Understanding Bayesian Nonparametric Methods Through Programming a Dirichlet Process Mixture Model

Yuelin Li, PhD.[a, b] and Elizabeth Schofield, MS.[a]

[a]: Department of Psychiatry & Behavioral Sciences
[b]: Department of Epidemiology & Biostatistics

2018 UConn Modern Modeling Methods (M3) Conference

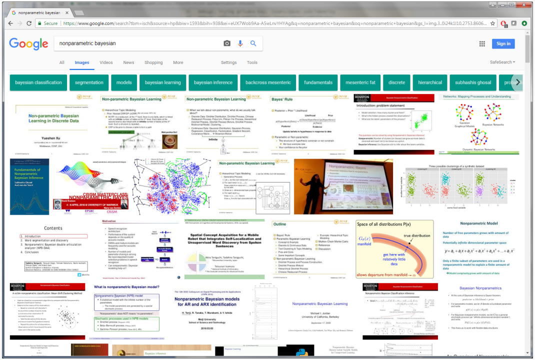
Tuesday May 22, 2018

Bayesian Nonparametric (BNP) Methods

- Bayesian "nonparametric" methods?
- But, what about $p(\theta|y) \propto L(\theta|y)p(\theta)$?
- Not finite parameters (growing/infinite number of parameters)

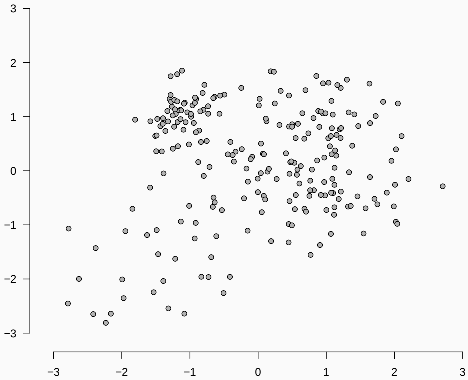
Bayesian Nonparametric (BNP) Methods

- Bayesian “nonparametric” methods?
- But, what about $p(\theta|y) \propto L(\theta|y)p(\theta)$?
- Not finite parameters (growing/infinite number of parameters)



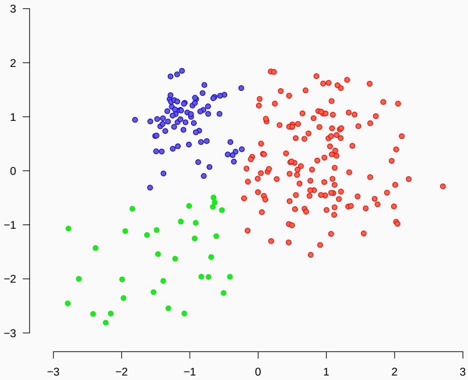
Bayesian Nonparametric (BNP) Methods

- Where to start for a beginner?
- Dirichlet Process Mixture Model (DPMM)



Bayesian Nonparametric (BNP) Methods

- Where to start for a beginner?
- Dirichlet Process Mixture Model (DPMM)



- Number of clusters grows when necessary

Outline

- What BNP looks like
- **Examples**: DPMM as an illustrative example
  1. `DPpackage` in R: Mixture of univariate Gaussians
  2. `scikit-learn` in Python: multivariate Gaussians
- **Computation**: DPMM by Chinese Restaurant Process
  1. How to program DPMM in R
- **Theory**: so that you can develop new methods
- **Real Example**: responders to end-of-life psychotherapy

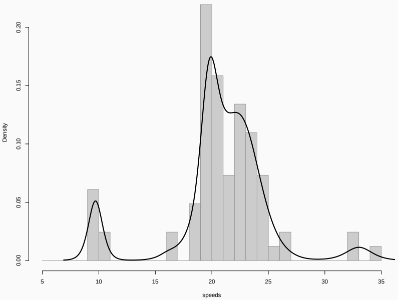
Examples:

`DPpackage` in R

`scikit-learn.org` in Python

Mixture of Univariate Normals

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

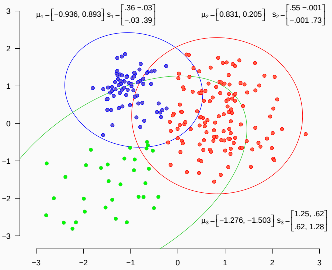



```
> library("DPpackage")
> prior2 <- list(alpha=1,m1=rep(0,1),
    psiinv2=solve(diag(0.5,1)),
    nu1=4,nu2=4,tau1=1,tau2=100)
> fit1.2 <- DPdensity(y=speeds,
    prior=prior2,mcmc=mcmc,
    state=state,status=TRUE)
> hist(speeds, probability = T,
    breaks = seq(5, 35, by =1),
    col = "grey80", border = "grey60")
> lines(cbind(fit1.2$x1,fit1.2$dens),
    lwd = 3)
```

