Partition models

Peter McCullagh
University of Chicago

November 4, 2015

1 Set partitions

For integer $n \geq 1$, a partition B of the finite set $[n] = \{1, \ldots, n\}$ is

(i) a collection $B = \{b_1, \ldots\}$ of disjoint non-empty subsets, called blocks, whose union is $[n]$;

(ii) an equivalence relation on $[n]$, i.e. a symmetric Boolean function $B\colon [n] \times [n] \to \{0, 1\}$ that is also reflexive and transitive;

(iii) a block factor or symmetric binary matrix of order n such that $B_{ij} = 1$ if i, j belong to the same block.

These equivalent representations are not distinguished in the notation, so B is a set of subsets, a Boolean function, a subset of $[n] \times [n]$, or a symmetric binary matrix, as the context demands. In practice, a partition is frequently written in an abbreviated form, such as $B = 2|13$ for a partition of [3] or $u_2|u_1, u_3$ for a partition of three objects $\{u_1, u_2, u_3\}$.. In this notation, the partitions of [2] are 12 and 1|2, and the five partitions of [3] are

$$123, \quad 12|3, \quad 13|2, \quad 23|1, \quad 1|2|3.$$

The blocks are unordered and unlabelled, so there is no concept of a first block or a last block, and 2|13 is the same partition as 13|2 and 2|31.

A partition B is a sub-partition of B^* if each block of B is a subset of some block of B^* or, equivalently, if $B_{ij} = 1$ implies $B^*_{ij} = 1$. This relationship is a partial order denoted by $B \leq B^*$, which can be interpreted as $B \subset B^*$ if each partition is regarded as a subset of $[n]^2$. The partition lattice $\mathcal{E}_n$ is the set of partitions of $[n]$ with this partial order. To each pair of partitions B, B' there corresponds a greatest lower bound $B \wedge B'$, which is the set intersection or Hadamard component-wise matrix product. The least upper bound $B \vee B'$ is the least element of $\mathcal{E}_n$ that is greater than or equal to both, the transitive completion of $B \cup B'$. The least element $\mathbf{0}_n \in \mathcal{E}_n$ is the partition with n singleton blocks, and the greatest element is the single-block partition denoted by $\mathbf{1}_n$. As matrices, $\mathbf{0}_n$ is the identity, whereas $\mathbf{1}_n = [n]^2$ is the matrix whose components are all one.

A permutation $\sigma\colon [n] \to [n]$ induces an action $B \mapsto B^\sigma$ by composition such that the transformed partition is $B^\sigma(i, j) = B(\sigma(i), \sigma(j))$ in the form of an equivalence relation. In matrix notation, $B^\sigma = \sigma B \sigma^{-1}$, so the action by conjugation maintains symmetry by permuting both the rows and columns of B in the same way. The block sizes are preserved and are maximally invariant under conjugation. In this way, the 15 partitions of [4] may be grouped into five orbits or equivalence classes as follows:

$$1234, \quad 123|4\,[4], \quad 12|34\,[3], \quad 12|3|4\,[6], \quad 1|2|3|4.$$

Thus, for example, 12|34 is the representative element for one orbit, which also includes 13|24 and 14|23.

The symbol # applied to a set denotes the number of its elements, so $\#B$ is the number of blocks, and $\#b$ is the size of block $b \in B$. As a matrix, B is positive semi-definite of rank $\#B$. A partition distribution is defined on the finite set $\mathcal{E}_n$, and the first few values of $\#\mathcal{E}_n$ are 1, 2, 5, 15, 52, called Bell numbers. More generally, $\#\mathcal{E}_n$ is the nth moment of the unit Poisson distribution whose exponential generating function is

$$\exp(e^t - 1) = 1 + \sum_{n=1}^{\infty} t^n \,\#\mathcal{E}_n/n!.$$

In the discussion and manipulation of explicit probability models on $\mathcal{E}_n$, it is helpful to use the ascending and descending factorial symbols

$$\begin{aligned} \alpha^{\uparrow r} &= \alpha(\alpha+1)\cdots(\alpha+r-1) = \Gamma(r+\alpha)/\Gamma(\alpha) \\ k^{\downarrow r} &= k(k-1)\cdots(k-r+1) \end{aligned}$$

for integer $r \ge 0$. Note that $k^{\downarrow r} = 0$ for positive integers $r > k$. By convention $\alpha^{\uparrow 0} = 1$. It is not a coincidence that $\alpha^{\uparrow r}$ is the ordinary generating function for the Stirling numbers of the first kind $S_{n,r}$, the number of permutations $[n] \to [n]$ having exactly r cycles.

