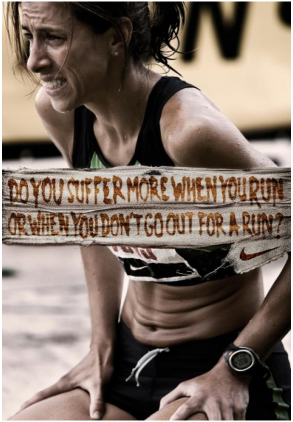
Bayesian Non-parametric Modeling for Marathon Age Grading

Melanie F. Pradier, Fernando Perez-Cruz

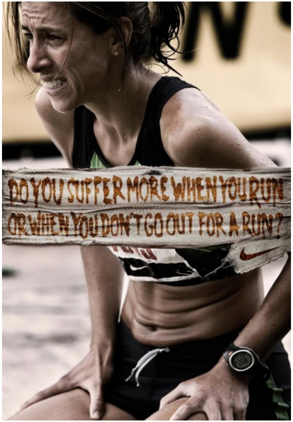
Universidad Carlos III of Madrid

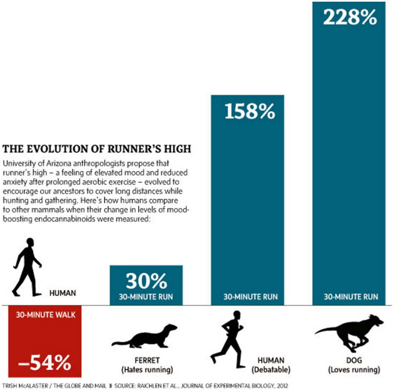
February 25, 2014

Interesting Fact: Runner's high



Interesting Fact: Runner's high





Outline

1. Problem Overview
2. Gentle Approach to Bayesian Non-Parametric (BNP)
3. A BNP model for Marathon data
4. Results

Outline

1. Problem Overview
2. Gentle Approach to Bayesian Non-Parametric (BNP)
3. A BNP model for Marathon data
4. Results

What is Age Grading?

Definition

System to normalize athlete's performance according to age and gender (since 1989)

$$score = \frac{your\ time}{world\ class\ time}$$

- Ill-defined metric, comparison against outlier
- Age Grading not really used in practice

What is Age Grading?

Definition

System to normalize athlete's performance according to age and gender (since 1989)

$$score = \frac{your\ time}{world\ class\ time}$$

- Ill-defined metric, comparison against outlier
- Age Grading not really used in practice

What is Age Grading?

Definition

System to normalize athlete's performance according to age and gender (since 1989)

$$score = \frac{your\ \ time}{world\ \ class\ \ time}$$

- Ill-defined metric, comparison against outlier
- Age Grading not really used in practice

What is Age Grading?

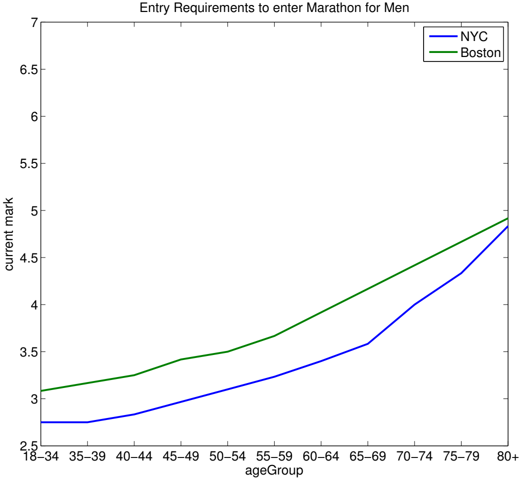
Definition

System to normalize athlete's performance according to age and gender (since 1989)

