JIRSS (2022)
Vol. 21, No. 02, pp 111-132
DOI: 10.22034/jirss.2022.707638

The Construction of Generalized Dirichlet Process Distributions via Pólya urn and Gibbs Sampling

Hassan Akell [1], Farkhondeh-Alsadat Sajadi [1], Iraj Kazemi [1]

[1] Department of Statistics, Faculty of Mathematics & Statistics, University of Isfahan.

Received: 15/02/2023, Accepted: 15/09/2023, Published online: 14 /01/2024

Abstract. Bayesian nonparametric inference is increasingly demanding in statistical modeling due to incorporating flexible prior processes in complex data analysis. This paper represents the Pólya urn scheme for the generalized Dirichlet process (GDP). It utilizes the partition analysis to construct the joint distribution of a random sample from the GDP as a mixture prior distribution of countable components. Using permutation theory, we present the components' weights in a computationally accessible manner to make the resulting joint prior equation applicable. The advantages of our findings include tractable algebraic operations that lead to closed-form equations. The paper recommends the Pólya urn Gibbs sampler algorithm, derive full conditional posterior distributions, and as an illustration, implement the algorithm for fitting some popular statistical models in nonparametric Bayesian settings.

Keywords. Exchangeability, GDP Mixture Model, Partition Analysis, Permutations, Pólya urn, Gibbs Sampler, Stick-breaking Priors.

MSC 2020: 60G20, 60J46, 62E15.

1 Introduction

Bayesian nonparametric models are flexible and robust alternatives to parametric inference. Several data analysis methods have been extensively developed in many fields, including health, biology, and financial studies. In the Bayesian framework, incorporating nonparametric priors into the data analysis process allows analysts to

Hassan Akell (h.akell@sci.ui.ac.ir)
Farkhondeh-Alsadat Sajadi (f.sajadi@sci.ui.ac.ir)
Corresponding Author: Iraj Kazemi (i.kazemi@stat.ui.ac.ir)

achieve the most information from an infinite mixture of stochastic priors. For this purpose, Ferguson (1973) introduces the Dirichlet process (DP) as a class of random measures wherein any realization of the DP constitutes a probability distribution. The Dirichlet process mixture model was then presented to facilitate computational aspects (Antoniak , 1974; Ferguson , 1983; Lo , 1984). Several researchers have developed the methodology for more practical models (Escobar and West , 1995; Neal , 2000), among many others.

The discreteness property of the DP was illustrated by Ferguson (1973) using the gamma representation. Blackwell and MacQueen (1973) drive underlying conditional distributions through the Pólya urn scheme (PUS). An interesting feature of the DP is the stick-breaking construction as an infinite series representation (Sethuraman and Tiwari , 1982; Sethuraman , 1994). Extending the DP to complex models requires knowledge of partitions, permutations, and modern algebraic operations. Combinatorial structures, such as Pitman partitions, Gibbs partitions, random trees, and Bell polynomials, need algebraic computations and are essential knowledge for the so-called combinatorial stochastic processes. The DP is crucial in revealing the relationship between combinatorial stochastic processes and nonparametric Bayesian inference. For a general overview, see Pitman (2006), Phadia (2016), Ghosal, and Van der Vaart (2017), and Mano (2018).

The stick-breaking representation demonstrates the Dirichlet measure through the construction of infinite series. Sethuraman's construction is the most attractive approach to developing a collection of nonparametric Bayesian priors (Hjort , 2000; Ishwaran and Zarepour , 2000; Pitman and Yor , 1997). The generalized Dirichlet process (GDP) presented by Hjort (2000) as a prior process allows further data analysis flexibility. Hjort (2000) computed theoretically the probability of an event on which all data points from the GDP were distinct. The beta two-parameter process (BTPP) was shown to be a particular case of the GDP (Ishwaran and Zarepour , 2000). Rodriguez and Dunson (2014) studied the clustering property of the GDP. Barcella (2017) discussed the truncated GDP for the random truncation. Barcella et al. (2018b) introduced the dependent GDP as a generalization of the dependent DP presented by MacEachern (1999). For recent applications of the DP, see Barcella et al. (2018a), Molinari, et al. (2021), and Aghabazaz et al. (2023).

Most applications of traditional statistical models require parametric specifications of probability distributions. A challenging issue arises if insufficient prior information involves justifying such parametric assumptions. In this case, Bayesian nonparametric methods allow adaptable specifications of prior distributions. The Dirichlet process mixture can be viewed as an infinite dimensional mixture model at the most basic level. It motivated us to offer a flexible class of priors with an explicit and convenient prediction rule in nonparametric settings. The class relaxes traditional parametric assumptions and adjusts model fitting, particularly once a generalized Pólya urn mechanism characterizes the prior. Mainly, our paper represents the Pólya urn for the GDP and uses the partition analysis to find the joint distribution of samples from the GDP.

An advantage of the findings includes tractable closed-form equations for the joint distribution as a countable mixture of accessible components' weights.

Bayesian nonparametric models can improve data analysis results in broad applications of various statistical models, including clustering and latent class analysis, density estimation, and prior specification when little information is available. An illustrative example used in this article is about the traditional Binomial/Beta model, which assumes that each experiment is a Binomial draw with unknown proportions X_i and known sample sizes n_i for $i = 1, \ldots, k$ groups, with a conjugate Beta prior for X_i. However, if the empirical evidence reveals bimodality, a single Beta cannot be a suitable prior choice. To address this concern, an appropriate method would involve effectively utilizing a Dirichlet process prior. Since proportions lie between 0 and 1, a pragmatic choice is a Dirichlet process mixture of Beta distributions. The working data set here is binary strings from rolls of common thumbtacks (Beckett and Diaconis , 1994; Liu , 1996). The flicks are presumed independent conditionally on the tack. There are 320 observations, relating to the thumbtack role, with each tack flipping nine times. The output variable Y_i is the number of times each tack landed point up. We extend model fitting to the data analysis in Section 6.

Section 2 exhibits preliminaries on the DP and its representations, the GDP, and the partition analysis. Based on permutation theory, section 3 provides a valuable expression for the joint distribution of a random sample $X_1, \ldots, X_n$ drew from the GDP, discusses its exchangeability property, and provides a simulation study to illustrate some properties of the process. Section 4 presents the Pólya urn representation of the GDP, computes the predictive and conditional distributions, shows the exchangeable partition probability function, and derives the PUS representation of the BTPP. Moreover, we verify our method for the DP, which matches the findings in the literature. Section 5 defines the GDP mixture models and offers the Pólya urn Gibbs sampler to insert the GDP into the nonparametric Bayesian approach. Section 6 analyzes the thumbtacks data using the Binomial/Beta GDP mixture model. Section 7 includes concluding remarks.

2 Background on the Dirichlet Process

Let $\mathcal{X}$ be a set and $\mathfrak{F}$ be a nonempty collection of subsets of $\mathcal{X}$, which is a $\sigma-$field. The Dirichlet process DP(ν) with base measure ν is a random probability measure G on the measurable space $(\mathcal{X}, \mathfrak{F})$, where the stochastic process $(G(A) : A \in \mathfrak{F})$ is indexed by measurable subsets, and sample paths are probability measures with probability one. It means that G is defined as a distribution over probability measures, such that for every finite measurable partition $A_1, \ldots, A_n$ of $\mathcal{X}$ (i.e., $A_i \in \mathfrak{F}$ for all i, $A_i \cap A_j = \emptyset$ for $i \neq j$, and $\bigcup_{i=1}^{n} A_i = \mathcal{X}$) and base measure ν, the finite-dimensional random vector $(G(A_1), \ldots, G(A_n))$ is distributed as a Dirichlet distribution with parameter $(\nu(A_1), \ldots, \nu(A_n))$. Usually, ν is governed by two parameters, $\nu(\cdot) = bG_0(\cdot)$. $G \sim \mathrm{DP}(b, G_0)$ denotes the DP measure, where the total mass $b = \nu(\mathcal{X})$ is the prior precision, and the probability measure $G_0(\cdot) = \nu(\cdot)/\nu(\mathcal{X})$ is the center measure. We have $\mathbb{E}[G(A)] = G_0(A)$ and $\mathrm{Var}[G(A)] = G_0(A)(1 - G_0(A))/(b + 1)$ for any $A \in \mathfrak{F}$. Accordingly, G_0 and b are named as the mean and inverse variance of the DP, respectively.

Blackwell and MacQueen (1973) present the generalized PUS representation of the DP. Through n steps of the PUS, they find the marginal distribution of a set of a random sample $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ from the DP(b, G_0), in which each step $1 \le i \le n$ provides the conditional distribution of X_i given the previous, i.e., $X_i|X_1, \ldots, X_{i-1} \sim (b + i - 1)^{-1}\Big(bG_0(\cdot) + \sum_{j=1}^{i-1} \delta_{X_j}(\cdot)\Big)$, where $\delta_X(\cdot)$ denotes a point mass centered at $\boldsymbol{X}$. Let $\boldsymbol{X}_{-n} = \{X_1, \ldots, X_{n-1}\}$ be $\boldsymbol{X}$ excluding X_n, and let $\boldsymbol{X}^*_m = \left\{X^*_1, \ldots, X^*_m\right\}$ be the set of distinct values among $\boldsymbol{X}_{-n}$ with occurrences $n_1, \ldots, n_m$ respectively; $1 \le m \le n$ and $\sum_{r=1}^{m} n_r = n - 1$. Following is the generalized PUS for the DP(b, G_0),

$$\begin{aligned} X_1 &\sim G_0, \\ X_n \mid X_1, \ldots, X_{n-1} &\begin{cases} = X^*_r & \text{with probability} \quad \frac{n_r}{b+n-1} \\ \sim G_0 & \text{with probability} \quad \frac{b}{b+n-1} \end{cases} . \end{aligned} \tag{2.1}$$

The discreteness property of the DP implies that the new observation X_n can be either equal to one of the distinct values or take a new value from G_0.

