Posterior Sampling from some Non-Exchangeable Priors

Shuhei Mano[1]
The Institute of Statistical Mathematics

November 2, 2018

1 Introduction

Definition 1 (Ferguson 1973)**.** Let ρ be a probability measure on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. A random probability measure F is called a Dirichlet process with base measure ρ, if F satisfies

$$(F(A_1),...,F(A_k)) \sim \mathrm{Dir}(\rho(A_1),...,\rho(A_k))$$

for every finite measurable partition $\{A_1,...,A_k\}$ of $\mathbb{R}$.

For a probability measure μ and $\theta>0$, a Dirichlet process with $\rho=\theta\mu$ will be denoted by $\mathrm{DP}(\theta;\mu)$. For simplicity, we assume μ is diffuse.

Theorem 1 (Ferguson 1973)**.** A Dirichlet process is constructed as follows.

1. Let $\{Y_t;t\geq 0\}$, $Y_0=0$ be the gamma process with $Y_t\sim\mathrm{Ga}(\theta t,1)$. The jump sizes $(J_1,J_2,...)$ with $\sum_{i=1}^\infty J_i=Y_1$.
2. For $X_i\overset{ind}{\sim}\mu(\cdot)$,
$$F(\cdot)=\sum_{i=1}^\infty \frac{J_i}{Y_1}\delta_{X_i}(\cdot)\sim \mathrm{DP}(\theta;\mu).$$

The prediction rule is well known. Let $F\sim\mathrm{DP}(\theta;\mu)$. By conjugacy of the Dirichlet distribution in multinomial sampling,

$$\begin{aligned}(F_n(A_1),...,F_n(A_k)) &:= (F(A_1),...,F(A_k))|(X_1,...,X_n)\\ &\sim \mathrm{Dir}\left(\theta\mu(A_1)+\sum_{i=1}^n\delta_{X_i}(A_1),...,\theta\mu(A_k)+\sum_{i=1}^n\delta_{X_i}(A_k)\right)\end{aligned}$$

and

$$F_n\sim\mathrm{DP}\left(\theta+n,\frac{\theta\mu+\sum_{i=1}^n\delta_{X_i}}{\theta+n}\right).$$

$\mathbb{P}(X_1\in\cdot)=\mathbb{E}\{\mathbb{P}(X_1\in\cdot|F)\}=\mathbb{E}(F(\cdot))=\mu(\cdot)$,

$$\mathbb{P}(X_{n+1}\in\cdot|X_1,...,X_n)=\frac{\theta}{\theta+n}\mu(\cdot)+\frac{n}{\theta+n}\Lambda_n(X_1,...,X_n)(\cdot).$$

Here, $\Lambda_n(X_1,...,X_n):=n^{-1}\sum_{i=1}^n\delta_{X_i}$ is the empirical distribution. In Bayesian context, F is called a prior process, and the posterior distribution is called the prediction rule. The sequential

[1]E-mail: smano@ism.ac.jp

sampling scheme is well known as the Blackwell-MacQueen urn scheme (1973), or the Chinese restaurant process.

The prediction rule induces measures on partitions. Let the j-th firstly appear value of $(X_1, ..., X_n)$ be X_j^*, $j \in \{1, 2, ..., k\}$. Then, $(n_1, ..., n_k)$, $n_j := \#\{i; X_i = X_j^*\}$ is a integer partition of a positive integer n. The prediction rule gives

$$\mathbb{P}(N_1 = n_1, ..., N_k = n_k) = \frac{\theta^k}{(\theta)_n} \prod_{j=1}^{k} (n_j - 1)!.$$

This is symmetric under permutations of $(n_1, ..., n_k)$, and the distribution of the multiplicities of integers $(c_1, ..., c_n)$, $c_i := \#\{j; n_j = i\}$, is

$$\mathbb{P}(C_1 = c_1, ..., C_n = c_n) = \frac{n!}{(\theta)_n} \prod_{i=1}^{n} \left(\frac{\theta}{i}\right)^{c_i} \frac{1}{c_i!},$$

where $(\theta)_n := \theta(\theta + 1) \cdots (\theta + n - 1)$. This measure on partitions is called the Ewens sampling formula (1972; Antoniak 1974). Sibuya (1993) considered the prediction rule as a random clustering process, and $(c_1, ..., c_n)$ are called size indices.

