An introduction to mixture modelling for unsupervised clustering

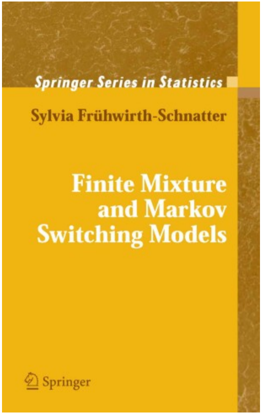
Mini-tutorial

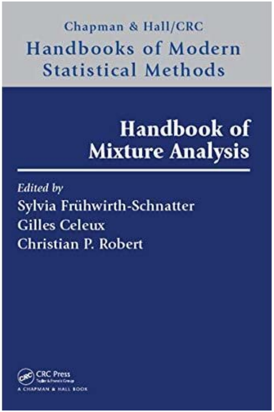
Nicole M White

Australian Centre for Health Services Innovation (AusHSI)
Queensland University of Technology

July 7, 2021

Further reading

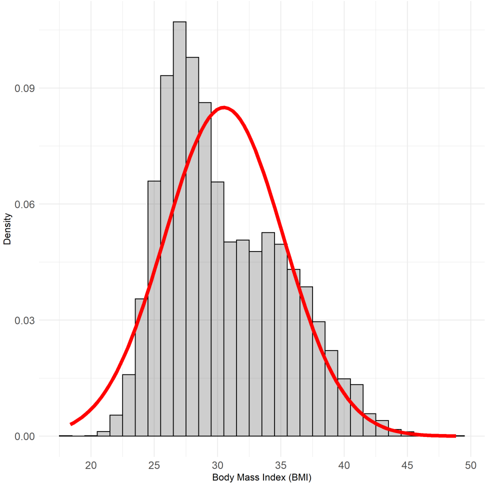




https://github.com/nicolemwhite/anzsc-mixture-modelling

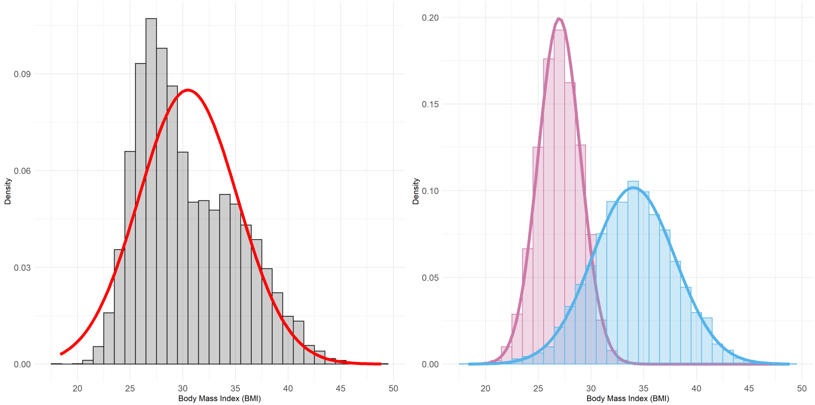
A motivating example

Distribution of body mass index (BMI) for 10,000 participants.



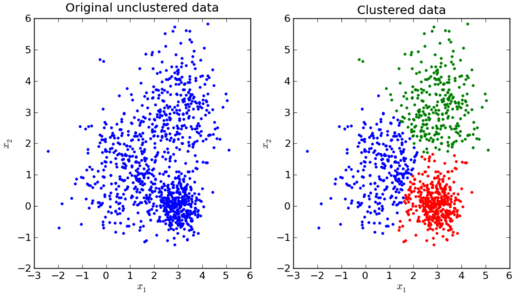
A motivating example

Distribution of body mass index (BMI) for 10,000 participants



Defining clustering

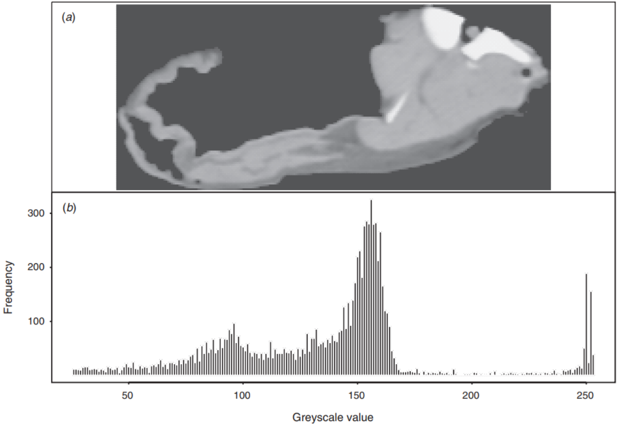
Unsupervised clustering $\leftrightarrow$ Identifying subgroups



Common approaches:

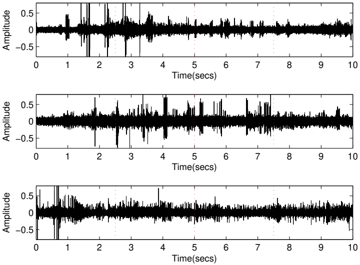
- Hierarchical clustering, K-means
- Mixture models

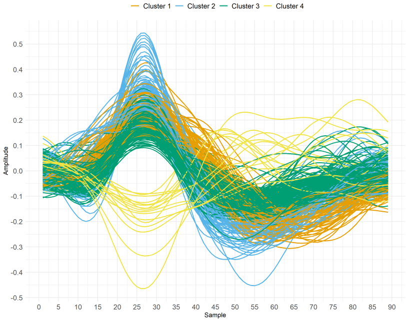
Examples of clustering using mixture models: Image classification



Alston et al (2005) DOI: 10.1071/AR04211

Examples of clustering using mixture models: Spike sorting





Mixture model ingredients

Data are drawn from a *convex combination of components*

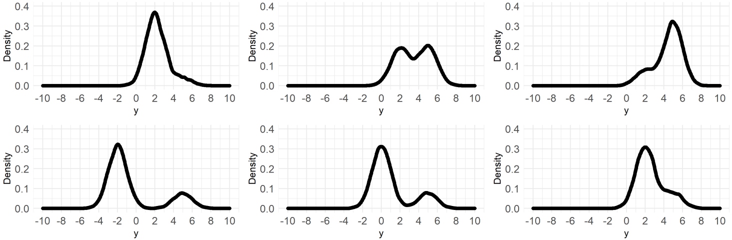
$$p(y) = \eta_1 f(y|\boldsymbol{\theta}_1) + \ldots + \eta_K f(y|\boldsymbol{\theta}_K)$$
$$= \sum_{k=1}^{K} \eta_k f(y|\boldsymbol{\theta}_k)$$