2.1 The Generalized Dirichlet Process

Let $V_1, V_2, \ldots$ be a sequence of independent random variates drawn from a distribution H with support on $(0, 1)$. Consider a sequence of random variates $\{\gamma_i\}_{i\ge1}$ obtained from the set $\{V_i\}_{i\ge1}$, where $\gamma_1 = V_1$, $\gamma_i = V_i \prod_{j=1}^{i-1} \overline{V}_j, i \ge 2$, and $\overline{V}_j = 1 - V_j$. The inequality $0 < \gamma_i < 1$ holds such that $\sum_{i=1}^{\infty} \gamma_i = 1$ a.s., (almost surely). Thus, a random probability measure G can be defined on $(\mathcal{X}, \mathfrak{F})$ as

$$G(\cdot) = \sum_{i=1}^{\infty} \gamma_i \delta_{\xi_i}(\cdot), \tag{2.2}$$

where the random elements ξ_i are independently and identically distributed (iid) drawn from the centered measure distribution G_0, independent of γ_i's. Hjort (2000) defined (2.2) as the GDP, denoted by $G \sim \mathrm{GDP}(H, G_0)$. It is described by the center measure G_0 and the distribution H over $(0, 1)$. Ishwaran and Zarepour (2000) consider H as Beta(a, b) and refer to (2.2) as the beta two-parameter process, denoted by $G \sim \mathrm{BTPP}(a, b, G_0)$. A particular case $a = 1$ to construct the stick-breaking representation of the DP (Sethuraman , 1994), denoted by $G \sim \mathrm{DP}(b, G_0)$, where b is the precision parameter.

Let G be a random probability measure of the form (2.2) and $A_1, \ldots, A_n$ be measurable sets. Hjort (2000) derived the marginal distribution of a random sample $X_1, \ldots, X_n$ from the GDP, satisfying $X_1 \in A_1, \ldots, X_n \in A_n$ as follows

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \mathbb{E}\Big[\prod_{i=1}^{n} G(A_i)\Big]. \tag{2.3}$$

Let $M_{i,j} := \mathbb{E}[V^i \overline{V}^j] = \int_0^1 u^i(1-u)^j dH(u)$ and $\rho_{k,j} = M_{k-j,j}/(1-M_{0,k})$, where i, j, k are nonnegative integers with $k \geq j$. Applying (2.3) for $n = 1, 2$,

$$\mathbb{P}(X_1 \in A_1) = \sum_{i=1}^{\infty} \mathbb{E}[\gamma_i]\mathbb{E}[\delta_{\xi_i}(A_1)] = G_0(A_1)\mathbb{E}\Big[\sum_{i=1}^{\infty} \gamma_i\Big] = G_0(A_1).$$

Therefore, the distribution of one observation is the center measure, i.e., $X_1 \sim G_0$,

$$\mathbb{P}\left(X_1 \in A_1, X_2 \in A_2\right) = \left[\rho_{2,0}\ 2\rho_{2,1}\right]\begin{bmatrix} G_0(A_1 \cap A_2) \\ G_0(A_1)G_0(A_2) \end{bmatrix}, \tag{2.4}$$

where $\mathbb{P}(X_1 = X_2) = \rho_{2,0}$, $\mathbb{P}(X_1 \neq X_2) = 2\rho_{2,1}$, and $\rho_{2,0} + 2\rho_{2,1} = 1$. Also, the expression for disjoint $A_1, \ldots, A_n$ (i.e., the distinct data points) is

$$\mathbb{E}\Big[\prod_{i=1}^{n} G(A_i)\Big] = \beta\, G_0(A_1) \ldots G_0(A_n)\,, \tag{2.5}$$

where β is the probability of drawing distinct data points computed by

$$\beta = \mathbb{P}(X_1 \neq \cdots \neq X_n) = n! \prod_{j=0}^{n-1} \rho_{j+1,j}, \tag{2.6}$$

One can derive the probability of drawing identical data points from GDP(H, G_0) (Ishwaran and Zarepour , 2000) as

$$\mathbb{P}(X_1 = \cdots = X_n) = \frac{M_{n,0}}{1 - M_{0,n}} = \rho_{n,0}. \tag{2.7}$$

The following sub-section is devoted to the partition analysis, which helps compute a random sample's joint distribution from the GDP.

2.2 Partition Analysis

Combinatorial stochastic processes can be better understood through the use of random partitions and random permutations. Consider a set $\boldsymbol{S}_n = \{1, \ldots, n\}$, and disjoint sets

$$E_1 = \{e_1, \ldots, e_{n_1}\}, \ldots, E_m = \{e_{n-n_m+1}, \ldots, e_n\}, \tag{2.8}$$

with sizes $|E_1| = n_1, \ldots, |E_m| = n_m$, where $\sum_{j=1}^{m} n_j = n$, $1 \leq m \leq n$ and each element $e_j \in \boldsymbol{S}_n; j = 1, \ldots, n$. Then $\boldsymbol{E} = \{E_1, \ldots, E_m\}$ refers to a collection of nonempty disjoint sets with $\bigcup_{i=1}^{m} E_i = \boldsymbol{S}_n$. The collection $\boldsymbol{E}$ is one partition of $\boldsymbol{S}_n$, and the sets $E_1, \ldots, E_m$ are called blocks of $\boldsymbol{E}$. We denote $\boldsymbol{n} = \{n_1, \ldots, n_m\}$ the block sizes, where $\sum_{\iota=1}^{m} n_\iota = n$ is fulfilled. Let $\mathfrak{P}(\boldsymbol{n})$ denote the collection of possible partitions of $\boldsymbol{S}_n$ with m blocks and size set $\boldsymbol{n}$. The cardinality of $\mathfrak{P}(\boldsymbol{n})$ (i.e., the number of partitions in $\mathfrak{P}(\boldsymbol{n})$) is

$$|\mathfrak{P}(\boldsymbol{n})| = \frac{n!}{\prod_{i=1}^{n} (i!)^{\gamma_i} \gamma_i!}\,, \tag{2.9}$$

where $\gamma_i = \sum_{j=1}^{m} I_{(n_j=i)}$ is the number of blocks with size i, see Pitman (2006) and Andrews (1998) for details.

It is essential to recognize that different types of $\boldsymbol{n}$ lead to varying $\mathfrak{P}(\boldsymbol{n})$ types. Let $\boldsymbol{N}_n = \{\{n\}, \{n-1, 1\}, \ldots, \{1, \ldots, 1\}\}$ denotes the collection of all formable types of $\boldsymbol{n}$ where $\sum_{\iota=1}^{m} n_\iota = n$ and $1 \le m \le n$. The cardinality of $\boldsymbol{N}_n$ is analogous to the number of all possible ways of expressing n as a sum of positive integers, denoted by $a(n) = |\{n, (n-1)+1, \ldots, 1+\ldots+1\}|$, and called the number of partitions of n, given by

$$a(n) = \sum_{k=1}^{n} \lambda(k) a(n-k), \tag{2.10}$$

which depends on Euler's recurrence relation, where $a(0) = 1$ and

$$\lambda(k) = \begin{cases} (-1)^{j+1} & ; \quad k = \frac{j(3j\mp1)}{2};\ j \in \mathbb{N}^+ \\ 0 & ; \quad o.w. \end{cases}.$$

Starting with $a(1) = 1$, the first few $a(n)$'s are $a(2) = 2$, $a(3) = 3$, $a(4) = 5$ and $a(5) = 7$; for online computation of $a(n)$ see https://oeis.org/A000041. The collection of all partitions of $\boldsymbol{S}_n$ can now be defined by

$$\mathfrak{P}(\boldsymbol{S}_n) = \bigcup_{\boldsymbol{n} \in \boldsymbol{N}_n} \mathfrak{P}(\boldsymbol{n}), \tag{2.11}$$

with cardinality $\mathcal{B}_n = |\mathfrak{P}(\boldsymbol{S}_n)| = \sum_{\boldsymbol{n}\in\boldsymbol{N}_n} |\mathfrak{P}(\boldsymbol{n})|$. In literature, the number of all partitions is called the *Bell number*, given by $\mathcal{B}_n = \sum_{k=0}^{n-1} C(n-1,k)\mathcal{B}_k$, or by $\mathcal{B}_n = \mathbb{E}\left[Z^n\right]$ where $Z \sim Poisson(1)$ and $\mathcal{B}_0 = 1$. The first few $\mathcal{B}_n$'s are $\mathcal{B}_1 = 1$, $\mathcal{B}_2 = 2$, $\mathcal{B}_3 = 5$, $\mathcal{B}_4 = 15$, and $\mathcal{B}_5 = 52$ (Pitman , 2006; Castellares, Ferrari and Lemonte , 2018). To clarify the above discussion, we present two cases, $n = 3, 4$, as follows.

Examples 2.1. Let 12|3 be an abbreviation for the partition $\{\{1,2\},\{3\}\}$, and 12|3 [3] be an abbreviation for three partitions $\{12|3, 13|2, 23|1\}$, and the same for others.

Case $n = 3$: We have $\boldsymbol{S}_3 = \{1,2,3\}$, $a(3) = 3$ and $\boldsymbol{N}_3 = \{\{3\},\{2,1\},\{1,1,1\}\}$.

(i) For $\boldsymbol{n} = \{3\}$, we have $\mathfrak{P}(\boldsymbol{n}) = \boldsymbol{S}_3$, with $|\mathfrak{P}(\boldsymbol{n})| = \frac{3!}{3!} = 1$.

(ii) For $\boldsymbol{n} = \{2,1\}$, we have $\mathfrak{P}(\boldsymbol{n}) = 12|3\ [3]$, with $|\mathfrak{P}(\boldsymbol{n})| = \frac{3!}{2!} = 3$.

(iii) For $\boldsymbol{n} = \{1,1,1\}$, we have $\mathfrak{P}(\boldsymbol{n}) = 1|2|3$, with $|\mathfrak{P}(\boldsymbol{n})| = \frac{3!}{3!} = 1$.

Thus, $\mathfrak{P}(\boldsymbol{S}_3) = \bigcup_{\boldsymbol{n}\in\boldsymbol{N}_3} \mathfrak{P}(\boldsymbol{n}) = \{\boldsymbol{S}_3, 12|3\ [3], 1|2|3\}$, with $\mathcal{B}_3 = \sum_{\boldsymbol{n}\in\boldsymbol{N}_n} |\mathfrak{P}(\boldsymbol{n})| = 5$.

Case $n = 4$: $\boldsymbol{S}_4 = \{1,2,3,4\}$, $a(4) = 5$ and $\boldsymbol{N}_4 = \{\{4\},\{3,1\},\{2,2\},\{2,1,1\},\{1,1,1,1\}\}$.

(i) For $\boldsymbol{n} = \{4\}$, we have $\mathfrak{P}(\boldsymbol{n}) = \boldsymbol{S}_4$, with $|\mathfrak{P}(\boldsymbol{n})| = \frac{4!}{4!} = 1$.

(ii) For $\boldsymbol{n} = \{3,1\}$, we have $\mathfrak{P}(\boldsymbol{n}) = 123|4\ [4]$, with $|\mathfrak{P}(\boldsymbol{n})| = \frac{4!}{3!} = 4$.

(iii) For $\boldsymbol{n} = \{2,2\}$, we have $\mathfrak{P}(\boldsymbol{n}) = 12|34\ [3]$, with $|\mathfrak{P}(\boldsymbol{n})| = \frac{4!}{(2!)^3} = 3$.

