Random partitions and other combinatorial objects

Peter McCullagh

Department of Statistics
University of Chicago

Hotelling Lecture II
UNC Chapel Hill
December 3 2008

Outline

Partitions

$[n] = \{1, \ldots, n\}$ a finite set

A partition B of the set $[n] = [6]$ is

(i) a set of disjoint non-empty subsets $b \subset [n]$ called blocks...

e.g. $B = \{\{2,4,6\},\{1,3\},\{5\}\} \equiv 246|13|5 \equiv 13|246|5$

(ii) an equivalence relaton $B\colon [n] \times [n] \to \{0,1\}$

s.t. $B(i,j) = 1$ if $i \sim j$ belong to the same block

(iii) a symmetric Boolean matrix

$$B = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

$\#B$: number of elements (no. of blocks)

for $b \in B$, $\#b > 0$ is the number of elements ($\#b > 0$)

Integer partition $\nu(B) = 1 + 2 + 3$ associated with $B = 246|13|5$

The set $\mathcal{E}_n$ of partitions of $[n]$

$\mathcal{E}_1$: 1
$\mathcal{E}_2$: 12, 1|2
$\mathcal{E}_3$: 123, 12|3, 13|2, 23|1, 1|2|3|
$\mathcal{E}_4$: 1234, 123|4[4], 12|34[3], 12|3|4[6], 1|2|3|4
$\mathcal{E}_5$: 12345, 1234|5[5], 123|45[10], 123|4|5[10], 12|34|5[15], 12|3|4|5[10], 1|2|3|4|5
$\#\mathcal{E}_n$: 1, 2, 5, 15, 52, 203, 877, 4140, 21147, 115975, 678570

Permutation map $\pi : [n] \to [n]$ also acts $\pi^* : \mathcal{E}_n \to \mathcal{E}_n$
partition type $\nu(B)$ is maximal invariant
Deletion map: $D_n : \mathcal{E}_n \to \mathcal{E}_{n-1}$ (onto)

D_4: $1234,\ 123|4 \mapsto 123$
$12|3|4,\ 12|34,\ 124|3 \mapsto 12|3$ (three types)
$1|2|3|4,\ 1|2|34,\ 1|24|3,\ 14|2|3 \mapsto 1|2|3$

Same as removal of last row and column from matrix

$\mathcal{E}$ represents the sets $\{\mathcal{E}_n\}$ with permutation and deletion maps

Probability distributions on partitions

P_n a probability distribution on $\mathcal{E}_n$
Finitely exchangeable if $\nu(B) = \nu(B')$ implies $P_n(B) = P_n(B')$
Examples:

$\mathcal{E}_3$	123	12|3	13|2	23|1	1|2|3
P_3	$1/3$	$1/6$	$1/6$	$1/6$	$1/6$
P_3'	$1/6$	$1/6$	$1/6$	$1/6$	$1/3$

$\mathcal{E}_4$	1234	123|4	13|24	23|1|4	1|2|3|4
P_4	$1/4$	$1/12$	$1/24$	$1/24$	$1/24$
P_4'	$1/10$	$1/15$	$1/30$	$1/15$	$2/15$

Compatibility:
$P_4(1234 \cup 123|4) = 1/4 + 1/12 = 1/3 = P_3(123)$
$P_4(12|3|4 \cup 12|34 \cup 124|3) = 2/24 + 1/12 = 1/6 = P_3(12|3)$
$P_4(1|2|3|4 \cup 1|2|34[3]) = 1/24 + 3/24 = 1/6 = P_3(1|2|3)$
P_3 is the marginal distribution of P_4 under deletion
P_3' is the marginal distribution of P_4' under deletion

Exchangeable partition process

An *exchangeable partition process* is a sequence P_n such that
each P_n is finitely exchangeable
$P_n(B)$ depends only on block sizes $\nu(B)$
P_n is the marginal distribution of P_{n+1}.

Kolmogorov compatibility condition:

$$P_n(B) = \sum_{B' :\, D_{n+1}B' = B} P_{n+1}(B')$$

Conditional distribution

$$P_{n+1}(B' \mid B \in \mathcal{E}_n) = \begin{cases} P_{n+1}(B')/P_n(B) & D_{n+1}B' = B \\ 0 & \text{otherwise.} \end{cases}$$

Kingman's paintbox characterization

The Ewens partition process

Ewens distribution with parameter $\lambda > 0$

$$P_n(B) = \frac{\Gamma(\lambda)\lambda^{\#B}}{\Gamma(n+\lambda)} \prod_{b \in B} \Gamma(\#b)$$

