Bayesian Nonparametric Mixture Models

Yuelin Li, PhD.

Department of Psychiatry & Behavioral Sciences
Department of Epidemiology & Biostatistics
Memorial Sloan Kettering Cancer Center
New York, NY, USA

政大 E **計畫課程: 貝氏資料分析 (楊立行/李玉麟)**

December 24, 2020

Outline

Outline

Outline

Outline

Outline

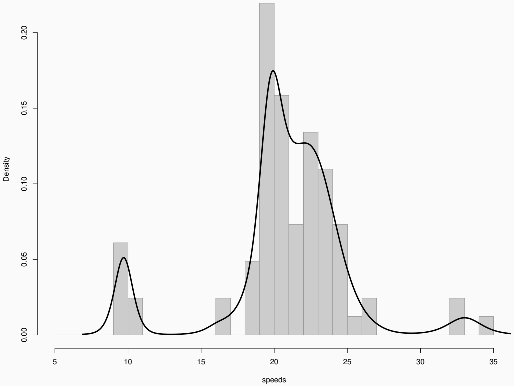
Outline

Outline

Outline

Simple Worked Examples

- `library("DPpackage")`:[1] dataset `galaxy`
- Receding velocities from 82 galaxies into 4 clusters
- Escobar & West (1995). *JASA*, 90, 577–88

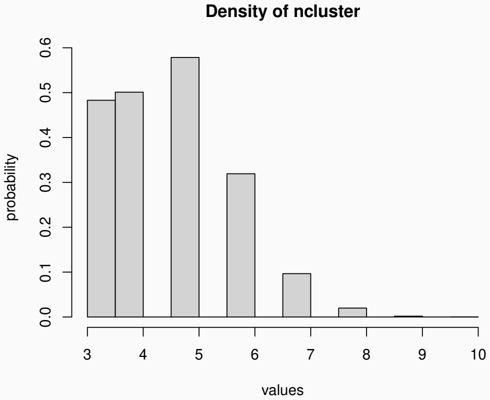



```
> # source('galaxy.R')
> library("DPpackage")
> prior2 <- list(alpha=1,m1=rep(0,1),
    psiinv2=solve(diag(0.5,1)),
    nu1=4,nu2=4,tau1=1,tau2=100)
> fit1.2 <- DPdensity(y=speeds,
    prior=prior2,mcmc=mcmc,
    state=state,status=TRUE)
> hist(speeds, probability = T,
    breaks = seq(5, 35, by =1),
    col = "grey80", border = "grey60")
> lines(cbind(fit1.2$x1,fit1.2$dens),
    lwd = 3)
```

[1] Jara et al. (2011), *J Stats Software*, 40, 1–30

Posterior distribution of K

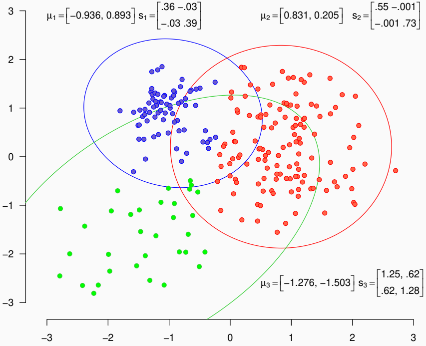
```
> plot(fit1.2, output = 'param')
```



- Mixture of up to $K = 10$ distributions

Mixture of Bivariate/Multivariate Normals

- `scikit-learn.org`: Machine Learning in Python
- Variational `BayesGaussianMixture()`[2]



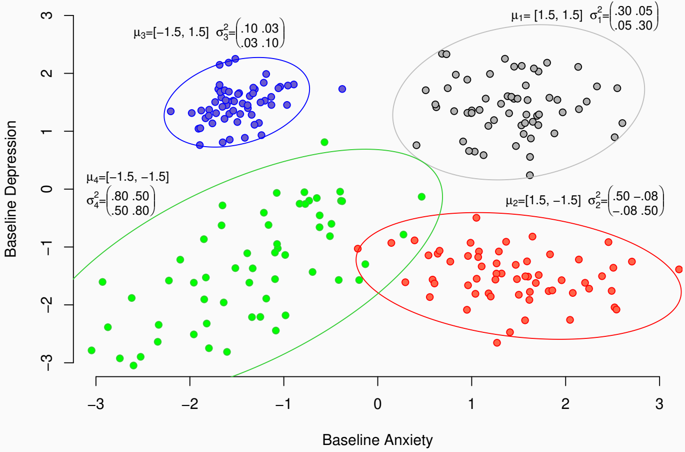

```
import numpy as np
from sklearn.mixture import BayesianGaussianMixture

K = 10
DProc = BayesianGaussianMixture(n_components=K,
  weight_concentration_prior_type="dirichlet_process",
  init_params='kmeans',
  covariance_prior = 10 * np.eye(n_cols),
  random_state=random_seed
  ).fit(csv_data)
results = DProc.predict(csv_data)
probs = DProc.predict_proba(csv_data)
print DProc.means_          # print cluster means
print DProc.covariances_    # and covariances
```

[2]scikit-learn.org/stable/modules/mixture.htm#bgmm

Theory

Hypothetical Data to Illustrate the Theory

- Hypothetical psychotherapy clinical trial for patients with advanced or terminal cancer



Conventional Clustering vs. BNP Clustering

- Conventional clustering
  - E.g., `kmeans()`, `hclust()`, or finite mixtures
  - You must specify what K is
  - Try $k = 1, 2, 3, \ldots, K$, then determine K by model comparison metrics
- Bayesian nonparametric clustering does things fundamentally differently
  - Posterior distribution of $k = 1, 2, 3, \ldots, \infty$ to a theoretically infinite number of clusters
  - A specific K may emerge out of a dataset
  - K is unbounded and may grow as the model encounters more data
  - Sometimes called an *infinite* mixture model [3]

---

[3] Rasmussen (2000).
http://papers.nips.cc/paper/1745-the-infinite-gaussian-mixture-model.pdf

- Neal (2000): one of the most cited papers on BNP[4]

$$
\begin{aligned}
y_i | c_i, \phi &\sim \mathcal{N}(\phi_k) \\
c_i | \boldsymbol{\pi} &\sim \mathrm{Discrete}(\pi_1, \pi_2, \ldots, \pi_k) \\
\phi_k &\sim G_0 \\
\boldsymbol{\pi} &\sim \mathrm{DP}(\alpha/K, \ldots, \alpha/K).
\end{aligned}
$$

[4] Neal (2000), J Compu and Graph Stats, 9(2): 249-265.

- Neal (2000): one of the most cited papers on BNP[4]

$$
\begin{aligned}
y_i | c_i, \phi &\sim \mathcal{N}(\phi_k) \\
c_i | \boldsymbol{\pi} &\sim \mathrm{Discrete}(\pi_1, \pi_2, \ldots, \pi_k) \\
\phi_k &\sim G_0 \\
\boldsymbol{\pi} &\sim \mathrm{DP}(\alpha/K, \ldots, \alpha/K). \quad \text{“Dirichlet Process Prior”}
\end{aligned}
$$

[4] Neal (2000), J Compu and Graph Stats, 9(2): 249-265.

Dirichlet Process Mixture Modeling

- Neal (2000): one of the most cited papers on BNP[4]

$$
\begin{aligned}
y_i | c_i, \phi &\sim \mathcal{N}(\phi_k) \\
c_i | \boldsymbol{\pi} &\sim \text{Discrete}(\pi_1, \pi_2, \ldots, \pi_k) \\
\phi_k &\sim G_0 \\
\boldsymbol{\pi} &\sim \text{DP}(\alpha/K, \ldots, \alpha/K). \quad \text{“Dirichlet Process Prior”}
\end{aligned}
$$

- Last two lines allow DP to spawn new clusters, look simple
  - But how? We went through many tutorial papers, books, & videos
  - Did not get very far with them
  - Details omitted entirely, assumed understood, or explained in abstract and unfamilar concepts

[4] Neal (2000), J Compu and Graph Stats, 9(2): 249-265.