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

Mixture of Bivariate Normals

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



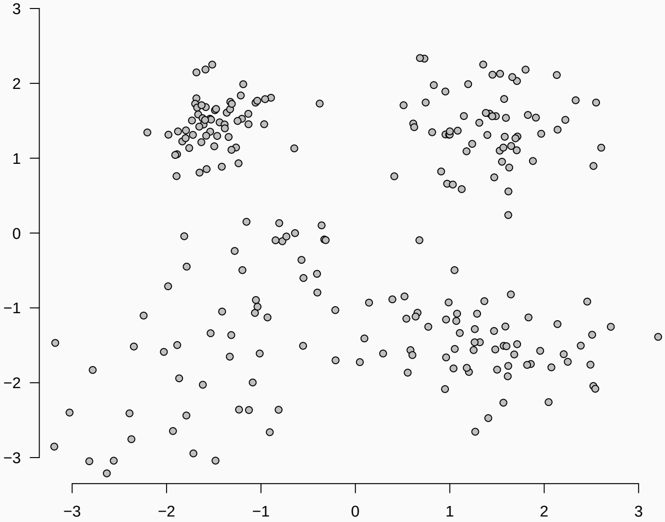

```
import numpy as np
from sklearn.mixture import BayesianGaussianMixture

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

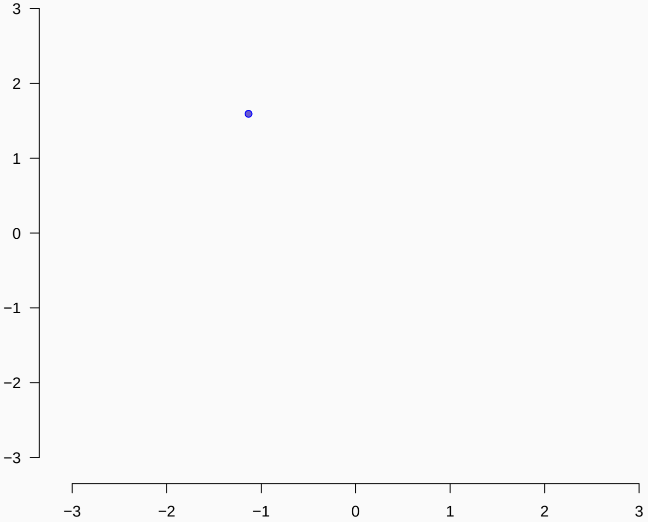
Computation:

What makes DP spawn new clusters when necessary?

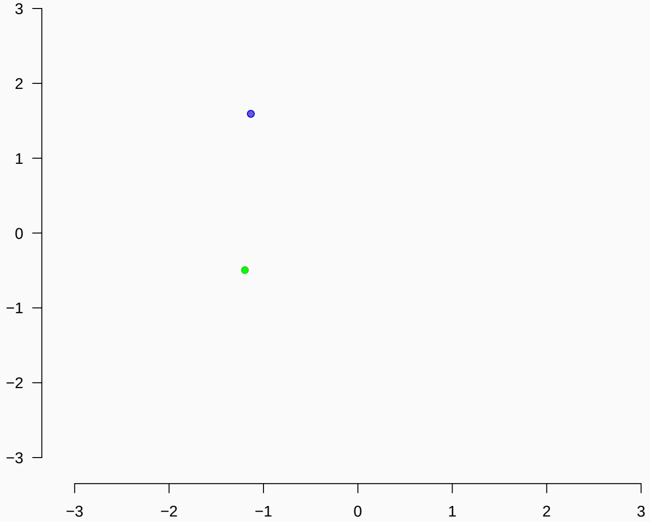
What DPMM Looks Like



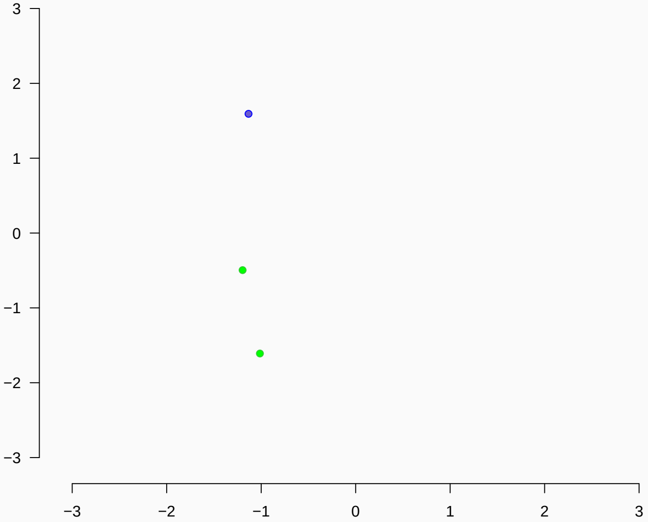
What DPMM Looks Like



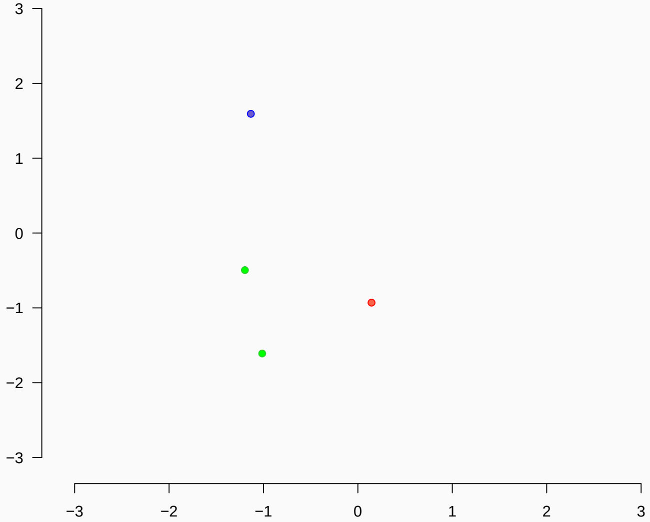
What DPMM Looks Like



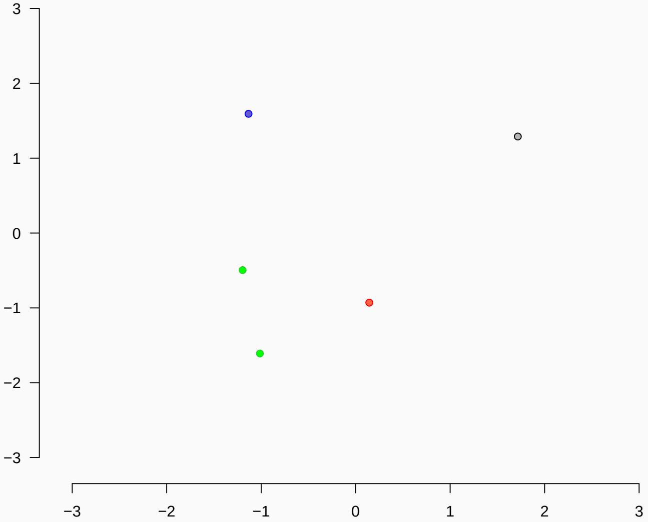
What DPMM Looks Like



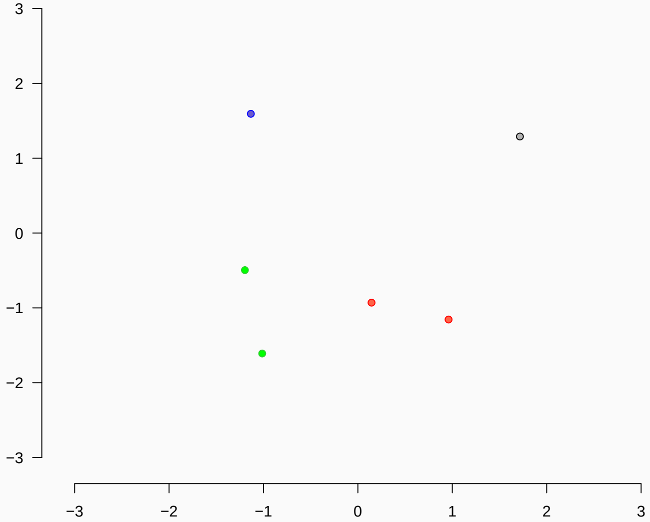
What DPMM Looks Like



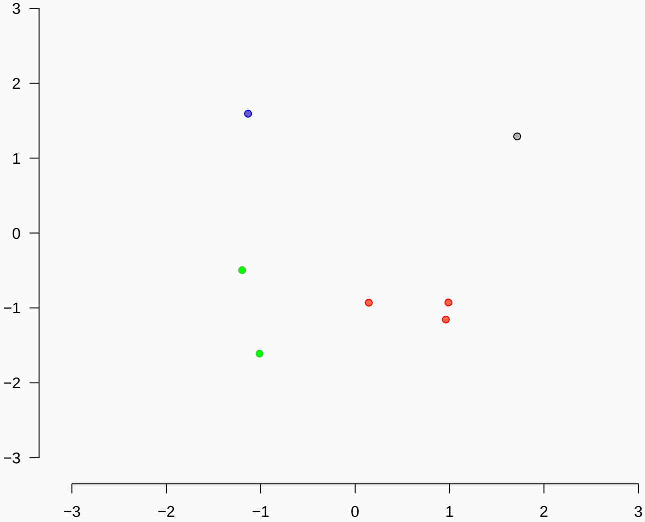
What DPMM Looks Like



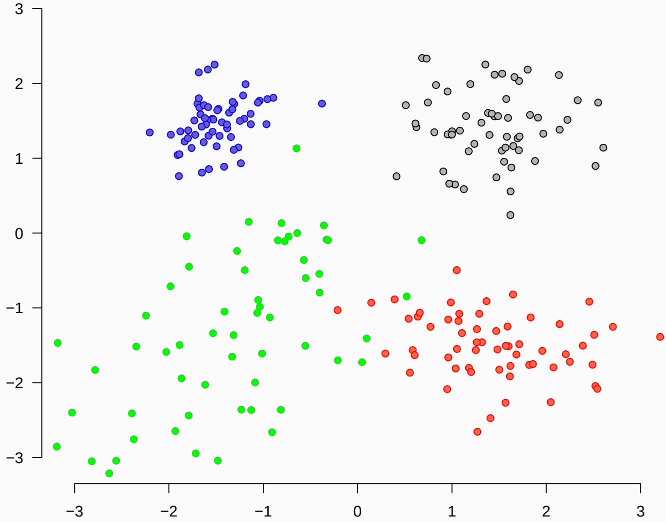
What DPMM Looks Like



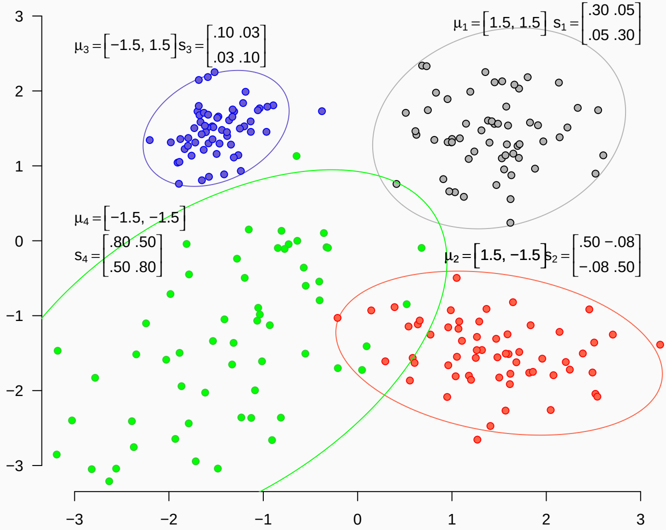
What DPMM Looks Like



What DPMM Looks Like

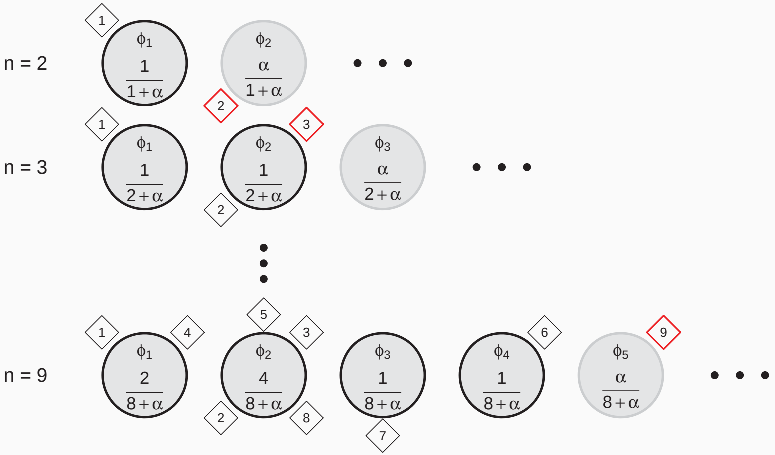


What DPMM Looks Like



DPMM Constructed from Chinese Restaurant Process (CRP)

- Pitman & Dubins (2002): Chinese Restaurant Process[3]
- Gershman & Blei (2012)



[3] UC Berkeley Dept Stats Tech Report 621

Theory

Finite Mixture Model

Mixture of Gaussians when k is fixed [4]

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

[4] Rasmussen (2000), The Infinite Gaussian Mixture Model. Advances in Neural Information Processing Systems 12. 554–560.

Mixture of Gaussians when k is fixed [4]

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

We introduce an indicator variable c_i representing each person's latent cluster membership. If $c_i = k$,

