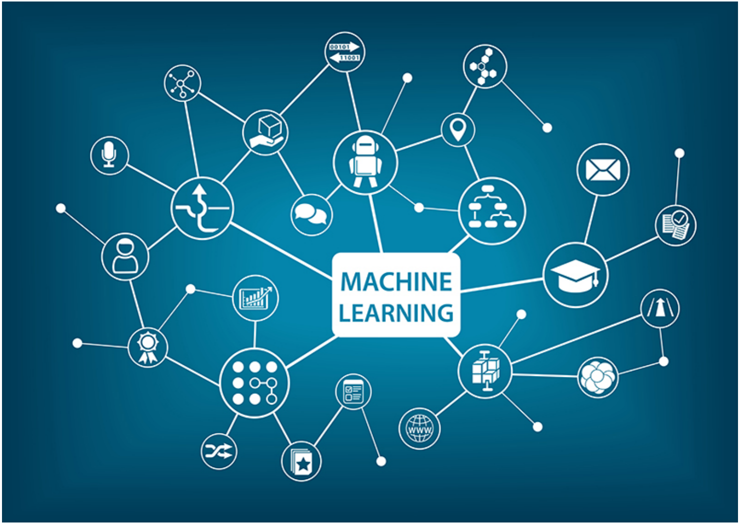
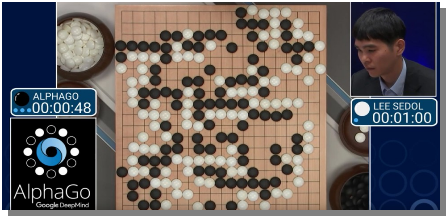
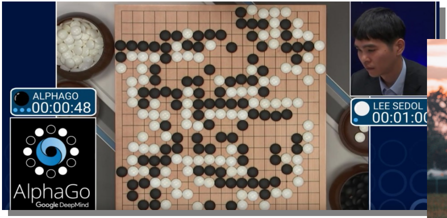
Bayesian Nonparametric Models for Data Exploration (CRCS Seminar)

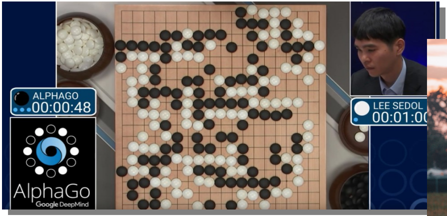
Melanie F. Pradier

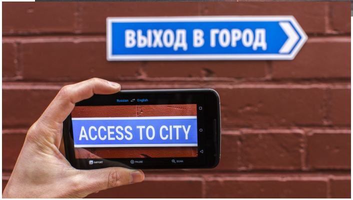
Monday 05th March, 2018

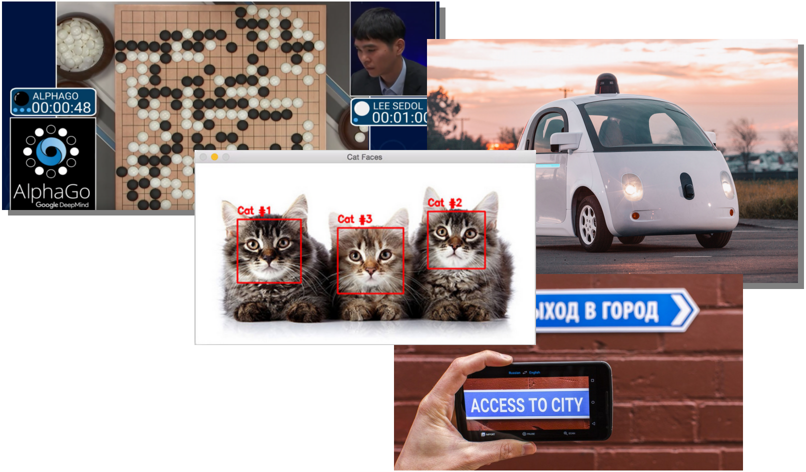

MACHINE
LEARNING

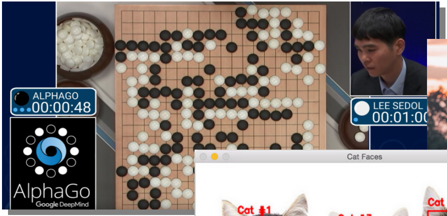

ALPHAGO
00:00:48
LEE SEDOL
00:01:00
AlphaGo
Google DeepMind

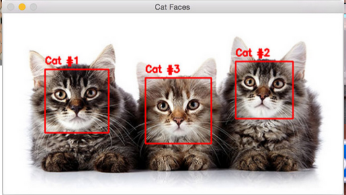

ALPHAGO
00:00:48
AlphaGo
Google DeepMind
LEE SEDOL

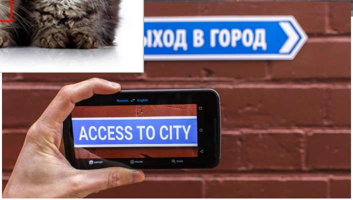

ALPHAGO
00:00:48
AlphaGo
Google DeepMind
LEE SEDOL


ВЫХОД В ГОРОД
ACCESS TO CITY

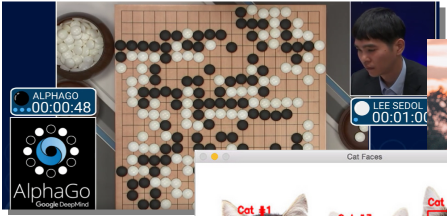

ALPHAGO
00:00:48
AlphaGo
Google DeepMind
LEE SEDOL
Cat Faces
Cat #1
Cat #3
Cat #2
ACCESS TO CITY

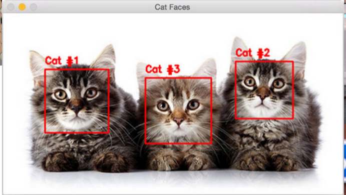


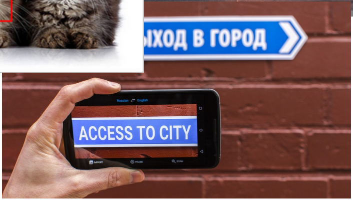




Data Exploitation Age







Data Exploitation Age

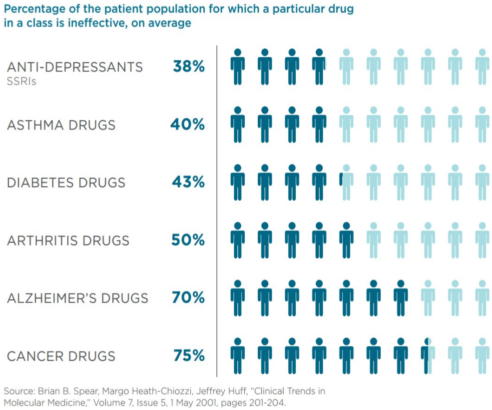
··· but are we making the utmost out of data?

Are we making the utmost out of data?

An example: personalized medicine

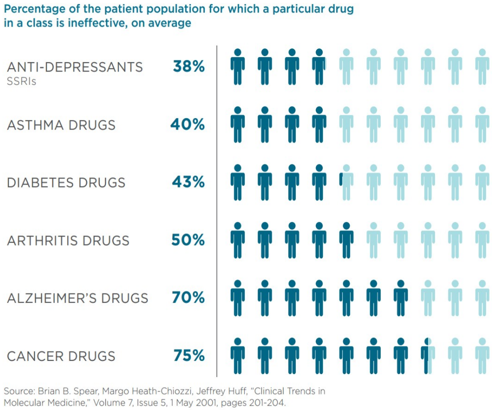
Are we making the utmost out of data?

An example: personalized medicine



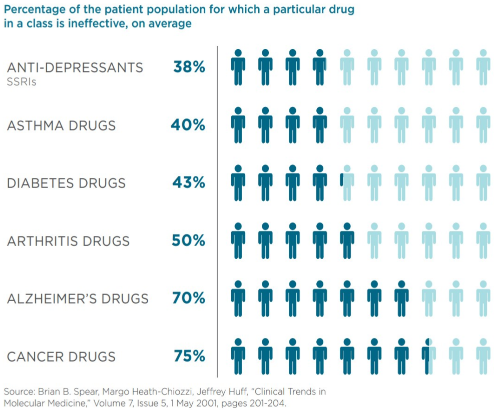
Are we making the utmost out of data?

An example: personalized medicine



Are we making the utmost out of data?

An example: personalized medicine

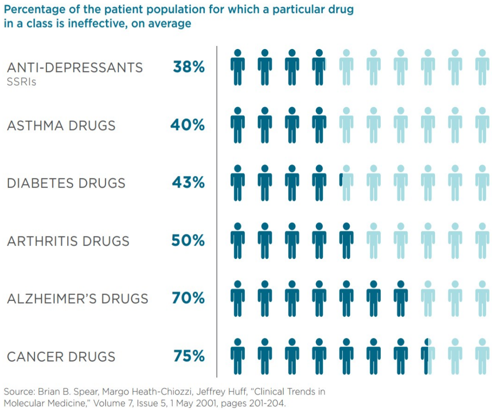


Challenges

- Complexity

Are we making the utmost out of data?

An example: personalized medicine

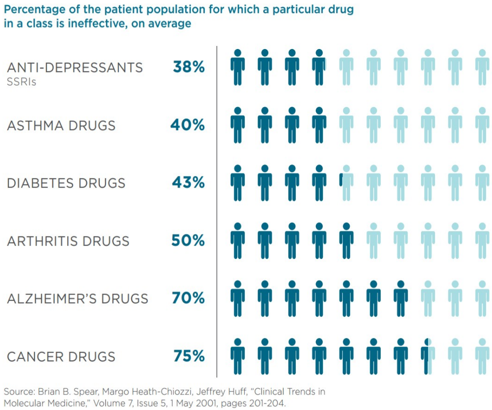


Challenges

- Complexity
- Missing data

Are we making the utmost out of data?