Conditional distributions

$$P_{n+1}(u_{n+1} \mapsto b \,|\, B) = \frac{P_{n+1}(B')}{P_n(B)} = \begin{cases} \#b/(n+\lambda) & b \in B \\ \lambda/(n+\lambda) & b = \emptyset \end{cases}$$

(Pitman's CRP description)

Induced distribution on integer partitions $\nu(B) = 1^{\nu_1} 2^{\nu_2} \cdots n^{\nu_n}$

$$Q_n(\nu) = \frac{\Gamma(\lambda)\lambda^{\nu_\cdot}}{\Gamma(n+\lambda)} \prod_{j=1}^{n} ((j-1)!)^{\nu_j} \times \frac{n!}{\prod (j!)^{\nu_j}\, \nu_j!}$$

No deletion operation for integer partitions
Hence no process on integer partitions

Permutation process

Exponential family of distributions on permutations $\sigma : [n] \to [n]$

$$P_n(\sigma;\lambda) = \lambda^{\#\sigma}/M_n(\lambda), \quad \#\sigma = \#\text{cycles}$$
$$M_n(\lambda) = \sum \lambda^r S_{n,r}$$
$$= \lambda(\lambda+1)\cdots(\lambda+n-1) = \Gamma(n+\lambda)/\Gamma(\lambda)$$
$$P_n(\sigma;\lambda) = \frac{\Gamma(\lambda)\lambda^{\#\sigma}}{\Gamma(n+\lambda)}$$

Exponential family with canonical statistic $\#\sigma$
Cumulant function $K(\lambda) = \log M(\lambda)$
determines the mean, variance,... of $\#\sigma$
(Goes back to Euler)

In what sense is this an exchangeable process?

Permutations $\{\Pi_n\}$ as a projective system

Projective system with respect to sub-sampling

A sub-sample of size m taken from $[n]$ (not random) is
- an ordered subset $\varphi_1, \ldots, \varphi_m$ distinct in $[n]$
- a 1–1 map $\varphi\colon [m] \to [n]$

sub sample $\varphi\colon [m] \to [n]$ $\longrightarrow$ deletion $\varphi^*\colon \Pi_n \to \Pi_m$
- φ^* in reverse direction on permutations
- by conjugation $\varphi^*\sigma = \varphi^{-1}\sigma\varphi$ if $m = n$
- by deletion from cycle representation if $m \leq n$
- $(i, j, \ldots)(\ldots) \mapsto (\varphi^{-1}(i), \varphi^{-1}(j), \ldots)(\ldots)$
- delete if $\varphi^{-1}(\{j\}) = \emptyset$

$(\varphi\psi)^* = \psi^*\varphi^*$ (composition in reverse order)

Exponential family distributions are compatible with these maps

Ewens permutation process

Ewens distribution on permutations Π_n

$$P_n(\sigma;\lambda) = \frac{\Gamma(\lambda)\lambda^{\#\sigma}}{\Gamma(n+\lambda)}$$

Induced distribution on partitions (cycles ignoring cyclic order)

$$P_n(B;\lambda) = \frac{\Gamma(\lambda)\lambda^{\#B}}{\Gamma(n+\lambda)} \prod_{b\in B} \Gamma(\#b)$$

Conditional distribution given Π_n

$$P_{n+1}(n+1 \mapsto (i, n+1, \sigma(i), \ldots) \,|\, \sigma) = 1/(n+\lambda) \quad 1 \le i \le n$$
$$P_{n+1}(n+1 \mapsto (n+1) \,|\, \sigma) = \lambda/(n+\lambda) \quad \text{new cycle}$$

Defines an infinite exchangeable random permutation
$\#\sigma$ is approximately $\mathrm{Po}(\lambda \log n)$
Note difference between permutation and a ranking

Other interpertations of the Ewens process

Conditional Poisson interpretation

$X_1, X_2, \ldots$ independent Poisson variables $X_j \sim \mathrm{Po}(\lambda/j)$
as multiplicities $1^{X_1} 2^{X_2} 3^{X_3} \cdots$ in integer partition

Conditional distribution of $X_1, \ldots, X_n$ given $\sum jX_j = n$

$$p(x) \propto e^{-\lambda \sum 1/j} \frac{\lambda^{x_\bullet}}{\prod j^{x_j} x_j!}$$

is exactly the Ewens partition

Negative binomial model for the number of species
Fisher 1943; Good 1953; Mosteller & Wallace 197?; Efron & Thisted 1976

Kendall's (1975) family-size process (Kelly's book)

Prime factorization of large integers (Billingsley 1972; Donnelly & Grimmett)

Partition induced by Dirichlet process

Characterization of the Ewens distribution

Why is the Ewens distribution ubiquitous?