(iv) For $\boldsymbol{n} = \{2, 1, 1\}$, we have $\mathfrak{P}(\boldsymbol{n}) = 12|3|4$ [6], with $|\mathfrak{P}(\boldsymbol{n})| = \frac{4!}{(2!)^2} = 6$.

(v) For $\boldsymbol{n} = \{1, 1, 1, 1\}$, we have $\mathfrak{P}(\boldsymbol{n}) = 1|2|3|4$, with $|\mathfrak{P}(\boldsymbol{n})| = \frac{4!}{4!} = 1$.

Therefore, $\mathfrak{P}(\boldsymbol{S}_4) = \{\boldsymbol{S}_4, 123|4\ [4], 12|34\ [3], 12|3|4\ [6], 1|2|3|4\}$, with $\mathcal{B}_4 = 15$.

We now derive a novel combinatorial formula for the joint distribution of a random sample from the GDP.

3 The Joint Distribution of GDP Samples

Let $X_1, \ldots, X_n$ be a random sample from the GDP, satisfying $X_1 \in A_1, \ldots, X_n \in A_n$ with arbitrary measurable sets $A_1, \ldots, A_n$. The joint distribution of the sample from GDP is obtained by substituting (2.2) in (2.3) and utilizing particular mathematical combinatorics. We set $B_1 = \{A_{e_1}, \ldots, A_{e_{n_1}}\}, \ldots, B_m = \{A_{e_{n-n_m+1}}, \ldots, A_{e_n}\}$ with $\{e_1, \ldots, e_n\} = \{1, \ldots, n\}$, $1 \le m \le n$. The collection $\{B_1, \ldots, B_m\}$ is a one partition of $\{A_1, \ldots, A_n\}$ with sizes $\boldsymbol{n} = \{n_1, \ldots, n_m\}$ and $\sum_{\iota=1}^{m} n_\iota = n$. For simplicity we replace $\{A_1, \ldots, A_n\}$ by the index set $\boldsymbol{S}_n = \{1, \ldots, n\}$, and $\{B_1, \ldots, B_m\}$ by the collection of subsets $\boldsymbol{E} = \{E_1, \ldots, E_m\}$ described in (2.8). As mentioned in Section 2.2, the $\boldsymbol{E}$ blocks are formed one partition of $\boldsymbol{S}_n$ with block sizes $\boldsymbol{n}$, i.e., $\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{S}_n)$. For each block E of $\boldsymbol{E}$, let $\Psi_E = G_0\left(\bigcap_{e \in E} A_e\right)$ and $Q_{\boldsymbol{E}} = \prod_{E \in \boldsymbol{E}} \Psi_E$. We present the general joint distribution for n observations in the following Theorem.

Theorem 3.1. *Let G follow $GDP(H, G_0)$ and $A_1, \ldots, A_n$ be measurable sets. Then*

$$\mathbb{P}(X_1 \in A_1, ..., X_n \in A_n) = \sum_{\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{S}_n)} \boldsymbol{w}_{\boldsymbol{E}} Q_{\boldsymbol{E}}\,, \tag{3.1}$$

where each $\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{S}_n)$ consists of blocks determined by (2.8) *and the weight*

$$\boldsymbol{w}_{\boldsymbol{E}} = \sum_{\substack{i_1, \ldots, i_m \\ \text{distinct}}} \mathbb{E}\Big[\prod_{\iota=1}^{m} \gamma_{i_\iota}^{n_\iota}\Big]. \tag{3.2}$$

Proof. We know $\prod_{i=1}^{k} \delta_\xi(A_i) = \delta_\xi\Big(\bigcap_{i=1}^{k} A_i\Big)$; $k \le n$, and

$$\prod_{i=1}^{n} G(A_i) = \sum_{i_1, \ldots, i_n} \prod_{\iota=1}^{n} \gamma_{i_\iota} \delta_{\xi_{i_\iota}}(A_\iota) := \sum_{\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{S}_n)} C_{\boldsymbol{E}}\,, \tag{3.3}$$

where $\prod_{i=1}^{n} G(A_i)$ is defined as what is on the right-hand side after expanding the sum $\sum_{i_1, \ldots, i_n}$, and every $C_{\boldsymbol{E}}$ constituted using one partition $\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{S}_n)$ as

$$C_{\boldsymbol{E}} = \sum_{\substack{i_1, \ldots, i_m \\ \text{distinct}}} \prod_{\iota=1}^{m} \gamma_{i_\iota}^{n_\iota} \delta_{\xi_{i_\iota}}\Big(\bigcap_{e \in E_\iota} A_e\Big). \tag{3.4}$$

Each $\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{S}_n)$ contains m blocks $\{E_1, \ldots, E_m\}$; $1 \leq m \leq n$ with sizes $|E_1| = n_1, \ldots,$ $|E_m| = n_m$; $\sum_{\iota=1}^{m} n_\iota = n$, and the elements of blocks given in (2.8). The proof is complete by taking the expectation of both sides of (3.3), where $\mathbb{E}\Big[\delta_{\xi_{i_\iota}}\Big(\bigcap_{e\in E_\iota} A_e\Big)\Big] = \Psi_{E_\iota}$ and

$$\mathbb{E}[C_{\boldsymbol{E}}] = \sum_{\substack{i_1,\ldots,i_m \\ \text{distinct}}} \mathbb{E}\Big[\prod_{\iota=1}^{m} \gamma_{i_\iota}^{n_\iota}\Big] \prod_{\iota=1}^{m} \Psi_{E_\iota} = \boldsymbol{w}_{\boldsymbol{E}} Q_{\boldsymbol{E}} \ ,$$

for each partition $\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{S}_n)$. □

Remark 1. As was mentioned in (3.4), every $C_{\boldsymbol{E}}$ is constituted from one partition $\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{S}_n)$. The weights $\boldsymbol{w}_{\boldsymbol{E}}; \boldsymbol{E} \in \mathfrak{P}(\boldsymbol{n}) \subset \mathfrak{P}(\boldsymbol{S}_n)$ in (3.2) are only related to the number of blocks m and their sizes $\boldsymbol{n}$. Therefore, these weights have the same value for all partitions $\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{n})$, denoted by $\boldsymbol{w}(\boldsymbol{n})$ as

$$\boldsymbol{w}(\boldsymbol{n}) \ = \sum_{\substack{i_1,\ldots,i_m \\ \text{distinct}}} \mathbb{E}\Big[\prod_{\iota=1}^{m} \gamma_{i_\iota}^{n_\iota}\Big], \tag{3.5}$$

while distributions $Q_{\boldsymbol{E}}; \boldsymbol{E} \in \mathfrak{P}(\boldsymbol{n})$ are different as they relate to each partition's elements within blocks (2.8).

Representation (3.5) is useful for explaining the exchangeability property, as will be seen later. For each $\boldsymbol{n} \in \boldsymbol{N}_n$, let $\mathfrak{Q}(\boldsymbol{n})$ denote the collection of distributions corresponding to partitions in $\mathfrak{P}(\boldsymbol{n})$, and $\mathfrak{Q}(\boldsymbol{S}_n) = \bigcup_{\boldsymbol{n}\in\boldsymbol{N}_n} \mathfrak{Q}(\boldsymbol{n})$. Using Remark 1, we can write the joint distribution in (3.1) as

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \sum_{\boldsymbol{n}\in\boldsymbol{N}_n} \boldsymbol{w}(\boldsymbol{n})\boldsymbol{Q}(\boldsymbol{n}), \tag{3.6}$$

where $\boldsymbol{Q}(\boldsymbol{n}) = \sum_{Q\in\mathfrak{Q}(\boldsymbol{n})} Q$. Theorem 3.1 shows that

$$\sum_{\boldsymbol{E}\in\mathfrak{P}(\boldsymbol{S}_n)} \boldsymbol{w}_{\boldsymbol{E}} = \sum_{\boldsymbol{n}\in\boldsymbol{N}_n} |\mathfrak{P}(\boldsymbol{n})|\boldsymbol{w}(\boldsymbol{n}) = 1, \tag{3.7}$$

and the joint distribution of a random sample from the GDP is a mixture distribution of countable components. There is a one-to-one correspondence between components and partitions. Accordingly, this mixture's number of components equals the Bell number $\mathcal{B}_n$. Now, we rewrite (3.5) without infinite summations using permutation theory to be computationally applicable for performance-enhancing. Initially, we propose the following lemma.

Lemma 3.1. *Let $m, n_1, \ldots, n_m$ be positive integers, then*

$$\Sigma_m \ := \sum_{i_1>\cdots>i_m} \prod_{\iota=1}^{m} M_{0,\eta_\iota}^{\zeta_\iota} \ = \ \prod_{\iota=1}^{m} \frac{1}{1 - M_{0,\eta_\iota}} \ , \tag{3.8}$$

where $\eta_\iota = \sum_{r=1}^{\iota} n_r$, $\zeta_\iota = i_\iota - i_{\iota+1} - 1$, and $i_{m+1} = 0$.

Proof. By induction, for $m = 1$, $\Sigma_1 = \sum_{i_1=1}^{\infty} M_{0,n_1}^{i_1-1} = 1/(1 - M_{0,n_1})$. Suppose (3.8) holds for $m = k$, we prove it for $m = k + 1$,

$$
\begin{aligned}
\Sigma_{k+1} &= \sum_{i_1>\cdots>i_{k+1}} \prod_{\iota=1}^{k+1} M_{0,\eta_\iota}^{\zeta_\iota} = \sum_{i_1>\cdots>i_k} \prod_{\iota=1}^{k-1} M_{0,\eta_\iota}^{\zeta_\iota} M_{0,\eta_k}^{i_k-2} \sum_{i_{k+1}=0}^{i_k-2} \rho^{i_{k+1}} \\
&= \sum_{i_1>\cdots>i_k} \prod_{\iota=1}^{k-1} M_{0,\eta_\iota}^{\zeta_\iota} M_{0,\eta_k}^{i_k-2} \frac{1-\rho^{i_k-1}}{1-\rho} = c \sum_{i_1>\cdots>i_k} \prod_{\iota=1}^{k-1} M_{0,\eta_\iota}^{\zeta_\iota} (M_{0,\eta_k}^{i_k-1} - M_{0,\eta_{k+1}}^{i_k-1}) \\
&= c(\Sigma_k - \Sigma_k^*) = \prod_{\iota=1}^{k+1} \frac{1}{1 - M_{0,\eta_\iota}} ,
\end{aligned}
$$

where $\rho = M_{0,\eta_{k+1}}/M_{0,\eta_k} \neq 1$, the third equality comes from the sum of first $i_k - 1$ terms of a geometric series, $c = 1/(M_{0,\eta_k} - M_{0,\eta_{k+1}})$, and Σ_k^* is (3.8) with $n_k := n_k + n_{k+1}$. □

The following theorem reveals that (3.5) can be derived without infinite summation.