Unknown parameters: $\boldsymbol{\nu} = (\boldsymbol{\eta}, \boldsymbol{\theta})$

- $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_K)$: Mixture weights; $\sum_{k=1}^{K} \eta_k = 1$
- $f(y|\boldsymbol{\theta}_k)$: k^{th} Mixture component; same parametric family

A simple 2-component mixture model

$$y_i \sim \eta_1 \mathcal{N}(\mu_1, 1) + \eta_2 \mathcal{N}(\mu_2, 1)$$



Mixture model examples

General formulation:

$$p(y_i) = \sum_{k=1}^{K} \eta_k f(y_i|\boldsymbol{\theta}_k)$$

Latent class analysis (J items)

$$f(\mathrm{y}_i|\boldsymbol{\theta}_k) = \prod_{j=1}^{J} f(y_{ij}|\theta_{jk})$$

Latent class regression: $\eta_k \rightarrow \eta_k(\mathrm{x}_i)$

$$\eta_k(\mathrm{x}_i) = \frac{exp\left(\mathrm{x}_i^T \beta_k\right)}{\sum_{l=1}^{K} exp\left(\mathrm{x}_i^T \beta_l\right)}$$

Mixture model examples

Focus of mini-tutorial: cross-sectional data

- Finite mixture model
- Dirichlet Process mixture model
- Profile regression

Bayesian approaches to inference: Markov chain Monte Carlo (MCMC)

1. Finite mixture models
2. Dirichlet Process Mixture models
3. Profile regression

Finite mixture model: Setup

Assume:

- K is fixed *a priori*
- Each observation has a probability of belonging to components $1, \ldots, K$

Likelihood for $\boldsymbol{y} = (y_1, \ldots, y_n)$

$$p(\boldsymbol{y}) = \prod_{i=1}^{n} \sum_{k=1}^{K} \eta_k f(y_i | \boldsymbol{\theta}_k)$$

Aim is to learn $\boldsymbol{\nu} = (\boldsymbol{\eta}_{1,\ldots,K}, \boldsymbol{\theta}_{1,\ldots,K})$

Finite Mixture Model: Setup

Likelihood:

$$p(\boldsymbol{y}|\boldsymbol{\nu}) = \prod_{i=1}^{n} \sum_{k=1}^{K} \eta_k f(y_i|\boldsymbol{\theta}_k)$$

Priors:

$$(\eta_1, \ldots, \eta_K) \sim \mathcal{D}(\gamma_1, \ldots, \gamma_K)$$
$$\theta_k \sim p(\theta_k|\delta)$$

Finite Mixture Model: Setup

Example: Finite mixture of Normal distributions

$$p(\boldsymbol{y}|\boldsymbol{\nu}) = \prod_{i=1}^{n} \sum_{k=1}^{K} \eta_k \mathcal{N}(y_i|\mu_k, \sigma_k^2)$$

$$\mu_k = \mathcal{N}\left(\mu_0, \frac{\sigma_k^2}{N_0}\right)$$

$$\sigma_k^2 = \mathcal{IG}\left(c_0, C_0\right)$$

How to estimate when membership of y_i to components $1, \ldots, K$ is not known?

Finite Mixture Model: Estimation

Enter data augmentation! (Tanner and Wong, 1987; *JASA*)
The idea:

- Introduce z_i = cluster membership for y_i and treat as missing data

$$p(y_i|\boldsymbol{\nu}) = \sum_{k=1}^{K} p(y_i|z_i = k, \boldsymbol{\nu})Pr(z_i = k|\boldsymbol{\nu})$$

$$Pr(z_i = k|\boldsymbol{\nu}) = \eta_k$$

$$p(y_i|z_i = k, \boldsymbol{\nu}) = f(y_i|\theta_k)$$

- Inference on z_i provides information on clustering

Finite Mixture Model: Estimation

Inference based on the complete-data likelihood

$$\begin{aligned} p(\mathrm{y}, \mathrm{z}|\boldsymbol{\nu}) &= \prod_{k=1}^{K} \prod_{i:z_i=k} \eta_k p(y_i|\theta_k) \\ &= \prod_{k=1}^{K} \eta_k^{\sum_{i=1}^{n} \mathbb{I}(z_i=k)} \prod_{i:z_i=k} p(y_i|\theta_k) \end{aligned}$$

Finite Mixture Model: Estimation by MCMC

1. Sample $\boldsymbol{z}$ (Bayes' rule)

$$Pr(z_i = k|y_i, \boldsymbol{\nu}) = \frac{\eta_k f(y_i|\boldsymbol{\theta}_k)}{\sum_{j=1}^{K} \eta_j f(y_i|\boldsymbol{\theta}_j)}$$

$$z_i \sim MN(1, Pr(z_i = 1|y_i, \boldsymbol{\nu}), \ldots, Pr(z_i = K|y_i, \boldsymbol{\nu}))$$

2. Conditional on $\boldsymbol{z}$: Update $\eta_1, \ldots, \eta_K$

$$\eta_1, \ldots, \eta_K \sim \mathcal{D}(\delta_1 + \sum_{i=1}^{n} \mathbb{I}(z_i = 1), \ldots, \delta_K + \sum_{i=1}^{n} \mathbb{I}(z_i = K))$$

3. Conditional on $\boldsymbol{z}$: Update $\theta_1, \ldots, \theta_K$

$$\theta_k \sim p(\boldsymbol{\theta}_k|\delta) \prod_{i:z_i=k} f(y_i|\boldsymbol{\theta}_k)$$

Finite Mixture Model: Estimation by MCMC

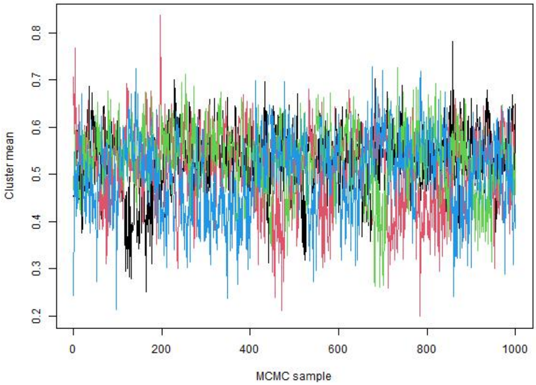
Available approaches in R:

- `R2OpenBUGS` (see `fmm_BUGS.R`)
- `bayesmix`
- Maximum likelihood: `mixtools`, `mclust`

Or code from scratch:

- `fmm_mvn.R`
- `fmm_multinomial.R`
- `fmm_pois.R`

Example MCMC output



Label switching

Issue arises as likelihood is invariant to permutations of k e.g. $K = 3$

$$
\begin{aligned}
p(y_i|\boldsymbol{\nu}) &= \eta_1 p(y_i|\boldsymbol{\theta}_1) + \eta_2 p(y_i|\boldsymbol{\theta}_2) + \eta_3 p(y_i|\boldsymbol{\theta}_3) \\
&= \eta_3 p(y_i|\boldsymbol{\theta}_3) + \eta_2 p(y_i|\boldsymbol{\theta}_2) + \eta_1 p(y_i|\boldsymbol{\theta}_1) \\
&= \eta_2 p(y_i|\boldsymbol{\theta}_2) + \eta_3 p(y_i|\boldsymbol{\theta}_3) + \eta_1 p(y_i|\boldsymbol{\theta}_1)
\end{aligned}
$$

When sampling z_i, components can be relabelled $\rightarrow$ affects posterior inference

Exploring posterior space $\leftrightarrow$ Posterior inference on components

Label switching: Possible solutions

Prior constraints (not a good idea):

$$\eta_1 < \eta_2 < \ldots < \eta_K$$

Relabelling algorithms:

- Loss functions: minimise over all MCMC samples of z (Stephens, 2000)
- MAP estimate $\widehat{z}$ as 'pivot' (Marin et. al, 2005)

Label switching: Possible solutions

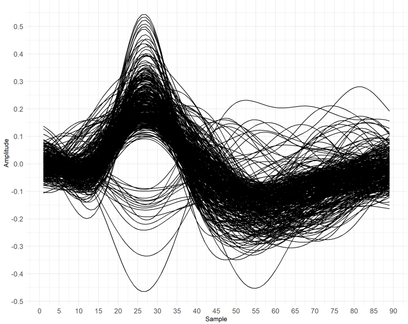
Similarity matrix, **S**:

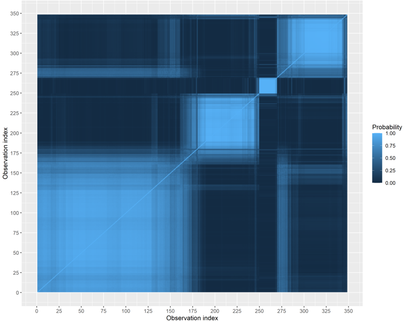
$$S_{ii'}^{(d)} = \begin{cases} 1 & \text{if } z_i^{(d)} = z_{i'}^{(d)} \\ 0 & \text{otherwise.} \end{cases}$$

$$\overline{S} = \frac{1}{D} \sum_{d=1}^{D} S^{(d)}$$

Label switching

Example: Spike sorting, $K = 4$ clusters





Loss functions to obtain likely classification z: e.g Least squares (Dahl, 2006) R packages: `mcclust`, `label.switching`

Choosing K

Common information criteria:

- Akaike's Information Criterion (AIC)

$$AIC_K = -\log p(\mathrm{y}|\boldsymbol{\eta}^*, \boldsymbol{\theta}^*) + 2p_k$$

- Bayesian Information Criterion (BIC)

$$BIC_K = -\log p(\mathrm{y}|\boldsymbol{\eta}^*, \boldsymbol{\theta}^*) + p_k \log n$$

- Deviance Information Criterion (DIC)

$$DIC_K = -4E_{\boldsymbol{\eta},\boldsymbol{\theta}|\mathrm{y}}\left[\log p(\mathrm{y}|\boldsymbol{\eta}, \boldsymbol{\theta})\right] + 2\log f(\mathrm{y})$$

$$f(\mathrm{y}) = \prod_{i=1}^{n} \frac{1}{D} \sum_{d=1}^{D} \sum_{k=1}^{K} \eta_k^{(d)} f(y_i|\theta_k^{(d)})$$

1. Finite mixture models
2. Dirichlet Process Mixture models
3. Profile regression

Dirichlet Process mixture model: Motivation

General formulation:

$$p(y_i|\boldsymbol{\nu}) = \sum_{k=1}^{K} \eta_k f(y_i|\boldsymbol{\theta}_k)$$

In a finite mixture - assume K as fixed $\rightarrow$ model comparison problem
Alternative: Infer K as part of modelling

$$p(y_i|\boldsymbol{\nu}) = \sum_{k \geq 1} \eta_k f(y_i|\boldsymbol{\theta}_k)$$

Dirichlet Process mixture model: Setup

- Nonparametric approach to mixture modelling
- Does not estimate K directly; focus on clustering of individual parameters, θ_i

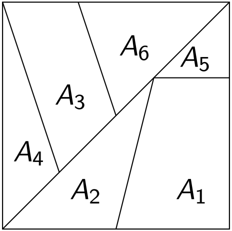
Dirichlet process (DP) prior:

$$\begin{aligned} G &\sim DP(\alpha, G_0) \\ G_0 &= E(G); \qquad \text{Base distribution} \\ \alpha &> 0; \qquad \text{Concentration parameter} \end{aligned}$$

Dirichlet Process mixture model: Setup

Each draw from a DP is itself a distribution:

$$G(A_1), \ldots, G(A_k) | \alpha, G_0 \sim D(\alpha G_0(A_1), \ldots, \alpha G_0(A_k))$$



"Discreteness property": multiple draws from $DP(\alpha, G_0)$ can take the same value; induces clustering behaviour

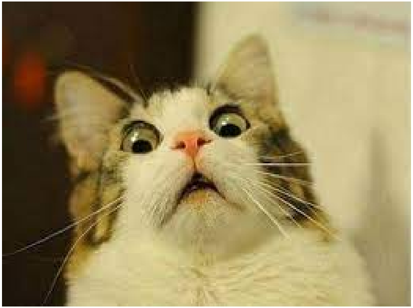
Dirichlet Process mixture model: Setup

DP as prior within mixture setting:

$$
\begin{aligned}
y_i|\boldsymbol{\theta}_i &\sim p(y_i|\boldsymbol{\theta}_i) \\
\boldsymbol{\theta}_i|G &\sim G \\
G &\sim DP(\alpha, G_0)
\end{aligned}
$$

- G is the mixing distribution
- G_0 prior distribution on unknown components
- α controls variation around G_0

Dirichlet Process mixture model: Estimation



How to sample from $DP(\alpha, G_0)$?!
What happened to η_k?!