2 Dirichlet partition model

The term *partition model* refers to a probability distribution, or family of probability distributions, on the set $\mathcal{E}_n$ of partitions of $[n]$. In some cases, the probability is concentrated on the the subset $\mathcal{E}_n^k \subset \mathcal{E}_n$ of partitions having k or fewer blocks. A distribution on $\mathcal{E}_n$ such that $p_n(B) = p_n(\sigma B \sigma^{-1})$ for every permutation $\sigma\colon [n] \to [n]$ is said to be finitely exchangeable. Equivalently, p_n is exchangeable if $p_n(B)$ depends only on the block sizes of B.

Historically, the most important examples are Dirichlet-multinomial partitions generated for fixed k in three steps as follows.

(i) First generate the random probability vector $\pi = (\pi_1, \ldots, \pi_k)$ from the Dirichlet distribution with parameter $(\theta_1, \ldots, \theta_k)$.

(ii) Given π, the sequence $Y_1, \ldots, Y_n, \ldots$ is independent and identically distributed, each component taking values in $\{1, \ldots, k\}$ with probability π. Each sequence $(y_1, \ldots, y_n)$ in which the value r occurs $n_r \geq 0$ times has probability

$$p_n(y) = E(\pi_1^{n_1} \cdots \pi_k^{n_k}) = \frac{\Gamma(\theta_\bullet) \prod_{j=1}^k \theta_j^{\uparrow n_j}}{\Gamma(n + \theta_\bullet)},$$

where $\theta_\bullet = \sum \theta_j$.

(iii) Now forget the labels $1, \ldots, k$ and consider only the partition $B(Y)$ generated by the sequence Y, i.e. $B_{ij}(Y) = 1$ if $Y_i = Y_j$. Since Y is an exchangeable sequence, the partition distribution is also exchangeable, but an explicit simple formula is available only for the uniform case $\theta_j = \lambda/k$, which is now assumed. The number of sequences generating the same partition $B \in \mathcal{E}_n$ is $k^{\downarrow \#B}$, and these have equal probability in the uniform case. Consequently, the induced partition has probability

$$p_{nk}(B, \lambda) = k^{\downarrow \#B} \frac{\Gamma(\lambda) \prod_{b \in B} (\lambda/k)^{\uparrow \#b}}{\Gamma(n + \lambda)}, \tag{1}$$

called the uniform Dirichlet-multinomial partition distribution. The factor $k^{\downarrow \#B}$ ensures that partitions having more than k blocks have zero probability.

In the limit as $k \to \infty$, the uniform Dirichlet-multinomial partition becomes

$$p_n(B, \lambda) = \frac{\lambda^{\#B} \prod_{b \in B} \Gamma(\#b)}{\lambda^{\uparrow n}}. \tag{2}$$

This is the celebrated Ewens distribution, or Ewens sampling formula, which arises in population genetics as the partition generated by allele type in a population evolving according to the Fisher-Wright model by random mutation with no selective advantage of allele types (Ewens, 1972). The preceding derivation, a version of which can be found in chapter 3 of Kingman (1980), goes back to Watterson (1974). The Ewens partition is the same as the partition generated by a sequence drawn according to the Blackwell-McQueen urn scheme (Blackwell and McQueen, 1973).

Although the derivation makes sense only if k is a positive integer, the distribution (1) is well defined for negative values $-\lambda < k < 0$. For a discussion of this and the connection with GEM distributions and Poisson-Dirichlet distributions, see Pitman (2006, section 3.2).

3 Partition processes and partition structures

Deletion of element n from the set $[n]$, or deletion of the last row and column from the matrix representation $B \in \mathcal{E}_n$, determines a map $D_n \colon \mathcal{E}_n \to \mathcal{E}_{n-1}$, a projection from the larger to the smaller lattice. Equivalently, $D_n B \equiv B[n-1]$ is the restriction of B to the subset $[n-1]$. These deletion maps preserve partial order and make the sets $\{\mathcal{E}_1, \mathcal{E}_2, \ldots\}$ into a projective system

$$\cdots \mathcal{E}_{n+1} \xrightarrow{D_{n+1}} \mathcal{E}_n \xrightarrow{D_n} \mathcal{E}_{n-1} \ \cdots$$