Theorem 3.2. *For $\boldsymbol{n} \in \mathbf{N}_n$, let $\mathfrak{P}(\boldsymbol{n}) \subset \mathfrak{P}(\boldsymbol{S}_n)$ be the collection of partitions with m blocks and sizes $\boldsymbol{n}$, and let* $\mathbf{perm}(\boldsymbol{n})$ *be the set of all m−permutations of $\boldsymbol{n}$. Then*

$$
\boldsymbol{w}(\boldsymbol{n}) = \sum_{\mathbf{perm}(\boldsymbol{n})} \prod_{\iota=1}^{m} \rho_{\eta_\iota,\eta_{\iota-1}} , \tag{3.9}
$$

where $\eta_\iota = \sum_{j=0}^{\iota} n_{\sigma(j)}$, $n_{\sigma(0)} = 0$, and $(n_{\sigma(1)}, \ldots, n_{\sigma(m)})$ is a one permutation from $\mathbf{perm}(\boldsymbol{n})$ *and* $|\mathbf{perm}(\boldsymbol{n})| = m!$.

Proof. Since the sequences $\overline{V}_1, \overline{V}_2, \ldots$ are independent, for $i \neq j$, variables V_i and $\overline{V}_j$ are independent. The indices $i_1, \ldots, i_m$ are distinct, and each permutation $(\sigma(1), \ldots, \sigma(m))$ of the set $\{1, \ldots, m\}$ can make one order as $i_{\sigma(1)} > \cdots > i_{\sigma(m)}$, and all orders cover (3.5). Calculating the expectation based on one order $i_1 > \cdots > i_m$ is

$$
\begin{aligned}
\sum_{i_1>\cdots>i_m} \mathbb{E}\Big[\prod_{\iota=1}^{m} \gamma_{i_\iota}^{n_\iota}\Big] &= \sum_{i_1>\cdots>i_m} \mathbb{E}\Big[\prod_{\iota=1}^{m} V_{i_\iota}^{n_\iota} \overline{V}_{i_\iota}^{\eta_{\iota-1}} \prod_{\iota=1}^{m} \prod_{j=i_{\iota+1}+1}^{i_\iota-1} \overline{V}_j^{\eta_\iota}\Big] \\
&= \prod_{\iota=1}^{m} M_{n_\iota,\eta_{\iota-1}} \sum_{i_1>\cdots>i_m} \prod_{\iota=1}^{m} M_{0,\eta_\iota}^{\zeta_\iota} \\
&= \prod_{\iota=1}^{m} \rho_{\eta_\iota,\eta_{\iota-1}} ,
\end{aligned}
$$

where $\gamma_{i_\iota}^{n_\iota} = V_{i_\iota}^{n_\iota} \prod_{j=1}^{i_\iota-1} \overline{V}_j^{n_\iota}$, $n_0 = 0$, $\eta_\iota = \sum_{r=0}^{\iota} n_r$, $\zeta_\iota = i_\iota - i_{\iota+1} - 1$, $i_{m+1} = 0$, and the last equation comes from Lemma 3.1. Note that the result for each order $i_{\sigma(1)} > \cdots > i_{\sigma(m)}$ is related to one permutation $(n_{\sigma(1)}, \ldots, n_{\sigma(m)}) \in \mathbf{perm}(\boldsymbol{n})$. Thus, the weight $\boldsymbol{w}(\boldsymbol{n})$ can be derived by summation over all $m!$ permutations. □

When the number of blocks m is large, $|\mathbf{perm}(\boldsymbol{n})|$ will be massive, and the weight $w(\boldsymbol{n})$ is more challenging to compute. Therefore, it is necessary to reduce the number of operations by carrying out the calculations only on the distinct permutations. Following this, we multiply each result by frequencies to get the weight, as explained below.

Proposition 3.1. *Let d be the number of distinct elements in $\{n_1, \ldots, n_m\}$, and the frequency of each one be $r_i; i = 1, \ldots, d$. Then, $w(\boldsymbol{n})$ reduces to*

$$w(\boldsymbol{n}) = \prod_{i=1}^{d} r_i! \sum_{\mathbf{words}(n)} \prod_{\iota=1}^{m} \rho_{\eta_\iota, \eta_{\iota-1}}, \tag{3.10}$$

where $\mathbf{words}(\boldsymbol{n})$ is just the set of all distinct permutations on $\{n_1, \ldots, n_m\}$, and the cardinality reduces to $|\mathbf{words}(\boldsymbol{n})| = m!/\prod_{i=1}^{d} r_i!$.

Proof. Imagine that each n_i; $i = 1, \ldots, m$ is analogous to a letter and that identical n_i's share the same letter. Each permutation corresponds to m letters that form a word. Let $\mathbf{words}(\boldsymbol{n})$ be the set of words with or without meaning that can be formed from these letters. This set is analogous to the distinct permutation set. In permutations theory, the cardinality of $\mathbf{words}(\boldsymbol{n})$ is $m!/\prod_{i=1}^{d} r_i!$, and each comes from $\prod_{i=1}^{d} r_i!$ permutations. □

The following example shows the joint distribution of random samples from $\mathrm{GDP}(H, G_0)$ in two cases, $n = 3, 4$.

Examples 3.1. Let $Q_{12|3} = \Psi_{\{1,2\}}\Psi_{\{3\}}$ be the distribution when $\boldsymbol{E} = \{\{1, 2\}, \{3\}\}$, and $Q_{12|3}$ [3] be an abbreviation for three distributions $\left\{Q_{12|3}, Q_{13|2}, Q_{23|1}\right\}$, and the same for others. Applying Theorem 3.1 and Proposition 3.1 and using Example 2.1, only the selection of all partitions is required.

Case $n = 3$: We have $\boldsymbol{S}_3 = \{1, 2, 3\}$ and $\boldsymbol{N}_3 = \{\{3\}, \{2, 1\}, \{1, 1, 1\}\}$.

(i) For $\boldsymbol{n} = \{3\}$, we have $\mathfrak{Q}(\boldsymbol{n}) = Q_{S_3}$ and $\mathbf{words}(\boldsymbol{n}) = \{3\}$, then using (3.10); $w(\boldsymbol{n}) = \mathbb{P}(X_1 = X_2 = X_3) = \rho_{3,0}$.

(ii) For $\boldsymbol{n} = \{2, 1\}$, we have $\mathfrak{Q}(\boldsymbol{n}) = Q_{12|3}$ [3] and $\mathbf{words}(\boldsymbol{n}) = \{(2, 1), (1, 2)\}$, then $w(\boldsymbol{n}) = \mathbb{P}(X_e = X_k \neq X_l) = \rho_{2,0}\rho_{3,2} + \rho_{3,1}$; $e, k, l \in \boldsymbol{S}_3$.

(iii) For $\boldsymbol{n} = \{1, 1, 1\}$, we have $\mathfrak{Q}(\boldsymbol{n}) = Q_{1|2|3}$ and $\mathbf{words}(\boldsymbol{n}) = \{(1, 1, 1)\}$, then $w(\boldsymbol{n}) = \mathbb{P}(X_1 \neq X_2 \neq X_3) = 3!\rho_{2,1}\rho_{3,2}$.

Therefore, $\mathfrak{Q}(\boldsymbol{S}_3) = \bigcup_{\boldsymbol{n} \in \boldsymbol{N}_3} \mathfrak{Q}(\boldsymbol{n}) = \left\{Q_{S_3}, Q_{12|3}\,[3], Q_{1|2|3}\right\}$. Substitute the above results into (3.1) or (3.6) to get the joint distribution $\mathbb{P}(X_1 \in A_1, X_2 \in A_2, X_3 \in A_3)$. According to (3.7), the equality $\rho_{3,0} + 3(\rho_{2,0}\rho_{3,2} + \rho_{3,1}) + 3!\rho_{2,1}\rho_{3,2} = 1$ always holds.

Case $n = 4$: $\boldsymbol{S}_4 = \{1, 2, 3, 4\}$ and $\boldsymbol{N}_4 = \{\{4\}, \{3, 1\}, \{2, 2\}, \{2, 1, 1\}, \{1, 1, 1, 1\}\}$.

(i) For $\boldsymbol{n} = \{4\}$, we have $\mathfrak{Q}(\boldsymbol{n}) = Q_{S_4}$ and $\mathbf{words}(\boldsymbol{n}) = \{4\}$, then using (3.10); $w(\boldsymbol{n}) = \mathbb{P}(X_1 = X_2 = X_3 = X_4) = \rho_{4,0}$.

(ii) For $\boldsymbol{n} = \{3,1\}$, we have $\mathfrak{Q}(\boldsymbol{n}) = Q_{123|4}$ [4] and $\mathbf{words}(\boldsymbol{n}) = \{(3,1),(1,3)\}$, then $\boldsymbol{w}(\boldsymbol{n}) = \mathbb{P}(X_e = X_k = X_l \neq X_r) = \rho_{3,0}\rho_{4,3} + \rho_{4,1}$; $e,k,l,r \in \boldsymbol{S}_4$.

(iii) For $\boldsymbol{n} = \{2,2\}$, we have $\mathfrak{Q}(\boldsymbol{n}) = Q_{12|34}$ [3] and $\mathbf{words}(\boldsymbol{n}) = \{(2,2)\}$, then $\boldsymbol{w}(\boldsymbol{n}) = \mathbb{P}(X_e = X_k \neq X_l = X_r) = 2\rho_{2,0}\rho_{4,2}$.

(iv) For $\boldsymbol{n} = \{2,1,1\}$, $\mathfrak{Q}(\boldsymbol{n}) = Q_{12|3|4}$ [6] and $\mathbf{words}(\boldsymbol{n}) = \{(2,1,1),(1,2,1),(1,1,2)\}$; $\boldsymbol{w}(\boldsymbol{n}) = \mathbb{P}(X_e = X_k \neq X_l \neq X_r) = 2(\rho_{2,0}\rho_{3,2}\rho_{4,3} + \rho_{3,1}\rho_{4,3} + \rho_{2,1}\rho_{4,2})$.

(v) For $\boldsymbol{n} = \{1,1,1,1\}$, we have $\mathfrak{Q}(\boldsymbol{n}) = Q_{1|2|3|4}$ and $\mathbf{words}(\boldsymbol{n}) = \{(1,1,1,1)\}$, then $\boldsymbol{w}(\boldsymbol{n}) = \mathbb{P}(X_1 \neq X_2 \neq X_3 \neq X_4) = 4!\rho_{2,1}\rho_{3,2}\rho_{4,3}$.