- Stick breaking representation
- Pòlya Urn scheme/Chinese restaurant process

Stick breaking representation

G replaced by an infinite sum of weighted point masses:

$$\begin{aligned} G &\sim DP(\alpha, G_0) \\ G &= \sum_{k=1}^{\infty} \eta_k \delta_{\theta_k} \\ \theta_k &\sim G_0. \end{aligned}$$

$\eta_k; k = 1, \ldots$ are the "stick breaking weights".
Weights are drawn sequentially:

$$\begin{aligned} w_k|\alpha &\sim Beta(1, \alpha) \\ \eta_1 &= w_1 \\ \eta_k &= w_k \prod_{l=1}^{k-1}(1 - w_l). \end{aligned}$$

Stick breaking process

1

- Draw $w_1 \sim Beta(1, \alpha)$ and set $\eta_1 = w_1$

w_1	$1 - w_1$

- Draw $w_2 \sim Beta(1, \alpha)$ and compute $\eta_2 = w_2(1 - w_1)$

w_1	w_2	$(1 - w_2)(1 - w_1)$

- Draw $w_3 \sim Beta(1, \alpha)$ and compute $\eta_3 = w_3(1 - w_1)(1 - w_2)$

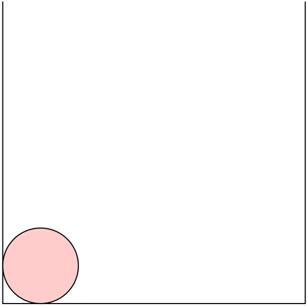
w_1	w_2	w_3	$(1 - w_3)(1 - w_2)(1 - w_1)$

As $K \to \infty$

$$p(y_i|\boldsymbol{\nu}) = \sum_{k=1}^{K} \eta_k p(y_i|\boldsymbol{\theta}_k)$$

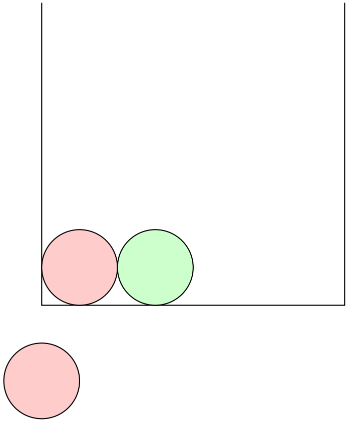
The Pòlya Urn scheme

Good analogy for implied clustering behaviour
Begin with α red balls in an urn:



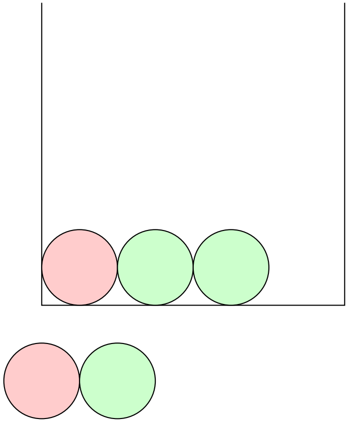
The Pòlya Urn scheme

If a red ball is drawn, record colour and replace with a ball of a new colour:

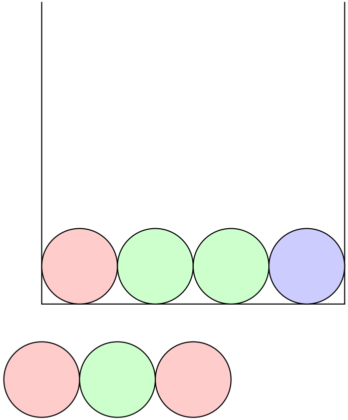


The Pòlya Urn scheme

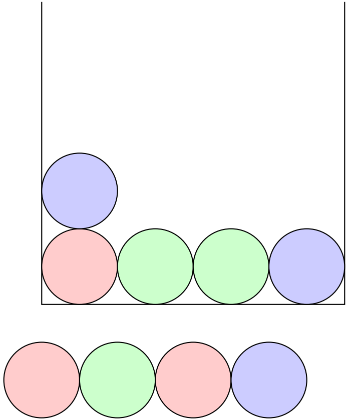
If a non-red ball is drawn, record colour and replace with a ball of the same colour:



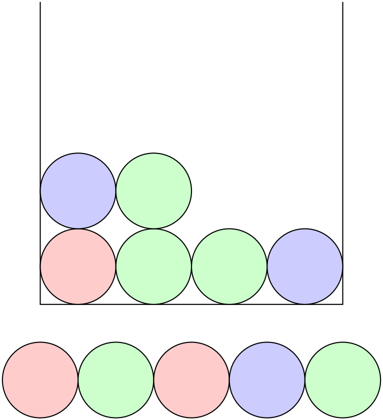
The Pòlya Urn scheme



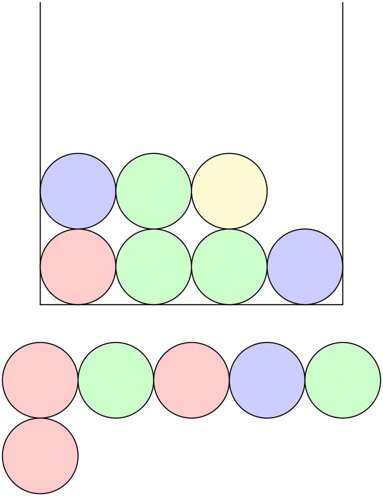
The Pòlya Urn scheme



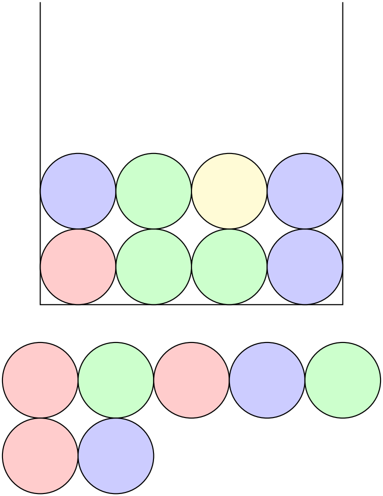
The Pòlya Urn scheme



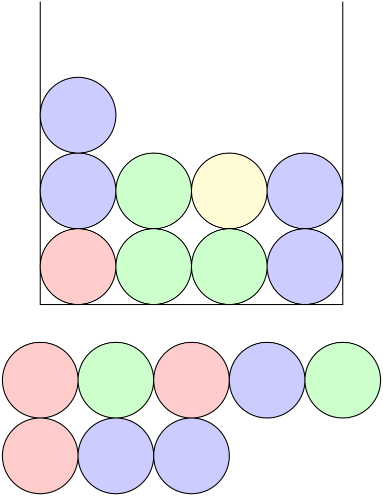
The Pòlya Urn scheme



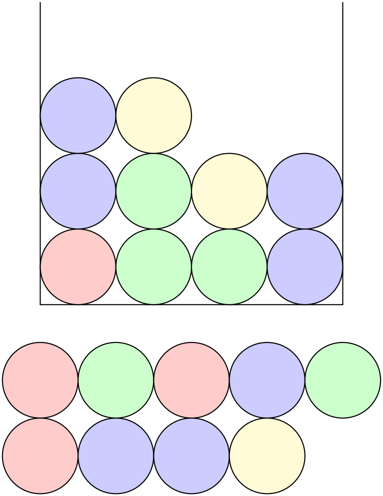
The Pòlya Urn scheme



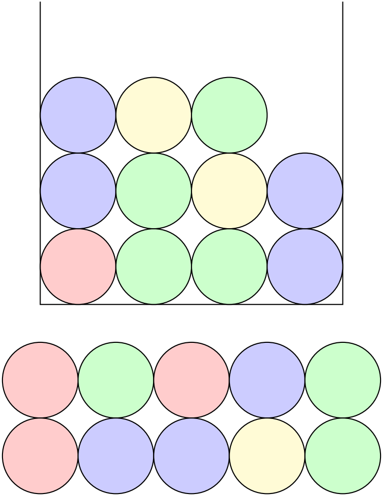
The Pòlya Urn scheme



The Pòlya Urn scheme



The Pòlya Urn scheme



The Pòlya Urn scheme

The more often a colour is drawn, the more likely it is to appear - induced clustering effect.
G is integrated out of the DP (Blackwell and MacQueen, 1973)

$$
\begin{aligned}
p(y_i|\theta_i) &\sim p(y_i) \\
\theta_1 &\sim G_0 \\
\theta_i|\theta_{i-1},\ldots,\theta_1,\alpha,G_0 &= \sum_{j=1}^{i-1}\frac{\delta_{\theta_j}}{\alpha+i-1}+\frac{\alpha G_0}{\alpha+i-1}
\end{aligned}
$$

The Pòlya Urn scheme

In terms of unobserved cluster memberships, z_1, z_2, etc

$$z_i = k | z_1, \ldots, z_{i-1}, \alpha = \begin{cases} \frac{N_k}{i-1+\alpha} & 1 \leq k \leq K \\ \frac{\alpha}{i-1+\alpha} & k = K+1. \end{cases}$$

For each observation:

- Chance of being assigned to a cluster $\propto$ number of observations already assigned (N_k)
- Always a chance of a new cluster being generated (α)

Dirichlet Process Mixture model: Role of α

α controls the level of variation about G_0

$$G \sim DP(\alpha, G_0)$$

- as $\alpha \to 0$, no clustering
- as $\alpha \to \infty$, singleton clusters

Assume α unknown and assign Gamma prior (Escobar and West, 1995)

$$\alpha \sim G(a, b)$$

Dirichlet Process Mixture model: Estimation

Stick breaking representation

- Blocked Gibbs sampler (Ishwaran and James, 2002)

$$p(\mathrm{y}|\boldsymbol{\nu}) = \prod_{i=1}^{n} \sum_{k=1}^{L} \eta_k p(y_i|\boldsymbol{\theta}_k); \qquad L >> E(K)$$

- Slice sampler (Walker, 2007); "adaptive truncation"

$$\begin{aligned} u_i &\sim \mathcal{U}(0, \eta_{z_i}) \\ \sum_{k \geq 1} \eta_k &> 1 - \min\{u_1, \ldots, u_n\} \end{aligned}$$

- R packages: `PReMiuM`
- see `spike_sorting_DPM.R`

Dirichlet Process Mixture model: Estimation

The Pòlya Urn scheme: collapsed Gibbs sampler
Each z_i is updated sequentially given current values for remaining observations, $\mathrm{z}^{(-i)}$

$$p(z_i = k|\mathrm{z}^{-i}, \mathrm{y}, \xi) = \begin{cases} \frac{N_k^{-i}}{n-1+\alpha} p(y_i|\mathrm{y}^{-i}, \mathrm{z}^{-i} = k, \xi) & 1 \leq k \leq K, \\ \frac{\alpha}{n-1+\alpha} p(y_i|\xi) & k = K+1. \end{cases}$$

$$p(y_i|\mathrm{y}^{-i}, \mathrm{z}^{-i} = k, \xi) = \int p(y_i|\theta, z_i = k) p(\theta|\mathrm{y}^{-i}, \mathrm{z}^{-i} = k, \xi)\, d\theta$$

see R package `dirichletprocess`

Dirichlet Process Mixture model: An example

Subgroup identification in Parkinson's disease based on symptom severity (White et al, 2012)

- Data collected using Unified Parkinson's Disease Rating Scale (UPDRS)
- Symptom scores: Activities of daily living (ADL), Tremor, Rigidity and Akinesia
- Responses coded as Low/Moderate/High Progression based on UPDRS score

Model specification

- For J symptoms, $y_i = (y_{i1}, \ldots, y_{iJ})$ is Multinomial with $R = 3$ levels

$$
\begin{aligned}
p(\mathrm{y}_i|\boldsymbol{\theta}_k &\sim \prod_{j=1}^{J} Mult(y_{ij}; \boldsymbol{\theta}_{jk}) \\
\boldsymbol{\theta}_{jk} &= (\theta_{jk1}, \ldots, \theta_{jkR})
\end{aligned}
$$