BNP by Dirichlet Process Mixtures: Our How-To Guide

- We worked out these missing details for non-technical readers

Journal of Mathematical Psychology 91 (2019) 128–144

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Tutorial

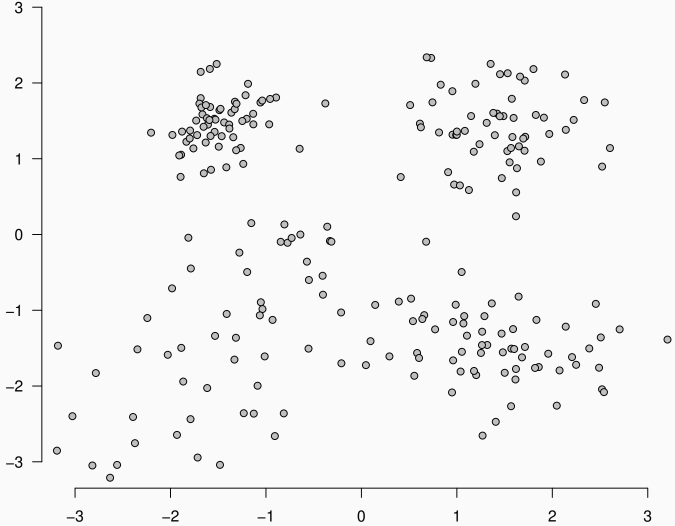
A tutorial on Dirichlet process mixture modeling

Yuelin Li [a,b,*], Elizabeth Schofield [a], Mithat Gönen [b]

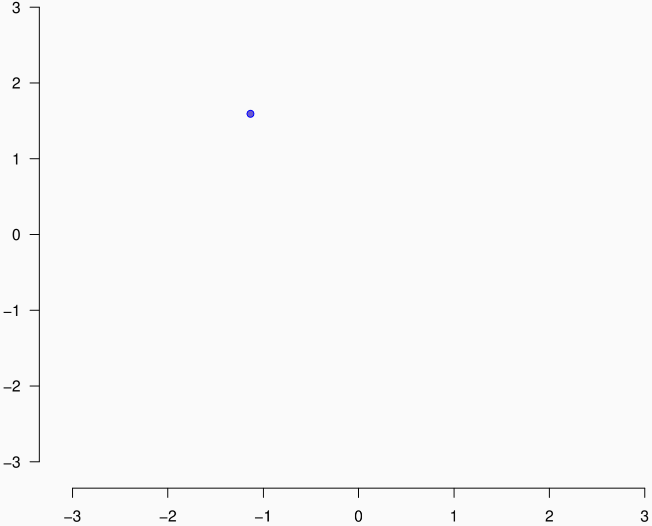
[a] *Department of Psychiatry & Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, NY 10022, USA*
[b] *Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY 10017, USA*

- Important equations carefully derived in great detail
- It may also help pave the way to other BNP methods