A family $p = (p_1, p_2, \ldots)$ in which p_n is a probability distribution on $\mathcal{E}_n$ is said to be mutually consistent, or Kolmogorov-consistent, if each p_{n-1} is the marginal distribution obtained from p_n under deletion of element n from the set $[n]$. In other words, $p_{n-1}(A) = p_n(D_n^{-1}A)$ for $A \subset \mathcal{E}_{n-1}$. Kolmogorov consistency guarantees the existence of a random partition B of the natural numbers whose finite restrictions $B[n]$ are distributed as p_n. The partition is infinitely exchangeable if each p_n is finitely exchangeable. Some authors, for example Kingman (1980), refer to p as *a partition structure*.

An exchangeable partition process may be generated from an exchangeable sequence $Y_1, Y_2, \ldots$ by the transformation $B_{ij} = 1$ if $Y_i = Y_j$ and zero otherwise. The Dirichlet-multinomial and the Ewens processes are generated in this way. Kingman's (1978) paintbox construction shows that every exchangeable partition process may be generated from an exchangeable sequence in this manner. Moreover, the list of relative block sizes in decreasing order has a limit, which may be random. In the case of the Ewens process, the relative size of the largest block, $X_n = \max_{b\in B} \#b/n$, has a limit $X_n \to X$ distributed as beta with parameter $(1, \lambda)$, i.e. with density $\lambda(1-x)^{\lambda-1}$ for $0 < x < 1$. Given the size of the largest block, the relative size of the next largest block as a fraction of the remaining elements has the same distribution, and so on.

Let B be an infinitely exchangeable partition, $B \sim p$, which means that the restriction $B[n]$ of B to $[n]$ is distributed as p_n. Let B^* be a fixed partition in $\mathcal{E}_n$, and suppose that the event $B[n] \le B^*$ occurs. Then $B[n]$ lies in the lattice interval $[\mathbf{0}_n, B^*]$, which means that $B[n] = B[b_1]|B[b_2]|\ldots$ is the concatenation (union) of partitions of the blocks $b \in B^*$. For each block $b \in B^*$, the restriction $B[b]$ is distributed as $p_{\#b}$, so it is natural to ask whether, and under what conditions, the blocks of B^* are partitioned independently given $B[n] \le B^*$. Conditional independence implies that

$$p_n(B \mid B[n] \le B^*) = \prod_{b\in B^*} p_{\#b}(B[b]), \tag{3}$$

which is a type of non-interference or lack-of-memory property not dissimilar to that of the exponential distribution on the real line. It is straightforward to check that the condition is satisfied by (2) but not by (1). Aldous (1996) shows that conditional independence uniquely characterizes the Ewens family. Mixtures of Ewens processes do not have this property.

4 Further exchangeable partition models

Although Dirichlet partition processes are the most common in applied work, it is useful to know that many alternative partition models exist. Although some of these are easy to simulate, most do not have simple expressions for the distributions, but there are exceptions of the form

$$p_n(B; \lambda) = \frac{\Gamma(B)\, Q_n(B; \lambda)}{\lambda^{\uparrow n}}, \tag{4}$$

for certain polynomials $Q_n(B;\lambda)$ of degree $\#B$ in λ. One such polynomial is

$$Q_n(B,\lambda) = \sum_{B\le B'\le \mathbf{1}_n} \lambda^{\#B'}/B',$$

which depends on B only through the block sizes. The functions $\Gamma(B) = \prod_{b\in B}\Gamma(\#b)$ and $B^\alpha = \prod_{b\in B}(\#b)^\alpha$ are multiplicative $\mathcal{E}_n \to \mathcal{R}$, and $1/B = B^{-1}$ is the inverse of the product of block sizes.

For each $\lambda > 0$, $p_n(B;\lambda)$ depends on B only through the block sizes, so the distribution is exchangeable. Moreover, it can be shown that the family is mutually consistent in the Kolmogorov sense. However, the conditional independence property (3) is not satisfied.

The expected number of blocks grows slowly with n, approximately $\lambda\log(n)$ for the Ewens process, and $\lambda\log^2(n)/\log\log(n)$ for the process shown above.