An example: personalized medicine

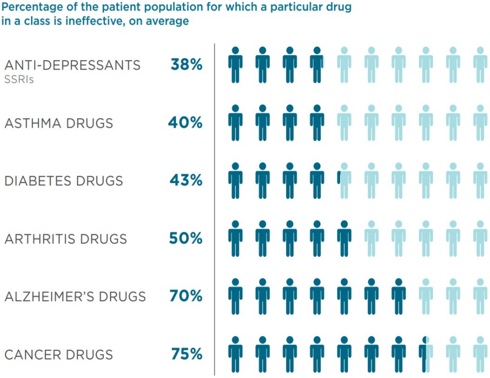


Challenges

- Complexity
- Missing data
- *Small data within big data*
- ...

Are we making the utmost out of data?

An example: personalized medicine



Source: Brian B. Spear, Margo Heath-Chiozzi, Jeffrey Huff, "Clinical Trends in Molecular Medicine," Volume 7, Issue 5, 1 May 2001, pages 201-204.

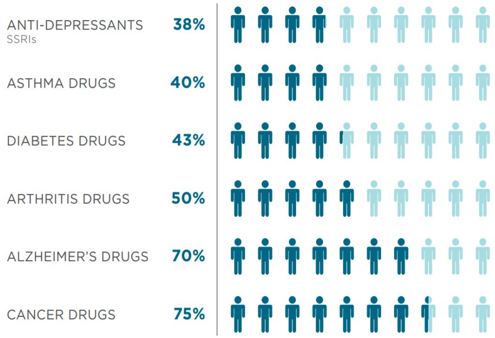
Challenges

- Complexity
- Missing data
- *Small data within big data*
- ...
- Final objective
  → data exploration

Are we making the utmost out of data?

An example: personalized medicine

Percentage of the patient population for which a particular drug in a class is ineffective, on average



Source: Brian B. Spear, Margo Heath-Chiozzi, Jeffrey Huff, "Clinical Trends in Molecular Medicine," Volume 7, Issue 5, 1 May 2001, pages 201-204.

Challenges

- Complexity
- Missing data
- *Small data within big data*
- ...
- Final objective
  $\rightarrow$ data exploration

How can ML systems help "understand" data?

Focus: data exploration

Interpretability

- "ability to explain or to present in understandable terms to a human" (Doshi-Velez and Kim, 2017)
- requirement in the 2018 EU General Data Protection Regulation (Goodman et.al. 2016)

Focus: data exploration

Interpretability

- "ability to explain or to present in understandable terms to a human" (Doshi-Velez and Kim, 2017)
- requirement in the 2018 EU General Data Protection Regulation (Goodman et.al. 2016)

Interpretable Machine Learning

- Interpretable models to explain black-boxes
  - Local Interpretable Explanations (Ribeiro et.al, 2016)
  - Interpretable Decision Sets (Lakkaraju et.al, 2016)
- Interpretable models from scratch
  - Tree-regularization of deep models (Wu et.al, 2017)
  - Input-gradient regularization (Ross et.al, 2017)

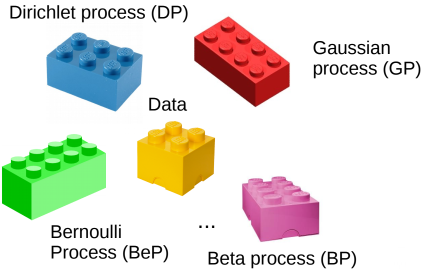
In this talk, interpretability via prob. graphical models

Why probabilistic graphical models?

- Generative model $\equiv$ unsupervised approach, model $p(\mathbf{X})$
- Graphical model for multidisciplinary research
- Latent variables

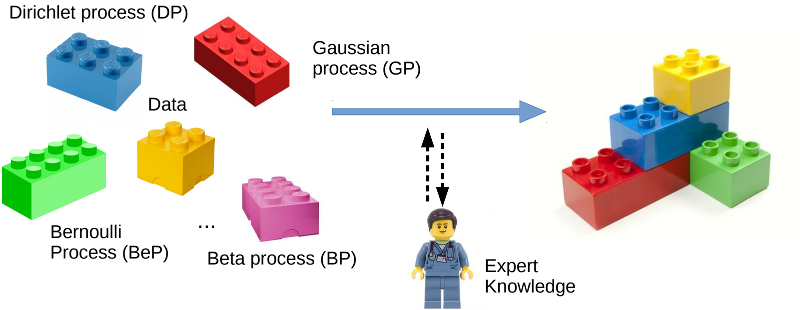
Why probabilistic graphical models?

- Generative model $\equiv$ unsupervised approach, model $p(\mathbf{X})$
- Graphical model for multidisciplinary research
- Latent variables



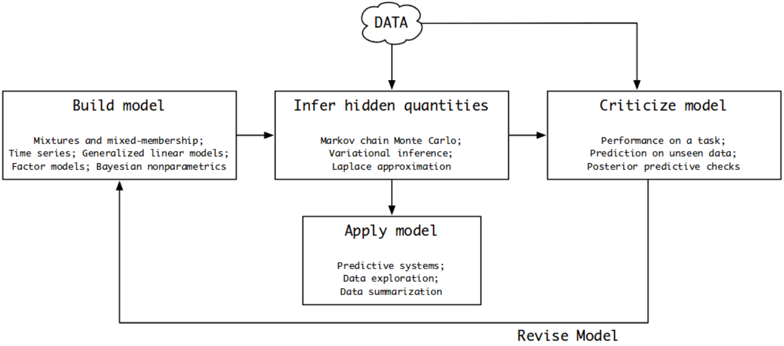
Why probabilistic graphical models?

- Generative model $\equiv$ unsupervised approach, model $p(\mathbf{X})$
- Graphical model for multidisciplinary research
- Latent variables



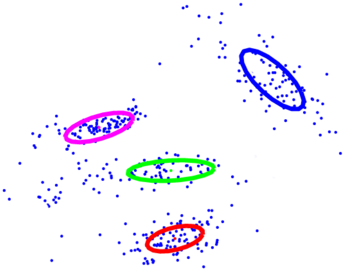
Why probabilistic graphical models?

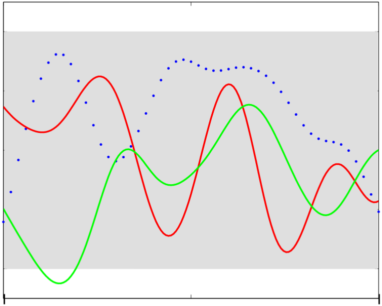
The "Box's loop" (Blei, 2014)



Why Bayesian nonparametrics?

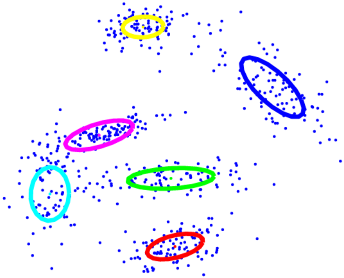
- Bayesian: to handle uncertainty $p(\mathbf{\Phi}|\mathbf{X}) \propto p(\mathbf{X}|\mathbf{\Phi})p(\mathbf{\Phi})$
- Nonparametric: to adapt model complexity depending on input data (hypothesis generation)

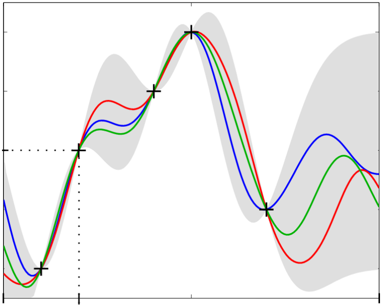




Why Bayesian nonparametrics?

- Bayesian: to handle uncertainty $p(\mathbf{\Phi}|\mathbf{X}) \propto p(\mathbf{X}|\mathbf{\Phi})p(\mathbf{\Phi})$
- Nonparametric: to adapt model complexity depending on input data (hypothesis generation)





Outline

1. Bayesian nonparametrics
2. Marathon modeling
3. Biomarker discovery in clinical trials

Outline

1. Bayesian nonparametrics
2. Marathon modeling
3. Biomarker discovery in clinical trials

Bayesian nonparametrics (BNPs)

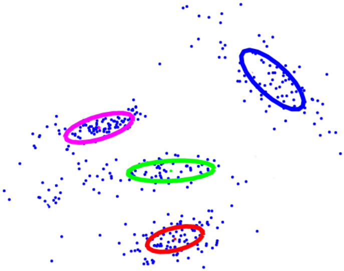
- Bayesian framework for **model selection**
- Nonparametric: number of parameters grows with the amount of data:
  - Prior over **infinite-dimensional** parameter space
  - Only a **finite subset** of parameters is used for any finite dataset

Bayesian nonparametrics (BNPs)

- Bayesian framework for **model selection**
- Nonparametric: number of parameters grows with the amount of data:
    - Prior over **infinite-dimensional** parameter space
    - Only a **finite subset** of parameters is used for any finite dataset

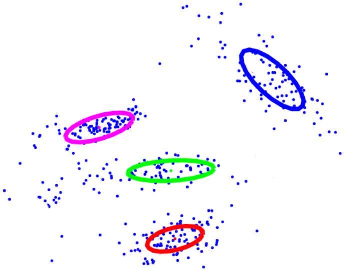
- Rely on stochastic processes:
    - Dirichlet process
    - Beta process
    - Gaussian process
    - ...