(i) Exchangeability: $B \sim P_n$ implies $B^\pi = \pi^{-1} B \pi \sim P_n$

(ii) Restriction to subsets $[m] \subset [n]$
if $B \sim P_n$, the restriction is $B[m] \sim P_m$ (process property)

(iii) Self-similarity (lack of memory)
Given that $B \leq b|b'$, conditional distn is $B \sim P_{\#b} \times P_{\#b'}$
(Aldous, 199?)

Leading to a theory of Markov fragmentation trees...
by recursive partitioning...

The Gauss-Ewens cluster process

Cluster process has following parts:

(i) An index set $\mathbb{N}$

(ii) A random sequence $Y_1, Y_2, \ldots$ with $Y_i \in \mathcal{S}$, $(\mathcal{S} = \mathcal{R}^d)$

(iii) A random partition B of $\mathbb{N}$ (not a partition of $\mathcal{S}$)

(iv) Finite-dimensional distributions such as

$$P_n(B) = \frac{\lambda^{\#B}\,\Gamma(\lambda)}{\Gamma(n+\lambda)} \prod_{b \in B} \Gamma(\#b)$$

$$Y[n] \mid B[\mathbb{N}] \sim N(\mathbf{1}\mu,\ I_n \otimes \Sigma + B[n] \otimes \Sigma')$$

Note $B \equiv B[\mathbb{N}]$ (no interference)

Parameters: $\mu \in \mathcal{R}^d$, $\lambda > 0$;

Σ, Σ' within- and between-cluster covariance matrices

Exchangeability of cluster process

Observation for a finite sample $[n] \subset \mathbb{N}$ is $(Y[n], B[n])$

Observation space is $\mathcal{S}^n \times \mathcal{E}_n$

Permutation $\pi : [n] \to [n]$ acts on observation space
$(Y[n], B[n]) \mapsto (Y^\pi, B^\pi)$ by composition $B^\pi(i,j) = B(\pi_i, \pi_j)$

Restriction $\varphi : [m] \to [n]$ acts on observation spaces
$(Y[n], B[n]) \mapsto (Y^\varphi, B^\varphi)$ by composition
$Y^\varphi(i) = Y(\varphi(i)), \qquad B^\varphi(i,j) = B(\varphi_i, \varphi_j)$

(i) Distribution Q_n on $\mathcal{E}_n \times \mathcal{S}^n$ unaffected by permutation π of $[n]$

(ii) Q_m on $\mathcal{E}_m \times \mathcal{S}^m$ is the marginal distn of Q_n

Hence there exists an infinite random clustering (Y, B)