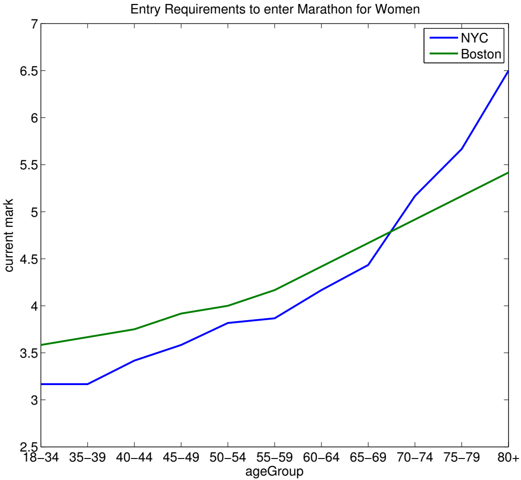
$$score = \frac{your\ time}{world\ class\ time}$$

- Ill-defined metric, comparison against outlier
- Age Grading not really used in practice

Entry Requirements



Entry Requirements



Objectives

Our Aim

- data exploration of marathon results
- propose a better Age Grading system to compare athletes
- analyze age impact on physical capabilities

How?

- BNP approach, generative model for $p(x)$
- Density Estimation using Gaussian Mixture Model

Objectives

Our Aim

- data exploration of marathon results
- propose a better Age Grading system to compare athletes
- analyze age impact on physical capabilities

How?

- BNP approach, generative model for $p(x)$
- Density Estimation using Gaussian Mixture Model

Objectives

Our Aim

- data exploration of marathon results
- propose a better Age Grading system to compare athletes
- analyze age impact on physical capabilities

How?

- BNP approach, generative model for $p(x)$
- Density Estimation using Gaussian Mixture Model

Outline

1. Problem Overview
2. Gentle Approach to Bayesian Non-Parametric (BNP)
3. A BNP model for Marathon data
4. Results

Bayesian Non-Parametric

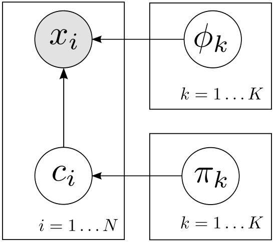
- Bayesian: Combine Prior Knowledge with Data Evidence

- Non-parametric
  - infinite number of parameters
  - complexity depends on input data
  - model selection avoided

Bayesian Non-Parametric

- Bayesian: Combine Prior Knowledge with Data Evidence
- Non-parametric
  - infinite number of parameters
  - complexity depends on input data
  - model selection avoided

Gaussian Mixture Model



π_k : mixture weights

ϕ_k : mixture parameters

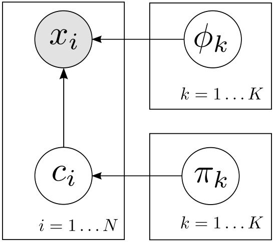
$$x_i | c_i, \pi_{1:K} \sim N\left(x_i | \mu_{c_i}, \Sigma_{c_i}\right)$$