An example: finite Gaussian mixture model



π_k : mixture weights
ϕ_k : mixture parameters

An example: finite Gaussian mixture model



π_k : mixture weights
ϕ_k : mixture parameters

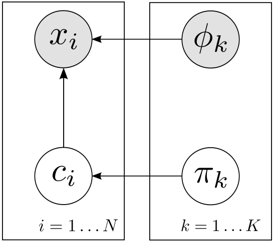
$$\phi_k \sim G_0$$

$$\pi_{1:K} \sim \text{Dirichlet}\left(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K}\right)$$

$$c_i \sim \text{Categorical}\left(\pi_1, \ldots, \pi_K\right)$$

$$x_i | c_i, \phi_{c_i} \sim F\left(\phi_{c_i}\right)$$

An example: finite Gaussian mixture model



$$p(x) = \sum_{k=1}^{K} \pi_k N(x; \mu_k, \Sigma_k)$$

π_k : mixture weights
ϕ_k : mixture parameters

$$\phi_k \sim G_0$$

$$\pi_{1:K} \sim \text{Dirichlet}\left(\frac{\alpha}{K}, \dots, \frac{\alpha}{K}\right)$$

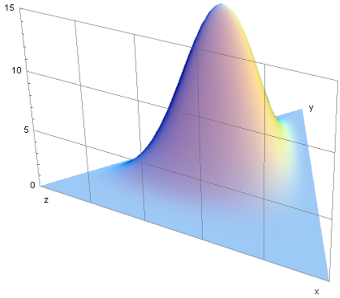
$$c_i \sim \text{Categorical}(\pi_1, \dots, \pi_K)$$

$$x_i | c_i, \phi_{c_i} \sim F(\phi_{c_i})$$

An example: finite Gaussian mixture model

Dirichlet distribution

$$f\left(x_1, \ldots, x_K; \alpha_1, \ldots, \alpha_K\right) = \frac{1}{B\left(\alpha\right)} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$$



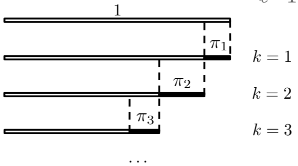
An example: infinite Gaussian mixture model

Stick-breaking process

(Ishwaran et.al, 2001)

For $k = 1, \cdots, \infty$

$$v_k \sim \text{Beta}(\alpha, 1),\ \pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$



$$\boldsymbol{\pi} \sim \text{GEM}(\alpha)$$

For $k = 1, \cdots, \infty$

$$\phi_k \sim G_0$$

Dirichlet Process

$$G \sim \text{DP}(\alpha, G_0)$$

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$

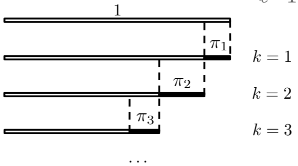
An example: infinite Gaussian mixture model

Stick-breaking process

(Ishwaran et.al, 2001)

For $k = 1, \cdots, \infty$

$$v_k \sim \text{Beta}(\alpha, 1),\ \pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$



$$\boldsymbol{\pi} \sim \text{GEM}(\alpha)$$

For $k = 1, \cdots, \infty$

$$\phi_k \sim G_0$$

Dirichlet Process

$$G \sim \text{DP}(\alpha, G_0)$$

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$

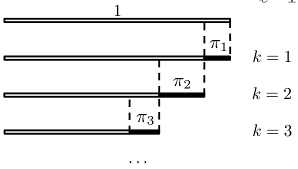
An example: infinite Gaussian mixture model

Stick-breaking process

(Ishwaran et.al, 2001)

For $k = 1, \cdots, \infty$

$$v_k \sim \text{Beta}(\alpha, 1),\ \pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$



$$\boldsymbol{\pi} \sim \text{GEM}(\alpha)$$

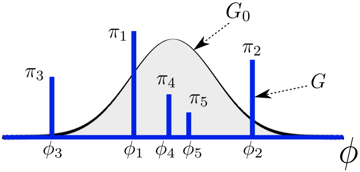
For $k = 1, \cdots, \infty$

$$\phi_k \sim G_0$$

Dirichlet Process

$$G \sim \text{DP}(\alpha, G_0)$$

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$



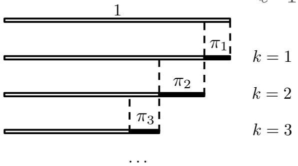
An example: infinite Gaussian mixture model

Stick-breaking process

(Ishwaran et.al, 2001)

For $k = 1, \cdots, \infty$

$$v_k \sim \mathrm{Beta}(\alpha, 1),\ \pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$



$$\boldsymbol{\pi} \sim \mathrm{GEM}(\alpha)$$

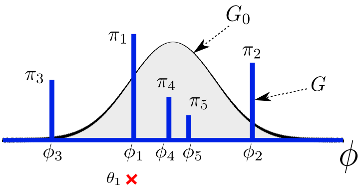
For $k = 1, \cdots, \infty$

$$\phi_k \sim G_0$$

Dirichlet Process

$$G \sim \mathrm{DP}(\alpha, G_0)$$

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$



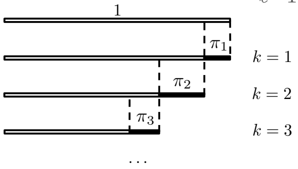
An example: infinite Gaussian mixture model

Stick-breaking process

(Ishwaran et.al, 2001)

For $k = 1, \cdots, \infty$

$$v_k \sim \mathrm{Beta}(\alpha, 1),\ \pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$



$$\boldsymbol{\pi} \sim \mathrm{GEM}(\alpha)$$

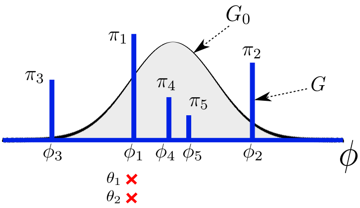
For $k = 1, \cdots, \infty$

$$\phi_k \sim G_0$$

Dirichlet Process

$$G \sim \mathrm{DP}(\alpha, G_0)$$

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$



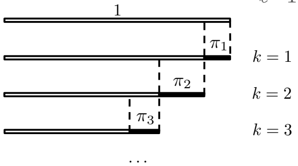
An example: infinite Gaussian mixture model

Stick-breaking process

(Ishwaran et.al, 2001)

For $k = 1, \cdots, \infty$

$$v_k \sim \mathrm{Beta}(\alpha, 1),\ \pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$



$$\boldsymbol{\pi} \sim \mathrm{GEM}(\alpha)$$

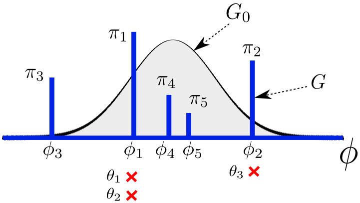
For $k = 1, \cdots, \infty$

$$\phi_k \sim G_0$$

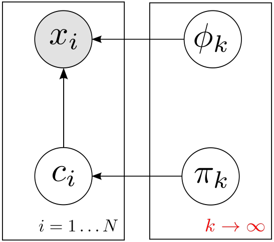
Dirichlet Process

$$G \sim \mathrm{DP}(\alpha, G_0)$$

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$



An example: infinite Gaussian mixture model

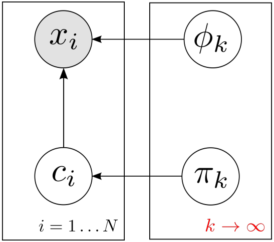


π_k : mixture weights
ϕ_k : mixture parameters

$$p(x) = \sum_{k=1}^{\infty} \pi_k N(x; \mu_k, \Sigma_k)$$

$$\begin{aligned} G &\sim \mathrm{DP}(\alpha, \mathrm{G}_0) \\ \theta_i | G &\sim G \\ x_i | \theta_i &\sim F(\theta_i) \end{aligned}$$

An example: infinite Gaussian mixture model



π_k : mixture weights
ϕ_k : mixture parameters

$$\phi_k \sim G_0$$

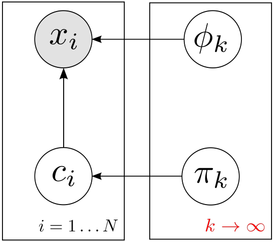
$$\pi_{1:K} \sim \text{Dirichlet}\left(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K}\right)$$

$$c_i \sim \text{Categorical}\left(\pi_1, \ldots, \pi_K\right)$$

$$x_i | c_i, \phi_{c_i} \sim F\left(\phi_{c_i}\right)$$

$$p\left(x\right) = \sum_{k=1}^{\infty} \pi_k N\left(x; \mu_k, \Sigma_k\right)$$

An example: infinite Gaussian mixture model



π_k : mixture weights

ϕ_k : mixture parameters

$$\phi_k \sim G_0$$

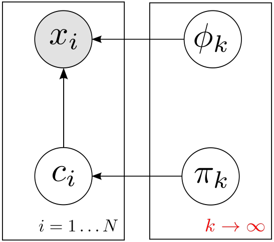
$$\boldsymbol{\pi} \sim \text{GEM}(\alpha)$$

$$c_i \sim \text{Categorical}(\boldsymbol{\pi})$$

$$x_i | c_i, \phi_{c_i} \sim F(\phi_{c_i})$$

$$p(x) = \sum_{k=1}^{\infty} \pi_k N(x; \mu_k, \Sigma_k)$$

An example: infinite Gaussian mixture model



π_k : mixture weights
ϕ_k : mixture parameters

$$\phi_k \sim G_0$$

$$\boldsymbol{c} \sim \text{CRP}(\alpha)$$

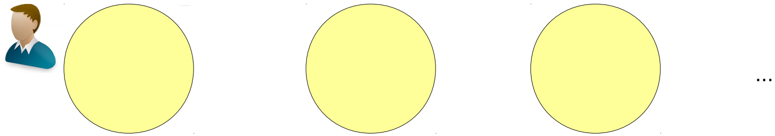
$$x_i | c_i, \phi_{c_i} \sim F(\phi_{c_i})$$

$$p(x) = \sum_{k=1}^{\infty} \pi_k N(x; \mu_k, \Sigma_k)$$

Chinese restaurant process (CRP)

$$\boldsymbol{c} \sim \mathrm{CRP}(\alpha)$$

where $\boldsymbol{c} \equiv$ infinite sequence of natural numbers.



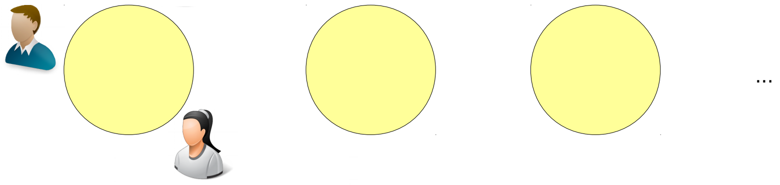
(Pitman et.al, 2002)

$$p\left(c_i = m | \boldsymbol{c}^{\neg i}, \alpha\right) \begin{cases} |m|^{\neg i}, & m \in \boldsymbol{c}^{\neg i} \\ \alpha, & m \notin \boldsymbol{c}^{\neg i} \end{cases}$$

Chinese restaurant process (CRP)

$$\boldsymbol{c} \sim \mathrm{CRP}(\alpha)$$

where $\boldsymbol{c} \equiv$ infinite sequence of natural numbers.