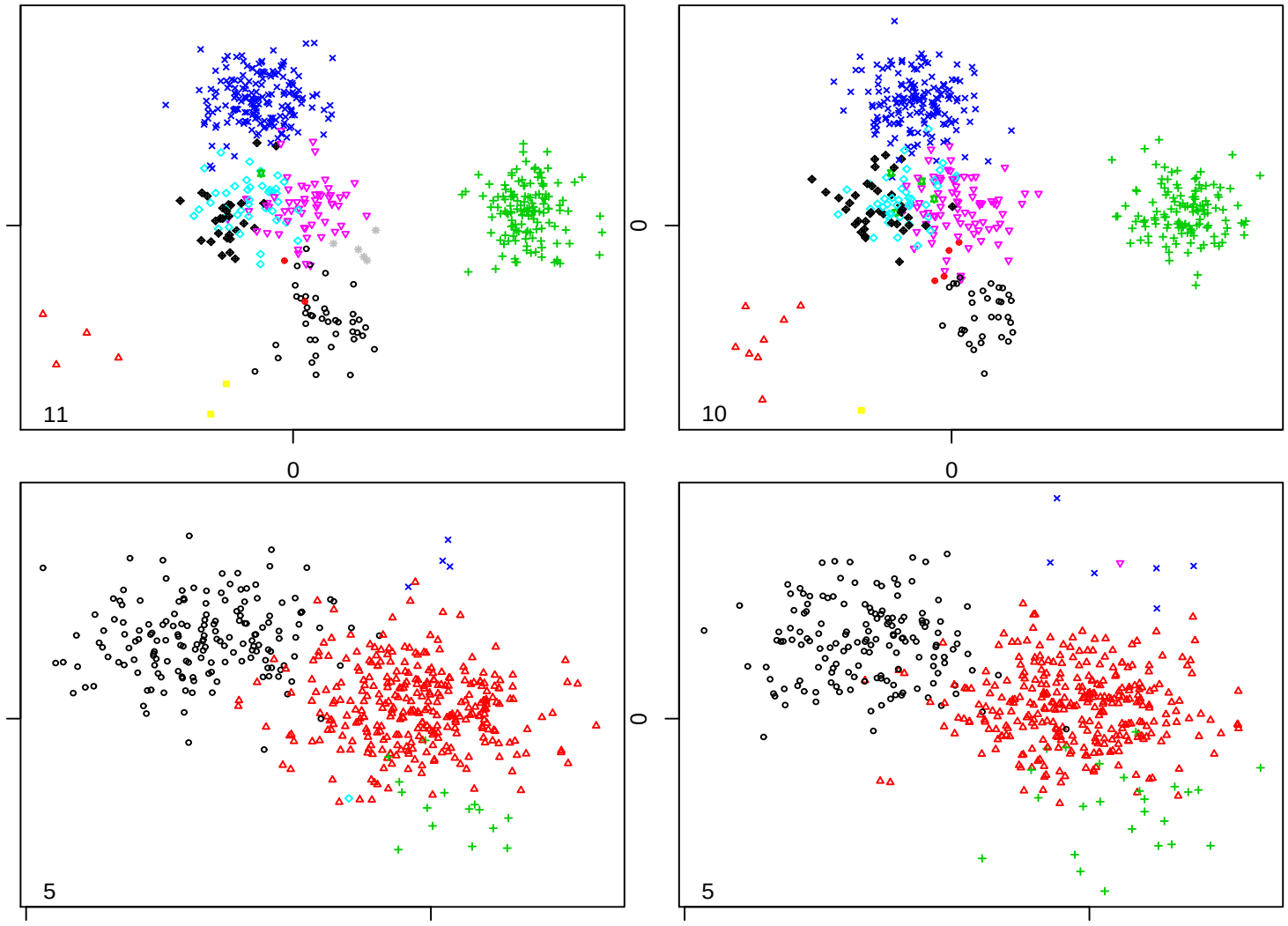
More conventional version

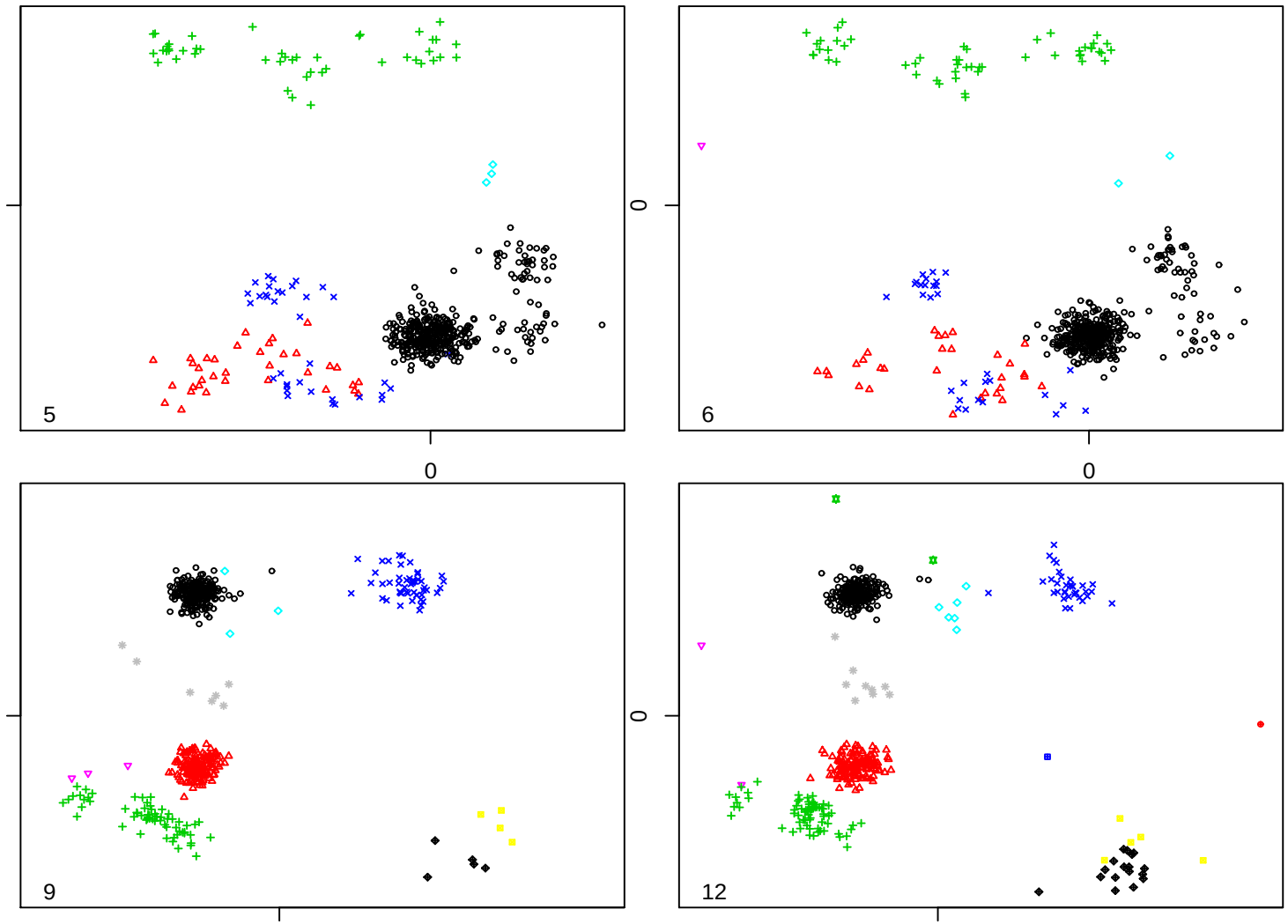
Gauss-Ewens cluster model is more or less equivalent to