A partition λ of n ($\lambda \vdash n$) is

$$\lambda = (\lambda_1, \lambda_2, ..., \lambda_k)$$

for some k with $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_k > 0$ with $\lambda_1 + \lambda_2 + \cdots + \lambda_k = n$. Here, $l(\lambda) = k$ is called the length of partition. Let

$$\mathcal{P}_n := \{\lambda; \lambda \vdash n\}, \qquad \mathcal{P}_{n,k} := \{\lambda; \lambda \vdash n, l(\lambda) = k\}.$$

For $\lambda \in \mathcal{P}_{n,k}$, the multiplicities $c_i(\lambda) := \#\{j; \lambda_j = i\}$ satisfies

$$1 \cdot c_1 + 2 \cdot c_2 + \cdots + n \cdot c_n = n, \qquad c_1 + c_2 + \cdots + c_n = k.$$

and determines the shape of a Young diagram uniquely. Figure 1 gives an example.

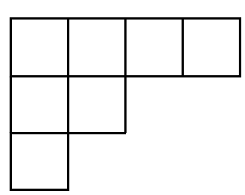

Figure 1: Young diagram of $\lambda = (4, 2, 1)$.

For $\theta = 1$, Ewens sampling formula is

$$\mathbb{P}(C = c) = \frac{1}{z_\lambda}, \qquad z_\lambda := \prod_{i=1}^{n} i^{c_i} c_i!, \qquad \lambda \in \mathcal{P}_n.$$

young diagram	cycle decomposition	probability ($1/z_\lambda$)
$c_3 = 1$	$(123), (132)$	$1/3$
$c_2 = 1, c_1 = 1$	$(12)(3), (23)(1), (31)(2)$	$1/2$
$c_1 = 3$	$(1)(2)(3)$	$1/6$

Table 1: S_3

This is the distribution of cycle lengths in cycle decomposition of random permutations. It is the uniform distribution with respect to cardinality of conjugacy class of the symmetric group. Table 1 gives an example.

For symmetric polynomials $\Lambda_k = \mathbb{Z}[x_1, ..., x_k]^{S_k}$, $\Lambda_k = \oplus_{n \geq 0} \Lambda_k^n$, where Λ_k^n consists of the homogeneous symmetric polynomials of degree n, together with the zero polynomial. A monomial symmetric function is

$$m_\lambda(x_1, ..., x_k) := \sum_\sigma \prod_{i=1}^k x_i^{\sigma_i}, \qquad \sigma \in \{\rho; \pi(\rho) = \lambda\}.$$

For example,

$$m_{(2,1)} = x_1^2 x_2 + x_1 x_2^2 \ \in \Lambda_2^3.$$

For each $r \geq 1$ the r-th power sum is

$$p_r := m_{(r)} = \sum_{i=1}^k x_i^r.$$

The power sum symmetric function is defined as

$$p_\lambda := p_{\lambda_1} \cdots p_{\lambda_{l(\lambda)}}$$

For example,

$$p_{(2,1)} = p_2 p_1 = (x_1^2 + x_2^2)(x_1 + x_2).$$

The Schur symmetric function is defined as

$$s_\lambda(x) := \frac{\det(x_i^{\lambda_j + k - j})_{1 \leq i,j \leq k}}{\det(x_i^{k-j})_{1 \leq i,j \leq k}}.$$

It satisfies Cauchy's identity

$$\prod_{i,j} (1 - x_i y_j)^{-1} = \sum_\lambda s_\lambda(x) s_\lambda(y),$$

the sum is over all partitions. Let us introduce the orthonormality $\langle s_\lambda, s_\mu \rangle = \delta_{\lambda,\mu}$. In terms of the power sum symmetric functions,

$$\prod_{i,j} (1 - x_i y_j)^{-1} = \sum_\lambda z_\lambda^{-1} p_\lambda(x) p_\lambda(y),$$

and it follows that

$$\langle p_\lambda, p_\mu \rangle = \delta_{\lambda\mu} z_\lambda.$$

The Jack symmetric function is a generalization of the Schur symmetric function. The Jack symmetric functions are derived by the orthogonality

$$\langle p_\lambda, p_\mu \rangle_\alpha = \delta_{\lambda,\mu} z_\lambda \alpha^{l(\lambda)},$$

coming from the identity

$$\prod_{i,j} (1 - x_i y_j)^{-1/\alpha} = \sum_\lambda (z_\lambda \alpha^{l(\lambda)})^{-1} p_\lambda(x) p_\lambda(y).$$