Thus, $\mathfrak{Q}(\boldsymbol{S}_4) = \left\{Q_{\boldsymbol{S}_4}, Q_{123|4}\ [4], Q_{12|34}\ [3], Q_{12|3|4}\ [6], Q_{1|2|3|4}\right\}$. Substitute the results into (3.1) or (3.6) to get the joint distribution $\mathbb{P}(X_1 \in A_1, X_2 \in A_2, X_3 \in A_3, X_4 \in A_4)$. According to (3.7), the equality $\sum_{\boldsymbol{n}\in N_4} |\mathfrak{P}(\boldsymbol{n})| \boldsymbol{w}(\boldsymbol{n}) = 1$ is always true.

Note that, to calculate $\boldsymbol{w}(\boldsymbol{n})$, for example, when $\boldsymbol{n} = \{1,1,1,1\}$, instead of applying (3.9) over 4! permutations, we used (3.10) over one word and multiplied it by replications.

A particular case of the GDP, when the distribution H is Beta(a,b), is BTPP(a,b,G_0), which is an extension of the DP. Ishwaran and Zarepour (2000) obtained the probability of drawing a sample of identical data points as $\rho_{n,0} = a^{\lceil n \rceil}\xi_n^{-1}$, where $\xi_n = (a+b)^{\lceil n \rceil} - b^{\lceil n \rceil}$; $a^{\lceil n \rceil} = a(a+1)\ldots(a+n-1)$ and $a^{\lceil 0 \rceil} = 1$. We are applying Theorem 3.1 to get the BTPP joint distribution, where $M_{i,j} = \Gamma(a+b)\Gamma(a)^{-1}\Gamma(b)^{-1}\Gamma(a+b+i+j)^{-1}\Gamma(a+i)\Gamma(b+j) = a^{\lceil i \rceil}b^{\lceil j \rceil}/(a+b)^{\lceil i+j \rceil}$ and $\rho_{n,j} = a^{\lceil n-j \rceil}b^{\lceil j \rceil}\xi_n^{-1}$. According to Proposition 3.1, each $\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{n})$ is weighted as

$$\boldsymbol{w}(\boldsymbol{n}) \;=\; \frac{b^m}{b^{\lceil n \rceil}} \prod_{\iota=1}^{m} a^{\lceil n_\iota - 1 \rceil} \tau_m^{a,b}(\boldsymbol{n})\,, \tag{3.11}$$

where by letting $f(k) = (b+1)^{\lceil k-1 \rceil}\,\xi_k^{-1}$ and $\eta_\iota = \sum_{j=1}^{\iota} n_{\sigma(j)}$,

$$\tau_m^{a,b}(\boldsymbol{n}) = \prod_{\iota=1}^{m}(n_\iota + a - 1)\prod_{i=1}^{d} r_i! \sum_{\mathbf{words}(\boldsymbol{n})} \prod_{\iota=1}^{m} f(\eta_\iota)\,.$$

For the DP with $a = 1$, we have $a^{\lceil k \rceil} = k!$, $f(k) = k^{-1}$. Using Lemma 3.1 in Miller (2019), $\prod_{i=1}^{d} r_i! \sum_{\mathbf{words}(\boldsymbol{n})} \prod_{\iota=1}^{m} \eta_\iota^{-1} = \prod_{\iota=1}^{m} n_\iota^{-1}$. Thus, $\tau_m^{1,b}(\boldsymbol{n}) = 1$ and

$$\boldsymbol{w}(\boldsymbol{n}) \;=\; \frac{b^m}{b^{\lceil n \rceil}} \prod_{\iota=1}^{m} (n_\iota - 1)!\,. \tag{3.12}$$

The weight (3.12) coincides with that presented by Blackwell and MacQueen (1973) and the Blackwell-MacQueen joint distribution can then be obtained as a particular case of Theorem 3.1.

Examples 3.2. Using Example 3.1 for $n = 3$, we can obtain the BTPP joint distribution. According to (3.11), $\boldsymbol{w}(\{3\}) = a^{\lceil 3 \rceil}\xi_3^{-1}$, $\boldsymbol{w}(\{2,1\}) = ab(a+1)(ab+\xi_2)\xi_2^{-1}\xi_3^{-1}$, and $\boldsymbol{w}(\{1,1,1\}) = 3!a^2b^2(b+1)\xi_2^{-1}\xi_3^{-1}$, while (3.12) gives the DP weights $\boldsymbol{w}(\{3\}) = 2(b+1)^{-1}(b+2)^{-1}$, $\boldsymbol{w}(\{2,1\}) = b(b+1)^{-1}(b+2)^{-1}$, and $\boldsymbol{w}(\{1,1,1\}) = b^2(b+1)^{-1}(b+2)^{-1}$.

Remark 2 (Moments). Let K_n be the number of distinct values among a sample of size n from the GDP. According to the literature on the GDP, only the probability of distinct and identical observations has been studied (see (2.6) and (2.7)). In general, we can present the probability of appearing $K_n = m; 1 \leq m \leq n$ as

$$\mathbb{P}(K_n = m) = \sum_{\boldsymbol{n} \in \boldsymbol{N}_n; |\boldsymbol{n}| = m} |\mathfrak{P}(\boldsymbol{n})| \boldsymbol{w}(\boldsymbol{n}). \tag{3.13}$$

Therefore, $\mathbb{E}[K_n] = \sum_{m=1}^{n} m\mathbb{P}(K_n = m)$ gives the expectation of the number of distinct values. Also, for every n, the raw moment $\mathbb{E}[G^n(\cdot)]$ is $\mathbb{E}[G^n(\cdot)] = \sum_{m=1}^{n} \mathbb{P}(K_n = m)G_0^m(\cdot)$. Consequently, $\mathbb{E}[G(\cdot)] = \mathbb{P}(K_1 = 1)G_0(\cdot) = G_0(\cdot)$, and

$$Var(G(\cdot)) = \rho_{2,0} G_0(\cdot)(1 - G_0(\cdot)), \tag{3.14}$$

where $\mathbb{P}(K_2 = 1) = 1 - \mathbb{P}(K_2 = 2) = \rho_{2,0}$.

We can evaluate the quantity and expectations of distinct values by conducting a simulation experiment for GDP samples. The graphical representation in Figure 1 shows cases with significant variations in expectations across different distribution types of H. Figure 1 shows that the number of distinct values increases with sample sizes, which is influenced by the $\rho_{2,0}$ values in variance. A small $\rho_{2,0}$ implies that G closely aligns with its mean G_0, presenting many distinct values. A closer examination of histograms reveals the convergence of G towards the central measure G_0 as $\rho_{2,0}$ decreases.

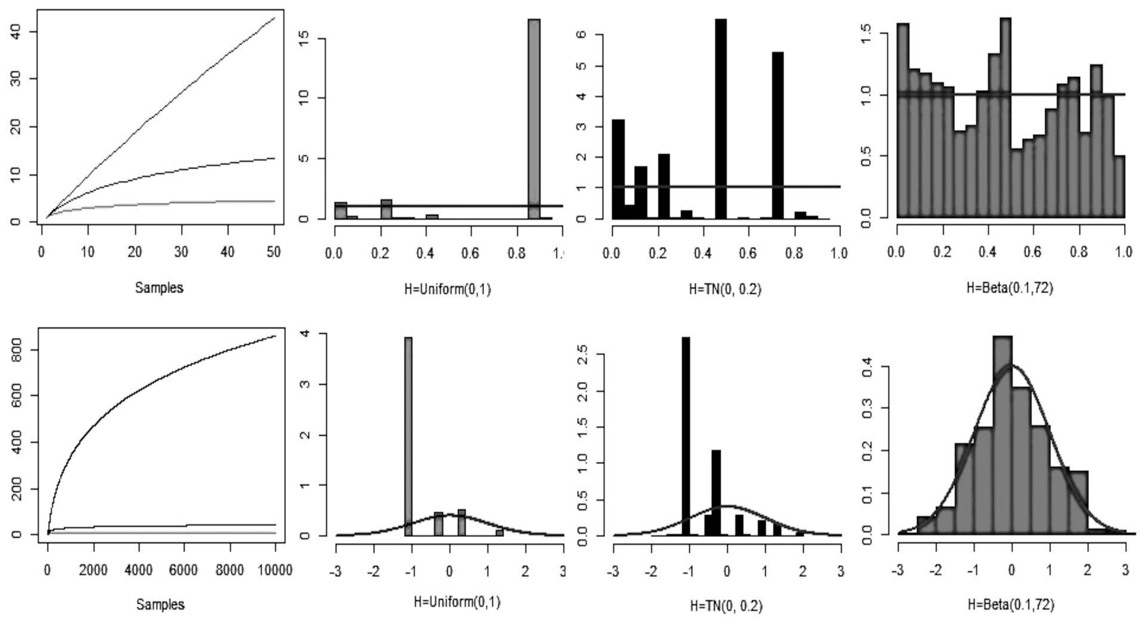

Figure 1: Histograms of samples from GDP with different types of H: the green histograms: H=Uniform(0,1) with $\rho_{2,0} = 0.5$; the black histograms: the truncated normal distribution over $(0, 1)$ with mean 0 and standard deviation 0.2, i.e., H=TN$(0, 0.2)$ with $\rho_{2,0} = 0.143$; the blue histograms: H=Beta(0.1,72) with $\rho_{2,0} = 0.008$. (Above) The center measures are the standard normal distributions; the red lines are their densities. (Below) The center measures are the standard uniform distributions; the red lines are their densities. (Left above) The expectation of the number of distinct values. (Left below) The number of distinct values in large samples from the GDP.

4 Generalized Pólya urn Mechanism for the GDP

Let $\boldsymbol{X}_{-n} = \{X_1, \ldots, X_{n-1}\}$ be a set of a random sample of size $n-1$ from GDP(H, G_0) with the nonatomic G_0. Since $\mathbb{P}(X_i = X_j) \neq 0$ for $i \neq j$, the GDP exhibits discreteness, i.e., by drawing a sample from the process, repetition of observations is expected, and only some distinct values will appear in $\boldsymbol{X}_{-n}$. Let $\boldsymbol{X}^*_m = \left\{X^*_1, \ldots, X^*_m\right\}$ be the set of distinct values among $\boldsymbol{X}_{-n}$, and $\#X^*_r = n_r$ be the number of repetitions of X^*_r; $\sum_{r=1}^{m} n_r = n-1$. Distinct values with repetitions are analogous to a partition $\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{n}) \subset \mathfrak{P}(\boldsymbol{S}_{n-1})$ with blocks

$$E_1 = \{e_1, \ldots, e_{n_1}\}, \ldots, E_m = \{e_{n-n_m}, \ldots, e_{n-1}\}\ , \tag{4.1}$$

where sizes $\boldsymbol{n} = \{n_1, \ldots, n_m\}$. Each block E_ι in (4.1) contains precisely the repetition indices for X^*_ι. For example, block E_1, contains indices of draws $X_{e_1}, \ldots, X_{e_{n_1}}$, which share a single distinct value, X^*_1.