$$p(y_i|c_i = k) = \mathcal{N}(y_i|\mu_j, s_j^{-1}),$$

$$\text{with priors}$$

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

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

[4] Rasmussen (2000), The Infinite Gaussian Mixture Model. Advances in Neural Information Processing Systems 12. 554–560.

Calculate Class Membership [5]

$$p(c_1, \ldots, c_k | \pi_1, \ldots, \pi_k) = \prod_{j=1}^{k} \pi_j^{n_j}.$$

$$p(\pi_1, \ldots, \pi_k | \alpha) \sim \text{Dirichlet}(\alpha/k, \ldots, \alpha/k) = \frac{\Gamma(\alpha)}{\Gamma(\alpha/k)^k} \prod_{j=1}^{k} \pi_j^{\alpha/k-1}.$$

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

[5] Li, Schofield & Gonen, in preparation.

Calculate Class Membership [5]

$$p(c_1, \ldots, c_k | \pi_1, \ldots, \pi_k) = \prod_{j=1}^{k} \pi_j^{n_j}.$$

$$p(\pi_1, \ldots, \pi_k | \alpha) \sim \text{Dirichlet}(\alpha/k, \ldots, \alpha/k) = \frac{\Gamma(\alpha)}{\Gamma(\alpha/k)^k} \prod_{j=1}^{k} \pi_j^{\alpha/k-1}.$$

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

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

[5] Li, Schofield & Gonen, in preparation.

Calculate Class Membership [5]

$$p(c_1, \ldots, c_k | \pi_1, \ldots, \pi_k) = \prod_{j=1}^{k} \pi_j^{n_j}.$$

$$p(\pi_1, \ldots, \pi_k | \alpha) \sim \text{Dirichlet}(\alpha/k, \ldots, \alpha/k) = \frac{\Gamma(\alpha)}{\Gamma(\alpha/k)^k} \prod_{j=1}^{k} \pi_j^{\alpha/k-1}.$$

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

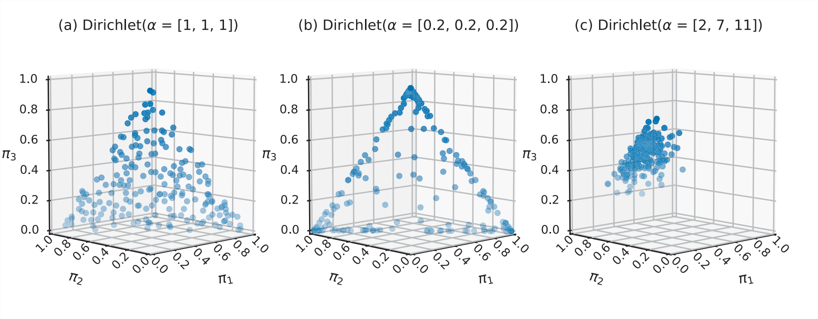
$$p(c_i = j | \boldsymbol{c}_{-i}, \alpha) = \frac{n_{-i,j} + \alpha/k}{n-1+\alpha}. \Longleftrightarrow \textit{exchangeability!}$$

