)}}{h_{j+1}} \right) dt_{j+1} \\ &= \frac{1}{h_{j+1}^j \sum_{i=1}^{I-j} K_i} \sum_{i=1}^{I-j} \sum_{k=1}^{K_i} \omega_j \left(\frac{t_1 - Z_{ik1}}{h_{j+1}}, \frac{t_2 - Z_{ik2}}{h_{j+1}}, \dots, \frac{t_j - Z_{ikj}}{h_{j+1}} \right), \\ \hat{\mu}_{j2}(t_1, \dots, t_j) &= \frac{1}{h_j^{j+1} K_{I-j+1}} \sum_{k=1}^{K_{I-j+1}} \omega_j \left(\frac{t_1 - Z_{(I-j+1)k1}}{h_j}, \dots, \frac{t_j - Z_{(I-j+1)kj}}{h_j} \right). \end{aligned} \tag{28}$$

Therefore, the density $\mu(t_1, t_2, \dots, t_j)$ for $j = 1, 2, \dots, J-1$ can be estimated by

$$\begin{aligned} \hat{\mu}(t_1, t_2, \dots, t_j) \\ = \frac{\sum_{i=1}^{I-j} K_i \hat{\mu}_{j1}(t_1, \dots, t_j) + K_{I-j+1} \hat{\mu}_{j2}(t_1, \dots, t_j)}{\sum_{i=1}^{I-j+1} K_i}. \end{aligned} \tag{29}$$

And $\mu(t_1, t_2, \dots, t_j)$ can be estimated by $\hat{\mu}_{j2}(t_1, t_2, \dots, t_j)$. Furthermore, the consistence of ω_j also indicates that

$$\begin{aligned} \hat{\mu}(t_j | t_1, t_2, \dots, t_{j-1}) &= \frac{\hat{\mu}(t_1, t_2, \dots, t_j)}{\int \hat{\mu}(t_1, t_2, \dots, t_j) dt_j} \\ &\text{for } j = 2, 3, \dots, J. \end{aligned} \tag{30}$$

By a decomposition of the joint density μ as products of a series of conditional densities, that is, $\mu(t_1, t_2, \dots, t_j) = \mu(t_1) \prod_{j=1}^J \mu(t_j | t_1, t_2, \dots, t_{j-1})$, one can estimate the density μ by

$$\hat{\mu}^m(t_1, t_2, \dots, t_j) = \hat{\mu}(t_1) \prod_{j=2}^J \hat{\mu}(t_j | t_1, t_2, \dots, t_{j-1}), \tag{31}$$

where the superscript “ m ” indicates that the estimate is obtained by the monotone missing data structure and $\hat{\mu}(t_1)$ and $\hat{\mu}(t_j | t_1, t_2, \dots, t_{j-1})$ are, respectively, proper estimates for $\mu(t_1)$ and $\mu(t_j | t_1, t_2, \dots, t_{j-1})$.

(2) *Titterington and Mill's Method.* For a vector $\mathbf{X} = (\mathbf{X}^o, \mathbf{X}^m)$, Titterington and Mill [22] used $\int K(\cdot | (\mathbf{x}^o, \mathbf{x}^m), \mathbf{h}) \hat{f}(\mathbf{x}^m | \mathbf{x}^o) d\mathbf{x}^m$ instead of the usual kernel $K(\cdot | \mathbf{x}^o)$ when estimating the density of \mathbf{X}^o , where \mathbf{X}^o and \mathbf{X}^m are, respectively, the observed part and missing part of \mathbf{X} and $\hat{f}(\mathbf{x}^m | \mathbf{x}^o)$ is a probability density function to estimate true $f(\mathbf{x}^m | \mathbf{x}^o)$.