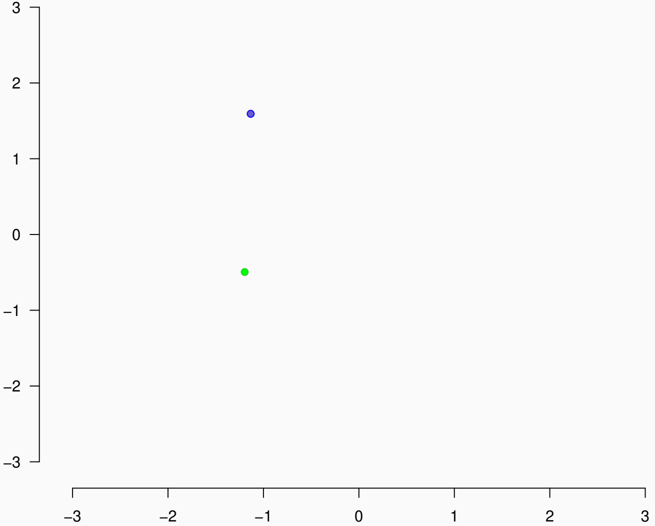
What DPMM Looks Like



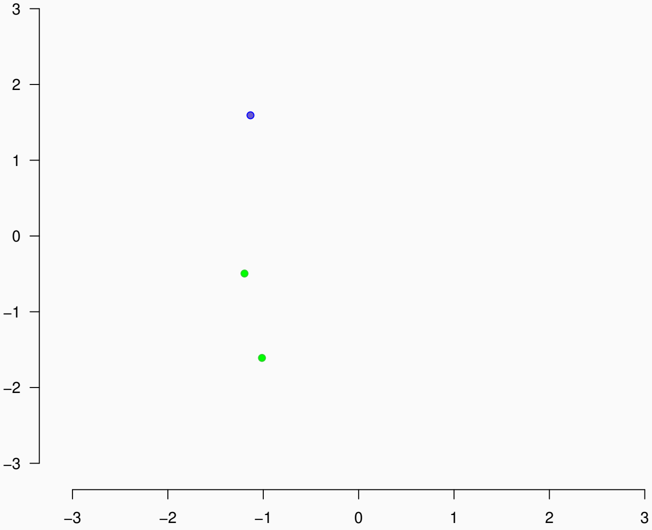
What DPMM Looks Like



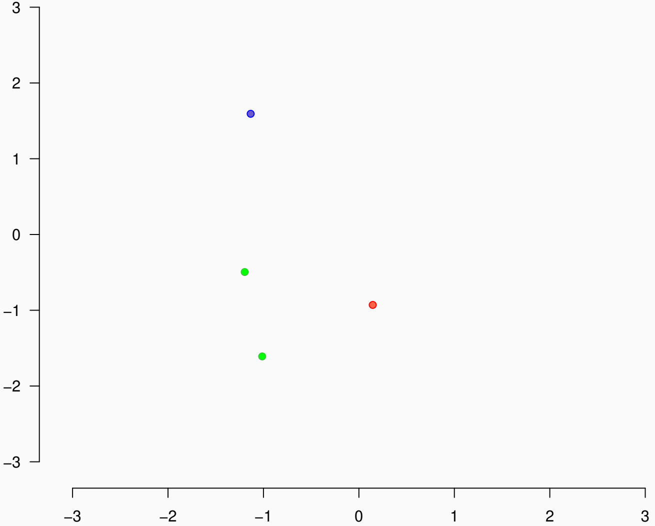
What DPMM Looks Like



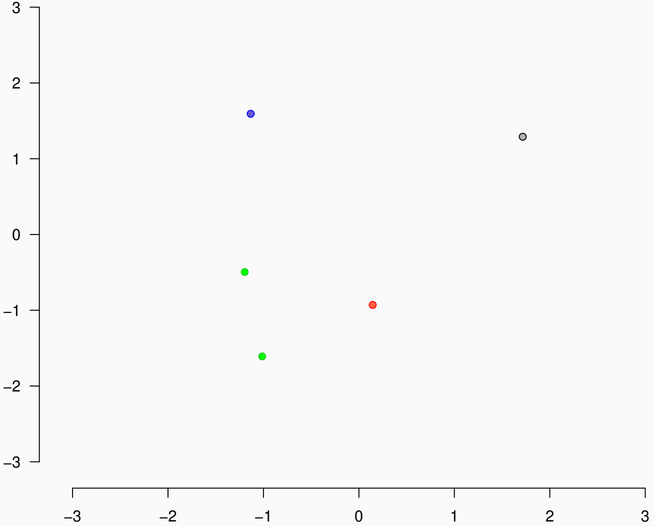
What DPMM Looks Like



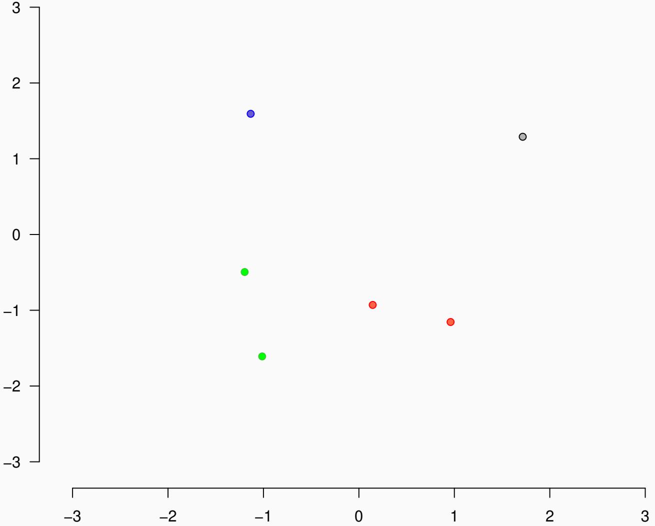
What DPMM Looks Like



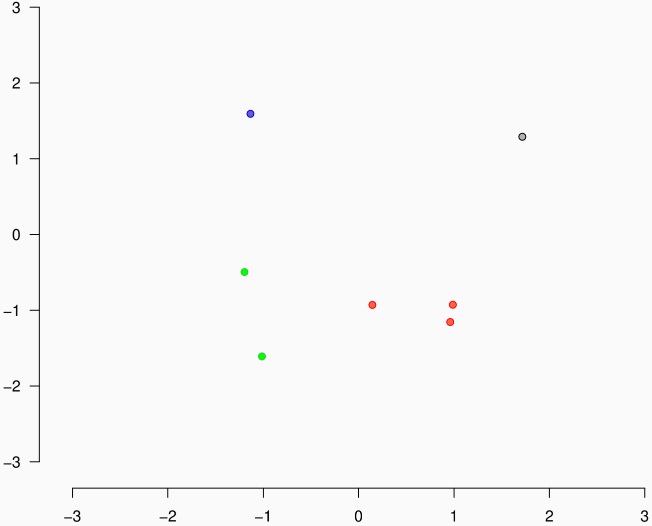
What DPMM Looks Like



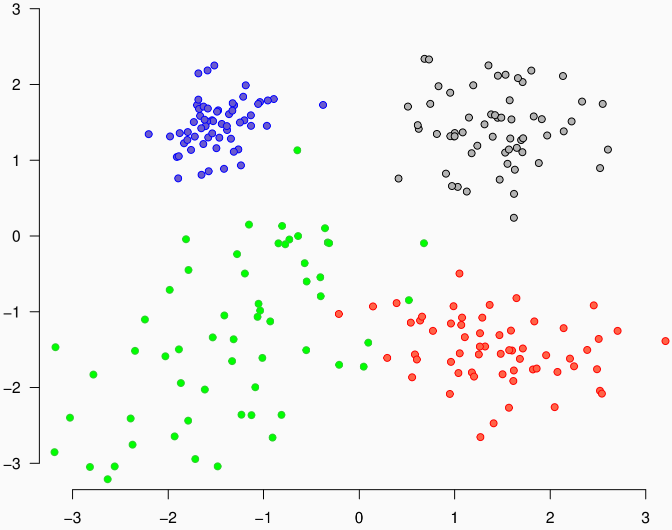
What DPMM Looks Like



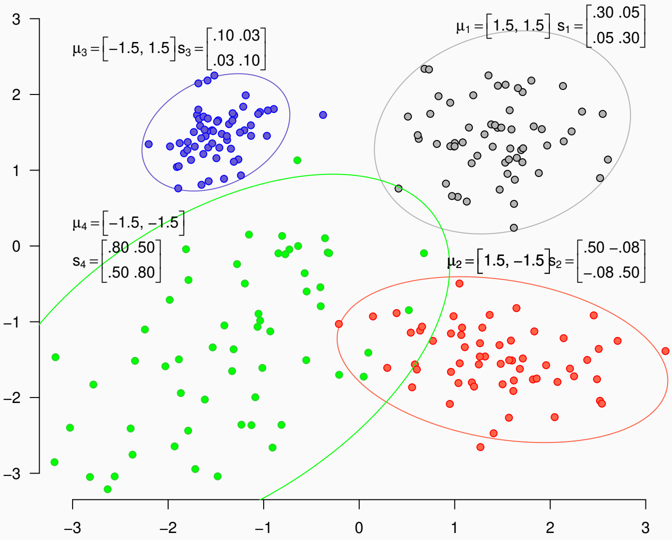
What DPMM Looks Like



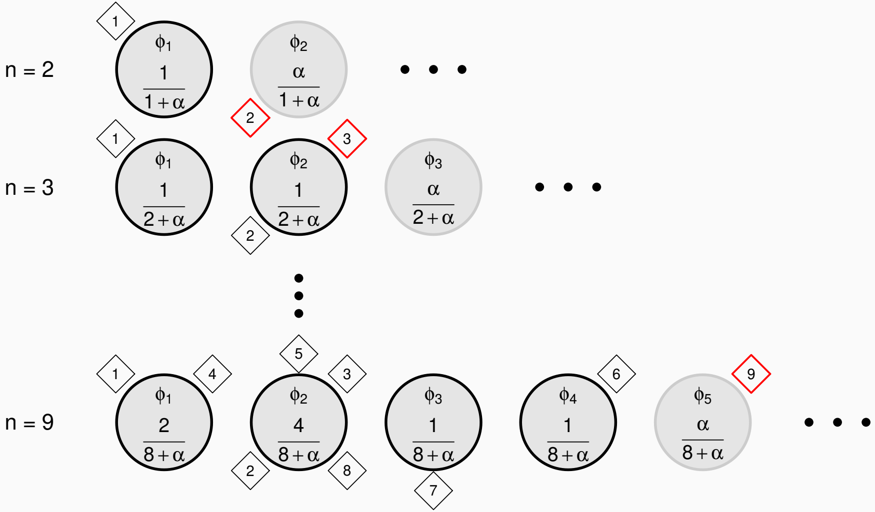
What DPMM Looks Like



What DPMM Looks Like



The Chinese Restaurant Process (CRP) [5]



[5] Gershman & Blei (2012), *J Math Psych*, 56, 1–12.

Finite Mixture Model

Data likelihood on mixture of Gaussians when k is fixed [6]

$$p(y|\mu_1, \ldots, \mu_k, s_1, \ldots, s_k, \pi_1, \ldots, \pi_k) = \sum_{j=1}^{k} \pi_j \, \mathcal{N}(\mu_j, s_j),$$

[6] Rasmussen (2000), The Infinite Gaussian Mixture Model.

Finite Mixture Model

Data likelihood on mixture of Gaussians when k is fixed [6]

$$p(y|\mu_1, \ldots, \mu_k, s_1, \ldots, s_k, \pi_1, \ldots, \pi_k) = \sum_{j=1}^{k} \pi_j \, \mathcal{N}(\mu_j, s_j),$$

Priors on cluster means and precisions:

$$p(\mu_j) \sim \mathcal{N}(\mu_0, s_0)$$

$$p_0(s_j|\gamma, \beta) \sim \mathcal{G}(\gamma, \beta) \propto s^{\gamma-1} \exp(-\beta s)$$

[6] Rasmussen (2000), The Infinite Gaussian Mixture Model.

Data likelihood on mixture of Gaussians when k is fixed [6]

$$p(y|\mu_1, \ldots, \mu_k, s_1, \ldots, s_k, \pi_1, \ldots, \pi_k) = \sum_{j=1}^{k} \pi_j \, \mathcal{N}(\mu_j, s_j),$$

Priors on cluster means and precisions:

$$p(\mu_j) \sim \mathcal{N}(\mu_0, s_0)$$

$$p_0(s_j|\gamma, \beta) \sim \mathcal{G}(\gamma, \beta) \propto s^{\gamma-1} \exp(-\beta s)$$

Prior on mixing proportions:

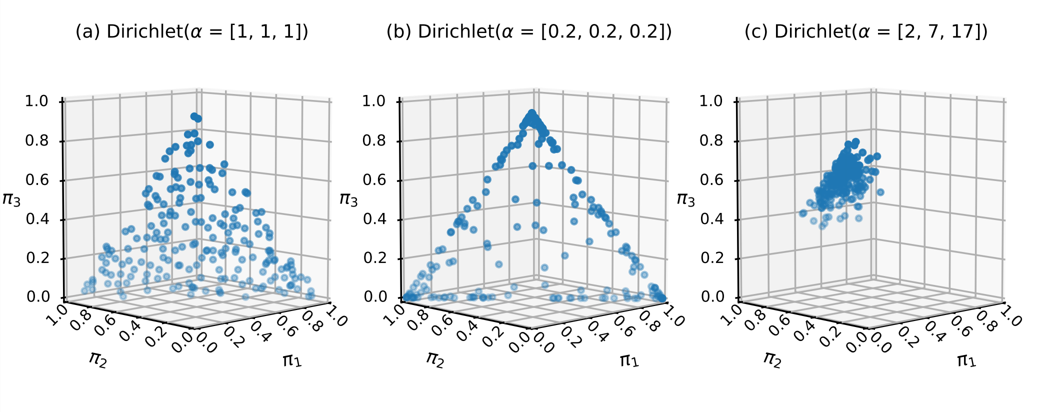
$$\begin{aligned} p(\pi_1, \ldots, \pi_k|\alpha) &\sim \mathrm{Dirichlet}(\alpha/k, \ldots, \alpha/k) \\ &= \frac{\Gamma(\alpha)}{\Gamma(\alpha/k)^k} \prod_{j=1}^{k} \pi_j^{\alpha/k-1}, \end{aligned}$$

where $\alpha/k = 1$ for a flat Dirichlet distribution for simplicity.