[5] Li, Schofield & Gonen, in preparation.

Dirichlet Distribution

- A *symmetric* Dirichlet distribution (uniform prior)

$$\text{Dirichlet}(\alpha/k, \ldots, \alpha/k) = \text{Dirichlet}(\alpha/k = [1, 1, 1]).$$



Finite Mixture to Infinite Mixture of DPMM

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

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

Finite Mixture to Infinite Mixture of DPMM

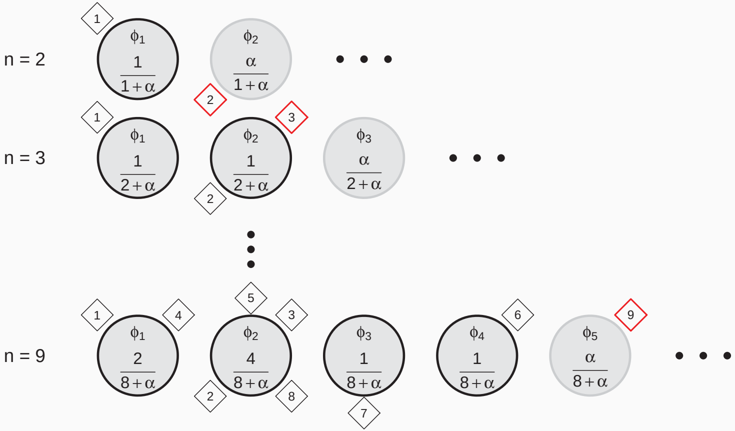
$$p(c_i = j | \boldsymbol{c}_{-i}, \alpha) = \frac{n_{-i,j} + \alpha/k}{n - 1 + \alpha}, \quad \text{letting } k \to \infty,$$

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

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

CRP Revisited

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



Conditional Posterior Distribution of c_i

$p(c_i|\boldsymbol{c}_{-i})$ has to be weighted by $\mathcal{N}(\bar{y}_j, s_j + s_0)$, the posterior probability of *newly observed* values given the data you have already seen:

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

$$\begin{aligned} p(c_i|\boldsymbol{c}_{-i}, \mu_j, s_j, \alpha) &\propto p(c_i|\boldsymbol{c}_{-i}, \alpha) p(\tilde{y}_i|\mu_j, s_j, \boldsymbol{c}_{-i}) \\ &\propto \frac{n_{-i,j}}{n-1+\alpha} \mathcal{N}(\bar{y}_j, s_j + s_0), \end{aligned}$$

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

$$\begin{aligned} & p(c_i \neq c_j \, \forall j \neq i | \boldsymbol{c}_{-i}, \mu_0, s_0, \gamma, \beta, \alpha) \propto \\ & p(c_i \neq c_j \, \forall j \neq i | \boldsymbol{c}_{-i}, \alpha) \int p(\tilde{y}_i|\mu_j, s_j) p(\mu_j, s_j|\mu_0, s_0, \gamma, \beta) \, d\mu_j ds_j \\ & \propto \frac{\alpha}{n-1+\alpha} \int \mathcal{N}(\tilde{y}_i, \phi) \, dG_0(\phi).^{6} \end{aligned}$$