$$\vdots$$

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

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

Gaussian Mixture Model



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

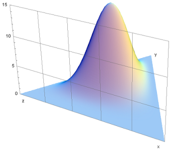
π_k : mixture weights

ϕ_k : mixture parameters

$$x_i | c_i, \pi_{1:K} \sim N(x_i | \mu_{c_i}, \Sigma_{c_i})$$

$$\vdots$$

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



Dirichlet Process

Dirichlet Process

- domain itself is a set of probability distributions

$$G \sim \mathrm{DP}\left(\mathrm{H}, \alpha\right)$$

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

$\mathrm{H}:$ base measure

$\alpha:$ concentration parameter

Dirichlet Process

Dirichlet Process

- domain itself is a set of probability distributions

$$G \sim \mathrm{DP}\left(\mathrm{H}, \alpha\right)$$

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

H : base measure

α : concentration parameter

Dirichlet Process

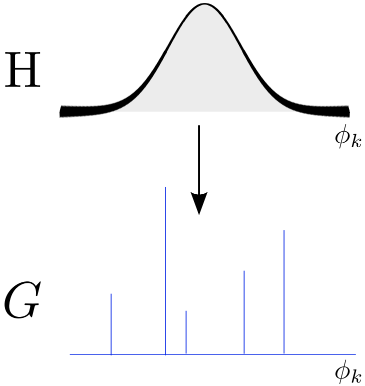
Dirichlet Process

- domain itself is a set of probability distributions

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

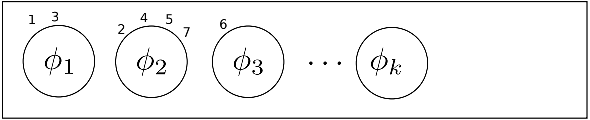
$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$

H : base measure
α : concentration parameter



Dirichlet Process

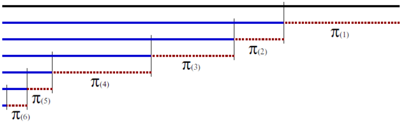
- Chinese Restaurant Process



- Stick Breaking Process

$$\pi_k = v_k \prod_{i=1}^{k-1} (1 - v_i)$$

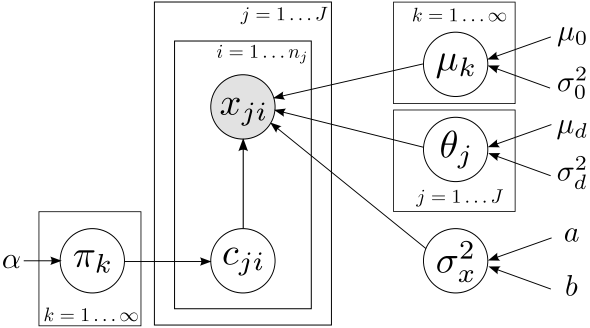
$$v_k \sim \mathrm{Beta}(1, \alpha)$$



Outline

1. Problem Overview
2. Gentle Approach to Bayesian Non-Parametric (BNP)
3. A BNP model for Marathon data
4. Results

Generative Model

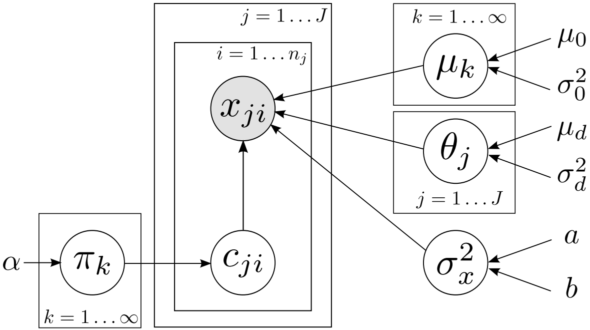


J: number of age groups

$$\begin{aligned} \mu_k &\sim N(\mu_0, \sigma_0^2) \\ \theta_j &\sim N(\mu_d, \sigma_d^2) \\ \sigma_x^2 &\sim \mathrm{IG}(a, b) \\ \pi_k &\sim \mathrm{GEM}(\alpha) \end{aligned}$$

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

Generative Model



J: number of age groups

$$\begin{array}{rcl} \mu_k & \sim & N\left(\mu_0, \sigma_0^2\right) \\ \theta_j & \sim & N\left(\mu_d, \sigma_d^2\right) \\ \sigma_x^2 & \sim & \mathrm{IG}\left(a, b\right) \\ \pi_k & \sim & \mathrm{GEM}\left(\alpha\right) \end{array}$$

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

Generative Model

- Model belongs to family of Dependent Dirichlet Process [MacEachern,2000]
- Prior for a set of mixture models (like the HDP)
- Weights π_k shared across groups
- Stick positions change: $\mu_{jk} = \mu_k + \theta_j$
- We call it: Atom-Dependent Dirichlet Process (ADDP)