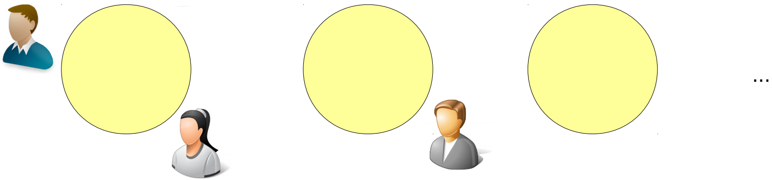
(Pitman et.al, 2002)

$$p\left(c_i = m|\boldsymbol{c}^{\neg i}, \alpha\right) \begin{cases} |m|^{\neg i}, & m \in \boldsymbol{c}^{\neg i} \\ \alpha, & m \notin \boldsymbol{c}^{\neg i} \end{cases}$$

Chinese restaurant process (CRP)

$$\boldsymbol{c} \sim \mathrm{CRP}(\alpha)$$

where $\boldsymbol{c} \equiv$ infinite sequence of natural numbers.



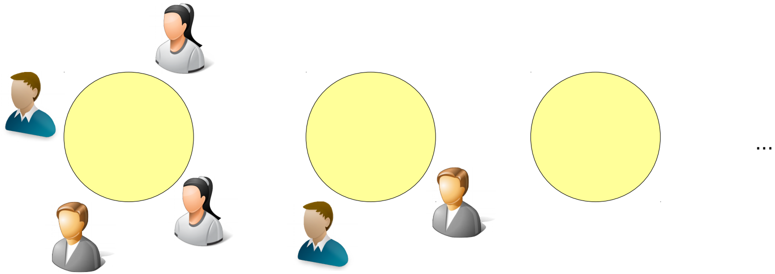
(Pitman et.al, 2002)

$$p\left(c_i = m | \boldsymbol{c}^{\neg i}, \alpha\right) \begin{cases} |m|^{\neg i}, & m \in \boldsymbol{c}^{\neg i} \\ \alpha, & m \notin \boldsymbol{c}^{\neg i} \end{cases}$$

Chinese restaurant process (CRP)

$$\boldsymbol{c} \sim \mathrm{CRP}(\alpha)$$

where $\boldsymbol{c} \equiv$ infinite sequence of natural numbers.



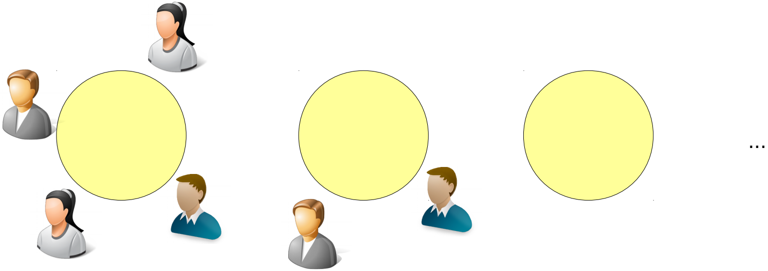
(Pitman et.al, 2002)

$$p\left(c_i = m | \boldsymbol{c}^{\neg i}, \alpha\right) \begin{cases} |m|^{\neg i}, & m \in \boldsymbol{c}^{\neg i} \\ \alpha, & m \notin \boldsymbol{c}^{\neg i} \end{cases}$$

Chinese restaurant process (CRP)

$$\boldsymbol{c} \sim \mathrm{CRP}(\alpha)$$

where $\boldsymbol{c} \equiv$ infinite sequence of natural numbers.



(Pitman et.al, 2002)

$$p\left(c_i = m|\boldsymbol{c}^{\neg i}, \alpha\right) \begin{cases} |m|^{\neg i}, & m \in \boldsymbol{c}^{\neg i} \\ \alpha, & m \notin \boldsymbol{c}^{\neg i} \end{cases}$$

Exchangeability and De Finetti's Theorem

Exchangeability (Pitman et.al, 2002)

An infinitely exchangeable sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ is a sequence whose probability is invariant under finite permutations ρ of the first N elements, for all $N \in \mathbb{N}$, i.e.,

$$p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = p(\mathbf{x}_{\rho(1)}, \mathbf{x}_{\rho(2)}, \ldots, \mathbf{x}_{\rho(n)}), \; \forall N \in \mathbb{N}.$$

De Finetti's Theorem (Foti et.al, 2012)

Any infinitely exchangeable sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ can be written as a mixture of i.i.d. samples as follows:

$$p(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N) = \int_\phi \prod_{n=1}^{N} \mathcal{Q}(\mathbf{x}_n|\phi)P(d\phi), \; \forall N \in \mathbb{N}, \tag{2.1}$$

Outline

1. Bayesian nonparametrics
2. Marathon modeling
3. Biomarker discovery in clinical trials

Motivation

1. What is the impact of age and gender on runners performance?
2. Can we compare different runners in a fair manner?
   - entry requirements
   - rewards

Motivation

1. What is the impact of age and gender on runners performance?
2. Can we compare different runners in a fair manner?
   - entry requirements
   - rewards

Our Approach

- dependent density estimation model
  - delivers scientific knowledge in sport sciences
  - constitutes a fair age-gender grading system
  - relies on dependent Dirichlet process

Dependent Dirichlet process (DDP)

(MacEachern,2000)

J: number of groups

$$G_j = \sum_{k=1}^{\infty} \pi_{jk} \delta_{\phi_{jk}}$$

Dependent Dirichlet process (DDP)

(MacEachern,2000)

J: number of groups

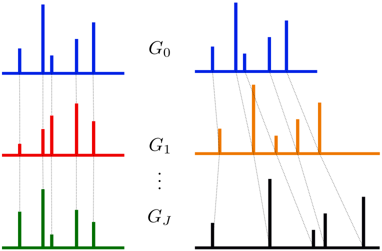
$$G_{j} = \sum_{k=1}^{\infty} \pi_{jk} \delta_{\phi_{jk}}$$

- hierarchical DP (Teh et.al, 2005)

$$G_{j} = \sum_{k=1}^{\infty} \pi_{jk} \delta_{\phi_{k}}$$

- single-p DDP (MacEachern, 2000)

$$G_{j} = \sum_{k=1}^{\infty} \pi_{k} \delta_{\phi_{jk}}$$



hierarchical DP

$$G_0 \sim \mathrm{DP}(\alpha, \mathrm{H})$$
$$G_j \sim \mathrm{DP}(\gamma, \mathrm{G}_0)$$

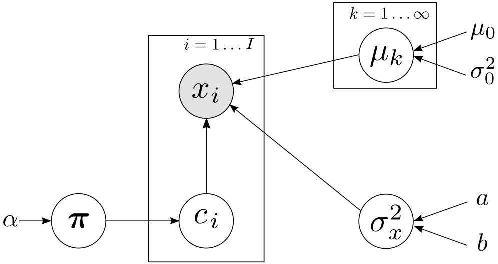
single-p DDP

$$G_0 \sim \mathrm{DP}(\alpha, \mathrm{H})$$
$$G_j = \mathrm{T}_j [G_0]$$

Atom-dependent DP mixture model

(Pradier et.al, 2016)

$x_i \equiv$ marathon finishing time for runner i



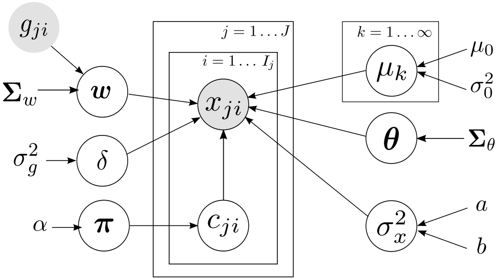
$$\boldsymbol{\pi}|\alpha \sim \mathrm{GEM}(\alpha)$$
$$c_i|\boldsymbol{\pi} \sim \mathrm{Cat}(\boldsymbol{\pi})$$
$$\mu_k \sim \mathcal{N}\left(\mu_0, \sigma_0^2\right)$$
$$\sigma_x^2 \sim \mathcal{IG}\left(a, b\right)$$

$$x_i|\text{other vars} \sim \mathcal{N}\left(x_i|\mu_{c_i}, \sigma_x^2\right)$$

Atom-dependent DP mixture model

(Pradier et.al, 2016)

$x_{ji} \equiv$ marathon finishing time for runner i in age group j



$$\boldsymbol{\pi}|\alpha \sim \mathrm{GEM}(\alpha)$$
$$c_{ji}|\boldsymbol{\pi} \sim \mathrm{Cat}(\boldsymbol{\pi})$$
$$\mu_k \sim \mathcal{N}\left(\mu_0, \sigma_0^2\right)$$
$$\sigma_x^2 \sim \mathcal{IG}\left(a, b\right)$$
$$\boldsymbol{\theta} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_\theta\right)$$

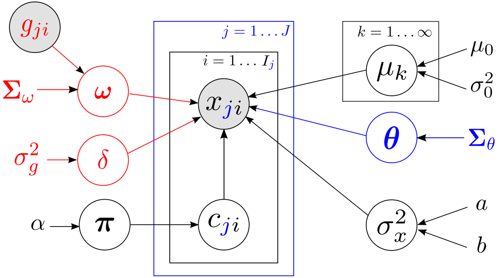
$$x_{ji}|\text{other vars} \sim \mathcal{N}\left(x_{ji}|\mu_{c_{ji}} + \theta_j, \sigma_x^2\right)$$

$$(\boldsymbol{\Sigma}_\theta)_{\ell q} = \sigma_\theta^2 \exp\left(-\frac{(\ell - q)^2}{2\nu^2}\right) + \kappa\delta(\ell - q)$$

Atom-dependent DP mixture model

(Pradier et.al, 2016)

$x_{ji} \equiv$ marathon finishing time for runner i in age group j $g_{ji} \equiv$ gender



$$\boldsymbol{\pi}|\alpha \sim \mathrm{GEM}(\alpha)$$
$$c_{ji}|\boldsymbol{\pi} \sim \mathrm{Cat}(\boldsymbol{\pi})$$
$$\mu_k \sim \mathcal{N}\left(\mu_0, \sigma_0^2\right)$$
$$\sigma_x^2 \sim \mathcal{IG}\left(a, b\right)$$
$$\boldsymbol{\theta} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_\theta\right)$$
$$\delta \sim \mathcal{N}\left(\mathbf{0}, \sigma_\omega^2\right)$$
$$\boldsymbol{\omega} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_\omega\right)$$