5 Chinese restaurant process

A partition process is a random partition $B \sim p$ of a countably infinite set $\{u_1, u_2, \ldots\}$, and the restriction $B[n]$ of B to $\{u_1, \ldots, u_n\}$ is distributed as p_n. The conditional distribution of $B[n+1]$ given $B[n]$ is determined by the probabilities assigned to those events in $\mathcal{E}_{n+1}$ that are compatible with $B[n]$, i.e. the events $u_{n+1} \mapsto b$ for $b \in B$ and $b = \emptyset$. For the uniform Dirichlet-multinomial model (1), these are

$$\mathrm{pr}(u_{n+1} \mapsto b \mid B[n] = B) = \begin{cases} (\#b + \lambda/k)/(n+\lambda) & b \in B \\ \lambda(1 - \#B/k)/(n+\lambda) & b = \emptyset. \end{cases} \tag{5}$$

In the limit as $k \to \infty$, we obtain

$$\mathrm{pr}(u_{n+1} \mapsto b \mid B[n] = B) = \begin{cases} \#b/(n+\lambda) & b \in B \\ \lambda/(n+\lambda) & b = \emptyset, \end{cases} \tag{6}$$

which is the conditional probability for the Ewens process.

To each partition process p there corresponds a sequential description called the Chinese restaurant process, in which $B[n]$ is the arrangement of the first n customers at $\#B$ tables. The placement of the next customer is determined by the conditional distribution $p_{n+1}(B[n+1] \mid B[n])$ (Pitman, 1996). For the Ewens process, the customer chooses a new table with probability $\lambda/(n+\lambda)$ or one of the occupied tables with probability proportional to the number of occupants. This description, which is due to Dubins and Pitman, first appears in print in section 11 of Aldous (1983). It was used initially in connection with the Ewens and Dirichlet-multinomial models, but has subsequently been applied more broadly to general partition models.

6 Random permutations

Beginning with the uniform distribution on the set Π_n of permutations of $[n]$, the exponential family with canonical parameter $\theta = \log(\lambda)$ and canonical statistic

$\#\sigma$ equal to the number of cycles is

$$q_n(\sigma) = \lambda^{\#\sigma}/\lambda^{\uparrow n}.$$

The Stirling number of the first kind, $S_{n,k}$, is the number of permutations of $[n]$ having exactly k cycles, for which $\lambda^{\uparrow n} = \sum_{k=1}^n S_{n,k}\lambda^k$ is the ordinary generating function. The cycles of the permutation determine a partition of $[n]$ whose distribution is (2), and a partition of the integer n whose distribution is (7). From the cumulant function

$$\log(\lambda^{\uparrow n}) = \sum_{j=0}^{n-1} \log(j + \lambda)$$

it follows that $\#\sigma = X_0 + \cdots + X_{n-1}$ is the sum of independent Bernoulli variables with parameter $E(X_j) = \lambda/(\lambda + j)$, which is evident also from the Chinese restaurant representation. For large n, the number of cycles is roughly Poisson with parameter $\lambda \log(n)$, implying that $\hat{\lambda} \simeq \#\sigma/\log(n)$ is a consistent estimate as $n \to \infty$, but practically inconsistent.

A minor modification of the Chinese restaurant process also generates a random permutation by keeping track of the cyclic arrangement of customers at tables. After n customers are seated, the next customer chooses a table with probability (5) or (6), as determined by the partition process. If the table is occupied, the new arrival sits to the left of one customer selected uniformly at random from the table occupants. The random permutation thus generated is $j \mapsto \sigma(j)$ from j to the left neighbour $\sigma(j)$.

The cycles of a permutation $\sigma\colon [n] \to [n]$ determine a partition $B_\sigma \in \mathcal{E}_n$, which is a mapping $\Pi_n \to \mathcal{E}_n$ from permutations to partitions. Thus, any probability distribution p_n on partitions can be lifted to a probability distribution $q_n(\sigma) = p_n(B_\sigma)/\Gamma(B_\sigma)$ on permutations. Provided that the partition process $\{p_n\}$ is consistent and exchangeable, the lifted distributions $\{q_n\}$ are exchangeable and mutually consistent under the projection $\Pi_n \to \Pi_{n-1}$ on permutations in which element n is deleted from the cycle representation (Aldous, 1983; Pitman, 2006, section 3.1). In this way, every infinitely exchangeable random partition also determines an infinitely exchangeable random permutation $\sigma\colon \mathbb{N} \to \mathbb{N}$ of the natural numbers. Since the group acts on itself by conjugation, distributional exchangeability in this context is not to be confused with uniformity on Π_n.