Generative Model

- Model belongs to family of Dependent Dirichlet Process [MacEachern,2000]
- Prior for a set of mixture models (like the HDP)
- Weights π_k shared across groups
- Stick positions change: $\mu_{jk} = \mu_k + \theta_j$
- We call it: Atom-Dependent Dirichlet Process (ADDP)

Generative Model

- Model belongs to family of Dependent Dirichlet Process [MacEachern,2000]
- Prior for a set of mixture models (like the HDP)
- Weights π_k shared across groups
- Stick positions change: $\mu_{jk} = \mu_k + \theta_j$
- We call it: Atom-Dependent Dirichlet Process (ADDP)

Generative Model

- Model belongs to family of Dependent Dirichlet Process [MacEachern,2000]
- Prior for a set of mixture models (like the HDP)
- Weights π_k shared across groups
- Stick positions change: $\mu_{jk} = \mu_k + \theta_j$
- We call it: Atom-Dependent Dirichlet Process (ADDP)

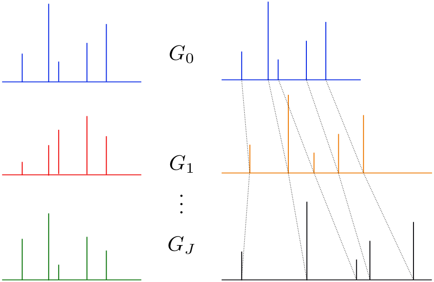
Generative Model

- Model belongs to family of Dependent Dirichlet Process [MacEachern,2000]
- Prior for a set of mixture models (like the HDP)
- Weights π_k shared across groups
- Stick positions change: $\mu_{jk} = \mu_k + \theta_j$
- We call it: Atom-Dependent Dirichlet Process (ADDP)

Generative Model

- Model belongs to family of Dependent Dirichlet Process [MacEachern,2000]
- Prior for a set of mixture models (like the HDP)
- Weights π_k shared across groups
- Stick positions change: $\mu_{jk} = \mu_k + \theta_j$
- We call it: Atom-Dependent Dirichlet Process (ADDP)

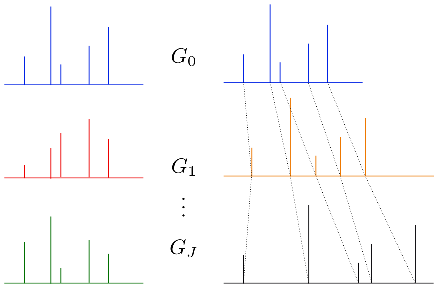
Comparison HDP Vs ADDP



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

Atom-Dependent DP
$G_0 \sim \mathrm{DP}(\alpha, \mathrm{H})$
$G_j = \mathrm{T}_j[G_0]$
$\mathrm{T}_j : [\mu_k] \to [\mu_k + \theta_j]$

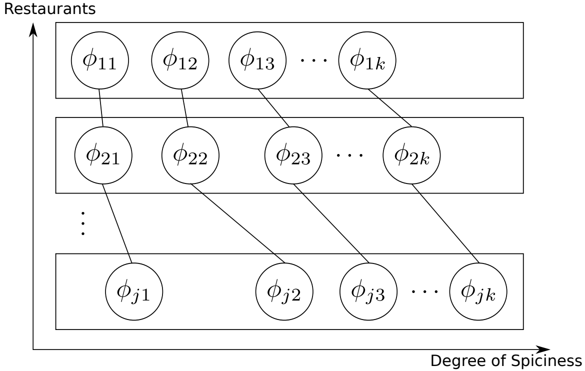
Comparison HDP Vs ADDP



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

Atom-Dependent DP
$G_0 \sim \mathrm{DP}(\alpha, \mathrm{H})$
$G_j = \mathrm{T}_j[G_0]$
$\mathrm{T}_j : [\mu_k] \rightarrow [\mu_k + \theta_j]$