With the above idea, data S_j , and group $(I-j+2)$, another estimate for μ_j is given by

$$\begin{aligned} \hat{\mu}(t_1, t_2, \dots, t_j) \\ = \frac{\sum_{i=1}^{I-j+1} K_i \hat{\mu}_{j3}(t_1, t_2, \dots, t_j) + K_{I-j+2} \hat{\mu}_{j4}(t_1, t_2, \dots, t_j)}{\sum_{i=1}^{I-j+2} K_i}, \end{aligned} \tag{32}$$

where

$$\begin{aligned} \hat{\mu}_{j3}(t_1, t_2, \dots, t_j) &= \frac{1}{h_j^j \sum_{i=1}^{I-j+1} K_i} \\ &\cdot \sum_{i=1}^{I-j+1} \sum_{k=1}^{K_i} \omega_j \left(\frac{t_1 - Z_{ik1}}{h_j}, \frac{t_2 - Z_{ik2}}{h_j}, \dots, \frac{t_j - Z_{ikj}}{h_j} \right), \\ \hat{\mu}_{j4}(t_1, t_2, \dots, t_j) &= \frac{1}{h_{j+1}^j K_{I-j+2}} \\ &\cdot \sum_{k=1}^{K_{I-j+2}} \int \omega_{j+1} \left(\frac{t_1 - Z_{(I-j+2)k1}}{h_{j+1}}, \dots, \right. \\ &\left. \frac{t_{j+1} - Z_{(I-j+2)k(j+1)}}{h_{j+1}} \right) \hat{f}_j(z_{(I-j+2)k(j+1)} | z_{(I-j+2)k1}, \\ & z_{(I-j+2)k2}, \dots, z_{(I-j+2)kj}) dz_{(I-j+2)k(j+1)}. \end{aligned} \tag{33}$$

We should note that $\hat{f}_j(t_{j+1} | t_1, t_2, \dots, t_j)$ is an appropriate estimate of the density of $Z_{(I-j+2)k(j+1)}$ given $(Z_{(I-j+2)k1}, Z_{(I-j+2)k2}, \dots, Z_{(I-j+2)kj})$.

5. Simulation

In this section, we report a small numerical simulation regarding the performance of the estimates of α and θ given in

TABLE 2: Performance of estimates for α with $\theta = 10$.

Size	$\hat{\alpha}$	Bias			SD			MSE		
		0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
Level 1	$\hat{\alpha}_{mle}$	0.091	0.001	0.078	0.043	0.014	0.036	0.010	0.000	0.007
	$\hat{\alpha}_w$	0.530	0.507	0.150	0.021	0.004	0.036	0.282	0.257	0.024
Level 2	$\hat{\alpha}_{mle}$	0.040	0.008	-0.001	0.005	0.005	0.018	0.002	0.000	0.000
	$\hat{\alpha}_w$	0.518	0.391	0.104	0.002	0.003	0.024	0.269	0.153	0.011
Level 3	$\hat{\alpha}_{mle}$	0.042	0.009	-0.001	0.006	0.005	0.004	0.002	0.000	0.090
	$\hat{\alpha}_w$	0.525	0.397	0.276	0.003	0.002	0.003	0.276	0.157	0.331

Section 3. The simulation was proceeded under the following settings:

- (i) $I = 10$ and because the estimates depend only on the observations of K_i s which can be determined by the first component, we set $J = 1$ for convenience.
- (ii) The standard normal distribution $N(0, 1)$ was taken as the base measure μ .
- (iii) The sizes of individuals $(n_1, n_2, \dots, n_{10})$ took three levels of values:
 - level 1: $n_1 = 50, n_2 = 51, \dots, n_9 = 58, n_{10} = 59$,
 - level 2: $n_1 = 500, n_2 = 510, \dots, n_9 = 580, n_{10} = 590$,
 - level 3: $n_1 = 100, n_2 = 200, \dots, n_9 = 900, n_{10} = 1000$.

The procedure of the simulation under those settings is simply described as follows.

Algorithm 9.