---

[6] priors μ_0, s_0^{-1} (MacEachern & Müller, 1998)

Conditional Posterior Distribution of c_i

$p(c_i|\boldsymbol{c}_{-i})$ has to be weighted by $\mathcal{N}(\bar{y}_j, s_j + s_0)$, the posterior probability of *newly observed* values given the data you have already seen:

$$\text{clusters where } n_{-i,j} > 0: \quad p(c_i|\boldsymbol{c}_{-i}, \mu_j, s_j, \alpha) \propto p(c_i|\boldsymbol{c}_{-i}, \alpha) p(\tilde{y}_i|\mu_j, s_j, \boldsymbol{c}_{-i}) \propto \frac{n_{-i,j}}{n-1+\alpha} \mathcal{N}(\bar{y}_j, s_j + s_0),$$

$$\text{all other clusters combined}: \quad p(c_i \neq c_j \; \forall j \neq i|\boldsymbol{c}_{-i}, \mu_0, s_0, \gamma, \beta, \alpha) \propto p(c_i \neq c_j \; \forall j \neq i|\boldsymbol{c}_{-i}, \alpha) \int p(\tilde{y}_i|\mu_j, s_j) p(\mu_j, s_j|\mu_0, s_0, \gamma, \beta) \, d\mu_j ds_j \propto \frac{\alpha}{n-1+\alpha} \int \mathcal{N}(\tilde{y}_i, \phi) \, dG_0(\phi).[6]$$

[6] priors μ_0, s_0^{-1} (MacEachern & Müller, 1998)

Conditional Posterior Distribution of c_i

$p(c_i|\boldsymbol{c}_{-i})$ has to be weighted by $\mathcal{N}(\bar{y}_j, s_j + s_0)$, the posterior probability of *newly observed* values given the data you have already seen:

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

$$\begin{aligned} p(c_i|\boldsymbol{c}_{-i}, \mu_j, s_j, \alpha) &\propto p(c_i|\boldsymbol{c}_{-i}, \alpha) p(\tilde{y}_i|\mu_j, s_j, \boldsymbol{c}_{-i}) \\ &\propto \frac{n_{-i,j}}{n-1+\alpha} \mathcal{N}(\bar{y}_j, s_j + s_0), \end{aligned}$$

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