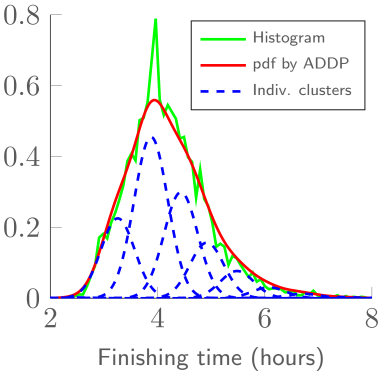
$$x_{ji}|\text{other vars} \sim \mathcal{N}\left(x_{ji}|\mu_{c_{ji}} + \theta_j + \mathbb{1}[g_{ji} = 1](\delta + \omega_j), \sigma_x^2\right)$$

$$(\boldsymbol{\Sigma}_\theta)_{\ell q} = \sigma_\theta^2 \exp\left(-\frac{(\ell - q)^2}{2\nu^2}\right) + \kappa\delta(\ell - q)$$

Results: impact of age

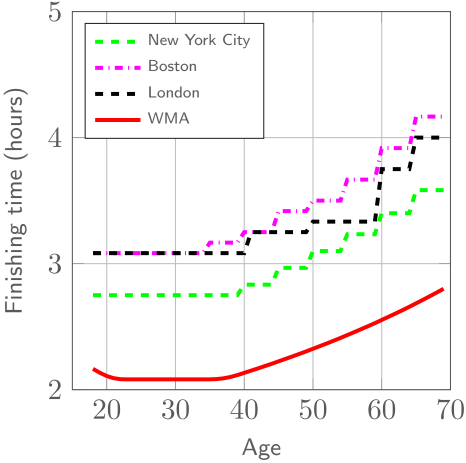
(Pradier et.al, 2016)

- MCMC approach
- block Gibbs sampler
- $1/4$ M runners

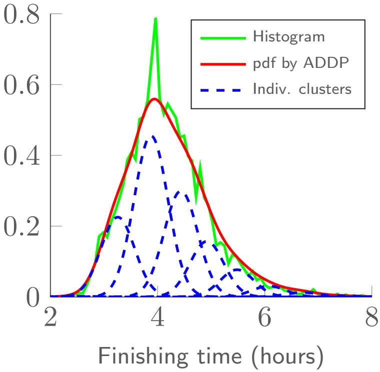


Results: impact of age

(Pradier et.al, 2016)

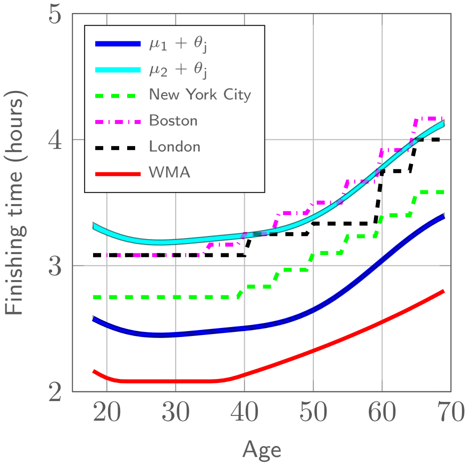


- MCMC approach
- block Gibbs sampler
- $1/4$ M runners

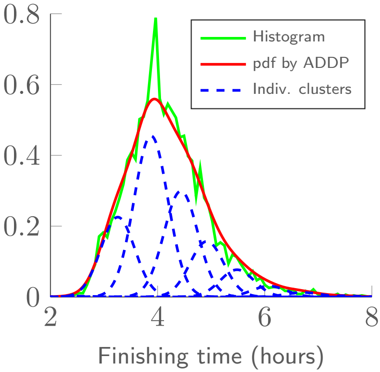


Results: impact of age

(Pradier et.al, 2016)

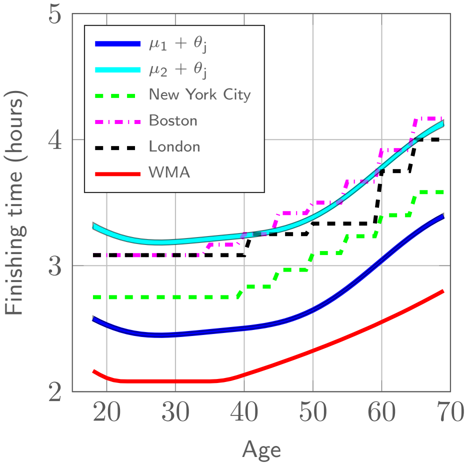


- MCMC approach
- block Gibbs sampler
- 1/4 M runners

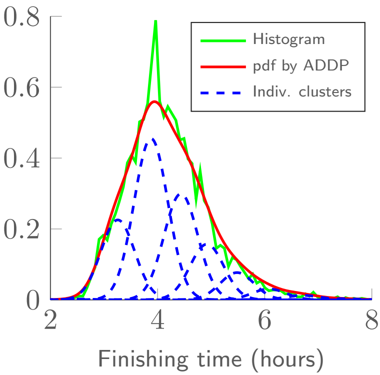


Results: impact of age

(Pradier et.al, 2016)

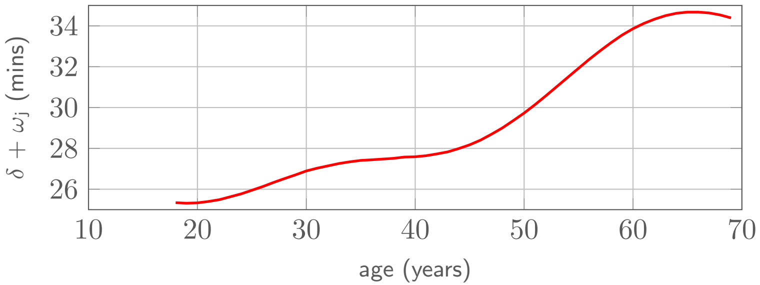


- MCMC approach
- block Gibbs sampler
- 1/4 M runners



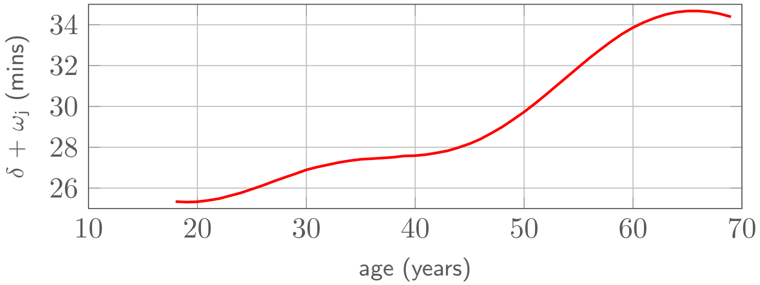
Results: impact of gender

(Pradier et.al, 2016)



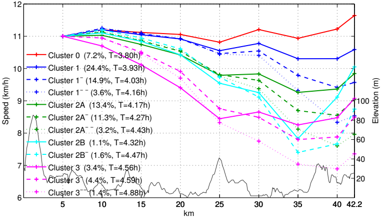
Results: impact of gender

(Pradier et.al, 2016)



Other Results (Pradier et.al, 2016)

- **Speed-dependent** cluster means
- Link to **mixture of experts**
- Analysis of **running patterns**
- **Prediction** of finishing time



Outline

1. Bayesian nonparametrics
2. Marathon modeling
3. Biomarker discovery in clinical trials

Our Focus: Biomarker discovery

Def: "any variable that can be used as an indicator of a particular disease state".

Our Focus: Biomarker discovery

Def: "any variable that can be used as an indicator of a particular disease state".

Biomarkers are used everywhere!!

Our Focus: Biomarker discovery

Def: "any variable that can be used as an indicator of a particular disease state".

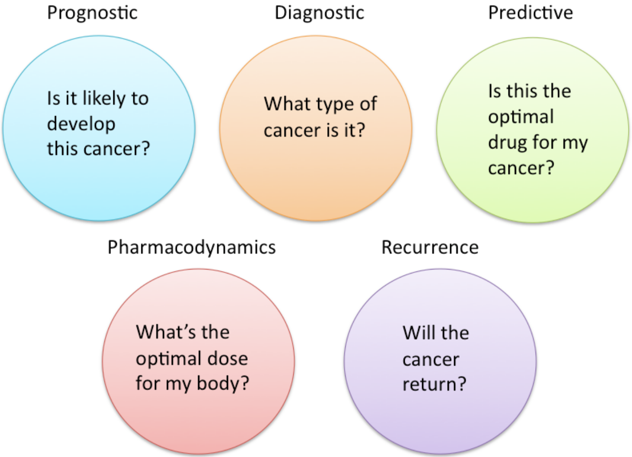
Biomarkers are used everywhere!!

Some examples

- Prostate-specific antigen (PSA) to diagnose prostate cancer
- Estrogen / progesterone to predict sensitivity to endocrine therapy in breast cancer
- KRAS mutation to predict resistance to EGFr antibody treatment

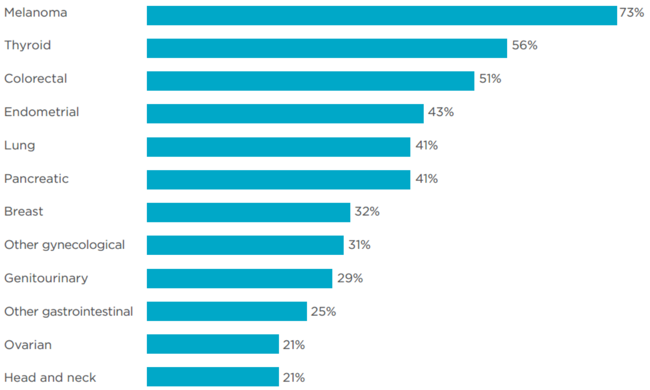
Our Focus: Biomarker discovery

Def: "any variable that can be used as an indicator of a particular disease state".



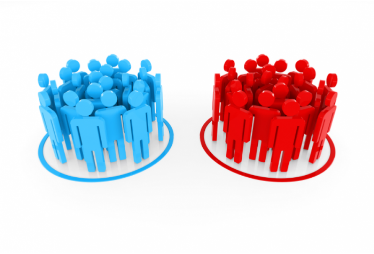
Biomarkers as potential targets for new drugs

TACKLING TUMORS: Percentage of patients whose tumors were driven by certain genetic mutations that could be targets for specific drugs, by types of cancer.