The GDP can be characterized by a generalized PUS mechanism using the above knowledge and the concept of conditional probabilities. Suppose $\boldsymbol{X}_{-n}$ has been drawn from the GDP(H, G_0). A newly drawn observation X_n, is equal to one element of $\boldsymbol{X}_m^*$ or drawing from the center measure G_0 as a new distinct value denoted by X_{m+1}^*. Our approach to these two cases will be as follows:

(i) For $X_n \in \boldsymbol{X}_m^*$, such as $X_n = X_r^*$; $r \leq m$, we have a sample $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ and a set of distinct values $\boldsymbol{X}_m^*$ with iterations $\boldsymbol{n}^r = \{n_1, \ldots, n_r + 1, \ldots, n_m\}$. The new state is precisely analogous to one partition of $\mathfrak{P}(\boldsymbol{n}^r) \subset \mathfrak{P}(\boldsymbol{S}_n)$, with probability

$$\boldsymbol{w}(\boldsymbol{n}^r) = \sum_{\substack{i_1,\ldots,i_m \\ \text{distinct}}} \mathbb{E}\Big[\gamma_{i_r} \prod_{\iota=1}^{m} \gamma_{i_\iota}^{n_\iota}\Big]. \tag{4.2}$$

(ii) For $X_n \notin \boldsymbol{X}_m^*$, it takes on a new distinct value drawn from G_0 denoted by X_{m+1}^*. Thus, we have a sample $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ and a set of distinct values $\boldsymbol{X}_{m+1}^* = \boldsymbol{X}_m^* \cup \{X_{m+1}^*\} = \{X_1^*, \ldots, X_m^*, X_{m+1}^*\}$ with iterations $\boldsymbol{n}^+ = \{n_1, \ldots, n_m, 1\}$. The new state is precisely analogous to one partition of $\mathfrak{P}(\boldsymbol{n}^+) \subset \mathfrak{P}(\boldsymbol{S}_n)$, with probability

$$\boldsymbol{w}(\boldsymbol{n}^+) = \sum_{\substack{i_1,\ldots,i_{m+1} \\ \text{distinct}}} \mathbb{E}\Big[\gamma_{i_{m+1}} \prod_{\iota=1}^{m} \gamma_{i_\iota}^{n_\iota}\Big]. \tag{4.3}$$

The weights (4.2) and (4.3) can be calculated immediately by (3.10). Since $\sum_{j=1}^{\infty} \gamma_j = 1$ a.s., then $1 - \sum_{r=1}^{m} \gamma_{i_r} = \sum_{j \notin I} \gamma_j$ a.s., where $I = \{i_1, \ldots, i_m\}$ is a set of distinct indices. Thus,

$$\boldsymbol{w}(\boldsymbol{n}) = \mathbb{E} \sum_{\substack{i_1,\ldots,i_m \\ \text{distinct}}} \prod_{\iota=1}^{m} \gamma_{i_\iota}^{n_\iota} \Big(\sum_{r=1}^{m} \gamma_{i_r} + 1 - \sum_{r=1}^{m} \gamma_{i_r}\Big) = \sum_{r=1}^{m} \boldsymbol{w}(\boldsymbol{n}^r) + \boldsymbol{w}(\boldsymbol{n}^+). \tag{4.4}$$

Based on Pitman's terminology, $\boldsymbol{w}(\boldsymbol{n})$ is an exchangeable partition probability function (EPPF), a symmetric function of $\boldsymbol{n}$. Now, as with the generalized PUS representations for the DP and the Pitman-Yor process (Pitman , 1995, 1996), we form the generalized PUS for the GDP as follows,

$$X_1 \sim G_0;$$

$$X_n \mid X_1, \ldots, X_{n-1} \begin{cases} = X_r^* & \text{with probability } \frac{\boldsymbol{w}(\boldsymbol{n}^r)}{\boldsymbol{w}(\boldsymbol{n})} \\ \sim G_0 & \text{with probability } \frac{\boldsymbol{w}(\boldsymbol{n}^+)}{\boldsymbol{w}(\boldsymbol{n})} \end{cases}. \tag{4.5}$$

In particular, for BTPP(a, b, G_0) we derive the PUS by (4.5) where

$$\frac{\boldsymbol{w}(\boldsymbol{n}^r)}{\boldsymbol{w}(\boldsymbol{n})} = \frac{n_r + a - 1}{b + n - 1} \frac{\tau_m^{a,b}(\boldsymbol{n}^r)}{\tau_m^{a,b}(\boldsymbol{n})}, \quad \frac{\boldsymbol{w}(\boldsymbol{n}^+)}{\boldsymbol{w}(\boldsymbol{n})} = \frac{b}{b + n - 1} \frac{\tau_{m+1}^{a,b}(\boldsymbol{n}^+)}{\tau_m^{a,b}(\boldsymbol{n})}.$$

Accordingly, the Blackwell-MacQueen urn scheme (2.1) is accurately reflected for $a = 1$, whereas we have $\tau_m^{1,b}(\boldsymbol{n}) = \tau_m^{1,b}(\boldsymbol{n}^r) = \tau_{m+1}^{1,b}(\boldsymbol{n}^+) = 1$ using Lemma 3.1 in Miller (2019). By (4.5), the GDP predictive distribution is emanated as

$$\mathbb{P}(X_n \in \cdot \mid X_1, \ldots, X_{n-1}) = \frac{\boldsymbol{w}(\boldsymbol{n}^+)\, G_0(\cdot) \;+\; \sum_{r=1}^{m} \boldsymbol{w}(\boldsymbol{n}^r)\, \delta_{X_r^*}(\cdot)}{\boldsymbol{w}(\boldsymbol{n})}. \tag{4.6}$$

Moreover, we can obtain the joint marginal distribution (3.1) by multiplying n successive conditional distributions (4.6) as

$$\mathbb{P}(X_1 \in A_1, ..., X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i \mid X_1, \dots, X_{i-1}). \tag{4.7}$$

Let $\boldsymbol{X}_{-i} = \{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n\}$ be $\boldsymbol{X}$ excluding X_i, with $|\boldsymbol{X}_{-i}| = n-1$ for $i = 1, \dots, n$, and denote its set of distinct values by $\boldsymbol{X}^*_m = \{X^*_1, \dots, X^*_m\}$ with iterations $\boldsymbol{n} = \{n_1, \dots, n_m\}$, where $\sum_{\iota=1}^{m} n_\iota = n-1$. The state of $\boldsymbol{X}_{-i}$ (i.e., its distinct values and iterations) is analogous to one partition of $\mathfrak{P}(\boldsymbol{n})$. The above discussion is relevant since the weight only relates to the number of blocks m and their sizes $\boldsymbol{n}$ (see Remark 1). Assume that $\boldsymbol{X}_{-i}$ is taken into account. The observation X_i will either be equal to one element $X^*_r \in \boldsymbol{X}^*_m; r \le m$ with the state probability (4.2) or drawn from G_0 and treated as a new distinct value denoted by X^*_{m+1} with the state probability (4.3). Thus, the conditional distribution of X_i given $\boldsymbol{X}_{-i}$ is represented by

$$\mathbb{P}(X_i \in \cdot \mid \boldsymbol{X}_{-i}) = \mathbb{P}\left(X_n \in \cdot \mid X_1, \dots, X_{n-1}\right). \tag{4.8}$$

According to Theorem 3.1, the joint distribution of a random sample from the GDP is a countable mixture of distributions $Q_{\boldsymbol{E}}$'s over each $\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{S}_n)$ with weights $\boldsymbol{w}_{\boldsymbol{E}}$'s, and (3.7) has to be satisfied. For $\boldsymbol{n} \in \boldsymbol{N}_n$, let $\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{n})$ be one partition of $\boldsymbol{S}_n$ defined in (2.8) with m blocks. Component $Q_{\boldsymbol{E}}$ can be rewritten as $\prod_{\iota=1}^{m} G_0\left(X^*_\iota\right) \prod_{j \in E_\iota} \delta_{X^*_\iota}(X_j)$ based on the above discussion. In addition, the drawing mechanism from $Q_{\boldsymbol{E}}$ can be represented as

$$\{\{X^*_\iota \sim G_0, X_j = X^*_\iota; j \in E_\iota\}; \iota = 1, \dots, m\}.$$

Therefore, observations are classified into m clusters

$$\{\{X_{e_1}, \dots, X_{e_{n_1}}\}, \dots, \{X_{e_{n-n_m+1}}, \dots, X_{e_n}\}\}, \tag{4.9}$$

which share $\boldsymbol{X}^*_m$ with iterations $\boldsymbol{n}$. All $Q_{\boldsymbol{E}}$; $\boldsymbol{E} \in \mathfrak{P}(\boldsymbol{n})$ make the same number of distinct values, iterations, and the weight $\boldsymbol{w}(\boldsymbol{n}) = \mathbb{P}\left(\#X^*_1 = n_1, ..., \#X^*_m = n_m\right)$, but observations are classified into different indices. The exchangeability property for the GDP is reinforced by (4.7) and (4.8), where the sequence of observations from GDP(H, G_0) is infinitely exchangeable. That is, for every n, the joint distribution of the original order is the same as that of $(X_{\sigma(1)}, \dots, X_{\sigma(n)})$ for any permutation $(\sigma(1), \dots, \sigma(n))$ of $\boldsymbol{S}_n$, as

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_{\sigma(1)} \in A_1, \dots, X_{\sigma(n)} \in A_n). \tag{4.10}$$

Alternatively, the de Finetti representation of infinitely exchangeable sequences can be derived directly from (2.3) as

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \int \prod_{i=1}^{n} G(A_i)\, \mathcal{G}(dG),$$

in which the probability distribution $\mathcal{G} := \mathrm{GDP}(H, G_0)$ exists, serves as a prior measure for G, and is often known as the de Finetti measure such that

$$\begin{aligned} X_i \mid G &\overset{\text{iid}}{\sim} G, \quad i = 1, \ldots, n \\ G &\sim \mathrm{GDP}(H, G_0), \end{aligned} \tag{4.11}$$

which makes (4.11) a nonparametric Bayesian model for $X_1, \ldots, X_n$.