$$\begin{aligned} &p(c_i \neq c_j \, \forall j \neq i|\boldsymbol{c}_{-i}, \mu_0, s_0, \gamma, \beta, \alpha) \propto \\ &p(c_i \neq c_j \, \forall j \neq i|\boldsymbol{c}_{-i}, \alpha) \int p(\tilde{y}_i|\mu_j, s_j) p(\mu_j, s_j|\mu_0, s_0, \gamma, \beta) d\mu_j ds_j \\ &\propto \frac{\alpha}{n-1+\alpha} \int \mathcal{N}(\tilde{y}_i, \phi) dG_0(\phi). \end{aligned}$$ [6]

[6] priors μ_0, s_0^{-1} (MacEachern & Müller, 1998)

Conditional Posterior Distribution of c_i

$p(c_i|\boldsymbol{c}_{-i})$ has to be weighted by $\mathcal{N}(\bar{y}_j, s_j + s_0)$, the posterior probability of *newly observed* values given the data you have already seen:

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

$$\begin{aligned} p(c_i|\boldsymbol{c}_{-i}, \mu_j, s_j, \alpha) &\propto p(c_i|\boldsymbol{c}_{-i}, \alpha) p(\tilde{y}_i|\mu_j, s_j, \boldsymbol{c}_{-i}) \\ &\propto \frac{n_{-i,j}}{n-1+\alpha} \mathcal{N}(\bar{y}_j, s_j + s_0), \end{aligned}$$

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

$$p(c_i \neq c_j \; \forall j \neq i | \boldsymbol{c}_{-i}, \mu_0, s_0, \gamma, \beta, \alpha) \propto$$

$$p(c_i \neq c_j \; \forall j \neq i | \boldsymbol{c}_{-i}, \alpha) \int p(\tilde{y}_i|\mu_j, s_j) p(\mu_j, s_j|\mu_0, s_0, \gamma, \beta) \, d\mu_j ds_j$$

$$\propto \frac{\alpha}{n-1+\alpha} \int \mathcal{N}(\tilde{y}_i, \phi) \, dG_0(\phi).^6$$

[6] priors μ_0, s_0^{-1} (MacEachern & Müller, 1998)

Conditional Posterior Distribution of c_i

$p(c_i|\boldsymbol{c}_{-i})$ has to be weighted by $\mathcal{N}(\bar{y}_j, s_j + s_0)$, the posterior probability of *newly observed* values given the data you have already seen:

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