Metaphor: Chinese Restaurant Franchise



Inference: Block Gibbs Sampling

- MCMC method: Gibbs Sampling to sample everything
- Slower convergence, but very fast computation

Inference: Block Gibbs Sampling

- MCMC method: Gibbs Sampling to sample everything
- Slower convergence, but very fast computation

Inference: Block Gibbs Sampling

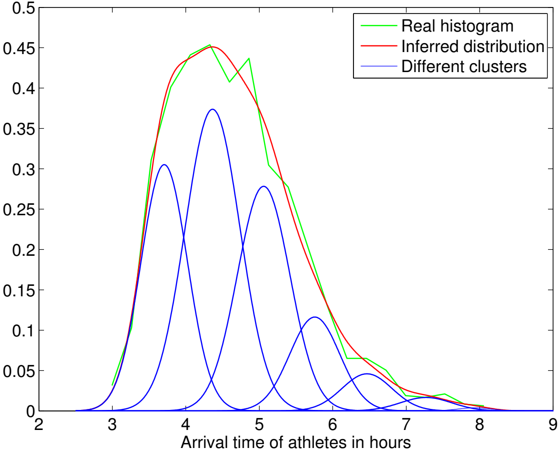
- MCMC method: Gibbs Sampling to sample everything
- Slower convergence, but very fast computation

N	10.000 iterations
47.095	15 min
249.899	1h05 min

Outline

1. Problem Overview
2. Gentle Approach to Bayesian Non-Parametric (BNP)
3. A BNP model for Marathon data
4. Results

Results
Model Fit



Results
Age Distribution per Cluster

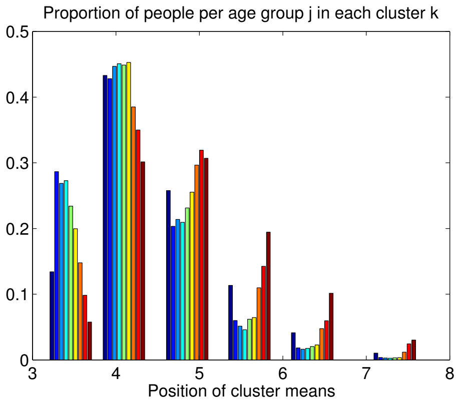


Figure : Dirichlet Process Prior

Results

Age Distribution per Cluster

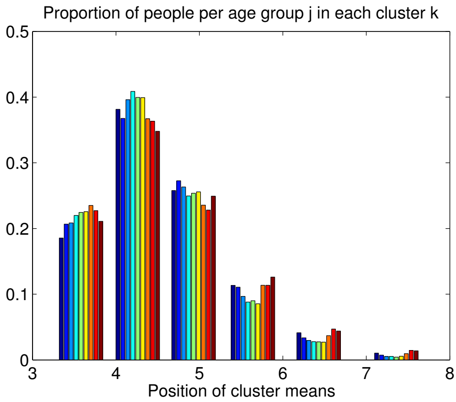
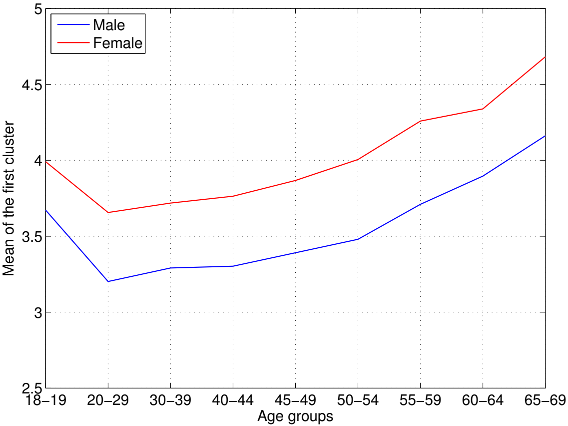