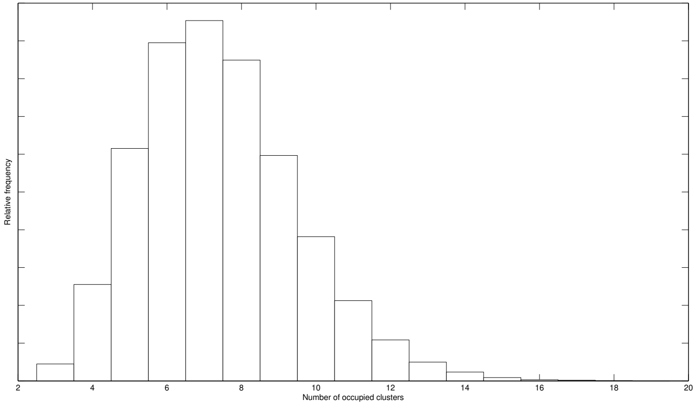
- Base distribution, G_0

$$
\begin{aligned}
G_0 &= p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J) \\
&= \prod_{j=1}^{J} D\left(\frac{1}{R}, \ldots, \frac{1}{R}\right)
\end{aligned}
$$

- Assume stick breaking representation with $\alpha \sim G(2, 2)$

Results - Clustering inference

Blocked Gibbs sampler: 100,000 iterations (50,000 burnin), and $L = 50$



Parameter inference

ADL	Cluster						
	1	2	3	4	5	6	7*
Low	0.09(0.04)	0.16(0.05)	0.85(0.06)	0.46(0.10)	0.62(0.17)	0.11(0.16)	0.17(0.21)
Moderate	0.20(0.06)	0.60(0.07)	0.11(0.05)	0.31(0.09)	0.19(0.14)	0.78(0.21)	0.66(0.27)
High	0.71(0.06)	0.24(0.06)	0.04(0.03)	0.23(0.07)	0.19(0.14)	0.11(0.18)	0.17(0.21)
Tremor	Cluster						
	1	2	3	4	5	6	7*
Low	0.10(0.04)	0.43(0.07)	0.69(0.08)	0.01(0.02)	0.05(0.07)	0.11(0.18)	0.17(0.21)
Moderate	0.26(0.06)	0.24(0.06)	0.30(0.08)	0.88(0.06)	0.62(0.17)	0.11(0.16)	0.66(0.27)
High	0.64(0.07)	0.33(0.07)	0.01(0.02)	0.11(0.06)	0.33(0.17)	0.78(0.21)	0.17(0.21)
Rigidity	Cluster						
	1	2	3	4	5	6	7*
Low	0.01(0.01)	0.52(0.07)	0.59(0.09)	0.06(0.05)	0.90(0.10)	0.44(0.25)	0.66(0.27)
Moderate	0.13(0.05)	0.30(0.06)	0.33(0.08)	0.93(0.05)	0.05(0.07)	0.12(0.16)	0.17(0.21)
High	0.86(0.05)	0.18(0.05)	0.08(0.05)	0.01(0.02)	0.05(0.07)	0.44(0.25)	0.17(0.21)
Akinesia	Cluster						
	1	2	3	4	5	6	7*
Low	0.07(0.04)	0.01(0.01)	0.95(0.04)	0.49(0.10)	0.90(0.10)	0.78(0.21)	0.17(0.21)
Moderate	0.05(0.03)	0.88(0.05)	0.04(0.03)	0.23(0.08)	0.05(0.07)	0.11(0.16)	0.17(0.21)
High	0.88(0.05)	0.11(0.04)	0.01(0.02)	0.28(0.09)	0.05(0.07)	0.11(0.16)	0.66(0.27)
η	0.30(0.04)	0.30(0.04)	0.19(0.03)	0.15(0.03)	0.04(0.01)	0.01(0.01)	0.01(0.01)

Cluster 1 - High severity

ADL	Cluster						
	1	2	3	4	5	6	7*
Low	0.09(0.04)	0.16(0.05)	0.85(0.06)	0.46(0.10)	0.62(0.17)	0.11(0.16)	0.17(0.21)
Moderate	0.20(0.06)	0.60(0.07)	0.11(0.05)	0.31(0.09)	0.19(0.14)	0.78(0.21)	0.66(0.27)
High	**0.71(0.06)**	0.24(0.06)	0.04(0.03)	0.23(0.07)	0.19(0.14)	0.11(0.18)	0.17(0.21)
Tremor	Cluster						
	1	2	3	4	5	6	7*
Low	0.10(0.04)	0.43(0.07)	0.69(0.08)	0.01(0.02)	0.05(0.07)	0.11(0.18)	0.17(0.21)
Moderate	0.26(0.06)	0.24(0.06)	0.30(0.08)	0.88(0.06)	0.62(0.17)	0.11(0.16)	0.66(0.27)
High	**0.64(0.07)**	0.33(0.07)	0.01(0.02)	0.11(0.06)	0.33(0.17)	0.78(0.21)	0.17(0.21)
Rigidity	Cluster						
	1	2	3	4	5	6	7*
Low	0.01(0.01)	0.52(0.07)	0.59(0.09)	0.06(0.05)	0.90(0.10)	0.44(0.25)	0.66(0.27)
Moderate	0.13(0.05)	0.30(0.06)	0.33(0.08)	0.93(0.05)	0.05(0.07)	0.12(0.16)	0.17(0.21)
High	**0.86(0.05)**	0.18(0.05)	0.08(0.05)	0.01(0.02)	0.05(0.07)	0.44(0.25)	0.17(0.21)
Akinesia	Cluster						
	1	2	3	4	5	6	7*
Low	0.07(0.04)	0.01(0.01)	0.95(0.04)	0.49(0.10)	0.90(0.10)	0.78(0.21)	0.17(0.21)
Moderate	0.05(0.03)	0.88(0.05)	0.04(0.03)	0.23(0.08)	0.05(0.07)	0.11(0.16)	0.17(0.21)
High	**0.88(0.05)**	0.11(0.04)	0.01(0.02)	0.28(0.09)	0.05(0.07)	0.11(0.16)	0.66(0.27)
η	**0.30(0.04)**	0.30(0.04)	0.19(0.03)	0.15(0.03)	0.04(0.01)	0.01(0.01)	0.01(0.01)