7 On the number of unseen species

A partition of the set $[n]$ is a set of blocks, and the block sizes determine a partition of the integer n. For example, the partition 15|23|4 of the set [5] is associated with the integer partition $2+2+1$, one singleton and two doubletons. An integer partition $m = (m_1, \ldots, m_n)$ is a list of multiplicities, also written as $m = 1^{m_1}2^{m_2}\cdots n^{m_n}$, such that $\sum jm_j = n$. The number of blocks, usually called the number of parts of the integer partition, is the sum of the multiplicities $m_{\bullet} = \sum m_j$.

Under the natural action $B \mapsto \pi B \pi^{-1}$ of permutations π on set partitions, each orbit is associated with a partition of the integer n. The multiplicity vector m contains all the information about block sizes, but there is a subtle transfer of emphasis from block sizes to the multiplicities of the parts.

By definition, an exchangeable distribution on set partitions is a function only of the block sizes, so $p_n(B) = q_n(m)$, where m is the integer partition corresponding to B. Since there are

$$\frac{n!}{\prod_{j=1}^n (j!)^{m_j} m_j!}$$

set partitions B corresponding to a given integer partition m, to each exchangeable distribution p_n on set partitions there corresponds a marginal distribution

$$q_n(m) = p_n(B) \times \frac{n!}{\prod_{j=1}^n (j!)^{m_j} m_j!}$$

on integer partitions. For example, the Ewens distribution on integer partitions is

$$\frac{\lambda^{m_{\bullet}} \Gamma(\lambda) \prod \Gamma(j)^{m_j}}{\Gamma(n+\lambda)} \times \frac{n!}{\prod_{j=1}^n (j!)^{m_j} m_j!} = \frac{\lambda^{m_{\bullet}}\, n!\, \Gamma(\lambda)}{\Gamma(n+\lambda) \prod_j j^{m_j} m_j!}, \tag{7}$$

where the combinatorial factor $n!/\prod_j j^{m_j} m_j!$ is the size of the conjugacy class m, i.e. the number of permutations whose cycle structure is m.

Arratia, Barbour and Tavaré, (1992) noted that this version leads naturally to an alternative description of the Ewens distribution in which the multiplicities $M = M_1, \ldots, M_n$ are independent Poisson random variables with mean $E(M_j) = \lambda/j$. Then the conditional distribution $\mathrm{pr}(M = m \mid \sum_{j=1}^n jM_j = n)$ is the Ewens integer-partition distribution with parameter λ (Kingman 1993, section 9.5). In fact, we may consider the more general two-parameter Poisson model with means $E(M_j) = \lambda\theta^j/j$ for $\lambda, \theta > 0$, in which case the pair $(\sum M_j, \sum jM_j)$ is minimal sufficient for (θ, λ), and the conditional distribution given $\sum_{j=1}^n jM_j$ is (7) independent of θ. For a response vector in the form of an integer partition, for example Fisher (1943) or Efron and Thisted (1976), this representation leads naturally to a simple method of estimation and testing, using Poisson log-linear models with model formula $1 + j$ and offset $-\log(j)$.

The problem of estimating the number of unseen species was first tackled in a paper by Fisher (1943), using an approach that appears to be entirely unrelated to partition processes. Specimens from species i occur as a Poisson process with rate ρ_i, the rates for distinct species being independent and identically distributed gamma random variables. The number $N_i \geq 0$ of occurrences of species i in an interval of length t is a negative binomial random variable

$$\mathrm{pr}(N_i = x) = (1-\theta)^\nu \theta^x \frac{\Gamma(\nu + x)}{x!\, \Gamma(\nu)}. \tag{8}$$

In this setting, $\theta = t/(1+t)$ is a monotone function of the sampling time, whereas $\nu > 0$ is a fixed number independent of t. Specimen counts for distinct species are independent and identically distributed random variables with parameters $\nu > 0$ and $0 < \theta < 1$.

The probability that no specimens from species i occur in the sample is $(1-\theta)^\nu$, the same for every species. Most species are unlikely to be observed if either θ is small, i.e. the time interval is short, or ν is small.