With normalization, $(z_\lambda \alpha^{l(\lambda)})^{-1}$, $\theta := 1/\alpha$ is the Ewens sampling formula. The Macdonald symmetric function is based on the identity

$$\prod_{i,j} \frac{(t x_i y_j; q)_\infty}{(x_i y_j; q)_\infty} = \sum_\lambda (z_\lambda(q,t))^{-1} p_\lambda(x) p_\lambda(y),$$

where

$$z_\lambda(q,t) := z_\lambda \prod_{i \geq 1} \left(\frac{1 - q^i}{1 - t^i} \right)^{c_i}, \qquad (x; y)_n := \prod_{i=0}^{n-1} (1 - x y^i),$$

which reduces to the Jack function in the limit $t = q^\theta$, $q \to 1$.

Diaconis and Lam (2012) discussed mixing of MCMC for a random walk on Young diagrams with the probability measure given by $(z_\lambda(q,t))^{-1}$. We will call the random partition Macdonald partition.

Algorithm 1 (Diaconis & Lam 2012)**.** Random walk for MCMC for the probability measure $(z_\lambda(q,t))^{-1}$ on Young diagrams.

1. Set $t = 0$ and pick an initial sample $\lambda^{(0)}$.
2. Pick parts σ with probability
$$\mathbb{P}(\Sigma = \sigma) = \frac{1}{q^n - 1} \prod_{i=1}^{n} \begin{pmatrix} c_i(\lambda^{(t)}) \\ c_i(\lambda^{(t)} \backslash \sigma) \end{pmatrix} (q^i - 1)^{c_i(\sigma)}.$$
3. Pick parts $\sigma' \vdash |\sigma|$ with probability
$$\mathbb{P}(\Sigma' = \sigma') = \frac{t}{t-1} \prod_{i=1}^{n} \left\{ \frac{1}{i} \left(1 - \frac{1}{t^i} \right) \right\}^{c_i(\sigma')} \frac{1}{c_i(\sigma')!}.$$
4. Set $\lambda^{(t+1)} = (\lambda^{(t)} \backslash \sigma) \cup \sigma'$, increment t to $t+1$, and go to Step 2.

This talk we will see

- MCMC seems to be inevitable because some random partitions including the Macdonald partition does not admit sequential sampling scheme.
- Nevertheless, we will see that a direct sampling from the random partition is possible.
- As a statistical application, we will discuss posterior sampling in a mixture model setting.

2 Gibbs partitions

Definition 2 (Pitman 2006; M 2018)**.** Gibbs partition is the probability measure on partitions $\lambda \vdash n \in \mathbb{N} := \{1, 2, ...\}$ of the form

$$\mathbb{P}(C = c) = \frac{v_{n,l(c)}}{B_n(v, w)} n! \prod_{i=1}^{n} \left(\frac{w_i}{i!}\right)^{c_i} \frac{1}{c_i!}.$$

Here, $l(c) = c_1 + \cdots + c_n$ is the length, and the normalization constant is written as

$$B_n(v, w) = \sum_{k=1}^{n} v_{n,k} B_{n,k}(w),$$

where

$$B_{n,k}(w) = \sum_{c \in \mathcal{P}_{n,k}} n! \prod_{i=1}^{n} \left(\frac{w_i}{i!}\right)^{c_i} \frac{1}{c_i!}$$

is known as the partial Bell polynomial.

Gibbs partitions are commonly used to characterize prior processes.

Remark 1 (Exponential structure)**.** If $(v_{n,k}) = 1$, $B_n(w) := B_n(1, w)$ is the Bell polynomial. The Gibbs partition reduces to the exponential structure (exponential generating function of $B_n(w)$ is $e^{W(x)}$, where $W(x)$ is exponential generating function of (w_i)). The exponential structure is a class of multiplicative measures defined by Vershik (1996) for study of limit shapes of random Young diagrams.

Example 1 (Pitman's partition, 1995)**.** Pitman's partition is the case with parameters

$$v_{n,k} = (\theta)(\theta + \alpha) \cdots (\theta + (k-1)\alpha), \quad w_i = (1 - \alpha)_{i-1}, \quad \alpha < 1,$$

and we have $B_n(v, w) = (\theta)_n$. This is a sample from the two-parameter Poisson-Dirichlet process (Pitman & Yor 1997), which is obtained from the α-stable subordinator. If $\alpha = 0$, this random partition reduces to the Ewens sampling formula.