$$\begin{aligned} p(c_i|\boldsymbol{c}_{-i}, \mu_j, s_j, \alpha) &\propto p(c_i|\boldsymbol{c}_{-i}, \alpha) p(\tilde{y}_i|\mu_j, s_j, \boldsymbol{c}_{-i}) \\ &\propto \frac{n_{-i,j}}{n-1+\alpha} \mathcal{N}(\bar{y}_j, s_j + s_0), \end{aligned}$$

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

$$p(c_i \neq c_j \, \forall j \neq i|\boldsymbol{c}_{-i}, \mu_0, s_0, \gamma, \beta, \alpha) \propto$$

$$p(c_i \neq c_j \, \forall j \neq i|\boldsymbol{c}_{-i}, \alpha) \int p(\tilde{y}_i|\mu_j, s_j) p(\mu_j, s_j|\mu_0, s_0, \gamma, \beta) d\mu_j ds_j$$

$$\propto \frac{\alpha}{n-1+\alpha} \int \mathcal{N}(\tilde{y}_i, \phi) dG_0(\phi).^6$$

[6] priors μ_0, s_0^{-1} (MacEachern & Müller, 1998)

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**

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

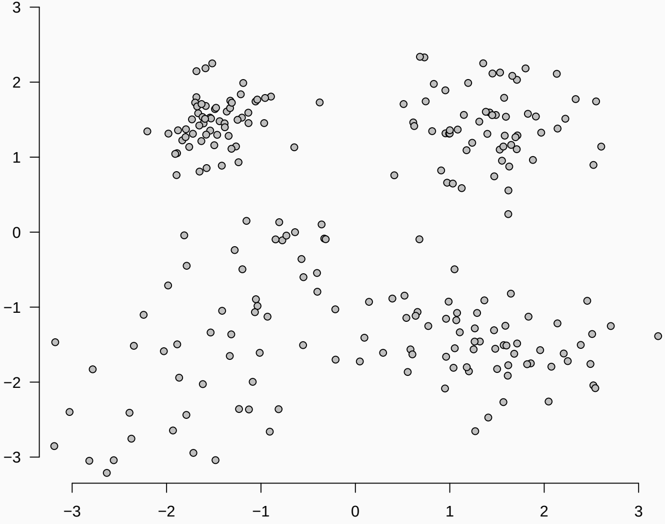
else

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

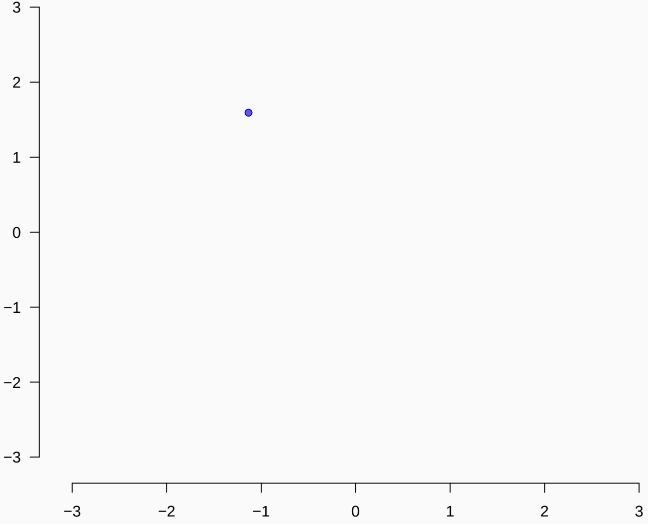
end

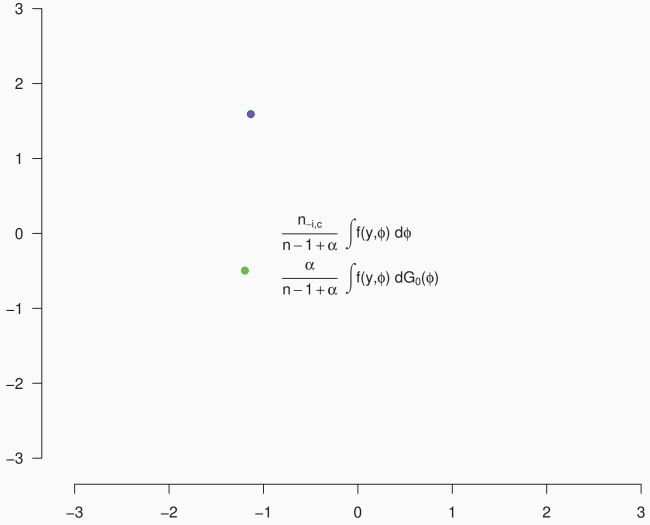
return c_i

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

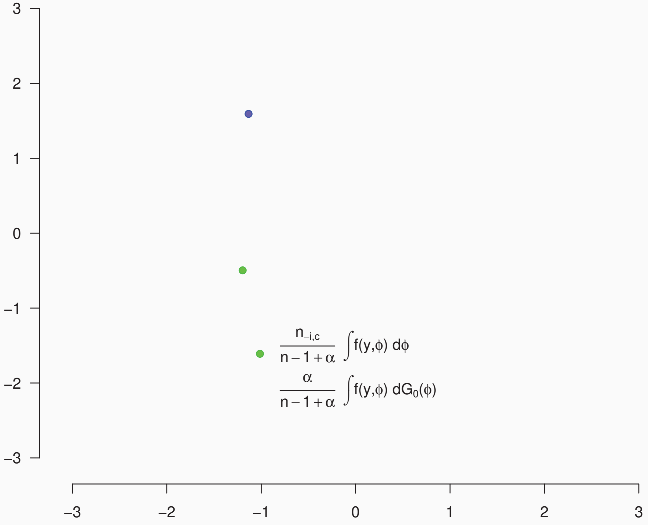


3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

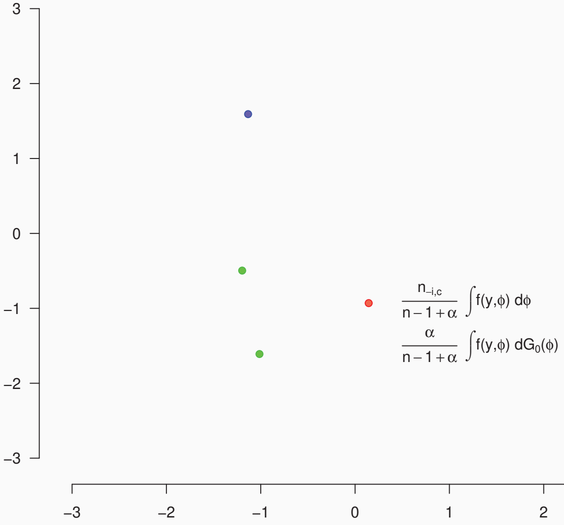

3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

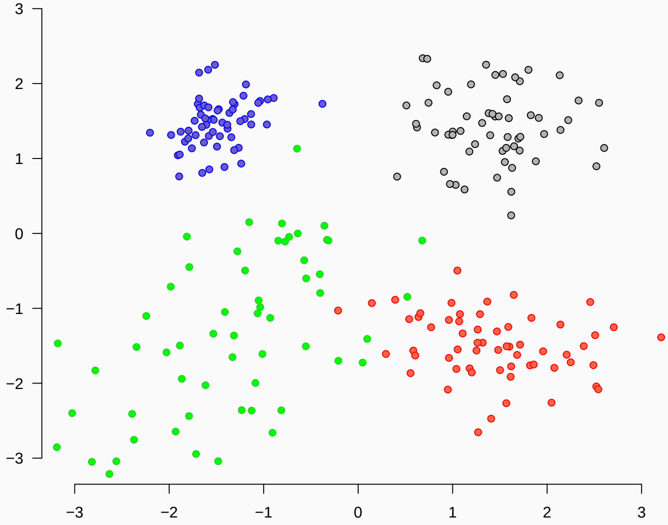


3
2
1
0
–1
–2
–3
$\frac{n_{-i,c}}{n-1+\alpha} \int f(y,\phi)\, d\phi$
$\frac{\alpha}{n-1+\alpha} \int f(y,\phi)\, dG_0(\phi)$
–3
–2
–1
0
1
2
3



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


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


3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

R Program

R Implementation of the CRP [8]