Step 1. Generate the data $X_{i1}, X_{i2}, \dots, X_{in_i}$ for every group $i = 1, 2, \dots, I$ by means of a generalized Polya urn model as follows:

- (i) $X_{i1} \mid \alpha, \theta, \mu \sim \mu(\cdot) = N(0, 1)$.
- (ii) $X_{ij}, j = 2, 3, \dots, n_i$, is generated by the prediction rule

$$\begin{aligned}
 & X_{i(j+1)} \mid X_{i1}, \dots, X_{ij}, \alpha, \theta, \mu \\
 & \sim \frac{\theta + \alpha K_{ij}}{\theta + j} \mu(\cdot) + \frac{1}{\theta + j} \sum_{l=1}^{K_{ij}} (m_l - \alpha) \delta_{Z_{il}}(\cdot), \tag{34}
 \end{aligned}$$

where Z_{il} is the l th distinct value among X_{i1}, \dots, X_{ij} , $m_l = \#\{r : X_{ir} = Z_{il}, 1 \leq r \leq j\}$ is the number of occurrences of Z_{il} , and K_{ij} is the number of distinct values of X_{i1}, \dots, X_{ij} .

Step 2. Compute the values of K_i and a_j^i from the data generated in Step 1.

Step 3. Compute the estimates of α, θ with Newton-Raphson method.

The simulation compared the estimates $\hat{\alpha}_{mle}$ and $\hat{\alpha}_w$ of α for known θ in which we set $\theta = 10$ and α took three values: 0.1, 0.3, and 0.5; the case where $\alpha > 0.5$ can simply be obtained

TABLE 3: Performance of the estimate of θ with $\alpha = 0.5$.

Sample size	θ				
	0.1	1	5	10	
Bias	Level 1	-0.018	0.067	0.107	0.047
	Level 2	0.019	0.052	0.022	0.081
	Level 3	0.012	0.012	0.114	0.032
SD	Level 1	0.230	0.429	0.971	1.653
	Level 2	0.210	0.369	0.746	1.105
	Level 3	0.202	0.369	0.779	1.187
MSE	Level 1	0.053	0.189	0.954	2.821
	Level 2	0.044	0.138	0.556	1.228
	Level 3	0.041	0.137	0.620	1.410

by replacing α with $1 - \alpha$. For each combination of the parameters, the simulation was proceeded 1000 replications. The performances of the two estimates are shown in Table 2, from which it is clearly shown that the simulated biases of $\hat{\alpha}_{mle}$ are all smaller than those of $\hat{\alpha}_w$. Although the simulated standard deviations of $\hat{\alpha}_w$ are stable and smaller in three levels under three different values, the MSEs of $\hat{\alpha}_{mle}$ are smaller than those of $\hat{\alpha}_w$. That is, $\hat{\alpha}_{mle}$ performs better than $\hat{\alpha}_w$ under the MSE criterion.

Next, the simulation studied the performance of the estimates of θ with $\alpha = 0.5$ at four true values: $\theta = 0.1, 1, 5$, and 10. The simulation was also proceeded 1000 replications for each combination of the parameters. The simulation results are summarized in Table 3, from which we can see that the estimates $\hat{\theta}_{mle}$ generally performed better with the increasing of the sample size.

Finally, we considered the MLEs for α and θ when neither of the parameters is known. The true values of the parameters and the performance of the estimates of the 1000 replications are shown in Table 4. The simulation used the Newton-Raphson iteration method.

Favaro et al. [10] adopted an empirical Bayes procedure for estimating (α, θ) , writing $(\hat{\alpha}, \hat{\theta})_F$, when the data is not divided into groups. They investigated the issue of prediction within species sampling problems and applications to genomics based on the estimates $(\hat{\alpha}, \hat{\theta})_F$ when both parameters are unknown. Favaro et al. [10] also studied the asymptotic behavior of the number of new species conditionally on the observed sample and derived asymptotic highest posterior density intervals for the estimates of their interest. The comparisons of simulation results between $(\hat{\alpha}, \hat{\theta})_F$ and

TABLE 4: Performance of the estimates for (α, θ) under level 1.