[6] Rasmussen (2000), The Infinite Gaussian Mixture Model.

Dirichlet Distribution

- Dirichlet distribution when $k = 3$ is fixed, e.g., $\mathrm{Dirichlet}(\alpha/k = [1, 1, 1])$ as a conjugate prior for a 3-category solution



Posteriors: [7]

$$
\begin{aligned}
p(\boldsymbol{\mu}_j|\boldsymbol{c}, \boldsymbol{y}, s_j, \mu_0, s_0) &\propto p(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{c}, s_j)p(\boldsymbol{\mu_0}) \\
&\sim \mathcal{N}\left(\frac{\bar{y}_j n_j s_j + \mu_0 s_0}{n_j s_j + s_0}, \frac{1}{n_j s_j + s_0}\right), \\
p(s_j|\boldsymbol{c}, \boldsymbol{y}, \boldsymbol{\mu}_j, \alpha, \beta) &\propto s^{n_j/2} \exp\left(-n_j \sum_{[i]\, c_j = j} (y_i - \bar{y}_j)^2\right) \times s^{\alpha-1} \exp(-\beta s) \\
&= s^{\alpha-1+n_j/2} \exp\left(-s(\beta + \frac{1}{2} \sum_{[i]\, c_j = j} (y_i - \bar{y}_j)^2)\right) \\
&= \mathcal{G}\Big(\gamma + n_j/2, \beta + \frac{\sum_{i=1}^{n_j} (y_i - \mu_j)^2}{2}\Big).
\end{aligned}
$$

---

[7] See Li, Schofield & Gönen (2019) in handouts for step by step detailed derivations.

Posteriors: [7]

$$
\begin{aligned}
p(\boldsymbol{\mu}_j|\boldsymbol{c}, \boldsymbol{y}, s_j, \mu_0, s_0) &\propto p(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{c}, s_j)p(\boldsymbol{\mu_0}) \\
&\sim \mathcal{N}\left(\frac{\bar{y}_j n_j s_j + \mu_0 s_0}{n_j s_j + s_0}, \frac{1}{n_j s_j + s_0}\right), \\
p(s_j|\boldsymbol{c}, \boldsymbol{y}, \boldsymbol{\mu}_j, \alpha, \beta) &\propto s^{n_j/2} \exp\left(-n_j \sum_{[i]\, c_j = j} (y_i - \bar{y}_j)^2\right) \times s^{\alpha-1} \exp(-\beta s) \\
&= s^{\alpha-1+n_j/2} \exp\left(-s(\beta + \frac{1}{2} \sum_{[i]\, c_j = j} (y_i - \bar{y}_j)^2)\right) \\
&= \mathcal{G}\left(\gamma + n_j/2, \beta + \frac{\sum_{i=1}^{n_j} (y_i - \mu_j)^2}{2}\right).
\end{aligned}
$$

Prior parameters updated by data likelihood when k is fixed.

[7] See Li, Schofield & Gönen (2019) in handouts for step by step detailed derivations.

Let an indicator variable c_i represent each person's latent cluster membership.

$$
\begin{aligned}
p(c_1, \ldots, c_k | \pi_1, \ldots, \pi_k) &= \prod_{j=1}^{k} \pi_j^{n_j}. \\
p(\pi_1, \ldots, \pi_k | \alpha) &\sim \mathrm{Dirichlet}(\alpha/k, \ldots, \alpha/k) = \frac{\Gamma(\alpha)}{\Gamma(\alpha/k)^k} \prod_{j=1}^{k} \pi_j^{\alpha/k-1}. \\
p(c_1, \ldots, c_k | \alpha) &= \int p(c_1, \ldots, c_k | \pi_1, \ldots, \pi_k) p(\pi_1, \ldots, \pi_k) \, d\pi_1 \cdots d\pi_k \\
&= \frac{\Gamma(\alpha)}{\Gamma(\alpha/k)^k} \int \prod_{j=1}^{k} \pi_j^{n_j+\alpha/k-1} \, d\pi_j \\
&= \frac{\Gamma(\alpha)}{\Gamma(n+\alpha)} \prod_{j=1}^{k} \frac{\Gamma(n_j+\alpha/k)}{\Gamma(\alpha/k)}.
\end{aligned}
$$

But How to Get k to Grow with Data?

Let an indicator variable c_i represent each person's latent cluster membership.

$$
\begin{aligned}
p(c_1, \ldots, c_k | \pi_1, \ldots, \pi_k) &= \prod_{j=1}^{k} \pi_j^{n_j}. \\
p(\pi_1, \ldots, \pi_k | \alpha) &\sim \text{Dirichlet}(\alpha/k, \ldots, \alpha/k) = \frac{\Gamma(\alpha)}{\Gamma(\alpha/k)^k} \prod_{j=1}^{k} \pi_j^{\alpha/k-1}. \\
p(c_1, \ldots, c_k | \alpha) &= \int p(c_1, \ldots, c_k | \pi_1, \ldots, \pi_k) p(\pi_1, \ldots, \pi_k) \, d\pi_1 \cdots d\pi_k \\
&= \frac{\Gamma(\alpha)}{\Gamma(\alpha/k)^k} \int \prod_{j=1}^{k} \pi_j^{n_j+\alpha/k-1} \, d\pi_j \\
&= \frac{\Gamma(\alpha)}{\Gamma(n+\alpha)} \prod_{j=1}^{k} \frac{\Gamma(n_j + \alpha/k)}{\Gamma(\alpha/k)}. \\
p(c_i = j | \boldsymbol{c}_{-i}, \alpha) &= \frac{n_{-i,j} + \alpha/k}{n - 1 + \alpha}.
\end{aligned}
$$

But How to Get k to Grow with Data?

Let an indicator variable c_i represent each person's latent cluster membership.

$$
\begin{aligned}
p(c_1, \ldots, c_k | \pi_1, \ldots, \pi_k) &= \prod_{j=1}^{k} \pi_j^{n_j}. \\
p(\pi_1, \ldots, \pi_k | \alpha) &\sim \text{Dirichlet}(\alpha/k, \ldots, \alpha/k) = \frac{\Gamma(\alpha)}{\Gamma(\alpha/k)^k} \prod_{j=1}^{k} \pi_j^{\alpha/k-1}. \\
p(c_1, \ldots, c_k | \alpha) &= \int p(c_1, \ldots, c_k | \pi_1, \ldots, \pi_k) p(\pi_1, \ldots, \pi_k) \, d\pi_1 \cdots d\pi_k \\
&= \frac{\Gamma(\alpha)}{\Gamma(\alpha/k)^k} \int \prod_{j=1}^{k} \pi_j^{n_j+\alpha/k-1} \, d\pi_j \\
&= \frac{\Gamma(\alpha)}{\Gamma(n+\alpha)} \prod_{j=1}^{k} \frac{\Gamma(n_j+\alpha/k)}{\Gamma(\alpha/k)}. \\
p(c_i = j | \boldsymbol{c}_{-i}, \alpha) &= \frac{n_{-i,j} + \alpha/k}{n-1+\alpha}. \iff \textit{exchangeability!}
\end{aligned}
$$