Cluster 2 - ADL/Akinesia

ADL	Cluster						
	1	2	3	4	5	6	7*
Low	0.09(0.04)	0.16(0.05)	0.85(0.06)	0.46(0.10)	0.62(0.17)	0.11(0.16)	0.17(0.21)
Moderate	0.20(0.06)	**0.60(0.07)**	0.11(0.05)	0.31(0.09)	0.19(0.14)	0.78(0.21)	0.66(0.27)
High	0.71(0.06)	0.24(0.06)	0.04(0.03)	0.23(0.07)	0.19(0.14)	0.11(0.18)	0.17(0.21)
Tremor	Cluster						
	1	2	3	4	5	6	7*
Low	0.10(0.04)	**0.43(0.07)**	0.69(0.08)	0.01(0.02)	0.05(0.07)	0.11(0.18)	0.17(0.21)
Moderate	0.26(0.06)	0.24(0.06)	0.30(0.08)	0.88(0.06)	0.62(0.17)	0.11(0.16)	0.66(0.27)
High	0.64(0.07)	0.33(0.07)	0.01(0.02)	0.11(0.06)	0.33(0.17)	0.78(0.21)	0.17(0.21)
Rigidity	Cluster						
	1	2	3	4	5	6	7*
Low	0.01(0.01)	**0.52(0.07)**	0.59(0.09)	0.06(0.05)	0.90(0.10)	0.44(0.25)	0.66(0.27)
Moderate	0.13(0.05)	0.30(0.06)	0.33(0.08)	0.93(0.05)	0.05(0.07)	0.12(0.16)	0.17(0.21)
High	0.86(0.05)	0.18(0.05)	0.08(0.05)	0.01(0.02)	0.05(0.07)	0.44(0.25)	0.17(0.21)
Akinesia	Cluster						
	1	2	3	4	5	6	7*
Low	0.07(0.04)	0.01(0.01)	0.95(0.04)	0.49(0.10)	0.90(0.10)	0.78(0.21)	0.17(0.21)
Moderate	0.05(0.03)	**0.88(0.05)**	0.04(0.03)	0.23(0.08)	0.05(0.07)	0.11(0.16)	0.17(0.21)
High	0.88(0.05)	0.11(0.04)	0.01(0.02)	0.28(0.09)	0.05(0.07)	0.11(0.16)	0.66(0.27)
η	0.30(0.04)	**0.30(0.04)**	0.19(0.03)	0.15(0.03)	0.04(0.01)	0.01(0.01)	0.01(0.01)

Cluster 3 - Low severity

ADL	Cluster						
	1	2	3	4	5	6	7*
Low	0.09(0.04)	0.16(0.05)	**0.85(0.06)**	0.46(0.10)	0.62(0.17)	0.11(0.16)	0.17(0.21)
Moderate	0.20(0.06)	0.60(0.07)	0.11(0.05)	0.31(0.09)	0.19(0.14)	0.78(0.21)	0.66(0.27)
High	0.71(0.06)	0.24(0.06)	0.04(0.03)	0.23(0.07)	0.19(0.14)	0.11(0.18)	0.17(0.21)
Tremor	Cluster						
	1	2	3	4	5	6	7*
Low	0.10(0.04)	0.43(0.07)	**0.69(0.08)**	0.01(0.02)	0.05(0.07)	0.11(0.18)	0.17(0.21)
Moderate	0.26(0.06)	0.24(0.06)	0.30(0.08)	0.88(0.06)	0.62(0.17)	0.11(0.16)	0.66(0.27)
High	0.64(0.07)	0.33(0.07)	0.01(0.02)	0.11(0.06)	0.33(0.17)	0.78(0.21)	0.17(0.21)
Rigidity	Cluster						
	1	2	3	4	5	6	7*
Low	0.01(0.01)	0.52(0.07)	**0.59(0.09)**	0.06(0.05)	0.90(0.10)	0.44(0.25)	0.66(0.27)
Moderate	0.13(0.05)	0.30(0.06)	0.33(0.08)	0.93(0.05)	0.05(0.07)	0.12(0.16)	0.17(0.21)
High	0.86(0.05)	0.18(0.05)	0.08(0.05)	0.01(0.02)	0.05(0.07)	0.44(0.25)	0.17(0.21)
Akinesia	Cluster						
	1	2	3	4	5	6	7*
Low	0.07(0.04)	0.01(0.01)	**0.95(0.04)**	0.49(0.10)	0.90(0.10)	0.78(0.21)	0.17(0.21)
Moderate	0.05(0.03)	0.88(0.05)	0.04(0.03)	0.23(0.08)	0.05(0.07)	0.11(0.16)	0.17(0.21)
High	0.88(0.05)	0.11(0.04)	0.01(0.02)	0.28(0.09)	0.05(0.07)	0.11(0.16)	0.66(0.27)
η	0.30(0.04)	0.30(0.04)	**0.19(0.03)**	0.15(0.03)	0.04(0.01)	0.01(0.01)	0.01(0.01)