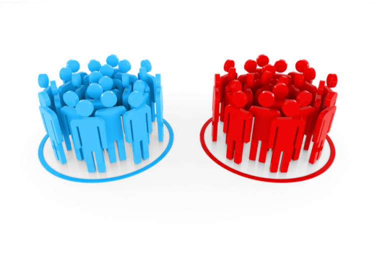
Motivation: biomarker discovery in clinical trials

Def: "any variable that can be used as an indicator of a particular disease state".



Motivation: biomarker discovery in clinical trials

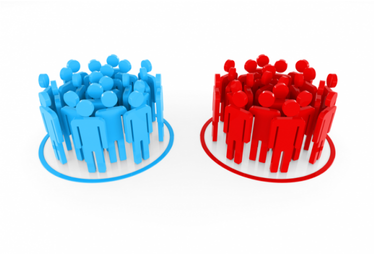
Def: "any variable that can be used as an indicator of a particular disease state".



We want to discover:

Motivation: biomarker discovery in clinical trials

Def: "any variable that can be used as an indicator of a particular disease state".

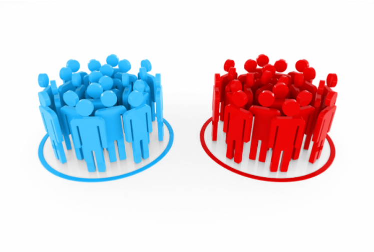


We want to discover:

1. Indicators of disease progression: prognostic biomarkers

Motivation: biomarker discovery in clinical trials

Def: "any variable that can be used as an indicator of a particular disease state".

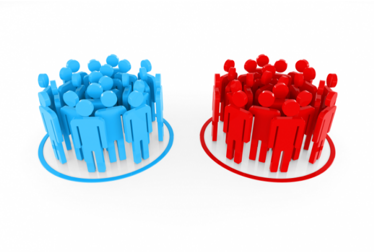


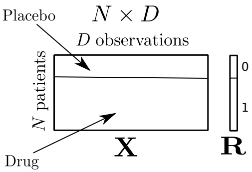
We want to discover:

1. Indicators of disease progression: prognostic biomarkers
2. Indicators of (positive) drug response: predictive biomarkers

Motivation: biomarker discovery in clinical trials

Def: "any variable that can be used as an indicator of a particular disease state".



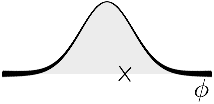


We want to discover:

1. Indicators of disease progression: prognostic biomarkers
2. Indicators of (positive) drug response: predictive biomarkers

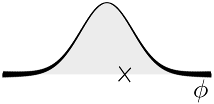
Indian Buffet Process (Ghahramani et.al, 2006)

Indian Buffet Process (Ghahramani et.al, 2006)

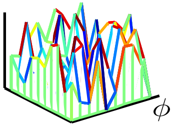


$$x \sim \mathcal{N}(\mu_0, \sigma_0^2) \longrightarrow x = 0.356$$

Indian Buffet Process (Ghahramani et.al, 2006)

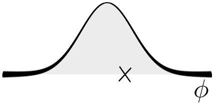


$$\xrightarrow[x \sim \mathcal{N}(\mu_0, \sigma_0^2)]{} \qquad x = 0.356$$

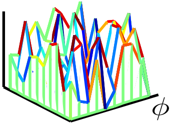


$$\xrightarrow[Z \sim \mathrm{IBP}(\alpha)]{} \qquad Z = \begin{bmatrix} 1 & 0 & 1 & \\ 0 & 1 & 0 & \vdots \\ 1 & 1 & 0 & \\ 1 & 0 & 1 & \end{bmatrix}$$

Indian Buffet Process (Ghahramani et.al, 2006)



$\longrightarrow$ $x \sim \mathcal{N}(\mu_0, \sigma_0^2)$ $\quad x = 0.356$



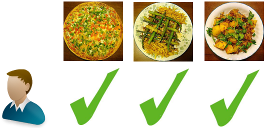
$\longrightarrow$ $Z \sim \text{IBP}(\alpha)$

$$Z = \begin{bmatrix} 1 & 0 & 1 & \\ 0 & 1 & 0 & \vdots \\ 1 & 1 & 0 & \\ 1 & 0 & 1 & \end{bmatrix}$$

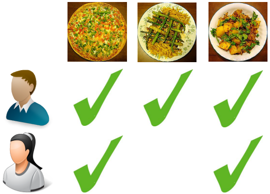
- Prior over binary matrices with infinite number of columns
- Rows $\equiv$ observations; columns $\equiv$ features
- $\mathbf{Z} \sim \text{IBP}(\alpha)$
- α: concentration parameter

Indian Buffet Process (Ghahramani et.al, 2006)

Indian Buffet Process (Ghahramani et.al, 2006)



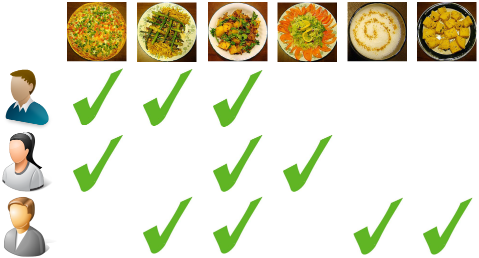
Indian Buffet Process (Ghahramani et.al, 2006)



Indian Buffet Process (Ghahramani et.al, 2006)

Indian Buffet Process (Ghahramani et.al, 2006)

Indian Buffet Process (Ghahramani et.al, 2006)



Indian Buffet Process (Ghahramani et.al, 2006)

Indian Buffet Process (Ghahramani et.al, 2006)

1	1	1	0	0	0
1	0	1	1	0	0
0	1	1	0	1	1

. . .

⋮

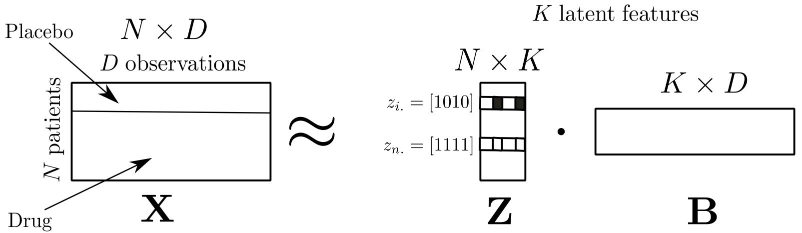
Indian Buffet Process (Ghahramani et.al, 2006)

. . .

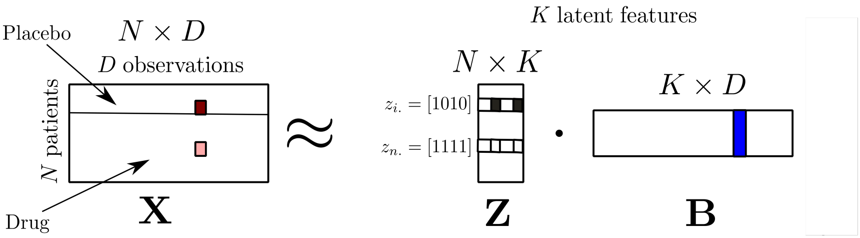
1	1	0	1	0	1
1	0	1	0	0	1
0	0	1	0	1	1

⋮

Infinite Latent Feature Model

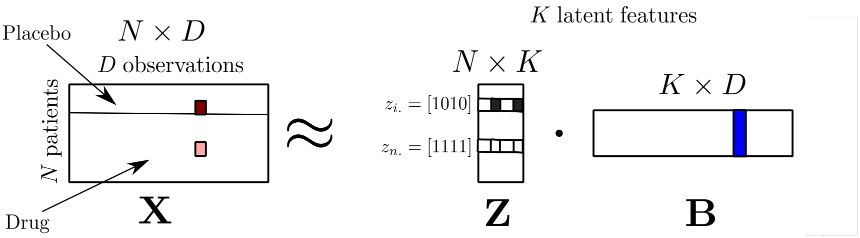


Infinite Latent Feature Model



- $x_{id} = 173\,\text{ml/dL} = 73 + 0 + 100\,\text{ml/dL}$
- $x_{nd} = 136\,\text{ml/dL} = 86 + 40 + 60 - 50\,\text{ml/dL}$

Infinite Latent Feature Model



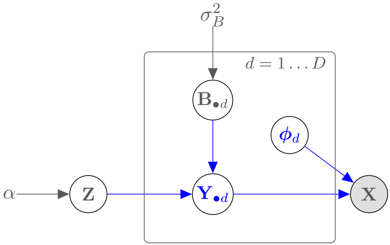
- $x_{id} = 173\,\mathrm{ml/dL} = 73 + 0 + 100\,\mathrm{ml/dL}$
- $x_{nd} = 136\,\mathrm{ml/dL} = 86 + 40 + 60 - 50\,\mathrm{ml/dL}$

Note: Correlation does not imply causality!