Letting $k \to \infty$

Continue from the previous slide,

$$p(c_i = j | \boldsymbol{c}_{-i}, \alpha) = \frac{n_{-i,j} + \alpha/k}{n - 1 + \alpha}, \quad \text{letting } k \to \infty,$$

$$\lim_{k \to \infty} \frac{n_{-i,j} + \alpha/k}{n - 1 + \alpha} = \frac{n_{-i,j}}{n - 1 + \alpha},$$

Letting $k \to \infty$

Continue from the previous slide,

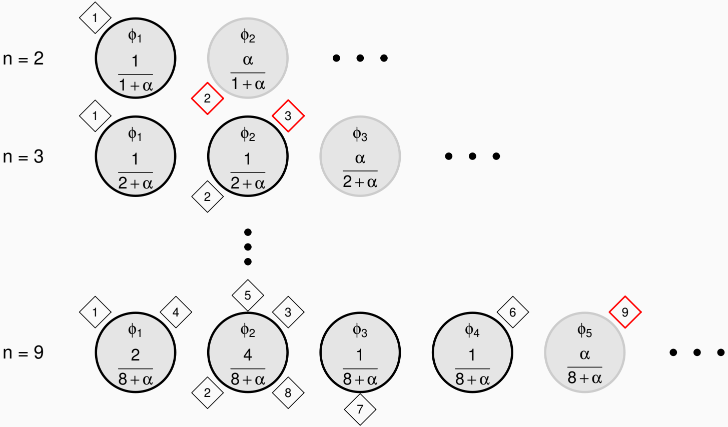
$$p(c_i = j | \boldsymbol{c}_{-i}, \alpha) = \frac{n_{-i,j} + \alpha/k}{n - 1 + \alpha}, \quad \text{letting } k \to \infty,$$

$$\lim_{k\to\infty} \frac{n_{-i,j} + \alpha/k}{n - 1 + \alpha} = \frac{n_{-i,j}}{n - 1 + \alpha},$$

$$\begin{cases} \text{already occupied clusters} : & p(c_i | \boldsymbol{c}_{-i}, \alpha) & = \dfrac{n_{-i,j}}{n - 1 + \alpha}, \\ \text{a new cluster} : & p(c_i \neq c_j \, \forall j \neq i | \boldsymbol{c}_{-i}, \alpha) & = \dfrac{\alpha}{n - 1 + \alpha}. \end{cases}$$

CRP Revisited

$$\begin{cases} \text{already occupied clusters} \; : & p(c_i | \boldsymbol{c}_{-i}, \alpha) & = \frac{n_{-i,j}}{n-1+\alpha}, \\ \text{a new cluster} \; : & p(c_i \neq c_j \, \forall j \neq i | \boldsymbol{c}_{-i}, \alpha) & = \frac{\alpha}{n-1+\alpha}. \end{cases}$$



CRP Equivalent to Stick-Breaking Process Construction [8]

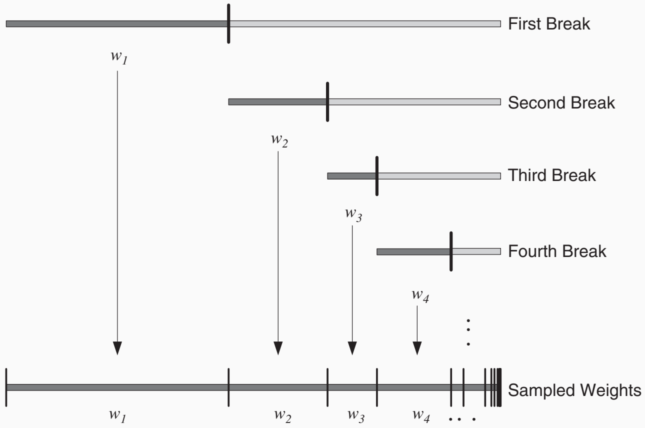


Fig. 5. A graphical depiction of the stick-breaking process, showing successive breaks of a stick with starting length one, and how the lengths of the pieces correspond to sampled weights.

[8] See Navarro et al., (2006) for a brief history and explanation

$p(c_i|\boldsymbol{c}_{-i})$ has to be weighted by the posterior probability of *newly observed* values given the data you have already seen:

$$\bullet \text{ clusters where } n_{-i,k} > 0 :$$

$$\begin{aligned} p(c_i|\boldsymbol{c}_{-i}, \mu_k, \tau_k, \alpha) &\propto p(c_i|\boldsymbol{c}_{-i}, \alpha) p(\tilde{y}_i|\mu_k, \tau_k, \boldsymbol{c}_{-i}) \\ &\propto \frac{n_{-i,k}}{n-1+\alpha} \mathcal{N}\left(\tilde{y}_i; \frac{\bar{y}_k n_k \tau_k + \mu_0 \tau_0}{n_k \tau_k + \tau_0}, \frac{1}{n_k \tau_k + \tau_0} + \sigma_y^2\right). \end{aligned}$$

$$\bullet \text{ all other clusters combined} :$$

$$\begin{aligned} &p(c_i \neq c_k \, \forall j \neq i | \boldsymbol{c}_{-i}, \mu_0, \tau_0, \gamma, \beta, \alpha) \propto \\ &\qquad p(c_i \neq c_k \, \forall j \neq i | \boldsymbol{c}_{-i}, \alpha) \int p(\tilde{y}_i|\mu_k, \tau_k) p(\mu_k, \tau_k|\mu_0, \tau_0, \gamma, \beta) \, d\mu_k d\tau_k \\ &\qquad \propto \frac{\alpha}{n-1+\alpha} \mathcal{N}(\tilde{y}_i; \mu_0, \sigma_0^2 + \sigma_y^2). \end{aligned}$$

Algorithm 1: DPMM algorithm

Let the state of the Markov chain consist of $\boldsymbol{c} = (c_1, \cdots, c_n)$ and $\boldsymbol{\phi} = (\phi_c \ : \ c \in \{c_1, \cdots, c_n\})$. Repeatedly sample:

for $i \leftarrow 1$ **to** n **do**

- Remove y_i from cluster c_i because we are going to sample a new c_i.
- draw $c_i | c_{-i}, y$ from:

if $c = c_j$ for some $j \neq i$ **then**

$p(c_i = c | c_{-i}, y_i) \propto \frac{n_{-i,c}}{n-1+\alpha} \int \mathcal{N}(\tilde{y}_i, \phi) dH_{-i,c}(\phi)$

else

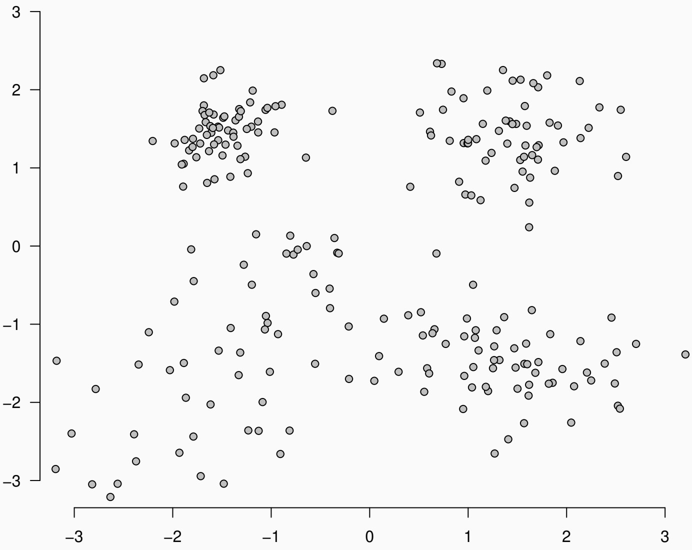
$p(c_i \neq c_j \, \forall j \neq i | c_i, y_i) \propto \frac{\alpha}{n-1+\alpha} \int \mathcal{N}(\tilde{y}_i, \phi) dG_0(\phi)$