Let M_x be the number of species occurring $x \ge 0$ times, so that $M_\bullet$ is the unknown total number of species of which $M_\bullet - M_0$ are observed. The approach followed by Fisher is to estimate the parameters θ, ν by conditioning on the number of species observed and regarding the observed multiplicities M_x for $x \ge 1$ as multinomial with parameter vector proportional to the negative binomial frequencies (8). For Fisher's entomological examples, this approach pointed to $\nu = 0$, consistent with the Ewens distribution (7), and indicating that the data are consistent with the number of species being infinite. Fisher's approach using a model indexed by species is less direct for ecological purposes than a process indexed by specimens. Nonetheless, subsequent analyses by Good and Toulmin (1956), Holgate (1969) and Efron and Thisted (1976) showed how Fisher's model can be used to make predictions about the likely number of new species in a subsequent temporal extension of the original sample. This amounts to a version of the Chinese restaurant process.

At this point, it is worth clarifying the connection between Fisher's negative binomial formulation and the Ewens partition formulation. The relation between them is the same as the relation between binomial and negative binomial sampling schemes for a Bernoulli process: they are not equivalent, but they are complementary. The partition formulation is an exchangeable process indexed by *specimens*: it gives the distribution of species numbers in a sample consisting of a fixed number of *specimens*. Fisher's version is also an exchangeable process, in fact an iid process, but this process is indexed by *species*: it gives the distribution of the sample composition for a fixed set of *species* observed over a finite period. In either case, the conditional distribution given a sample containing k species and n specimens is the distribution induced from the uniform distribution on the set of $S_{n,k}$ permutations having k cycles. For the sorts of ecological or literary applications considered by Good and Toulmin (1956) or Efron and Thisted (1976), the partition process indexed by specimens is much more direct than one indexed by species.

Fisher's finding that the multiplicities decay as $E(M_j) \propto \theta^j/j$, proportional to the frequencies in the log-series distribution, is a property of many processes describing population structure, either social structure or genetic structure. It occurs in Kendall's (1975) model for family sizes as measured by surname frequencies. One explanation for universality lies in the nature of the transition rates for Kendall's process, a discussion of which can be found in section 2.4 of Kelly (1978).

8 Equivariant partition models

A family $p_n(\sigma;\theta)$ of distributions on permutations indexed by a parameter matrix θ, is said to be equivariant under the induced action of the symmetric group if $p_n(\sigma;\theta) = p_n(g\sigma g^{-1}; g\theta g^{-1})$ for all σ, θ, and for each group element $g\colon [n] \to [n]$. By definition, the parameter space is closed under conjugation: $\theta \in \Theta$ implies $g\theta g^{-1} \in \Theta$. The same definition applies to partition models. Unlike exchangeability, equivariance is not a property of a distribution, but a property of the family. In this setting, the family is indexed by $\theta \in \Theta$ for some fixed n. There is no implication that the family p_n is the same as the family of marginal distributions induced by deletion from $[n+1]$.

Exponential family models play a major role in both theoretical and applied work, so it is natural to begin with such a family of distributions on permutations of the matrix-exponential type

$$p_n(\sigma;\theta) = \alpha^{\#\sigma} \exp(\mathrm{tr}(\sigma\theta))/M_\alpha(\theta),$$

where $\alpha > 0$ and $\mathrm{tr}(\sigma\theta) = \sum_{j=1}^n \theta_{\sigma(j),j}$ is the trace of the ordinary matrix product. The normalizing constant is the α-permanent

$$M_\alpha(\theta) = \mathrm{per}_\alpha(K) = \sum_\sigma \alpha^{\#\sigma} \prod_{j=1}^n K_{\sigma(j),j}$$

where $K_{ij} = \exp(\theta_{ij})$ is the component-wise exponential matrix. This family of distributions on permutations is equivariant.

The limit of the α-permanent as $\alpha \to 0$ gives the sum of cyclic products

$$\mathrm{cyp}(K) = \lim_{\alpha\to 0} \alpha^{-1}\,\mathrm{per}_\alpha(K) = \sum_{\sigma:\#\sigma=1} \prod_{j=1}^n K_{\sigma(j),j},$$

giving an alternative expression for the α-permanent

$$\mathrm{per}_\alpha(K) = \sum_{B\in\mathcal{E}_n} \alpha^{\#B} \prod_{b\in B} \mathrm{cyp}(K[b])$$

as a sum over partitions. The induced marginal distribution (11) on partitions is of the product-partition type recommended by Hartigan (1990), and is also equivariant. Note that the matrix θ and its transpose determine the same distribution on partitions, but they do not usually determine the same distribution on permutations. The α-permanent has a less obvious convolution property that helps to explain why this function might be expected to occur in partition models:

$$\sum_{b\subset[n]} \mathrm{per}_\alpha(K[b])\,\mathrm{per}_{\alpha'}(K[\bar b]) = \mathrm{per}_{\alpha+\alpha'}(K). \tag{9}$$

The sum extends over all 2^n subsets of $[n]$, and $\bar b$ is the complement of b in $[n]$. A derivation can be found in section 2.4 of McCullagh and Møller (2006).