Cluster 4 - Tremor/Rigidity

ADL	Cluster						
	1	2	3	4	5	6	7*
Low	0.09(0.04)	0.16(0.05)	0.85(0.06)	**0.46(0.10)**	0.62(0.17)	0.11(0.16)	0.17(0.21)
Moderate	0.20(0.06)	0.60(0.07)	0.11(0.05)	0.31(0.09)	0.19(0.14)	0.78(0.21)	0.66(0.27)
High	0.71(0.06)	0.24(0.06)	0.04(0.03)	0.23(0.07)	0.19(0.14)	0.11(0.18)	0.17(0.21)
Tremor	**Cluster**						
	1	2	3	4	5	6	7*
Low	0.10(0.04)	0.43(0.07)	0.69(0.08)	0.01(0.02)	0.05(0.07)	0.11(0.18)	0.17(0.21)
Moderate	0.26(0.06)	0.24(0.06)	0.30(0.08)	**0.88(0.06)**	0.62(0.17)	0.11(0.16)	0.66(0.27)
High	0.64(0.07)	0.33(0.07)	0.01(0.02)	0.11(0.06)	0.33(0.17)	0.78(0.21)	0.17(0.21)
Rigidity	**Cluster**						
	1	2	3	4	5	6	7*
Low	0.01(0.01)	0.52(0.07)	0.59(0.09)	0.06(0.05)	0.90(0.10)	0.44(0.25)	0.66(0.27)
Moderate	0.13(0.05)	0.30(0.06)	0.33(0.08)	**0.93(0.05)**	0.05(0.07)	0.12(0.16)	0.17(0.21)
High	0.86(0.05)	0.18(0.05)	0.08(0.05)	0.01(0.02)	0.05(0.07)	0.44(0.25)	0.17(0.21)
Akinesia	**Cluster**						
	1	2	3	4	5	6	7*
Low	0.07(0.04)	0.01(0.01)	0.95(0.04)	**0.49(0.10)**	0.90(0.10)	0.78(0.21)	0.17(0.21)
Moderate	0.05(0.03)	0.88(0.05)	0.04(0.03)	0.23(0.08)	0.05(0.07)	0.11(0.16)	0.17(0.21)
High	0.88(0.05)	0.11(0.04)	0.01(0.02)	0.28(0.09)	0.05(0.07)	0.11(0.16)	0.66(0.27)
η	0.30(0.04)	0.30(0.04)	0.19(0.03)	**0.15(0.03)**	0.04(0.01)	0.01(0.01)	0.01(0.01)

Cluster 5 to 7 - small weights

ADL	Cluster						
	1	2	3	4	5	6	7*
Low	0.09(0.04)	0.16(0.05)	0.85(0.06)	0.46(0.10)	0.62(0.17)	0.11(0.16)	0.17(0.21)
Moderate	0.20(0.06)	0.60(0.07)	0.11(0.05)	0.31(0.09)	0.19(0.14)	0.78(0.21)	0.66(0.27)
High	0.71(0.06)	0.24(0.06)	0.04(0.03)	0.23(0.07)	0.19(0.14)	0.11(0.18)	0.17(0.21)
Tremor	Cluster						
	1	2	3	4	5	6	7*
Low	0.10(0.04)	0.43(0.07)	0.69(0.08)	0.01(0.02)	0.05(0.07)	0.11(0.18)	0.17(0.21)
Moderate	0.26(0.06)	0.24(0.06)	0.30(0.08)	0.88(0.06)	0.62(0.17)	0.11(0.16)	0.66(0.27)
High	0.64(0.07)	0.33(0.07)	0.01(0.02)	0.11(0.06)	0.33(0.17)	0.78(0.21)	0.17(0.21)
Rigidity	Cluster						
	1	2	3	4	5	6	7*
Low	0.01(0.01)	0.52(0.07)	0.59(0.09)	0.06(0.05)	0.90(0.10)	0.44(0.25)	0.66(0.27)
Moderate	0.13(0.05)	0.30(0.06)	0.33(0.08)	0.93(0.05)	0.05(0.07)	0.12(0.16)	0.17(0.21)
High	0.86(0.05)	0.18(0.05)	0.08(0.05)	0.01(0.02)	0.05(0.07)	0.44(0.25)	0.17(0.21)
Akinesia	1	2	3	4	5	6	7
Low	0.07(0.04)	0.01(0.01)	0.95(0.04)	0.49(0.10)	0.90(0.10)	0.78(0.21)	0.17(0.21)
Moderate	0.05(0.03)	0.88(0.05)	0.04(0.03)	0.23(0.08)	0.05(0.07)	0.11(0.16)	0.17(0.21)
High	0.88(0.05)	0.11(0.04)	0.01(0.02)	0.28(0.09)	0.05(0.07)	0.11(0.16)	0.66(0.27)
η	0.30(0.04)	0.30(0.04)	0.19(0.03)	0.15(0.03)	**0.04(0.01)**	**0.01(0.01)**	**0.01(0.01)**

Comparison with finite mixture modelling

- DIC provided support for a $K = 4$

K	3	4	5
DIC	1248.7	1241.1	1247.9

- Identified subgroups had same characteristics as DPM Clusters 1-4
- DPM versus FMM: 1 model versus comparison of many models

1. Finite mixture models
2. Dirichlet Process Mixture models
3. Profile regression

Profile regression: Motivation

- Extends Dirichlet Processes to account for joint clustering in outcomes and covariates
- Developed as a solution for dealing with multi-collinearity within standard regression framework

Profile regression: Setup

Assignment submodel:

$$p\left(\mathrm{x}_i|\boldsymbol{\xi}\right) = \sum_{k\geq 1} \eta_k \prod_{j=1}^{J} p(x_{ij}|\phi_k)$$

Outcome submodel:

$$p\left(y_i|\boldsymbol{\theta}, z_i = k\right) = f\left(y_i|\boldsymbol{\theta}_k\right)$$

Profile regression: An example

Assessing falls in early stage Parkinson's disease (Sarini, 2018)
Assignment submodel: Disease severity profile

$$p(x_i|z_i = k) = \prod_{j=1}^{J} \mathcal{N}(x_{ij}|\mu_k, \sigma_k^2)$$

Outcome submodel: Poisson distribution

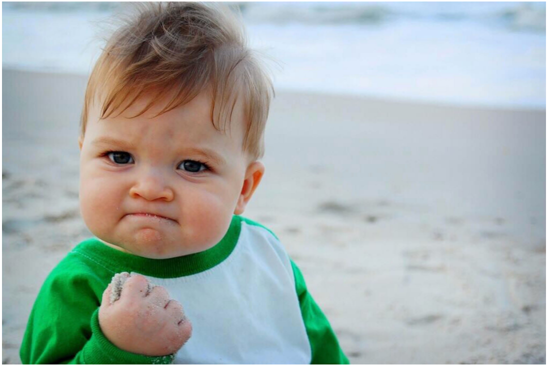
$$p(y_i|z_i = k) = Po\left(\theta_k\right))$$

Profile regression: The PReMiuM package

https://cran.r-project.org/web/packages/PReMiuM/index.html

- Outcome submodel: Continuous (Normal), binary/categorical (Multinomial), count (Poisson)
- Assignment submodel: Continuous (Normal), binary/categorical (Multinomial)
- Estimation based on stick-breaking representation
- Simulated datasets available as part of package

Done!



https://github.com/nicolemwhite/anzsc-mixture-modelling

E: nm.white@qut.edu.au Twitter: @nicolem_white