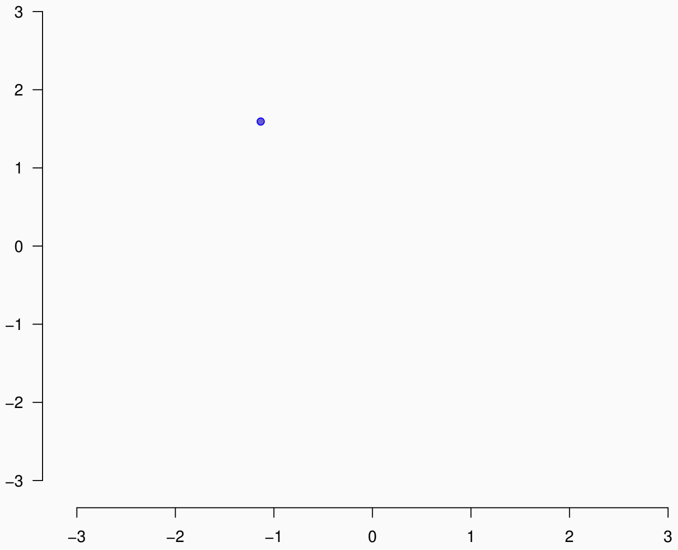
end

return c_i

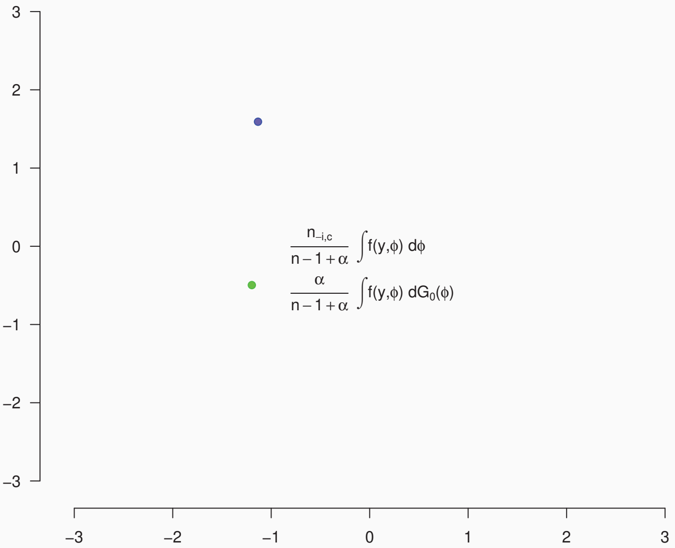
[9] Neal (2000), algorithm 3. *J Compu Graph Stats*, 9(2), 249–265.

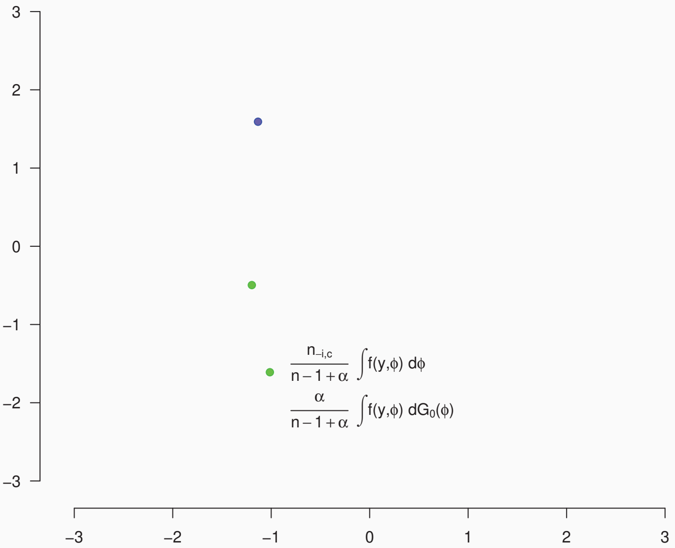
Chinese Restaurant Process Construction


3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

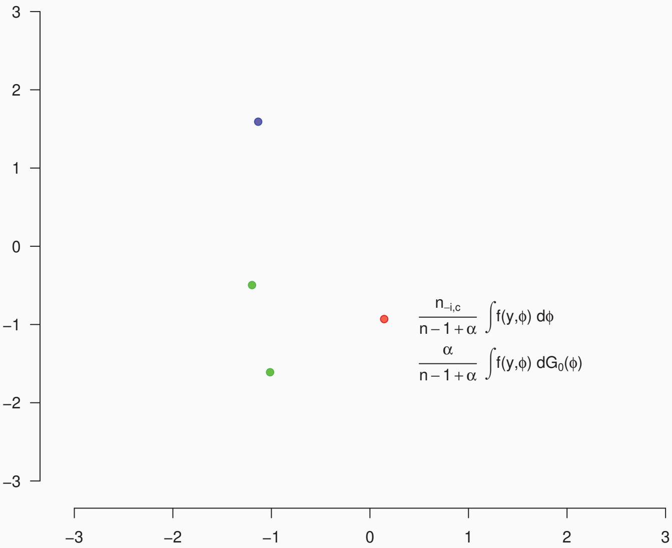


3
2
1
0
–1
–2
–3
–3
–2
–1
0
1
2
3

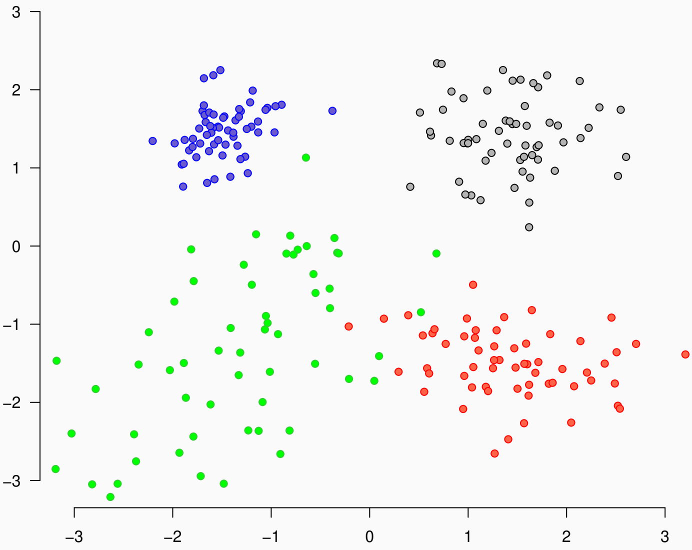

3
2
1
0
–1
–2
–3
$\frac{n_{-i,c}}{n-1+\alpha} \int f(y,\phi)\, d\phi$
$\frac{\alpha}{n-1+\alpha} \int f(y,\phi)\, dG_0(\phi)$
–3
–2
–1
0
1
2
3



3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3
$\frac{n_{-i,c}}{n-1+\alpha}\int f(y,\phi)\,d\phi$
$\frac{\alpha}{n-1+\alpha}\int f(y,\phi)\,dG_0(\phi)$



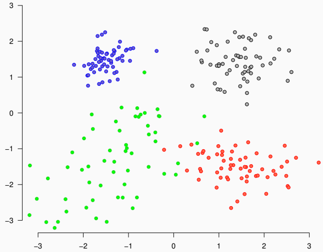
$\frac{n_{-i,c}}{n-1+\alpha} \int f(y,\phi)\, d\phi$
$\frac{\alpha}{n-1+\alpha} \int f(y,\phi)\, dG_0(\phi)$


3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

R Implementation of the CRP [10]

```
....
# already occupied table, n/(N + alpha) * mvnorm() density
log_weights[c.idx] <- log(counts[c.idx]) +
      dmvnorm(data[n, ], mean = c_mean,
              sigma = c_Sig + Sig, log = TRUE)
....
# new table, alpha/(N + alpha) weighted by mvnorm() density
log_weights[Nclust + 1] <- log(alpha) +
      dmvnorm(data[n, ], mean = mu0, sigma = Sig0 + Sig,
      log = TRUE)
....
```

[10] Modified from the R program by Tamara Broderick
https://people.csail.mit.edu/tbroderick/tutorial_2016_mlss_cadiz.html



		k-means				crp_gibbs()				Python			
		$k = 4$				$k = 10, \alpha = 1/k$				$k = 10, \alpha = 1/k$			
		●	●	●	●	●	●	●	●	●	●	●	●
Estimated Cluster	1	0	0	60	13	0	60	0	8	0	0	60	1
	2	0	60	0	4	0	0	0	36	0	1	0	55
	3	0	0	0	43	0	0	60	16	0	59	0	4
	4	60	0	0	0	60	0	0	0	60	0	0	0
		$k = 4$				$k = 5, \alpha = 1/k$				$k = 5, \alpha = 1/k$			
Estimated Cluster	1					0	0	60	18	0	0	60	1
	2	same as above				0	0	0	35	60	0	0	0
	3					0	60	0	7	0	60	0	4
	4					60	0	0	0	0	0	0	55

Why BNP in PRO Research?