(α, θ)	$(\hat{\alpha}, \hat{\theta})$	Bias	SD	MSEs
(0.1, 10)	$(\hat{\alpha}, \hat{\theta})_{mle}$	(0.064, -1.209)	(0.095, 2.098)	(0.013, 5.864)
	$(\hat{\alpha}, \hat{\theta})_F$	(0.153, -3.762)	(0.380, 3.644)	(0.038, 27.429)
(0.2, 4)	$(\hat{\alpha}, \hat{\theta})_{mle}$	(0.171, -1.621)	(0.100, 1.193)	(0.039, 4.052)
	$(\hat{\alpha}, \hat{\theta})_F$	(0.207, -2.200)	(0.106, 1.757)	(0.054, 7.925)
(0.25, 6)	$(\hat{\alpha}, \hat{\theta})_{mle}$	(0.049, -0.663)	(0.109, 2.080)	(0.014, 4.765)
	$(\hat{\alpha}, \hat{\theta})_F$	(0.105, -1.637)	(0.118, 3.457)	(0.025, 14.626)
(0.3, 7)	$(\hat{\alpha}, \hat{\theta})_{mle}$	(0.011, -0.094)	(0.100, 2.234)	(0.010, 5.001)
	$(\hat{\alpha}, \hat{\theta})_F$	(0.039, -0.121)	(0.0.117, 4.852)	(0.015, 23.557)
(0.4, 8)	$(\hat{\alpha}, \hat{\theta})_{mle}$	(-0.009, 0.429)	(0.095, 2.867)	(0.009, 8.405)
	$(\hat{\alpha}, \hat{\theta})_F$	(-0.020, 1.083)	(0.103, 5.242)	(0.011, 28.653)
(0.5, 5)	$(\hat{\alpha}, \hat{\theta})_{mle}$	(-0.001, 0.263)	(0.077, 2.160)	(0.006, 4.734)
	$(\hat{\alpha}, \hat{\theta})_F$	(-0.008, 1.434)	(0.084, 4.528)	(0.007, 22.562)

the MLEs which are under the multigroups data structure are presented in Table 4 with the total size of individuals in the perspective of bias, SD, and MSEs. From Table 4, we can see that the MSE of $\hat{\alpha}$ is very small and the MSE of $\hat{\theta}$ is a bit bigger. Apparently, the simulation shows that the MLEs generally are better than $(\hat{\alpha}, \hat{\theta})_F$ of single group data under the criteria of bias, SD, and MSEs.

6. Conclusion

In this paper, we studied the estimates of the unknown parameters as well as the density of the base measure in Poisson-Dirichlet process priors under monotonically missing data. The parameters are estimated by the method of MLEs and a set of simulations shows that the estimates perform well.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by NSFC under Grant no. 71371074.

References

[1] T. S. Ferguson, "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, vol. 1, pp. 209–230, 1973.

[2] T. S. Ferguson, "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, vol. 2, no. 4, pp. 615–629, 1974.

[3] M. Lavine, "Some aspects of polya tree distributions for statistical modelling," *The Annals of Statistics*, vol. 20, no. 3, pp. 1222–1235, 1992.

[4] M. Lavine, "More aspects of Pólya tree distributions for statistical modelling," *The Annals of Statistics*, vol. 22, no. 3, pp. 1161–1176, 1994.

[5] R. D. Mauldin, W. D. Sudderth, and S. C. Williams, "Pólya trees and random distributions," *The Annals of Statistics*, vol. 20, no. 3, pp. 1203–1221, 1992.

[6] J. F. C. Kingman, "Random discrete distributions," *Journal of the Royal Statistical Society: Series B*, vol. 37, pp. 1–22, 1975.