5 Pólya urn Gibbs Sampler

Let $\boldsymbol{Y} = \{Y_1, \ldots, Y_n\}$ be a conditionally independent data set distributed by $f(Y_i \mid X_i)$, which is parametrized by $\boldsymbol{X}$ from the model (4.11), and let y_i and x_i be the observed values of Y_i and X_i, respectively. The model then specifies hierarchically as

$$\begin{aligned} Y_i \mid X_i = x_i &\overset{ind}{\sim} f(y_i \mid x_i), \; i = 1, \ldots, n \\ X_i \mid G &\overset{iid}{\sim} G \\ G \mid H, G_0 &\sim \mathrm{GDP}(H, G_0), \end{aligned} \tag{5.1}$$

representing a GDP mixture model (MacEachern and Müller , 1998). Based on the generalized PUS (4.5), the following theorem offers the conditional posterior distribution of X_i given $\boldsymbol{X}_{-i}$ and $\boldsymbol{Y}$ and displays the Gibbs sampling scheme.

Theorem 5.1. *For $i \in \boldsymbol{S}_n$, the conditional posterior distribution of $X_i \mid \boldsymbol{X}_{-i}, \boldsymbol{Y}$ is*

$$X_i \mid \boldsymbol{X}_{-i}, \boldsymbol{Y} \sim q_0^* G_i(\cdot) + \sum_{r=1}^{m} q_r^* \delta_{X_r^*}(\cdot)\,. \tag{5.2}$$

Here, $G_i(\cdot) = f(Y_i \mid X_i)G_0(dX_i) / \int f(Y_i \mid X_i)G_0(dX_i)$, $q_0^* = c\boldsymbol{w}(\boldsymbol{n}^+) \int f(Y_i \mid X_i)G_0(dX_i)$, and $q_r^* = c\boldsymbol{w}(\boldsymbol{n}^r) f(Y_i \mid X_r^*)$, where c is subject to the constraint $\sum_{r=0}^{m} q_r^* = 1$.

Proof. We have $f(\boldsymbol{Y} \mid \boldsymbol{X}) = \prod_{j=1}^{n} f(Y_j \mid X_j)$, and $f(\boldsymbol{X}) = f(\boldsymbol{X}_{-i}) f(X_i \mid \boldsymbol{X}_{-i})$, where $f(X_i \mid \boldsymbol{X}_{-i})$ is based on the generalized PUS in (4.8). Therefore,

$$\begin{aligned} f(X_i \mid \boldsymbol{X}_{-i}, \boldsymbol{Y}) &= \frac{f(\boldsymbol{X}) f(\boldsymbol{Y} \mid \boldsymbol{X})}{f(\boldsymbol{X}_{-i}, \boldsymbol{Y})} = \frac{f(Y_i \mid X_i) f(X_i \mid \boldsymbol{X}_{-i})}{\int f(Y_i \mid X_i) d f(X_i \mid \boldsymbol{X}_{-i})} \\ &= \frac{\boldsymbol{w}(\boldsymbol{n}^+) f(Y_i \mid X_i) G_0(dX_i) + \sum_{r=1}^{m} \boldsymbol{w}(\boldsymbol{n}^r) f(Y_i \mid X_r^*) \delta_{X_r^*}(dX_i)}{\boldsymbol{w}(\boldsymbol{n}^+) \int f(Y_i \mid X_i) G_0(dX_i) + \sum_{r=1}^{m} \boldsymbol{w}(\boldsymbol{n}^r) f(Y_i \mid X_r^*)}\,. \end{aligned}$$

The proof is complete, where the constant c^{-1} is the denominator of the last equation. □

Remark 3. Let

$$Y_i \mid X_i \sim f; \quad X_i \sim G_0\,, \tag{5.3}$$

and

$$Y_i \mid X_i \sim f; \quad X_i \sim \delta_{X_r^*}, \tag{5.4}$$

be two Bayesian models. The posterior distribution (5.2) is a mixture of distributions:

(i) The baseline measure $G_i(\cdot)$ is the posterior distribution of X_i given the observation Y_i if the prior of X_i is the center measure G_0, i.e., $G_i(\cdot)$ is the posterior distribution of $X_i \mid Y_i$ in (5.3). The weight q_0^* is proportional to the marginal distribution of Y_i in (5.3) multiplied by the resulting state probability of $\boldsymbol{X}$ in (4.3).

(ii) The point mass measure $\delta_{X_r^*}(\cdot)$ is the posterior distribution of X_i given Y_i if the prior of X_i is the point mass on X_r^*, i.e., $\delta_{X_r^*}(\cdot)$ is the posterior distribution of $X_i \mid Y_i$ in (5.4). The weight q_r^* is proportional to the marginal distribution of Y_i in (5.4) multiplied by the resulting state probability of $\boldsymbol{X}$ in (4.2).

Similar to Ishwaran and James (2001), we first insert the GDP into the Bayesian approach through the following basic Gibbs sampling algorithm:

Algorithm.

Step 1. Start by choosing the initial values of $\boldsymbol{X}$. Usually, we sample $X_i^{(0)}, i = 1, \ldots, n$ individually from the posterior distribution $G_i(\cdot)$ shown above.

Step 2. Sample $\boldsymbol{X}$ by drawing sequentially from the conditional posterior distribution of $(X_i \mid \boldsymbol{X}_{-i}, \boldsymbol{Y})$ in (5.2) for $i = 1$, then $i = 2$, and so on up to $i = n$. At each stage of the drawing, the $\boldsymbol{X}_{-i}$ contains the most recent values of elements.

Step 3. Return to step 2 until convergence.

The algorithm is a straightforward posterior sampler with the convergence discussion in Escobar (1994) and Escobar and West (1995). We slightly generalize the Gibbs sampler algorithm to conform to the semiparametric hierarchical model

$$\begin{aligned} Y_i \mid X_i = x_i, \varphi &\overset{ind}{\sim} f(y_i \mid x_i, \varphi),\ i = 1, \ldots, n \\ X_i \mid G &\overset{iid}{\sim} G \\ \varphi &\sim \pi \\ G \mid H, G_0 &\sim \mathrm{GDP}(H, G_0). \end{aligned} \tag{5.5}$$

Here the model depends on an additional finite-dimensional parameter φ distributed by $\pi(\varphi)$. The conditional posterior distributions of X_i and φ are given by

$$X_i \mid \boldsymbol{X}_{-i}, \varphi, \boldsymbol{Y} \sim q_0^* G_i(\cdot) + \sum_{r=1}^{m} q_r^* \delta_{X_r^*}(\cdot), \tag{5.6}$$

and

$$\varphi \mid \boldsymbol{X}, \boldsymbol{Y} \sim \varsigma\pi(\cdot)\prod_{i=1}^{n} f(Y_i \mid X_i, \varphi)\,, \tag{5.7}$$

where $G_i(\cdot) = f(Y_i \mid X_i, \varphi)G_0(\cdot)/\int f(Y_i \mid X_i, \varphi)G_0(dX_i)$, $q_0^* = c\boldsymbol{w}(\boldsymbol{n}^+)\int f(Y_i \mid X_i, \varphi)G_0(dX_i)$, $q_r^* = c\boldsymbol{w}(\boldsymbol{n}^r)f(Y_i \mid X_r^*, \varphi)$, $\sum_{r=0}^m q_r^* = 1$, and $\varsigma^{-1} = \int \prod_{i=1}^n f(Y_i \mid X_i, \varphi)\pi(\varphi)d\varphi$. The Gibbs sampler for the semiparametric model (5.5) may be modified by Step 2 to

Step 2′. Sample $\boldsymbol{X}$ by drawing from $(X_i \mid \boldsymbol{X}_{-i}, \varphi, \boldsymbol{Y})$ in (5.6) for $i = 1$, then $i = 2$, and so on up to $i = n$. At each stage, the $\boldsymbol{X}_{-i}$ contains the most recent values of elements.

Step 2″. Sample φ by drawing from the conditional distribution of $(\varphi \mid \boldsymbol{X}, \boldsymbol{Y})$ in (5.7).

The computation of weights $q_0^*, q_1^*, \ldots, q_m^*$ and drawing samples from $G_i(\cdot)$ are direct. The Gibbs sampler can be implemented easier with the conjugacy of (5.3). It requires only computing the weights and determining the posterior distribution $G_i(\cdot)$, as illustrated in the following GDP mixture models.

(i) **Binomial/Beta GDP Mixture Model:**
Let $Y_i \mid X_i \sim \mathrm{Bin}\,(L_i, X_i)$ $(i = 1, \ldots, n)$ and $G_0 = \mathrm{Beta}(a, b)$ in the nonparametric Bayesian model (5.1), where the parameters L_i and X_i denote the number of Bernoulli trials and the probability of success for the i-th binomial observation, respectively. Using Remark 3, for the Gibbs sampler implementation, we obtain $G_i(\cdot) = \mathrm{Beta}(a + Y_i, b + L_i - Y_i)$, $q_0^* = c\boldsymbol{w}(\boldsymbol{n}^+)B^{-1}(a, b)B(a + Y_i, b + L_i - Y_i)$, and $q_r^* = c\boldsymbol{w}(\boldsymbol{n}^r)X_r^{*Y_i}(1 - X_r^*)^{L_i - Y_i}$, with $c^{-1} = \sum_{j=0}^m c^{-1}q_j^*$ and $B(a, b) = \int_0^1 t^{a-1}(1 - t)^{b-1}dt$.

(ii) **Poisson/Gamma GDP mixture models:**
Let $Y_i \mid X_i \sim \mathrm{Poi}\,(X_i)$ with the mean parameter X_i $(i = 1, \ldots, n)$, and $G_0 = \mathrm{Gamma}(a, b)$. For the Gibbs sampler, we drive the baseline posterior $G_i(\cdot) = \mathrm{Gamma}(Y_i{+}a, 1{+}b)$, $q_0^* = c\boldsymbol{w}(\boldsymbol{n}^+)\Gamma(Y_i{+}a)\Gamma^{-1}(a)b^a(b{+}1)^{-Y_i - a}$ and $q_r^* = c\boldsymbol{w}(\boldsymbol{n}^r)X_r^{*Y_i}e^{-X_r^*}$, with $c^{-1} = \sum_{j=0}^m c^{-1}q_j^*$.