Why BNP in PRO Research?

- BNP offers solutions to enduring statistical challenges in the field of symptom science
    - Identification of symptom clusters (e.g., NINR Workshop; Nho et al, (2018))
- Informing personalized, patient-centered treatment decisions
    - Narrow down intervention options most likely to be successful for individual patients
    - NIH PA-18-139: Innovative Questions in Symptom Science and Genomics

Real-World PRO Examples

Individual Meaning-Centered Psychotherapy (IMCP)

- A randomized controlled trial (R01 CA128134, PI: Breitbart)
- Breitbart, Pessin, Rosenfeld, et al. (2018) *Cancer*, 124, 3231-3239
- Patients with advanced or terminal cancer were randomized
    1. Individual Meaning Centered Psychotherapy (IMCP, $n = 109$)
    2. Supportive Psychotherapy (SP, $n = 108$)
    3. Enhanced Usual Care (EUC, $n = 104$)
- Help patients develop/increase sense of meaning near end of life

Individual Meaning-Centered Psychotherapy (IMCP)

- A randomized controlled trial (R01 CA128134, PI: Breitbart)
- Breitbart, Pessin, Rosenfeld, et al. (2018) *Cancer*, 124, 3231-3239
- Patients with advanced or terminal cancer were randomized
  1. Individual Meaning Centered Psychotherapy (IMCP, $n = 109$)
  2. Supportive Psychotherapy (SP, $n = 108$)
  3. Enhanced Usual Care (EUC, $n = 104$)
- Help patients develop/increase sense of meaning near end of life
- Psychological outcome measures
  1. Meaning Making, Hopelessness, Desire for Hastened Death, Anxiety and Depression
  2. Pre-intervention baseline, mid-intv (week 4), post-intv (week 7), and 2-months post-intv (week 15)

DP Mixtures on Baseline Psychosocial Profiles

- Before psychotherapy, do patients show different psychosocial characteristics?
- Can we cluster baseline characteristics into profiles?
- If so, do different clusters respond to IMCP differently?
- Notion of precision psychooncology
- Psychological interventions tailored to baseline profiles

Baseline Psychosocial Profiles by DPMM

- `BayesianGaussianMixture()` in Python
- Constraining $k \leq 5$ to control sparseness

Baseline psychosocial profiles identified by `BayesianGaussianMixture()`

	(n)	age	KPRS	Hopelessness	Hastened Death	Anxiety	Depression	Personal Meaning	Existential Transcendence
1	(22)	63.9	73.7	5.9	2.5	8.5	8.0	82.5	86.4
2	(17)	56.8	77.4	10.7	5.6	13.0	9.9	47.0	20.9
3	(131)	58.2	81.8	3.1	1.3	7.0	3.7	82.4	88.2
4	(10)	63.5	81.2	4.1	3.4	6.0	5.4	93.5	123.5
5	(72)	56.3	81.5	6.7	4.4	10.0	7.6	66.8	65.6

2: “Acutely Distressed” cluster

5: “Moderately Distressed” cluster

- Personal Meaning subscale scores at post-Tx

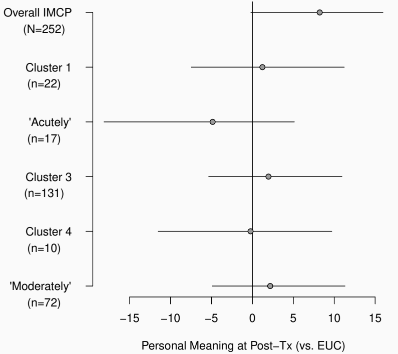
Post-Tx Personal Meaning subscale scores cf. baseline										
		Baseline		Post-Tx (week 7)						
		(N)		EUC	(n)	Meaning	(n)	Suprt	(n)
	1	(22)	82.5	78.5	6	88.8	5	83.3	11
"Acutely"	2	(17)	47.0	51.3	5	51.4	6	61.0	6
	3	(131)	82.4	81.9	46	92.0	48	87.5	37
	4	(10)	93.5	95.0	1	84.0	2	91.0	7
"Moderately"	5	(72)	66.8	71.4	16	84.1	31	75.8	25

Clusters Responded to IMCP Differently

- "random intervention effects" model [11]

$$y_{c[i]} = \beta_0 + \beta \mathrm{Tx}_{c[i]} + u_{0c} + u_{1c}\mathrm{Tx}_{c[i]} + \epsilon_{c[i]}, \qquad [u_{0c}, u_{1c}] \sim \mathcal{N}(0, \Sigma).$$

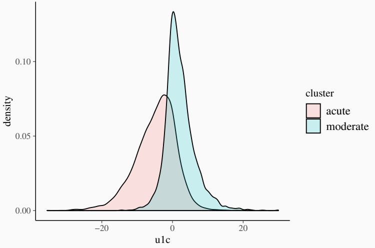
- `stan_lmer(PersMeaningT3 ~ Tx + (1 + Tx | clus), prior = NULL)`



[11] Lee & Thompson (2005), *Clin Trials*, 2(2), 163-73.

Probability of Intervention Fit

- Posterior $\Pr(u_{1c_i} > u_{1c_j} | c_i \neq c_j)$



clusters	1	2	3	4	5
1	-	0.81			
“Acutely” 2	0.19	-			
3	0.58	0.84	-		
4	0.42	0.74	0.34	-	
“Moderately” 5	0.58	0.88	0.52	0.65	-

Example 1: IMCP Summary

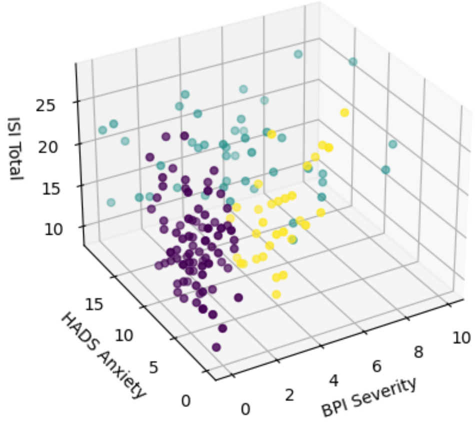
- Five baseline psychosocial profiles before therapy
- Most (n=131) appeared to be coping relatively well
- Few (n=17) appeared to be acutely distressed
- Some (n=72) moderately distressed but with meaning
- Heterogeneity of Treatment Effects (HTE)
- IMCP works better in moderately than acutely distressed
- Personalized interventions algorithm
    - We can derive probability of latent class, and
    - Probability of treatment response given latent class