$$\eta \sim \mathit{IIDN}(0, \Sigma'), \quad \epsilon \sim \mathit{IIDN}(0, \Sigma) \quad \text{independent}$$
$$\mathrm{tbl}_i = 1 + \max(\mathrm{tbl}[i-1]) \quad \text{w.p. } \lambda/(i-1+\lambda);$$
$$\text{else one of } (\mathrm{tbl}_1, \ldots, \mathrm{tbl}_{i-1}) \text{ with equal prob}$$
$$Y_i = \mu + \epsilon_i + \eta_{\mathrm{tbl}(i)}$$

with $(Y_1, \mathrm{tbl}_1), \ldots, (Y_n, \mathrm{tbl}_n)$ observed
. . .except that this version is not exchangeable

Can fix this by forgetting/ignoring table numbers
i.e. by defining `B = outer(tbl, tbl, "==")`
and saying that $(Y[n], B[n])$ is observed.


11
10
5
5
0
0
0
0


5
0
6
0
0
9
0
12

Statistical classification (aka supervised learning)

Feature $Y_u \in \mathcal{S}$ in feature space and class $t_u \in \mathcal{C}$

Training sample $u_1, \ldots, u_n$:
observed features $(Y_1, \ldots, Y_n)$ and types $(t_1, \ldots, t_n)$
and $Y_{n+1} = y^*$, to which class does u_{n+1} belong?

Deterministic interpretation: (forced choice of one $t^* \in \mathcal{C}$)
Statistical classification: probability distribution on $\mathcal{C}$
Enormous literature going back to Fisher (1936)

Logistic classification model (more recent; 1970s?)

$$\log \mathrm{pr}(t_u = r \mid y_u = y) = \frac{e^{\beta_r' y}}{\sum_r e^{\beta_s' y}}$$

for $r \in \mathcal{C}$, independently for distinct units

Cluster models for classification w/o classes

Problem: No set $\mathcal{C}$ of classes in a cluster process (Y, B)
Observation $(Y, B)[n]$ in training sample $u_1, \ldots, u_n$
How can we assign new unit u_{n+1} to classes?

Conditional distribution

$$\text{pr}(u_{n+1} \mapsto b \,|\, (Y, B)[n], y^*) = \begin{cases} f(\ldots) & b \in B \\ \ldots & \text{otherwise.} \end{cases}$$

Blocks of B are the classes!
Also need parameter estimates (at least $\lambda, \theta = \Sigma' \Sigma^{-1}$)

Lack of $\mathcal{C}$ is a big advantage!
possibility of assigning u_{n+1} to a previously unseen class

Explicit calculation of conditional distribution

Simplification $\Sigma' = \theta\Sigma$ in $\mathcal{S} = \mathcal{R}^d$

$$\text{pr}(u' \mapsto b \,|\, ...) \propto \begin{cases} \#b\, \phi_d(y(u') - \tilde{\mu}_b; \tilde{\Sigma}_b) & b \in B \\ \lambda\, \phi_d(y(u'); \Sigma(1+\theta)) & b = \emptyset \end{cases}$$