If B is a partition of $[n]$, the symbol $K \cdot B = B \cdot K$ denotes the Hadamard component-wise matrix product for which

$$\operatorname{per}_\alpha(K \cdot B) = \prod_{b \in B} \operatorname{per}_\alpha(K[b])$$

is the product over the blocks of B of α-permanents restricted to the blocks. Thus the function $B \mapsto \operatorname{per}_\alpha(K \cdot B)$ is of the product-partition type.

With α, K as parameters, we may define a family of probability distributions on $\mathcal{E}_n^k$, i.e. partitions of $[n]$ having k or fewer blocks, as follows:

$$p_{nk}(B) = k^{\downarrow \#B} \operatorname{per}_{\alpha/k}(K \cdot B) / \operatorname{per}_\alpha(K). \tag{10}$$

The fact that (10) is a probability distribution on $\mathcal{E}_n$ follows from the convolution property of permanents. The limit as $k \to \infty$

$$p_n(B) = \alpha^{\#B} \prod_{b \in B} \operatorname{cyp}(K[b]) / \operatorname{per}_\alpha(K), \tag{11}$$

is a product-partition model satisfying the conditional independence property (3).

Properties of the α-permanent are discussed by Vere-Jones (1997) and by McCullagh and Møller (2006) in the context of point processes. For $K = \mathbf{1}_n$, the $n \times n$ matrix whose elements are all one, the α-permanent is, by definition, the generating function for the Stirling numbers of the first kind. Thus, $\operatorname{per}_\alpha(\mathbf{1}_n) = \alpha^{\uparrow n}$ is the ascending factorial function, and for this exchangeable case, the distributions (10) and (11) coincide with (1) and (2).

9 Further applications of partition models

Partition models are used to construct cluster processes for use in classification and cluster analysis. Cluster analysis means a partitioning of the sample units into non-overlapping blocks such that the Y-values in $\mathcal{R}^d$ (feature values) are more similar within blocks than between blocks. It is important to remember that the goal of cluster analysis is not a partition of the feature space $\mathcal{R}^d$, but a partition of the finite set of units or specimens.

Exchangeable partition models are used to construct non-trivial, processes suitable for cluster analysis. See Richardson and Green (1997), Fraley and Raftery (2002) or Booth, Casella and Hobert (2008) for a discussion of computational techniques. The simplest of these models is the marginal Gauss-Ewens process in which the sample partition $B[n]$ is to be inferred from the finite sequence $Y[n]$. The conditional distribution $p_n(B \mid Y[n])$ on $\mathcal{E}_n$ is the posterior distribution on clusterings or partitions of $[n]$, and $E(B \mid Y[n])$ is the array of one-dimensional marginal distributions for pairs of units, i.e. $E(B_{ij} \mid Y[n])$ is the posterior probability that units i, j belong to the same block. The conditional distribution $p_n(B \mid Y[n])$ contains further information about triplets and k-tuples of units, from which it is possible in principle to compute the posterior distribution for the number of clusters or blocks. In estimating the number of clusters, it is important to distinguish between the sample number $\#B[n]$, which is necessarily finite, and the population number $\#B[\mathbb{N}]$, which could be infinite (McCullagh and Yang, 2008). The latter problem is essentially the same as estimating the number of unseen species given that the blocks are so well separated that $Y[n]$ determines $B[n]$.

The same Gauss-Ewens model may be used for density estimation, which refers to the conditional distribution of Y_{n+1} given the sample values. Usually, this is to be done for an exchangeable process in the absence of external covariate or relational information about the units. In the computer-science literature, cluster detection is also called unsupervised learning.