General latent feature model (GLFM)

(Valera et.al, 2017)

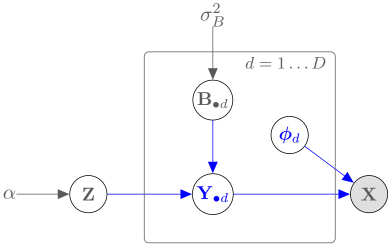
Latent feature model for heterogeneous datasets



General latent feature model (GLFM)

(Valera et.al, 2017)

Latent feature model for heterogeneous datasets



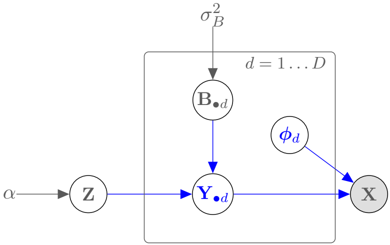
- Link functions T_d depend on type of data for each dimension d

$$\begin{aligned} x_{nd} &= T_d(y_{nd}; \boldsymbol{\phi_d}) \\ y_{nd}|\mathbf{Z},\mathbf{B} &\sim \mathcal{N}(\mathbf{Z}_{n\bullet}\mathbf{B}_{\bullet d}, \sigma_y^2) \\ B_{kd} &\sim \mathcal{N}(0, \sigma_B^2) \\ \mathbf{Z} &\sim \text{IBP}(\alpha) \end{aligned}$$

General latent feature model (GLFM)

(Valera et.al, 2017)

Latent feature model for heterogeneous datasets



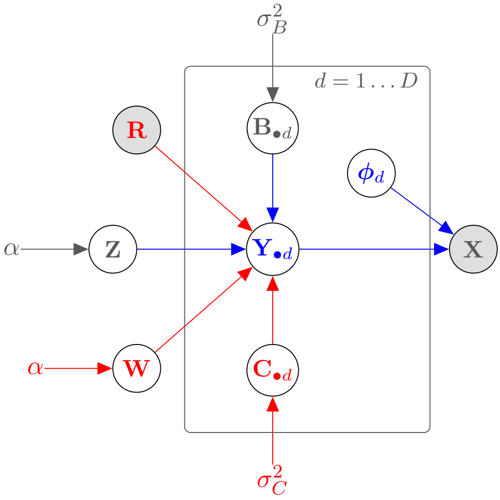
- Link functions T_d depend on type of data for each dimension d

$$
\begin{aligned}
x_{nd} &= T_d(y_{nd}; \boldsymbol{\phi_d}) \\
y_{nd}|\mathbf{Z}, \mathbf{B} &\sim \mathcal{N}(\mathbf{Z}_{n\bullet}\mathbf{B}_{\bullet d}, \sigma_y^2) \\
B_{kd} &\sim \mathcal{N}(0, \sigma_B^2) \\
\mathbf{Z} &\sim \mathrm{IBP}(\alpha)
\end{aligned}
$$

Open-source python code

https://github.com/ivaleraM/GLFM

Case-control IBP (Pradier et.al, 2018)



R_n: drug indicator por patient n

$$x_{nd} = T_d(y_{nd}; \boldsymbol{\phi_d})$$
$$y_{nd}|\mathbf{Z}, \mathbf{W}, \mathbf{B}, \mathbf{C}, \mathbf{R} \sim$$
$$\mathcal{N}(\mathbf{Z}_{n\bullet}\mathbf{B}_{\bullet d} + \mathbb{1}[\mathbf{R}_n = 1]\mathbf{W}_{n\bullet}\mathbf{C}_{\bullet d}, \sigma_y^2)$$
$$B_{kd} \sim \mathcal{N}(0, \sigma_B^2)$$
$$\mathbf{Z} \sim \text{IBP}(\alpha)$$
$$C_{kd} \sim \mathcal{N}(0, \sigma_C^2)$$
$$\mathbf{W} \sim \text{IBP}(\alpha)$$

- **Inference**: MCMC approach with accelerated Gibbs sampling
- **Biomarker discovery**: statistical multiple hypothesis testing

Results: subpopulations

GPC3 Antibody Treatment against Liver Cancer (J. Hepatology. 2016 Apr, Abou-Alfa et.al.)

- 180 patients: 60 took a placebo, 120 took the drug
- PFS: Progression Free Survival

Sub-population	Drug Identifier	F1	F2	F3	Size (number of patients)	Mean PFS (months)	Median PFS (months)
1.	0	0	0	0	33.37	3.06	1.65
2.	0	0	1	0	4.07	2.29	2.24
3.	0	1	0	0	17.84	2.72	1.81
4.	0	1	1	0	4.72	7.05	7.18
5.	1	0	0	0	51.52	3.22	2.55
6.	1	0	0	1	16.77	4.17	3.65
7.	1	0	1	0	8.38	1.74	1.33
8.	1	0	1	1	2.07	2.69	2.65
9.	1	1	0	0	29.88	3.36	2.03
10.	1	1	0	1	4.90	4.44	4.34
11.	1	1	1	0	4.53	6.31	5.31
12.	1	1	1	1	1.94	10.04	10.01

Results: subpopulations

GPC3 Antibody Treatment against Liver Cancer (J. Hepatology. 2016 Apr, Abou-Alfa et.al.)

- 180 patients: 60 took a placebo, 120 took the drug
- PFS: Progression Free Survival

Sub-population	Drug Identifier	F1	F2	F3	Size (number of patients)	Mean PFS (months)	Median PFS (months)
1.	0	0	0	0	33.37	3.06	1.65
2.	0	0	1	0	4.07	2.29	2.24
3.	0	1	0	0	17.84	2.72	1.81
4.	0	1	1	0	4.72	7.05	7.18
5.	1	0	0	0	51.52	3.22	2.55
6.	1	0	0	1	16.77	4.17	3.65
7.	1	0	1	0	8.38	1.74	1.33
8.	1	0	1	1	2.07	2.69	2.65
9.	1	1	0	0	29.88	3.36	2.03
10.	1	1	0	1	4.90	4.44	4.34
11.	1	1	1	0	4.53	6.31	5.31
12.	1	1	1	1	1.94	10.04	10.01

Results: subpopulations

GPC3 Antibody Treatment against Liver Cancer (J. Hepatology. 2016 Apr, Abou-Alfa et.al.)

- 180 patients: 60 took a placebo, 120 took the drug
- PFS: Progression Free Survival

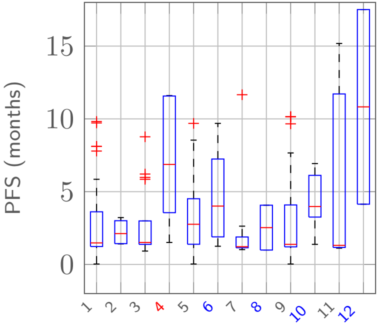
Sub-population	Drug Identifier	F1	F2	F3	Size (number of patients)	Mean PFS (months)	Median PFS (months)
1.	0	0	0	0	33.37	3.06	1.65
2.	0	0	1	0	4.07	2.29	2.24
3.	0	1	0	0	17.84	2.72	1.81
4.	0	1	1	0	4.72	7.05	7.18
5.	1	0	0	0	51.52	3.22	2.55
6.	1	0	0	1	16.77	4.17	3.65
7.	1	0	1	0	8.38	1.74	1.33
8.	1	0	1	1	2.07	2.69	2.65
9.	1	1	0	0	29.88	3.36	2.03
10.	1	1	0	1	4.90	4.44	4.34
11.	1	1	1	0	4.53	6.31	5.31
12.	1	1	1	1	1.94	10.04	10.01

Results: subpopulations

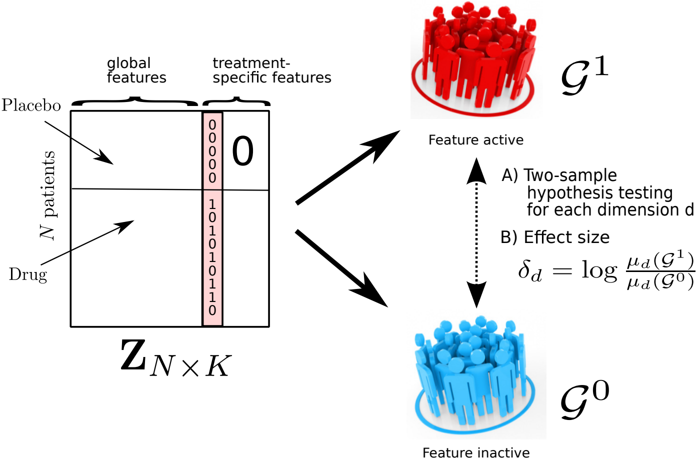
GPC3 Antibody Treatment against Liver Cancer (J. Hepatology. 2016 Apr, Abou-Alfa et.al.)

- 180 patients: 60 took a placebo, 120 took the drug
- PFS: Progression Free Survival

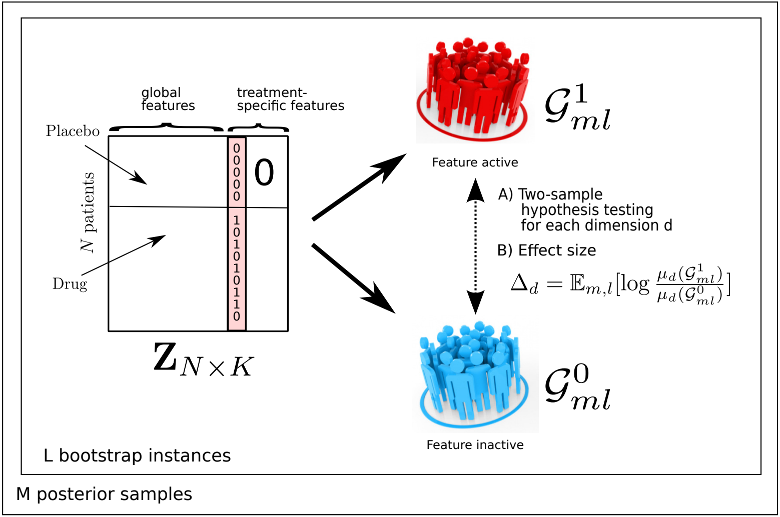
Sub-population	Drug Identifier	F1	F2	F3	Size (number of patients)	Mean PFS (months)	Median PFS (months)
1.	0	0	0	0	33.37	3.06	1.65
2.	0	0	1	0	4.07	2.29	2.24
3.	0	1	0	0	17.84	2.72	1.81
4.	0	1	1	0	4.72	7.05	7.18
5.	1	0	0	0	51.52	3.22	2.55
6.	1	0	0	1	16.77	4.17	3.65
7.	1	0	1	0	8.38	1.74	1.33
8.	1	0	1	1	2.07	2.69	2.65
9.	1	1	0	0	29.88	3.36	2.03
10.	1	1	0	1	4.90	4.44	4.34
11.	1	1	1	0	4.53	6.31	5.31
12.	1	1	1	1	1.94	10.04	10.01