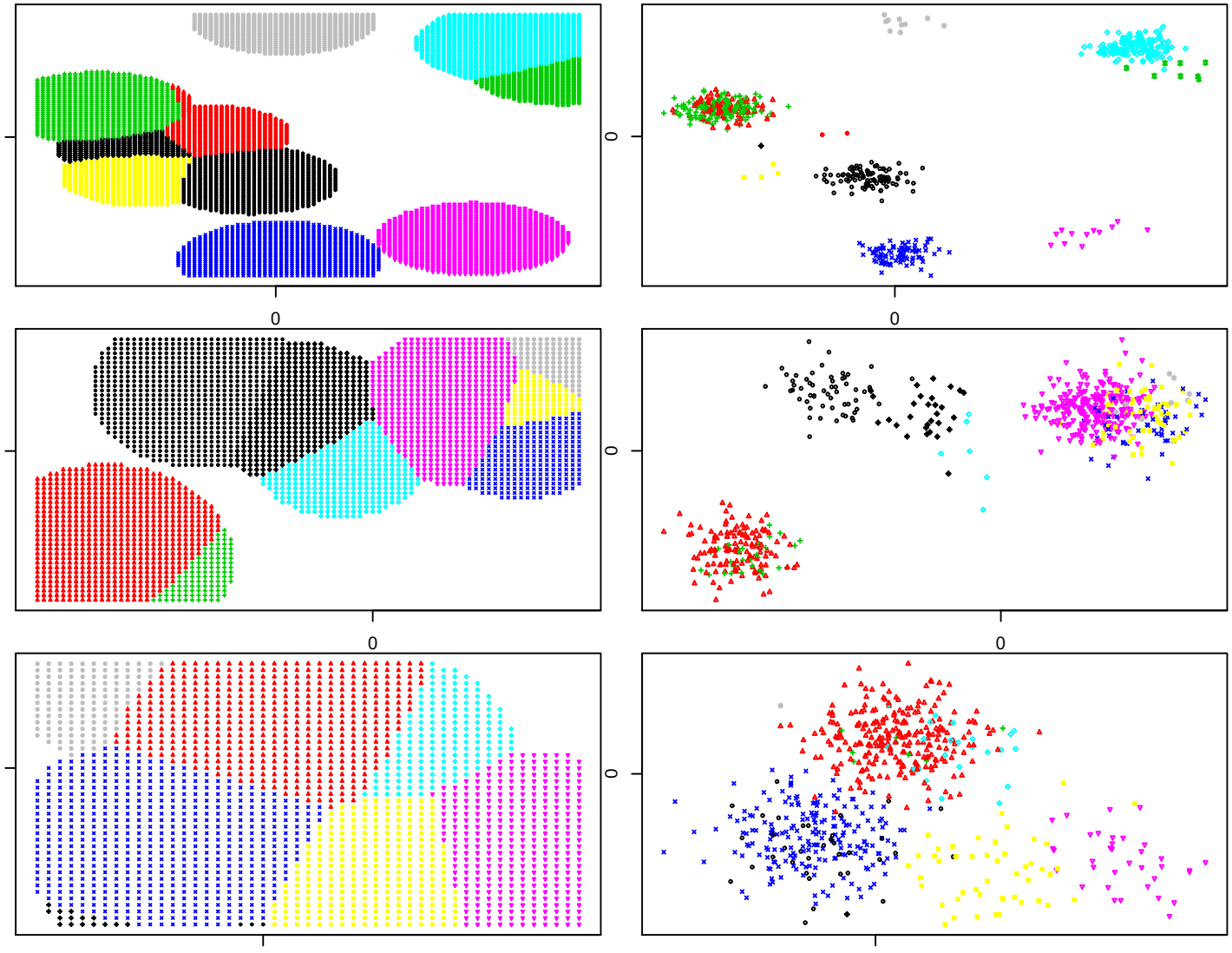
$$\tilde{\mu}_b = (\mu + n_b\theta\bar{y}_b)/(1 + n_b\theta), \quad \tilde{\Sigma}_b = \Sigma(1 + \theta/(1 + n_b\theta))$$

Typical values $\theta \geq 5$ and $n_b \geq 5$
so $\tilde{\mu}_b/\bar{y}_b = n_b\theta/(1 + n_b\theta) \geq 0.96$

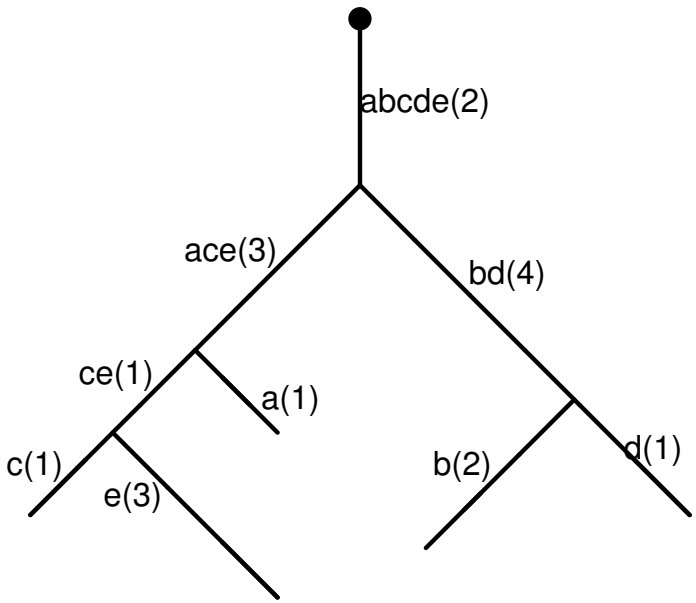
(similar to Fisher discriminant model, but with shrinkage)