Exchangeable partition models are also used to provide a Bayesian solution to the multiple comparisons problem (Gopalan and Berry 1998). In this setting k is the number of distinct treatments, and the key idea is to associate with each partition B of $[k]$ a subspace $V_B \subset \mathcal{R}^k$ equal to the span of the columns of B. Thus, V_B consists of vectors x such that $x_r = x_s$ if $B_{rs} = 1$. For a treatment factor having k levels with values $\tau_1, \ldots, \tau_k$, the Gauss-Ewens prior distribution on R^k puts positive mass on the subspaces V_B for each $B \in \mathcal{E}_k$. Likewise, the posterior distribution also puts positive probability on these subspaces, which enables us to compute in a coherent way the posterior probability $\mathrm{pr}(\tau \in V_B \mid y)$ or the marginal posterior probability $\mathrm{pr}(\tau_r = \tau_s \mid y)$.

10 Acknowledgements

This is a revised version of the article *Random permutations and partition models* from Lovric, Miodrag (2011), International Encyclopedia of Statistical Science, Heidelberg: Springer Science +Business Media, LLC. Support for this research was provided in part by NSF Grant DMS-0906592.

References

[1] Aldous, D.J. (1983) Exchangeability and Related Topics. In *École d'Éte de Probabilités de Saint-Flour XIII* Springer Lecture Notes in Mathematics vol 1117, 1–198.

[2] Aldous, D.J. (1996) Probability distributions on cladograms. In Random Discrete Structures. IMA Vol. Appl. Math **76**. Springer, New York, 1–18.

[3] Arratia, R., Barbour, A.D. and Tavaré, S. (1992) Poisson process approximations for the Ewens sampling formula. *Advances in Applied Probability* **2**, 519–535.

[4] Booth, J.G., Casella, G. and Hobert, J.P. (2008) Clustering using objective functions and stochastic search. *J. Roy. Statist. Soc. B* **70**, 119–139.

[5] Blackwell, D. and MacQueen, J. (1973) Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1**, 353–355.

[6] Efron, B. and Thisted, R.A. (1976) Estimating the number of unknown species: How many words did Shakespeare know? *Biometrika* **63**, 435–447.

[7] Ewens, W.J. (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87–112.

[8] Fraley, C. and Raftery, A.E. (2002) Model-based clustering, discriminant analysis and density estimation. *J. Amer. Statist. Assoc.* **97**, 611–631.

[9] Good, I.J. and Toulmin, G.H. (1956) The number of new species, and the increase in population coverage when a sample is increased. *Biometrika* **43**, 45–63.

[10] Gopalan, R. and Berry, D.A. (1998) Bayesian multiple comparisons using Dirichlet process priors. *J. Amer. Statist. Assoc.* **93**, 1130–1139.

[11] Hartigan, J.A. (1990) Partition models. *Communications in Statistics: Theory and Methods* **19**, 2745–2756.

[12] Holgate, P. (1969) Species frequency distributions. *Biometrika* **65**, 651–660.

[13] Kelly, F.P. (1978) *Reversibility and Stochastic Networks.* Wiley, Chichester.

[14] Kingman, J.F.C. (1975) Random discrete distributions (with discussion). *J. Roy. Statist. Soc. B* **37**, 1–22.

[15] Kingman, J.F.C. (1977) The population structure associated with the Ewens sampling formula. *Theoretical Population Biology* **11**, 274–283.

[16] Kingman, J.F.C. (1978) The representation of partition structures. *J. Lond. Math. Soc.* **18**, 374–380.

[17] Kingman, J.F.C. (1980) *Mathematics of Genetic Diversity.* CBMS-NSF conference series in applied math, **34** SIAM, Philadelphia.

[18] McCullagh, P. and Møller, J. (2006) The permanental process. *Adv. Appl. Prob.* **38**, 873–888.

[19] McCullagh, P. and Yang, J. (2008). How many clusters? *Bayesian Analysis* **3**, 1–19.

[20] Pitman, J. (1996) Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory: Papers in Honor of David Blackwell,* T.S. Ferguson et al editors. IMS Lecture Notes Monograph Series No. 30, 245–267.

[21] Pitman, J. (2006) *Combinatorial Stochastic Processes.* Springer-Verlag, Berlin.

[22] Richardson, S. and Green, P.J. (1997) On Bayesian analysis of mixtures with an unknown number of components (with discussion). *J. Roy. Statist. Soc. B* **59**, 731–792.

[23] Vere-Jones, D. (1997) Alpha-permanents and their application to multivariate gamma, negative binomial and ordinary binomial distributions. *New Zealand J. Math.* **26** 125–149.

[24] Watterson, G.A. (1974) The sampling theory of selectively neutral alleles. *Adv. Appl. Prob.* **6**, 217–250.