Statistical procedure for biomarker discovery

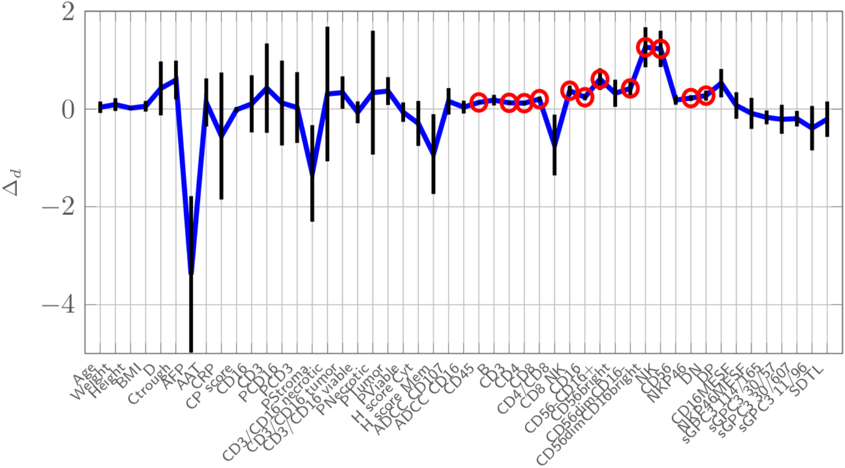


Statistical procedure for biomarker discovery



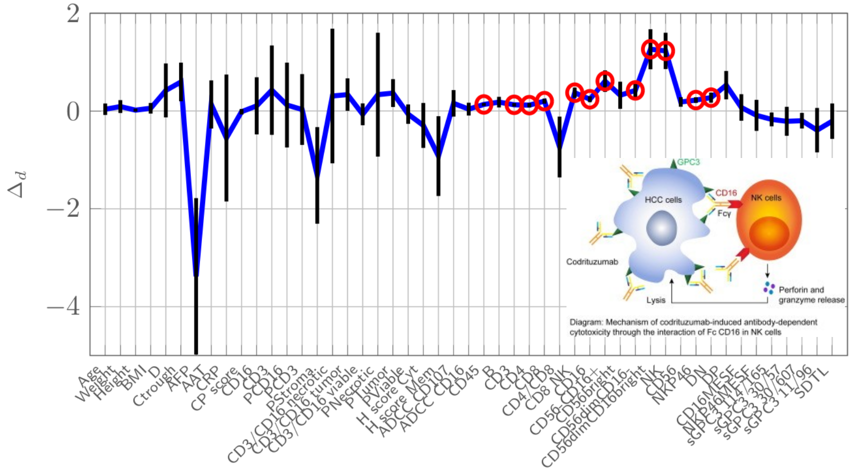
Results: biomarker discovery (Pradier et.al, 2018)

Treatment-specific feature F3



Results: biomarker discovery (Pradier et.al, 2018)

Treatment-specific feature F3



Other works using BNP models for data exploration

- Psyquiatric disorders (Rodriguez Ruiz et.al, 2014)
- Text analysis via topic models (Hughes et.al 2015)
- Economic complexity (Pradier et.al, 2018)

Software available

- General latent feature model:
  https://github.com/ivaleraM/GLFM
- Bayesian nonparametric for python:
  https://github.com/bnpy/bnpy

Conclusions

In this talk...

Bayesian non-parametrics

- useful for data exploration
  - Fair density estimation model
  - Structured latent feature model (global and group-specific factors)

Conclusions

In this talk...

Bayesian non-parametrics

- useful for data exploration
  - Fair density estimation model
  - Structured latent feature model (global and group-specific factors)

Sports science

- age-gender curves
- fair grading system
- running patterns over time

Conclusions

In this talk...

Bayesian non-parametrics

- useful for data exploration
  - Fair density estimation model
  - Structured latent feature model (global and group-specific factors)

Sports science

- age-gender curves
- fair grading system
- running patterns over time

Cancer research

- subpopulation learning
- biomarker discovery

Conclusions

Future work and discussion

1. Modeling
   - How to relax model assumptions?
   - How to incorporate prior knowledge?

Conclusions

Future work and discussion

1. Modeling
   - How to relax model assumptions?
   - How to incorporate prior knowledge?
2. Inference
   - scale algorithms (e.g., variational inference)
   - better exploration of posterior

Conclusions

Future work and discussion

1. Modeling
   - How to relax model assumptions?
   - How to incorporate prior knowledge?
2. Inference
   - scale algorithms (e.g., variational inference)
   - better exploration of posterior
3. Validation
   - new "data exploration" metrics
   - how to quantify model utility?

Acknowledgements

Special thanks to:

- Finale Doshi-Velez
- Weiwei Pan
- Michael Hughes
- All members of dtak!

- Francisco Rodriguez Ruiz
- Fernando Perez-Cruz

- Isabel Valera
- Maria Lomeli
- Zoubin Ghahramani

- Oscar Puig
- Francesca Milletti

CRCS Center for Research on Computation and Society
at Harvard John A. Paulson School of Engineering and Applied Sciences

HDSI | Harvard Data Science Initiative

Roche

MAX-PLANCK-GESELLSCHAFT

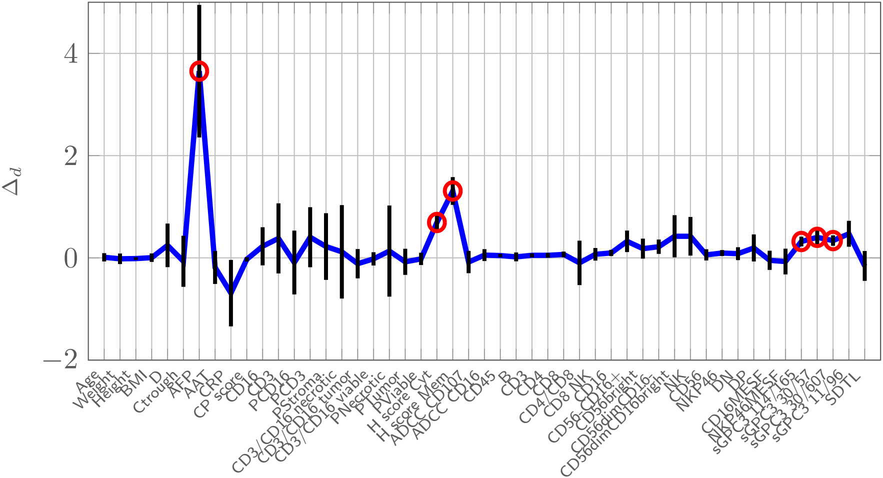
Thank you for listening!

Looking forward to your questions!

http://www.melaniefpradier.work

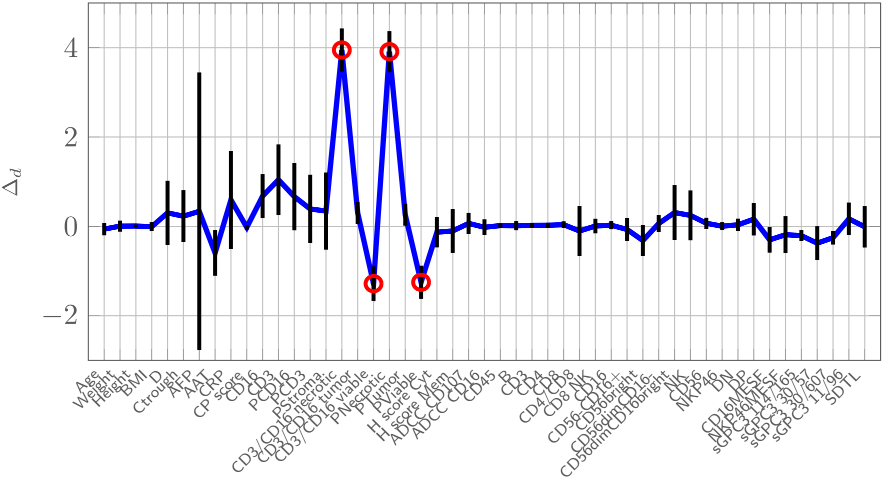
Results: biomarker discovery

Global feature F1



Results: biomarker discovery

Global feature F2



Indian buffet process (IBP)

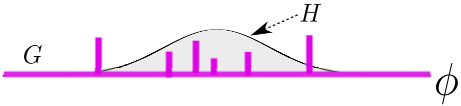
An alternative construction

- underlying block for infinite latent feature models

Indian buffet process (IBP)

An alternative construction

- underlying block for infinite latent feature models
- hierarchy of a Beta process (BP) with multiple Bernoulli processes (BeP)

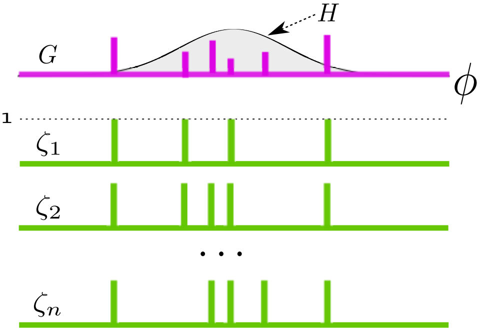


$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k} \sim \mathrm{BP}(c, \alpha, H)$$

Indian buffet process (IBP)

An alternative construction

- underlying block for infinite latent feature models
- hierarchy of a Beta process (BP) with multiple Bernoulli processes (BeP)



$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k} \sim \mathrm{BP}(c, \alpha, H)$$

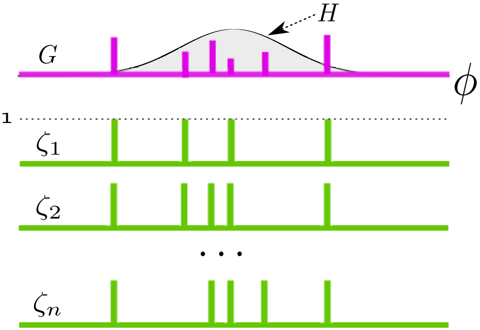
For $n = 1, \cdots, \infty$

$$\zeta_n = \sum_{k=1}^{\infty} z_{nk} \delta_{\phi_k} \sim \mathrm{BeP}(G)$$

Indian buffet process (IBP)

An alternative construction

- underlying block for infinite latent feature models
- hierarchy of a Beta process (BP) with multiple Bernoulli processes (BeP)



$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k} \sim \mathrm{BP}(c, \alpha, H)$$

For $n = 1, \cdots, \infty$

$$\zeta_n = \sum_{k=1}^{\infty} z_{nk} \delta_{\phi_k} \sim \mathrm{BeP}(G)$$

$$\mathbf{Z} \sim \mathrm{IBP}(\alpha)$$