Example 2 (Macdonald's partition)**.** Macdonald's partition is the case with parameters

$$v_{n,k} = 1, \qquad w_i = \frac{t^i - 1}{q^i - 1}(i - 1)!.$$

From a q-analogue of the negative binomial theorem, we have

$$B_n(w) = \frac{(t;q)_n}{(q;q)_n} n!, \qquad (x;y)_n := \prod_{i=0}^{n-1} (1 - xy^i).$$

Taking $q \to 1$ with $t = q^\theta$, this random partition reduces to the Ewens sampling formula.

Both of them are variations of sampling from the Dirichlet process, but the former is exchangeable, while the latter is not exchangeable.

Definition 3 (Kingman 1978)**.** A random partition Π_n of a finite set $[n]$ is called exchangeable, if for each partition $\{A_1, ..., A_k\}$ of $[n]$

$$\mathbb{P}(\Pi_n = \{A_1, ..., A_k\}) = p_n(|A_1|, ..., |A_k|)$$

for some symmetric function p_n. Moreover, if p_n is consistent, namely,

$$p_n(n_1, ..., n_k) = p_{n+1}(n_1, ..., n_k, 1) + \sum_{i=1}^{k} p_{n+1}(n_1, ..., n_i + 1, ..., n_k),$$

for all $n \in \mathbb{N}$, the sequence of random partitions is called (infinite) exchangeable.

Theorem 2 (Gnedin & Pitman 2005)**.** Gibbs partition is consistent iff $w_i = (1 - \alpha)_{i-1}$, $\alpha < 1$.

In general, a prediction rule is described as

$$\mathbb{P}(X_{n+1} \in \cdot | X_1, ..., X_n) = \left(1 - \sum_{i=1}^{k} P_i\right) \mu(\cdot) + \sum_{j=1}^{k} P_j \delta_{X_j{}^*}(\cdot)$$

for some random sequence $(P_1, P_2, ...)$.

Theorem 3 (Pitman 1995; Lee et al. 2013)**.** A random partition is given by a prediction rule is equivalent to partially exchangeability of the random partition. Moreover, a partial exchangeable partition is exchangeable iff p_n is symmetric.

Corollary 1. A Gibbs partition is symmetric. If it is not consistent, a prediction rule is not available.

Next is another example of non-exchangeable Gibbs partition. An advantage in applications is that the normalization constants have closed forms.

Example 3 (Hoshino's partition)**.** The limiting quasi-multinomial distribution by Hoshino (2005) is a random partition obtained via tilted random forests of labeled rooted trees. This is the case with parameters

$$v_{n,k} = \theta^k, \qquad w_i = i^{i-1}.$$

The partial bell polynomial has a colosed form

$$B_{n,k}(w) = \begin{pmatrix} n-1 \\ k-1 \end{pmatrix} n^{n-k}$$

and we have

$$B_n(v, w) = \theta(\theta + n)^{n-1}.$$

3 Direct sequential sampler via A-hypergeometric systems

Definition 4 (Gel'fand, Kapranov, Zelevinski 1990)**.** For an non-negative integer valued $d \times m$ matrix A of rank d and a vector $b \in \mathbb{C}^m$, the system of linear PDEs with annihilators

$$\sum_{j=1}^{m} a_{ij}\theta_j - b_j, \qquad i \in \{1, ..., d\}, \qquad \theta_j := x_j \partial_j,$$
$$\partial^{c^+} - \partial^{c^-}, \qquad c \in \mathsf{ker} A \cap \mathbb{Z}^m,$$

is called the A(GKZ)-hypergeometric system $H_A(b)$. Here, $c_i^+ := c_i \vee 0$, $c_i^- := (-c_i) \vee 0$. $H_A(b)$ is a left ideal of the Weyl algebra and called A-hypergeometric ideal. The series solution around the origin

$$Z_A(b; x) := \sum_{\{c; Ac=b, c \in \mathbb{N}_0{}^m\}} \frac{x^c}{c!}, \qquad x^c := \prod_{i=1}^{m} x_i^{c_i}, \qquad c! := \prod_{i=1}^{m} c_i!$$

is called the A-hypergeometric series. Here, $Z_A(b; x) = 0$ if $b \notin A\mathbb{N}_0{}^m$, $\mathbb{N}_0 := 0 \cup \mathbb{N}$.