Tree version with classes and sub-classes

Block having maximum conditional probability



Trees (rooted and leaf-labelled)



$$T = \begin{array}{c|ccccc} & a & b & c & d & e \\ a & 6 & 2 & 5 & 2 & 5 \\ b & 2 & 8 & 2 & 6 & 2 \\ c & 5 & 2 & 7 & 2 & 6 \\ d & 2 & 6 & 2 & 7 & 2 \\ e & 5 & 2 & 6 & 2 & 9 \end{array}$$

$$T_{ij} \geq \min(T_{ik}, T_{jk})$$

Set of rooted leaf-labelled trees

Each tree T is a positive definite matrix

$$T = \begin{array}{c|ccccc} & d & b & a & c & e \\ d & 7 & 6 & 2 & 2 & 2 \\ b & 6 & 8 & 2 & 2 & 2 \\ a & 2 & 2 & 6 & 5 & 5 \\ c & 2 & 2 & 5 & 7 & 6 \\ e & 2 & 2 & 5 & 6 & 9 \end{array}$$

$$T = \sum_e \lambda_e \, e \otimes e$$

$$\mathrm{sk}(T) = \sum e \otimes e$$

$$T_{ij} \geq \min(T_{ik}, T_{jk})$$

Skeleton tree $\mathrm{sk}(T)$ is the set of edges (also called topology)

$\mathrm{sk}(T) = \{abcde, ace, bd, ce, a, b, c, d, e\}$
No. of edges $\leq 2n - 1$.

Structure of $\mathcal{T}_n$

Symmetric, non-negative, $T_{ij} \geq \min(T_{ik}, T_{jk})$

(i) Closed under component-wise monotone transformation
$T \mapsto g(T)$ with $g(0) \geq 0$

(ii) closed under component-wise scalar multiplication

(iii) Define $(AB)_{ij} = \max_k\{\min(A_{ik}, B_{kj})\}$
then $T^2 = T$ if and only if $T \in \mathcal{T}_n$.

(iv) if A is non-negative, then $\lim_{n\to\infty} A^n = T$ exists

(v) T_n is the union of intersecting manifolds of dimension $2n-1$.
How many? $1 \cdot 3 \cdots (2n-3)$

(vi) Contrast unrooted trees: $U_{ij} = T_{ii} + T_{jj} - 2T_{ij}$
$U_{ij} + U_{kl} \leq \max\{U_{ik} + U_{jl}, U_{il} + U_{jk}\}$ (Buneman)

Exchangeable fragmentation trees

Gibbs fragmentation trees:

Exchangeable random tree:
(i) distribution P_n on $\mathcal{T}_n$ invariant under permutation
(ii) P_n is marginal distribution of P_{n+1}

Markov property:
splits independent of waiting times
edge lengths independent exponential
Branches behave independently following split
Branch of size r distributed as P_r (self-similarity)

Binary splits

Kolmogorov consistency for the Gibbs skeletal tree

Gibbs skeletal tree T: a random collection of subsets of $[n]$ satisfying certain tree conditions