Figure : Atom Dependent Dirichlet Process Prior

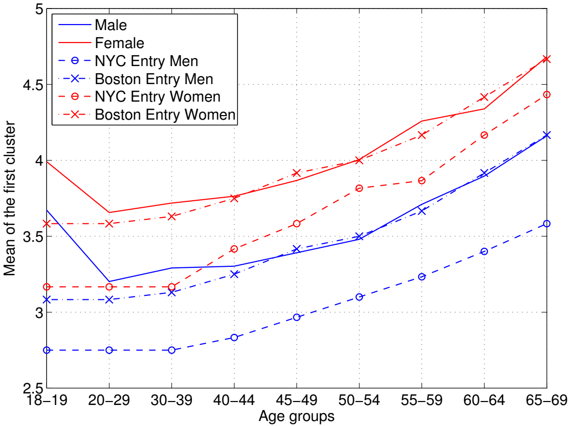
Basic Results

Age Grading Curves: θ_j



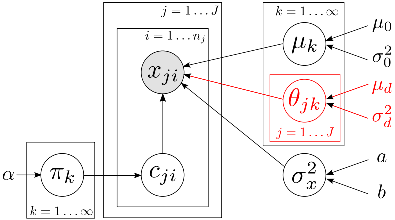
Basic Results

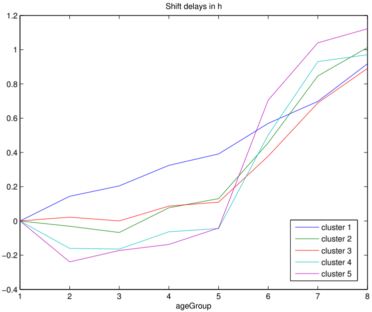
Age Grading Curves: θ_j



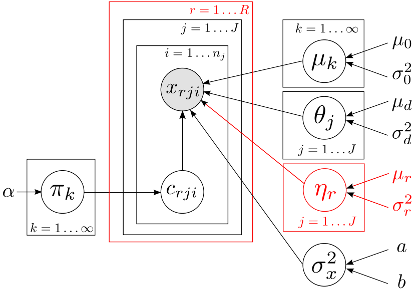
Model Improvements

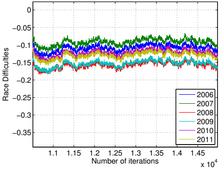
Age Delays θ_{jk} dependent on cluster k

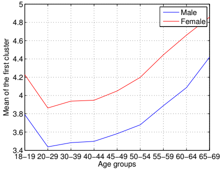




Model Improvements
Comparison of Multiple Races







Conclusion

Outlook

- Non-parametric model to compare different group distributions
- Modeling of the NYC Marathon
- Inference of robust age grading curves

Further research topics

1. Introduce correlation over age delays
2. Deal with temporal evolution: Inference of running patterns
3. Other applications
    - Pediatrics, Social Studies, Pharmaceutics...

Conclusion

Outlook

- Non-parametric model to compare different group distributions
- Modeling of the NYC Marathon
- Inference of robust age grading curves

Further research topics

1. Introduce correlation over age delays
2. Deal with temporal evolution: Inference of running patterns
3. Other applications
   - Pediatrics, Social Studies, Pharmaceutics...

Conclusion

Outlook

- Non-parametric model to compare different group distributions
- Modeling of the NYC Marathon
- Inference of robust age grading curves

Further research topics

1. Introduce correlation over age delays
2. Deal with temporal evolution: Inference of running patterns
3. Other applications
    - Pediatrics, Social Studies, Pharmaceutics...

Bibliography

- World Masters Athletics homepage, http://www.world-masters-athletics.org/about-us
- Bishop, Pattern Recognition and Machine Learning, 2006
- Whye Teh, et. al, Hierarchical Dirichlet Process, 2005
- MacEachern, Dependent Dirichlet Process, 2000
- Neal, Markov Chain Sampling Methods for Dirichlet Process Mixture Models, 2000

Thank you!

Looking forward to your questions...