Definition 5 (Takayama, Kuriki, Takemura 2018)**.** Consider m cells and let $t_i \in [m]$ be the cell of the $i \in [n]$-th observation of a sample of size k. For the count vector

$$(c_1, ..., c_m), \qquad c_j := \#\{i; t_i = j\},$$

the probability distribution with mass function of the form

$$\mathbb{P}(C_1 = c_1, ..., C_m = c_m) = \frac{1}{Z_A(b; x)} \frac{x^c}{c!}$$

is called the A-hypergeometric distribution. The support is $\{c; Ac = b, c \in \mathbb{N}_0{}^m\}$ and $Z_A(b; x)$ is the A-hypergeometric polynomial.

Remark 2. An A-hypergeometric distribution is the conditional distribution of multinomial sampling from log-affine models given $Ac = b$.

Homogeneity of the polynomial (row-space of A contains $(1, ..., 1)$) demands an annihilator

$$\sum_{i=1}^{m} \theta_i - n.$$

Using the contiguity relation of the A-hypergeometric polynomial

$$\theta_i Z_A(b;x) = x_i Z_A(b-a_i;x), \qquad i \in [m],$$

where a_i is the i-th column vector, we have

$$\sum_{i=1}^{m} x_i Z_A(b-a_i;x) = n Z_A(b;x),$$

or $\sum_{i=1}^{m} e_A(b;i)/n = 1$, where

$$e_A(b;i) := \mathbb{E}(C_i|AC=b) = \frac{Z_A(b-a_i;x)}{Z_A(b;x)} x_i.$$

Here, $e_A(b;i)/n$ can be regarded as the transition probability from $Z_A(b;x)$ to $Z_A(b-a_i;x)$ in a Markov chain with reducing degree of polynomial by one. This observation gives a direct sequential sampling algorithm for A-hypergeometric distributions.

Algorithm 2 (M 2017)**.** Direct sequential sampling from A-hypergeometric distributions.

1. Pick $t_1 = j$ with probability $e_A(b;j)/n$.

2. For $i = 2, ..., n$, pick $t_i = j$ with probability

$$\frac{e_A(b-(a_{t_1}+\cdots+a_{t_{i-1}});j)}{n-i+1}.$$

Remark 3. To compute the expectations $e_A(b;j)$, we can use a Pfaffian system for a holonomic ideal I

$$\theta_i \bullet Q = P_i Q, \qquad i \in [\mathrm{rank}(I)],$$

where Q is the standard monomial. It is generally possible to obtain the Pfaffian system by the theory of Gröbner bases, but more efficient ways are devised for actual purpose.

A homogeneous 2-row matrix A is generally given as

$$A = \begin{pmatrix} 0 & i_1 & i_2 & \cdots & i_{m-1} \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}.$$

Here, $0 < i_1 < i_2 < \cdots < i_{m-1}$ are relatively prime integers. If $i_{m-1} = m-1$, $b_1, b_2 \in \mathbb{N}$, the support is $\mathcal{P}_{b_1+b_2,b_2}$:

$$1 \cdot c_1 + 2 \cdot c_2 + \cdots + m \cdot c_m = b_1 + b_2, \; c_1 + c_2 + \cdots + c_m = b_2.$$

Here, $b_1 + b_2$ and b_2 are the weight and length of a partition.

Remark 4. The toric ideal of the polynomial ring with A determines an algebraic curve called monomial curve. In particular, if $i_{m-1} = m-1$ the curve is the rational normal curve. The system was considered in the context (Cattani et al. 1999; Saito et al. 2010).

For a Gibbs partition

$$\mathbb{P}(C=c)=\frac{v_{n,l(c)}}{B_n(v,w)}n!\prod_{i=1}^{n}\left(\frac{w_i}{i!}\right)^{c_i}\frac{1}{c_i!},\qquad c\in\mathcal{P}_n,$$

the length $l(c)=k$ is the sufficient statistic for parameters $(v_{n,k})$, and the conditional distribution is an A-hypergeometric distribution

$$\mathbb{P}(C=c|AC=b)=\frac{1}{Z_A(b;x)}\frac{x^c}{c!},\qquad c\in\mathcal{P}_{n,k},$$

where

$$A=\begin{pmatrix}0 & 1 & 2 & \cdots & n-k\\ 1 & 1 & 1 & \cdots & 1\end{pmatrix},\quad b=\begin{pmatrix}n-k\\ k\end{pmatrix},\quad x_i=\frac{w_i}{i!},$$

and $n!Z_A(b;x)=B_{n,k}(w)$.