```
....
# 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)
....
```

[8] Modified from the R program by Tamara Broderick
https://people.csail.mit.edu/tbroderick/tutorial_2017/mitll.html

Application: Individual Meaning-Centered Psychotherapy Trial (IMCP)

IMCP Design and Data

- A randomized controlled trial (R01 CA128134, PI: Breitbart)
- Patients with advanced or terminal cancer
- Randomization
  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 a sense of meaning near end of life

IMCP Design and Data

- A randomized controlled trial (R01 CA128134, PI: Breitbart)
- Patients with advanced or terminal cancer
- Randomization
    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 a 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)

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

Responders to IMCP?

- Personal Meaning subscale scores at post-Tx

Post-Tx Personal Meaning subscale scores cf. baseline

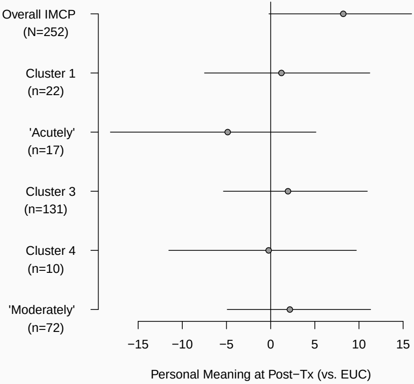
		Baseline		Post-Tx (week 7)					
		(N)		EUC	(n)	Meaning	(n)	Suprt	(n)
	1	(22)	82.5	78.5	6	88.8	5	83.3	11
"Acutely"	2	(17)	47.0	51.3	5	51.4	6	61.0	6
	3	(131)	82.4	81.9	46	92.0	48	87.5	37
	4	(10)	93.5	95.0	1	84.0	2	91.0	7
"Moderately"	5	(72)	66.8	71.4	16	84.1	31	75.8	25

Clusters Responded to IMCP Differently

- “random intervention effects” model [9]

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

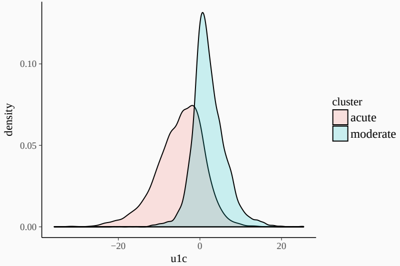
- `stan_lmer(PersMeaningT3 ~ Tx + (1 + Tx | clus), prior = NULL)`



[9] 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	-

Summary on the IMCP Trial

- Overall IMCP effect [10]
- Subtle effect: IMCP works better in moderately than acutely distressed
- Help guide future interventions
- Randomization is important: DPMM cannot replace it

[10] Rosenfeld, Breitbart *et al.*, *Cancer*, in press.

Conclusions

- I hope I have given you enough info on DPMM
- DPMM and R to make things explicit
- Derivations from finite mixture to infinite mixture
- DP prior $DP(\alpha G_0)$ to posterior
    - CRP yields DP posterior
    - A stochastic process controlled by α, G_0, and $\mathcal{N}(\tilde{y}_i|\mu_j, s_j^{-1})$
- Scratched the surface only
- I have not covered
    - Hyperpriors
    - Posterior on α (Escobar & West, 1995)
    - Other construction methods (e.g., stick-breaking)

Next steps?

- Gershman & Blei (2012); Neal (2000); Rasmussen (2000)
- Explore other tools in BNP, e.g., George Karabatsos
- BNP by Measure Theory (Jara, 2016, Int J Approx Reasoning)
- More abstract representations (Ferguson, 1973)

$$
\begin{aligned}
y_i|\theta_i &\sim \mathcal{N}(\theta_i) \\
\theta_i|G &\sim G \\
G &\sim DP(G_0, \alpha)
\end{aligned}
$$

Thanks

- Funding
  - NIH R01 CA128134 (W. B.)
  - NIH P30 CA008748 for Memorial Sloan Kettering Cancer Center
- DP mixture clustering computer program in R
  - Tamara Broderick https://people.csail.mit.edu/tbroderick/tutorial_2017/mitll.html