Real-World Example 2: CHOICE Study

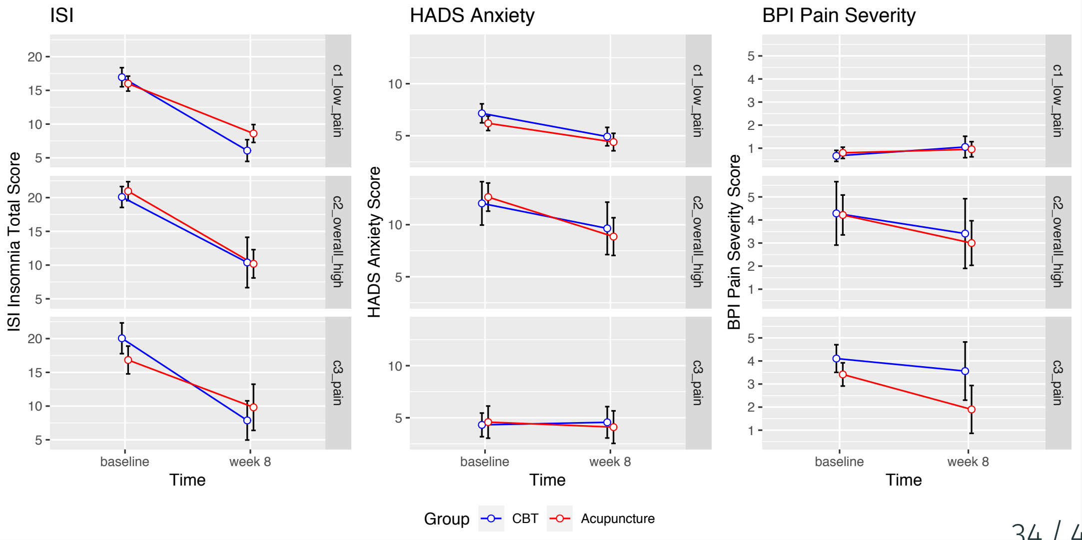
- A Comparative Effectiveness Trial (PCORI CER-1403-14292-IC, PI: Mao)
- Cancer survivors with Insomnia Severity Index > 7 (‘mild’), randomized
  1. Cognitive Behavioral Therapy for Insomnia (CBTI, $n = 79$)
  2. Acupuncture (Acupuncture, $n = 80$)
- Overall, Garland & Mao et al. (2019) reported that CBTI was better than acupuncture on insomnia

Real-World Example 2: CHOICE Study

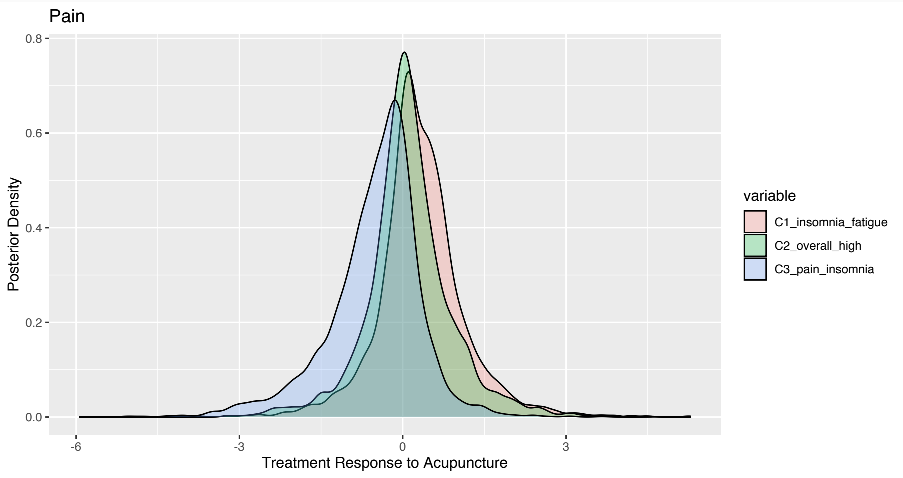
- A Comparative Effectiveness Trial (PCORI CER-1403-14292-IC, PI: Mao)
- Cancer survivors with Insomnia Severity Index > 7 ('mild'), randomized
  1. Cognitive Behavioral Therapy for Insomnia (CBTI, $n = 79$)
  2. Acupuncture (Acupuncture, $n = 80$)
- Overall, Garland & Mao et al. (2019) reported that CBTI was better than acupuncture on insomnia
- However, were there symptom clusters that responded differently?
  1. We applied DPMM to baseline insomnia, fatigue, pain, and anxiety
  2. Pre-intervention baseline and post-intv (week 8)
  3. PCORI Methdology Standards: 5. Heterogeneity of Treatment Effects (HTE)


ISI Total
25
20
15
10
HADS Anxiety
15
10
5
0
BPI Severity
0
2
4
6
8
10

Which is Better and for What Symptoms?



Probability of Intervention Success



- $p(c_3 < c_1) = 0.87 \quad p(c_3 < c_2) = 0.79 \quad p(c_2 < c_1) = 0.64$

R Programs Explained Line by Line

R Programs Explained Line by Line

Key Features of BNP

Key Features of BNP

- Number of profiles should be able to grow as the model encounters more data,
- Model should introduce no more profiles than are necessary to explain the data
- Identify the subset of individuals who showed subtle and nuanced differences

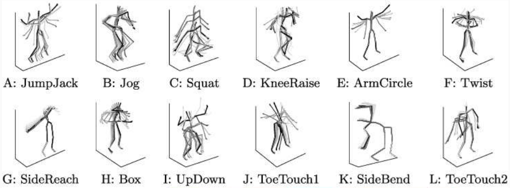
Numerous Other Applications of BNP and Future Directions (end at 10:00 CDT)

Bayesian Nonparametric (BNP) Methods

- Bayesian “nonparametric” methods?
- Not finite parameters (growing/infinite number of parameters)

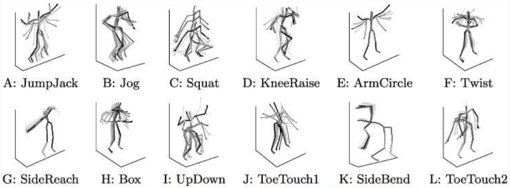
Bayesian Nonparametric (BNP) Methods

- Bayesian “nonparametric” methods?
- Not finite parameters (growing/infinite number of parameters)
- Motion capture data (`https://arxiv.org/pdf/1308.4747.pdf`)



Bayesian Nonparametric (BNP) Methods

- Bayesian "nonparametric" methods?
- Not finite parameters (growing/infinite number of parameters)
- Motion capture data (https://arxiv.org/pdf/1308.4747.pdf)



- Protein interactome data (Lloyd et al, *NIPS* 2012)

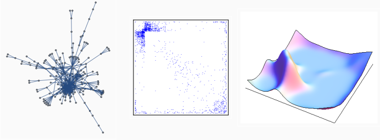
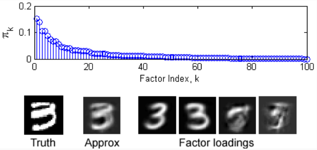


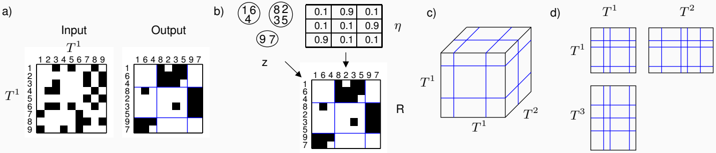
Figure 2: Protein interactome data. *Left:* Interactome network. *Middle:* Sorted adjacency matrix. The network exhibits stochastic equivalence (visible as block structure in the matrix) and homophily (concentration of points around the diagonal). *Right:* Maximum a posteriori estimate of the function Θ, corresponding to the function in Fig. 1 (middle).

Bayesian Nonparametric (BNP) Methods

- BNP as the engine for versatile analytic tools, e.g.,
  - Paisley et al. (2009): BNP factor analysis



  - Kemp et al. (2006): Infinite Relational Model



Bayesian Nonparametric (BNP) Methods

- Beyond infinite clusters, e.g.,
  - Navarro et al. (2006): web browsing habits
  - Karabatsos & Walker (2009): BNP Test Equating
  - Tanenbaum et al. (2011): “How to Grow a Mind”
  - Austerweil & Griffiths (2013): feature learning/representation
  - De Iorio et al. (2009) Biometrics, 65, 762–771.
  - Graziani et al. (2015) Biometrics, 71, 188–197.

Vast Literature on BNP

- Countless other points on BNP not covered here
- But I hope I have given you a fundamental intuition on BNP using the DP
- Make you feel more confident and efficient in learning more on your own
- Potentially rediscover finer details of BNP yourself

Thanks

- Funding
  - NIH P30 CA008748 for MSKCC
  - NIH R01 CA128134 (PI: Breitbart)
- DP mixture clustering computer program in R
  - Tamara Broderick at MIT[12]

[12] https://people.csail.mit.edu/tbroderick/tutorial_2016_mlss_cadiz.html