[7] J. Pitman and M. Yor, "The Poisson-Dirichlet distribution derived from a stable subordinator," *Annals of Probability*, vol. 25, no. 2, pp. 855–900, 1997.

[8] S. Goldwater, T. Griffiths, and M. Johnson, "Interpolating between types and tokens by estimating power-law generators," *Advances in Neural Information processing Systems*, vol. 18, pp. 459–466, 2006.

[9] E. B. Sudderth and M. Jordan, *Shared Segmentation of Natural Scenes Using Dependent Pitman-Yor Processes*, Advances in Neural Information Processing Systems 21, MIT Press, 2009.

[10] S. Favaro, A. Lijoi, R. H. Mena, and I. Prünster, "Bayesian non-parametric inference for species variety with a two-parameter Poisson-Dirichlet process prior," *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, vol. 71, no. 5, pp. 993–1008, 2009.

[11] N. Hoshino, "Applying Pitmans sampling formula to microdata risk assessment," *Journal of Official Statistics*, vol. 17, pp. 499–520, 2001.

[12] M. Johnson, T. L. Griffiths, and S. Goldwater, "Adaptor grammars: a framework for specifying compositional nonparametric Bayesian models," *Advances in Neural Information processing Systems*, vol. 19, pp. 641–648, 2007.

[13] F. Wood and Y. W. Teh, "A hierarchical nonparametric Bayesian approach to statistical language model domain adaptation," *Journal of Machine Learning Workshop and Conference Proceedings: Artificial Intelligence in Statistics*, vol. 5, pp. 607–614, 2009.

[14] H. Wallach, C. Sutton, and A. McCallum, "Bayesian modeling of dependency trees using hierarchical Pitman-Yor priors," in *Proceedings of the Workshop on Prior Knowledge for Text and Language*, pp. 15–20, 2008.

[15] L. Yang and X. Wu, "Estimation of Dirichlet process priors with monotone missing data," *Journal of Nonparametric Statistics*, vol. 25, no. 4, pp. 787–807, 2013.

[16] M. A. Carlton, *Applications of the Poisson-Dirichlet distribution [Dissertation]*, University of California, 1999.

[17] J. Huang, C. Qiu, X. Wu, and X. Zhou, "An individual loss reserving model with independent reporting and settlement," *Insurance: Mathematics and Economics*, vol. 64, pp. 232–245, 2015.

[18] J. Huang, X. Wu, and X. Zhou, "Asymptotic behaviors of stochastic reserving: aggregate versus individual models," *European Journal of Operational Research*, vol. 249, no. 2, pp. 657–666, 2016.

[19] M. M. Marini, A. R. Olsen, and D. B. Rubin, "Maximum-likelihood estimation in panel studies with missing data," *Sociological Methodology*, vol. 11, pp. 314–357, 1980.

[20] J. Hao and K. Krishnamoorthy, "Inferences on a normal covariance matrix and generalized variance with monotone missing data," *Journal of Multivariate Analysis*, vol. 78, no. 1, pp. 62–82, 2001.

[21] V. M. Raats, B. B. van der Genugten, and J. J. Moors, "Multivariate regression with consecutively added dependent variables," *Linear Algebra and its Applications*, vol. 410, pp. 170–197, 2005.

[22] D. M. Titterington and G. M. Mill, "Kernel-based density estimates from incomplete data," *Journal of the Royal Statistical Society B: Methodological*, vol. 45, no. 2, pp. 258–266, 1983.

- [23] T. Cacoullos, "Estimation of a multivariate density," *Annals of the Institute of Statistical Mathematics*, vol. 18, pp. 179–189, 1966.
- [24] M. P. Wand and M. C. Jones, "Multivariate plug-in bandwidth selection," *Computational Statistics*, vol. 9, pp. 97–166, 1994.
- [25] R. M. Korwar and M. Hollander, "Contributions to the theory of dirichlet processes," *The Annals of Probability*, vol. 1, no. 4, pp. 705–711, 1973.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