$$p_n(T) = K_n^{-1} \prod_{b \in T} \psi(\#b) \tag{1}$$

$\psi_(n), n = 1, 2, 3, \ldots$ Gibbs potential function

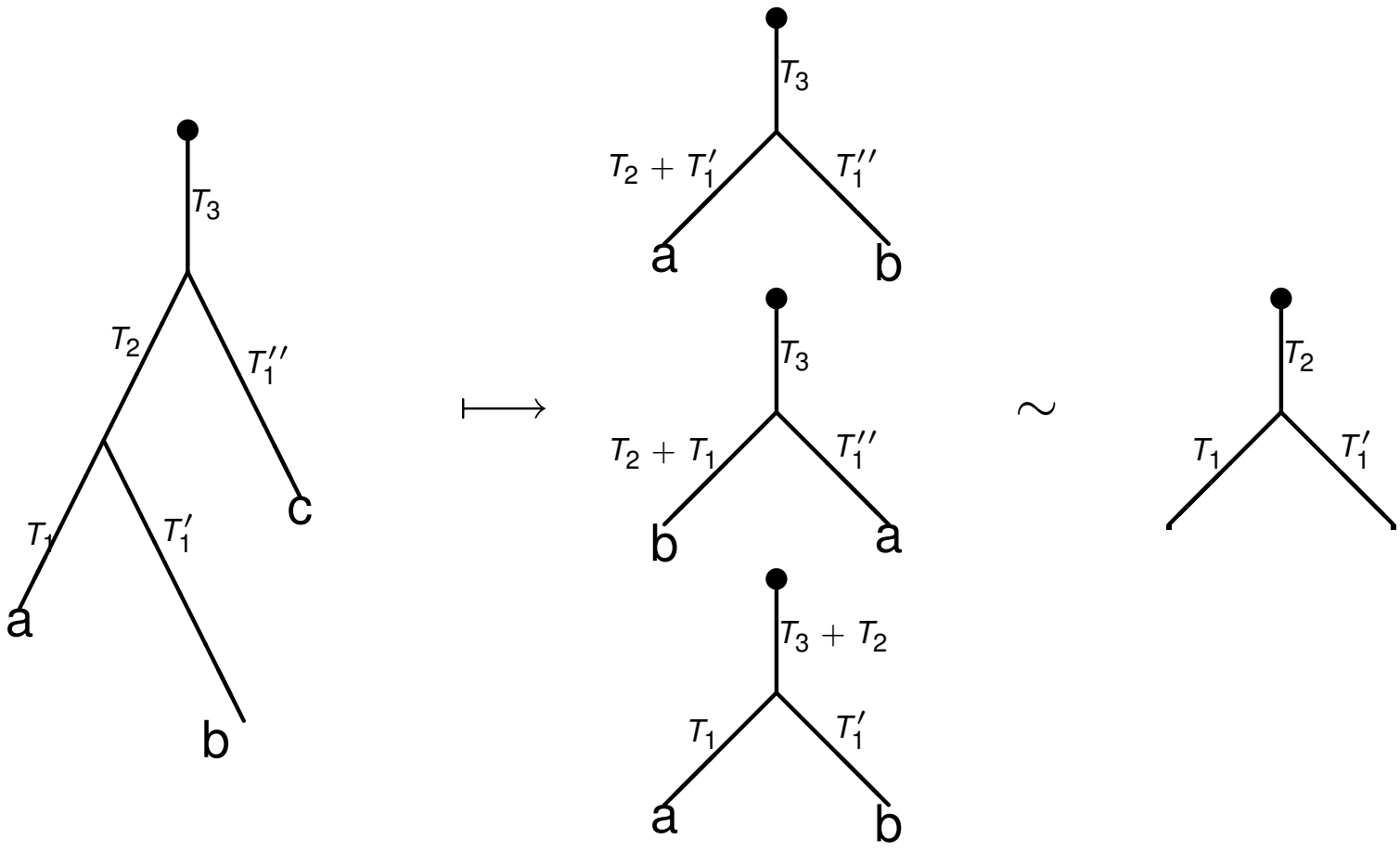
Kolmogorov consistency condition P_n is marginal distn of P_{n+1}

Can take $\psi_1 = \psi_2 = 1$ w.l.o.g.
Consistency implies

$$\psi_{n+1} = \frac{\psi_n(1 + (n-1)\gamma)}{2 + \psi_n + (n-2)\gamma}$$

for some $\gamma > 0$.
$\psi_3 = (1 + \gamma)/3$ determines the entire sequence

Leaf deletion for $n = 3$



Gibbs fragmentation trees

The class of exchangeable homogeneous Markovian trees
one-dimensional family (Aldous's beta-splitting rules $\beta > -2$)
(Bertoin, Le Gall, Berestyki, McC, Pitman, Winkel,)

One member ($\beta = -1$)

Waiting time exponential with rate
$\varphi(n) = 1 + 1/2 + \cdots + 1/(n-1)$

mean waiting time $1/\varphi(n) \simeq 1/\log(n)$ $\quad(\varphi(1) = 0)$

splitting distribution $n \mapsto r + s$

$$p_n(r,s) = \frac{n}{2\, r\, s\, \varphi(n)} \quad 1 \le r \le n-1, \quad r+s=n$$

Exercises connected with Gibbs trees

(i) Beginning with $[n]$ at $t = 0$, find the partition at time t

(ii) Description of behaviour as $n \to \infty$

e.g. size of largest block at time t

(iii) Time to complete fragmentation: $T = 2\log(n) + O_p(1)$

($\beta = -1$)

Density: $f(t) \propto (1 - e^{-t/2})^{n-2} e^{-t}$

(iv) Distn of time until u_1 is isolated (leaf height)

Density: $\sum_{r=1}^{n} \binom{n-1}{r-1} (-1)^r \varphi_r e^{-\varphi_r t}$

(v) Expected fragmentation rate given $[n]$ at time 0

$\lambda_n(t) = \sum_{r=1}^{n} \binom{n}{r} (-1)^r \varphi_r e^{-\varphi_r t}$

(vi) Analogous theory for unrooted trees

(vii) Applications of Gibbs trees in statistical models