Now we have a direct sampler for Gibbs partitions.

Algorithm 3. Direct sampling from Gibbs partitions.

1. Pick a length with $\mathbb{P}(l(\lambda)=k)=v_{n,k}B_{n,k}(w)/B_n(w)$.

2. Pick the rows $(n_1,...,n_k)$ by the direct sequential sampler.

Example 4 (A path of $\mathcal{P}_{7,3}$).

$$\begin{array}{ccccccc}
(4,\underline{2},1) & \to & (\underline{4},1) & \to & (\underline{1}) & \to & 0\\
x_4x_2x_1 & & x_4x_1 & & x_1 & & \\
(3,\underline{2},2) & & (3,2) & & & & \\
+x_3x_2^2/2! & & +\,x_3x_2 & & & & \\
(5,1,1) & & & & & & \\
+x_5x_1^2/2! & & & & & &
\end{array}$$

$$\mathbb{P}(\text{path})=\frac{x_1x_4+x_2x_3}{x_1x_2x_4+x_1^2x_5/2!+x_2^2x_3/2!}\frac{x_2}{3}\frac{x_1}{x_1x_4+x_2x_3}\frac{x_4}{2}=\frac{\mathbb{P}((4,2,1))}{3!}.$$

Remark 5. Hoshino's partition and Pitman's partition with $\alpha=-1,1/2$ are the cases that the sampler works without resorting to use of the Pfaffian system, because we know closed forms of the partial Bell polynomials.

4 Posterior sampling

A typical Bayesian semiparametric setting is data $(Y_1,...,Y_n)$ derived from a hierarchical model (MacEachern 1994; Escober & West 1995)

$$Y_i|X_i,\sigma\overset{ind}{\sim}\mathrm{N}(Y_i;X_i,\sigma),\qquad \sigma\sim\pi(\sigma),$$
$$X_i|P\overset{ind}{\sim}P,\qquad P\sim\mathrm{DP}(\theta;\mu).$$

The marginalized model for the Dirichlet process mixture

$$(X_1, ..., X_n) \sim \mathbb{P}(X_1, ..., X_n)$$

is sampled directly by the Blackwell-MacQueen urn scheme. To draw from posterior $\pi(Y, \sigma|X)$, a Gibbs sampler is used, where we iteratively draw values from conditional distributions

$$X_i|(X_{-i}, \sigma, Y), \qquad i \in [n], \qquad \sigma|(X, Y).$$

Thanks to the exchangeability, the prediction rule gives

$$\mathbb{P}(X_i = \cdot|X_{-i}, \sigma, Y) \propto \mathrm{N}(Y_i; X_i = \cdot, \sigma)\theta\mu(\cdot) + \sum_{j=1}^{k} \mathrm{N}(Y_i; X_j{}^*, \sigma)n_j\delta_{X_j{}^*}(\cdot).$$

Algorithm 4 (Prior sampling)**.** Sampling from a prior associated with a Gibbs partition.

1. Pick a Gibbs partition $(n_1, ..., n_k)$ by the direct sampler.
2. Pick $(X_1^*, ..., X_k^*)$ as i.i.d. from the base measure $\mu(\cdot)$.
3. Pick $(X_1, ..., X_n)$ such that $n_j = \#\{i; X_i = X_j^*\}$.

Remark 6. A generalization of explicit allocation prior given by Green & Richardson (2001) for the Dirichlet multinomial model.

Posterior sampling will be discussed.

5 Summary

- A direct sampling from Gibbs partitions including the Macdonald partition was introduced.
- The sampler is based on the direct sequential sampler from the A-hypergeometric distribution associated with monomial curves.
- As a statistical application, posterior sampling in a mixture model setting was discussed.
- Beyond exchangeability has been attract interests of Bayesians, but probably outside of partial exchangeability has not been discussed.

References can be found in Mano S (2018) Partitions, hypergeometric systems and Dirichlet processes in statistics, Springer Briefs in Statistics, JSS Research Series in Statistics, Springer.