(iii) **Normal/Normal GDP mixture models:**
Let $Y_i \mid X_i, \tau \sim \mathrm{Nor}\,(X_i, \tau)$ $(i = 1, \ldots, n)$, $G_0 = \mathrm{Nor}\,(\mu_0, \tau_0)$, and $\tau \sim \mathrm{Gamma}\,(a, b)$ in (5.5), where the parameters X_i and τ denote the mean and the inverse-variance of the normal distribution, respectively, and μ_0, τ_0, a, b are known. To implement the Gibbs sampler, we obtain $G_i(\cdot) = \mathrm{N}(\mu', \tau')$, $q_0^* = c\boldsymbol{w}(\boldsymbol{n}^+)\tau_1^{1/2}exp(-0.5\tau_1(Y_i - \mu_0)^2)$, and $q_r^* = c\boldsymbol{w}(\boldsymbol{n}^r)\tau^{1/2}exp(-0.5\tau(Y_i - X_r^*)^2)$, with $c^{-1} = \sum_{j=0}^m c^{-1}q_j^*$, $\mu' = \tau'(\tau Y_i + \tau_0\mu_0)$, $\tau' = (\tau + \tau_0)^{-1}$, $\tau_1^{-1} = (\tau^{-1} + \tau_0^{-1})$. Also, the posterior distribution of τ is given by $\tau \mid \boldsymbol{X}, \boldsymbol{Y} \sim \mathrm{Gamma}(a + n/2, b + \sum(Y_i - X_i)/2)$.

6 An Empirical Study for Binomial Data

As mentioned, the Binomial/Beta model assumes that each experiment is a Binomial draw with unknown proportions and fixed sample sizes for several groups. An illustrative example to fit this model is a study conducted by Beckett and Diaconis (1994). The data set is binary strings generated by rolling thumbtacks. The flicks are presumed independent conditionally on the tack. There are 320 thumbtacks flicked 9 times each. The data set is available in Table 1 of Liu (1996), showing the observations Y_i for $i = 1, \ldots, 320$. The output variable Y_i is the number of times each tack landed point up. A Binomial/Beta model was fitted to analyze the data set by employing a Dirichlet process prior, with G_0 to be a standard uniform center distribution. Several precision parameter values were tested, revealing that the approximate posterior mean displayed bimodality for some precision parameter options, although with less pronounced bimodality for others. One can find the posterior densities of the Binomial/Beta GDP mixture model, where $Y_i \mid X_i \sim \text{Bin}(9, X_i)$ and $G_0 = \text{Beta}(1, 1)$. In which $G_i(\cdot)$, q_0^*, and q_r^* are reduced to $\text{Beta}(1 + Y_i, 10 - Y_i)$, $c\boldsymbol{w}(\boldsymbol{n}^+)B(1 + Y_i, 10 - Y_i)$, and $c\boldsymbol{w}(\boldsymbol{n}^r)X_r^{*Y_i}(1 - X_r^*)^{9-Y_i}$, respectively.

We performed a sensitivity analysis by incorporating various H forms with supports on $(0, 1)$. Our study noticed a significant difference in examining the estimated marginal posteriors for each X_i, where $i = 1, \ldots, n$. As illustrated in Figure 2, the observed bimodality is more pronounced for higher values of $\rho_{2,0} = 0.5$ (i.e., H=uniform$(0, 1)$ shown in Figure 1), suggesting a limited number of distinct values. The bimodality reduces as $\rho_{2,0}$ decreases, and the number of distinct values increases. This trend continues to improve as $\rho_{2,0} = 0.143$ (i.e., H=TN$(0, 0.2)$ shown in Figure 1), as the updated posterior distributions show agreement and have moved from bimodal to unimodal.

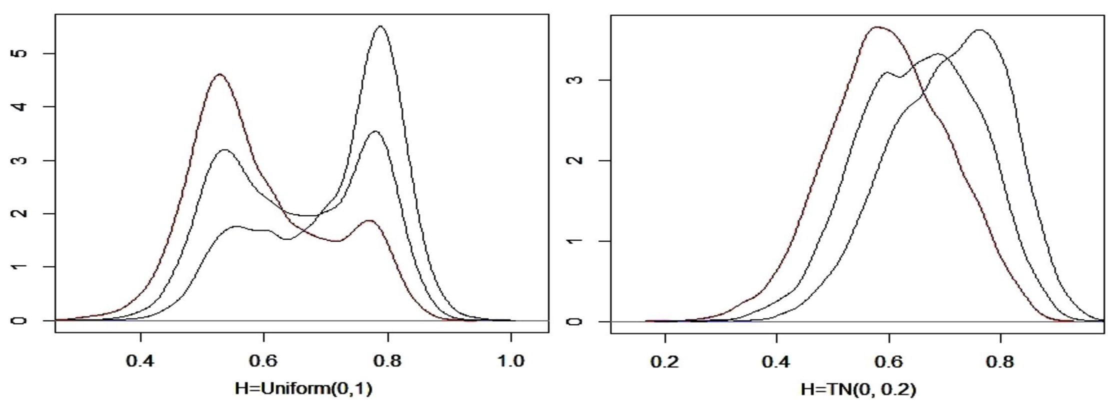

Figure 2: Posterior densities of X_i of the Binomial/Beta GDP mixture model, where $G_0 = \text{Beta}(1, 1)$; The red lines: X_{50}; the black lines: X_{100}; the blue lines: X_{200}. (Left): H=uniform$(0, 1)$; (Right): H=TN$(0, 0.2)$.

7 Concluding Remarks

The partition analysis offers an adaptable strategy for constructing the joint distribution of a random sample from the GDP process. Representing the Pólya urn scheme for the GDP makes the process more beneficial for nonparametric Bayesian purposes.

The distribution can be formed as a mixture distribution of countable components, and it is thus straightforward to implement in most applications. The construction delivers the probability of appearing any number of distinct values among a sample from the process. The expectation of this number is accessible. As a particular case, we find the distribution for the beta two-parameter process and discuss it for the DP, which gives the Blackwell-MacQueen urn scheme. To highlight the theoretical parts of modeling topics, examples of the Gibbs sampler implementation were presented for Binomial/Beta, Poisson/Gamma, and Normal/Normal GDP mixture models. The paper represented the Hjort path through the Sethuraman random probability measure for the GDP. Future studies are demanding to give directions in increasing distributions' flexibility, in particular, extensions on the stick-breaking construction since the sequence $V_1, V_2, \ldots$ can be selected from any prior having support over (0,1), not restricted only to the Beta distribution. Further features of our proposed distribution need more work which is the aim of future study.

References

Aghabazaz, Z., Kazemi, I. and Nematollahi, A. (2023). Dynamic mixed models with heterogeneous covariance components using multivariate GARCH innovations and the Dirichlet process mixture. *Journal of Computational and Applied Mathematics*, Forthcoming.

Aldous, D. J. (1985). Exchangeability and related topics. *In: Hennequin, P.L. (eds) École d'Été de Probabilités de Saint-Flour XIII—1983*, Lecture Notes in Mathematics, vol 1117. Springer, Berlin, Heidelberg.

Andrews, G. E. (1998), *The Theory of Partitions (Encyclopedia of Mathematics and its Applications)*. Cambridge: Cambridge University Press.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, **2**(**6**), 1152-1174.

Barcella, W. (2017). *Covariate Dependent Random Measures With Applications in Biostatistics*. Ph.D. Thesis, UCL (University College London).

Barcella, W., De Iorio, M., Favaro, S. and Rosner, G.L. (2018). Dependent generalized Dirichlet process priors for the analysis of acute lymphoblastic leukemia. *Biostatistics*, **19**(3), 342-358.

Barcella, W., Iorio, M. D. and Malone-Lee, J. (2018). Modelling correlated binary variables: an application to lower urinary tract symptoms. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **67**(4), 1083-1100.

Beckett, L., and Diaconis, P. (1994). Spectral analysis for discrete longitudinal data. *Advances in Mathematics*, **103**(1), 107–128.

Blackwell, D., and MacQueen, J.B. (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics*, **1(2)**, 353-355.

Castellares, F., Ferrari, S. L. and Lemonte, A.J. (2018). On the Bell distribution and its associated regression model for count data. *Applied Mathematical Modelling*, **56**, 172-185.

Escobar, M.D. (1994). Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association*, **89(425)**, 268-277.

Escobar, M.D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, **90(430)**, 577-588.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 209-230.

Ferguson, T.S. (1983). Bayesian density estimation by mixtures of normal distributions. *In Rizvi, M.H., Rustagi, J.S., and Siegmund, D. (Eds), Recent Advances in Statistics*, Academic Press, pp 287-302.

Ghosal, S. and Van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge: Cambridge University Press.

Hjort, N.L. (2000). Bayesian analysis for a generalised Dirichlet process prior. *Preprint Series. Statistical Research Report http://urn.nb.no/URN: NBN: no-23420*.

Hjort, N.L., Holmes, C., Müller, P. and Walker, S. G. (2010). *Bayesian Nonparametrics*. Cambridge: Cambridge University Press.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, **96(453)**, 161-173.

Ishwaran, H. and Zarepour, M. (2000). Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*, **87(2)**, 371-390.

Liu, J. S. (1996). Nonparametric hierarchical Bayes via sequential imputations. *The Annals of Statistics*, **24(3)**, 911–930.

Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, **12(1)**, 351-357.

MacEachern, S. N. (1999). Dependent nonparametric processes. *In ASA Proceedings of the Section on Bayesian Statistical Science, Alexandria, VA*. American Statistical Association.

MacEachern, S.N. and Müller, P. (1998). Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics*, **7**(2), 223-238.

Mano, S. (2018), *Partitions, Hypergeometric Systems, and Dirichlet Processes in Statistics*. Springer Verlag, Japan.

Miller, J.W. (2019). An elementary derivation of the Chinese restaurant process from Sethuraman's stick-breaking process. *Statistics and Probability Letters*, **146**, 112-117.

Molinari, M., de Iorio, M., Chaturvedi, N., Hughes, A. and Tillin, T. (2021). Modelling ethnic differences in the distribution of insulin resistance via Bayesian nonparametric processes: an application to the SABRE cohort study. *The International Journal of Biostatistics*, **17**(1), 153-164.

Müller, P., Quintana, F. A., Jara, A. and Hanson, T. (2015). *Bayesian Nonparametric Data Analysis*, Springer International Publishing Switzerland.

Neal, R. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational & Graphical statistics*, **9**(2), 249-265.

Phadia, E.G. (2016). *Prior Processes and Their Applications*. Springer International Publishing.

Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields*, **102**(2), 145-158.

Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. *Lecture Notes-Monograph Series*, **30**, 245-267.

Pitman, J. (2006). *Combinatorial Stochastic Processes*. Springer Berlin, Heidelberg.

Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, **25**(**2**), 855-900.

Rodriguez, A. and Dunson, D. B. (2014). Functional clustering in nested designs: modeling variability in reproductive epidemiology studies. *The Annals of Applied Statistics*, **8**(**3**), 1416–1442.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, **8**, 639-650.

Sethuraman, J. and Tiwari, R. C. (1982). Convergence of Dirichlet measures and the interpretation of their parameter. *In Gupta, S.S. and Berger, J.O. (eds). Statistical Decision Theory and Related Topics III*, Academic Press, pp 305-315.